GENOMICS-GUIDED PERSONALIZED CANCER TREATMENT

BY MEANS OF TUMOR TISSUE, LIQUID BIOPSY AND RADIOMICS ANALYSES

LINDSAY ANGUS

GENOMICS-GUIDED PERSONALIZED CANCER TREATMENT

BY MEANS OF TUMOR TISSUE, LIQUID BIOPSY AND RADIOMICS ANALYSES

LINDSAY ANGUS

Printing of this thesis was financially supported by: the department of Medical Oncology of the Erasmus MC Cancer Institute, Erasmus University Rotterdam, ChipSoft and Pfizer.

ISBN: 978-94-6419-288-9
Cover: Manouk Bos
Lay-out: Ilse Modder | www.ilsemodder.nl
Print: Gildeprint | www.gildeprint.nl

Genomics-guided Personalized Cancer Treatment
by means of tumor tissue, liquid biopsy and radiomics analyses

Behandeling op maat voor patiënten met kanker op basis van genomische analyses van tumorweefsel, “vloeibare biopten” en radiomics

Proefschrift
ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. F.A. van der Duijn Schouten

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 22 september 2021 om 13.00 uur

door

Lindsay Angus
geboren te ‘s-Gravenhage

Erasmus University Rotterdam

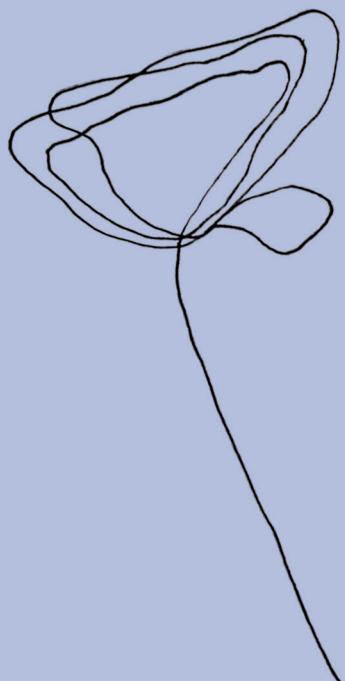
Promotiecommissie:

Promotoren: prof. dr. S. Sleijfer
prof. dr. J.W.M. Martens

Overige leden: prof. dr. W.N.M. Dinjens
prof. dr. E. van der Wall
prof. dr. S. Van Laere

Copromotor: dr. A. Jager

More than words is all you need to do to make it real (Extreme, 1991)



Contents

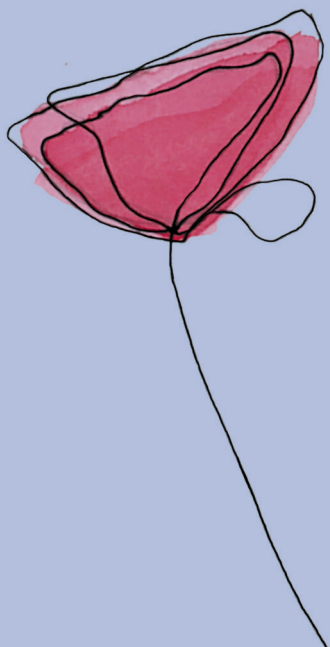
Chapter 1	General introduction and outline of the thesis	13
Part I: Tumor tissue biopsies		29
Chapter 2	The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies *Nature Genetics. 2019 Oct;51(10):1450-1458.*	31
Chapter 3	Whole genome sequencing of metastatic tissue reveals clinical and molecular variables associated with capecitabine response in patients with breast cancer *In preparation*	85
Chapter 4	*ESR1* mutations: moving towards guiding treatment decision-making in metastatic breast cancer patients *Cancer Treatment Reviews. 2017 Jan;52:33-40.*	109
Chapter 5	Genomic alterations associated with endocrine resistance in metastatic breast cancer have a differential impact on downstream ER signaling *Submitted*	131
Part II: Liquid Biopsies		149
Chapter 6	*RAS* and *BRAF* mutations in cell-free DNA are predictive for outcome of cetuximab monotherapy in patients with tissue-tested *RAS* wild-type advanced colorectal cancer *Molecular Oncology. 2019 Nov;13(11):2361-2374.*	151
Chapter 7	Whole exome sequencing of cell-free DNA - A systematic review and Bayesian individual patient data meta-analysis *Cancer Treatment Reviews. 2020 Feb;83:101951.*	179
Chapter 8	Novel methods to diagnose leptomeningeal metastases in breast cancer *Neuro Oncology. 2019 Mar 18;21(4):428-439.*	211
Chapter 9	Detection of aneuploidy in cerebrospinal fluid from patients with breast cancer can improve diagnosis of leptomeningeal metastases *Clinical Cancer Research. 2021 May 15;27(10):2798-2806.*	235

Chapter 10 Expression levels of TSPAN13, CD44, MAGEA3 and PLAU in circulating tumor cells are associated with brain metastases development in patients with metastatic breast cancer 265
Submitted

Part III: Radiomics 281
Chapter 11 The *BRAF* p.V600E mutation status of melanoma lung metastases cannot be discriminated on computed tomography by LIDC criteria nor radiomics using machine learning 283
Journal of Personalized Medicine. 2021 Apr 1;11(4):257.

Part IV: Integration of modalities 309
Chapter 12 Non-invasive estrogen receptor assessment by [^{18}F]-fluorestradiol (^{18}F-FES) PET/CT or circulating tumor cells predicts estrogen receptor status in patients with metastatic breast cancer 311
In preparation
Chapter 13 Summary and general discussion 331

Appendices Nederlandse samenvatting 356
Curriculum vitae 366
PhD portfolio 368
List of publications 372
Author Affiliations 376
Dankwoord 382



CHAPTER 1

General introduction and outline of the thesis

Genomics-guided personalized cancer treatment

During the past decades the treatment of patients with cancer has gradually shifted from "a one-size-fits-all" approach to a more personalized approach in which specific patient- and tumor characteristics are taken into account. Traditionally, patients with metastatic cancer were and still are treated depending on the tissue of origin, however, improved understanding of molecular mechanisms driving cancer progression and mechanisms underlying treatment resistance has increasingly led to trials stratifying based on molecular profiles rather than only on the anatomical origin of the primary tumor.

Sequencing efforts of tumor tissue have led to extensive characterization of mostly primary tumors[1-4]. For example for breast cancer, these efforts have identified a heterogeneous repertoire of disease drivers, evidence of clonal evolution and underlying mutational processes[2,5]. The analyses of the first cohort of whole genomes of primary breast cancer have revealed that the vast majority of aberrations are passenger events and it has been estimated that primary breast cancer on average "only" harbors four driver mutations per tumor[2]. Importantly, no two tumors seems to be genomically identical[6] and, although originating from the same organ, might need a completely different therapeutic approach.

Focusing on molecular characteristics, certain predictive and prognostic factors have already been incorporated into clinical care. Here, the molecular tests which are currently performed for metastatic breast- and colorectal cancer, and melanoma – tumor types described in this thesis – will be summarized.

Breast cancer

For metastatic breast cancer, predictive factors such as the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) have already been implemented in routine workup for many years[7,8] dividing breast cancer into four – clinically used – subtypes: hormone receptor (HR)-positive/HER2-negative, HR-positive/HER2-positive, HR-negative/HER2-positive and triple negative when both hormone receptors and the HER2 receptor are absent on tumor cells. Depending on breast cancer subtype, tumor load, presence of visceral disease and rate of disease progression patients are either treated with endocrine and/or HER2 directed therapy, or chemotherapy with or without HER2 blockage. Next to these established biomarkers, the OlympiAD trial showed that patients with HER2-negative metastatic breast cancer and germline mutations in *BRCA1* and *BRCA2* benefitted from treatment

with the oral poly (adenosine diphosphate-ribose) polymerase (PARP) inhibitor olaparib[9]. The median progression free survival was significantly longer in the olaparib arm than in the standard therapy arm receiving physician's choice (capecitabine, eribuline, or vinorelbine): 7.0 versus 4.2 months. Comparable results were observed for talazoparib versus physician's choice of chemotherapy[10]. Studies investigating the effectiveness of PARP-inhibitors in patients with non-BRCA metastatic breast cancer but mutations in other genes related to homologous recombination deficiency are ongoing (NCT02401347 and NOBROLA study (NCT03367689)).

In addition to germline *BRCA1/2* mutations, recently the predictive value of *PIK3CA* mutations has been shown for response to alpelisib – a PI3Kα-specific inhibitor[11]. Patients with HR-positive/HER2-negative disease harboring a hotspot mutation in *PIK3CA* (exons 7, 9, and 20) had a longer progression free survival when treated with alpelisib plus fulvestrant (an estrogen receptor degrader) versus placebo plus fulvestrant. Patients with wild type *PIK3CA* status did not seem to have clinically meaningful benefit from the combination alpelisib plus fulvestrant[11]. Likewise, testing the *PIK3CA* mutation status will probably enter the diagnostic arena soon and will be tested when treatment with alpelisib is considered.

Colorectal cancer

Molecular subtyping in patients diagnosed with colorectal cancer occurs in patients <70 years old to distinguish mismatch repair deficient from proficient tumors[12]. To this end, immunohistochemistry (IHC) of mismatch repair proteins, MLH1, PSM2, MSH2 and MSH6 or microsatellite instability (MSI) analysis via polymerase chain reaction is performed. Initially these analyses were performed to diagnose the hereditary nonpolyposis colorectal cancer syndrome, also referred to as Lynch syndrome[13]. However, more recently the presence of defective mismatch repair (dMMR)/MSI-high has proven to have important therapeutically implications as well. In patients with dMMR/MSI-high chemo refractory metastatic colorectal cancer, immune therapy with immune checkpoint inhibitors has shown an overall response rate of 40% with pembrolizumab monotherapy[14] and an objective response rate of 51% using nivolumab monotherapy[15] and up to 60% using the combination of nivolumab and ipilimumab[16]. Importantly, these responses were not observed in patients with MMR proficient/MSI-low tumors, emphasizing the need for upfront molecular characterization. Based on these results the Food and Drug Administration (FDA) granted accelerated approval for pembrolizumab, nivolumab and the combination of nivolumab/ipilimumab for the treatment of dMMR/MSI-high refractory colorectal cancer[15,17]. Next to evaluation of MSI-status, mutations in *KRAS, NRAS*[18] and *BRAF*[19,20] have been associated with

resistance to anti-epidermal growth factor receptor (EGFR) monoclonal antibodies such as cetuximab and panitumumab. Hence, the mutational status of these genes need to be routinely tested preceding prescription of EGFR-inhibtors[18].

Melanoma

Recent developments in the treatment of advanced metastatic melanoma, including immune checkpoint inhibitors and BRAF inhibitors, have significantly improved the perspectives of patients with this disease[21]. Approximately half of patients with metastatic melanoma harbor a mutation in the *BRAF* gene, with c.1799T>A (p.V600E) being the most common variant[22-24]. Patients harboring this mutation may significantly benefit from single agent therapy with BRAF inhibitors or combination treatment with BRAF and MEK inhibitors[25-28]. Hence, determination of the *BRAF* status in patients with advanced disease is mandatory according to ESMO guidelines[29].

Next to abovementioned targets for genomics-guided treatment, there is a swiftly growing list of targets for which targeted treatments are approved[30]. An overview of targets and matching treatments is shown in **Table 1**.

Although targeted treatment has been increasingly adopted in clinical care, there are still important challenges:

1. The molecular characteristics between primary tumors and metastatic lesions can substantially differ, but genomic analysis thus far have mainly been performed on primary tumor tissue and the discovery of targets for treatment is almost entirely limited to this setting;
2. Genomic profiles are not static but change over time and under treatment pressure, which requires collection of molecular profiles over time;
3. Molecular profiles are heterogeneous within lesions and between metastatic lesions, respectively intra-tumor and inter-tumor heterogeneity, requiring analysis of multiple tumor specimens per patient or body fluids likely representing tumor specific alterations from all metastatic sites (i.e., plasma-derived cell-free DNA);
4. The majority of currently used anti-cancer therapies is not administered based on tumor- or patient characteristics. For example, chemotherapy is still an important backbone in the treatment of cancer, but its administration is often not based on predictive factors. Identification of predictors will allow better patient stratification.

Since each tumor has its own unique genomic profile, which might have implications for treatment choices, more and more techniques – from sequencing of tumor tissue and liquid biopsies to analysis of big data using artificial intelligence – are being used to help

deliver the right drug at the right moment until the right moment to the right patient. This thesis aims to enhance the genomic understanding of metastatic disease, focusing on three common tumor types: breast cancer, colorectal cancer and melanoma. To this end, we used different techniques to obtain genomic information from tumor tissue, liquid biopsies and radiomics – all having its own pros and cons.

Table 1 – EMA and/or FDA approved drugs for genomics-guided cancer treatment

Gene/mechanism	Type of aberration	Tumor type	Drug(s)
BCR-ABL1	Fusion/mutation	Leukemia	Imatinib, Dasatinib, Nilotinib, Bosutinib, Ponatinib
ALK	Fusion/mutation	Lung	Crizotinib, Ceritinib, Alectinib, Lorlatinib, Brigatinib
BRAF	Mutation	Melanoma Lung, Thyroid	Vemurafenib, Dabrafenib, Encorafenib, Trametinib, Cobimetinib, Binimetinib Dabrafenib, Trametinib
BRCA1 and BRCA2	Germline/somatic mutation	Ovarian Breast	Olaparib, Rucaparib, Niraparib Olaparib, Talazoparib
EGFR	Germline mutation Mutation	Lung	Gefitinib, Erlotinib, Afatinib, Dacomitinib, Osimertinib
ERBB2	Amplification	Breast Gastric, gastroesophageal junction adenocarcinoma	Trastuzumab, Pertuzumab, Ado-trastuzumab emtansine, Lapatinib, Neratinib Trastuzumab
FGFR2 and FGFR3	Fusions/mutation	Urothelial	Erdafitinib
FLT3	Mutation	Leukemia	Midostaurin, Gilteritinib
IDH1	Mutation	Leukemia	Ivosidenib
IDH2	Mutation	Leukemia	Enasidenib
KIT	Mutation	GIST	Imatinib, Sunitinib, Regorafenib
KRAS, NRAS and BRAF	Wild-type	CRC	Cetuximab, Panitumumab
MET	Amplification/Exon 14 Skipping	Lung, Renal	Crizotinib, Cabozantinib
NTRK1, NTRK2 and NTRK3	Fusion	All solid tumors	Larotrectinib, Entrectinib
PDGFRA/PDGFB	Fusion	Leukemia	Imatinib, Dasatinib
PIK3CA	Mutation	Breast	Alpelisib
ROS1	Fusion	Lung	Crizotinib, Entrectinib
TSC1 and TSC2	Mutation	SEGA	Everolimus
Defective DNA mismatch repair	Defective MMR/ MSI-High	All solid tumors CRC	Pembrolizumab Pembrolizumab, Nivolumab

Abbreviations: CRC, colorectal cancer; GIST, gastrointestinal stromal tumor; SEGA, subependymal giant cell astrocytoma

Part I: Tumor tissue biopsies

In order to allow genomics-guided therapy it is of utmost importance to determine which genomic alterations are present in a tumor. Currently, the gold standard to determine which genomic alterations are present in a patient with metastatic disease is preferably to take a metastatic biopsy and perform molecular characterization on this prospectively collected material. Although guidelines[31,32] recommend to take biopsies from metastatic lesions, in daily clinical practice, treatment is still often based on the molecular characteristics of the primary tumor.

Currently, primary tumors of almost all tumor types have been characterized quite well at the genomic level[33,34]. However, patients do not die as a result of their primary tumor but due to metastatic disease. Hence, the urge to characterize metastatic cancer in detail has led to several biopsy protocols[35-37] of which the Center for Personalized Cancer Treatment Biopsy Protocol (CPCT-02; NCT01855477) represents the largest whole genome sequencing database available world-wide[38]. In the CPCT-02 study, patients with metastatic disease, regardless of tumor type, starting with a new line of systemic therapy were prospectively recruited. Fresh-frozen tissue biopsies and matched germline DNA were collected and analyzed by whole genome sequencing to reveal the genomic landscape of metastatic disease and its clinical significance.

To gain further insight into the genomic make-up of metastatic breast cancer, we analyzed whole sequencing data from the CPCT-02 cohort and described the genomic landscape of fresh frozen tissue biopsies of 442 patients with metastatic breast cancer. In **chapter 2** we compared various genomic features between our metastatic breast cancer cohort and previously described primary breast cancer cohorts: BASIS[2] and TCGA[4]. Hence, the frequency in which driver genes are affected, contribution of mutational signatures and the number of mutations per tumor were compared. Next, the available clinical data allowed to us to associate pre-treatment with the presence of mutational signatures and affected genes. Finally, the potential clinical implications of whole genome sequencing were evaluated.

In addition to the wealth of data whole genome sequencing provides to enhance the understanding of metastatic tumor biology, the ultimate goal of the CPCT-02 study[38] is to predict outcome to treatment based on the genomic profile of the tumor. As already mentioned above, many chemotherapeutics are prescribed without upfront selection. To this end, we selected a cohort of patients with metastatic breast cancer who all received capecitabine monotherapy, a fluoropyrimidine, directly after their

tumor biopsy. This allowed us to associate clinico-genomic features with response to capecitabine monotherapy and the results are described in **chapter 3**.

Since genomic profiles are not static but change over time and under treatment pressure, genomic analysis of tissue or liquid biopsies at disease progression will learn us more about resistance mechanisms. In breast cancer, the occurrence of mutations in the gene encoding for the estrogen receptor, *ESR1*, has sparked significant interest as mechanism for endocrine resistance in metastatic breast cancer. Functional studies have shown that activating mutations in the ligand-binding domain of *ESR1*[39], result in constitutive activity of ERα thereby conferring resistance to endocrine therapy. Here, the characterization of metastatic tissue and liquid biopsies has shown its importance, since these *ESR1* mutations are rarely present in primary breast cancer and are mainly detected in the metastatic setting in patients who received prior endocrine therapies[40,41]. **Chapter 4** reviews the methods used to identify *ESR1* mutations and gives an overview of the most relevant pre-clinical and clinical studies on *ESR1* mutations with respect to the occurrence of endocrine resistance.

In addition to *ESR1* mutations, other mechanisms of endocrine resistance have been elucidated recently. Razavi *et al.*[42] identified an increased number of alterations in genes involved in the mitogen-activated protein kinase (MAPK) pathway and estrogen receptor transcriptional machinery in patients previously exposed to endocrine therapy. Alterations in the MAPK pathway and transcriptional regulators were mutually exclusive with *ESR1* mutations and are associated with shorter durations of response on subsequent lines of endocrine therapies. To what extent these three mechanisms of resistance (activating *ESR1* mutations, MAPK-pathway activation or activation of transcriptional regulation) impacts downstream gene expression of ER and other genes is unknown. In **chapter 5** we integrated whole genome sequencing with RNA sequencing to explore the effect of these different resistance mechanisms with downstream gene expression.

Part II: Liquid Biopsies

Although tissue biopsies provide ground for extensive molecular analyses, taking tissue biopsies is a cumbersome procedure and repeated sampling is often not feasible. Next to that, tumor cells are plastic and change over time and under treatment pressure. Therefore, genomic profiling of plasma derived cell-free DNA (cfDNA) or circulating tumor cells (CTCs) are considered as real-time minimally-invasive surrogates for tumor tissue analysis. Both analytes[43,44] have shown to yield prognostic value[45,46] and are also

being recognized as tools to monitor recurrence[47-49], resistance and treatment effect[50,51] in many tumor types.

CfDNA is thought to be derived from apoptotic cells consisting of short fragments (~160-180 base pairs) derived from both normal- and tumor cells[52], whereas most CTCs are intact tumor cells representing the genomic information of single tumor cells. Hence, CTCs and ctDNA reflect distinct and complementary information. Since CTCs are intact cells, this enables, beyond DNA analysis, characterization at the RNA and protein level either using pools of cells[53,54] or single cells. In addition, once CTCs are enriched from blood, functional assays can be performed as well[55]. Both CTCs and cfDNA can be used for characterization at the DNA level, providing information on mutations, epigenetic and copy number alterations and fusion genes. Although CTC analysis allows for a more comprehensive genomic analysis, the detection and isolation of CTCs requires special enrichment strategies such as the FDA-approved EpCAM based enrichment method: the CellSearch® system[56]. Yet, the major advantage of cfDNA over CTCs is that its isolation does not require any special enrichment methods and therefore might be more clinically applicable.

Predictive value of liquid biopsies

Currently, there are three FDA-approved assays for analysis of cfDNA. These include two tests with predictive value. First, the cobas® *EGFR* mutation test v2 detecting *EGFR* exon 19 deletions and hotspot mutations in *EGFR* p.L858R and p.T790M, the first two alterations predicting response to erlotinib and the latter response to osimertinib in non-small cell lung cancer[57]. Second, the therascreen® *PIK3CA* polymerase chain reaction kit detecting hotspot mutations in exons 7, 9 and 20 of *PIK3CA*, predictive for response to alpelisib[11]. The third FDA-approved cfDNA test is Epi proColon test for the detection of methylated *SEPT9*, a gene associated with the presence of colorectal cancer[58].

As abovementioned, the list of targets for treatment is not limited to drugs targeting alterations in *EGFR* and *PIK3CA* and thus cfDNA testing for other indications might be of clinical relevance as well. For example, in patients with metastatic colorectal cancer, the *RAS* mutation status is routinely tested in tumor tissue before starting treatment with anti-EGFR monoclonal antibodies. However, despite tissue testing, still only 40-45% of patients with *RAS* wild-type metastatic colorectal cancer will respond to this treatment[59,60]. A potential explanation for non-response in wild-type patients might be the presence of intra-tumor and inter-tumor heterogeneity, also known as spatial heterogeneity, of the *RAS* mutation status. Tumor heterogeneity could lead to *RAS* and *BRAF* mutated subclones, which are missed due to either being under the detection

limit of the assay or not being present at all in the evaluated part of the tumor. Since the analysis of cfDNA might be more representative of the entire mutational burden within a patient, we studied the presence of *RAS* and *BRAF* mutations in cfDNA of patients starting with third line cetuximab monotherapy. In **chapter 6** we evaluated the genetic profiles in cfDNA of 34 patients with *RAS* wild type metastatic colorectal cancer, before the start of treatment, using a next generation sequencing approach covering 14 genes, including *KRAS, NRAS, EGFR* and *PIK3CA*. The aim of this study was to assess if cfDNA analysis could improve upfront patient selection for anti-EGFR monoclonal therapy. Next, the collection of cfDNA samples at disease progression supported the analysis of underlying resistance mechanisms.

Feasibility of whole exome sequencing on plasma-derived cfDNA

Notwithstanding the value of detection of mutations using targeted gene panels of single gene assays, a more comprehensive view on the tumor genome will lead to better insight of tumor biology, including the genomic mechanisms by which tumor cells can confer resistance. Sequencing techniques covering the exome or entire genome provide a view on the complex landscape of somatic alterations. This also enables the identification of mutational signatures, pinpointing microsatellite instability and tumor mutational burden, of which the latter two are recognized as predictive biomarkers for response to immune checkpoint inhibitors. So, compared to targeted panels comprising a relatively limited number of genes, whole exome sequencing (WES) of ctDNA holds great promise to identify emerging genes that are of interest in treatment resistance and to capture DNA signatures important for treatment decision making. However, WES of cfDNA is technically challenging due to the often low tumor fractions in a high background of normal cfDNA and the detection of these low fractions is hampered by the error-rate of currently used sequencing methods. So, the aim of **chapter 7** was to evaluate to what extent WES of cfDNA in cancer patients is technically feasible by performing a systematic review of the literature. In addition, a meta-analysis on the sensitivity of WES-detected single nucleotide variants in cfDNA using tumor tissue as reference as well as the agreement between cfDNA and tumor tissue was performed.

Liquid biopsies to enhance detection of leptomeningeal metastasis and prediction of developing brain metastasis

Although, plasma-derived cfDNA is the most extensively studied source to analyze tumor-derived DNA, all body fluids such as urine, ascites or cerebrospinal fluid (CSF) can serve as a liquid biopsy. CSF has shown to contain genomic information from primary brain tumors[61-63] as well as metastases[61,63] in the central nervous system.

In patients with breast cancer, leptomeningeal metastasis (LM) is rare but is often accompanied with devastating neurological symptoms and has a very poor prognosis. One of the explanations of the dismal prognosis is the delay in diagnosing LM. Currently, two diagnostic methods are used to detect LM: gadolinium MRI having a sensitivity of 53 - 80% and CSF cytology having a sensitivity of 45 - 75%[64-67]. To improve the detection of LM, a plethora of biomarkers has been investigated. In **chapter 8**, we describe the current diagnostic work-up to diagnose LM, give an overview of novel techniques to diagnose LM and provide requirements for new biomarker to be able to reach clinical use.

Using these criteria to the markers described in **chapter 8**, two tumor specific candidates to detected LM at an earlier stage were identified: identification and enumeration of tumor cells present in CSF by EpCAM-based assays and the detection of CSF cfDNA.

Henceforth in **chapter 9**, we performed copy number analysis of cfDNA isolated from CSF in a large unique cohort of breast cancer patients who underwent a lumbar puncture for the clinical suspicion of LM. In this retrospective study, copy number variation in CSF cfDNA was measured by the "Fast Aneuploidy Screening Test-Sequencing System" (FAST-SeqS)[68-70]. The aim of this study was to assess aneuploidy in CSF cfDNA of breast cancer patients suspected of LM, and correlate the cfDNA analysis to clinical parameters such as neurological signs and symptoms and routinely performed diagnostic tests such as CSF chemistry, CSF cytology and imaging.

In addition to diagnosing patients with LM at an earlier stage, it would be of clinical relevance to predict which patients have the highest chance to develop CNS metastasis. Identification of patients at high risk might lead to preventive approaches as is the case for small cell lung cancer patients who receive prophylactic cranial irradiation or could lead to prescription of treatments with a higher penetration through the blood brain barrier. However, one of the major challenges is the availability of human samples available for molecular profiling during the course of the disease which can be used to predict those patients at risk. In a cohort of patients with metastatic breast cancer, who were starting with either first line endocrine or first line chemotherapy, CTCs were enriched for molecular profiling. We hypothesized that patients who developed brain metastasis might harbour CTCs with a different gene expression profile than patients who did not develop these metastases. In **chapter 10**, we compared the gene expression profiles measured in CTCs between patients who did develop brain metastasis with patients who did not.

Part III: Radiomics

In addition to tissue- and liquid biopsies as means to characterize cancer at the genomic level, radiomics is recognized as a non-invasive computational analysis of features captured in radiological images[71]. From a region of interest, multiple features (i.e., shape and texture) can be extracted. Subsequently, these features can be used in a model which can be used to predict molecular profiles and subtypes. Radiomic imaging biomarkers have proven to be effective in the prediction of the molecular status of several tumor types[72-74].

For melanoma, the *BRAF*-mutation status is determined by molecular analysis of tumor tissue, often using primary tumor material or a metastatic tissue biopsy. Next to the invasive nature of taking biopsies, it can take several days before the *BRAF*-mutation status becomes available by next generation sequencing of the primary tumor or metastasis. However, immediate treatment with BRAF/MEK inhibitors is required in patients with advanced disease and life-threatening symptoms. Consequently, a faster and less invasive diagnostic method might improve patient management. Therefore, in **chapter 11** we used a radiomics approach to non-invasively differentiate between *BRAF*-mutant and wild type lung metastases of patients with metastatic melanoma.

Part IV: Integration of modalities

As tissue, liquid biopsies and imaging analyses all represent distinct information, the ultimate goal is to integrate information of all diagnostic methods to optimize prognostication and prediction of treatment effects. As of yet, most studies on genomics only included one or two of these modalities in their study design. The "Imaging patients for cancer drug selection – metastatic breast cancer" (IMPACT-MBC) (NCT01832051) study evaluates the clinical utility of experimental PET-scans. In this multicenter prospective study patients with newly diagnosed metastatic breast cancer, PET-scans evaluating the estrogen receptor (^{18}F-FES-PET) and HER2 receptor (^{89}Zr-Trastuzumab-PET) are performed and combined with conventional imaging modalities as CT imaging, bone scintigraphy and FDG-PET. Next to focusing on imaging, CTC enumeration and characterization is performed, as well as tissue analyses of the primary tumor and metastatic biopsies. This approach enables the integrative analysis of information obtained from different modalities to predict therapy response. Also, the collection of multimodal information provides the opportunity to compare different read-outs of the ER status. In **chapter 12,** we investigated the concordance for the assessment of the ER

status measured by routine IHC of a metastatic lesion, by *ESR1* gene expression analysis in CTCs and by imaging using ^{18}F-FES-PET.

Finally, **chapter 13** discusses the pros and cons of using tumor tissue, liquid biopsies and/or radiomics for genomic characterization of solid tumors. Furthermore, it provides a summary of the thesis and reflects my view on future research in this field.

References

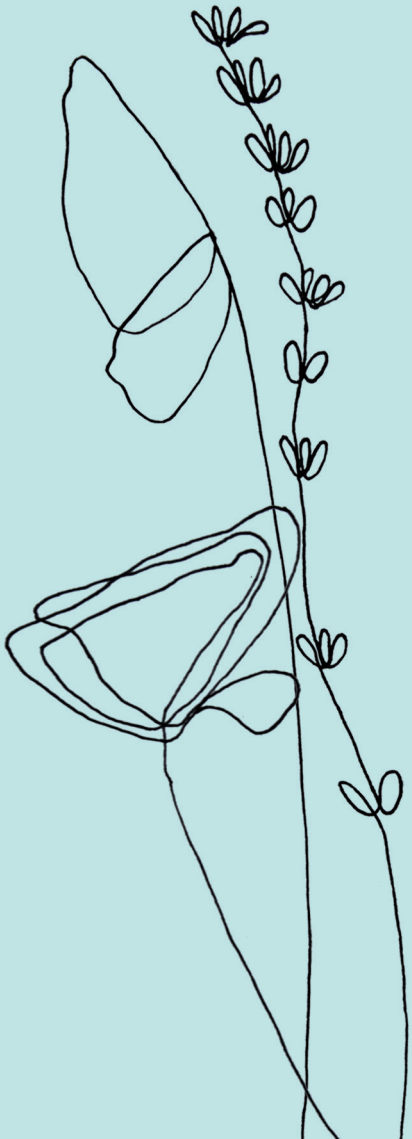
1. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93 (2012).
2. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
3. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-7 (2012).
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
5. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
6. McGranahan, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15-26 (2015).
7. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J Clin Oncol* **14**, 2843-77 (1996).
8. Harris, L. *et al.* American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* **25**, 5287-312 (2007).
9. Robson, M. *et al.* Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation. *N Engl J Med* **377**, 523-533 (2017).
10. Litton, J.K. *et al.* Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *N Engl J Med* **379**, 753-763 (2018).
11. Andre, F. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* **380**, 1929-1940 (2019).
12. Stjepanovic, N. *et al.* Hereditary gastrointestinal cancers: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **30**, 1558-1571 (2019).
13. Lynch, H.T. & de la Chapelle, A. Hereditary Colorectal Cancer. *N Engl J Med* **348**, 919-932 (2003).
14. Le, D.T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509-2520 (2015).
15. Overman, M.J. *et al.* Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol* **18**, 1182-1191 (2017).
16. Lenz, H.J.J. *et al.* Durable clinical benefit with nivolumab (NIVO) plus low-dose ipilimumab (IPI) as first-line therapy in microsatellite instability-high/mismatch repair deficient (MSI-H/dMMR) metastatic colorectal cancer (mCRC). *Ann Oncol* **29**(2018).
17. Marcus, L., Lemery, S.J., Keegan, P. & Pazdur, R. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. *Clin Cancer Res* **25**, 3753-3758 (2019).
18. Sorich, M.J. *et al.* Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann Oncol* **26**, 13-21 (2015).
19. Pietrantonio, F. *et al.* Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: a meta-analysis. *Eur J Cancer* **51**, 587-94 (2015).
20. Rowland, A. *et al.* Meta-analysis of BRAF mutation as a predictive biomarker of benefit from anti-EGFR monoclonal antibody therapy for RAS wild-type metastatic colorectal cancer. *Br J Cancer* **112**, 1888-94 (2015).
21. Luke, J.J., Flaherty, K.T., Ribas, A. & Long, G.V. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* **14**, 463-482 (2017).
22. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-54 (2002).
23. Curtin, J.A. *et al.* Distinct sets of genetic alterations in melanoma. *N Engl J Med* **353**, 2135-47 (2005).
24. Colombino, M. *et al.* BRAF/NRAS mutation frequencies among primary tumors and metastases in patients with melanoma. *J Clin Oncol* **30**, 2522-9 (2012).
25. Long, G.V. *et al.* Dabrafenib plus trametinib versus dabrafenib monotherapy in patients with metastatic BRAF V600E/K-mutant melanoma: long-term survival and safety analysis of a phase 3 study. *Ann Oncol* **28**, 1631-1639 (2017).

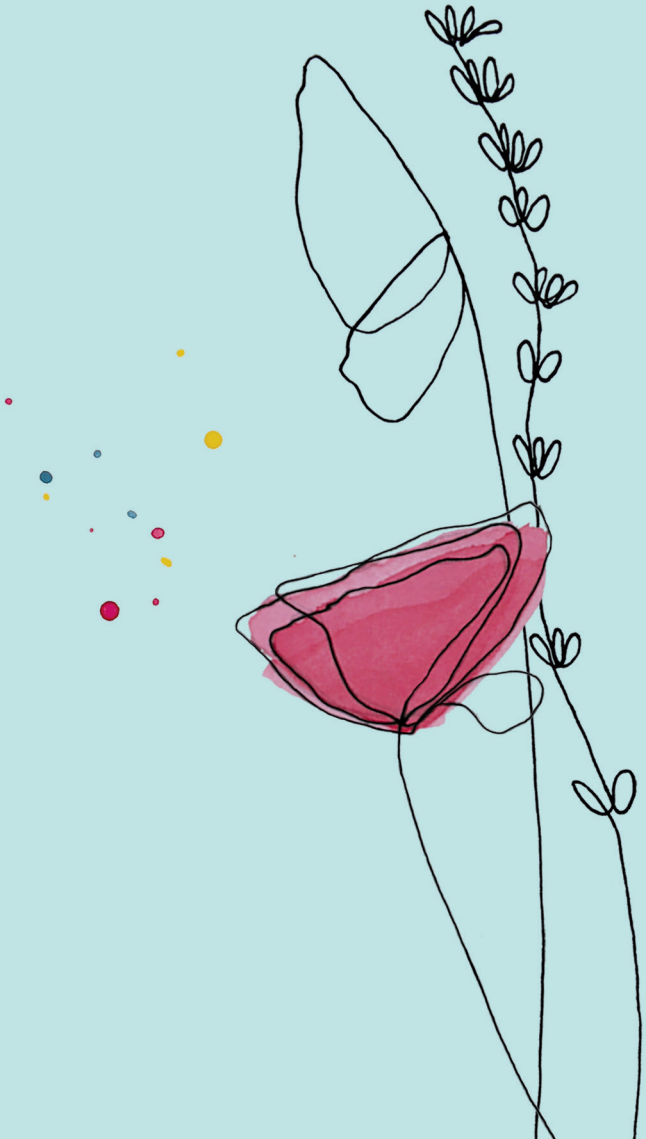
26. Long, G.V. *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med* **371**, 1877-88 (2014).
27. Larkin, J. *et al.* Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med* **371**, 1867-76 (2014).
28. Ascierto, P.A. *et al.* Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol* **17**, 1248-60 (2016).
29. Dummer, R. *et al.* Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **26 Suppl 5**, v126-32 (2015).
30. Marquart, J., Chen, E.Y. & Prasad, V. Estimation of the Percentage of US Patients With Cancer Who Benefit From Genome-Driven Oncology. *JAMA Oncol* **4**, 1093-1098 (2018).
31. Cardoso, F. *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)dagger. *Ann Oncol* **29**, 1634-1657 (2018).
32. Van Poznak, C. *et al.* Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol* **33**, 2695-704 (2015).
33. International Cancer Genome Consortium *et al.* International network of cancer genome projects. *Nature* **464**, 993-8 (2010).
34. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-20 (2013).
35. Andre, F. *et al.* Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/UNICANCER). *Lancet Oncol* **15**, 267-74 (2014).
36. Massard, C. *et al.* High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. *Cancer Discov* **7**, 586-595 (2017).
37. Le Tourneau, C. *et al.* Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol* **16**, 1324-34 (2015).
38. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* (2019).
39. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45 (2013).
40. Robinson, D.R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51 (2013).
41. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* **73**, 6856-64 (2013).
42. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438 e6 (2018).
43. Heitzer, E. Circulating Tumor DNA for Modern Cancer Management. *Clin Chem* (2019).
44. Pierga, J.Y. Circulating Tumor Cells for Cancer Management. *Clin Chem* (2019).
45. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* **351**, 781-91 (2004).
46. Cohen, S.J. *et al.* Relationship of circulating tumor cells to tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J Clin Oncol* **26**, 3213-21 (2008).
47. Sparano, J. *et al.* Association of Circulating Tumor Cells With Late Recurrence of Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncol* **4**, 1700-1706 (2018).
48. Tie, J. *et al.* Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med* **8**, 346ra92 (2016).
49. Tie, J. *et al.* Circulating Tumor DNA Analyses as Markers of Recurrence Risk and Benefit of Adjuvant Therapy for Stage III Colon Cancer. *JAMA Oncol* (2019).
50. Dawson, S.J., Rosenfeld, N. & Caldas, C. Circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **369**, 93-4 (2013).
51. Bidard, F.C. *et al.* Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled

analysis of individual patient data. *Lancet Oncol* **15**, 406-14 (2014).
52. Diaz, L.A., Jr. & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579-86 (2014).
53. Onstenk, W. *et al.* Efficacy of Cabazitaxel in Castration-resistant Prostate Cancer Is Independent of the Presence of AR-V7 in Circulating Tumor Cells. *Eur Urol* **68**, 939-45 (2015).
54. Beije, N. *et al.* Prognostic Impact of HER2 and ER Status of Circulating Tumor Cells in Metastatic Breast Cancer Patients with a HER2-Negative Primary Tumor. *Neoplasia* **18**, 647-653 (2016).
55. Pantel, K. & Alix-Panabieres, C. Functional Studies on Viable Circulating Tumor Cells. *Clin Chem* **62**, 328-34 (2016).
56. Beije, N., Jager, A. & Sleijfer, S. Circulating tumor cell enumeration by the CellSearch system: the clinician's guide to breast cancer treatment? *Cancer Treat Rev* **41**, 144-50 (2015).
57. cobas EGFR Mutation Test v2. 2016.
58. Song, L., Jia, J., Peng, X., Xiao, W. & Li, Y. The performance of the SEPT9 gene methylation assay and a comparison with other CRC screening tests: A meta-analysis. *Sci Rep* **7**, 3032 (2017).
59. Karapetis, C.S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**, 1757-65 (2008).
60. Lievre, A. *et al.* KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol* **26**, 374-9 (2008).
61. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, 8839 (2015).
62. Mattox, A.K., Yan, H. & Bettegowda, C. The potential of cerebrospinal fluid-based liquid biopsy approaches in CNS tumors. *Neuro Oncol* **21**, 1509-1518 (2019).
63. Pentsova, E.I. *et al.* Evaluating Cancer of the Central Nervous System Through Next-Generation Sequencing of Cerebrospinal Fluid. *J Clin Oncol* **34**, 2404-15 (2016).
64. Olson, M.E., Chernik, N.L. & Posner, J.B. Infiltration of the leptomeninges by systemic cancer. A clinical and pathologic study. *Arch Neurol* **30**, 122-37 (1974).
65. Glass, J.P., Melamed, M., Chernik, N.L. & Posner, J.B. Malignant cells in cerebrospinal fluid (CSF): the meaning of a positive CSF cytology. *Neurology* **29**, 1369-75 (1979).
66. Balm, M. & Hammack, J. Leptomeningeal carcinomatosis. Presenting features and prognostic factors. *Arch Neurol* **53**, 626-32 (1996).
67. Wasserstrom, W.R., Glass, J.P. & Posner, J.B. Diagnosis and treatment of leptomeningeal metastases from solid tumors: experience with 90 patients. *Cancer* **49**, 759-72 (1982).
68. Belic, J. *et al.* Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clin Chem* **61**, 838-49 (2015).
69. Belic, J. *et al.* mFast-SeqS as a Monitoring and Pre-screening Tool for Tumor-Specific Aneuploidy in Plasma DNA. *Adv Exp Med Biol* **924**, 147-155 (2016).
70. Suppan, C. *et al.* Untargeted Assessment of Tumor Fractions in Plasma for Monitoring and Prognostication from Metastatic Breast Cancer Patients Undergoing Systemic Treatment. *Cancers (Basel)* **11**(2019).
71. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* **48**, 441-6 (2012).
72. Vos, M. *et al.* Radiomics approach to distinguish between well differentiated liposarcomas and lipomas on MRI. *Br J Surg* **106**, 1800-1809 (2019).
73. Rios Velazquez, E. *et al.* Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* **77**, 3922-3930 (2017).
74. Grimm, L.J. Breast MRI radiogenomics: Current status and research implications. *J Magn Reson Imaging* **43**, 1269-78 (2016).

PART I

Tumor tissue biopsies





CHAPTER 2

The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies

Nature Genetics. *2019 Oct;51(10):1450-1458*

Lindsay Angus, Marcel Smid, Saskia M. Wilting, Job van Riet, Arne Van Hoeck, Luan Nguyen, Serena Nik-Zainal, Tessa G. Steenbruggen, Vivianne C.G. Tjan-Heijnen, Mariette Labots, Johanna M.G.H. van Riel, Haiko J. Bloemendal, Neeltje Steeghs, Martijn P. Lolkema, Emile E. Voest, Harmen J.G. van de Werken, Agnes Jager, Edwin Cuppen, Stefan Sleijfer, John W.M. Martens

Abstract

The whole-genome sequencing of prospectively collected tissue biopsies from 442 patients with metastatic breast cancer reveals that, compared to primary breast cancer, tumor mutational burden doubles, the relative contributions of mutational signatures shift and the mutation frequency of six known driver genes increases in metastatic breast cancer. Significant associations with pre-treatment are also observed. The contribution of mutational signature 17 is significantly enriched in patients pretreated with fluorouracil, taxanes, platinum and/or eribulin, whereas the de novo mutational signature I identified in this study is significantly associated with pretreatment containing platinum-based chemotherapy. Clinically relevant subgroups of tumors are identified, exhibiting either homologous recombination deficiency (13%), high tumor mutational burden (11%) or specific alterations (24%) linked to sensitivity to FDA-approved drugs. This study provides insights into the biology of metastatic breast cancer and identifies clinically useful genomic features for future improvement of patient management.

Editor's Summary

Whole-genome sequencing of metastatic biopsies from 442 patients with breast cancer provides insight intro metastatic disease, including associations of genomic features with prior treatments and identification of therapeutic vulnerabilities.

Introduction

Breast cancer is the most common malignancy among women worldwide[1]. In-depth analyses of primary breast cancer have provided clear evidence of clonal evolution and have resulted in the identification of a heterogeneous repertoire of nearly 100 disease-causing genes and passenger events, both resulting from various underlying mutational processes[2-6] including age-related deterioration, homologous recombination deficiency (HRD)[7] and apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) mutagenesis[8,9].

However, patients do not die from their primary breast tumor but as a consequence of metastases. Due to tumor evolution and treatment pressure, the genomic alterations in metastatic breast cancer can differ substantially from the primary tumor[10-15]. Therefore, thorough genomic characterization of metastases will yield valuable insights into the active molecular processes in metastatic disease. This is crucial to understand the effects of systemic treatment on the tumor genome and ultimately improve the treatment of patients with metastatic breast cancer.

To date, in-depth analyses of metastatic breast cancer lesions are limited to studies using either whole exome sequencing[16,17] in relatively small cohorts or targeted sequencing of cancer-associated genes in a larger cohort[18]. These studies have suggested that metastatic breast cancer largely carries the same drivers seen in primary breast cancer; however, they also show clear differences in the numbers and types of genes that are affected.

To obtain an unbiased and complete picture of the genomic landscape of metastatic breast cancer and its underlying mutational processes, as reflected by mutational signatures, we performed whole-genome sequencing (WGS) on a large multicenter, prospective collection of snap-frozen metastatic tissue biopsies from 442 patients with breast cancer starting a new line of systemic treatment. These data enabled us to investigate the potential for patient-specific patterns of aberration that can be used to inform treatment choices. We performed an in-depth characterization of the genomic landscape of these patients with metastatic breast cancer and report on the presence of genomic alterations, and mutational and rearrangement signatures compared to a well-characterized cohort of primary breast cancer (the BASIS cohort)[6]. The available clinical data allowed us to associate genomic features with clinical information such as prior treatment. Finally, we identified subgroups of patients with specific and targetable genomic features who might be eligible for established or experimental therapies.

Results

The metastatic biopsies and matched germline DNA (peripheral blood) of 625 patients with metastatic breast cancer were analyzed (**Fig. 1a**). Patients with metastatic breast cancer who were biopsied in their primary tumor (*n* = 55) were excluded from the metastatic analyses, but were used as an additional control group. Metastatic biopsy sites mainly included liver, lymph nodes, bone and soft tissue (**Fig. 1b).** Twenty-two percent of all metastatic biopsies could not be evaluated, while lesions obtained from bone metastases had a failure rate of 33% (**Supplementary Table 1**). Breast cancer subtype distribution did not differ between nonevaluable and evaluable biopsies. Metastatic tumor biopsies and paired normal of the remaining 442 patients were sequenced at a median read coverage of 107 (interquartile range (IQR) = 98–114) and 38 (IQR = 35–42), respectively.

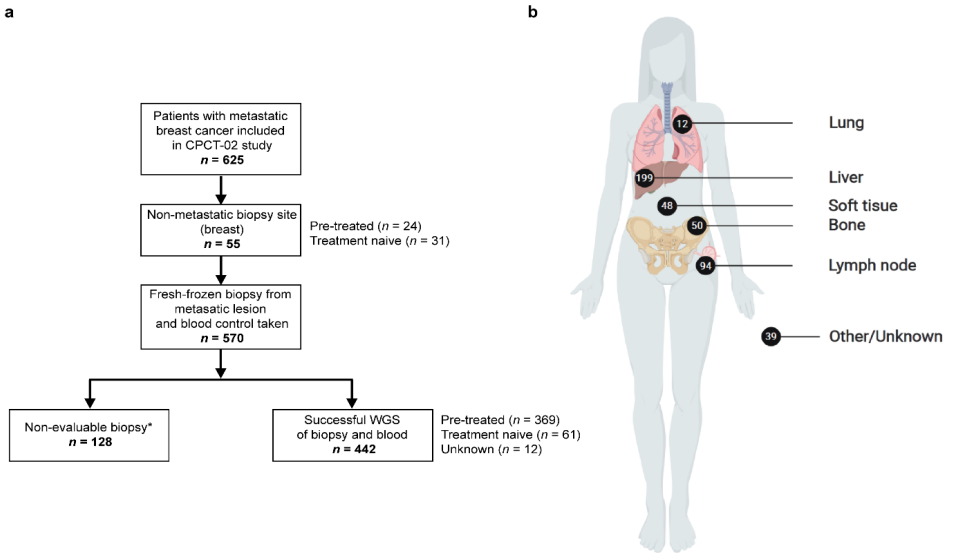


Figure 1 - Overview of study design and biopsy sites (*n* = 442).
a, Flowchart of patient inclusion. Patients with metastatic breast cancer were selected from the CPCT-02 cohort (*n* = 442). Patients were excluded if the only available biopsy was from the primary lesion. The asterisk indicates biopsies that could not be evaluated defined as no biopsy taken, <30% tumor cells or DNA yield too low for WGS.
b, Overview of biopsy sites. The number of biopsies per metastatic site analyzed with WGS are shown. Credit: Created with BioRender (https://biorender.com/).

The somatic landscape of metastatic breast cancer differs from primary breast cancer

Metastatic lesions showed a median of 7,661 single-nucleotide variants (SNVs; IQR = 4,607–14,417), 57 multiple-nucleotide variants (MNVs; IQR = 32–106), 689 small insertions and deletions (indels; IQR = 443–1,084), and 214 structural variants (SVs; IQR = 99–392 (**Supplementary Fig. 1**). Estrogen receptor $(ER)^-$ tumors had a 1.6 fold higher SV count than ER^+ tumors (95% confidence interval (CI) 1.3–2.0, $P < 0.001$) and, human epidermal growth factor receptor 2 $(HER2)^+$ tumors had higher SV counts than $HER2^-$ cases ($P = 0.013$).

Compared to WGS from 560 primary breast cancer samples (BASIS cohort)[6], the median numbers of SNVs, indels and SVs were significantly higher in metastatic breast cancer: 3,491 SNVs/MNVs (IQR = 2,075–6,911; 2.2x; 95% CI 1.9–2.4; $P < 1 \times 10^{-5}$), 204 indels (IQR = 133–365; 3.3x; 95% CI 3.0–3.6; $P < 1 \times 10^{-5}$), and 85 SVs (IQR = 25–208; 2.4x; 95% CI 2.1–2.8; $P < 1 \times 10^{-5}$). Consequently, the median tumor mutational burden (TMB) of 2.97 per million base pairs (Mbp) (IQR = 1.84–5.44) in metastatic breast cancer was significantly higher than that observed in the BASIS primary breast cancer cohort (**Supplementary Table 2**) (1.29 Mbp^{-1}; IQR = 0.78–2.56; 2.2x 95% CI 2.0–2.5, $P < 1 \times 10^{-5}$). In line with our findings, another cohort of patients with metastatic breast cancer (**Supplementary Table 2**) also reported an elevated median TMB of 3.19 Mbp^{-1} [17]. In our metastatic breast cancer cohort, we did not observe differences in median TMB between breast cancer subtypes or biopsy sites (**Supplementary Fig. 2**).

To ensure that the higher TMB we observed in our metastatic breast cancer cohort compared to primary disease was not due to methodological differences (**Supplementary Table 2**), we used the data of the 55 patients in our cohort who were biopsied in their primary tumor (**Fig. 1a**), including 31 patients who were treatment-naïve (group 1) and 24 patients who were pretreated (group 2). We compared the TMB of these primary tumors with the TMB of the metastatic biopsies of 61 patients who were treatment-naïve (group 3) and 369 patients who were pretreated (group 4) (**Supplementary Fig. 3**). In a multivariate linear regression model using these four groups, both type of tissue (metastatic/primary) and pretreatment (yes/no) were associated with TMB ($P < 1 \times 10^{-5}$ for the model; the estimated coefficients were 0.3212 ($P = 0.02$) and 0.3664 ($P = 0.001$), respectively). After stratifying for ER status, both pretreatment (0.4404, $P = 0.014$) and type of tissue (0.5208, $P = 0.0003$) were associated with TMB in ER^+ cases but not in ER^- cases. However, low numbers (only 8 pretreated primary ER^- tumors) make the interpretation of the results of this regression inconclusive. This suggests that, next to the disease course, treatment pressure is a major contributor to TMB.

Mutational signatures are associated with pre-treatment

To investigate which mutational processes operate in metastatic breast cancer and to what extent pretreatment is associated with the resulting mutational patterns, we applied the mathematical approach proposed by Alexandrov *et al.*[2] to categorize mutational signatures. De novo signature calling revealed 10 signatures in metastatic breast cancer, all of which could be mapped back to the already known Catalogue of Somatic Mutations in Cancer (COSMIC) signatures (cosine similarities ranging from 0.79 to 0.99; **Fig. 2a** and **Supplementary Fig. 4**). Except for de novo signatures I and J, all identified signatures have been previously described in primary breast cancer.

De novo signature J was very similar to COSMIC mutational signature 7, which is probably due to ultraviolet (UV) exposure. Detailed evaluation showed that the algorithm only identified this signature in one patient with a very high contribution (>98%), suggesting that this liver biopsy, containing mostly UV-induced DNA damage, had been misclassified as metastatic breast cancer.

De novo signature I (221 patients with >10% contribution, 27 patients with >25% contribution) was very similar to COSMIC mutational signatures 4 and 8 (**Fig. 2b**) and was more frequently observed in patients pretreated with platinum-based chemotherapy (P = 0.001) in our cohort. Signature 4 has been associated with tobacco mutagens[15] and is characterized by C>A substitutions and CC>AA dinucleotide substitutions[5]. The etiology of signature 8 is still unknown but its presence has been observed in primary breast cancer and was recently linked to BRCA deficiency[7]. This signature also shows C>A substitutions and has the CC>AA characteristic. Cisplatin mainly forms Pt-d(GpG) di-adducts[16] and patients pretreated with platinum-based chemotherapy showed higher levels of CC>AA substitutions (1.8x; 95% CI 1.2–2.5; P = 0.0013) than patients who did not receive platinum treatment. Also, patients with at least a 10% contribution of de novo signature I had higher levels of CC>AA (2x; 95% CI 1.6–2.2; $P < 1 \times 10^{-5}$; **Fig. 2c**), but patients with at least a 10% contribution of signature 8 did not have elevated CC>AA levels (P = 0.706). A previously published cisplatin signature[19] with characteristic C>T conversion peaks, which are absent in de novo signature I, had a higher cosine similarity in patients pretreated with platinum-based chemotherapy than in patients who did not receive this pretreatment (1.2x; 95% CI 1.1–1.3; P = 0.0008; **Fig. 2d**). Furthermore, when samples were dichotomized in two groups based on their similarity to the cisplatin signature identified by Boot *et al.*,[19] (permutations $P < 0.05$ and $P > 0.05$, respectively), 23 out of 27 samples with at least 25% contribution of de novo signature I had a high similarity to this signature (2.6x; 95% CI 2.4–3.0; $P < 1 \times 10^{-5}$; **Fig. 2e**). Next, since de novo signature I resembles

signature 8, which in turn is linked to BRCA deficiency[7], we analyzed germline BRCA mutation status in this context. A multivariate regression model showed that both germline *BRCA* mutation status and pretreatment with platinum-containing drugs were significantly associated with the relative contribution of de novo signature I (P <1 x 10^{-5} for the model; estimated coefficients for germline *BRCA* mutation status 10.36 (P = 5.4 x 10^{-7}) and 4.13 (P = 0.0014) for pretreatment with platinum).

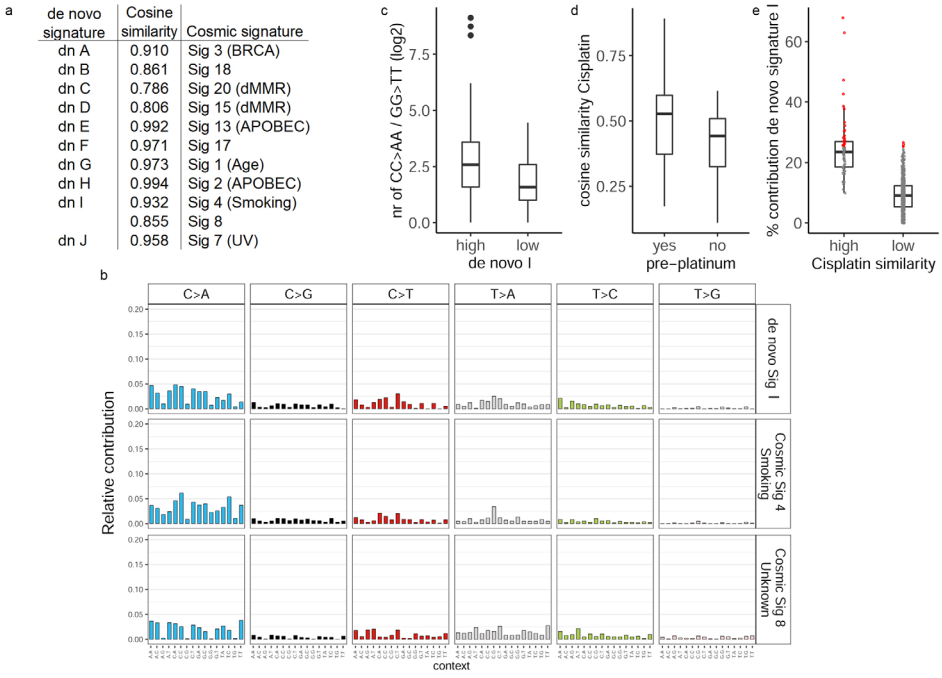


Figure 2 - De novo signature I is associated with prior platinum-based chemotherapy.

a, De novo signature calling revealed ten mutational processes involved in metastatic breast cancer. These de novo mutational signatures have high cosine similarities with known COSMIC signatures. dMMR, defective mismatch repair.

b, The mutational spectrum of de novo signature I and COSMIC signatures 4 and 8.

c, The number of CC>AA or GG>TT mutations in patients with a low (<10%, n = 221) or high (≥10%, n = 221) relative contribution of de novo signature I.

d, Box plot of the cosine similarity of the cisplatin signature defined by Boot *et al.*[19] and samples of patients who did (n = 43) or did not (n = 385) receive prior treatment with platinum-based chemotherapy. **e**, Box plot of the contribution of mutational signature I and samples with a high (permutation P < 0.05, n = 59) or low (permutation P > 0.05, n = 383) similarity to the cisplatin signature identified by Boot *et al.* [19]. The red dots indicate samples with >25% contribution of de novo signature I.

c-e, The box is bounded by the 25th and 75th percentiles, with the horizontal line in the box depicting the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentiles.

Since the observed de novo signatures largely overlapped known COSMIC signatures, we also determined the contributions of these 30 known signatures to the mutational landscape of our cohort (**Supplementary Figs. 5 and 6**). Out of the 30 COSMIC signatures, 12 contributed to ≥10% of the observed mutations in at least 5 patients; therefore, they were defined as dominant signatures (**Fig. 3**). The most frequently represented signatures in metastatic breast cancer were signature 8 (64% of patients), signature 1 (59%), which is related to aging; signature 2 (43%) and 13 (36%), which are related to APOBEC mutagenesis, and signature 3 (41%), which is associated with HRD. Analyses according to breast cancer subtype revealed that signatures 3 and 9 mutations were significantly more often present (2.7x; 95% CI 1.9–3.9 and 1.3x; 95% CI 1.1–1.6, respectively) in ER$^-$ compared to ER$^+$ metastatic breast cancer, whereas signature 2 (APOBEC) mutations were significantly more frequent (2.1x; 95% CI 1.5–2.9) in ER$^+$ metastatic breast cancer (all $P < 0.05$).

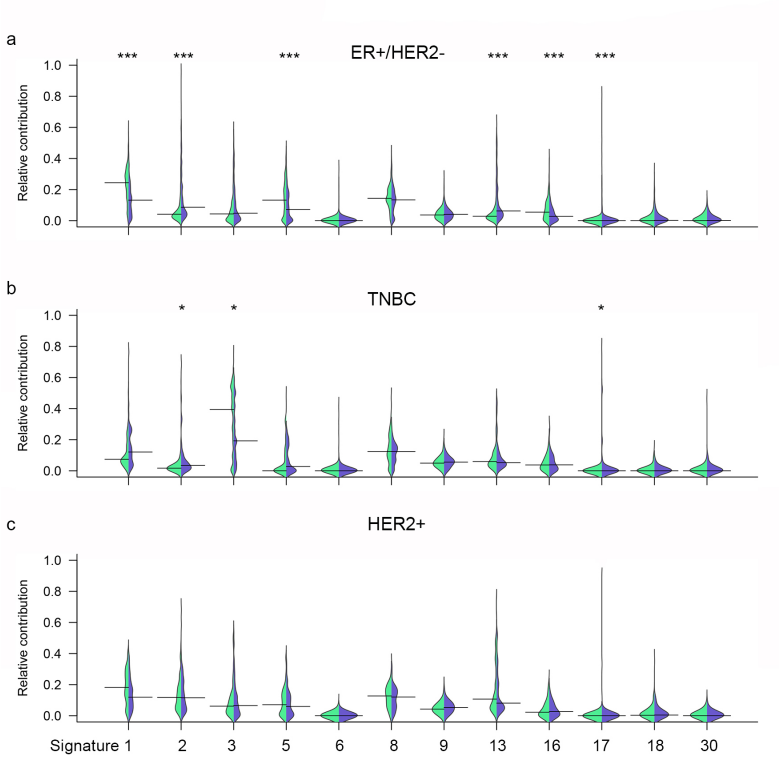


Figure 3 - Mutational signatures: metastatic breast cancer versus primary breast cancer.
a-c, Bean plots showing the relative contribution of 12 COSMIC signatures that dominantly contribute to the total number of SNVs in the metastatic cohort. Relative contributions between metastatic breast cancer and primary breast cancer samples from the BASIS cohort were compared and are shown per breast cancer subtype: ER$^+$/HER2$^-$ (BASIS, n = 320; CPCT, n = 279) (**a**), TNBC (BASIS, n = 167; CPCT, n = 58) (**b**), HER2$^+$ (BASIS, n = 73; CPCT, n = 77) (**c**). Left of center (green) indicates the distribution of primary tumors from

the BASIS cohort, whereas right of center (purple) indicates metastatic biopsy. The width of the bean plot depicts the density of the observations in each group; the horizontal line shows the median. The length of the bean plot shows the full range of observations. Mann–Whitney *U* (two-sided, Benjamini–Hochberg-corrected): * $P < 0.05$, *** $P < 0.001$. Exact Benjamini–Hochberg-corrected *P* values: TNBC signature 2, 0.03; signature 3, 0.03; signature 17, 0.02; $ER^+/HER2^-$: signature1, 2.40 $x10^{-20}$; signature 2, 5.79 $x10^{-12}$; signature 5, 2.63 $x10^{-5}$; signature13, 1.64 $x10^{-10}$; signature 16, 2.75 $x10^{-6}$; signature 17, 2.24 x 10^{-10}.

Ten of the 12 COSMIC signatures detected in our metastatic breast cancer cohort were previously described in the BASIS primary breast cancer cohort[6] (signatures 1–3, 5, 6, 8, 13, 17, 18 and 30), whereas signature 9 (8%) and 16 (14%) were not reported in the BASIS cohort[6]. After reevaluation of the previously published BASIS data[6], we found that the latter two signatures were actually present, but at relatively low levels (median relative contribution <5%). Subsequently, we compared the absolute and relative contributions of the 12 dominant COSMIC signatures between our metastatic breast cancer cohort and the BASIS primary breast cancer cohort per breast cancer subtype (**Fig. 3**; see **Supplementary Fig. 5** for all relative comparisons of the 30 COSMIC signatures). Irrespective of breast cancer subtype, the median absolute number of mutations was higher in metastatic breast cancer compared to primary breast cancer for almost all signatures (**Supplementary Fig. 6**), reflecting the significantly higher TMB in metastatic breast cancer and ongoing mutagenic processes.

On a relative scale we found a decrease in signatures 1 and 5 (age) and signature 16 (reported in liver cancer), as well as an increase in signatures 2 and 13 (APOBEC), and signature 17 (unknown etiology) in $ER^+/HER2^-$ metastatic disease compared to $ER^+/HER2^-$ primary breast cancer from the BASIS cohort. In triple negative breast cancer (TNBC), a decrease in signature 3 (HRD) and an increase in signatures 2 and signature 17 was seen in metastatic lesions compared to primary breast cancer. In patients with $HER2^+$ disease, no differences in the relative contributions of the 12 dominant signatures were found between primary and metastatic disease (**Fig. 3**), irrespective of taking ER status into account.

To determine whether these differences between primary breast cancer and metastatic breast cancer were driven by disease course or pretreatment, we performed a multivariate linear regression analysis using the previously defined four groups of primary and metastatic lesions with and without pretreatment. This showed a significantly lower (signature 1) and higher (signature 17) contribution in patients who were pretreated, irrespective of disease course. Thus, pretreatment in itself – regardless of treatment type – causes a limited shift in certain signature patterns.

Within metastatic breast cancer lesions, we also investigated the potential role of specific pretreatments on the relative signature contributions for all 12 dominant COSMIC signatures as defined earlier. Pretreatment with fluorouracil (5-FU), taxanes, platinum-containing chemotherapy and/or eribulin was associated with significantly higher relative contributions of signature 17 (all false discovery rate (FDR) *P* values < 0.05, with 5-FU the most significant at an FDR $P = 2.0 \times 10^{-9}$) (**Supplementary Fig. 7**). These treatments have been given to 40, 58, 10, and 3% of patients, respectively. The large overlap in patients who were pretreated with all these therapies hampered further specification of which of these therapies is directly associated with signature 17. Although signature 17 is present in primary breast cancer due to endogenous processes, the fact that signature 17 is mainly characterized by T>G and T>C in a CTT context might implicate 5-FU, inhibiting thymidylate synthase and thus synthesis of thymidine[20], as a probable drug contributing to this pattern. Finally, we investigated the association between the mutational signatures and response to the line of therapy that was initiated directly after sampling tumor material. Patients with progression at first response evaluation after twelve weeks of treatment had a significantly higher relative contribution of signature 17 ($P = 0.0012$). However, we also observed that the number of pretreatments given was higher in patients with ≥10% signature 17 contribution, making it hard to distinguish whether or not signature 17 is truly a biomarker for poor response to therapy or a marker of poor outcome in general.

In conclusion, virtually all mutational processes present in primary breast cancer contribute to the observed increased TMB in metastatic breast cancer. On a relative scale, we observed a shift from more indolent age-related mutagenesis in primary disease towards more APOBEC-driven processes in metastatic breast cancer. Additionally, previously given lines of therapy can impose specific mutational profiles in breast cancer cells.

Structural variation and homologous recombination deficiency

To evaluate structural variation in metastatic lesions we extracted the six rearrangement signatures described previously[6]. Rearrangement signatures 1 and 3 (SV1, SV3) were the least prevalent (both 6% of all rearrangements) in metastatic lesions while SV2, SV4, SV6 contributed 20, 14 and 19%, respectively. SV5 was most dominant and contributed to 36% of all rearrangements. Compared to primary breast cancer, the relative contribution of SV3, related to *BRCA1* gene abrogation, was significantly decreased (2.9x; 95% CI 1.5–7.1; $P < 1 \times 10^{-5}$) while *BRCA2*-related SV5 increased (3.2x; 95% CI 2.7–3.8, $P < 1 \times 10^{-5}$) in metastatic lesions regardless of breast cancer subtype (**Supplementary Fig. 8**).

To investigate the presence of an HRD phenotype based on somatic alterations, we applied the recently developed Classifier of Homologous recombination Deficiency (CHORD) (Nguyen, van Hoeck and Cuppen, manuscript in preparation). This algorithm predicts HRD and assigns the *BRCA* gene most probably responsible based on a combination of rearrangement signatures (SV1, SV3 and SV5), a specific type of indels flanked by microhomology and mutational signature 3. In our cohort of 442 patients, 18 had a germline loss of *BRCA1* or *2* (*BRCA1*, *n* = 5; *BRCA2*, *n* = 13). CHORD identified 39 additional patients carrying a HRD tumor next to all 18 germline *BRCA*-mutation carriers.

Unsupervised clustering reveals eight distinct genomic clusters in metastatic breast cancer

Based on the genomic characteristics of our metastatic breast cancer cohort comprising 442 metastatic lesions, we performed an unsupervised clustering analysis, which revealed 8 clusters representing tumors with distinct genomic phenotypes (**Fig. 4**). Biopsy site and treatment outcome were evenly distributed among the eight clusters. Clusters A and B were both characterized by mutational signature 3. Cluster A was further characterized by short tandem duplications and by SV3; cluster B was characterized by large deletions and SV5. In addition, these two clusters are enriched for HRD ($P < 1 \times 10^{-5}$) as predicted by the CHORD algorithm. In cluster A, HRD was predicted to be based on *BRCA1* deficiency; in cluster B, it was predicted to be based on *BRCA2* deficiency. However, clusters A and B also contained one and four patients, respectively, who were predicted to be homologous recombination-proficient. In these patients, we checked for mutated genes that are known in HR (as described in the Methods); however, none of these genes were homozygously affected.

Clusters C, D and E were characterized by mutational signatures 17, 18 and 16, respectively. Cluster F was mainly based on insertions. Cluster G showed a low TMB, few SVs and a relatively high proportion of mutational signature 5. Finally, cluster H represented tumors predominantly harboring mutational signatures 2 and 13 related to APOBEC mutagenesis, a relative high TMB and kataegic events. Kataegis was observed in 177 (40%) patients with metastatic breast cancer (ranging from 1 to 144 events), with 15 patients exhibiting 10 or more foci. In kataegic foci, mainly APOBEC mediated mutagenesis occurred ($P < 0.001$; **Supplementary Fig. 9**). Patients exhibiting kataegis frequently harbored *ATR* mutations (21 out of 25 identified patients with an *ATR* mutation showed kataegis), suggesting that kataegis might be associated with collapsing replication forks in these patients.

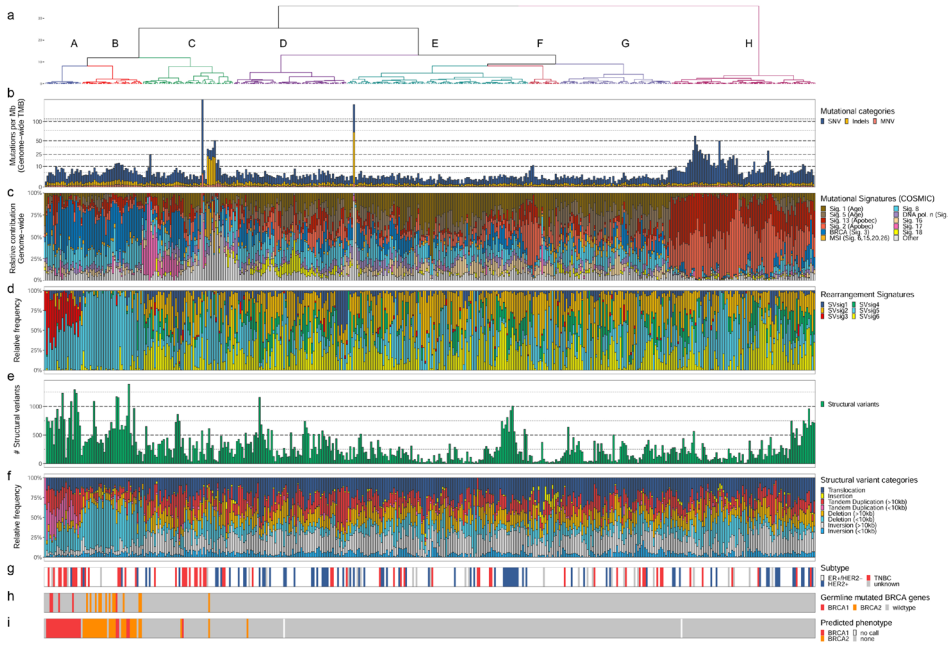


Figure 4 - Unsupervised clustering reveals distinct genomic phenotypes in metastatic breast cancer.
a, Dendrogram of unsupervised clustering. The eight clusters are denoted (A–H). The *Y* axis displays the clustering distance (Pearson and ward.D).
b, Number of genomic mutations per Mbp (TMB) divided into mutational categories SNV, indels and MNV. All genome-wide somatic mutations were taken in to consideration.
c, Relative contribution of COSMIC mutational signatures.
d, Relative contribution of rearrangement signatures.
e, Absolute number of unique SVs per sample.
f, Relative frequency per SV category; tandem duplications; deletions and inversions are subdivided into <10kb and >10kb categories.
g, Breast cancer subtype subdivided in ER^{+}/HER2^{-}, HER2^{+}, TNBC and unknown at the time of the analysis.
h, Germline *BRCA*1/2 mutational status.
i, HRD score as assessed by CHORD. The predicted phenotypes BRCA1 and BRCA2 deficiency are depicted.

Somatic drivers of metastatic breast cancer: SNVs and copy number alterations

Using the ratio of nonsynonymous and synonymous mutations (dN/dScv)[21], we identified 21 potential driver genes, including known key drivers of breast cancer. The top five driver genes were *TP53* (42.8%), *PIK3CA* (42.3%), *ESR1* (14.3%), *GATA3* (11.3%) and *KMT2C* (11.3%) (**Supplementary Fig. 10**). With regard to breast cancer subtypes, we observed that *TP53* was enriched in TNBC (P <1 x10^{-5}), whereas *ESR1*, *PIK3CA,* and *GATA3* were more often mutated in ER^{+} metastatic breast cancer (all P values <0.001). *ESR1* mutations were, as expected, more frequently present in patients pretreated with aromatase inhibitors (26.9% in patients who were pretreated versus 2.7% in patients without aromatase inhibitor pretreatment).

In addition to nonsynonymous mutations, we observed 44 rearrangements involving *ESR1* in 34 patients and deep gains of *ESR1* in 29 patients. Fusions, mutations and deep gains were not mutually exclusive, but were specific to ER$^+$ breast cancer. No amplifications of *cis*-acting enhancers of *ESR1* were observed.

We compared the frequency of alterations in our 21 identified potential drivers in metastatic breast cancer with two primary breast cancer cohorts: The Cancer Genome Atlas (TCGA)[4] and BASIS cohorts[6] (**Fig. 5**). Using an FDR <0.05, six genes, *ESR1, TP53, NF1, AKT1, KMT2C* and *PTEN* (**Table 1** and **Supplementary Table 3**), were more frequently mutated in ER$^+$/HER2$^-$ metastatic lesions than in primary breast cancer. Except for *ESR1,* these genes were not associated with pretreatment, nor with response. Individual analysis did not reveal mutual exclusivity of these genes; however grouping of mitogen-activated protein kinase pathway and ER transcriptional regulator genes (*NF1*, *TBX3*, *ERBB2, CTCF*, *EGFR*, *KRAS*, *BRAF*, *ERBB3, HRAS*, *MYC*) showed mutual exclusivity with *ESR1*, as shown by Razavi et al[18]. In patients with HER2$^+$ disease (irrespective of subdivision by ER status) or TNBC, no significant differences were observed. A bootstrap analysis to better estimate the distribution of gene mutation frequencies in primary disease (TCGA and BASIS cohorts combined) confirmed that observed enrichments of *ESR1, TP53, NF1, AKT1, KMT2C* and *PTEN* mutations were unlikely to be explained by sampling bias.

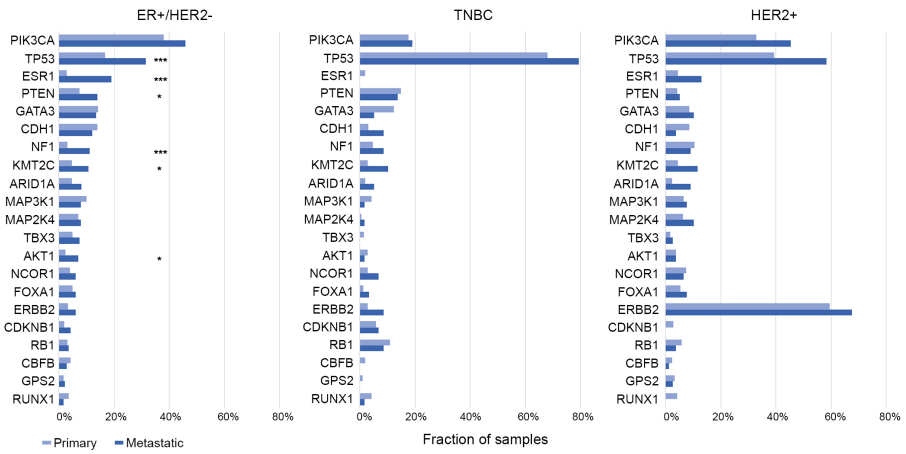


Figure 5 - Driver genes in metastatic breast cancer versus primary breast cancer.
Bar chart showing the frequency of affected driver genes in metastatic breast cancer (dark blue) versus primary breast cancer (light blue) subdivided in ER$^+$/HER2$^-$ (BASIS + TCGA, *n* = 816; CPCT, *n* = 279), TNBC (BASIS + TCGA, *n* = 310; CPCT, *n* = 58) and HER2$^+$ (BASIS + TCGA, *n* = 239; CPCT, *n* = 77). Two-sided Fisher's exact-test (Benjamini–Hochberg FDR-corrected): * P < 0.05, *** P < 0.001. The exact FDR-corrected *P* values are: *ESR1*, 8.46 x 10^{-16}; *TP53*, 6.92 x 10^{-6}; *NF1* 4.31 x 10^{-5}; *AKT1*, 0.012; *KMT2C*, 0.015; *PTEN*, 0.03.

Table 1 - Frequency of affected driver genes and type of genomic alteration in metastatic breast cancer

	ER+/HER2- (n=279)				HER2+ (n=77)				TNBC (n=58)			
	Total	Gain	Deletion	SNV/ InDels/ SVs	Total	Gain	Deletion	SNV/ InDels/ SVs	Total	Gain	Deletion	SNV/ InDels/ SVs
TP53	88	0	0	88	45	0	0	45	46	0	0	46
ESR1	53	6	0	47	10	0	0	10	0	0	0	0
CDH1	34	0	0	34	3	0	0	3	5	0	1	4
MAP3K1	22	0	1	21	6	1	0	5	1	0	0	1
GATA3	38	1	0	37	9	2	0	7	2	2	0	0
CBFB	8	0	0	8	1	0	0	1	0	0	0	0
ARID1A	23	0	1	22	7	0	0	7	3	0	0	3
ERBB2	17	2	0	15	51	46	0	5	4	1	0	3
RUNX1	5	0	0	5	0	0	0	0	1	0	0	1
MAP2K4	22	0	6	16	8	2	2	4	1	0	0	1
GPS2	6	0	0	6	2	0	0	2	0	0	0	0
FOXA1	17	2	0	15	6	1	0	5	2	0	0	2
TBX3	21	1	1	19	2	0	0	2	0	0	0	0
NCOR1	17	1	0	16	5	2	0	3	4	0	0	4
PTEN	39	0	7	32	4	0	0	4	8	0	3	5
PIK3CA	128	0	0	128	35	1	0	34	11	0	0	11
KMT2C	30	1	0	29	9	0	0	9	6	0	0	6
RB1	10	0	1	9	3	0	1	2	5	0	0	5
AKT1	20	2	0	18	3	2	0	1	1	0	0	1
CDKN1B	12	1	1	10	0	0	0	0	4	2	0	2
NF1	31	2	1	28	6	3	0	3	5	1	0	4

The dN/dScv analysis identified an additional potential driver gene, *GPS2*, which was not identified as a driver in primary breast cancer[6], but was recently described by Martincorena *et al.*[21] in primary breast cancer. The GPS2 protein forms a complex with NCOR1 and HDAC3. The three genes encoding these proteins were affected in an almost mutually exclusive fashion; 35 out of 36 patients harboring mutations in these genes had only one gene affected (CoMetExactTest, $P < 1 \times 10^{-5}$), indicating that the loss of either gene in this complex is sufficient. Alterations in GPS2–NCOR1–HDAC3 complex are enriched in metastatic breast cancer compared to primary breast cancer ($P = 0.004$), but not associated with a specific prior treatment or breast cancer subtype.

Regarding the primary 93 breast cancer driver genes reported by Nik-Zainal *et al.*[6] we found that, in addition to the above described differences between $ER^+/HER2^-$ primary and metastatic disease for *ESR1, NF1* and *TP53* mutations, *KMT2D* was also more frequently affected in metastatic disease whereas *AX1N1* was less frequently altered compared to primary breast cancer (FDR < 0.05) (**Supplementary Table 4**). Again, no differences for $HER2^+$ (irrespective of subdivision by ER status) and TNBC were observed.

Copy number analyses identified 51 narrow regions with somatic copy number alterations, including amplification peaks containing known driver genes such as *ERBB2*, *MYC* and *CCND1* and deletion peaks containing known tumor suppressor genes such as *PTEN*, *CKDN2A, RB1* and *NF1*. Using an FDR < 0.05, 29 regions were associated with ER status, that is, *MYC, SLC1A2* and *HOOK3* were more frequently amplified in ER$^-$ metastatic breast cancer and *PLK2* was more frequently deleted in ER$^-$ metastatic breast cancer. All amplification and deletion peaks in relation to ER status are shown in **Supplementary Table 5.** The total number of copy number alterations within these 51 regions was not associated with metastatic site or prognosis after 12 weeks of treatment. In addition we observed 6 focal amplification peaks (< 5 kilobases (kb)) in noncoding parts near three known breast cancer driver genes – *ZNF217*, *ZNF703* and *MYC* – and three other genes *LINC00266-1*, *TRPS1* and *KCNMB2*.

Potential clinical implications of WGS

To evaluate whether WGS may be used to improve treatment choices for future patients with metastatic breast cancer, we specifically focused on (1) high TMB/microsatellite instability (MSI) as a potential biomarker to select patients for immunotherapy, (2) HRD for poly-ADP ribose polymerase inhibitors and/or double stranded DNA break-inducing chemotherapy and (3) specific genomic alterations for which Food and Drug Administration (FDA)-approved drugs are already available (**Fig. 6**).

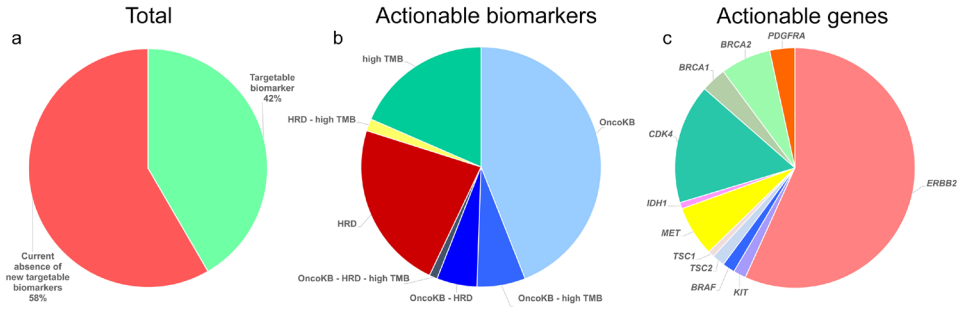


Figure 6 - Actionability.
a, Percentage of patients with and without an actionable target for treatment.
b, Actionable targets by type: HRD, high TMB (≥10 mutations Mbp^{-1}); and/or targetable alterations for which an FDA-approved drug is available (OncoKB knowledge base).
c, Genes indicated by OncoKB knowledge base for which targeted drugs are FDA-approved (*ERBB2* for breast cancer; all other genes for other cancer types).

Using a threshold of ≥10 mutations per Mbp in our cohort, previously used to distinguish responding from nonresponding patients with lung cancer receiving nivolumab plus ipilimumab[22], we identified 50 patients (11%) in our cohort with a high TMB, which, in most patients (70%), could be largely attributed to APOBEC-related mutations (≥50% of all mutations). In primary breast cancer APOBEC mutagenesis was previously associated with the presence of tumor-infiltrating lymphocytes further confirming antigenicity of APOBEC mutant cancers[23,24]. High TMB was not associated with breast cancer subtype, suggesting that inclusion of patients into future clinical trials investigating check point inhibitors should potentially be based on their genomic landscape rather than on tumor subtype (**Supplementary Fig. 2a**). Notably, five patients had a high TMB and a mutation in either *JAK2* or *STAT3*. As these latter mutations could help to evade the native immune response this might be of clinical relevance as well[25]. We also identified 7 (1.5%) patients with MSI according to MSI-seq[26,27], which is currently not tested in standard care but for which pembrolizumab has been approved for use in all tumor types[28].

Using CHORD, we identified 39 additional patients with HRD (9%) who did not harbor germline alterations in *BRCA1/2*. Based on their HRD phenotype, these patients might benefit from poly-ADP ribose polymerase inhibitors and/or chemotherapeutic agents that induce double-stranded DNA breaks[29].

Finally, we analyzed which patients could be treated with FDA-approved drugs based on the alterations present in their genome using the clinical annotation database OncoKB[30]. One hundred and five patients (24%) had at least one actionable event for which an FDA approved drug is currently available. 67 (15%) of all patients had an *ERBB2* amplification, 7 of which were clinically known as HER2^{-}. These patients might benefit from anti-HER2 therapies, which are already approved for breast cancer. Additionally, 47 patients had at least 1 alteration predicting response to a drug registered for other tumor types than breast cancer (**Fig. 6** and **Supplementary Table S6**). In summary, WGS provides us with a valuable tool to determine clinically relevant molecular features for informed treatment choices, such as TMB, HRD, MSI, and actionable mutations in one assay.

Discussion

An important feature of the current study is that all patients starting a new line of systemic treatment could participate. Thus, our cohort consists of patients treated with a heterogeneous repertoire of treatments leading to large differences in progression-free and overall survival. Expanding the number of patients combined with registration

of already known clinical prognostic factors, such as clinical performance scores and number of metastatic sites, allows for future reliable association analyses between genomic alterations and outcome stratified for relevant patient characteristics including pretreatment, breast cancer subtype and line of treatment.

We have identified differences in metastatic breast cancer compared to primary breast cancer regarding TMB, the frequency in which driver genes are affected and relative contribution of mutational signatures. Moreover, we have shown that the use of WGS enables to identify subgroups of patients (42% of all patients with metastatic breast cancer) for personalized treatment. Therefore, future clinical trials should incorporate tissue biopsies for sequencing and base treatment stratification on 'clinical genomics'.

Based on the current knowledge and treatment armamentarium, we still have a substantial number of patients with metastatic breast cancer (58%) without currently known targetable genomic features. Further exploration of large copy number changes, specific combinations of mutated genes and RNA sequencing will potentially unravel new actionable targets or profiles. The development and approval of new drugs which are currently under investigation, such as phosphoinositide 3- kinase inhibitors potentially relevant to a large subset of metastatic breast cancer patients (42% harbor a *PIK3CA* mutation in our cohort), will further increase the targetability of the tumor's genome.

Overall, our study provides significant insight into the biology of metastatic breast cancer and generates useful genomic information for future improvement of patient management.

Acknowledgements
We thank Barcode for Life, and Stichting Hetty Odink for financial support of clinical studies and WGS analyses. This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation and the Center of Personalized Cancer Treatment (CPCT) have made available to the study. We would like to thank all local principal investigators and medical specialists, and the nurses of all contributing centers for their help with patient recruitment. We are particularly grateful to all participating patients and their families.

Grants
This work was supported in parts by grants from Pink Ribbon (no. 204-184) and CZ health insurance (no. CZ-201300460). M.S. was supported by Cancer Genomics Netherlands through a grant from the Netherlands Organization of Scientific Research.

H.V.D.W, J.V.R and the Erasmus MC Cancer Computational Biology Center (CCBC) were financed through a grant for the Daniel den Hoed foundation.

References (main text only)

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7-30 (2018).
2. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
3. Nik-Zainal, S. & Morganella, S. Mutational Signatures in Breast Cancer: The Problem at the DNA Level. *Clin Cancer Res* **23**, 2617-2629 (2017).
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
5. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93 (2012).
6. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
7. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**, 517-525 (2017).
8. Swanton, C., McGranahan, N., Starrett, G.J. & Harris, R.S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov* **5**, 704-12 (2015).
9. Burns, M.B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366-70 (2013).
10. Brown, D. *et al.* Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun* **8**, 14944 (2017).
11. Brastianos, P.K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* **5**, 1164-1177 (2015).
12. Savas, P. *et al.* The Subclonal Architecture of Metastatic Breast Cancer: Results from a Prospective Community-Based Rapid Autopsy Program "CASCADE". *PLoS Med* **13**, e1002204 (2016).
13. Fumagalli, D. *et al.* Somatic mutation, copy number and transcriptomic profiles of primary and matched metastatic estrogen receptor-positive breast cancers. *Ann Oncol* **27**, 1860-6 (2016).
14. Ng, C.K.Y. *et al.* Genetic Heterogeneity in Therapy-Naive Synchronous Primary Breast Cancers and Their Metastases. *Clin Cancer Res* **23**, 4402-4415 (2017).
15. Schrijver, W. *et al.* Mutation Profiling of Key Cancer Genes in Primary Breast Cancers and Their Distant Metastases. *Cancer Res* **78**, 3112-3121 (2018).
16. Lefebvre, C. *et al.* Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med* **13**, e1002201 (2016).
17. Robinson, D.R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297-303 (2017).
18. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438 e6 (2018).
19. Boot, A. *et al.* In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res* **28**, 654-665 (2018).
20. Wyatt, M.D. & Wilson, D.M., 3rd. Participation of DNA repair in the response to 5-fluorouracil. *Cell Mol Life Sci* **66**, 788-99 (2009).
21. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **173**, 1823 (2018).
22. Ramalingam, S.S. *et al.* Abstract CT078: Tumor mutational burden (TMB) as a biomarker for clinical benefit from dual immune checkpoint blockade with nivolumab (nivo) + ipilimumab (ipi) in first-line (1L) non-small cell lung cancer (NSCLC): identification of TMB cutoff from CheckMate 568. *Cancer Research* **78**, CT078-CT078 (2018).
23. Smid, M. *et al.* Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nat Commun* **7**, 12910 (2016).
24. Pitt, J.J. *et al.* Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. *Nat Commun* **9**, 4181 (2018).
25. Yates, L.R. *et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. *Cancer Cell* **32**, 169-184 e7 (2017).
26. Huang, M.N. *et al.* MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci Rep* **5**, 13321 (2015).
27. Priestley, P. *et al.* Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv*, 415133 (2018).

28. Brahmer, J.R. *et al.* Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* **366**, 2455-65 (2012).
29. Lord, C.J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152-1158 (2017).
30. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**(2017).
31. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-47 (2009).
32. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
33. Casper, J. *et al.* The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* **46**, D762-D769 (2018).
34. Forbes, S.A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45**, D777-D783 (2017).
35. Tamborero, D. *et al.* Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* **10**, 25 (2018).
36. Griffith, M. *et al.* CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* **49**, 170-174 (2017).
37. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
38. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
39. Spurdle, A.B. *et al.* ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* **33**, 2-7 (2012).
40. Landrum, M.J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-8 (2016).
41. Huber, W., Toedling, J. & Steinmetz, L.M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22**, 1963-70 (2006).
42. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017).
43. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine* **10**, 33 (2018).
44. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
45. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-2 (2016).
46. Morganella, S. *et al.* The topography of mutational processes in breast cancer genomes. *Nat Commun* **7**, 11383 (2016).
47. Hahsler, M., Hornik, K. & Buchta, C. Getting Things in Order: An Introduction to the R Package seriation. *2008* **25**, 34 (2008).
48. Hodges, J.L. & Lehmann, E.L. Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics* **34**, 598-611 (1963).
49. Lehmann, E.L. Nonparametric Confidence Intervals for a Shift Parameter. **34**, 1507-1512 (1963).

Methods

Methods including statements of data availability are available in the online version of the paper.

Online Methods

Patient cohort and study procedures

For our analyses, we selected patients with metastatic breast cancer who were included under the protocol of the Center for Personalized Cancer Treatment (CPCT) consortium (CPCT-02 Biopsy Protocol, ClinicalTrial.gov no. NCT01855477), which was approved by the medical ethics committee of the University Medical Center Utrecht, the Netherlands. A detailed description of the consortium and the whole patient cohort has been described in detail recently[27]. This consortium consists of 49 oncology centers in the Netherlands and aims to analyze the cancer genome of patients with advanced cancer, irrespective of cancer type, to develop predictors for outcome to systemic treatment. Patients of ≥ 18 years, with incurable locally advanced or metastatic solid tumors, for whom a histological biopsy could be safely obtained and systemic treatment with anticancer agents was indicated, were eligible for inclusion. All patients gave written informed consent prior to any study procedure; the study complies with all relevant ethical regulations. We performed an in-depth analysis of all included patients with metastatic breast cancer. Patients who were biopsied in their primary breast tumor (n = 55), were excluded from the metastasis analyses, but were used as an additional control group. Patients with evaluable biopsies were classified according to the ER and HER2 status (**Supplementary Table 1**). Collection and sequencing of samples was performed as described previously[27].

Treatment outcome

Clinical outcome was evaluated according to RECIST v1.1 after 12 weeks of treatment and was defined as stable disease, partial response, complete response or progressive disease[31]. To relate outcome to genomic data, we defined response to therapy as CR of PR after 12 weeks of treatment and nonresponse as progressive disease after 12 weeks of treatment.

Detection of somatic changes

Detailed methods on calling of somatic SNVs, MNVs and SVs were previously described[27]. Additional annotation of somatic variants and heuristic filtering was performed: Heuristic filtering removed somatic SNV, indel and MNV variants based

on the following criteria: (1) minimal alternative reads observations ≤3; (2) Genome Aggregation Database (gnoMAD) exome (ALL) allele frequency ≥ 0.001 (corresponding to approximately 62 gnoMAD individuals); and (3) gnoMAD genome (ALL) ≥ 0.005 (approximately 75 gnoMAD individuals)[32]. GnoMAD database v2.0.2 was used. When multiple variants on the same genomic position were present, the most deleterious mutation was used to annotate the overlapping gene. SVs a with B-allele frequency ≥ 0.1 were further annotated by retrieving overlapping and nearest up-stream and downstream annotations using custom R scripts based on GRCh37 canonical University of California, Santa Cruz (UCSC) promoter and gene annotations with regard to their respective upstream or downstream orientation (if known)[33]. Only potential fusions with only two different gene partners were considered; SVs with both breakpoints falling within the same gene were simply annotated as SV mutations. Fusion annotation from the COSMIC (v.85), Cancer Genome Interpreter (CGI, November 6, 2018) and Clinical interpretation of Variants in Cancer (CIViC, accessed November 6, 2018) databases were used to assess known fusions[34-36]. The COSMIC, OncoKB (accessed July 12, 2018), CIVIC (accessed November 6, 2018), CGI (accessed November 6, 2018) and the list from Martincorena *et al.* (dN/dS) were used to classify known oncogenic or cancer-associated genes[21,34-36].

Ploidy and copy number analysis

Ploidy and copy number analysis was performed using a custom pipeline as previously described[27]. Briefly, this pipeline combines B-allele frequency, read depth and SVs to estimate the purity and copy number profile of a tumor sample. Recurrent focal and broad copy number alterations were identified with GISTIC v.2.0.23[37]. GISTIC was run with the following parameters: genegistic 1; gcm extreme; maxseg 4000; broad 1; brlen 0.98; conf 0.95; rx 0; cap 3; saveseg 0; armpeel 1; smallmem 0; res 0.01; ta 0.1; td 0.1; savedata 0; savegene 1; qvt 0.1. Categorization of shallow and deep copy number aberration per gene was based on thresholded GISTIC calls. Focal peaks detected by GISTIC were re-annotated, based on overlapping genomic coordinates, using custom R scripts and UCSC gene annotations. GISTIC peaks were annotated with all overlapping canonical UCSC genes within the narrow peak limits. If a narrow GISTIC peak overlapped with ≤ 3 genes, the most probably targeted gene was selected based on oncogenic or tumor suppressor annotation in the COSMIC, OncoKB, CIViC and CGI lists[21,34-36]. Peaks in gene deserts were annotated with their nearest gene.

Putative enhancer regions (as detected by GISTIC; focal amplification peaks with a width < 5,000 bp) were retrieved per sample. If regions overlapped multiple distinct copy number segments, the maximum copy number value of the overlapping segments

was used to represent the region. Samples with gene-to-enhancer ratios deviating >1 studentized residual from equal 1:1 gene-to-enhancer ratios (linear model: $\log_2$(copy number of enhancer) – $\log_2$(copy number of gene locus) = approximately 0) were categorized as gene- or enhancer-enriched. Based on the direction of the ratio, samples were either denoted as enhancer- (if the ratio was positive) or gene-enriched (if the ratio was negative).

Estimation of tumour mutational burden

The mutation rate per million base pairs (Mbp) of genomic DNA was calculated as the total genome-wide amount of SNV, MNV and indels divided by the total amount of mappable nucleotides (ACTG) in the human reference genome (hg19) FASTA sequence file:

$$TMB_{genomic} = \frac{(SNV_{genomic} + MNV_g + indels_{genomic})}{\left(\frac{2858674662}{10^6}\right)} \quad (1)$$

The mutation rate per Mb of coding mutations was calculated as the amount of coding SNV, MNV and indels divided by the summed lengths of distinct nonoverlapping coding regions, as determined on the subset of protein-coding and fully supported (transcript support level (TSL) = 1) transcripts in GENCODE v.28 (hg19)[38]:

$$TMB_{coding} = \frac{(SNV_{coding} + MNV_{coding} + indels_{coding})}{\left(\frac{28711682}{10^6}\right)} \quad (2)$$

MSI and HRD prediction

HRD/'BRCAness' was estimated using the CHORD classifier (Nguyen, van Hoeck and Cuppen, manuscript in preparation). This classifier was based on the HRDetect[7] algorithm; however, redesigned to improve its performance beyond primary breast cancer. The binary prediction score (ranging from 0 to 1) was used to indicate BRCAness level within sample. A *BRCA1*/2 variant was assigned as pathogenic when annotated by ENIGMA[39] (accessed February 26, 2018) or ClinVar[40] (accessed January 28, 2018).

We used the following gene list to check whether homologous recombination-related genes were mutated in samples that clustered in clusters A and B and were classified as homologous recombination-proficient (**Fig. 4**): *ATM; BARD1; BLM; BRCA1; BRCA2; BRIP1; EME1; ERCC1; ERCC4; EXO1; GEN1; H2AFX; MRE11A; MUS81; NBN; NSMCE1; NSMCE2; PALB2; PCNA; RAD18; RAD21; RAD50; RAD51; RAD51AP1; RAD51C; RAD51L1; RAD51L3; RAD52; RAD54B; RAD54L; RECQL4; RECQL5; RTEL1; SLX1A; SLX4; TDP1; WRN; XRCC2;* and *XRCC3.*

MSI status was determined using the MSIseq score[26,27]. Briefly, this validated score

classifies a sample based on the number of indels per Mbp occurring in homopolymers of ≥5 bases or dinucleotide, trinucleotide and tetranucleotide sequences of repeat count ≥4. A sample with an MSI-seq score ≥4 was considered MSI.

Detection of (onco)genes under selective pressure

To detect (onco)genes under tumor evolutionary mutational selection, we employed a Poisson-based dN/dS model (under the full trinucleotide model; 192 rate parameters) using the R package dndscv v0.0.0.9[21]. Briefly, this model tests the normalized ratio of nonsynonymous (missense, nonsense and splicing) over background (synonymous) mutations while correcting for sequence composition and mutational signatures. A global $q \leq 0.1$ (with and without taking indels into consideration) was used to identify statistically significant driver genes.

Identification of hypermutated foci (kataegis)

Putative kataegis events were detected using a dynamic programming algorithm that determines a globally optimal fit of a piecewise constant expression profile along genomic coordinates as described by Huber *et al.* [41] and implemented in the tilingArray R package v.1.56.0. Only SNVs were used to detect kataegis. Each chromosome was assessed separately and the maximum number of segmental breakpoints was based on a maximum of 5 consecutive SNVs (maximum 5,000 segments per chromosome). Fitting was performed on $\log_{10}$-transformed intermutational distances. Per segment, we assessed if the mean intermutational distance was ≤ 2,000 bp and at least 5 SNVs were used in the generation of the segment. Samples with >200 distinct observed events were set to zero-observed events as these were found to be hypermutated throughout the entire genome rather than locally. Kataegis was visualized using the R package karyoploteR v1.4.1[42].

Mutational and structural rearrangement signatures analysis

Mutational signature analysis using the MutationalPatterns R package v.1.4.2 was performed as described previously[43]. The thirty COSMIC mutational signatures, as established by Alexandrov *et al*[2], (matrix S*ij*; *i* = 96 trinucleotide motifs; *j* = number of signatures) were downloaded from COSMIC (accessed 23 May 2018). For de novo signature calling, between 2 and 20 signatures were assessed using the NMF package v.0.21.0 with 500 iterations[44]. By comparing the cophenetic correlation coefficient over the range of possible signatures, we opted to assign ten de novo signatures. We used the cosine similarity metric to compare the de novo with COSMIC signatures. Structural rearrangement signatures were established as previously described[6]. Briefly, SVs were called using Manta v.1.0.3[45] and default parameters, after which additional filters were

applied[27]. The reported tandem duplications, deletions, inversions, insertions and translocations were then categorized by size (<10kb, 10–100kb, 100kb–1Mb, 1–10Mb and >10Mb). Interrearrangement distances were calculated and rearrangements were labelled as clustered if the average interrearrangement distance of a segment was at least ten times less than the whole-genome average for a patient sample. The segments were determined using a piecewise constant fitting function ('exactPcf' from the copynumber R package) using a minimum of ten events in a segment (K_{min}) and a γ of 25 (smoothness of segmentation). To calculate an a posteriori probability for each substitution, we implemented the method described previously[46], that assigns the most probable causative signature. In short, this method uses the contribution of each signature in each sample in conjunction with the probability of a signature to generate the particular substitution in its trinucleotide context.

Unsupervised clustering of metastatic breast cancer WGS characteristics

Samples were clustered using Pearson correlation coefficient ($1 - r$), as distance metric, and Ward.D hierarchical clustering based the following whole-genome characteristics: number of SNV; indels and MNVs per Mbp, total number and number by type of SVs; and relative frequencies of mutational signatures. Data were scaled but not centered (root mean square) before calculating the Pearson correlation coefficients. After clustering, optimal leaf ordering was performed using the seriation package v.1.2.3[47]. Gap statistic method was employed to determine optimal number of discriminating clusters.

Comparison with primary breast cancer

BASIS cohort

The somatic mutations for the BASIS cohort were extracted from the European Genome-phenome Archive (accession code EGAS00001001178). This cohort of the complete genomes of 560 primary breast cancers and paired nonneoplastic tissue as reference, consists of 320 patients with $ER^+/HER2^-$, 46 patients $ER^+/HER2^+$, 27 patients with $ER^-/HER2^+$ and 167 patients with TNBC[6]. To allow comparison of mutational loads, mutational signatures and somatic mutations between the BASIS cohort and our cohort, we compared whether the calling from both pipelines yielded comparable results for eight patients from the BASIS cohort. Since the mutational load (linear regression $R^2 = 0.9987$), mutational signatures (average similarity of 0.90 (s.d. 0.08), which are significantly higher (one-sample t-test, $P = 1.57 \times 10^{-5}$) than the similarity between nonmatching samples), and detected driver genes were very similar between both pipelines, we considered the results from the pipelines to be comparable.

TCGA cohort

Breast cancer data (*n* = 805) were downloaded from cBioPortal.org (https://www.cbioportal.org/; accessed April 2018). Synonymous mutations were removed and multiple mutations in the same gene/patient were combined. For the copy number data, a –2 call was used as deletion, +2 for amplification. This cohort consists of 143 patients who were triple negative, 496 patients with $ER^+/HER2^-$, 39 patients with $ER^-/HER2^+$, and 127 patients with $ER^+/HER2^+$.

Selection of cohort per analysis

For the most optimal comparison of TMB, and absolute and relative contributions of COSMIC signatures between primary breast cancer and metastatic breast cancer, we selected the BASIS cohort, since this cohort also used WGS making it the most suitable dataset for these comparisons. For the comparison of driver genes, which are located in the coding parts of the genome, we decided to use both cohorts (TCGA and BASIS) to increase power.

Bootstrapping of primary cohort

To investigate whether the enrichment of driver genes in metastatic breast cancer compared to primary breast cancer was influenced by population differences or sampling bias, we performed a bootstrap analysis to better estimate the distribution of gene mutation frequencies in primary breast cancer (TCGA[4] and BASIS[6] cohorts combined). For each of the driver genes, a bootstrap analysis was performed by taking the actual mutated frequency in primary breast cancer within a subtype, randomly selecting 80% of cases using sampling with replacement and counting the number of times a sample was selected that was mutated for that gene. This was repeated 100,000 times to obtain an estimated distribution for a gene in primary breast cancer. Then, we determined whether the mutation frequency of that gene in the metastatic cohort in the same subtype fell outside the 99% percentile of the estimated primary breast cancer distribution.

Statistics

Pearson's chi-squared test or Fisher's exact test (in case of too few expected events) was used to evaluate categorical data (for example, prior treatment versus the occurrence of a certain mutation). To compare continuous variables (for example the relative contribution of mutational signatures versus breast cancer subtype or RECIST v1.1 response category (complete response/ partial response or progressive disease)) a Mann–Whitney *U*-test or a Kruskal–Wallis H test was performed. Where suitable, effect sizes and CIs were estimated using Hodges–Lehmann estimator[48,49]. All statistical

tests were two-sided and considered statistically significant when $P < 0.05$. Stata 13.0 (StataCorp), R v.3.4.4. or SPSS 24 (IBM Corporation) were used for the statistical analyses. We used the Benjamini–Hochberg procedure to correct *P* values for multiple hypothesis testing when appropriate.

Code availability

Full codes are available at https://github.com/hartwigmedical/ and https://bitbucket.org/ccbc/r2ccbc.

Data availability

WGS data and corresponding clinical data have been requested from Hartwig Medical Foundation and provided under data request number DR-026. The clinical data provided by CPCT have been locked at 1st of June 2018. Both WGS and clinical data are freely available for academic use from the Hartwig Medical Foundation through standardized procedures and request forms can be found at https://www.hartwigmedicalfoundation.nl[27].

Methods-only References

37. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
38. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-1774 (2012).
39. Spurdle, A. B. *et al.* ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Hum Mutat* **33**, 2-7 (2012).
40. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**, D862-868 (2016).
41. Huber, W., Toedling, J. & Steinmetz, L. M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22**, 1963-1970 (2006).
42. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017).
43. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine* **10**, 33, doi:10.1186/s13073-018-0539-0 (2018).
44. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
45. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
46. Morganella, S. *et al.* The topography of mutational processes in breast cancer genomes. *Nat Commun* **7**, 11383 (2016).
47. Hahsler, M., Hornik, K. & Buchta, C. Getting Things in Order: An Introduction to the R Package seriation. *2008* **25**, 34, doi:10.18637/jss.v025.i03 (2008).
48. Hodges, J. L. & Lehmann, E. L. Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics* **34**, 598-611 (1963).
49. Lehmann, E. L. Nonparametric Confidence Intervals for a Shift Parameter. **34**, 1507-1512, doi:10.1214/aoms/1177703882 (1963).

Supplementary tables

Supplementary Table S1 - Patient characteristics

	Patients (n = 442)	
Age	N	%
Median	58	
Range	28 - 86	
Gender		
Female	438	99.10
Male	4	0.90
Breast cancer subtype		
ER-positive/HER2-negative	279	63.12
ER-positive/HER2-positive	49	11.09
ER-negative/HER2-positive	28	6.33
Triple negative	58	13.12
Unknown at time of analysis	28	6.33
Prior systemic therapy		
Yes	367	83.03
Endocrine therapy only	46	12.53
Chemotherapy only	59	16.08
Endocrine and chemotherapy	162	44.14
Endocrine, chemo and targeted therapy	72	19.62
Endocrine and targeted	4	1.09
Chemo and targeted therapy	24	6.54
If Targeted		
Anti-HER2	55	55.00
Everolimus	43	43.00
Anti-HER2 and everolimus	2	2.00
No	61	13.80
Unknown at time of analysis	14	3.17
Prior radiotherapy		
Yes	284	64.25
No	144	32.58
Unknown at time of analysis	14	3.17
Started therapy after biopsy for WGS		
Yes	340	76.92
Endocrine based		
Aromatase inhibitor	38	11.18
Aromatase inhibitor + CDK4/6 inhibitor	22	6.47
Aromatase inhibitor + everolimus	13	3.82
Fulvestrant	2	0.59
Fulvestrant + CDK4/6 inhibitor	21	6.18
Tamoxifen	6	1.76
Other	4	1.18
Chemotherapy based		
Anthracycline based	24	7.06
Taxane based	32	9.41
Platinum based	21	6.18
Platinum + Taxane	3	0.88
Anthracycline + Taxane	2	0.59
Single agent *	81	23.82
Other chemo	2	0.59
Anti- HER2 based therapy	54	15.88
Other	15	4.41
No	21	4.75
Unknown at time of analysis	81	18.33

Supplementary Table S1 - Continued

	Patients (n = 442)	
Biopsy site		
Liver	199	45.02
Lymph node	94	21.27
Bone	50	11.31
Lung	12	2.71
Soft tissue **	48	10.86
Other ***	18	4.07
Unknown at time of analysis	21	4.75

* Single agent: capecitabine, vinorelbine, cyclophosphamide, eribuline, gemcitabine

** Soft tissue: (sub)cutis, muscle

***Other, including: brain, omentum, peritoneum, adrenal gland, ovarium

Supplementary Table S2 - Cohorts for the comparison of genomic alterations

	TCGA N = 805		BASIS N = 560		Robinson N = 91		CPCT N = 442	
Setting	Primary breast cancer		Primary breast cancer		Metastatic breast cancer		Metastatic breast cancer	
Breast cancer subtype	N	%	N	%	N	%	N	%
ER+/HER2-	496	61.6	320	57.1	NA	NA	279	63.1
ER+/HER2+	127	15.8	46	8.2	NA	NA	49	11.1
ER-/HER2+	39	4.8	27	4.8	NA	NA	28	6.3
Triple negative	143	17.8	167	29.8	NA	NA	58	13.1
Unknown	-	-	-	-	91	100	28	6.3
Sequencing method	Whole exome		Whole genome		Whole exome		Whole genome	
Median sequencing depth (IQR)								
Normal	Not reported		29.9 (22.2-35.4)		125 (110-138)		38 (35-42)	
Tumour	At least 70% 20x depth of ~34 Mbp coding region target		41.7 (38.2-45.8)		178 (157-202.5)		107 (98-114)	
Median tumour content (%) assessed by pathologist (IQR)	Not reported		70 (60-80)		68 (50-79)		60 (45-75)	

Supplementary Table S3 - Frequency of affected driver genes (defined by dN/dScv) per breast cancer subtype

				Frequency in metastatic breast cancer			Frequency in primary breast cancer (TCGA and BASIS)		
Gene	FDR* HER2+	FDR* ER+/HER2-	FDR* TNBC	HER2+ N=77	ER+/HER2- N=279	TNBC N=58	HER2+ N=239	ER+/HER2- N=816	TNBC N=310
ESR1	0.33	8.46E-16	1	13.0%	19.0%	0.0%	4.6%	2.8%	1.9%
TP53	0.08	6.92E-06	1	58.4%	31.5%	79.3%	39.3%	16.8%	68.1%
NF1	1	4.31E-05	1	9.1%	11.1%	8.6%	10.5%	3.2%	4.8%
AKT1	1	0.0123	1	3.9%	7.2%	1.7%	3.8%	2.5%	2.9%
KMT2C	0.60	0.0147	0.40	11.7%	10.8%	10.3%	4.6%	4.8%	2.9%
PTEN	1	0.0270	1	5.2%	14.0%	13.8%	4.2%	7.5%	14.8%
PIK3CA	0.96	0.3514	1	45.5%	45.9%	19.0%	33.1%	38.0%	17.7%
ARID1A	0.36	0.5004	1	9.1%	8.2%	5.2%	2.5%	4.8%	1.9%
CDKN1B	1	0.5932	1	0.0%	4.3%	6.9%	2.9%	2.0%	5.8%
ERBB2	1	0.6134	1	67.5%	6.1%	8.6%	59.4%	3.3%	2.9%
CBFB	1	0.8425	1	1.3%	2.9%	0.0%	2.5%	4.3%	1.9%
CDH1	1	0.8425	1	3.9%	12.2%	8.6%	8.8%	14.0%	3.2%
FOXA1	1	0.8425	1	7.8%	6.1%	3.5%	5.4%	4.9%	1.3%
GATA3	1	0.8425	1	10.4%	13.6%	5.2%	8.8%	14.2%	12.3%
GPS2	1	0.8425	1	2.6%	2.2%	0.0%	3.4%	1.7%	1.0%
MAP2K4	1	0.8425	1	10.4%	7.9%	1.7%	6.3%	7.1%	0.7%
MAP3K1	1	0.8425	1	7.8%	7.9%	1.7%	6.7%	10.1%	4.2%
NCOR1	1	0.8425	1	6.5%	6.1%	6.9%	7.5%	4.2%	2.9%
RB1	1	0.8425	1	3.9%	3.6%	8.6%	5.9%	3.2%	11.0%
RUNX1	1	0.8425	1	0.0%	1.8%	1.7%	4.2%	3.7%	4.2%
TBX3	1	0.8425	1	2.6%	7.5%	0.0%	1.7%	4.9%	1.6%

*Hochberg corrected FDR of a two-sided Fisher's exact-test

Supplementary Table S4 - Frequency of driver genes (93 breast cancer driver genes reported by Nik-Zainal *et al.*) per breast cancer subtype

				Frequency in metastatic breast cancer			Frequency in primary breast cancer (TCGA and BASIS)		
Gene	**FDR* HER2+**	**FDR* ER+/HER2-**	**FDR* TNBC**	**HER2+ N=77**	**ER+/HER2- N=279**	**TNBC N=58**	**HER2+ N=239**	**ER+/HER2- N=816**	**TNBC N=310**
ESR1	1	3.75E-15	1	13.0%	19.0%	0.0%	4.6%	2.8%	1.9%
KMT2D	1	1.32E-05	1	5.2%	8.6%	1.7%	4.2%	1.5%	4.5%
TP53	0.35	3.15E-05	1	58.4%	31.5%	79.3%	39.3%	16.8%	68.1%
NF1	1	0.00	1	9.1%	11.1%	8.6%	10.5%	3.2%	4.8%
AXIN1	1	0.03	1	2.6%	0.7%	1.7%	3.8%	5.3%	3.2%
AKT1	1	0.06	1	3.9%	7.2%	1.7%	3.8%	2.5%	2.9%
KMT2C	1	0.08	1	11.7%	10.8%	10.3%	4.6%	4.8%	2.9%
ATR	1	0.09	1	7.8%	5.4%	6.9%	2.9%	1.5%	5.2%
MDM2	1	0.12	1	2.6%	9.0%	1.7%	5.0%	3.8%	1.6%
PTEN	1	0.14	1	5.2%	14.0%	13.8%	4.2%	7.5%	14.8%
MLLT4	1	1.00	1	0.0%	0.0%	0.0%	5.9%	2.7%	4.8%
CIC	1	0.28	1	2.6%	3.6%	5.2%	1.7%	0.9%	2.9%
APC	1	0.66	1	2.6%	5.0%	3.5%	2.5%	1.8%	3.2%
ATRX	1	0.74	1	3.9%	5.7%	5.2%	5.9%	2.3%	3.6%
PALB2	1	0.93	1	2.6%	1.1%	1.7%	4.2%	4.2%	2.3%
ZNF217	1	0.94	1	13.0%	13.3%	6.9%	16.3%	8.0%	5.2%
AKT2	1	1	1	2.6%	0.4%	5.2%	0.8%	1.8%	3.9%
ARID1A	1	1	1	9.1%	8.2%	5.2%	2.5%	4.8%	1.9%
ARID1B	1	1	1	3.9%	2.5%	3.5%	4.6%	1.5%	6.5%
ASXL1	1	1	1	3.9%	2.9%	0.0%	1.7%	1.0%	6.5%
ATM	1	1	1	6.5%	6.1%	3.5%	4.6%	2.8%	3.9%
BCOR	1	1	1	2.6%	2.5%	0.0%	3.8%	1.8%	4.5%
BRAF	1	1	1	1.3%	2.9%	1.7%	2.1%	1.2%	3.9%
BRCA1	1	1	1	6.5%	2.2%	1.7%	6.3%	2.0%	3.9%
BRCA2	1	1	1	10.4%	6.1%	3.5%	4.6%	3.9%	5.2%
BUB1B	1	1	1	1.3%	1.8%	0.0%	2.1%	0.3%	4.2%
CASP8	1	1	1	2.6%	1.1%	3.5%	2.5%	1.0%	2.9%
CBFB	1	1	1	1.3%	2.9%	0.0%	2.5%	4.3%	1.9%
CBLB	1	1	1	1.3%	0.7%	3.5%	2.5%	1.7%	2.3%
CCND1	1	1	1	18.2%	25.1%	8.6%	22.6%	19.1%	4.2%
CCND3	1	1	1	0.0%	1.4%	5.2%	1.3%	1.4%	8.1%

Supplementary Table S4 - Continued

Gene	FDR* HER2+	FDR* ER+/HER2-	FDR* TNBC	Frequency in metastatic breast cancer			Frequency in primary breast cancer (TCGA and BASIS)		
				HER2+ N=77	ER+/HER2- N=279	TNBC N=58	HER2+ N=239	ER+/HER2- N=816	TNBC N=310
CCNE1	1	1	1	1.3%	1.4%	10.3%	3.4%	2.1%	9.4%
CDH1	1	1	1	3.9%	12.2%	8.6%	8.8%	14.0%	3.2%
CDK6	1	1	1	1.3%	0.7%	5.2%	0.0%	1.7%	3.6%
CDKN1B	1	1	1	0.0%	4.3%	6.9%	2.9%	2.0%	5.8%
CDKN2A	1	1	1	10.4%	6.5%	15.5%	5.0%	3.2%	8.1%
CDKN2B	1	1	1	10.4%	5.4%	15.5%	4.6%	2.8%	7.7%
CNOT3	1	1	1	2.6%	1.8%	1.7%	4.6%	2.7%	2.3%
CREBBP	1	1	1	3.9%	2.2%	1.7%	5.9%	5.0%	4.8%
CTCF	1	1	1	1.3%	3.2%	1.7%	2.9%	3.7%	1.3%
CUX1	1	1	1	2.6%	1.8%	1.7%	2.1%	1.4%	3.6%
DNMT3A	1	1	1	2.6%	1.4%	3.5%	1.3%	1.5%	3.6%
ECT2L	1	1	1	6.5%	3.2%	3.5%	3.4%	1.5%	4.8%
EGFR	1	1	1	1.3%	2.9%	5.2%	3.4%	1.6%	5.2%
ERBB2	1	1	1	67.5%	6.1%	8.6%	59.4%	3.3%	2.9%
ERBB3	1	1	1	1.3%	3.6%	5.2%	5.0%	1.4%	2.3%
ERCC4	1	1	1	2.6%	1.8%	1.7%	5.0%	3.9%	2.3%
FBXW7	1	1	1	3.9%	1.4%	0.0%	2.1%	1.7%	5.8%
FGFR1	1	1	1	11.7%	19.7%	6.9%	14.6%	14.0%	8.4%
FGFR2	1	1	1	0.0%	4.3%	6.9%	2.9%	3.1%	2.9%
FOXA1	1	1	1	7.8%	6.1%	3.5%	5.4%	4.9%	1.3%
FOXP1	1	1	1	3.9%	1.4%	0.0%	3.4%	1.2%	2.6%
GATA3	1	1	1	10.4%	13.6%	5.2%	8.8%	14.2%	12.3%
GNAS	1	1	1	7.8%	10.8%	3.5%	9.6%	6.7%	4.8%
HRAS	1	1	1	0.0%	0.7%	3.5%	0.8%	0.6%	2.9%
IGF1R	1	1	1	1.3%	2.5%	3.5%	5.4%	4.5%	4.8%
KDM6A	1	1	1	5.2%	1.1%	0.0%	2.9%	2.8%	4.8%
KRAS	1	1	1	6.5%	2.9%	8.6%	1.3%	1.6%	7.1%
MAP2K4	1	1	1	10.4%	7.9%	1.7%	6.3%	7.1%	0.7%
MAP3K1	1	1	1	7.8%	7.9%	1.7%	6.7%	10.1%	4.2%
MED23	1	1	1	2.6%	5.0%	1.7%	2.5%	2.9%	4.8%
MEN1	1	1	1	2.6%	1.1%	1.7%	4.2%	0.5%	1.0%

Supplementary Table S4 - Continued

Gene	FDR* HER2+	FDR* ER+/HER2-	FDR* TNBC	Frequency in metastatic breast cancer			Frequency in primary breast cancer (TCGA and BASIS)		
				HER2+ N=77	ER+/HER2- N=279	TNBC N=58	HER2+ N=239	ER+/HER2- N=816	TNBC N=310
MLH1	1	1	1	2.6%	1.4%	1.7%	1.3%	1.1%	1.6%
MSH2	1	1	1	1.3%	2.2%	3.5%	2.9%	0.7%	1.6%
MYC	1	1	1	18.2%	16.1%	29.3%	29.3%	13.2%	30.0%
NCOR1	1	1	1	6.5%	6.1%	6.9%	7.5%	4.2%	2.9%
NF2	1	1	1	0.0%	0.4%	0.0%	1.7%	0.9%	1.9%
NOTCH1	1	1	1	1.3%	2.9%	3.5%	3.8%	1.1%	3.9%
NOTCH2	1	1	1	3.9%	3.2%	13.8%	13.4%	7.4%	17.1%
NRAS	1	1	1	1.3%	1.1%	3.5%	0.4%	1.0%	2.9%
PBRM1	1	1	1	1.3%	1.8%	8.6%	2.1%	1.0%	1.9%
PDGFRA	1	1	1	0.0%	2.9%	1.7%	2.9%	1.1%	3.9%
PHF6	1	1	1	0.0%	1.4%	0.0%	3.4%	1.6%	4.5%
PIK3CA	1	1	1	45.5%	45.9%	19.0%	33.1%	38.0%	17.7%
PIK3R1	1	1	1	2.6%	2.5%	5.2%	2.9%	2.1%	5.8%
PMS2	1	1	1	0.0%	1.1%	0.0%	2.1%	1.6%	2.6%
PRDM1	1	1	1	10.4%	2.9%	10.3%	7.5%	1.8%	4.5%
PREX2	1	1	1	13.0%	9.7%	12.1%	12.6%	9.4%	11.6%
RB1	1	1	1	3.9%	3.6%	8.6%	5.9%	3.2%	11.0%
RHOA	1	1	1	1.3%	1.4%	1.7%	2.1%	1.0%	1.6%
RUNX1	1	1	1	0.0%	1.8%	1.7%	4.2%	3.7%	4.2%
SETD2	1	1	1	5.2%	1.8%	1.7%	2.9%	2.3%	2.3%
SF3B1	1	1	1	1.3%	3.2%	0.0%	2.1%	3.6%	1.3%
SMAD4	1	1	1	0.0%	1.4%	0.0%	2.9%	2.0%	1.6%
SMARCA4	1	1	1	5.2%	1.4%	10.3%	3.8%	1.5%	5.8%
SPEN	1	1	1	3.9%	5.7%	0.0%	5.4%	3.6%	4.8%
STAG2	1	1	1	2.6%	1.1%	5.2%	4.2%	1.5%	4.5%
STK11	1	1	1	0.0%	1.4%	8.6%	2.1%	0.9%	2.9%
TBX3	1	1	1	2.6%	7.5%	0.0%	1.7%	4.9%	1.6%
TET2	1	1	1	2.6%	2.5%	0.0%	1.7%	1.2%	2.9%
USP9X	1	1	1	3.9%	3.2%	0.0%	4.6%	2.5%	5.5%
XBP1	1	1	1	1.3%	2.9%	1.7%	3.8%	1.0%	1.3%
ZFP36L1	1	1	1	1.3%	1.1%	0.0%	1.3%	0.9%	1.0%

*Hochberg corrected FDR of a two-sided Fisher's exact-test

Supplementary Table S5 - Gains and losses defined by GISTIC2.0 (v2.0.23)

Unique Name	Descriptor	Chromosome	start	end	frequency shallow gain (%)	frequency deep gain (%)	Residual q values after removing segments shared with higher peaks*	Overlapping Genes	frequency ER+ (%) N=328	frequency ER- (%) N=86	FDR**
					Gains						
Amplification Peak 1	1q32.1	chr1	206509602	206757551	16.29	65.84	2,76E-13	*IKBKE*	84.76	74,42	0,04
Amplification Peak 2	3q26.32	chr3	178539144	178985909	21.49	21.95	0,088179	*PIK3CA*	40.55	55,81	0,02
Amplification Peak 3	6q21	chr6	107143141	107313999	11.09	12.67	5,53E-10	*ATG5*	20.73	37,21	0,00
Amplification Peak 4	8p11.23	chr8	37441298	37561815	11.54	31.90	1,06E-57	*ZNF703*	46.34	32,56	0,04
Amplification Peak 5	8q11.1	chr8	43811001	47040098	23.98	33.71	0,0047755	*HOOK3*	54.27	75,58	0,00
Amplification Peak 6	8q23.3	chr8	115544191	117523477	15.38	58.14	0,00022117	*RAD21*	71.04	86,05	0,01
Amplification Peak 7	8q24.21	chr8	128692053	128765288	13.12	59.50	6,01E-34	*MYC*	69.51	86,05	0,01
Amplification Peak 8	11p13	chr11	35245819	35381001	14.93	13.57	0,0014368	*SLC1A2*	24.09	44,19	0,00
Amplification Peak 9	11q13.3	chr11	69412252	69502689	16.74	38.01	1,38E-65	*CCND1*	54.88	61,63	0,35
Amplification Peak 10	12q15	chr12	69484002	69824999	21.72	20.36	3,95E-14	*FRS2*	44.21	34,88	0,18
Amplification Peak 11	17q12	chr17	37814431	37950999	13.12	21.95	4,12E-36	*ERBB2, IKZF3*	33.84	39,53	0,41
Amplification Peak 12	17q23.1	chr17	57879002	58010595	18.55	30.77	8,88E-19	*CLTC*	52.13	38,37	0,04
Amplification Peak 13	20q13.2	chr20	52146334	52396999	23.53	42.31	1,60E-22	*ZNF217*	65.24	72,09	0,32
Amplification Peak 14	20q13.33	chr20	62717002	63025520	25.57	38.24	0,0037973	*ARFRP1*	60.98	74,42	0,04
					Losses						
Deletion Peak 1	1p36.22	chr1	6234713	12630697	51.13	1.13	6,54E-07	*CAMTA1, MTOR, PIK3CD, ERRFI1, RPL22*	52.13	50,00	0,83
Deletion Peak 2	1p36.11	chr1	17761302	29662075	52.04	0.90	0,056733	*EPHB2, MDS2, ID3, PAX7, ARHGEF10L, ARID1A, SESN2, MAP3K6, CDC42*	53.96	47,67	0,41
Deletion Peak 3	1p21.3	chr1	79470793	113255512	43.44	0.68	7,97E-05	*CSF1, DPYD, RPL5, RBM15, BCL10*	47.56	30,23	0,01
Deletion Peak 4	2q22.1	chr2	139648821	143638367	37.33	0.90	2,63E-06	*LRP1B*	37.50	41,86	0,58
Deletion Peak 5	2q37.3	chr2	242028208	243199373	38.24	1.81	3,29E-06	*PASK, PDCD1*	37.50	48,84	0,09
Deletion Peak 6	3p14.2	chr3	59027297	61552450	50.68	1.81	5,84E-13	*FHIT*	50.00	63,95	0,03
Deletion Peak 7	3p13	chr3	70060535	72804696	50.00	2.94	8,05E-07	*RYBP, FOXP1, SHQ1*	52.74	54,65	0,82

Supplementary Table S5 - Continued

Unique Name	Descriptor	Chromosome	start	end	frequency shallow gain (%)	frequency deep gain (%)	Residual q values after removing segments shared with higher peaks*	Overlapping Genes	frequency ER+ (%) N=328	frequency ER- (%) N=86	FDR**
Deletion Peak 8	4p15.2	chr4	1	37247310	41.63	0.90	0,0081839	*SLC34A2, FGFR3, SLIT2*	34.15	70,93	0,00
Deletion Peak 9	4q22.1	chr4	90872446	93227353	35.29	1.58	0,00866	*AFF1*	29.88	60,47	0,00
Deletion Peak 10	4q35.2	chr4	188924862	191154276	36.88	3.39	1,88E-08	*FAT1*	35.06	60,47	0,00
Deletion Peak 11	5q12.1	chr5	58140413	59786052	26.92	1.58	2,16E-07	*PLK2*	19.82	59,30	0,00
Deletion Peak 12	6q15	chr6	75396297	116267927	42.53	2.04	1,34E-08	*PNRC1, EPHA7, FOXO3, FYN, SESN1, HDAC2, TMEM30A, BACH2, PRDM1, MAP3K7, WISP3, CCNC, ATG5*	46.65	44,19	0,75
Deletion Peak 13	6q27	chr6	167795078	168376999	43.89	3.17	1,18E-11	*FGFR1OP*	47.26	53,49	0,41
Deletion Peak 14	7q36.2	chr7	151563571	159138663	26.24	0.90	1,43E-05	*MNX1, KMT2C, XRCC2*	28.66	23,26	0,41
Deletion Peak 15	8p21.2	chr8	21164188	24771333	59.05	11.76	1,40E-99	*NKX3-1*	67.68	87,21	0,00
Deletion Peak 16	9p23	chr9	7798746	10628684	44.34	5.43	0,02366	*PTPRD*	51.22	48,84	0,75
Deletion Peak 17	9p21.3	chr9	21965002	21995139	42.99	10.63	3,21E-21	*CDKN2A*	54.27	53,49	0,91
Deletion Peak 18	10q23.31	chr10	89597723	90040374	33.03	3.85	1,73E-09	*PTEN*	32.62	52,33	0,00
Deletion Peak 19	10q26.3	chr10	128201980	135534747	37.33	0.90	0,040054	*MGMT, MKI67*	33.84	54,65	0,00
Deletion Peak 20	11p15.5	chr11	1	219774	44.57	1.81	8,83E-15	*HRAS*	42.68	65,12	0,00
Deletion Peak 21	11q23.3	chr11	108364019	117857394	63.35	1.81	6,98E-46	*DDX10, PAFAH1B2, POU2AF1, USP28, SDHD, ZBTB16, PCSK7*	71.95	46,51	0,00
Deletion Peak 22	11q25	chr11	132206090	135006516	59.50	2.94	3,20E-08	*OPCML*	67.68	50,00	0,01
Deletion Peak 23	12p13.1	chr12	12796998	12956397	27.60	2.04	1,25E-08	*CDKN1B*	28.96	32,56	0,64
Deletion Peak 24	12q23.1	chr12	99119111	100435321	19.68	1.13	0,00091601	*IGF1*	13.72	45,35	0,00
Deletion Peak 25	13q14.2	chr13	48832931	49065950	58.60	3.39	2,93E-08	*RB1*	62.20	65,12	0,74

Supplementary Table S5 - Continued

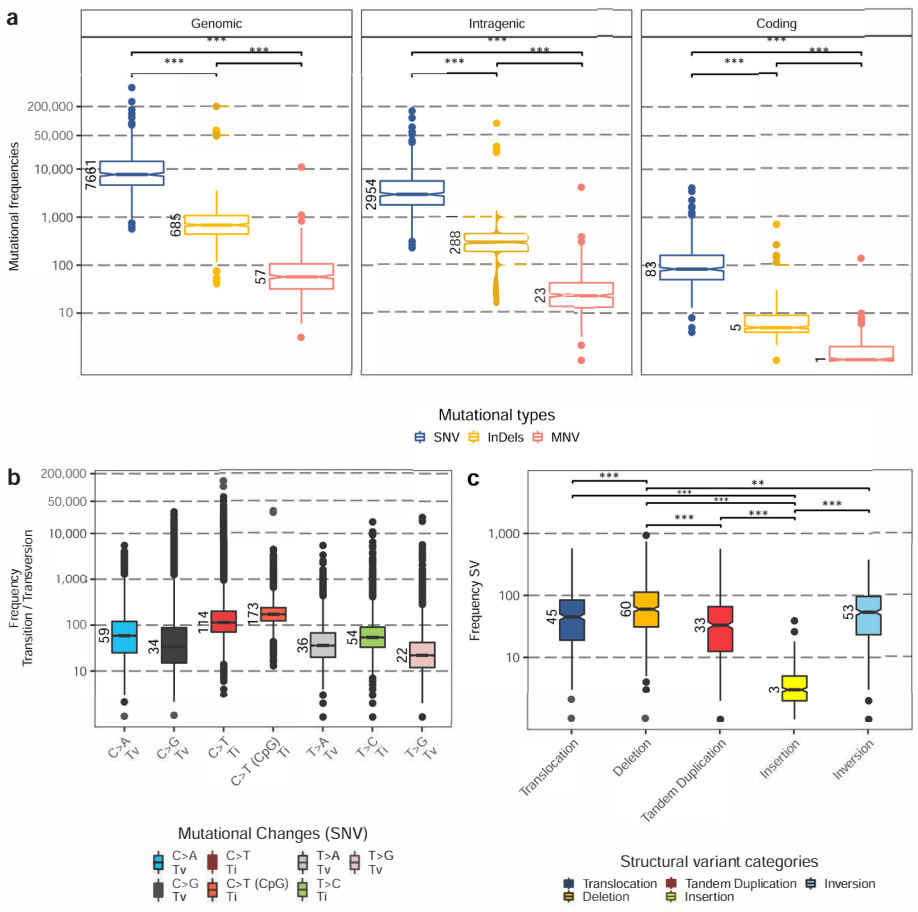
Unique Name	Descriptor	Chromosome	start	end	frequency shallow gain (%)	frequency deep gain (%)	Residual q values after removing segments shared with higher peaks*	Overlapping Genes	frequency ER+ (%) N=328	frequency ER- (%) N=86	FDR**
Deletion Peak 26	14q31.1	chr14	63509918	102025696	37.33	0.90	0,00053469	*GPHN, DICER1, TCL6, MLH3, MAX, YLPM1, RAD51B, BCL11B, ZFP36L1, TSHR, TCL1A, SETD3, TRIP11, GOLGA5*	35.37	55,81	0,00
Deletion Peak 27	16q23.1	chr16	78063174	79635015	56.79	0.90	1,34E-11	*MAF, WWOX*	60.06	50,00	0,15
Deletion Peak 28	17p12	chr17	11898769	12460718	69.23	5.20	3,07E-51	*MAP2K4*	74.39	75,58	0,90
Deletion Peak 29	17q11.2	chr17	29324421	29725411	40.50	1.13	0,013501	*NF1*	37.80	55,81	0,01
Deletion Peak 30	18p11.31	chr18	3475383	8361339	39.82	2.26	0,0038007	*TGIF1*	43.29	38,37	0,53
Deletion Peak 31	18q23	chr18	47719488	78077248	44.57	2.26	4,03E-07	*MALT1, DCC, KDSR, SMAD4, PMAIP1, BCL2, SERPINB3, TNFRSF11A*	44.51	56,98	0,08
Deletion Peak 32	19p13.3	chr19	972151	2037529	62.22	1.58	1,32E-26	*ARID3A, DAZAP1, STK11, TCF3*	61.28	73,26	0,06
Deletion Peak 33	19q13.43	chr19	59068208	59128983	32.58	1.81	0,001086	*U2AF2*	29.57	47,67	0,01
Deletion Peak 34	20p12.1	chr20	13956349	16038662	21.95	2.49	1,62E-08	*CRNKL1*	22.26	36,05	0,02
Deletion Peak 35	22q13.33	chr22	50446002	51304566	54.52	1.36	0,003135	*NA*	58.84	47,67	0,09
Deletion Peak 36	Xp22.33	chrX	1	592349	42.76	3.39	0,064028	*PIGA*	39.94	67,44	0,00
Deletion Peak 37	Xq23	chrX	80547132	115570202	44.80	3.62	0,00063416	*PAK3, BTK, IRS4*	43.29	63,95	0,00

* GISTIC2.0 derived Benjamini-Hochberg FDR-corrected permuted probability value

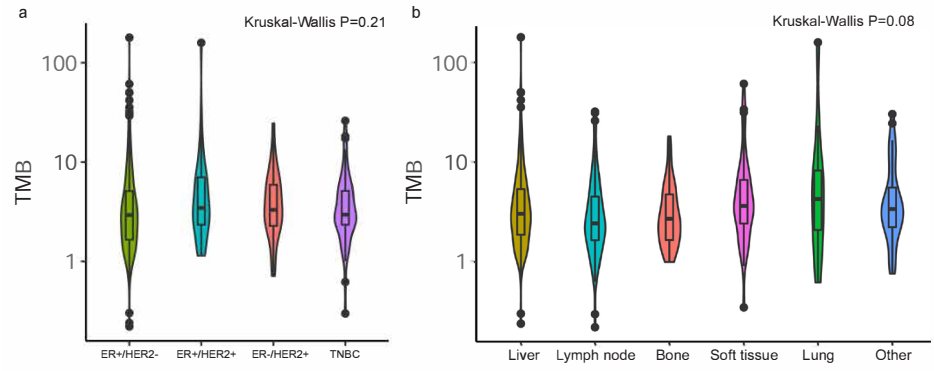
** Hochberg corrected FDR of a two-sided Fisher's exact-test

Supplementary Table S6 - Actionable alterations according to OncoKB (July 12, 2018)

	N	%
Total number of patients with actionable alterations	105	23.70
On label (N=67)		
ERBB2 amplification - anti-HER2 therapies (lapatinib. trastuzumab)	67	15.00
Off label (N=47)		
BRAF p.V600E - BRAF inhibitors (vemurafenib. dabrafenib)	2	0.45
BRCA1 mutation - PARP inhibitors (niraparib. rucaparib)	4	0.90
BRCA2 mutation - PARP inhibitors (niraparib. rucaparib)	8	1.81
CDK4 amplification - CKD4/6 inhibitors (Abemaciclib. Palbociclib)	19	4.30
IDH1 mutation - Ivosidenib	1	0.23
KIT mutation - Tyrosine kinase inhibitors (Sorafenib. sunitinib)	2	0.45
MET amplification - Crizotinib. cabozantinib	8	1.81
PDGFRA mutation - Imatinib	4	0.90
TSC1 mutation - Everolimus	1	0.23
TSC2 mutation - Everolimus	2	0.45



Supplementary Figure 1 - Overview of somatic characteristics.
(a) Number of Single Nucleotide Variant (SNV; blue), InDels (orange) and Multi NV (red) per wholegenome sequenced sample over three resolutions; genome-wide, within intragenic regions and within coding regions (n=442 patients).
(b) Type of genome-wide SNVs. Transitions (Ti) and transversion (Tv) (n=442 patients).
(c) Frequency of structural variation such as translocations (n=440), deletions (n=442), tandem duplication (n=439), insertions (n=343) and inversions (n=438).
(a,b,c) The box is bounded by the 25th and 75th percentile, with the horizontal line in the box depicting the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentile.
Outliers lie >1.5 IQR beyond either end of the box. Statistical significance: two-sided Mann-Whitney U test (FDR corrected): * P < 0.05, ** P < 0.01, *** P < 0.001.

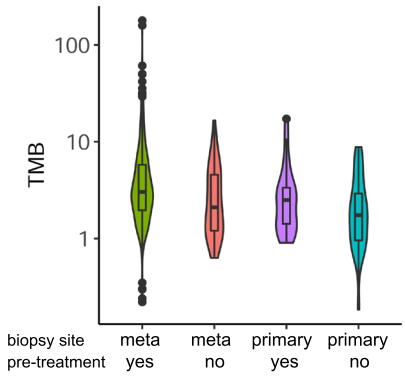


Supplementary Figure 2 - Tumour mutational burden (TMB) is equal between of breast cancer subtypes and biopsy sites.

(a) Violin plot showing the distribution of TMB on a log-scale per breast cancer subtype. Black dash indicates median value. ER+/HER2- (n=279), ER+/HER2+ (n=49), ER-/HER2+(n=28), TNBC (n=58).

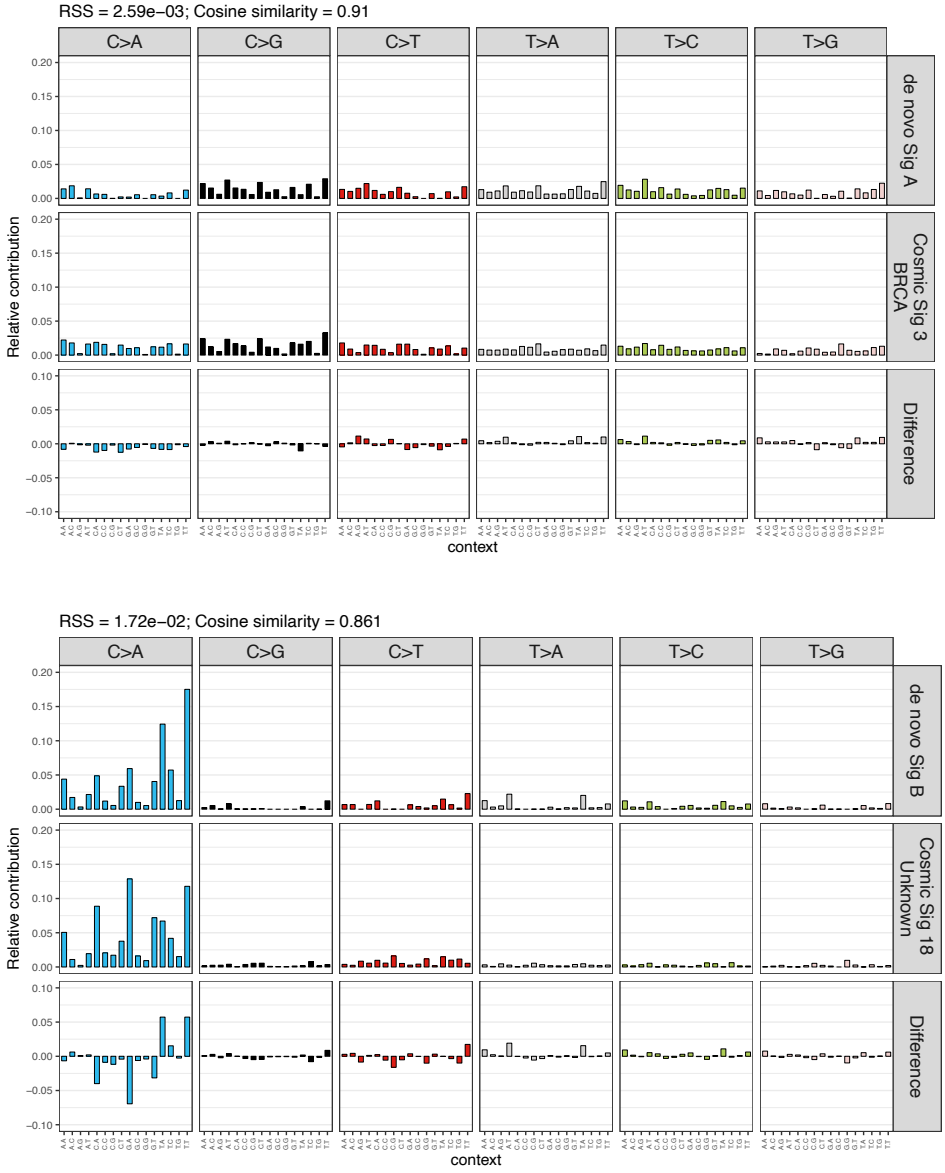
(b) Violin plot showing the distribution of TMB on a log-scale per biopsy site. Liver (n=199), Lymph node (n=94), Bone (n=50), Soft tissue (n=48), Lung (n=12), Other (n=18).

(a,b) Combined violin plot/box plot where the violin-width depicts the density and the length the range of data. The box is bounded by the 25th and 75th percentile, with the horizontal line in the box depicting the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentile. Outliers lie >1.5 IQR beyond either end of the box.

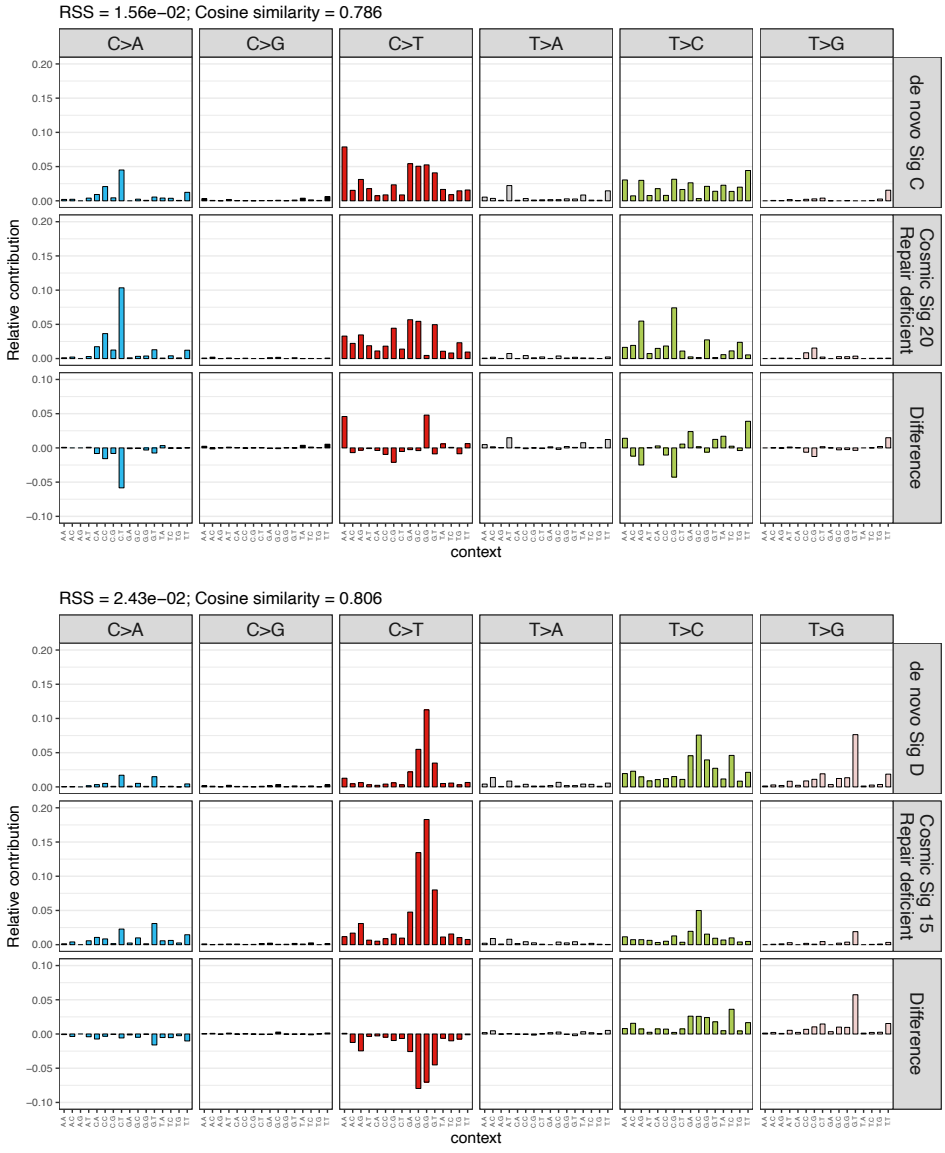


Supplementary Figure 3 - Pre-treatment and metastatic tissue are both associated with TMB.

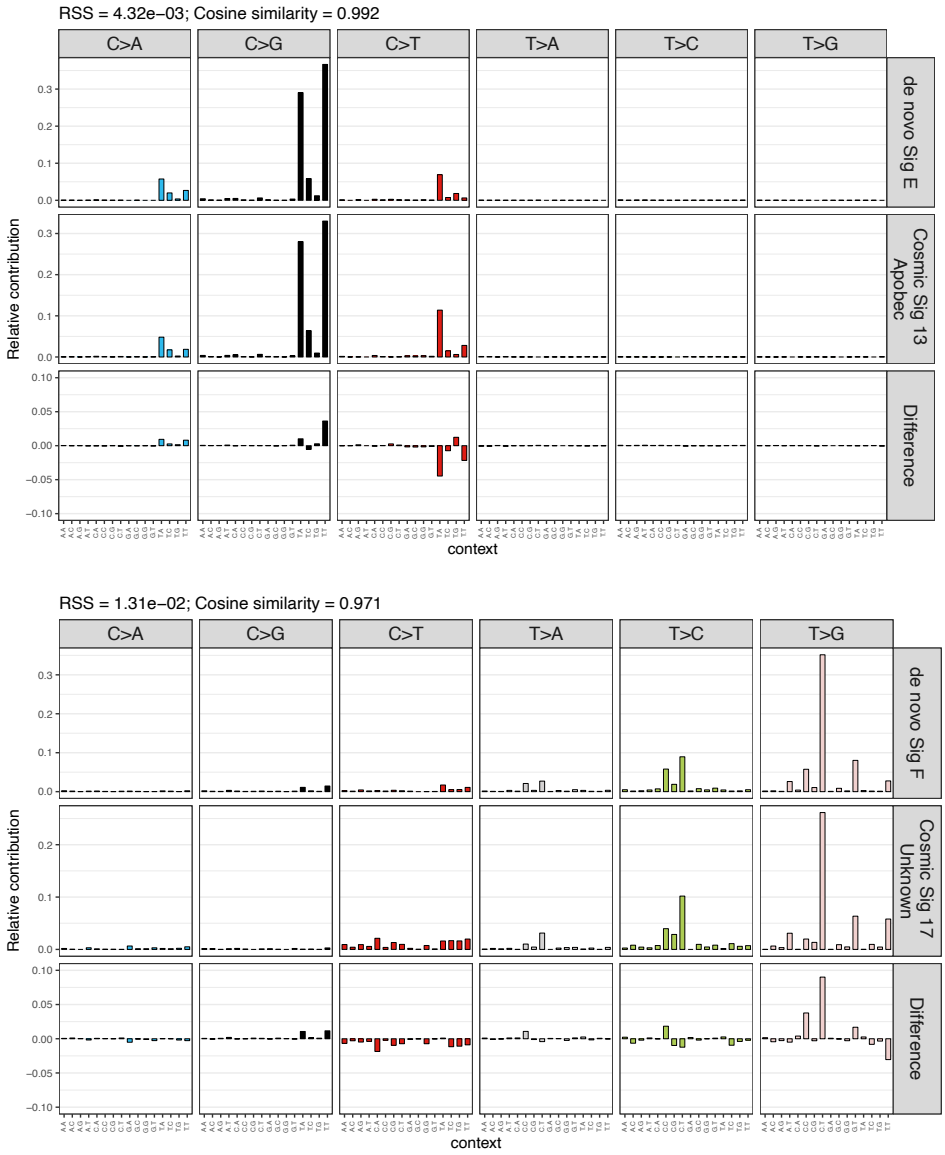
Violin plot showing the distribution of TMB on a log-scale per biopsy location (metastasis or primary tumour) and prior treatment (yes or no). Metastatic biopsy site (received pre-treatment n=349; received no pre-treatment n=56); primary biopsy site (received pre-treatment n=23; received no pre-treatment n=30). Combined violin plot/box plot where the violin-width depicts the density and length the range of data. The box is bounded by the 25th and 75th percentile, with the horizontal line in the box depicting the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentile. Outliers lie >1.5 IQR beyond either end of the box.



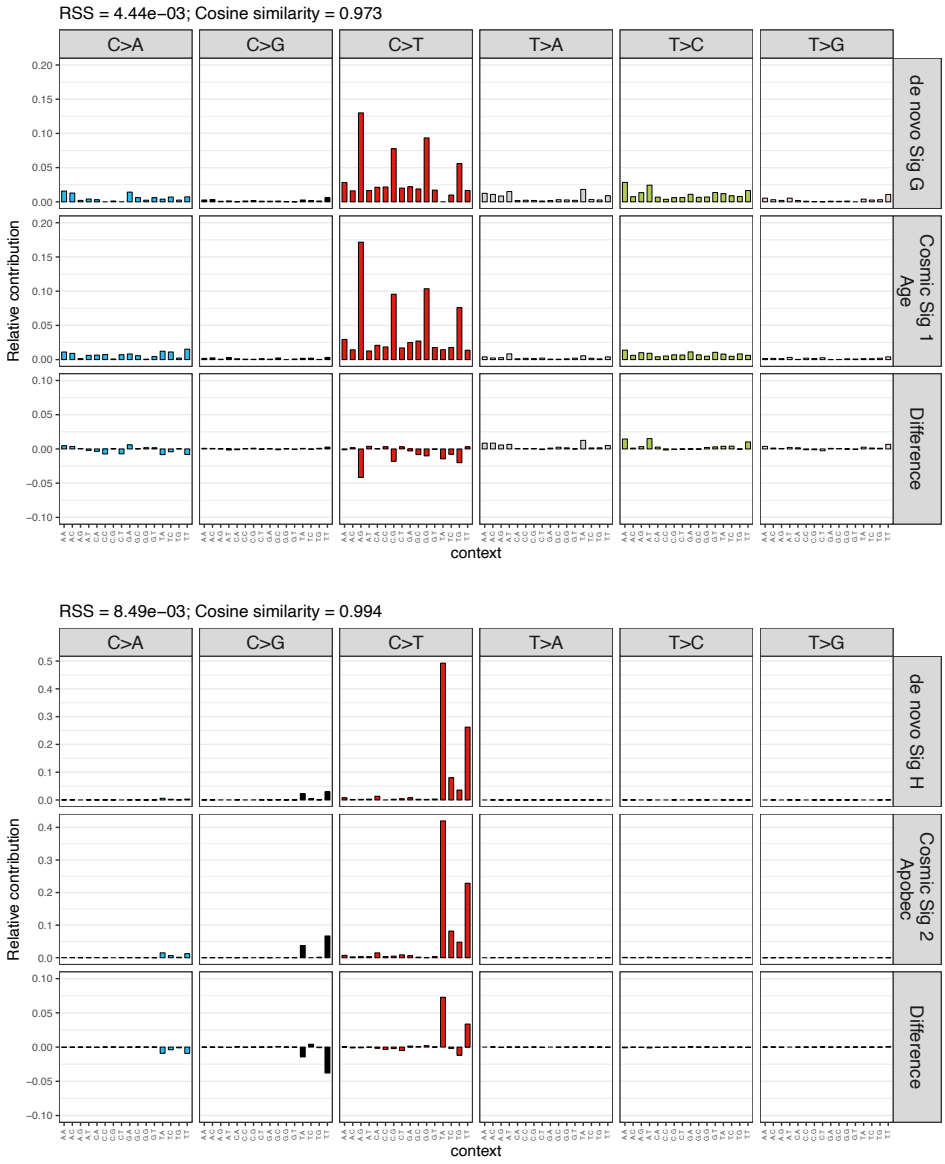
Supplementary Figure 4 - Mutational spectrum of de novo signatures and Cosmic signatures with the highest cosine similarity.



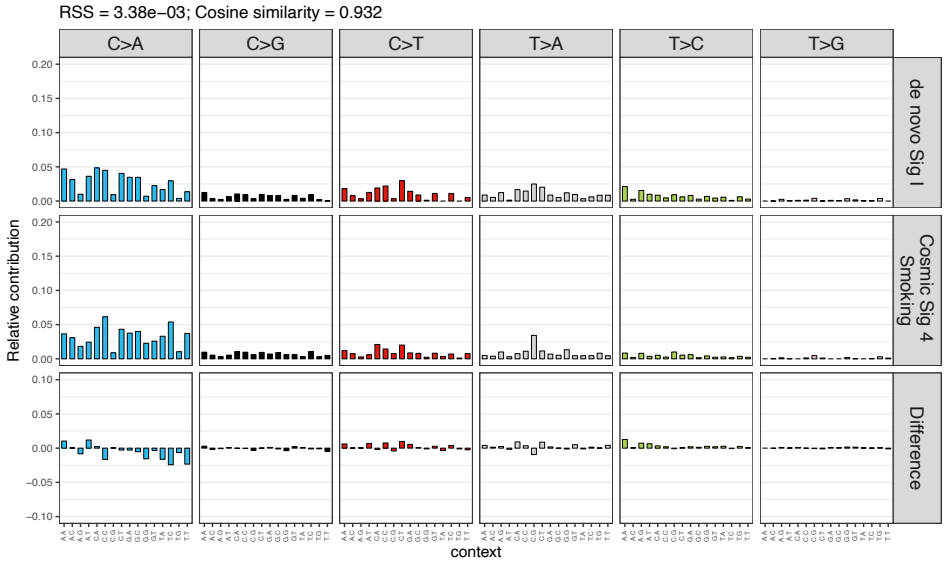
Supplementary Figure 4 - Continued

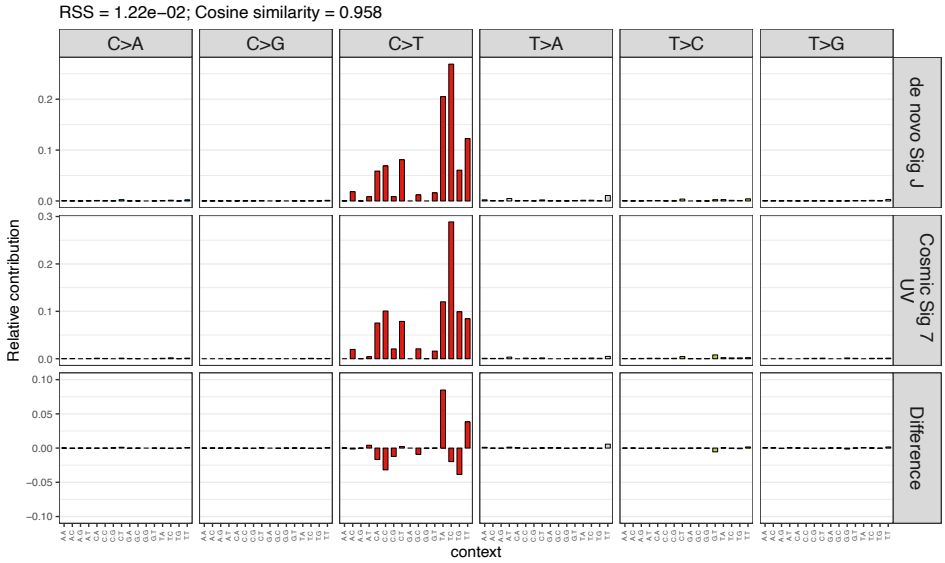


Supplementary Figure 4 - Continued

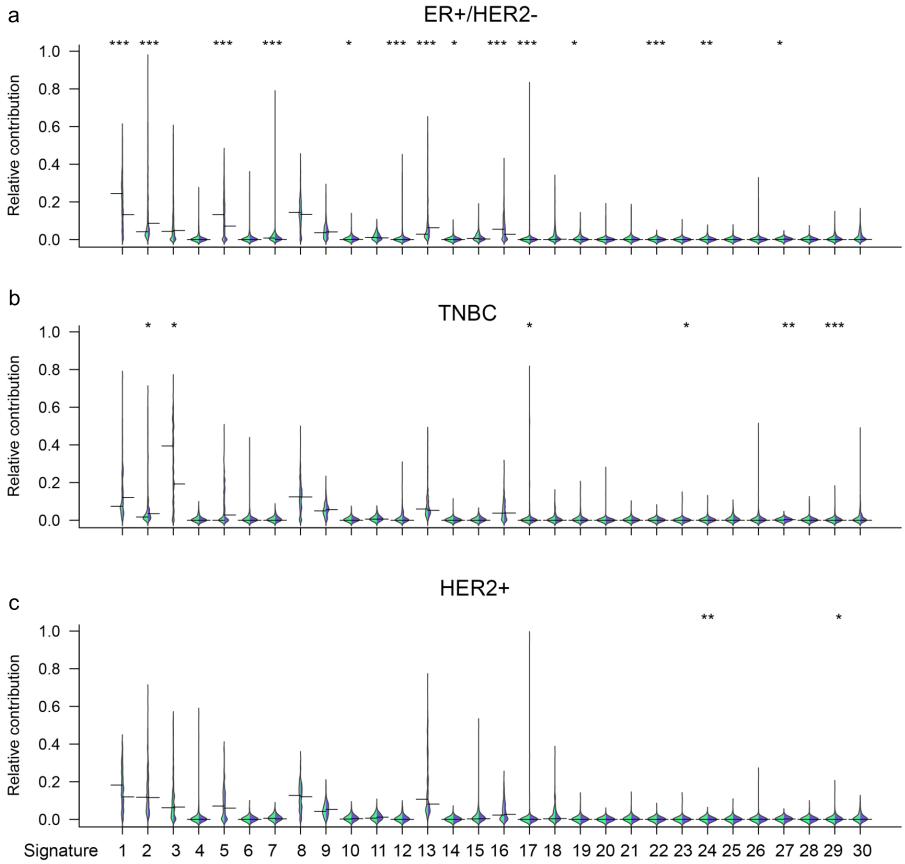


Supplementary Figure 4 - Continued





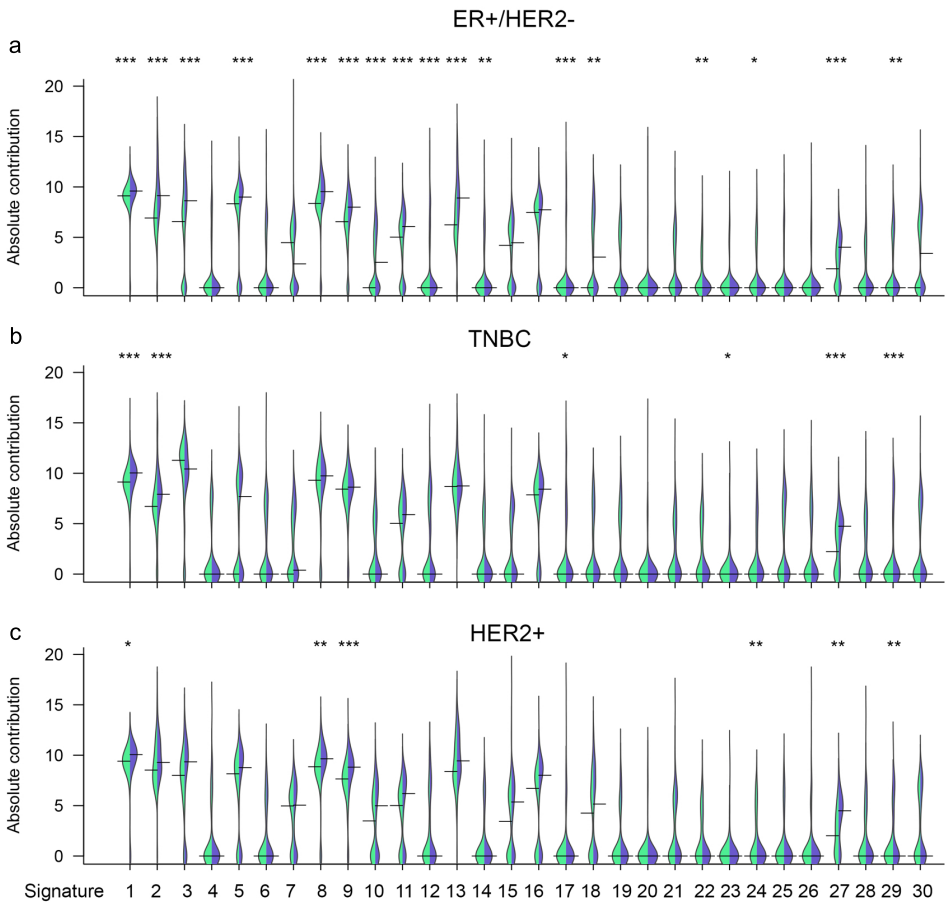
Supplementary Figure 4 - Continued



Supplementary Figure 5 - Relative contributions of all 30 Cosmic mutational signatures: metastatic breast cancer versus primary breast cancer.

Bean plots showing the relative contribution of all 30 Cosmic signatures. Relative contributions were compared between metastatic breast cancer and primary breast cancer per subtype: ER+/HER2- (a) (BASIS n=320; CPCT n=279), TNBC (b) (BASIS n=167; CPCT n=58), HER2+ (c) (BASIS n=73; CPCT n=77). Per graph, left of centre (green) indicates the distribution of primary tumours from the BASIS cohort, right of centre (purple) metastatic biopsy.

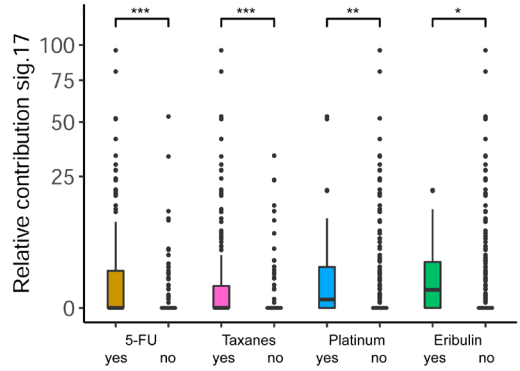
(a,b,c) The width of the bean plot depicts the density of the observations in each group, the horizontal line shows the median. The length of the bean plot shows the full range of observations. Statistical significance: two-sided Mann-Whitney U (FDR corrected): * P < 0.05, ** P
< 0.01, *** P < 0.001.



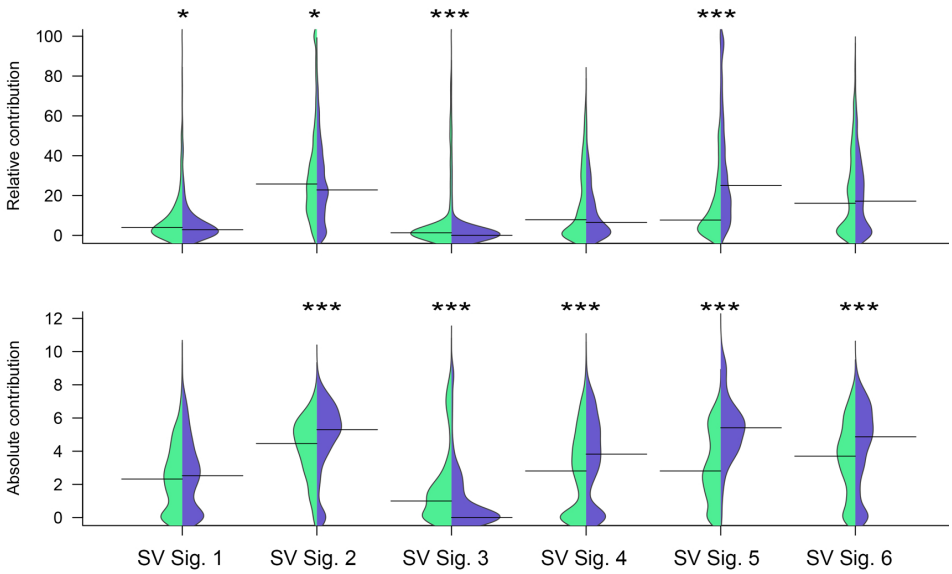
Supplementary Figure 6 - Absolute numbers of mutations of all 30 Cosmic mutational signatures: metastatic breast cancer versus primary breast cancer.

Bean plots showing the absolute numbers of mutations of all 30 Cosmic signatures. Absolute numbers were compared between metastatic breast cancer and primary beast cancer per subtype: ER+/HER2- (a) (BASIS n=320; CPCT n=279), TNBC (b) (BASIS n=167; CPCT n=58), HER2+ (c) (BASIS n=73; CPCT n=77). Per graph, left of centre (green) indicates the distribution of primary tumours from the BASIS cohort, right of centre (purple) metastatic biopsy.

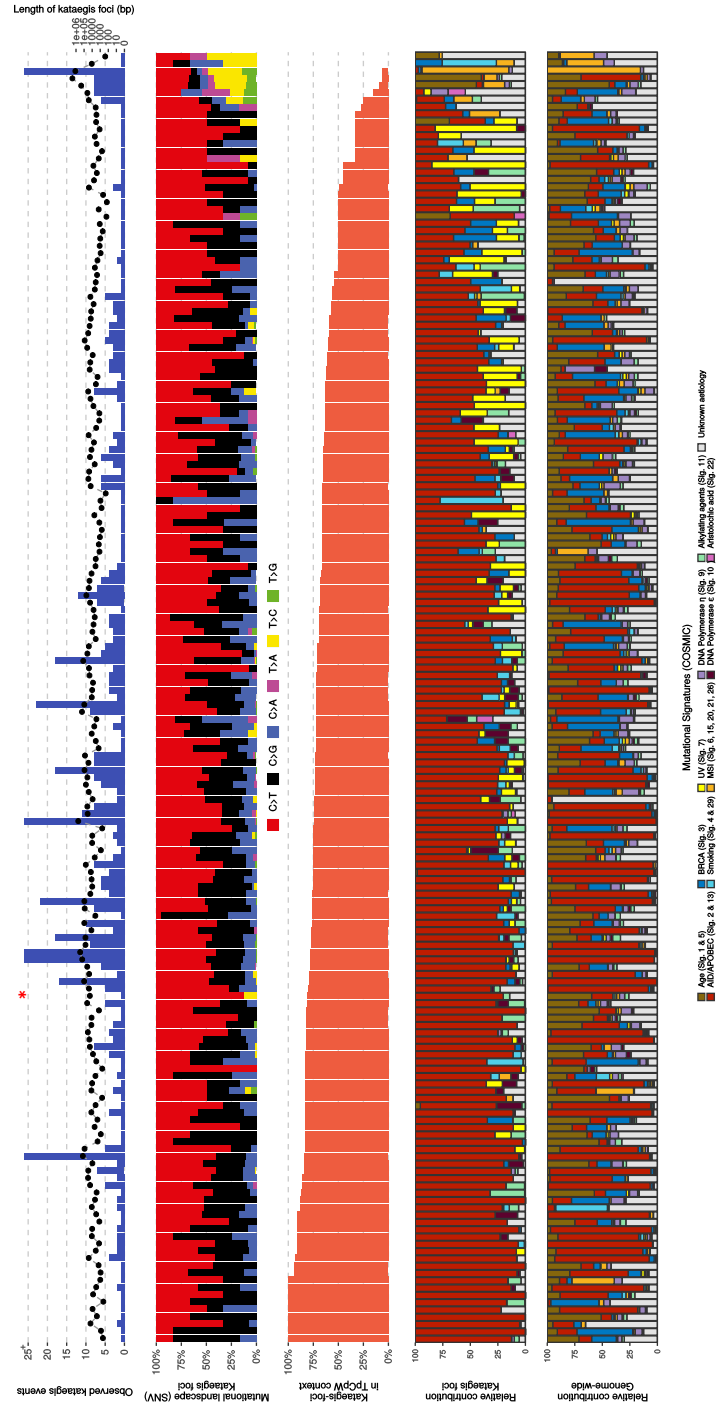
(a,b,c) The width of the bean plot depicts the density of the observations in each group, the horizontal line shows the median. The length of the bean plot shows the full range of observations. Statistical significance: two-sided Mann-Whitney U (FDR corrected) * P < 0.05, ** P < 0.01, *** P < 0.001

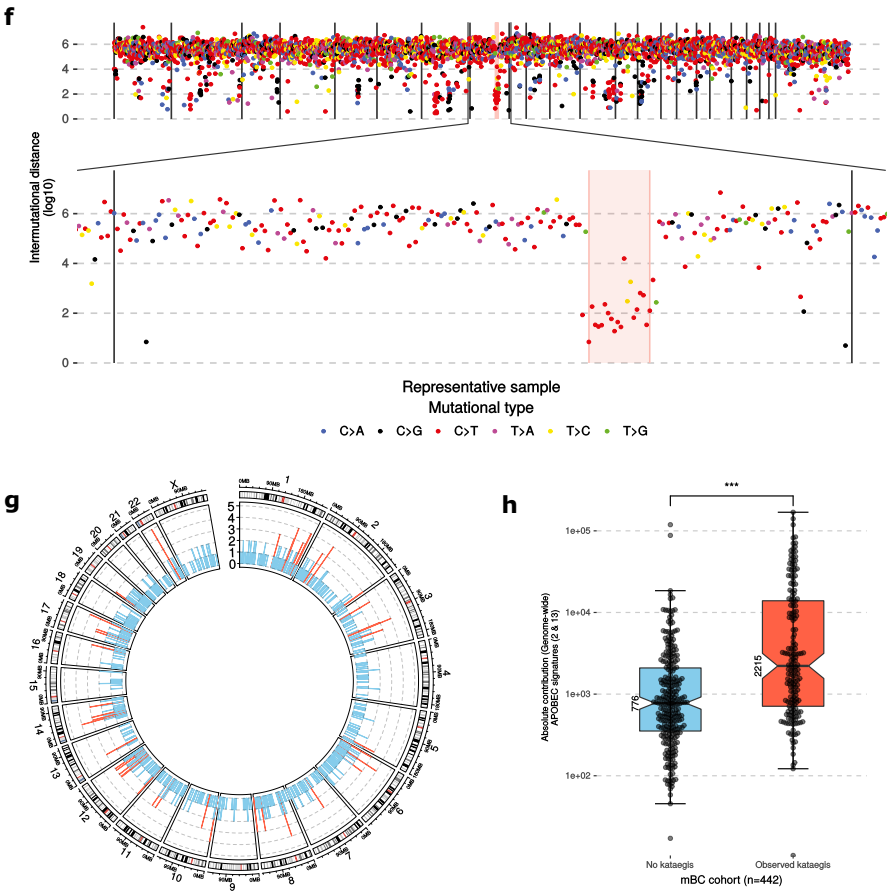


Supplementary Figure 7 - Signature 17 and its association with pre-treatment.
Boxplots showing the distribution of relative contribution of mutational signature 17 and its association with pre-treatments as 5-FU (yes n=170; no n=258), taxanes (yes n=247; no n=181), platinum-based chemotherapy (yes n=43; no n=385) and eribulin (yes n=14; no n=414). Y-axis in square root-scale. The box is bounded by the 25th and 75th percentile, with the horizontal line in the box depicting the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentile. Outliers lie >1.5 IQR beyond either end of the box. Statistical significance: two-sided Mann-Whitney U (FDR corrected) * P < 0.05, ** P < 0.01, *** P < 0.001.



Supplementary Figure 8 - Relative contributions of six rearrangement signatures: metastatic breast cancer versus primary breast cancer.
Bean plots showing the relative contribution of 6 rearrangement signatures. Relative contributions were compared between metastatic breast cancer (n=442) and primary breast cancer (n=560). The coloured dot in the Tufte plot indicates the median value, the lines show the range of observations. Per graph, left of centre (green) indicate the distribution of primary tumours from the BASIS cohort, right of centre (purple) metastatic biopsy. The width of the bean plot depicts the density of the observations in each group, the horizontal line shows the median. The length of the bean plot shows the full range of observations. Statistical significance: two-sided Mann-Whitney U (FDR corrected) * P < 0.05, ** P < 0.01, *** P < 0.00

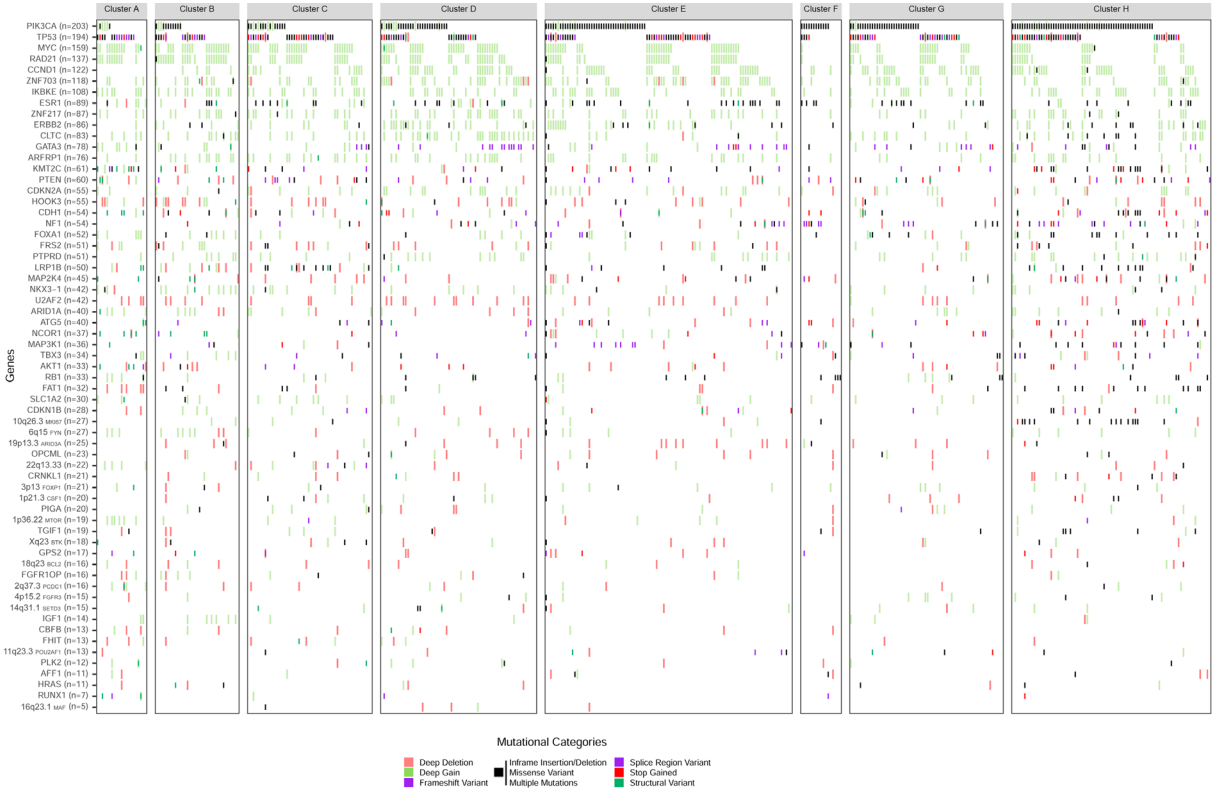

Length of kataegis focl (bp)
1e+06
100000
10000
1000
100
10
0
a
Observed kataegis events
25
20
15
10
5
0
*
b
Mutational landscape (SNV)
Kataegis foci
100%
75%
50%
25%
0%
C>T
C>G
C>A
T>A
T>C
T>G
c
Kataegis-foci
in TpCpW context
d
Relative contribution
Kataegis foci
e
Relative contribution
Genome-wide
Mutational Signatures (COSMIC)
Age (Sig. 1 & 5)
APOBEC (Sig. 2 & 13)
BRCA (Sig. 3)
Smoking (Sig. 4 & 29)
UV (Sig. 7)
MSI (Sig. 6, 15, 20, 21, 26)
DNA Polymerase η (Sig. 9)
DNA Polymerase ε (Sig. 10)
Alkylating agents (Sig. 11)
Aristocholic acid (Sig. 22)
Unknown aetiology



Supplementary Figure 9 - Kataegis prevalence.
(a) Number of observed kataegis events in CPCT-02 cohort samples (n = 177, blue bars) and the respective genomic width of all observed kataegis foci per sample (right y-axis; black points). The sample that is denoted with an asterisk is shown in more detail in panel f.
(b) Relative frequency of mutational contexts (of SNV) found in all observed kataegis foci per sample.
(c) Relative frequency of SNV in observed kataegis foci in APOBEC-related TpCpW mutational context. W stands for T or A.
(d) Relative contribution to mutational signatures (COSMIC) within the kataegis foci.
(e) Relative contribution to mutational signatures (COSMIC) of all genome-wide events of the sample.
(f) Representation of one distinct kataegis foci on chromosome 8 within a single respective sample (highlighted with *). SNV (colored on Ti/Tv type) are shown with relative genomic distances (in log10) to neighboring SNV. Observed kataegis foci are highlighted with a transparent red background.
(g) Frequency and locations of cohort-wide observed kataegis foci, binned per 1Mb. Bins with >1 kataegis events in distinct samples are coloured red, else blue.
(h) Absolute contribution of APOBEC signatures (2 & 13) in samples without (n = 234) and with observed kataegis (n = 177). The box is bounded by the 25th and 75th percentile, with the horizontal line in the box depicting the median. The notch in the box displays the 95% confidence interval of the median. The whiskers extend to 1.5 of the IQR above the 75th and below the 25th percentile. Outliers lie >1.5 IQR beyond either end of the box. Statistical significance: two-sided Mann-Whitney U (FDR corrected) * P < 0.05, ** P < 0.01, *** P < 0.001.

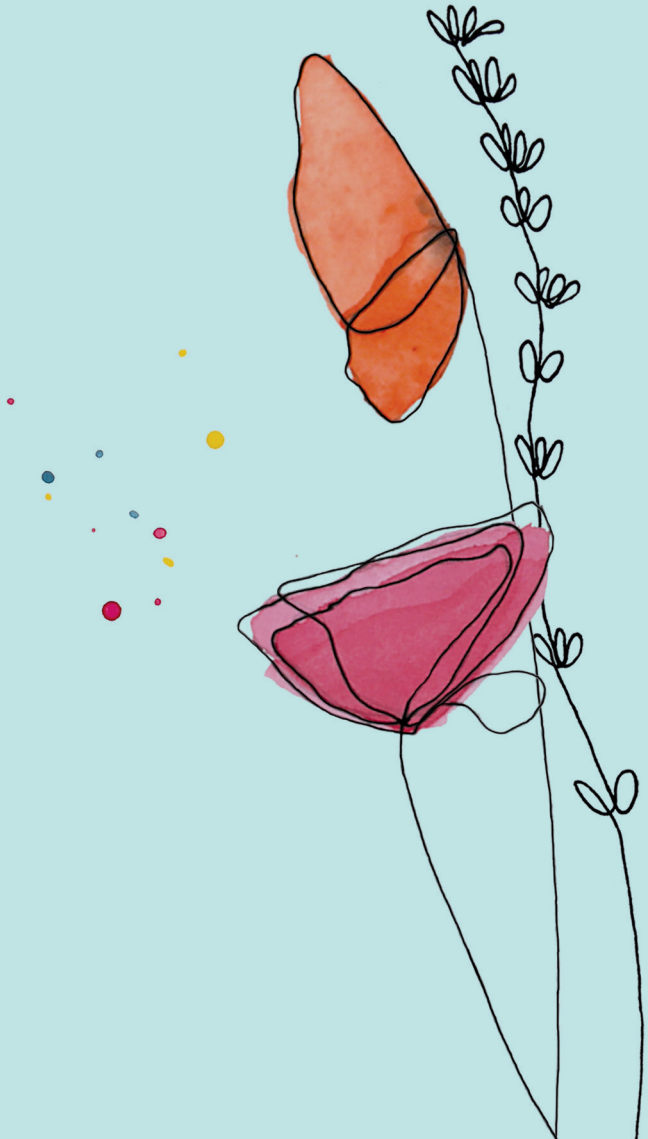
Supplementary Figure 10 A.

Supplementary Figure 10 B.



Supplementary Figure 10 - Oncoplot.

Based on dN/dScv ($q \leq 0.01$) and GISTIC2 focal peak ($q \leq 0.1$) criteria, we show the genes and focal genomic regions which are recurrently affected in our metastatic breast cancer cohort of 442 patients per breast cancer subtype (a) and per cluster (b) identified by our unsupervised clustering.



CHAPTER 3

Whole genome sequencing of metastatic tissue reveals clinical and molecular variables associated with capecitabine response in patients with breast cancer

In Preparation

Lindsay Angus, Saskia M. Wilting, Marcel Smid, Kazem Nasserinejad, Neeltje Steeghs, C. Willemien Menke - van der Houven van Oordt, Vivianne C.G. Tjan-Heijnen, Laurens V. Beerepoot, Joan B. Heijns, Emile E. Voest, Edwin Cuppen, Martijn P. Lolkema, John W.M. Martens, Agnes Jager, Stefan Sleijfer

Abstract

Purpose

Capecitabine is frequently used for the treatment of patients with metastatic breast cancer. Although the toxicity profile and response rates are relatively favorable, many patients will not benefit from capecitabine stressing the need for factors distinguishing responders from non-responders. This study aimed to identify clinical and genomic variables associated with response to capecitabine monotherapy.

Patients and Methods

Patients who underwent a tumor biopsy prior to capecitabine monotherapy were included (CPCT-02 study, NCT01855477). Clinical and genomic variables, as derived from whole genome sequencing (WGS), were associated with response to therapy. Univariable associations were assessed by ordered logistic regression in which best response was defined as ordinal dependent variable: partial response (PR), stable disease (SD), and progressive disease (PD). An ordered least absolute shrinkage and selection operator (LASSO)-based logistic regression was performed for multivariable analysis.

Results

In total, 73 capecitabine treated patients were included of which 18 patients achieved a PR, 28 patients SD and 27 patients PD. In univariable analysis, 27 variables were associated with a worse response category. From these, 14 remained significant after multivariable analysis, amongst which were ER-negative status (odds ratio (OR) 1.45), mutations in *TP53* (OR 1.91) and *PTPRS* (OR 2.63), wild type *CEP350* (OR 0.30), *HMCN1* (OR 0.87), and *ADGRG4* (OR 0.79), amplification of 17q23.1 (OR 1.60), loss of 4p16.3 (OR 1.16) and higher numbers of mutations related to mutational signature 16 (OR 1.08) and rearrangement signature 1 (OR 1.04).

Conclusions

This study illustrates that the use of WGS can identify genomic profiles in subgroups of patients with or without response from capecitabine, which need further validation in an independent cohort.

Introduction

For patients with metastatic breast cancer only palliative treatment is available. One of the agents frequently used in this setting is single agent capecitabine. Capecitabine is an orally administered fluoropyrimidine which is preferentially converted into its active metabolite, 5-fluorouracil (5-FU), in tumor tissue and only partly in normal tissues[1,2]. The reported objective response rates for capecitabine monotherapy range from 9-36% and disease stabilization occurs in another 23-46% of patients[3-9]. Capecitabine has a relatively favorable toxicity profile, however, still around 30% of patients suffer from grade 3 and 4 adverse events, with hand-foot syndrome (10-24%), gastrointestinal side effects (8-16%) and neutropenia (6-7%) being the most frequently observed ones[4-6,9]. Although the toxicity profile and response rates are relatively favorable, many patients will not benefit from capecitabine stressing the need for factors identifying which patients will benefit and who will not. Clinical parameters that have previously been associated with improved objective responses to capecitabine were first line of therapy, absence of lung metastasis, predominance of visceral metastasis, and estrogen receptor (ER) and/or progesterone receptor (PR) positivity[5]. Also, prior treatment with anthracyclines has been associated with a worse overall survival on capecitabine[5].

In addition to these clinical-pathological factors, several reports have described associations between genomic alterations – both somatic and germline – and response to capecitabine. Pharmacogenetic studies have identified predictive single nucleotide polymorphisms (SNP) in genes encoding for enzymes which play a role in the conversion of capecitabine into 5-FU[10,11]. SNPs in genes involved in 5-FU metabolism can lead to increased enzyme activity, leading to higher 5-FU levels and consequently better response rates[10,12-14]. Contrarily, overexpression of astrocyte elevated gene-1 (*AEG-1*) has been linked to 5-FU resistance in hepatocellular carcinoma by enhancing the expression of dihydropyrimidine dehydrogenase (DPYD) that catalyzes the conversion from 5-FU into its inactive metabolite[15].

Besides genes involved in 5-FU metabolism, several other genes have been suggested to play a role in 5-FU sensitivity or resistance, mainly investigated in colorectal cancer cell line models. *In vitro* studies have shown that in response to genotoxic stresses such as 5-FU, wild type p53 is induced. This induction leads to p53-mediated cell cycle arrest and/or induction of apoptosis and suggests that wild type p53 could be important factor in obtaining tumor cell death and consequently to response[16,17]. In addition to wild type p53, low-level *MYC* amplifications in colorectal cancer patients have been associated with improved disease free survival when treated with 5-FU-based adjuvant therapy[18].

Furthermore, the interaction between *MYC* and p53 has been studied as well, in vitro as well as in patients. In colorectal cancer, patients with tumors showing both amplified *MYC* and wild type *TP53* benefitted most from 5-FU based adjuvant therapy[19]. Moreover, *TP53* wild type colorectal cells resistant to 5-FU, it has been shown that activation of the Wnt pathway led to suppression of *CHK1* which subsequently led to lower levels of p53 which resulted in cell survival[20]. So, wild type p53 seems to be an important factor for obtaining responses on 5-FU-based therapies.

Although analyses have been performed based on single genes or combinations of few genes – mostly involving *in vitro* studies – a genome wide approach to predict response on capecitabine monotherapy in patients with metastatic breast cancer has never been performed before and might reveal genomic predictors for response. Here, we report a cohort of 73 patients with metastatic breast cancer treated with capecitabine monotherapy in a real-life setting. All patients underwent a tumor biopsy prior to capecitabine monotherapy to correlate their tumor genome as investigated by whole genome sequencing (WGS) with subsequent response to this therapy.

Methods

Patient cohort and study procedures

For our analyses, we selected patients who were included under the CPCT-02 biopsy protocol (ClinicalTrial.gov no; NCT01855477), which was approved by the medical ethics committee of the University Medical Center Utrecht, the Netherlands. A detailed description of the consortium[21] and the breast cancer cohort has been described in detail recently[22]. In short, patients ≥18-years-old with incurable locally advanced or metastatic solid tumors, from whom a histological biopsy could be safely obtained and systemic treatment with anticancer agents was indicated, were eligible for inclusion. For genomic characterization, tissue biopsies and matched normal (leukocytes) were subsequently whole genome sequenced at the Hartwig Medical Foundation. All patients gave written informed consent before any study procedure. For this study, we selected patients with metastatic breast cancer who received capecitabine monotherapy directly after the tissue biopsy. Clinical parameters that were available included gender, birth year, breast cancer subtype, biopsy date, biopsy location, type of prior systemic treatment, prior radiotherapy, treatment after the tissue biopsy with start and stop dates and response assessments every 8-12 weeks.

Treatment outcome

Tumor responses were evaluated every 8-12 weeks, according RECIST version 1.1(ref.[23]). To relate outcome to genomic and clinical data, we used best overall response: complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD). For patients without a response assessment according to RECIST, treatment duration was used as a surrogate for outcome. In these patients, "clinical benefit" was defined as a treatment duration of ≥12 weeks and "no clinical benefit" as a treatment duration of <12 weeks. For the response analyses, patients with clinical benefit were pooled with the "stable disease" group and patients without clinical benefit with the "progressive disease" group. Patients of whom no information of response or treatment duration was available, were excluded from the analysis (**Supplementary Figure S1**). Objective response was defined as having CR or PR as best response.

Somatic mutation calling

Detailed methods on calling somatic single nucleotide variants (SNVs), multiple nucleotide variants (MNVs) and structural variants (SVs) were described previously[21,22]. For mutation analysis versus response to capecitabine monotherapy, we only included genes that were mutated in at least five patients.

Copy number analysis

Copy number analysis was performed as previously described[22,24]. In brief, recurrent and broad copy number alterations were identified with GISTIC v.2.0.23 (ref.[25]) on WGS data from a large cohort of 637 biopsies obtained from patients with metastatic breast cancer. For association with response to capecitabine monotherapy we tested 15 amplification peaks and 39 deletion peaks. These peaks were analyzed as categorical predictors, for deletion peaks either being deleted versus non-deleted and for amplification peaks being amplified versus non-amplified.

Mutational signature and rearrangement signature analysis

Catalogue Of Somatic Mutations In Cancer (COSMIC) mutational signatures were established as previously described using the MutationalPatterns R package v.1.4.2.(ref.[26]). Rearrangement signatures analysis was performed as previously described[27] using Manta v.1.0.3.(ref.[28]). For association with response to treatment we tested all six rearrangement signatures and only those COSMIC mutational signatures which were dominantly present in our patient cohort; i.e. signatures that contributed to ≥10% of the observed mutations in at least 5 patients.

Statistical analyses

First, univariable associations between the clinicogenomic characteristics and best response were assessed by an ordered logistic regression analysis in which best response was defined as ordinal dependent variable (PR, SD and PD). A list of all tested clinicogenomic variables is depicted in **Supplementary Table S1**. To reduce the skewness, continuous genomic variables were log-transformed. Second, for variable selection, variables that were significant in the univariable ordered logistic regression analysis, using a threshold of $P<0.05$, were tested in the multivariable analysis using a penalized logistic regression, i.e. an ordered least absolute shrinkage and selection operator (LASSO)-based logistic regression analysis. This regression method is suitable for datasets with a relatively high number of variables in comparison with the number of cases, and protects against overfitting[29]. In order to choose the most appropriate penalty value (lambda λ) for the LASSO method, a 10-fold cross-validation was performed and λ_{min}, which gives the minimum mean cross-validated error, was selected.

All statistical analyses were performed in R statistical program version 3.6.1. The *MASS* package for the univariable ordered logistic regression, and the *ordinalNet* package for ordered LASSO logistic regression analysis were used. All statistical tests were two-sided and considered statistically significant at $P<0.05$.

Results

Patient characteristics

In total, 73 capecitabine treated patients were included from July 2016 until October 2018. Patient demographics are summarized in **Table 1**. At the time of data extraction (July 2019), 63 patients had progressed, of which 27 patients had died (37%). Of all 73 patients, 18 patients achieved a PR, 28 patients had SD or clinical benefit, 27 patients PD or no clinical benefit. The objective response rate was 25%, which is in line with previous studies[5].

Univariable analysis – ordinal regression

Clinical parameters versus response

To discover factors associated with capecitabine response we first performed univariable analysis. In univariable analysis, ER-negative status was associated with a worse response ($P=0.001$; odds ratio(OR) 6.44; 95% confidence interval (CI): 2.12-19.56) with no objective responses being reported in any of the patients with ER-negative disease. Contrary to previously described in literature, we observed that a

lower number of different anti-tumor therapies was associated with a worse response to therapy (*P*=0.002; OR 0.74; 95% CI: 0.61-0.90) (**Table 2**).

Table 1 - Patient characteristics

	Patients (n = 73)		Specification of prior treatments		
Age	N	%		N	%
Median	56		Aromatase inhibitor	42	57.5
Range	33 - 76		Tamoxifen	49	67.1
Gender			Fulvestrant	19	26.0
Female	73	100	Everolimus	8	11.0
Male	0	0	CDK4/6	13	17.8
Breast cancer subtype			5-FU	15	20.5
ER-positive/HER2-negative	47	64.4	Taxanes	50	68.5
ER-positive/HER2-positive	6	8.2	Platinum/Parp	14	19.2
ER-negative/HER2-positive	0	0	Anthracyclines	55	75.3
Triple negative	18	24.7	Cyclophosphamide	55	75.3
Unknown at time of analysis	2	2.7	Eribulin	2	2.7
Prior systemic therapy			Vinorelbine	3	4.1
Yes			Anti-HER2	10	13.7
Endocrine therapy only	7	9.6			
Chemotherapy only	14	19.2			
Endocrine and chemotherapy	38	52.1			
Endocrine, chemo and anti-HER2	6	8.2			
Endocrine and anti-HER2	2	2.7			
Chemo and anti-HER2	3	4.1			
Nr of lines (median, range)	3 (0 - 9)				
Nr of drugs (median, range)	5 (0 -11)				
No prior treatment	3	4.1			
Prior radiotherapy					
Yes	54	74.0			
No	19	26.0			
Biopsy site					
Liver	38	52.1			
Bone	9	12.3			
Lymph node	8	11.0			
Breast	5	6.8			
Lung	2	2.7			
Soft tissue	8	11.0			
Other (peritoneum)	3	4.1			
Best response					
According to RECIST1.1.					
Partial response	18	24.7			
Stable disease	23	31.5			
Progressive disease	17	23.3			
According to treatment duration					
Clinical benefit (<12 weeks treated)	5	6.8			
No clinical benefit (>12 weeks treated)	10	13.7			

Table 2 - Clinical parameters versus response

Clinical parameter	Status	Response category PR (n=18) N	%	SD (n=28) N	%	PD (n=27) N	%	Ordered logistic regression Univariable P-value	Coefficient	Odds ratio	95% CI odds ratio
ER status (positive vs. negative)	Positive	18	100	22	78.6	15	55.6	0.001	1.86	6.44	2.12-19.56
	Negative	0	0	6	21.4	12	44.4				
Number of prior drugs	Median	7	-	5	-	4	-	0.002	-0.31	0.74	0.61-0.90

Table 3 - Mutation status versus response

Gene	Status	Response category PR (n=18) N	%	SD (n=28) N	%	PD (n=27) N	%	Ordered logistic regression Univariable P-value	Coefficient	Odds ratio	95% CI odds ratio
MUC16	Mutant	6	33.3	4	14.3	1	3.7	0.009	-1.69	0.18	0.05-0.65
(wild type vs. mutant)	Wild type	12	66.7	24	85.7	26	96.3				
TP53	Mutant	4	22.2	10	35.7	16	59.3	0.011	1.18	3.27	1.32-8.11
(wild type vs. mutant)	Wild type	14	77.8	18	64.3	11	40.7				
CEP350	Mutant	4	22.2	1	3.6	0	0	0.015	-2.80	0.06	0.01-0.58
(wild type vs. mutant)	Wild type	14	77.8	27	96.4	27	100				
PTPRS	Mutant	0	0	1	3.6	6	22.2	0.019	2.59	13.4	1.53-117.25
(wild type vs. mutant)	Wild type	18	100	27	96.4	21	77.8				
ADGRG4	Mutant	5	27.8	3	10.7	1	3.7	0.022	-1.61	0.20	0.05-0.79
(wild type vs. mutant)	Wild type	13	72.2	25	89.3	26	96.3				
PIK3CA	Mutant	11	61.1	9	32.1	7	25.9	0.022	-1.05	0.35	0.14-0.87
(wild type vs. mutant)	Wild type	7	38.9	19	67.9	20	74.1				
HMCN1	Mutant	5	27.8	6	21.4	1	3.7	0.029	-1.28	0.28	0.09-0.88
(wild type vs. mutant)	Wild type	13	72.2	22	78.6	26	96.3				
MUC4	Mutant	3	16.7	1	3.6	0	0	0.036	-2.46	0.09	0.01-0.85
(wild type vs. mutant)	Wild type	15	83.3	27	96.4	27	100				
SPHKAP	Mutant	3	16.7	2	7.1	0	0	0.041	-1.86	0.16	0.03-0.93
(wild type vs. mutant)	Wild type	15	83.3	26	92.9	27	100				

Mutation analysis versus response

To investigate whether alterations in specific genes and response to capecitabine were associated, we tested genes that were mutated in at least five patients (n=84). From these genes *MUC16, TP53, CEP350, PTPRS, ADGRG4, PIK3CA, HMCN1, MUC4,* and *SPHKAP,* were associated with response to therapy (**Table 3**). Mutations in *TP53* and *PTPRS* were associated with a higher likelihood of having a worse response. Thirty patients harbored a *TP53* mutation, including 16 out of 27 patients with PD (59%), 10 out of 28 patients with SD (36%) and 4 out of 18 patients with PR (22%). Only seven patients harbored a mutation in *PTPRS,* none of which had PR as best response. On the other hand, *MUC16, CEP350, ADGRG4*, *PIK3CA*, *HMCN1*, *MUC4,* and *SPHKAP* were most frequently mutated in patients with PR as best response and hence wild type status of these genes was significantly associated with a higher likelihood of having a worse response (**Table 3**). For example, 27 patients harbored a mutation in *PIK3CA*, including 11 out of 18 (61%) patients with PR, 9 out of 28 (32%) patients with SD and 7 out of 27 patients with PD (26%). For *CEP350* – mutated in 5 patients – no mutations were observed in patients with PD.

Copy number variation versus response

Copy number regions, identified by GISTIC2.0, were also tested versus best response. Four amplification regions and two deletion regions were associated with response (**Table 4**). Amplifications of 8q22.2, 8q24.21, 20q13.2 and 17q23.1 were associated with having a worse response. For example, 54 patients harbored an amplification of 8q22.2, including 8 out of 18 patients (44%) with PR, 23 out of 28 (82%) patients with SD, and 23 out of 27 (85%) patients with PD. Losses of 8p21.3 and 4p16.3 were also associated with a higher likelihood of having a worse response. Twenty-five patients harbored a loss of 4p16.3, including 2 out of 18 (11%) of patients with PR, 11 out of 28 (39%) patients with SD, and 12 out of 27 (44%) with PD.

Mutational signatures, rearrangement signatures and type of structural variants

To investigate the contribution of mutational signatures and response to capecitabine, we used the 30 COSMIC mutational signatures[30]. Out of the 30 COSMIC signatures, signatures 1 and 5 (age), 2 and 13 (APOBEC), 3 (defective homologous recombination DNA-damage repair), 8 (linked to BRCA deficiency), 9 (linked to polymerase eta), 16 (unknown etiology) and 18 (possible damage by reactive oxygen species), contributed to ≥10% of the observed mutations in at least 5 patients and were selected for our analysis. The most frequently represented signatures in this cohort were signature 8 (66% of patients), signature 1 (63%) and signature 3 (49%). A higher contribution of COSMIC signatures 3 and 16 was associated with worse response; the highest median

absolute numbers of mutations related to these signatures were observed in patients with PD as best response (**Table 5**). Regarding the structural variant (SV) signatures, SV1, which is characterized by large (>100 kb) tandem duplications, had the highest contribution in patients with PD as best response. Besides, SV3, characterized by short (<10 kb) tandem duplications and related to *BRCA1* gene abrogation, also had the highest contribution in patients with PD as best response (**Table 5**). An increasing number of mutations related to SV1 and SV3 was associated with the likelihood of having a worse response (P=0.006; OR 1.33 and P=0.039; OR 1.17). With respect to the type of SV an increasing number of large duplications, large deletions and inversions were associated with a worse response (**Table 5**).

Multivariable penalized (LASSO) ordinal regression

All variables which had a P-value of <0.05 in the univariable ordinal regression analyses were tested in multivariable penalized (LASSO) ordinal regression model. From the 31 univariable variables which were included in the LASSO model, 14 variables remained significant. Variables which remained significant in the LASSO model are shown in **Table 6**. Clinical variables which remained significantly associated with worse response in the multivariable analysis were ER-negative status (OR 1.45), a lower number of prior drugs (OR 0.83). Regarding mutation status, *TP53* (OR 1.91) and *PTPRS* (OR 2.63) mutations remained significantly associated with a worse response, whereas wild type status of *CEP350* (OR 0.30), *HMCN1* (OR 0.87), and *ADGRG4* (OR 0.79) were significantly associated with a worse response. With respect to copy number regions, amplification of 17q23.1 (OR 1.24) and loss of 4p16.3 (OR 1.16) were significantly associated with a worse response. Finally, an increasing number of large duplications (OR 1.24), deletions (OR 1.09), mutations related to COSMIC mutational signature 16 (OR 1.08), and SV signature 1 (OR 1.04) remained significantly associated with worse response.

Table 4 - CNV versus response

Amplification region	Status	Response category PR n=18) N	%	SD (n=28) N	%	PD (n=27) N	%	Ordered logistic regression Univariable P-value	Coefficient	Odds ratio	95% CI odds ratio
8q22.2	Normal	10	55.6	5	17.9	4	14.8	0.005	1.53	4.63	1.61-13.36
(normal vs. gain)	Gain	8	44.4	23	82.1	23	85.2				
8q24.21	Normal	10	55.6	6	21.4	4	14.8	0.005	1.49	4.41	1.58-12.37
(normal vs. gain)	Gain	8	44.4	22	78.6	23	85.2				
20q13.2	Normal	9	50.0	4	14.3	5	18.5	0.031	1.17	3.21	1.11-9.29
(normal vs. gain)	Gain	9	50.0	24	85.7	22	81.5				
17q23.1	Normal	14	77.8	13	46.4	12	44.4	0.048	0.88	2.41	1.01-5.76
(normal vs. gain)	Gain	4	22.2	15	53.6	15	55.6				
Deletion region											
8p21.3	Normal	7	38.9	5	17.9	3	11.1	0.031	1.22	3.37	1.12-10.13
(normal vs. loss)	Loss	11	61.1	23	82.1	24	88.9				
4p16.3	Normal	16	88.9	17	60.7	15	55.6	0.036	0.98	2.66	1.07-6.65
(normal vs. loss)	Loss	2	11.1	11	39.3	12	44.4				

Table 5 - Mutational signatures, rearrangement signatures and SVs versus response

COSMIC mutational signature	Response category PR (n=18) Median number mutations	SD (n=28) Median number mutations	PD (n=27) Median number mutations	Ordered logistic regression Univariable P-value	Coefficient	Odds ratio	95% CI odds ratio
Signature 3	7	938	985	0.004	0.15	1.16	1.05-1.29
Signature 16	0	172	235	0.008	0.13	1.14	1.03-1.25
Rearrangement signature							
SV Signature 1	3	5	7	0.006	0.28	1.33	1.09-1.63
SV Signature 3	0	0	1	0.039	0.16	1.17	1.01-1.36
Structural variants	**Median number SVs**	**Median number SVs**	**Median number SVs**				
Number of large duplications (>10 kb)	13	35	32	0.002	0.69	1.98	1.28-3.06
Number of duplications	18	42	39	0.003	0.71	2.04	1.26-3.29
Number of deletions	37	103	86	0.008	0.70	2.01	1.20-3.36
Number of structural variants	144	336	264	0.020	0.62	1.86	1.10-3.15
Number of large deletions (>10k kb)	20	43	42	0.032	0.57	1.77	1.05-2.98
Number of inversions	35	76	64	0.044	0.46	1.59	1.01-2.49

Table 6 - Variables associated with worse response

			Univariable analysis *Ordinal regression*			Multivariable analysis *Penalized (LASSO) ordinal regression*		
Category	**Variable**	***P*-value**	**Coefficient**	**Standard Error**	**Odds ratio**	**Coefficient**	**Odds ratio**	**Associated with worse response**
Clinical	ER status (positive vs. negative)	0.001	1.86	0.57	6.44	0.36	1.43	ER-negative status
	Number of prior drugs	0.002	-0.31	0.10	0.74	-0.10	0.91	Less prior drugs
	Prior treatment with CDK4/6-inhibitor (no vs. yes)	0.010	-1.57	0.61	0.21	-0.52	0.60	No prior treatment
	Prior treatment with tamoxifen (no vs. yes)	0.012	-1.22	0.48	0.29	.	.	No prior treatment
	Prior treatment with cyclophosphamide (no vs. yes)	0.044	-1.06	0.53	0.34	-0.81	0.45	No prior treatment
	Prior treatment with anthracyclines (no vs. yes)	0.044	-1.06	0.53	0.34	.	.	No prior treatment
Gene	MUC16	0.009	1.69	0.65	0.18	.	.	Wild type
wild-type vs. mutant	TP53	0.011	1.18	0.46	3.27	0.73	2.08	Mutation
	CEP350	0.015	-2.80	1.15	0.06	-1.07	0.34	Wild type
	PTPRS	0.019	2.59	1.11	13.4	1.19	3.27	Mutation
	ADGRG4	0.022	-1.61	0.70	0.20	.	.	Wild type
	PIK3CA	0.025	-1.05	0.47	0.35	.	.	Wild type
	HMCN1	0.029	-1.28	0.59	0.28	-0.18	0.83	Wild type
	MUC4	0.036	-2.46	1.17	0.09	.	.	Wild type
	SPHKAP	0.041	-1.86	0.91	0.16	.	.	Wild type
Regions for CNV	8q22.2	0.005	1.53	0.54	4.63	.	.	Amplification
normal vs. gain	8q24.21	0.005	1.48	0.53	4.41	.	.	Amplification
	17q23.1	0.048	0.88	0.44	2.41	0.43	1.53	Amplification
	20q13.2	0.031	1.17	0.54	3.21	.	.	Amplification
normal vs. loss	4p16.3	0.036	0.98	0.47	2.66	0.18	1.19	Loss
	8p21.3	0.031	1.21	0.56	3.37	.	.	Loss
Structural variant	Number of large duplications (>10 kb)	0.002	0.69	0.22	1.98	0.35	1.42	Higher number
	Number of duplications	0.003	0.71	0.24	2.04	.	.	Higher number
	Number of deletions	0.008	0.70	0.26	2.04	0.04	1.04	Higher number
	Number of structural variants	0.020	0.62	0.27	1.86	.	.	Higher number
	Number of large deletions (>10k kb)	0.032	0.57	0.27	1.77	.	.	Higher number
	Number of inversions	0.044	0.46	0.23	1.59	.	.	Higher number
SV signatures	Structural variant signature 1 (absolute contribution)	0.006	0.28	0.10	1.33	0.02	1.02	Higher number
	Structural variant signature 3 (absolute contribution)	0.039	0.16	0.08	1.17	.	.	Higher number
COSMIC mutational	Signature 3 (absolute contribution)	0.004	0.15	0.05	1.16	.	.	Higher number
signatures	Signature 16 (absolute contribution)	0.008	0.13	0.05	1.14	0.07	1.07	Higher number

Discussion

This is the first study which examined the relation between clinical variables and genomic alterations as found by WGS of metastatic biopsies, taken just before starting capecitabine monotherapy, and response to this therapy in a cohort of 73 patients with metastatic breast cancer. Our multivariable penalized (LASSO) ordinal regression analyses identified factors, which were associated with response.

Amongst the clinical parameters that were associated with a worse response was ER-negative status, which is in line with a pooled analysis of individual patient data from trials investigating the effect of capecitabine monotherapy[5]. Blum *et al*. showed that ER- and/or PR-positive status was associated with improved response, progression free survival and overall survival. Regarding the objective response rate, receptor status (ER and/or PR) was found to be a predictive factor in their analysis: one/both positive vs. both negative, OR 2.03 (95% CI: 1.26-3.26; *P*=0.0036). In addition, in a cohort of 102 patients with metastatic breast cancer treated with capecitabine monotherapy, hormone receptor positivity was associated with a longer time to treatment failure (hazard ratio 0.59; 95% CI: 0.359-0.971; *P*=0.038)[31].

Contrary to a previous report that showed that objective responses were more frequently observed in patients who received capecitabine in first-line than second-line or later-line (second or later vs. first, OR 0.58; 95%CI: 0.37-0.90; *P*=0.0142)[5], we found that a lower number of previously received drugs was associated with a worse response. Of note, the odds ratio for worse response was relatively close to 1 (OR 0.91) indicating a small effect.

Regarding mutations, the link between *TP53* and outcome on 5-FU has been shown previously[32]. In cell line models, the protein expression of p53 increased in reaction to exposure to 5-FU[16]. Since p53 plays a pivotal role in cell cycle arrest and programmed cell death in response to exposure to DNA damaging agents, it has been suggested that wild type p53 is important to activate apoptosis, henceforth leading to response[16,20]. In patients with lymph node positive (N1) colorectal cancer who were treated with adjuvant 5-FU, *TP53* wild type status was associated with better overall survival (5-year overall survival of 81.0% for *TP53* wild type versus 62.0% for *TP53* mutant status)[32]. In addition, Pilat *et al*.[33] showed that patients harboring a *TP53* mutation who were treated with neo-adjuvant fluorouracil-based chemotherapy for colorectal liver metastasis had a worse survival than wild type patients (five-year survival 22% versus 60%, respectively). This effect was not observed in patients who were treated with

surgery alone, suggesting that *TP53* is a predictive biomarker for fluorouracil-based neo-adjuvant chemotherapy. Here we show that in addition to previously described associations with progression free and overall survival, *TP53* mutation status is also associated with worse response to capecitabine monotherapy. Moreover, in a recent study of six patients with ER-positive/HER2-negative metastatic breast cancer who had durable responses on capecitabine (54 to >122 months), none of the patients had a *TP53* mutation. Next to wild type *TP53* status, aberrations in genes involved in homologous recombination (*CHEK2, PALB2, ATM*) and chromatin remodeling (*NCOR1, TET2*) were observed in these patients. Defects in these pathways might have prevented repair of 5-FU-induced DNA damage[34]. In our cohort, these genes were affected in less than five patients and hence not analyzed in relation to capecitabine response. In addition to *TP53* mutations, *PTPRS* mutations were also associated with a worse response. *PTPRS* is a receptor-type protein tyrosine phosphatase and deletions in this gene have been associated with activation of the EGFR/PI3K pathway in head and neck cancers[35]. In colorectal cancer, inactivating mutations in *PTPRS* occur in around 10% of patients and the presence of inactivating mutations in this gene have recently been associated with *RAS/ERK* pathway activation[36], which promotes cell proliferation and reduces apoptosis[37].

Contrary to *TP53* and *PTPRS, CEP350, HMCN1* and *ADGRG4* were more frequently mutated in patients experiencing PR or SD. The *CEP350* protein is predicted to play a role in organizing, binding and anchoring microtubules at the centrosome[38]. In addition to the role at the centrosome, *CEP350* might also play a role in regulating nuclear hormone receptor signaling and recently a tumor suppressive role in melanoma has been described[38]. Of note, in our cohort this gene was mutated in only five patients (all ER-positive) and its potential role in response to capecitabine needs further investigation. Next, *HMCN1* encodes an extracellular protein of the immunoglobulin superfamily. Analysis of TCGA data has shown that *HMCN1* is mutated in 6.1% of primary breast tumors. A high variant allele frequency (>30%) in the TCGA cohort, was associated with a higher number of positive lymph nodes and a shorter overall survival[39]. Although *HMCN1* mutations in primary breast cancer are associated with worse survival, its exact role – in particular in relation with capecitabine response – remains to be elucidated. Moreover, *ADGRG4* encodes an adhesion G protein–coupled receptor and has thus far not been linked with cancer or breast cancer in particular.

With respect to mutational signatures, higher contribution of signature 16 was associated with a worse response. The proposed etiology of this signature is currently unknown, but its existence has been described in hepatocellular carcinoma[30]. This

signature is characterized by T>C mutations which are almost exclusively occurring on the transcribed strand[30]. Assuming these mutations are introduced as a result of defective transcription coupled nucleotide excision repair (NER), defective NER might play a role in resistance to capecitabine. With regard to SV1, the cause of its phenotypic characteristic – large tandem duplications – is currently unknown. It has been described that this signature is mainly observed in patients with triple negative breast cancer and tumors exhibiting *TP53* mutations, having an enrichment of signature 3 and a high homologous recombination deficiency (HRD) index[27]. In our cohort we observed that the median contribution of SV1 was also higher in patients with ER-negative tumors than ER-positive tumors (median 20 versus 5 mutations, $P<0.001$). In addition, in samples with >10% contribution of SV1 versus <10%, we also observed a higher frequency of *TP53* mutations (50% versus 36%, $P=0.183$) and a higher absolute number of signature 3 mutations (median 1201 mutations versus 517 mutations, $P=0.056$). Although, factors such as *TP53* and ER-status are also significantly associated with worse response in the multivariable model, the absolute number of SV1 signature events seems to contribute as well.

Loss of 4p16.3 and amplification of 17q23.1 were both associated with a worse response. In breast cancer, loss of heterozygosity of 4p is more frequently observed in basal-like tumors than other breast cancer subtypes[40]. In a small study of Shivapurkar *et al.*[41] a loss of 4p16.3 – on which several tumor suppressor genes are located – was observed in 22/38 of breast cancer patients. Exploratory analysis (data not shown) of 164 patients with matched WGS and RNA sequencing data revealed that a loss of 4p16.3 was associated with a significantly decreased expression (>1.5 fold) of 33 out of 225 potential transcripts, indicating that the loss of this region indeed has an effect on transcriptomic level. In addition, in the wide peak limits of 17q23.1 seven potential transcripts were located of which 6 were significantly higher expressed in samples with a gain, suggesting that the amplification of this region also affects gene expression.

With respect to SNPs in the *DPYP* gene, only three patients harbored a germline variant in one the four most clinically relevant variants[42] rendering this subgroup too small for statistical analyses. Additionally, as in the Netherlands *DPYD* variants are routinely tested prior to treatment start and dose reductions are applied accordingly, we do not expect that these three patients experienced higher fluoropyrimidine exposure. Best responses for these patients were SD (2x) and PD (1x).

Recently performed mutation analysis of intestinal organoids treated with 5-FU showed a characteristic mutational pattern that is dominated by T>G substitutions in

a CTT context, which had a strong resemblance to COSMIC mutational signature 17. Genome-wide mutation analysis of tumor tissue biopsies from patients pre-treated with 5-FU with breast and colorectal cancer confirmed that 5-FU also caused this specific mutational pattern in vivo[43]. Interestingly, tumors harboring *TP53* mutations accumulated higher numbers of 5-FU mutations after 5-FU exposure. This finding raises the question whether these 5-FU mutations do only occur due to a deficient p53 damage regulatory pathway or that these mutations are also linked to resistance. Reasonably, tumors intrinsically harboring signature 17 might therefore be less sensitive to 5-FU-based therapies. However, in our cohort of patients starting with capecitabine, we observed that the majority of patients (61 out of 73) had no COSMIC mutational signature 17-related mutations in their tumor biopsies and that the remaining 12 patients had low (<5%) relative contributions of this signature, rendering this marker unsuitable for upfront patient selection.

The current study has a number of limitations. First, although all patients received capecitabine monotherapy, the study population was heterogeneous regarding the number and type of previously received treatments and breast cancer subtype. Second, despite the fact that, to our knowledge, this study is the largest study analyzing WGS in relation to response to capecitabine treatment, our sample size might still be too small to reveal less frequently occurring genomic events. A larger sample size will enable analysis using combinations of more rarely affected genes. We now focused on genes which were affected in at least five patients. Third, since our cohort was heterogeneous regarding previous therapies and since survival is influenced by disease stage and later-line therapies, we were not able to associate WGS with overall survival. Since, progression free survival was not the primary outcome of the CPCT-02 study, this parameter is less reliably measured, hence we solely focused on best response to capecitabine treatment.

In conclusion, the current study illustrates the use of WGS to identify subgroups of patients with or without response from capecitabine in a real-life setting. The biomarkers identified in our exploratory study are considered hypothesis generating and need further investigation and validation in independent cohorts to determine their value for selection of patients for capecitabine therapy.

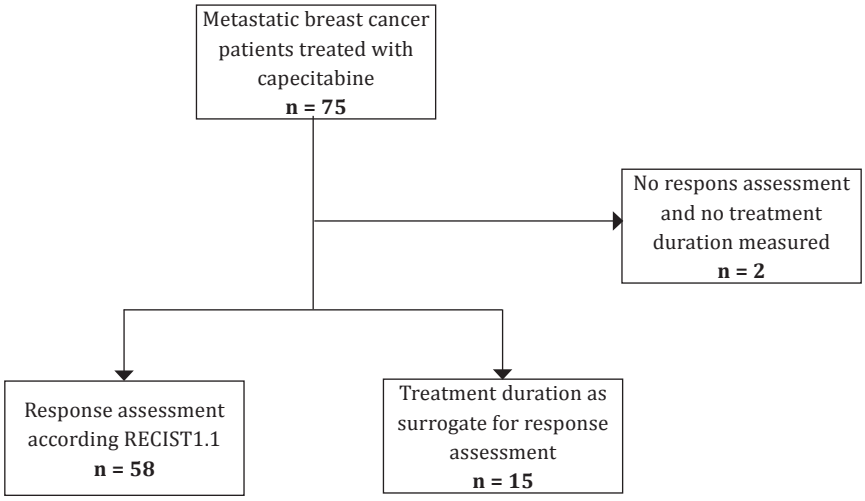
Acknowledgements
This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation and the Center of Personalized Cancer Treatment (CPCT) have made available to the study.

References

1. Miwa, M. *et al.* Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *Eur J Cancer* **34**, 1274-81 (1998).
2. Schuller, J. *et al.* Preferential activation of capecitabine in tumor following oral administration to colorectal cancer patients. *Cancer Chemother Pharmacol* **45**, 291-7 (2000).
3. Fumoleau, P. *et al.* Multicentre, phase II study evaluating capecitabine monotherapy in patients with anthracycline- and taxane-pretreated metastatic breast cancer. *Eur J Cancer* **40**, 536-42 (2004).
4. Blum, J.L. *et al.* Multicenter phase II study of capecitabine in paclitaxel-refractory metastatic breast cancer. *J Clin Oncol* **17**, 485-93 (1999).
5. Blum, J.L. *et al.* Pooled analysis of individual patient data from capecitabine monotherapy clinical trials in locally advanced or metastatic breast cancer. *Breast Cancer Res Treat* **136**, 777-88 (2012).
6. Oshaughnessy, J.A. *et al.* Randomized, open-label, phase II trial of oral capecitabine (Xeloda) vs. a reference arm of intravenous CMF (cyclophosphamide, methotrexate and 5-fluorouracil) as first-line therapy for advanced/metastatic breast cancer. *Ann Oncol* **12**, 1247-54 (2001).
7. Reichardt, P. *et al.* Multicenter phase II study of oral capecitabine (Xeloda(")) in patients with metastatic breast cancer relapsing after treatment with a taxane-containing therapy. *Ann Oncol* **14**, 1227-33 (2003).
8. Talbot, D.C. *et al.* Randomised, phase II trial comparing oral capecitabine (Xeloda) with paclitaxel in patients with metastatic/advanced breast cancer pretreated with anthracyclines. *Br J Cancer* **86**, 1367-72 (2002).
9. Miller, K.D. *et al.* Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *J Clin Oncol* **23**, 792-9 (2005).
10. Lam, S.W. *et al.* Single-nucleotide polymorphisms in the genes of CES2, CDA and enzymatic activity of CDA for prediction of the efficacy of capecitabine-containing chemotherapy in patients with metastatic breast cancer. *Pharmacol Res* **128**, 122-129 (2018).
11. Bonotto, M. *et al.* Making capecitabine targeted therapy for breast cancer: which is the role of thymidine phosphorylase? *Clin Breast Cancer* **13**, 167-72 (2013).
12. Ribelles, N. *et al.* A carboxylesterase 2 gene polymorphism as predictor of capecitabine on response and time to progression. *Curr Drug Metab* **9**, 336-43 (2008).
13. Zhao, H.Y. *et al.* Evaluations of biomarkers associated with sensitivity to 5-fluorouracil and taxanes for recurrent/advanced breast cancer patients treated with capecitabine-based first-line chemotherapy. *Anticancer Drugs* **23**, 534-42 (2012).
14. Lee, S.J. *et al.* Thymidylate synthase and thymidine phosphorylase as predictive markers of capecitabine monotherapy in patients with anthracycline- and taxane-pretreated metastatic breast cancer. *Cancer Chemother Pharmacol* **68**, 743-51 (2011).
15. Yoo, B.K. *et al.* Identification of genes conferring resistance to 5-fluorouracil. *Proc Natl Acad Sci U S A* **106**, 12938-43 (2009).
16. Ju, J. *et al.* Regulation of p53 expression in response to 5-fluorouracil in human cancer RKO cells. *Clin Cancer Res* **13**, 4245-51 (2007).
17. Yang, B. *et al.* Wild-type p53 protein potentiates cytotoxicity of therapeutic agents in human colon cancer cells. *Clin Cancer Res* **2**, 1649-57 (1996).
18. Augenlicht, L.H. *et al.* Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53-independent mutation associated with improved outcome in a randomized multi-institutional trial. *Cancer Res* **57**, 1769-75 (1997).
19. Arango, D. *et al.* c-myc/p53 interaction determines sensitivity of human colon carcinoma cells to 5-fluorouracil in vitro and in vivo. *Cancer Res* **61**, 4910-5 (2001).
20. He, L. *et al.* Wnt pathway is involved in 5-FU drug resistance of colorectal cancer cells. *Exp Mol Med* **50**, 101 (2018).

21. Priestley, P. *et al.* Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv*, 415133 (2018).
22. Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* (2019).
23. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-47 (2009).
24. Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* (2019).
25. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
26. Blokzijl, F. *et al.* MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Medicine* **10**, 33 (2018).
27. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
28. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-2 (2016).
29. Vos, S. *et al.* Comprehensive Proteomic Profiling-derived Immunohistochemistry-based Prediction Models for BRCA1 and BRCA2 Germline Mutation-related Breast Carcinomas. *Am J Surg Pathol* **42**, 1262-1272 (2018).
30. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
31. Osako, T. *et al.* Predictive factors for efficacy of capecitabine in heavily pretreated patients with metastatic breast cancer. *Cancer Chemother Pharmacol* **63**, 865-71 (2009).
32. Kandioler, D. *et al.* TP53 Mutational Status and Prediction of Benefit from Adjuvant 5-Fluorouracil in Stage III Colon Cancer Patients. *EBioMedicine* **2**, 825-30 (2015).
33. Pilat, N. *et al.* Assessing the TP53 marker type in patients treated with or without neoadjuvant chemotherapy for resectable colorectal liver metastases: a p53 Research Group study. *Eur J Surg Oncol* **41**, 683-9 (2015).
34. Levin, M.K. *et al.* Genomic alterations in DNA repair and chromatin remodeling genes in estrogen receptor-positive metastatic breast cancer patients with exceptional responses to capecitabine. *Cancer Med* **4**, 1289-93 (2015).
35. Morris, L.G. *et al.* Genomic dissection of the epidermal growth factor receptor (EGFR)/PI3K pathway reveals frequent deletion of the EGFR phosphatase PTPRS in head and neck cancers. *Proc Natl Acad Sci U S A* **108**, 19024-9 (2011).
36. Davis, T.B. *et al.* PTPRS Regulates Colorectal Cancer RAS Pathway Activity by Inactivating Erk and Preventing Its Nuclear Translocation. *Sci Rep* **8**, 9296 (2018).
37. Salaroglio, I.C. *et al.* ERK is a Pivotal Player of Chemo-Immune-Resistance in Cancer. *Int J Mol Sci* **20**(2019).
38. Mann, M.B. *et al.* Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma. *Nat Genet* **47**, 486-95 (2015).
39. Kikutake, C., *et al.* Intratumor heterogeneity of HMCN1 mutant alleles associated with poor prognosis in patients with breast cancer. *Oncotarget* **9**, 33337-33347 (2018).
40. Wang, Z.C. *et al.* Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res* **64**, 64-71 (2004).
41. Shivapurkar, N. *et al.* Multiple regions of chromosome 4 demonstrating allelic losses in breast carcinomas. *Cancer Res* **59**, 3576-80 (1999).
42. Henricks, L.M. *et al.* DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *Lancet Oncol* **19**, 1459-1467 (2018).
43. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat Commun* **10**, 4571 (2019).

Supplementary figure



Supplementary Figure S1 - Flowchart of patients selected for analyses.

Supplementary Table S1.

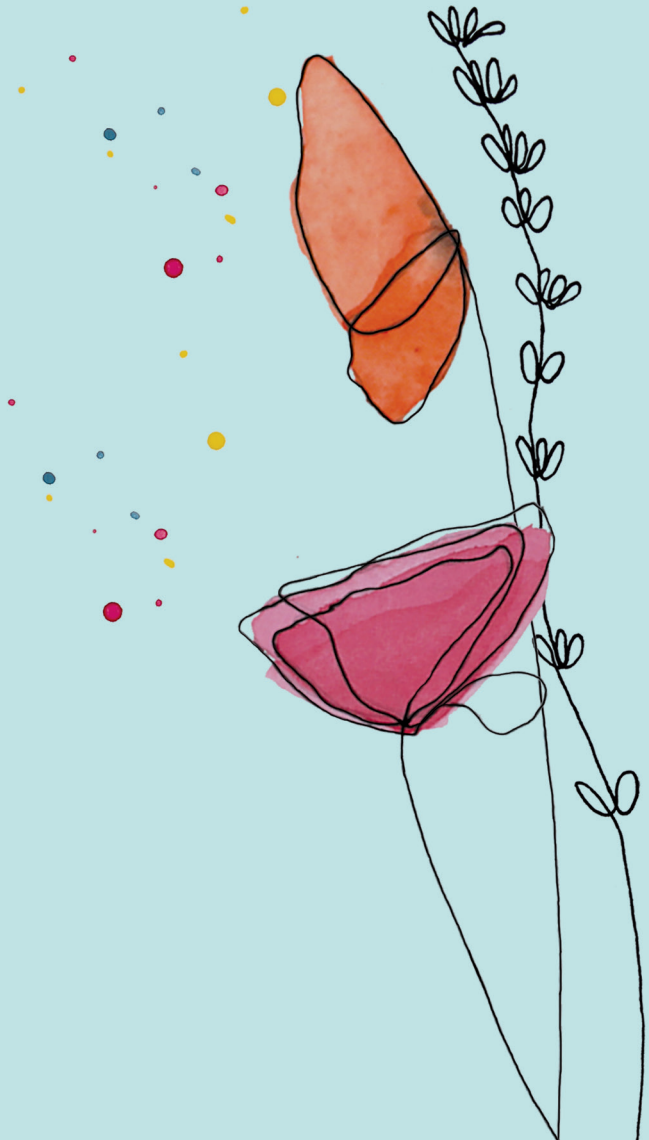
Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable
Clinical parameters		**Genes**		**Copy number regions**	
Age	Continuous	*CEP350*	Categorical	***Amplification peak***	
ER status	Categorical	*FLG*	Categorical	1p12	Categorical
HER2 status	Categorical	PTPRS	Categorical	1q32.1	Categorical
Prior radiotherapy	Categorical	MUC16	Categorical	6q21	Categorical
Number of prior drugs	Continuous	TP53	Categorical	8p11.23	Categorical
Prior treatment contains CDK4/6 inhibitor	Categorical	PKHD1	Categorical	8q11.1	Categorical
Prior treamtent contains tamoxifen	Categorical	MROH2B	Categorical	8q22.2	Categorical
Prior treamtent contains platinum/parp	Categorical	HMCN1	Categorical	8q24.21	Categorical
Prior treamtent contains anthracyclines	Categorical	PIK3CA	Categorical	11p13	Categorical
Prior treamtent contains cyclophosphamide	Categorical	ADGRG4	Categorical	11q13.3	Categorical
Prior treamtent contains anti-her2	Categorical	SPHKAP	Categorical	12q15	Categorical
Prior treamtent contains everolimus	Categorical	MUC4	Categorical	16p13.13	Categorical
Prior treamtent contains eribuline	Categorical	DNAJC13	Categorical	17q12	Categorical
Prior treamtent contains vinorelbine	Categorical	CUBN	Categorical	17q23.1	Categorical
Prior treamtent contains aromatase inhibitor	Categorical	OTOG	Categorical	20q13.2	Categorical
Prior treamtent contains 5-FU	Categorical	RYR2	Categorical	20q13.33	Categorical
Prior treamtent contains taxanes	Categorical	FRAS1	Categorical	***Deletion peak***	
Prior treamtent contains fulvestrant	Categorical	KRAS	Categorical	1p36.23	Categorical
		NCOR2	Categorical	1p36.11	Categorical
Structural variants		TSHZ2	Categorical	1p22.3	Categorical
Number of structural variants	Continuous	DMD	Categorical	2q22.1	Categorical
Number of duplications	Continuous	LYST	Categorical	2q37.3	Categorical
Number of deletions	Continuous	NEAT1	Categorical	3p13	Categorical
Number of inversions	Continuous	PC	Categorical	4p16.3	Categorical
Number of breakend variants	Continuous	ABCA12	Categorical	4q22.1	Categorical
Number of insertions	Continuous	ERBB2	Categorical	4q35.2	Categorical
Number of large duplications (>10 kb)	Continuous	ANKHD1	Categorical	5q12.1	Categorical
Number of small duplications (<10 kb)	Continuous	ANKHD1EIF4EBP3	Categorical	5q21.3	Categorical
Number of large deletions (>10 kb)	Continuous	ABCA13	Categorical	6p25.3	Categorical
Number of small deletions (<10 kb)	Continuous	RB1	Categorical	6q14.3	Categorical
		BIRC6	Categorical	6q26	Categorical

Supplementary Table S1 - Continued

Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable
COSMIC mutational signatures		CLTC	Categorical	7q11.22	Categorical
Absolute contribution		KNL1	Categorical	7q36.1	Categorical
Signature 1	Continuous	CDH1	Categorical	8p21.3	Categorical
Signature 2	Continuous	FSIP2	Categorical	9p23	Categorical
Signature 3	Continuous	SYNE1	Categorical	9p21.3	Categorical
Signature 5	Continuous	TTN	Categorical	10q23.31	Categorical
Signature 8	Continuous	GREB1L	Categorical	11p15.5	Categorical
Signature 9	Continuous	DNAH17	Categorical	11q23.2	Categorical
Signature 13	Continuous	DCHS2	Categorical	11q25	Categorical
Signature 16	Continuous	PBRM1	Categorical	12p13.1	Categorical
Signature 18	Continuous	C5ORF42	Categorical	12q23.1	Categorical
		FOXA1	Categorical	13q14.2	Categorical
Rearrangement signatures		KIAA1549	Categorical	14q24.3	Categorical
Absolute contribution		CSMD3	Categorical	16q23.1	Categorical
SV signature 1	Continuous	MED13	Categorical	17p13.1	Categorical
SV signature 2	Continuous	CTCF	Categorical	17p12	Categorical
SV signature 3	Continuous	PKD1	Categorical	17q21.2	Categorical
SV signature 4	Continuous	USH2A	Categorical	18q23	Categorical
SV signature 5	Continuous	TENM2	Categorical	19p13.3	Categorical
SV signature 6	Continuous	AHNAK2	Categorical	19q13.43	Categorical
		LRP2	Categorical	20p12.1	Categorical
Other		RNF111	Categorical	22q12.1	Categorical
CHORD - score for homologous recombination	Categorical	UNC79	Categorical	22q13.33	Categorical
Number of mutations per Mb	Continuous	VPS13D	Categorical	Xp21.1	Categorical
		PRKDC	Categorical	Xq21.33	Categorical
		DNAH11	Categorical		
		ARID1A	Categorical		
		KMT2D	Categorical		
		CCDC105	Categorical		
		UPF2	Categorical		
		KMT2C	Categorical		
		MAP2K4	Categorical		
		UNC80	Categorical		
		PTEN	Categorical		
		ZNFX1	Categorical		

Supplementary Table S1 - Continued

Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable	Variables tested in univariate analysis	Type of variable
		DYNC2H1	Categorical		
		XIRP2	Categorical		
		ATRX	Categorical		
		HIVEP1	Categorical		
		ABI3BP	Categorical		
		DNAH14	Categorical		
		ESR1	Categorical		
		GATA3	Categorical		
		CSMD1	Categorical		
		OBSCN	Categorical		
		MDN1	Categorical		
		LRP1B	Categorical		
		SPTA1	Categorical		
		WDFY4	Categorical		
		APOB	Categorical		
		RAD18	Categorical		
		SCN9A	Categorical		
		ZNF562	Categorical		



CHAPTER 4

ESR1 mutations: moving towards guiding treatment decision-making in metastatic breast cancer patients

Cancer Treatment Reviews. *2017 Jan;52:33-40.*

Lindsay Angus*, Nick Beije*, Agnes Jager, John W. M. Martens, Stefan Sleijfer

* Both authors contributed equally to this work

Abstract

Mutations in the gene coding for the estrogen receptor (ER), *ESR1*, have been associated with acquired endocrine resistance in patients with ER-positive metastatic breast cancer (MBC). Functional studies revealed that these *ESR1* mutations lead to constitutive activity of the ER, meaning that the receptor is active in absence of its ligand estrogen, conferring resistance against several endocrine agents. While recent clinical studies reported that the occurrence of *ESR1* mutations is rare in primary breast cancer tumors, these mutations are more frequently observed in metastatic tissue and circulating cell-free DNA of MBC patients pretreated with endocrine therapy. Given the assumed impact that the presence of *ESR1* mutations has on outcome to endocrine therapy, assessing *ESR1* mutations in MBC patients is likely to be of significant interest to further individualize treatment for MBC patients. Here, *ESR1* mutation detection methods and the most relevant pre-clinical and clinical studies on *ESR1* mutations regarding endocrine resistance are reviewed, with particular interest in the ultimate goal of guiding treatment decision-making based on *ESR1* mutations.

Introduction

Endocrine therapy with selective estrogen receptor modulators/downregulators (SERMs/SERDs) or by estrogen deprivation using aromatase inhibitors (AIs), is the most important treatment modality for estrogen receptor (ER)-positive metastatic breast cancer (MBC) patients[1]. Unfortunately, 40% of patients do not benefit from first-line endocrine therapy due to intrinsic resistance, and the remainder of patients initially responding will eventually develop resistance during therapy[1]. Several mechanisms have been linked to endocrine resistance, however, no marker for resistance has reached wide clinical use yet[2-4]. Recently, mutations in the gene encoding ERα, *ESR1*, have attracted particular interest as a mechanism for endocrine resistance in MBC. Large-scale next-generation sequencing (NGS) efforts on MBC tissues revealed that these mutations are enriched in MBC patients treated with endocrine agents while these variants are not or only at very low frequencies present in primary tumor tissue[5,6]. Importantly, this implies that their presence has to be assessed in metastatic lesions, or in "liquid biopsies" such as circulating cell-free DNA (cfDNA) as a representative of metastatic tumor cells. Here we review the pros and cons of current detection methods for *ESR1* mutations, the pre-clinical and clinical studies investigating *ESR1* mutations and highlight its potential role in treatment decision-making in MBC patients.

Functional studies on *ESR1* mutations

The ER belongs to the nuclear hormone receptor superfamily[7] and consists of two activation function (AF)-1/2 domains, DNA binding and hinge domains, and a ligand binding domain (LBD) (**Figure 1**). The ER functions as a ligand-dependent transcription factor. Binding of estradiol to the LBD leads to a conformational change of helix 12, resulting in recruitment of coregulatory proteins[8]. This eventually yields transcription of genes important in normal physiological processes but also for breast tumorigenesis and breast cancer (BC) progression[9].

Recent NGS efforts revealed that somatic *ESR1* mutations in the LBD were more frequently present in metastatic lesions than previously thought. In preclinical models to evaluate the role of *ESR1* mutations in endocrine resistance, it was demonstrated that cell lines transfected with a D538G, Y537S, L536Q, Y537N, Y537C, S463P or E380Q *ESR1* mutation exert activity in the absence of estrogen[6,10-15] (**Figure 1**)**.** This constitutive activity suggests that estrogen-depriving therapies such as AIs are not or less effective in patients with activating *ESR1* mutations. Cell lines transfected with mutant *ESR1* variants were however still responsive to treatment with tamoxifen and fulvestrant, though sensitivity to these drugs was relatively impaired compared to *ESR1*

wildtype transfected cell lines[5,6,12,13]. Similar observations were recently made for novel SERM/SERD hybrid endocrine therapies pipendoxifene and bazedoxifene[16].

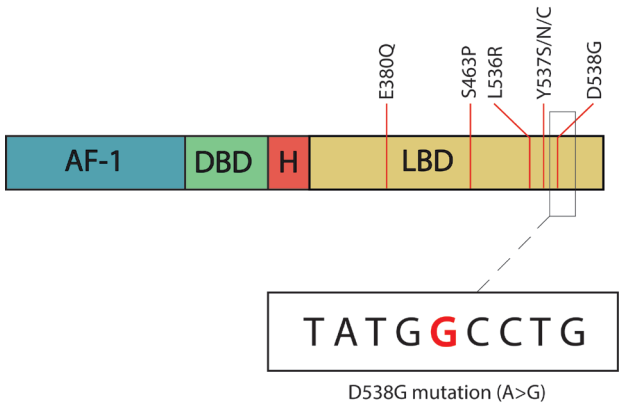


Figure 1 - Schematic overview of the different domains of the ER.
Activation function (AF) domain-1 present at the N-terminus acts in a ligand-independent manner, whereas, the AF-2 within the ligand binding domain (LBD) is dependent on estradiol for its activation [53]. The DNA binding domain encodes two zinc finger molecules, playing an important role in receptor dimerization and binding of the ER to specific DNA sequences: the estrogen response element (ERE)[54]. H=hinge region. *ESR1* mutations, some hotspot mutations shown as vertical red lines, mainly occur in the C-terminal domain of the receptor encoding for the LBD of the ER.

Techniques to detect *ESR1* mutations

Several techniques can be used to assess *ESR1* mutations in tissue or cfDNA (**Figure 2**), all having their own advantages and disadvantages. Importantly, these techniques widely vary in their sensitivity. NGS can be performed either in the context of whole genome sequencing, as part of a whole exome panel, or as part of a targeted *ESR1* panel. While NGS is an established and widely used approach for mutation detection in tumor tissue, mutation detection in cfDNA is more challenging, as the relative number of mutant to wildtype DNA alleles has to be taken into account. Frequencies of circulating tumor DNA (ctDNA) vary largely between patients, frequently being below 1% of the total cfDNA[17], which is beyond the sensitivity of conventional NGS. Therefore, techniques based on digital PCR (dPCR) have been introduced enabling the detection of ctDNA in frequencies as low as 0.001%[18,19]. In dPCR-based techniques, each individual DNA molecule, within its own partition, is able to react with a specific probe for wildtype *ESR1* and another probe for a specific *ESR1* mutant. There are several commercially available dPCR-based assays (e.g. digital PCR, droplet digital PCR (ddPCR), BEAMing), differing in used reagents and sample readouts, but generally having similar sensitivity[17,20]. In a

study comparing conventional targeted NGS with dPCR to detect mutations in cfDNA, threefold more D538G *ESR1* mutations in cfDNA were observed using dPCR than with NGS[21]. One disadvantage of dPCR assays is however that only a subset of hotspot mutations can be evaluated. Other assays, using some sort of target-enrichment prior to analysis, can be used to detect multiple hotspot mutations (OnTarget assay[22,23]) or multiple frequently mutated genes (e.g. SafeSeqS[24], CAPP-Seq[25]), however to date these assays have not yet been reported to be used to detect *ESR1* mutations.

Clinical studies on the significance of *ESR1* mutations

ESR1 mutations in primary and metastatic tumor tissue

Although already described anecdotally in the 90s[11,26,27], *ESR1* mutations were thought to be rare in BC. They occur only in up to 3% of primary tumors using NGS (**Supplementary Table 1**)[5,6,12,13]. Using more sensitive dPCR-based techniques, the *ESR1* mutation rate in primary BC tumors may mildly increase[28,29], however, only at very low variant allele frequencies (VAF; 0.07-0.2%)[29].

In contrast to mutation rates in primary BC, the landmark papers of Toy *et al*.[6] and Robinson *et al.*[13] showed much higher *ESR1* mutation rates in metastatic lesions (**Supplementary Table 2**). Toy and colleagues[6] found *ESR1* mutations (predominantly D538G and Y537S) in metastatic tissues in 9/36 ER-positive MBC patients who had received at least 3 months of endocrine therapy. All patients with an *ESR1* mutation were at least treated with two lines of endocrine therapy; all containing an AI. In an independent cohort of 44 metastatic tumors from patients pretreated with endocrine therapy, 5 metastases (11%) harbored an *ESR1* mutation.

Likewise, Robinson *et al.*[13] demonstrated *ESR1* mutations in 6/11 (55%) evaluated metastatic biopsies of ER-positive MBC patients. All patients with an *ESR1* mutation were pretreated with AIs and SERMs or SERDs. None of three available matched primary tumors of patients with a metastatic *ESR1* mutation harbored an *ESR1* mutation. Based on these findings and the accompanied functional studies, both groups hypothesized that *ESR1* mutations are a common mechanism underlying endocrine resistance, developing during estrogen deprivation, especially in the context of AI treatment.

Prompted by these findings, several studies investigated *ESR1* mutations in metastatic tissue of MBC patients. In 5/13 (38%) ER-positive MBC patients, who failed on multiple lines of endocrine treatment, a D538G *ESR1* mutation was reported[12]. Furthermore, Jeselsohn *et al.*[5] detected in 9/76 (12%) ER-positive metastatic tumors *ESR1* mutations (Y537N/C/S and D538G) using NGS, whereas none of the 115 ER-negative tumors they

assessed had such mutations. In both studies no *ESR1* mutations in matched primary tumors were detected[5,12].

In a study using dPCR, an *ESR1* mutation was revealed in metastatic tissue of 11/55 ER-positive MBC patients[28]. Notably, polyclonal *ESR1* mutations (multiple *ESR1* mutations in one sample) were observed in 4/11 (36%) patients. Also of particular interest was that two patients with *ESR1* mutations were not pretreated with any therapies at all and 4/11 only received prior treatment with tamoxifen, supporting a previous observation[5] that *ESR1* mutations are not exclusively found following AI treatment. In another study[29] applying dPCR, *ESR1* mutations were found in 3/43 primary tumors, 1/12 bone metastasis tissues and 3/38 brain metastasis tissues in ER-positive MBC patients. The prevalence of *ESR1* mutations and their VAF were higher in bone (1.4% VAF) and especially in brain metastases (34.3-44.9% VAF) compared to primary tumors (0.07-0.2% VAF), suggesting an enrichment of *ESR1*-mutant subclones in metastatic tissue. All these tissue-based studies provided important insights into the prevalence of *ESR1* mutations and the population of patients in which they occur. However, the biggest disadvantage of these studies is that they concerned mostly small, heterogeneously treated, and retrospectively selected patient cohorts. Furthermore, of note is that biopsies were usually taken at various time points and therefore the evidence at which moment *ESR1* mutations emerge, which is suggested to be mainly after AI treatment, is indirect. The majority of the above mentioned drawbacks are mainly driven by the fact that taking metastatic biopsies is a cumbersome procedure and even impossible in some patients, not easily allowing the assessment of *ESR1* mutations over time. In addition, taking metastatic biopsies may lead to sample bias due to tumor heterogeneity[30]. Therefore, recent studies have focused on *ESR1* mutation detection in "liquid biopsies" as a patient-friendly alternative to taking biopsies from metastatic lesions.

ESR1 mutations in 'liquid biopsies'

Circulating blood biomarkers such as circulating tumor cells (CTCs) and cfDNA are increasingly used as non-invasive surrogate "liquid biopsies", and are thought to represent the most important metastatic tumor sites[31,32]. Both these types of liquid biopsies can be measured in peripheral blood, with CTCs being intact tumor cells and cfDNA being DNA mainly derived from apoptotic tumor cells. Recently, several studies investigating the presence of *ESR1* mutations in liquid biopsies, particularly in cfDNA, have been published (**Table 1**).

Table 1 - Overview of *ESR1* mutation analysis in "liquid biopsies" of metastatic breast cancer patients

Patients	Method	*ESR1* mutation	Substrate	Number of patients with *ESR1* mutations	D538G	Y537S	Y537N	Y537C	Other	Ref
6 pts with ER- positive MBC either off or progressing on therapy	RNA sequencing	whole transcriptome	Cultured CTCs	3/6 (50%)	1/6 (17%)	1/6 (17%)	-	-	1/6 (17%) L536P	44
48 pts with ER-positive MBC receiving endocrine therapy	NGS	E380Q, V392I, P535H, Y537C/N/S, D538G	cfDNA	3/48 (6%)	1/48 (2%)	1/48 (2%)	-	-	1/48 (2%) E380Q	21
48 pts with ER-positive MBC receiving endocrine therapy	ddPCR	D538G	cfDNA	9/48 (19%)	9/48 (19%)	NP	NP	NP	NP	
3 pts with *ESR1*mutation detected in cfDNA by NGS	NGS	E380Q, V392I, P535H, Y537C/N/S, D538G	CTCs	1/3 (33%)	1/3 (33%)	-	-	-	-	
128 pts with ER-positive MBC, progression on therapy	ddPCR	D538G, Y537C/N/S, L536R	cfDNA	18/128 (14%)	14/128 (11%)	3/128 (2%)	4/128 (3%)	2/128 (2%)	2/128 (2%) L536R	34
11 pts with ER-positive MBC (8 with known *ESR1* mutation in metastatic biopsy by NGS)	ddPCR	D538G, Y537N/S	cfDNA	9/11 (82%)	6/11 (55%)	3/11 (27%)	1/11 (9%)	NP	NP	33
8 pts with ER-positive MBC	ddPCR	D538G, Y537N/S	cfDNA	6/8 (75%)	4/8 (50%)	2/8 (25%)	1/8 (13%)	NP	NP	
29 pts with MBC	ddPCR	K303R, S463P, Y537C/N/S, D538G	cfDNA	7/29 (24%)	6/29 (21%)	2/29 (7%)	-	1/29 (3%)	-	29
161 pts ER-positive MBC with prior sensitivity to nonsteroidal AI (SoFEA)	ddPCR	E380Q, L536R, Y537C/N/S, D538G, S463P	cfDNA	63/161 (39%)	29/161 (18%)	16/161 (10%)	23/161 (14%)	3/161 (2%)	6/161 (4%) E380Q, 6/161 (4%) S463P, 2/161(1%) L536R	38
360 pts with ER-positive MBC with progression on endocrine therapy (PALOMA3)	ddPCR	E380Q, L536R, Y537C/N/S, D538G, S463P	cfDNA	91/360 (25%)	51/360 (14%)	23/360 (6%)	14/360 (4%)	5/360 (1%)	22/360 (6%) E380Q, 4/360 (1%) S463P, 1/360 (1%) L536R	
153 pts with ER-positive MBC pre-treated with AI (FERGI)	BEAMing	E380Q, S463P, V524E, P535H, L536H/P/Q/R, Y537C/N/S, D538G	cfDNA	57/153 (37%)	31/153 (20%)	19/153 (12%)	10/143 (7%)	6/143 (4%)	15/153 (26%) E380Q, 5/143 (3%) S463P, 7/143 (5%) L536P	36
5 pts with MBC (4 ER+, 1 TN), with ≥100 CTCs	NGS	*ESR1* exome	40 single CTCs	10/40 (25%)	7/40 (18%)	-	-	-	3/40 (8%) E380Q	45
5 pts with MBC (4 ER+, 1 TN), with ≥100 CTCs	NGS	*ESR1* exome	cfDNA	3/5 (60%)	2/5 (40%)	-	-	1/5 (20%)	1/5 (20%) E380Q	
541 pts with ER-positive MBC with progression after nonsteroidal AI (BOLERO-2)	ddPCR	D538G, Y537S	cfDNA	156/541 (29%)	113/541 (21%)	72/541 (13%)	-	-	-	40

ddPCR= droplet digital PCR, NP= Not performed. Number of patients with a *ESR1* mutation in different study cohorts are listed. The specific mutations observed in these patients are also shown; in case of polyclonality, these numbers may exceed the total number of patients with a *ESR1* mutation.

To evaluate NGS and dPCR techniques to detect *ESR1* mutations in plasma, Guttery *et al.* examined cfDNA of 48 ER-positive MBC patients[21]. In 3/48 patients (6%), they observed an *ESR1* mutation in cfDNA using NGS. In one patient with a D538G mutation also CTCs, isolated by the CellSearch system, were sequenced, and the same mutation was detected in CTCs. When dPCR was performed in the same cohort for the D538G mutation only, the D538G mutation was found in 6 additional patients (15%) at VAF below 1%, underlining the limited sensitivity of NGS to detect low frequent mutations. In eleven patients, serial plasma samples were available. Interestingly, in one patient an *ESR1* mutation was present at baseline and was further enriched (0.4% VAF to 13.6% 3 months later) while treated with chemotherapy (docetaxel/vinorelbine).

To further explore whether *ESR1* mutations present in metastases are also represented in the cfDNA, Chu *et al*.[33] assessed *ESR1* mutations in plasma cfDNA in 11 ER-positive MBC patients in whom the *ESR1* mutation status in a metastatic lesion was assessed by NGS. All *ESR1* mutations (8/8) observed in the metastatic lesions were also observed in the cfDNA using dPCR. In one patient with an *ESR1* wildtype metastatic lesion, a low frequency *ESR1* mutation was observed in the cfDNA. It should however be noted that the cfDNA was obtained two months after the biopsy, meaning that changes in *ESR1* mutation status could also be due to therapy-related effects emerging after the initial biopsy. In an independent cohort consisting of 8 ER-positive patients, dPCR was once more demonstrated to be able to detect *ESR1* mutations in cfDNA, and in two more patients an *ESR1* mutation was observed in the cfDNA but not in the metastatic lesion. This study further underscored that dPCR is able to readily detect *ESR1* mutations in the cfDNA and that cfDNA seems to represent *ESR1* mutations in the metastatic lesions. Also, strikingly, *ESR1* mutations were detected in cfDNA but not in metastatic lesions, which may be indicative of heterogeneity within the metastatic lesion or between multiple metastases.

Another study only used dPCR to detect *ESR1* mutations[29], and *ESR1* mutations were detected in 7/29 MBC patients (24%), with one patient having polyclonal *ESR1* mutations. All patients with an *ESR1* mutation in cfDNA received at least one line of endocrine treatment, mainly AIs or tamoxifen. In this series, also an *ESR1* mutation was seen in a patient who had only received prior treatment with fulvestrant. Of particular interest were the serial blood draws in the patient with the polyclonal *ESR1* mutations, which revealed that two mutations were enriched during AI treatment and chemotherapy, while one mutation was absent after treatment. This may suggest that different mutations react differently to different treatments.

Schiavon and colleagues[34] were the first to present a study in which *ESR1* mutations were assessed in a relatively large cohort of MBC patients. With dPCR to examine cfDNA from MBC patients at the time of progression under endocrine therapy, *ESR1* mutations were observed in 18/128 patients (14%), with D538G mutations comprising 56% of all observed *ESR1* mutations. Polyclonality of *ESR1* mutations was observed in 21% of the patients. All patients in whom *ESR1* mutations were observed had received prior AI treatment, while no *ESR1* mutations were observed in a subset of 22 patients who had only received tamoxifen treatment. Interestingly, *ESR1* mutations were mainly detected in patients who received AIs only in the metastatic setting (36%), and not in patients who received AIs only in the adjuvant setting (4%) or in the adjuvant and metastatic setting (8%). In accordance were observations in two relatively small independent cohorts, in which no *ESR1* mutations were observed in 32 BC biopsies taken at recurrence after adjuvant AI treatment or in 7 cfDNA samples of MBC patients who received adjuvant AI treatment only. Regarding the outcome of patients with *ESR1* mutations, subgroup analyses in *ESR1* mutant versus wildtype patients revealed a significantly poorer progression-free survival (PFS) on subsequent AI treatment in patients harboring an *ESR1* mutation, although these analyses should be seen as exploratory given the small number of patients eligible for such analyses.

The observations by Schiavon *et al.* suggests that AI treatment in the metastatic setting, but not in adjuvant setting, causes *ESR1* mutations. This may suggest selection of subclones already present in the primary tumor, or in the metastases when the tumor load is increased and the probability of acquiring mutations increases[35]. This first observation could be in line with the previously mentioned findings by Wang *et al.* whom found *ESR1* mutations at extremely low VAF in primary tumors of MBC patients with *ESR1* mutations. While the study by Schiavon and colleagues also provided evidence for an impaired response to AI treatment, larger studies were needed to confirm these findings and to examine whether MBC patients with *ESR1* mutations will have improved responses on alternative therapies.

ESR1 mutations and outcome on endocrine therapies

In the randomized phase II FERGI trial, baseline plasma samples of patients failing to AI treatment randomized either to fulvestrant combined with the pan-PI3K inhibitor pictilisib or to the combination of fulvestrant and placebo, were examined for *ESR1* and *PIK3CA* mutations in tissue and cfDNA using BEAMing[36]. They detected *ESR1* mutations in cfDNA in 57/153 (37%) of patients at baseline; 13 patients (23%) harbored polyclonal mutations. Surprisingly, the prevalence of the E380Q mutation was rather high (26%), while this mutation was previously not often observed. No *ESR1* mutations

were detected in 42 matched primary tumors of patients with *ESR1* mutations in cfDNA. *PIK3CA* mutations were observed in the cfDNA of 40% of the patients and were generally concordant with findings in matched metastatic tissue. For the *ESR1* mutations, discordances between the cfDNA and metastatic biopsies occurred more frequently and cfDNA sometimes harbored more *ESR1* mutations than the metastatic biopsies. These analyses were however limited by the fact that metastatic tissue and cfDNA were generally not collected on the same day. Of note was that the median VAF of *PIK3CA* mutations was markedly higher than for *ESR1* mutations (3.6% versus 0.45%). The higher VAFs and concordance with tissue probably reflect that *PIK3CA* mutations usually occur in earlier stages of BC[37], in contrast to *ESR1* mutations. Similar to Wang and colleagues[29], it was observed in multiple longitudinal samples in patients with polyclonal *ESR1* mutations that different *ESR1* mutations reacted differently under treatment.

The clinical analyses in the fulvestrant/placebo arm of the FERGI study revealed that patients with an *ESR1* mutation in ctDNA had no impaired PFS on fulvestrant compared to *ESR1* wildtype. When the analyses were further restricted to those patients with polyclonal *ESR1* mutations or *ESR1* mutation with VAF above the median, also no effect on PFS was observed. Also no differences in PFS were observed in patients with and without *ESR1* mutations receiving fulvestrant and pictilisib.

The data from the FERGI study suggested that fulvestrant does not have reduced activity in patients with *ESR1* mutations. However, data on the impact of *ESR1* mutations on outcome to fulvestrant versus AI treatment and the addition of other agents to fulvestrant treatment were still missing. These gaps were filled by data from two phase III randomized trials, reported by Fribbens *et al*. whom assessed *ESR1* mutations in cfDNA by dPCR[38]. In the SoFEA study, patients who had previously benefited from a non-steroidal AI were randomly assigned to fulvestrant combined with anastrozole, fulvestrant with placebo, or exemestane alone. Mutations were detected at baseline in 63/161 (39%) patients; 27/55 (49%) patients evaluable for polyclonal mutations had such mutations. Patients with an *ESR1* mutation had an improved PFS after taking a fulvestrant-containing regimen versus exemestane (median PFS fulvestrant-containing 5.7 versus exemestane 2.6 months, P=0.02), in contrast to *ESR1* wildtype patients in whom a similar PFS was found (5.4 months versus 8.0 months, P=0.77). Within the exemestane-treated patients, patients with *ESR1* mutations (n=18) had a worse PFS compared to patients having an *ESR1* wildtype (n=39), (median PFS 2.6 versus 8.0 months P=0.01).

In the PALOMA3 study, patients who failed on prior endocrine therapy were randomized to fulvestrant in combination with the CDK4/6-inhibitor palbociclib or to fulvestrant and placebo. In 91/360 patients (25%), *ESR1* mutations were detected with polyclonal mutations observed in 26/91 (29%). The main study revealed a significant PFS benefit in patients receiving fulvestrant/palbociclib versus patients receiving fulvestrant alone (median 9.5 versus 4.6 months, *P*=0.0001)[39]. This PFS benefit was maintained in patients with *ESR1* mutations (median 9.4 versus 3.6 months, *P*=0.002), while no PFS difference was observed between *ESR1* mutants and wildtype in patients treated with fulvestrant/palbociclib (median 9.4 versus 9.5 months, respectively). Although PFS seemed to be slightly worse in the *ESR1* mutated patients treated with fulvestrant alone (3.6 months 95% CI, 2.0-5.5) compared to *ESR1* wildtype (5.4 months 95% CI 3.5-7.4), this was not statistically significant, which is in line with the results of the FERGI study[36].

So far, the only large study providing overall survival (OS) data with respect to *ESR1* mutations is the phase III BOLERO-2 study[40]. In this study, postmenopausal women who progressed on an AI were randomized to the AI exemestane combined with everolimus, or exemestane and placebo. Overall, 156/541 (28.8%) of evaluable patients had either a D583G and/or Y537S *ESR1* mutation detected in their cfDNA, with double-mutations detected in 30/541 (5.5%) patients. *ESR1* mutations were more prevalent in patients who had previously received AI treatment for metastatic disease (33%) than in patients who had received AIs as adjuvant therapy (11%), supporting previous data from Schiavon *et al*[34]. The results of the main study revealed that PFS was significantly improved in patients treated with everolimus and exemestane compared to exemestane and placebo (7.8 months versus 3.2 months), though the combination therapy did not result in improved OS[41,42]. In the *ESR1* mutation-driven subgroup analyses for PFS in the exemestane arm, patients with a mutation in D538G had a shorter PFS than *ESR1* wildtype patients (2.7 versus 3.9 months), which is in accordance with the findings of the SoFEA study[40,43]. When the analysis was restricted to patients with an Y537S mutation only, this association was not observed, which may be related to the limited sample size for these subgroup analyses. Of note is that the PFS of *ESR1* wildtype patients was 3.9 months in this study, while in the SoFEA study this was 8 months. This discrepancy in PFS might be due to differences in selection criteria of both studies. In the SoFEA trial only patients who received a non-steroidal AI as adjuvant therapy or as first-line therapy for MBC were included whereas patients in the BOLERO-2 trial were also included after receiving more lines of therapy for MBC representing a more advanced disease stage. When everolimus was added to exemestane this resulted in an improved PFS in both D538G mutated (5.8 months; HR 0.34, 95% CI 0.02-0.6) and wildtype patients (8.5 months; HR 0.4, 95% CI 0.3-0.5), suggesting that *ESR1* mutated

patients could still benefit from the addition of everolimus. Of note is that benefit of the addition of everolimus was not demonstrated for patients with an Y537S mutation alone (4.2 months; HR 0.98, 95% CI 0.5-1.9), or with both an Y537S and D538G mutation (5.4 months; HR 0.53, 95% CI 0.2-1.3). Again, one should keep in mind that these analyses may have suffered from the limited sample size of patients with only an Y537S mutation or a polyclonal *ESR1* mutation. If larger future studies confirm that patients with an Y537S indeed do not benefit from the addition of everolimus, this mutation might be used to select for patients who should be treated with other treatment modalities. Overall, the absolute median PFS interval seemed to be shorter in patients with an *ESR1* mutation than in *ESR1* wildtype patients, however, no formal analyses on these potential differences were observed. In this context, it was intriguing that OS analyses according to *ESR1* mutation status showed that patients with an *ESR1* mutation had a worse OS compared to wildtype patients (median OS 22 versus 32 months). Noteworthy, the type of individual mutations was also suggested to influence OS, with a median OS of 26 months for patients with a D538G mutation only and 20 months for the Y537S mutation alone. In patients harboring both mutations the OS was even worse with a median OS of 15 months. Overall, these results may indicate that *ESR1* mutations are associated with more aggressive disease biology.

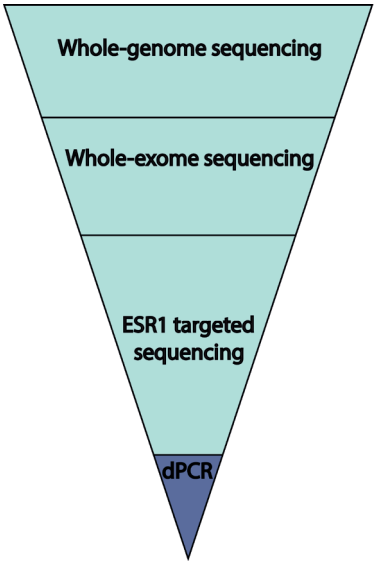


Figure 2 - Various techniques for *ESR1* mutation detection.
The pyramid represents the range in which the genome is investigated. *ESR1* mutations can be detected by large-scale NGS efforts such as whole-genome sequencing or whole-exome sequencing, or by more targeted methods as targeted sequencing of the *ESR1* gene only, or by the interrogation of individual mutations in *ESR1* by digital PCR.

Discussion

The putative role of *ESR1* mutations in endocrine resistance has sparked a wide interest in techniques enabling their detection, the conditions under which they appear, and whether their detection can ultimately assist treatment decision-making in MBC patients.

Regarding the best substrate for *ESR1* mutation detection, data from multiple studies suggests that the cfDNA compartment sometimes provides additional mutations compared to matched metastatic tumor material. This may indicate that cfDNA is more representative of the whole somatic tumor landscape. Another obvious advantage of cfDNA over metastatic biopsies is that it can easily be obtained repeatedly during treatment. Therefore, future studies on the clinical relevance of *ESR1* mutations should preferably be performed using cfDNA, measuring mutations not only at baseline but also sequentially during treatment. Of note, *ESR1* mutations can also be detected in CTCs[21,44,45], but at this point it is unclear how *ESR1* mutation detection in CTCs relates to *ESR1* mutation detection in cfDNA, and if this adds anything to *ESR1* mutation analyses in cfDNA.

Assessing *ESR1* mutations in tissue and cfDNA provided clues as to how these *ESR1* mutations are enriched in MBC patients. Very strong indirect evidence exists for the enrichment of these *ESR1* mutations during treatment with AIs in the metastatic setting. However, to date no direct evidence for the enrichment of *ESR1* mutations under AI treatment has been presented yet. In this context it is also of note that several studies observed *ESR1* mutations in metastases or cfDNA from patients treated with SERMs or SERDs only, or from patients not treated with endocrine therapy at all[5,21,28,29]. This further underlines that the understanding on how *ESR1* mutations exactly occur is still limited. *ESR1* mutations are present at very low frequencies in primary BC tumors using dPCR[29], supporting the hypothesis that *ESR1* mutations may already be subclonally present in the primary tumor, and because of growth advantages, become the more prominent clone under treatment pressure[34]. *ESR1* mutations might also occur as a result of mutational processes such as initiated by the APOBEC enzymes, however the mutational pattern of the hotspot *ESR1* mutations (T>A/C/G) does not follow an APOBEC pattern or the pattern of any other mutational signature known to date[46,47].

While the exact mechanism behind the enrichment of *ESR1* mutations in MBC is still unknown, the clinical relevance of *ESR1* mutations being present in cfDNA becomes evident. PFS after treatment with the AI exemestane was impaired in the patients

harboring an *ESR1* mutation[38,40], while fulvestrant had similar efficacy in patients with an *ESR1* mutation versus patients without an *ESR1* mutation. Given these results with fulvestrant, efficacy of tamoxifen may also be unaffected in patients harboring *ESR1* mutations, however, no clinical data on this is present as of yet. For the addition of other agents to endocrine treatment, for example palbociclib or everolimus, the question remains whether the presence of *ESR1* mutations is of any predictive significance for the efficacy of these agents. While the *ESR1* mutation status did not seem to impact median PFS in patients receiving fulvestrant and palbociclib, the presence of an *ESR1* mutation in patients treated with exemestane and everolimus might be associated with decreased PFS compared to *ESR1* wildtype patients.

Since a raise in *ESR1* mutation ratio during the course of treatment may be indicative of progressive disease[48] and *ESR1* mutations in general are associated with poor outcome[40], it will be of particular interest to see whether certain treatments (for example fulvestrant combined with palbociclib or specific chemotherapeutic regimen) are able to select against *ESR1* mutant subclones. Recently, it was shown that upon the discontinuation of anti-epidermal growth factor receptor (EGFR) antibodies, resistant KRAS mutant clones decay, allowing re-challenges with anti-EGFR antibodies in particular patients[49]. If *ESR1* mutations are lost with certain treatment regimen, this could potentially allow re-challenges with AIs in a subset of patients.

Also currently unknown is whether the different *ESR1* mutations result in distinctive phenotypes. Functional studies on *ESR1* mutations did not specifically focus on differences between various *ESR1* mutations, and for some *ESR1* mutations that have been measured in clinical studies (e.g. K303R, V524E, P535H, L536H/P/R), very little functional evaluation of its constitutive activity and potential role in endocrine resistance has been performed at all. In addition, clinical studies to date have generally been underpowered for subgroup analyses evaluating differential effects of different *ESR1* mutations. Even further complicating such analyses is the described polyclonality of *ESR1* mutations. Multiple studies with anecdotal longitudinal sampling data suggested that in patients with polyclonal *ESR1* mutations there are differential effects of therapy on different *ESR1* mutations. This suggests that *ESR1* mutations are present in different subclones, and not in the same cell. Theoretically, this may mean that patients with polyclonal mutations are more difficult to treat given the wider repertoire of resistance mutations. However, in rather small analyzed groups of patients with polyclonal *ESR1* mutations treated with fulvestrant such effects were not observed. Given that some *ESR1* mutations are rarer than others, the most pragmatic way to evaluate the prognostic value of these rare *ESR1* mutations will likely be in the form

of a meta-analysis in due time, as it is virtually impossible to evaluate the prognostic value of these mutations in single studies. In addition, functional studies evaluating all LBD *ESR1* mutations described in patients to date, validating their constitutive activity and exploring potential differential effects of different *ESR1* mutations are of interest.

The current evidence on *ESR1* mutations warrants prospective studies in which patients are randomized and treated according to the *ESR1* mutation status in cfDNA. Therefore, standardized methods to process plasma, to isolate cfDNA and to prepare and analyze the dPCR chips are needed. A lot of the recent *ESR1* mutation research was performed on cfDNA samples that were suboptimally collected. For example, in the SoFEA trial, plasma was collected in EDTA tubes and processed up to 9 days after sample collection which may have consequences for the sensitivity to detect *ESR1* mutations, especially in the context of longitudinal sampling[50,51]. Recently, it was demonstrated that blood collected in CellSave or BCT blood tubes assures optimal quality of cfDNA for dPCR or NGS analyses for up to 96 hours after the blood draw[50-52], providing opportunities to send blood samples to remote locations for plasma isolation. In addition, it is of utmost importance to assess variables such as intra-assay, inter-assay, inter-lab and inter-observer variability when using dPCR, which are currently not only poorly studied for *ESR1* mutations, but also for cfDNA analyses in general.

In conclusion, the presence of *ESR1* mutations in patients with ER-positive MBC has high potential for clinical validity and utility. Prospective studies in which the exact role of how *ESR1* mutations can be used to guide treatment decision-making have to be initiated, but firstly standardization of protocols to assess these mutations will be necessary to eventually allow clinical implementation.

References

1. Pritchard, K.I. Endocrine therapy: is the first generation of targeted drugs the last? *J Intern Med* **274**, 144-52 (2013).
2. Osborne, C.K. & Schiff, R. Mechanisms of endocrine resistance in breast cancer. *Annu Rev Med* **62**, 233-47 (2011).
3. De Laurentiis, M. *et al.* A meta-analysis on the interaction between HER-2 expression and response to endocrine treatment in advanced breast cancer. *Clin Cancer Res* **11**, 4741-8 (2005).
4. Amir, E. *et al.* Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J Clin Oncol* **30**, 587-92 (2012).
5. Jeselsohn, R. *et al.* Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* **20**, 1757-67 (2014).
6. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45 (2013).
7. Olefsky, J.M. Nuclear receptor minireview series. *J Biol Chem* **276**, 36863-4 (2001).
8. Kojetin, D.J.*et al.* Implications of the binding of tamoxifen to the coactivator recognition site of the estrogen receptor. *Endocr Relat Cancer* **15**, 851-70 (2008).
9. Heldring, N. *et al.* Estrogen receptors: how do they signal and what are their targets. *Physiol Rev* **87**, 905-31 (2007).
10. Weis, K.E. *et al.* Constitutively active human estrogen receptors containing amino acid substitutions for tyrosine 537 in the receptor protein. *Mol Endocrinol* **10**, 1388-98 (1996).
11. Zhang, Q.X. *et al.* An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer Res* **57**, 1244-9 (1997).
12. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* **73**, 6856-64 (2013).
13. Robinson, D.R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51 (2013).
14. Li, S. *et al.* Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep* **4**, 1116-30 (2013).
15. Pakdel, F., Reese, J.C. & Katzenellenbogen, B.S. Identification of charged residues in an N-terminal portion of the hormone-binding domain of the human estrogen receptor important in transcriptional activity of the receptor. *Mol Endocrinol* **7**, 1408-17 (1993).
16. Wardell, S.E. *et al.* Efficacy of SERD/SERM Hybrid-CDK4/6 Inhibitor Combinations in Models of Endocrine Therapy-Resistant Breast Cancer. *Clin Cancer Res* **21**, 5121-30 (2015).
17. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-90 (2008).
18. Diehl, F. *et al.* BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* **3**, 551-9 (2006).
19. Vogelstein, B. & Kinzler, K.W. Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-41 (1999).
20. Hindson, B.J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-10 (2011).
21. Guttery, D.S. *et al.* Noninvasive detection of activating estrogen receptor 1 (ESR1) mutations in estrogen receptor-positive metastatic breast cancer. *Clin Chem* **61**, 974-82 (2015).
22. Kidess, E. *et al.* Mutation profiling of tumor DNA from plasma and tumor tissue of colorectal cancer patients with a novel, high-sensitivity multiplexed mutation detection platform. *Oncotarget* **6**, 2549-61 (2015).
23. Thompson, J.D. *et al.* Winnowing DNA for rare sequences: highly specific sequence and methylation based enrichment. *PLoS One* **7**, e31597 (2012).
24. Kinde, I. *et al.* Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 9530-5 (2011).

25. Newman, A.M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548-54 (2014).
26. Karnik, P.S. *et al.* Estrogen receptor mutations in tamoxifen-resistant breast cancer. *Cancer Res* **54**, 349-53 (1994).
27. Roodi, N. *et al.* Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *J Natl Cancer Inst* **87**, 446-51 (1995).
28. Takeshita, T. *et al.* Droplet digital polymerase chain reaction assay for screening of ESR1 mutations in 325 breast cancer specimens. *Transl Res* **166**, 540-553 e2 (2015).
29. Wang, P. *et al.* Sensitive Detection of Mono- and Polyclonal ESR1 Mutations in Primary Tumors, Metastatic Lesions, and Cell-Free DNA of Breast Cancer Patients. *Clin Cancer Res* **22**, 1130-7 (2016).
30. Yates, L.R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751-9 (2015).
31. Haber, D.A. & Velculescu, V.E. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov* **4**, 650-61 (2014).
32. Onstenk, W. *et al.* Molecular characteristics of circulating tumor cells resemble the liver metastasis more closely than the primary tumor in metastatic colorectal cancer. *Oncotarget* (2016).
33. Chu, D. *et al.* ESR1 Mutations in Circulating Plasma Tumor DNA from Metastatic Breast Cancer Patients. *Clin Cancer Res* **22**, 993-9 (2016).
34. Schiavon, G. *et al.* Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* **7**, 313ra182 (2015).
35. Lipinski, K.A. *et al.* Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer* **2**, 49-63 (2016).
36. Spoerke, J.M. *et al.* Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat Commun* **7**, 11579 (2016).
37. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93 (2012).
38. Fribbens, C. *et al.* Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *J Clin Oncol* (2016).
39. Cristofanilli, M. *et al.* Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol* **17**, 425-39 (2016).
40. Chandarlapaty, S. *et al.* Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA Oncol* (2016).
41. Yardley, D.A. *et al.* Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis. *Adv Ther* **30**, 870-84 (2013).
42. Piccart, M. *et al.* Overlies plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2. *Ann Oncol* **25**, 2357-62 (2014).
43. Fribbens, C. *et al.* Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *J Clin Oncol* **34**, 2961-8 (2016).
44. Yu, M. *et al.* Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216-20 (2014).
45. Shaw, J.A. *et al.* Mutation analysis of cell-free DNA and single circulating tumor cells in metastatic breast cancer patients with high CTC counts. *Clin Cancer Res* (2016).
46. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
47. Morganella, S. *et al.* The topography of mutational processes in breast cancer genomes. *Nat Commun* **7**, 11383 (2016).
48. Takeshita, T. *et al.* Clinical significance of monitoring ESR1 mutations in circulating cell-free DNA in estrogen receptor positive breast cancer patients. *Oncotarget* **7**, 32504-18 (2016).

49. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 827 (2015).
50. Rothwell, D.G. *et al.* Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample. *Mol Oncol* **10**, 566-74 (2016).
51. Van Dessel, L.F. *et al.* Application of circulating tumor DNA in prospective clinical oncology trials: standardization of pre-analytical conditions. *Submitted for publication* (2016).
52. Kang, Q. *et al.* Comparative analysis of circulating tumor DNA stability In K3EDTA, Streck, and CellSave blood collection tubes. *Clin Biochem* (2016).
53. Ruff, M. *et al.* Estrogen receptor transcription and transactivation: Structure-function relationship in DNA- and ligand-binding domains of estrogen receptors. *Breast Cancer Res* **2**, 353-9 (2000).
54. Nilsson, S. *et al.* Mechanisms of estrogen action. *Physiol Rev* **81**, 1535-65 (2001).

Supplementary Table 1 - Overview of *ESR1* mutation analysis in primary breast cancer

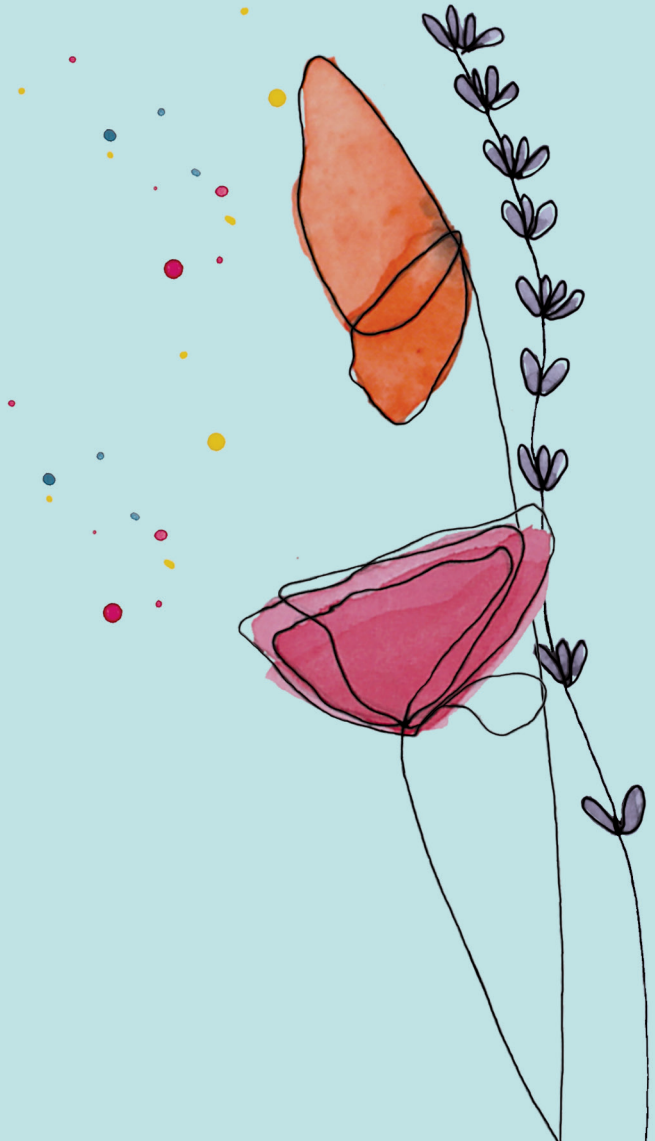
Patients	Method	*ESR1* mutation	No. *ESR1* mutations	Ref
183 pts with ER-positive MBC participating in BOLERO-2	NGS	*ESR1* exome	6/183 (3%)	1
390 ER-positive tumors resected before endocrine therapy (TCGA sequenced tumors)	NGS		0/390	2,3
80 ER-negative tumors			0/80	
134 pts with ER-positive/HER2-negative BC	NGS	*ESR1* exome	0/58	4
104 pts with ER-negative BC			0/115	
270 pts with ER-positive BC	ddPCR	*ESR1* D538G, Y537S/N	7/270 (0.03%)	5
43 ER-positive primary tumors	ddPCR	D538G, Y537S/N/C, S463P, K303R	3/43 (7%)	6

Supplementary Table 2 - Overview of *ESR1* mutation analysis in metastatic breast cancer

Patients	Method	*ESR1* mutation	Samples with *ESR1* mutations	D538G	Y537S	Y537N	Y537C	Other	Ref
11 ER-positive MBC patients	NGS	*ESR1* exome	6/11 (55%)	2/11 (18%)	3/11 (27%)	-	-	1/11 (9%) L536Q	3
36 ER-positive MBC patients, with progressive disease while at least 3 months treated with endocrine treatment	NGS	*ESR1* exome	9/36 (25%)	3/36 (8%)	4/36 (11%)	1/36 (3%)	-	3/36 (8%) S463P/L536R/ V534E	1
44 ER-positive MBC patients, participating in the BOLERO-2, with progressive disease on an AI			5/44 (11%)	1/44 (2%)	1/44 (2%)	1/44 (2%)	1/44 (2%)	2/44 (5%) S463P/P535H	
13 ER-positive MBC patients, failing several lines of treatment	NGS	*ESR1* exome	5/13 (38%)	5/13 (38%)	-	-	-	-	7
76 ER-positive MBC patients	NGS	*ESR1* exome	9/76 (12%)	3/76 (4%)	2/76 (3%)	3/76 (4%)	1/76 (1%)	-	4
31 ER-positive MBC patients with progression on therapy	ddPCR	D538G, Y537S/ N/C, L536R	4/31 (13%)	1/31 (3%)	1/31 (3%)	2/31 (6%)	1/31 (3%)	-	8
55 ER-positive MBC patients	ddPCR	D538G, Y537S/ N/C, L536R	11/55 (20%)	4/55 (7%)	5/55 (9%)	4/55 (7%)	4/55 (7%)	-	5
11 bone metastasis of MBC patients	ddPCR	D538G, Y537S/ N/C, S463P, K303R	1/12 (8%)	1/12 (8%)	-	-	-	-	6
38 brain metastasis of MBC patients			3/38 (8%)	3/38 (8%)	1/38 (3%)	-	-	-	
7 metastases of ER-positive MBC patients	Sanger sequencing		4/7 (57%)	2/7 (29%)	1/7 (14%)	1/7 (14%)	-	-	9
	dPCR	Y537N, Y537S, D538G	6/7 (71%)	3/7 (29%)	2/7 (29%)	2/7 (29%)	-	-	

Supplemental references

1. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45 (2013).
2. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
3. Robinson, D.R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51 (2013).
4. Jeselsohn, R. *et al.* Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* **20**, 1757-67 (2014).
5. Takeshita, T. *et al.* Droplet digital polymerase chain reaction assay for screening of ESR1 mutations in 325 breast cancer specimens. *Transl Res* **166**, 540-553 e2 (2015).
6. Wang, P. *et al.* Sensitive Detection of Mono- and Polyclonal ESR1 Mutations in Primary Tumors, Metastatic Lesions, and Cell-Free DNA of Breast Cancer Patients. *Clin Cancer Res* **22**, 1130-7 (2016).
7. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* **73**, 6856-64 (2013).
8. Schiavon, G. *et al.* Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* **7**, 313ra182 (2015).
9. Sefrioui, D. *et al.* Short report: Monitoring ESR1 mutations by circulating tumor DNA in aromatase inhibitor resistant metastatic breast cancer. *Int J Cancer* **137**, 2513-9 (2015).



CHAPTER 5

Genomic alterations associated with endocrine resistance in metastatic breast cancer have a differential impact on downstream ER signaling

Submitted

Lindsay Angus, Marcel Smid, Saskia M. Wilting, Manouk K. Bos, Neeltje Steeghs, Inge R.H.M. Konings, Vivianne C.G. Tjan-Heijnen, Johanna M.G.H. van Riel, Agnes J. van de Wouw, CPCT Consortium, Edwin Cuppen, Martijn P. Lolkema, Agnes Jager, Stefan Sleijfer, John W.M. Martens

Abstract

Purpose

Mutations in the estrogen receptor gene (*ESR1*), its transcriptional regulators, and the mitogen-activated protein kinase (MAPK) pathway are enriched in patients with endocrine resistant metastatic breast cancer (MBC). Here, we integrated whole genome sequencing with RNA sequencing data from the same samples to analyze the downstream effects of DNA alterations previously linked to endocrine resistance thereby gaining a better understanding of the associated mechanisms.

Experimental design

ER-positive/HER2-negative (n=101) MBC patients who underwent a tumor biopsy prior to start of a new line of treatment for MBC (CPCT-02 study, NCT01855477) were included. Unsupervised clustering was performed using expression of *ESR1* target genes. Genomic alterations at the DNA level, gene expression levels, and last administered therapy were compared between the identified clusters.

Results

Unsupervised clustering revealed two distinct clusters, one of which was characterized by increased expression of *ESR1* and its target genes. Samples in this cluster were significantly enriched for mutations in *ESR1* and amplifications in *FGFR1* and *TSPYL.* Patients in the other cluster showed relatively lower expression levels of *ESR1* and its target genes, comparable to ER-negative samples, and more often received endocrine therapy as their last treatment before biopsy. Genes in the MAPK-pathway, including *NF1*, and *ESR1* transcriptional regulators were evenly distributed.

Conclusion

RNA sequencing identified a subgroup of patients with clear expression of *ESR1* and its downstream targets, probably still benefiting from ER-targeting agents. The lower ER expression in the other subgroup, might be partially explained by ER activity still being blocked by recently administered endocrine treatment.

Introduction

Breast cancer is the most common cancer among women world-wide[1]. Although the majority of breast cancer patients are cured, 20-30% of patients will develop incurable metastatic disease[2]. As 60-70% of the patients with metastatic breast cancer (MBC) have tumors expressing estrogen receptor alpha (ER), endocrine therapy has become the mainstay treatment for these patients[3]. Despite the success of endocrine treatment, unfortunately, 20-30%[4-7] of patients have no clinical benefit from first-line endocrine therapy due to intrinsic resistance, whereas the remainder of initially responding patients will eventually develop resistance during therapy. Once resistant, tumors become more aggressive and more difficult to treat.

Recently performed sequencing efforts on metastatic tumor biopsies have revealed several mechanisms conferring resistance against endocrine treatment[8-10]. Activating mutations in the ligand binding domain of the gene encoding the ER, *ESR1,* lead to constitutive activity of ER and have been related to shorter progression free survival on single agent aromatase inhibitors (AIs)[11,12]. After exposure to nonsteroidal AIs for metastatic disease 29-39%[11-13] of patients harbor *ESR1* mutations, whereas *ESR1* mutations are quite rare in primary tumors (only ~3% of patients[14,15]) and newly diagnosed metastatic disease after adjuvant treatment with AIs (5.3-6.4% of patients[16-18]). The most common *ESR1* mutations have been functionally annotated whereby D538G and Y537 are known for their constitutive activity, whereas the E380Q variant "only" leads to estrogen hypersensitivity[15].

Additionally, mutations in the mitogen-activated protein kinase (MAPK)-pathway are enriched in MBC as well, including alterations in *ERBB2*, *NF1*, *EGFR*, *ERBB3*, *KRAS, BRAF*, *MAP2K1* and *HRAS*[8]. Of these genes, inactivating mutations in *NF1* are most frequently observed and are mutually exclusive with activating mutations in *ESR1*[8,19]. Recently, Zheng *et al.* have shown that *NF1* acts as a co-repressor of ER-α transcription, so when *NF1* gets inactivated this leads to an increased expression of ER target genes such as *GREB1* and *TFF1*[20]. In addition to mutations in *ESR1* and MAPK-pathway genes, alterations in ER transcriptional regulators have been associated with endocrine resistance as well, including *MYC*, *FOXA1*, *CTCF,* and *TBX3*[8].

Although whole genome and whole exome sequencing (WGS and WES, respectively) efforts have revealed enriched gene alterations in metastatic tumors compared to primary breast cancer, integration with gene expression is necessary to enable subsequent analysis of the downstream effects of these alterations. Here, we integrated WGS data

with matched RNA sequencing data obtained from biopsies of 101 patients with ER-positive/HER2-negative metastatic breast cancer to assess the relation between *ESR1* mutations, alterations in *ESR1*-transcriptional regulators, and MAPK-pathway mutations and the activity of the ER-pathway. Samples and associated genomic and clinical features were ordered by unsupervised clustering using RNA expression of ER-pathway genes to get a better understanding on how these mutations associate with *ESR1* expression and its target genes. Finally, we performed an exploratory analysis to correlate the tumors' transcriptome with subsequent best response to endocrine therapies.

Methods

Study design and patients

For the current analyses, we selected patients with metastatic breast cancer who were included under the protocol of the Center for Personalized Cancer Treatment (CPCT) consortium (CPCT-02 Biopsy Protocol, ClinicalTrial.gov no. NCT01855477), which was approved by the medical ethics committee of the University Medical Center Utrecht, the Netherlands. The consortium and the whole patient cohort have been described in detail recently[10,21]. Briefly, patients of ≥ 18 years old, with incurable locally advanced or metastatic solid tumors, from whom a histological biopsy could be safely obtained and systemic treatment with anticancer agents was indicated, were eligible for inclusion. Biopsies of metastatic lesions from patients with ER-positive/HER2-negative breast cancer (obtained via pathology reports of the primary tumor), from which both WGS and RNA sequencing data were available, were included (n=101). A cohort of 63 patients with ER-negative metastatic breast cancer was included as readout for low/absent ER expression.

Prior endocrine therapy

As we investigated the expression of estrogen regulated genes, it is important to establish in which patients expression of these genes was potentially influenced by endocrine therapy. We expected that endocrine therapies that were given as last systemic therapy before the tissue biopsy was taken, could still impact the expression of estrogen regulated genes. Therefore, we registered the last treatment that was administered before biopsy and grouped these treatments as follows: 1) tamoxifen; 2) aromatase inhibitors; 3) fulvestrant; 4) chemotherapy or other non-endocrine therapy; 5) combination endocrine therapy (endocrine backbone combined with either CDK-4/6 inhibitors or everolimus); 6) no therapy within one year before the biopsy.

Treatment outcome and response to endocrine therapies

Tumor responses were evaluated according to RECIST v1.1 every 8-12 weeks of treatment and best overall response was defined as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD)[22]. For the subset of patients that started with endocrine therapy after their tumor biopsy, RNA profiles were associated with the best overall response. Patients without response information or patients who did not start with endocrine therapy after the investigated biopsy were excluded from this analysis.

WGS and data analyses

Genomic features, such as somatic single nucleotide variants and copy number alterations, were extracted from WGS data as previously described[10]. Mutational signatures v3[23] were called using R package MutationalPatterns v1.10.0[24], focusing on single and double base signatures. Alterations in genes within the same pathway were grouped based on the findings of Razavi *et al*.[8]. In short, we used the following definitions: ESR1 hotspot mutations: mutations in codons 536-538 and 380; MAPK pathway alterations: mutations in *BRAF*, *ERBB3*, *HRAS*, *KRAS, MAP2K1,* mutations and deep gains in *EGFR* and *ERBB2*, and mutations and deep deletions in *NF1*; ER transcriptional regulators: mutations in *CTCF* and *TBX3*, mutations and deep gains in *FOXA1* and deep gains of *MYC*.

RNA sequencing: RNA isolation, library preparation and sequencing

RNA was isolated from fresh frozen tissue biopsies using the QiaSympony RNA kit (Qiagen, Venlo, the Netherlands) as per manufacturer's instructions and quantified using the Qubit RNA IQ Assay (Invitrogen, Life Technologies, Carlsbad, CA) according to the manufacturer's instructions using the Qubit fluorometer (Invitrogen). The RNA yield from the tissue biopsies ranged between 50 and 5,000 ng. Library preparation was performed using the KAPA RNA HyperPrep kit (Roche) with RiboErase (Human/Mouse/Rat) on an automated liquid handing platform (Beckman Coulter) using a total of 50-100 ng RNA. RNA was fragmented at 85°C for 6 minutes in the presence of magnesium to a target fragment length of 300bp. Barcoded libraries were sequenced as pools on NextSeq 500 (V2.5 reagents) generating 2 x 75 read pairs or at a later stage on a NovaSeq 6000 generating 2 x 150 read pairs using standard settings (Illumina). Binary base call (BCL) output from the sequencing platform was converted to FASTQ using Illumina bcl2fastq tool (versions 2.17 to 2.20 have been used) using default parameters

Processing of RNA sequencing data

Next, sequence-reads in the FASTQ files were trimmed for adapter-sequences using fastp v0.20.0[25]. The resulting FASTQ files were mapped to GRCh38 using STAR v2.6.1d9[26], and

Sambamba v0.7.0[27] was used to mark duplicates and index the resulting BAM files. Gene annotation was derived from GENCODE Release 30 (https://www.gencodegenes.org/). To obtain gene expression levels (raw read counts) featureCounts v1.6.3[28] was used. Finally, the count matrix was normalized using the GeTMM method[29], using R v3.6.0[30].

Unsupervised clustering on RNA expression levels of ER-regulated genes

Gene expression values were available as log2 values for 19,986 protein coding transcripts. Since unsupervised clustering of the top 5,000 variable genes was largely driven by the site of metastatic biopsy – i.e. all liver biopsies clustered together (**Supplementary Figure 1**), we corrected for this by performing a ComBat correction[31]. After correction, we performed unsupervised clustering on the expression levels of estrogen regulated transcripts of the following genes: *AP1B1*; *CA12*; *CDH26*; *CELSR2*; *COL18A1*; *COX7A2L*; *CTSD*; *DSCAM*; *EBAG9*; *ERBB2*; *ESR1*; *GREB1*; *HSPB1*; *IGFBP4*; *KRT19*; *MYC*; *NRIP1*; *PGR*; *PISD*; *PTMA*; *RARA*; *SGK3*; *SOD1*; *TFF1*; *TRIM25*; *CCN5*; *XBP1*[32]. Expression levels were first median-centered before clustering and then hierarchically clustered using average linkage and uncentered correlation distance metric[33] and visualized using Treeview[34].

Statistical methods

Pearson's chi-squared test or Fisher's exact test (in case of too few expected events) was used to evaluate categorical data. To compare continuous variables a Mann-Withney *U*-test or a Kruskal-Wallis H test was performed. All statistical tests were considered statistically significant at a two-sided $P < 0.05$. Stata 13.0 (StataCorp) and R v.3.6.0. were used for the statistical analyses. We used the Benjamini–Hochberg procedure to correct *P* values for multiple hypothesis testing when appropriate.

Results

Patient cohort

We integrated prospectively collected WGS and RNA sequencing data from tissue biopsies of 101 patients with ER–positive/HER2–negative metastatic breast cancer (**Table 1**). Seventy-six (75%) patients received one or more prior treatments. Focusing on the treatment last administered before biopsy, 8 patients received tamoxifen, 20 an AI, 2 fulvestrant, 6 combination therapy with endocrine backbone, 22 chemotherapy or other non-endocrine-containing therapy, and 43 received no prior treatment within one year before the biopsy.

Table 1 - Patient characteristics

	Patients (n = 101)		Specification of prior treatments		
Age	N	%		N	%
Median (Inter quartile range)	59 (52 - 64)		Aromatase inhibitor	53	52.5
Gender			Tamoxifen	57	56.4
Female	101	100	Fulvestrant	16	15.8
Male	0	0	Everolimus	15	14.9
Prior systemic therapy			CDK4/6	6	5.9
Yes			5-FU	35	34.7
Endocrine therapy only	14	13.7	Taxanes	44	43.6
Chemotherapy only	11	10.9	Platinum/Parp	8	7.9
Endocrine and chemotherapy	51	50.5	Anthracyclines	52	51.5
Nr of lines (median, IQR)	3 (2-5)		Cyclophosphamide	50	49.5
Nr of drugs (median, IQR)	5 (4-8)		Eribulin	3	3.0
No prior treatment	25	27.8	Vinorelbine	3	3.0
Last treatment before biopsy			Anti-HER2	3	3.0
Tamoxifen	8	7.9			
Aromatase inhibitor	20	19.8			
Fulvestrant	2	2.0			
Combination endocrine therapy	6	5.9			
Chemotherapy or non-containing endocrine therapy	22	21.8			
No treatment within one year before biopsy	43	42.6			
Prior radiotherapy					
Yes	62	61.4			
No	39	38.6			
Biopsy site					
Liver	36	35.6			
Bone	8	7.9			
Lymph node	20	19.8			
Breast	14	13.9			
Other	20	19.8			
Unknown	3	3.0			

Regarding the alterations in pathways associated with endocrine resistance, 16 patients (16%) had a mutation in *ESR1* of which 15 were located in the ligand binding domain (p.D538G (n=8); p.Y537S (n=4); p.Y537N (n=1); p.L536P (n=1); p.E380Q (n=1)) and one was a nonsense mutation located in the activation function 1 domain (p.Q17*). Next to *ESR1* mutations, 27 patients (27%) had alterations at the DNA level in the MAPK pathway and 45 (45%) patients in the ER transcriptional regulators (**Figure 1**).

Unsupervised clustering reveals distinct RNA expression of ER-target genes between ESR1 mutant and ESR1 wild type samples

We performed unsupervised clustering (**Figure 2**) using the gene expression levels of ER target genes (see methods for gene list) of 101 metastatic lesions, which revealed two clusters of samples. Compared to cluster B, samples in cluster A (n=47) had a higher average gene expression of all ER-target genes (Mann-Whitney *P*=4.95E-17) (**Figure 3**). *ESR1* itself and known ER-target genes such as *GREB1* (Mann-Whitney *P*=1.32E-16) and *PGR* (Mann-Whitney *P*=6.44E-12) also had a higher expression in cluster A than

in cluster B. As the samples in cluster B were derived from patients with ER-positive primary breast cancer, we investigated whether the average gene expression of the ER-target genes were at the same level as ER-negative samples. Comparing samples from cluster B (n=54) with 63 ER-negative samples, we found that the *ESR1* expression itself was higher in samples from cluster B. However, the average expression level of the ER-target genes was similar in the samples in cluster B as in ER-negative samples.

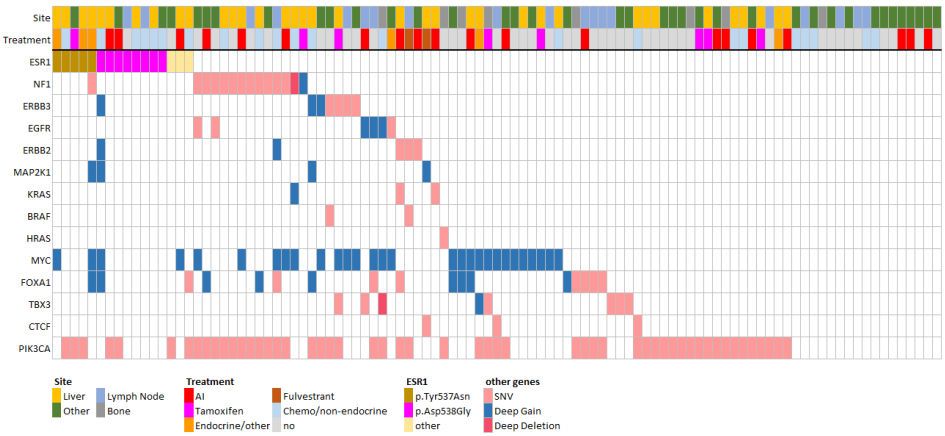


Figure 1 - Overview of genomic alterations previously associated with endocrine resistance in our cohort. The treatment depicted in figure is the latest administered therapy prior to biopsy if the metastatic lesion.

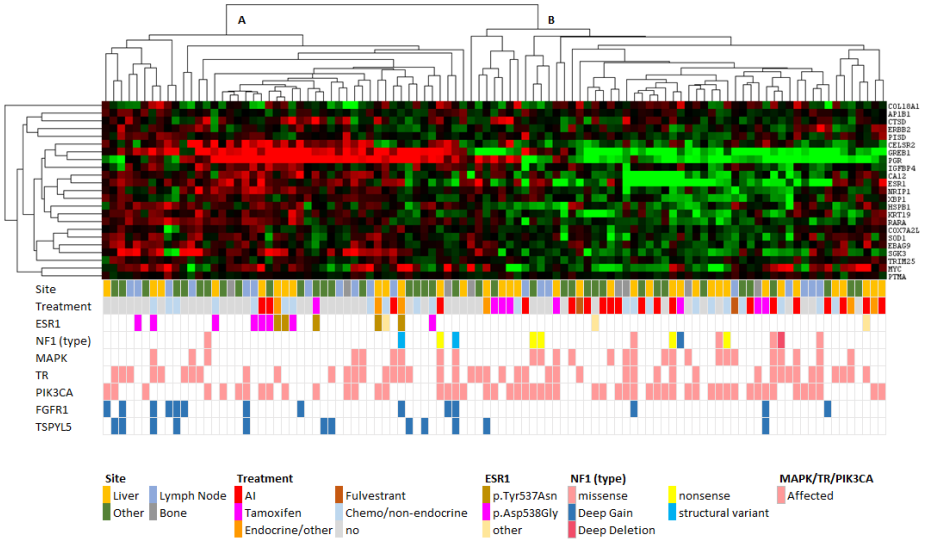


Figure 2 - Unsupervised clustering based on ER-target genes.

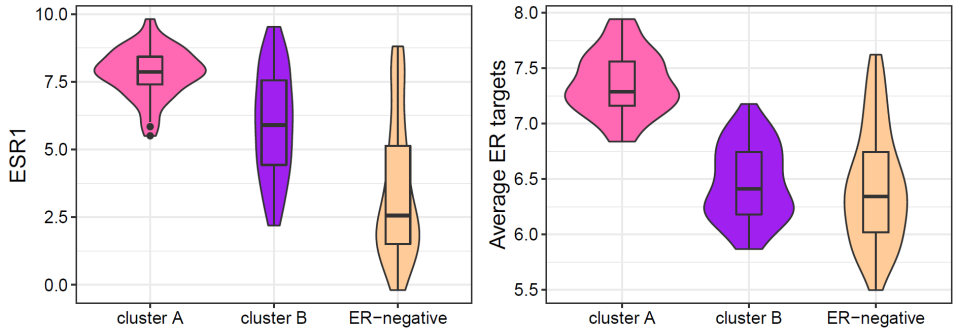


Figure 3 - Log expression values of *ESR1* (A) and the average expression of the ER-target genes (B) in samples in cluster A, B and a control group of ER-negative samples.

As cluster A represented samples with higher expression of *ESR1* and its target genes, we investigated whether specific underlying genomic alterations were enriched in cluster A versus cluster B. First, we focused on genes and gene-pathways previously associated with endocrine resistance: *ESR1* mutations, *ESR1*-transcriptional regulators and MAPK-pathway alterations. Cluster A was characterized by an enrichment of samples with hotspot mutations in *ESR1* (Fisher's Exact, $P<0.001$) (p.D538G n=8; p.Y537S n=4; p.Y537N (n=1); and p.E380Q (n=1)). Cluster B only contained two samples with an *ESR1* mutation, one harbored a truncating mutation p.Q17* and the other one contained mutation p.L536P. However, the first mutation is not located in the ligand binding domain and importantly is an inactivating mutation. For the latter variant, constitutive activity has been shown but in this one patient the "ER-low" profile was found[15]. Overall, these results support the observation that activating *ESR1* mutations result in a distinct profile with increased expression of *ESR1* and ER target genes, as was the case in cluster A. Alterations in *ESR1* transcriptional regulator genes and MAPK pathway genes, were equally distributed over the two clusters. Considering *NF1* and the RAS pathway genes (*BRAF/HRAS/KRAS*) separately, these seemed more frequently affected in cluster B, but this difference was not statistically significant (**Supplementary Figure 2**). Since tumors with *PIK3CA* mutations recently have been shown to yield sensitivity to alpelisib, a PI3K inhibitor, we compared its frequency between both clusters and observed that cluster B (n=54) was characterized by a modest but significant enrichment of *PIK3CA* mutations (Pearson Chi-Square, $P=0.029$).

Furthermore, we compared the relative contributions of mutational signatures between cluster A and B and observed that the contribution of COSMIC mutational signature 3 (associated with homologous recombination deficiency) was higher in cluster A than

in B (*P*=0.011). Conversely, COSMIC mutational signature 2, associated with APOBEC mutagenesis was enriched in cluster B (Mann-Whitney *P*=0.011). Since we expected that endocrine therapies that were discontinued just before the tissue biopsy was taken could still impact the expression of estrogen regulated genes, we compared the last administered treatment versus the clusters. The type of last administered treatment was different between both clusters (Fisher's Exact, *P*=0.016) and we observed that patients in cluster B had more frequently received an AI or tamoxifen directly prior to their biopsy, whereas samples from patients in cluster A more frequently received chemotherapy or no treatment at all within one year prior to biopsy.

***FGFR1* and *TSPYL5* amplifications are enriched beyond *ESR1* mutations**

To investigate whether other DNA alterations, next to *ESR1* mutations, were enriched in cluster A that could drive the high expression of *ESR1* and its target genes, we compared samples that were *ESR1* wild type from cluster A with all samples from cluster B. Firstly, focusing on mutations, we observed that three genes were less frequently mutated in cluster A: *MUC16, MAST4*, and *CACNA1E*. Secondly, focusing on copy number alterations, we observed that amplifications of fibroblast growth factor 1 (*FGFR1*) occurred more frequently in cluster A (Pearson Chi-Square, *P*=0.008) in a virtually exclusive manner with *ESR1* mutations (only 2 of 14 *ESR1* mutants also had an *FGFR1* amplification). In addition, amplifications of *TSPYL5* were enriched in cluster A (Fisher's Exact, *P*=0.006), which occurred only in one of the *ESR1* mutated samples. Thirdly, focusing on differential gene expression between both clusters, 17 genes were found significantly differentially expressed and with at least a two-fold change in median expression level (Mann-Whitney, FDR corrected) (**Table 2**).

Best response on endocrine therapies versus expression clusters

Lastly, we investigated whether the obtained clusters were associated with outcome on subsequent endocrine therapy. Twenty-one patients started with endocrine therapy after their biopsy; 17 patients started with an AI, and 4 patients with fulvestrant. Exploratory analyses on best response versus the above described clusters showed that the responses (PR, SD and PD) were equally distributed over the two clusters (Fisher's Exact *P*=1.00).

Table 2 - Genes differentially expressed (beyond the ER-target genes from figure 2) between cluster A (*ESR1* wild type) and cluster B

Gene	P-value FDR Hochberg	Fold change*
TPBG	1.88E-08	2.6
IGF1R	4.30E-05	3.2
CYP2T1P	5.02E-05	4.4
SIAH2	0.00011	2.3
FMN1	0.00039	2.1
THSD4	0.00043	3.1
AC0647992	0.00141	3.1
CUEDC1	0.00213	2.0
SUSD3	0.00402	5.3
PARD6B	0.00715	3.5
ZNF516	0.00746	2.2
PREX1	0.01012	2.6
IL6ST	0.01655	2.3
STC2	0.01764	4.6
MYBL1	0.02001	2.5
EGLN2	0.02175	2.0
COX6C	0.03034	3.5

** Median fold change cluster A over cluster B*

Discussion

To investigate the effects of DNA alterations associated with resistance against endocrine treatment on the expression of ER target genes, we here present the integration of RNA sequencing with WGS data obtained from metastatic lesions of a relatively large cohort of patients with ER-positive/HER2-negative breast cancer. We demonstrate that *ESR1* mutations, and *FGFR1* and *TSPYL* amplifications are associated with an increased expression of *ESR1* and its target genes, while mutations in genes involved in the MAPK-pathway or in genes encoding ER transcriptional regulators, previously associated with endocrine resistance[8], did not correlate with increased expression of ER-target genes.

Our findings on the association between *ESR1* mutations and high *ESR1* expression and its target genes adds to the compelling evidence that mutations in *ESR1* lead to constitutive activity of ER[14,35,36]. Interestingly, *FGFR1* amplifications were also enriched in samples with high ER pathway activity and occurred almost mutually exclusively with samples having an *ESR1* mutations. Our observation is in line with previous work showing that *FGFR1* amplification leads to ligand-independent ER-target gene transcription[37] and mediates endocrine therapy resistance[37-39]. High expression of *TSPYL* has previously been linked to poor outcome in breast cancer patients by suppressing p53[40], however, a link with increased ER pathway activation as found here, has to our knowledge not been described before.

Taking cases with *ESR1* mutations and *FGFR1/TSPYL* amplifications together, 28 out of 47 (59.6%) samples had DNA alterations associated with increased ER-target gene expression, still leaving 19 samples without a recurrent genomic event that might explain the higher expression of ER-target genes. The analysis of differentially expressed genes between cluster A and B did not provide additional clues why these samples had a higher ER- target gene expression since most of the differentially expressed genes between cluster A and B (**Table 2**) are direct target genes of ER, such as *STC2* and *EGLN2*.

The observation that there is a subgroup of patients having an increased ER expression profile underlines the remaining need of blocking the ER pathway in these patients since it is very likely that these tumors are still heavily rely on the ER pathway. Currently there are interventional trials underway investigating the effectiveness of next generation selective ER modulators and selective ER down-regulators targeting both wild type and mutant ER[41]. The question remains how to optimally select patients who will have the largest benefit from these new and existing endocrine treatment strategies. The integrative analysis of RNA sequencing and WGS in relation with response to these treatments might reveal a benefit for most of the patients in cluster A.

In contrast with *ESR1* mutations and *FGFR1* amplifications, *NF1* mutations and alterations in other MAPK-pathway genes were not significantly enriched in the "active ER cluster". This finding is in contrast with recent work[20] showing that *NF1* is a transcriptional co-repressor of ER-α and once inactivated leads to increased ER expression in cell line models. In our clinical samples we did not observe enrichment of *NF1* mutations in the cluster with increased ER expression. The number of nonsense mutations and structural variations in *NF1* in our cohort was low (n=7). So, assessment of the relation between *NF1* mutation status and *ESR1* regulated gene expression should be assessed in a larger clinical cohort with a higher number *NF1* mutations.

Importantly, we observed that tumor biopsies of patients who were receiving endocrine therapies such as tamoxifen and AIs as last treatment before their biopsy were significantly more frequently observed in "ER-low" cluster B. We speculate that, since the average ER-target gene expression in the "ER-low" cluster is at the same level as ER-negative tumors, these tumors classified as ER-positive might have turned into phenotypically ER-negative tumors, not being dependent on the ER-pathway anymore. These cells could revert to ER dependent growth once endocrine treatment pressure is lifted. However, given the fact that the last given treatment is associated with lower expression of *ESR1* and its target genes, one should be aware that these samples could

be wrongly classified as "ER-low". Moreover, samples in the "ER-low" cluster do show some degree of heterogeneity in the expression of the ER-target genes. A detailed look at the sub-clusters within the "ER-low" cluster shows that some samples with high *PGR* expression have an overall low expression of ER-target genes. Although progesterone receptor (PR) activity has long been considered as an indicator of functional estrogen response pathway[42], a small subset of primary breast cancer is classified as ER-negative/PR-positive with large studies showing an incidence of 3.4-3.8%[43,44]. A second sub-cluster within the "ER-low' cluster shows high *MYC* expression. Although known as *ESR1* target gene, *MYC* is also often found activated via amplification. Taken together, samples in these two sub-clusters may be indeed independent of the ER-pathway and switched to *PGR* or *MYC* enabled growth.

Our dataset has limitations; we selected a cohort of metastatic breast cancer patients with an ER-positive/HER2-negative primary tumor. As standard pathology assessment such as ER and HER2 expression status has not been investigated on these metastatic samples, and receptor conversion does occur[45], we do not have factual evidence that all samples were still ER-positive/HER2-negative upon biopsy. Moreover, our cohort is heterogeneous regarding the number and type of prior lines of treatments received which could have influenced the presence of specific genetic alterations at the DNA level as well as affected gene expression. Also, in the CPCT-02 study tumor biopsies were taken prior to a next line of therapy for metastatic disease. Although most patients will have had their tumor biopsy upon disease progression of the prior line of therapy or at diagnosis of metastatic disease, there might be a small subgroup of patients that underwent a biopsy after experiencing toxicity. Importantly, our cohort could comprise a mix of patients that are either still sensitive or already resistant to endocrine therapies.

In conclusion, we here show the potential of the integrative analysis of RNA sequencing and WGS and demonstrate that within the subgroup of ER-positive/HER2-negative breast cancers there are substantial differences in *ESR1* expression and its target genes. We further show associations between ER transcriptomic profiles and possible underlying DNA alterations. Given the fact that there is a subgroup of tumors with an increased expression of *ESR1* and its target genes, these tumors might still be responsive to drugs that target the ER-pathway. To identify the largest possible group of patients that could benefit from existing endocrine treatments or new drugs targeting both wild type and mutant ER, one should employ upfront DNA and RNA sequencing on metastatic tumor tissue to optimally relate treatment response to genomic and transcriptomic profiles.

Conflict of Interest

Dr. Lolkema reports grants from KWF (Dutch Cancer Society) and NWO (Dutch Scientific Council), during the conduct of the study; grants and personal fees from Sanofi, grants and personal fees from Johnson & Johnson, grants and personal fees from Merck, grants and personal fees from Astellas, personal fees from Incyte, personal fees from Amgen, personal fees from Janssen Cilag, personal fees from Bayer, personal fees from Servier, personal fees from Pfizer.

Dr. Tjan-Heijnen reports institutional grants from Roche, Novartis, E Lilly, Sanofi, AstraZeneca, Daiichi Sankyo and honoraria from Novartis, E Lilly, Roche.

Data availability statement

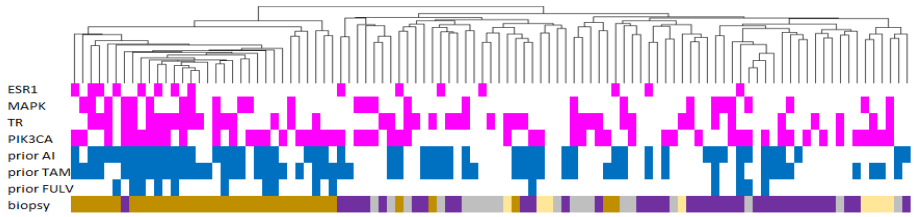
WGS data, RNA sequencing data and corresponding clinical data have been requested from Hartwig Medical Foundation and provided under data request number DR-068. The clinical data provided by CPCT have been locked at 30th of March 2020. Both WGS and clinical data are freely available for academic use from the Hartwig Medical Foundation through standardized procedures and request forms can be found at https://www.hartwigmedicalfoundation.nl.

References

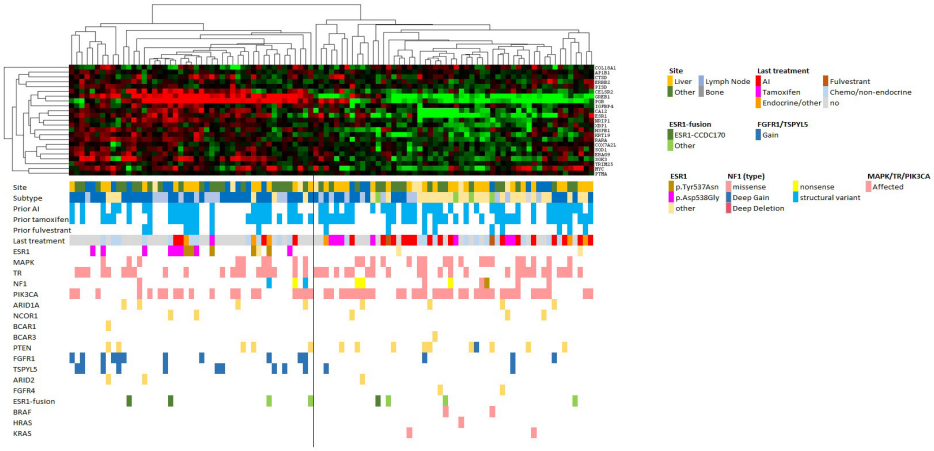
1. Siegel, R.L. *et al.* Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30 (2020).
2. Pan, H. *et al.* 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N Engl J Med* **377**, 1836-1846 (2017).
3. Cardoso, F. *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4). *Ann Oncol* **29**, 1634-1657 (2018).
4. Finn, R.S. *et al.* Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* **375**, 1925-1936 (2016).
5. Robertson, J.F.R. *et al.* Fulvestrant 500 mg versus anastrozole 1 mg for hormone receptor-positive advanced breast cancer (FALCON): an international, randomised, double-blind, phase 3 trial. *Lancet* **388**, 2997-3005 (2016).
6. Hortobagyi, G.N. *et al.* Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. *Ann Oncol* **29**, 1541-1547 (2018).
7. Goetz, M.P. *et al.* MONARCH 3: Abemaciclib As Initial Therapy for Advanced Breast Cancer. *J Clin Oncol* **35**, 3638-3646 (2017).
8. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438 e6 (2018).
9. Bertucci, F. *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564 (2019).
10. Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* (2019).
11. Chandarlapaty, S. *et al.* Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA Oncol* (2016).
12. Fribbens, C. *et al.* Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer. *J Clin Oncol* **34**, 2961-8 (2016).
13. Spoerke, J.M. *et al.* Heterogeneity and clinical significance of ESR1 mutations in ER-positive metastatic breast cancer patients receiving fulvestrant. *Nat Commun* **7**, 11579 (2016).
14. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45 (2013).
15. Toy, W. *et al.* Activating ESR1 Mutations Differentially Affect the Efficacy of ER Antagonists. *Cancer Discov* **7**, 277-287 (2017).
16. Schiavon, G. *et al.* Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci Transl Med* **7**, 313ra182 (2015).
17. Allouchery, V. *et al.* Circulating ESR1 mutations at the end of aromatase inhibitor adjuvant treatment and after relapse in breast cancer patients. *Breast Cancer Res* **20**, 40 (2018).
18. Bidard, F.C. *et al.* Abstract PD2-06: Circulating *ESR1* mutation detection rate and early decrease under first line aromatase inhibitor and palbociclib in the PADA-1 trial (UCBG-GINECO). *Cancer Research* **79**, PD2-06-PD2-06 (2019).
19. Sokol, E.S. *et al.* Loss of function of NF1 is a mechanism of acquired resistance to endocrine therapy in lobular breast cancer. *Ann Oncol* **30**, 115-123 (2019).
20. Zheng, Z.Y. *et al.* Neurofibromin Is an Estrogen Receptor-alpha Transcriptional Co-repressor in Breast Cancer. *Cancer Cell* (2020).
21. Priestley, P. *et al.* Pan-cancer whole genome analyses of metastatic solid tumors. *bioRxiv*, 415133 (2018).
22. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-47 (2009).
23. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).
24. Blokzijl, F. *et al.* MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**, 33 (2018).

25. Chen, S. *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
26. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
27. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-4 (2015).
28. Liao, Y. *et al.* featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
29. Smid, M. *et al.* Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics* **19**, 236 (2018).
30. R Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria 2019).
31. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
32. Verhaegh, W. *et al.* Selection of personalized patient therapy through the use of knowledge-based computational models that identify tumor-driving signal transduction pathways. *Cancer Res* **74**, 2936-45 (2014).
33. de Hoon, M.J.L. *et al.* Open source clustering software. *Bioinformatics* **20**, 1453-1454 (2004).
34. Saldanha, A.J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248 (2004).
35. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* **73**, 6856-64 (2013).
36. Robinson, D.R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51 (2013).
37. Formisano, L. *et al.* Association of FGFR1 with ERalpha Maintains Ligand-Independent ER Transcription and Mediates Resistance to Estrogen Deprivation in ER(+) Breast Cancer. *Clin Cancer Res* **23**, 6138-6150 (2017).
38. Turner, N. *et al.* FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* **70**, 2085-94 (2010).
39. Drago, J.Z. *et al.* FGFR1 Amplification Mediates Endocrine Resistance but Retains TORC Sensitivity in Metastatic Hormone Receptor-Positive (HR(+)) Breast Cancer. *Clin Cancer Res* **25**, 6443-6451 (2019).
40. Epping, M.T. *et al.* TSPYL5 suppresses p53 levels and function by physical interaction with USP7. *Nat Cell Biol* **13**, 102-8 (2011).
41. Fanning, S.W. & Greene, G.L. Next-Generation ERalpha Inhibitors for Endocrine-Resistant ER+ Breast Cancer. *Endocrinology* **160**, 759-769 (2019).
42. Horwitz, K.B. *et al.* Estrogen control of progesterone receptor in human breast cancer: role of estradiol and antiestrogen. *Endocrinology* **103**, 1742-51 (1978).
43. Rakha, E.A. *et al.* Biologic and clinical characteristics of breast cancer with single hormone receptor positive phenotype. *J Clin Oncol* **25**, 4772-8 (2007).
44. Colditz, G.A. *et al.* Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst* **96**, 218-28 (2004).
45. Lindstrom, L.S. *et al.* Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* **30**, 2601-8 (2012).

Supplementary Figures



Supplementary Figure 1.
When applying unsupervised clustering on the top 5000 variable genes, the RNA sequencing data revealed two large clusters of tumors that segregated primarily by the biopsy site. The first cluster of samples mainly consisted of biopsies obtained from liver metastases (brown) and the second cluster included samples from various organ sites consisting on bone (yellow), lymph node (grey), and other biopsy sites (purple). Gene expression in the first cluster was mainly driven by gene expression of normal liver tissue. We therefore corrected for this bias by performing an ComBat correction.



Supplementary Figure 2 - Unsupervised clustering based on ER-target genes also including genes which were not significantly different between both clusters.

PART II

Liquid Biopsies

CHAPTER 6

RAS and BRAF mutations in cell-free DNA are predictive for outcome of cetuximab monotherapy in patients with tissue-tested RAS wild-type advanced colorectal cancer

Molecular Oncology. *2019 Nov;13(11):2361-2374.*

Erik J. van Helden*, **Lindsay Angus***,
C. Willemien Menke - van der Houven van Oordt, Daniëlle A.M. Heideman, Eline Boon, Suzanne C. van Es, Sandra A. Radema, Carla M.L. van Herpen, Derk Jan A. de Groot, Elisabeth G.E. de Vries, Maurice P.H.M. Jansen, Stefan Sleijfer, Henk M.W. Verheul

* Both authors contributed equally to this work

Abstract

In metastatic colorectal cancer, *RAS* and *BRAF* mutations cause resistance to anti-EGFR therapies, such as cetuximab. Heterogeneity in *RAS* and *BRAF* mutations might explain non-response in a subset of patients receiving cetuximab. Analyzing mutations in plasma-derived circulating tumor DNA (ctDNA) could provide a more comprehensive overview of the mutational landscape as compared to analyses of primary and or metastatic tumor tissue. Therefore, this prospective multicenter study followed 34 patients with metastatic colorectal cancer who were tissue-tested as *RAS* wild-type (exons 2-4) during routine work-up and received third-line cetuximab monotherapy. *BRAF* mutation status was also tested but did not exclude patients from therapy. At baseline and upon disease progression, cell-free DNA (cfDNA) was isolated for targeted next-generation sequencing (NGS). At 8 weeks, we determined which patients had benefited from treatment. NGS of cfDNA identified three patients with *RAS* mutations not detected in tumor tissue during routine work-up. Another six patients had a *BRAF* or rare *RAS* mutation in ctDNA and/or tumor tissue. Relative to patients without mutations in *RAS/BRAF*, patients with mutations at baseline had shorter progression-free survival (1.8 versus 4.9 months (P <0.001)) and overall survival (3.1 versus 9.4 months (P=0.001)). In patients with clinical benefit (progressive disease after 8 weeks), ctDNA testing revealed previously undetected mutations in *RAS/BRAF* (71%) and *EGFR* (47%), which often emerged polyclonally. Our results indicate that baseline NGS of ctDNA can identify additional *RAS* mutation-carriers which could improve patient selection for anti-EGFR therapies. Acquired resistance, in patients with initial treatment benefit, is mainly explained by polyclonal emergence of *RAS*, *BRAF* and *EGFR* mutations in ctDNA.

Introduction

Patients with metastatic colorectal cancer (mCRC), harboring *RAS* mutations, do not benefit from anti-epidermal growth factor receptor (EGFR) monoclonal antibodies (MoAbs) such as cetuximab and panitumumab[1]. Despite patient selection for anti-EGFR MoAbs based on *RAS* mutations in the tumor, only 40-45% of patients with wild-type mCRC have clinical benefit resulting in partial responses in 8-13% and stable disease in 32% of patients[2-5]. Alternative biomarkers to predict treatment benefit are under investigation, including imaging of tumor uptake of cetuximab and early response evaluation with [^{18}F]FDG PET, but have not led to clinical implementation so far[6,7]. In addition to *RAS* mutations, recent meta-analyses demonstrated that *BRAF* mutated mCRC – which occurs in 8-10% of patients with *RAS* wild-type mCRC – also fails to respond to anti-EFGR MoAbs[8,9]. Consequently, patients with somatic *BRAF* p.V600E mutations are currently excluded from these therapies in clinical practice as well as in prospective clinical trials.

A potential explanation for the lack of response in patients with *RAS* and *BRAF* wild-type tumors is the presence of intralesional and interlesional differences in mutational status. Although high concordance rates have been described in some studies[10], others do report heterogeneity in *RAS* and *BRAF* mutations ranging from 5 to 32% between the primary tumor and metastatic sites[10-14]. Tumor heterogeneity could result in missed *RAS* and *BRAF* mutated sub clones, present under the detection limit of the assay or not present in the evaluated part of the tumors. In particular, the potential difference between primary tumor and metastatic site is of high relevance since in daily clinical practice primary tumor tissue is frequently being used to assess the mutational status of an individual's tumor, leaving mutations in metastatic cells undetected. This may result in non-response when a patient is treated in the metastatic setting.

Consequently, assessment of the mutational status of metastatic tissue prior to treatment with anti-EGFR MoAbs is important. Although a biopsy from a metastatic lesion can be taken, this is a cumbersome procedure for patients and repetitive sampling is frequently not feasible. An alternative approach to identify the complexity and heterogeneity of all metastatic lesions in a minimally invasive manner is the analysis of plasma derived circulating tumor DNA (ctDNA) in cell-free DNA (cfDNA) which consists of both healthy and tumor derived DNA. ctDNA comprises of short DNA fragments derived from tumor cells and theoretically represents the whole mutational landscape of all metastatic sites. Consequently, ctDNA might give a more accurate representation of the entire mutational profile than a single tumor tissue biopsy.

In untreated patients who started with anti-EGFR blockade in combination with chemotherapy, it has been shown that oncogenic mutations as *KRAS* and *BRAF* can be detected in ctDNA[15]. In addition, it has been described that mutations can appear in the circulation after acquired resistance in patients with initially wild-type disease[16,17]. However, most studies have described the mutational status in ctDNA by analyzing a limited number of genes and in patients treated with combination therapies of a chemotherapy backbone combined with cetuximab[18-20], which makes the interpretation of results with respect to anti-EGFR MoAbs alone difficult.

In this prospective multicenter study, we report the mutational analyses of ctDNA in a unique cohort of 34 tissue-tested *RAS* wild-type (codon 12, 13 (exon 2), 59, 61 (exon 3), 117, 146 (exon 4)) mCRC patients treated with third-line cetuximab monotherapy. Blood samples were collected prior to cetuximab therapy, during therapy and at disease progression. Mutations in ctDNA were measured by a large panel of 14 genes (236 hotspots), including *KRAS, NRAS, EGFR* and *PIK3CA,* using a targeted Next Generation Sequencing (NGS) approach with molecular barcoding. This approach allowed us to evaluate genetic profiles under the sole effect of cetuximab therapy. The aim of this study was to assess if ctDNA could further improve patient selection for anti-EGFR MoAb therapy. In addition, we aimed to gain more insight into the underlying mechanisms for acquired resistance to anti-EGFR MoAb monotherapy.

Materials and Methods

Study design and patients

The IMPACT-CRC is a prospective phase I – II multicenter interventional study (registered with ClinicalTrials.gov, number NCT02117466) to evaluate the predictive value of [^{89}Zr]cetuximab PET scans for cetuximab treatment response. As part of this study, plasma for cfDNA analyses was collected at baseline, after 2 weeks of treatment and at disease progression. All patients received cetuximab monotherapy as third-line palliative systemic treatment. All 34 patients started with 500mg/m^2 every other week. Based on the [^{89}Zr]cetuximab PET/CT eight patients received a higher dose cetuximab (750 – 1250 mg/m^2), whereas 26 patients continued with 500 mg/m^2 (manuscript in preparation). Patients were included in Amsterdam UMC, Vrije Universiteit Amsterdam, University Medical Center Groningen and Radboud University Medical Center. The study was performed in accordance with the Declaration of Helsinki and approved by the Medical Research Ethics Committee of the Amsterdam UMC, Vrije Universiteit Amsterdam. All patients gave written informed consent prior to study procedures.

Patients were eligible for inclusion if they had unresectable *RAS* wild-type metastatic colorectal cancer, had been treated with or had contra-indications for standard chemotherapy (fluoropyrimidine, irinotecan and oxaliplatin), and were naive for anti-EGFR MoAbs. In all patients, mutational analysis was performed as part of routine clinical work-up on either primary or metastatic tumor tissue and had to be *RAS* wild-type. *RAS* wild-type was defined as wild-type in codon 12, 13 (exon 2), 59, 61 (exon 3), 117, 146 (exon 4) of *KRAS* and *NRAS*. Patients with *BRAF* p.V600E mutations were allowed per protocol to participate, since only recently became clear that these patients do also not respond to anti-EGFR MoAbs[8,9].

Clinical outcome was defined as no clinical benefit for patients having progressive disease at 8 weeks and as clinical benefit for patients with stable disease or partial response according to RECISTv1.1 at 8 weeks[21]. Additionally, progression-free survival (PFS) and overall survival (OS) were evaluated, defined as the period between the first treatment cycle until progressive disease or death, respectively. Patients that were still on-treatment and/or alive at the last follow-up date (1st of December 2017) were censored.

Plasma sample collection and handling

Prior to the first cetuximab cycle (baseline), after 2 weeks of treatment and at progressive disease 18 ml of blood was drawn in Vacutainer® EDTA tubes (BD, Franklin Lakes, NJ). Plasma was isolated within 1 hour after blood collection performing two sequential centrifugation steps: 10 minutes 820g at room temperature (RT) with brakes off, and 20,000g for 10 minutes at RT. After centrifugation, plasma was snap frozen and stored at -80 °C until further handling.

Tumor tissue handling

According to standard of care, before start with cetuximab therapy, formalin-fixed paraffin-embedded material of the primary tumor and/ or metastasis was tested for *RAS* (exon 2-4) and *BRAF* (exon 15) if the tumor percentage was ≥ 20% on hematoxylin eosin immunohistochemistry staining. For all patients included in the Amsterdam UMC, Vrije Universiteit Amsterdam, a TruSeq Amplicon Cancer Panel (TSACP; Illumina Inc, San Diego, CA) was used as described previously[22]. In case tumor tissue was of insufficient quality for TSACP-MiSeq-NGS, a high resolution melting technology-based approach followed by direct sequencing to determine *RAS* and *BRAF* mutations was performed[23,24]. For all patients included in University Medical Center Groningen and Radboud University Medical Center multiplex PCR and PGM/ Ion Torrent (Life Technologies) sequence analyses was used as described previously[25]. Multicenter

comparison of mutation testing for *RAS* and *BRAF* previously demonstrated an excellent reproducibility between these Dutch centers[25].

In addition to routine work-up, some patients underwent an additional biopsy prior to cetuximab therapy, which was analyzed via the Center for Personalized Cancer Treatment (CPCT; NCT01855477). This Dutch consortium offers next generation whole genome sequencing of snap-frozen tumor material for the discovery of tumor mutations. To identify true somatic mutations, germline DNA collected from whole blood was sequenced in the same fashion as reference to tumor tissue[26]. The sequencing data of this CPCT biopsy came available after start of cetuximab therapy and did not influence clinical decision-making.

cfDNA isolation and quantification

For cfDNA isolation, plasma samples were thawed and 4 mL plasma was used. cfDNA isolation was performed for all 34 patients at baseline and 27 patients at disease progression. Additionally, for nine patients with clinical benefit, cfDNA was isolated from plasma collected after two weeks of treatment. cfDNA was isolated and eluted in 60µL buffer using the QiaSymphony Circulating DNA kit (Qiagen, Venlo, The Netherlands) as per manufacturer's instructions and stored at -20°C. CfDNA concentrations were quantified using the Quant-iT dsDNA high-sensitivity assay (Invitrogen, Life Technologies, Carlsbad, CA) according to the manufacturer's instructions, and the Qubit fluorometer (Invitrogen) was used as read out.

Targeted NGS and digital PCR

A targeted NGS approach with molecular barcoding using Oncomine™ Colon cfDNA Assay (Thermo Fisher Scientific) was applied for low limit (down to 0.1%) somatic variant detection according the manufacturer's instructions. This assay consists of 14 colorectal cancer-specific genes covering 236 hotspots and indels in 49 amplicons, including *AKT1, APC, BRAF, CTNNB1, EGFR, FBXW7, GNAS, HER2, KRAS, MAP2K1, NRAS, PIK3CA, SMAD4* and *TP53*. CfDNA samples were thawed at RT and a maximum volume input of 13µL of the cfDNA eluate was used, unless the amount of cfDNA in this volume exceeded an input of 20ng cfDNA, then 20 ng cfDNA was used. This amount was used to standardize cfDNA input for targeted NGS between patients and allowed us to achieve a limit of detection of 0.1% (1 mutant copy in a background of 1,000 wild-type copies). Samples with cfDNA concentrations <1,5ng/µL (33/69 (48%) samples), were concentrated using the Eppendorf™, Vacufuge™ Concentrator (Fisher Scientific, MA). Baseline and PD samples originating from the same patient were sequenced within the same run.

Analyses were done as previously reported, using Ion S5 XL sequencing system and 540 chips, and evaluated with a standard variant calling pipeline[27]. First, raw Ion S5 sequencing results with the Oncomine cfDNA assays were loaded into the TorrentSuite variant caller 5.6. Applying additional filtering, hotspot variants were called when at least 1,000 unique molecules for that particular position were sequenced to achieve sufficient coverage for a limit of detection of 0.1% and if the mutant sequence was covered in 3 unique molecules and 10 reads (i.e., 3 reads per unique molecule).

cfDNA samples from two patients who harbored a *BRAF* p.V600E mutation in their tumor tissue and of whom the cfDNA analyses was negative according to targeted NGS (one sample failed during NGS, the other one tested wild-type), were additionally tested for this mutation using a validated digital polymerase chain reaction (dPCR) assay (TaqMan® SNP genotyping assays (ThermoFisher Scientific, Waltham, MA)), as described previously[28].

Tumor load

To compare the total measured cfDNA and ctDNA (mutant copies/mL plasma) with the tumor burden in a patient, we evaluated tumor load on CT and [^{18}F]FDG PET/CT scan. On the baseline diagnostic CT scan, the total number of metastases was evaluated per patient. Additionally, the sum of diameters of all tumor lesions was calculated.

Baseline [^{18}F]FDG PET scan was performed within 2 weeks before the first treatment with cetuximab. The PET scans were created according to EANM guidelines[29]. Briefly, patients fasted 6 hours before tracer injection (target serum glucose ≤7 mmol/l). Mid-femur-skull vertex PET-CT was performed 60 minutes (±5 min) after injection of [^{18}F]FDG (3-4 MBq/kg), combined with low-dose CT (120 kVp, 50 mAs). PET data were normalized and corrected for scatter and randoms, attenuation and decay. Tumor load on [^{18}F]FDG PET scan is expressed as metabolically active tumor volume (MATV), which was calculated using a threshold of 50% of peak standard uptake value to define tumor volume.

Statistical methods

All statistical analyses were performed using IBM SPSS version 24. A *P*-value below 0.05 was used as cut-off for significance. To compare the presence of a mutation with treatment benefit a Fisher's Exact test was used. For survival analysis, patients without progression and patients that are still alive on December 1st 2017 were censored.

Univariate analysis was done using Kaplan-Meier curves and Log Rank tests. With univariate and multivariate Cox regression, Hazard ratios (HRs) were calculated (enter method). To correlate the concentration of ng cfDNA per mL plasma with the total volume of tumor load a Spearman ρ was used.

Results

Patients, plasma and tumor tissue characteristics

In total 34 patients were included from May 2014 until December 2016, patient characteristics are described in **Table 1**. At the time of analyses (December 2017) all patients had progressed and 29 (85.3%) had died. Of all patients, 13 (38%) did not have treatment benefit. The median PFS of the whole cohort was 4.0 months (95% CI 2.7 – 5.2) and median OS was 9.0 months (95% CI 6.0 – 12.1).

Table 1 - Baseline patient characteristics

Characteristics	Clinical benefit (%)	No clinical benefit (%)	Total (%)
No. patients	21 (62)	13 (38)	34 (100)
Median age (range)	64 (50-82)	64 (55-78)	64 (50-82)
Male gender	17 (81)	8 (62)	25 (73.5)
WHO performance status			
0	6 (28.6)	3 (23.1)	9 (26.5)
1	14 (66.7)	8 (61.5)	22 (64.7)
2	1 (4.8)	2 (15.4)	3 (8.8)
Primary tumor			
Right-sided	1 (4.8)	8 (61.5)	9 (26.5)
Left-sided	20 (95.2)	5 (38.5)	25 (73.5)
Previous treatments			
Fluoropyrimidine	21 (100)	13 (100)	34 (100)
Oxaliplatin	21 (100)	13 (100)	34 (100)
Irinotecan	18 (85.7)	13 (100)	31 (91.4)
Bevacizumab	15 (71.4)	8 (61.5)	23 (67.6)
Sunitinib	1 (4.8)	0	1 (2.9)
RECIST evaluation after 8 weeks			
PD	0	13 (100)	13 (38.2)
SD	18 (85.7)	0	18 (52.9)
PR	3 (14.3)	0	3 (8.8)
cfDNA			
median cfDNA concentration in ng/mL plasma (range)	46.5 (6.6-111)	54 (5.5-174)	49.4 (5.5-174)
KRAS/BRAF mutations	1 (4.8)	7 (53.8)	8 (23.5)
Median MATV on [^{18}F] FDG PET (range)	148 (14-1189)	156 (40-805)	152 (14-1189)
PD at time of analysis	21 (100)	13 (100)	34 (100)
Deceased at time of analysis	16 (76.2)	13 (100)	29 (85.3)

Abbreviations: PD, progressive disease; SD, stable disease; PR, partial response

Plasma isolation and raw analysis of samples

The median cfNDA concentration at baseline was 49.4 ng/mL plasma (range 5.5 – 784 ng/mL plasma) and at progressive disease 30.8 ng/mL (range 4.91 – 228 ng/mL plasma). A median of 20 ng (range 11.5 – 33.6 ng) was sequenced on the Ion S5 platform (**Supplementary Table S1**). Variants were called based on our definition of a true positive (molecular coverage of ≥1000, and ≥ 10 mutant reads, and ≥3 mutated unique molecules). Five hotspots variants, which had a molecular coverage <1,000 were also considered true positives as these variants were detected in another sample collected at a different time point as well or if the hotspot was also detected in tumor tissue. The median molecular coverage of all amplicons was 2,851 (range 0 – 20,000) and the median molecular coverage of mutated hotpots was 3,436 (range 71 – 9,641). In total 3 samples failed during the sequencing process and were omitted from further analyses (**Supplementary Table S2**). In summary, successful sequencing results were obtained from 33 of 34 baseline samples, from 7 of 9 2-week samples and all 26 samples at progression (**Fig. 1**).

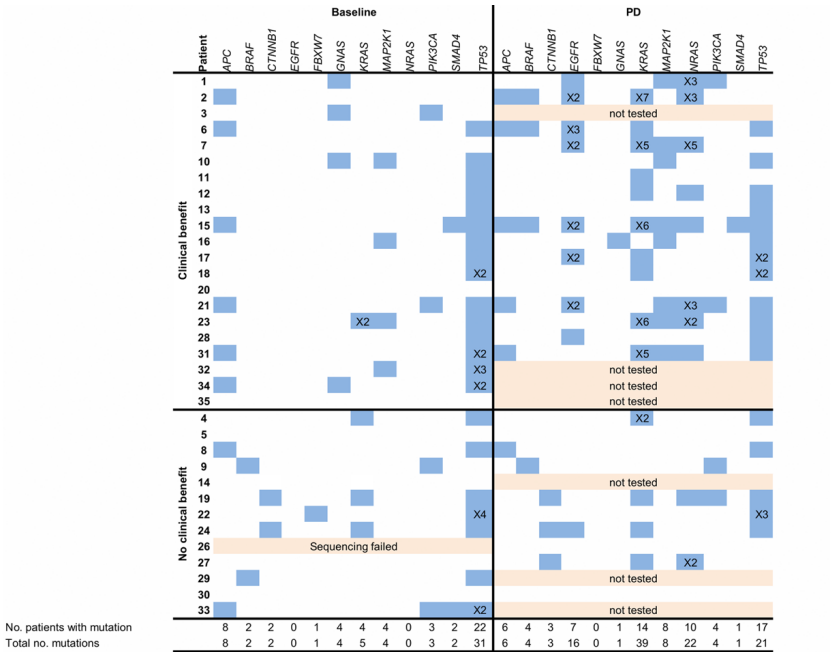


Figure 1 - Comparison of mutational status as determined by ctDNA analyses at baseline and progressive disease (PD) in patients with and without clinical benefit. The number behind the "X" indicates the number of hotspot mutations within a gene.

RAS and *BRAF* mutations in tissue and ctDNA at baseline

Tissue versus ctDNA

Sequencing results of baseline ctDNA obtained prior to start of cetuximab therapy were compared to the mutational status found in routinely tested tumor tissue (**Table 2**). In patients with treatment benefit (n=21), no mutations in *KRAS*, *NRAS* and *BRAF* were detected in tumor tissue. In ctDNA, however, a polyclonal mutation in codon 12 and 61 of *KRAS* was found in one patient (no. 23).

In patients without treatment benefit (n=13), four *BRAF* p.V600E mutations and one rare *KRAS* p.G60D mutation were detected in tumor tissue. Three of four *BRAF* p.V600E mutations were also detected in baseline ctDNA. In one patient (no. 26), sequencing of baseline ctDNA failed, but *BRAF* p.V600E status was assessed by a dPCR confirming the presence of the *BRAF* mutation at a mutant allele frequency (MAF) of 6.6%. In another patient (no. 14), the *BRAF* mutation was not detected in ctDNA by both sequencing and dPCR. No additional *BRAF* mutations over tumor tissue-testing were identified in baseline ctDNA. The *KRAS* p.G60D mutation was confirmed in ctDNA and two additional *KRAS* mutations were detected in ctDNA of patients 4 and 19 which were not detected in tumor tissue.

Table 2 - Baseline mutations in genes: *BRAF, KRAS, NRAS*

	Non Responders (*n* = 13)			**Responders (*n* = 21)**		
Genes	*Patient*	*Tissue (MAF%)*	*cfDNA (MAF%)*	*Patient*	*Tissue (MAF%)*	*cfDNA (MAF%)*
BRAF	9	p.V600E (13)	p.V600E (1.97)		-	-
	14	p.V600E (29)	-			
	26	p.V600E (34)	p.V600E (6.6)[a]			
	29	p.V600E (44)	p.V600E (46.49)			
KRAS	4	-	p.G12A (1.34)	23	-	p.Q61H (0.38)
	19	-	p.Q61H (0.06)	23	-	p.G12A (0.15)
	24	p.G60D (43)	p.G60D (25.97)[b]			
	33	p.S89P (44)[c]	-			
NRAS		-	-		-	-

Mutations detected in tumor tissue during routine work-up and in cfDNA prior to start of cetuximab monotherapy. Mutations detected in tumor tissue and ctDNA are expressed in mutant allele frequency (MAF).
[a] NGS failed, *BRAF* p.V600E was detected by dPCR.
[b] This patient received cetuximab despite having a *KRAS* mutation, as mutations in codon 60 were not an exclusion criteria.
[c] *KRAS* mutation detected by WGS, this test result came available after treatment initiation. This hotspot is not covered by the Oncomine™ Colon cfDNA Assay.

Additional tissue analysis

For eight patients, mutational analyses were performed on two tumor tissue samples obtained prior to start of treatment (**Supplementary Table S3**). Additional sequencing

results came available after start of treatment and therefore did not influence clinical decision making.

In two patients a *KRAS* mutation was found after an initially *RAS* wild-type test. Both *KRAS* mutations were rare and not known as resistance inducing mutations, i.e., codon 89 (*KRAS* p.S89P) and codon 60 (*KRAS* p.G60D). The first mutation was not covered by the initial *RAS* analysis; the latter was covered by the initial sequencing panel, but was not detected in the initial sample.

RAS/BRAF mutations in ctDNA and tumor tissue are predictive for treatment response

Patients with any *RAS/BRAF* mutations in either tumor tissue or ctDNA had less treatment benefit than patients who had a negative test result. Eight of 13 (61.5%) patients without clinical benefit had a *RAS/BRAF* mutation versus one out of 21 (4.8%) patients with clinical benefit ($P = 0.001$). PFS was shorter for patients with *RAS*/*BRAF* mutations, with a median PFS of 1.8 months versus 4.9 months in wild-type patients ($P < 0.001$, HR 4.3; 95% CI 1.8 – 10.0, **Fig. 2A**). In multivariate analysis, correcting for WHO performance status (0 versus 1-2) and left versus right-sidedness, any *RAS* or *BRAF* mutation remained correlated with PFS ($P = 0.004$, HR 4.3; 95% CI 1.6 – 11.6). In line with PFS, OS was shorter in patients with *RAS*/*BRAF* mutated disease, with a median of 3.1 versus 9.4 months ($P = 0.001$, HR 3.9; 95% CI 1.6 – 9.3, **Fig. 2B**). Also, with multivariate analysis, corrected for sidedness and WHO performance status, any *RAS*/*BRAF* mutation remained correlated with OS ($P = 0.007$, HR 5.8; 95% CI 1.6 – 20.7).

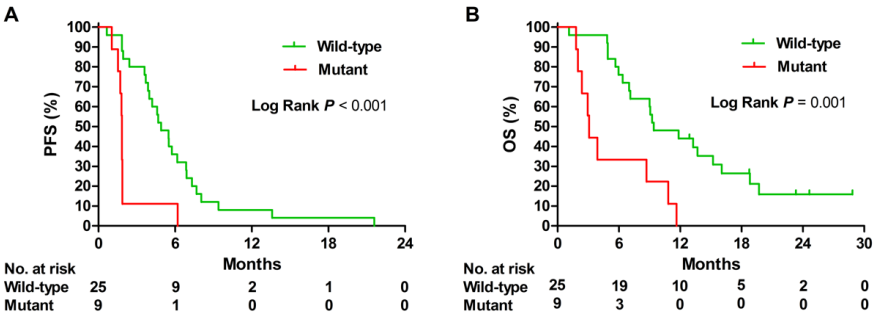


Figure 2 - Progression-free survival (PFS) (A) and overall survival (OS) (B) for patients with *RAS* and/or *BRAF* mutations (mutant) versus patients without *RAS/BRAF* mutations (wild-type) in tissue and ctDNA.

Comparison of mutations in ctDNA: baseline, 2 weeks on treatment and at progressive disease

ctDNA mutations at baseline versus 2 weeks

For nine patients with clinical benefit plasma obtained after 2 weeks of treatment was available for cfDNA analyses. CfDNA concentrations decreased from a median of 44.7 ng/mL plasma (range 13.3 – 784 ng/mL plasma) at baseline to 18.9 ng/mL plasma (range 7.4 – 41.7 ng/mL plasma) after 2 weeks of cetuximab treatment (P = 0.008), **Supplementary Fig. S1**. Paired sequencing results showed that the MAF of dominant tumor clones present at baseline decreased after 2 weeks of treatment, suggesting a reduction in ctDNA load (**Fig. 3**). Detailed information on positions of mutations, MAF and number of mutant molecules per mL plasma is available in **Supplementary Table S4**.

ctDNA mutations at baseline versus at progressive disease

To explore mechanisms of resistance, we compared the mutational signature at baseline and at disease progression. Paired cfDNA sequencing results were available for 17 patients with clinical benefit and 8 patients without clinical benefit.

In 17 patients with initial clinical benefit an evident increase in mutations in well-known resistance inducing genes as *KRAS*, *NRAS* and *BRAF* was observed at the time of progression (median sampling after 25 weeks (range 16 – 94 weeks)) (**Fig. 1**). Twelve patients (71%) had mutations in *KRAS* (n=10) either or not combined with a mutation in *NRAS* (n=8) and/or *BRAF* (n=3) at disease progression. The total number of mutations in *KRAS* increased from 2 at baseline to 34 at progressive disease, for *NRAS* from 0 to 19 and for *BRAF* from 0 to 3, respectively. Polyclonal *KRAS* mutations were present in one patient at baseline and in five patients at progressive disease. Polyclonal mutations in *NRAS* were present in five patients at progressive disease. For example, patient 23, who already harbored two *KRAS* mutations next to a dominant mutation in *TP53* at baseline (21%), showed a marked decrease of the dominant *TP53* mutation after 2 weeks (2%) of treatment and gained 4 *KRAS* and 2 *NRAS* mutations next to a clear increase of the *TP53* mutation (15%) at progressive disease (**Supplementary Fig. S2**).

In addition to the already established resistance inducing genes, the progression samples of patients with initial response to anti-EGFR MoAbs were also enriched for *EGFR* mutations. Mutations in *EGFR* were detected in 8/17 (47%) patients at disease progression, which were not present at baseline, neither in ctDNA nor in tumor tissue. In 6/8 of patients with an *EGFR* mutation, polyclonal mutations occurred. These *EGFR* mutations were located in codon 464, 465, and 492, and code for the epitope binding site of cetuximab[30]. In addition, the number of patients harboring *MAP2K1*

mutations increased from four at baseline to eight at progression. Taken together, at disease progression 15/17 patients (88%) had a mutation related to anti-EGFR MoAbs resistance (12 patients with *RAS* mutations, two patients with only *MAP2K1* mutations and one patient with only an *EGFR* mutation). Mutated genes and the number of unique mutations per gene at baseline and progressive disease are depicted in **Fig. 4A**.

In patients without clinical benefit, baseline and progressive disease (median sampling after 8 weeks, range 3 – 10 weeks) ctDNA mutation analyses demonstrated only a few differences (**Fig. 1**). Only one patient, without baseline mutations in ctDNA nor tumor tissue gained mutations in *KRAS*, *NRAS* and *BRAF* at progression in ctDNA. Patients 4, 19 and 24 gained all one additional mutation at progression; a *KRAS, NRAS,* and *EGFR* mutation, respectively (**Fig. 4B**).

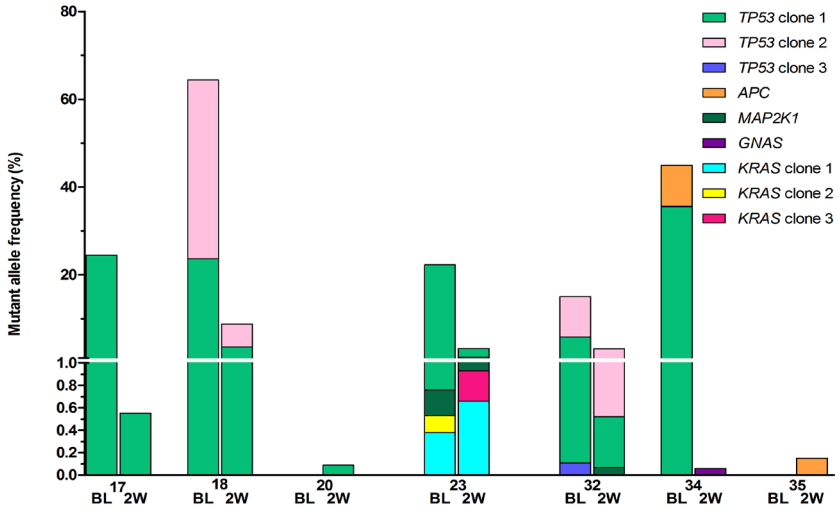


Figure 3 - Paired baseline and 2 week sequencing results of patients with clinical benefit. Mutations were grouped per gene and if patients harbored polyclonal mutations, the clones were numbered. For example, in patient 18 two *TP53* mutations were detected at baseline, clone 1 and 2, which both decreased in MAF at 2 weeks.

Baseline ctDNA mutations: clinical benefit versus no clinical benefit

Baseline ctDNA of patients without clinical benefit was compared to baseline ctDNA of patients with initial clinical benefit to define whether there were differences in affected genes beyond *KRAS, NRAS* and *BRAF* mutations. *APC*, *TP53*, *MAP2K1*, *SMAD4* and *PIK3CA* mutations were present in baseline ctDNA samples of both patient groups. Mutations in *CTNNB1* were only present in baseline samples of two patients without

treatment benefit. However, both *CTNNB1* mutations were present together with a *KRAS* mutation. *CTNNB1* is associated with constitutive RAF/MEK/ERK pathway activation [31]. An overview of all mutations, in tissue and ctDNA from all time points is shown in **Supplementary Table S4**.

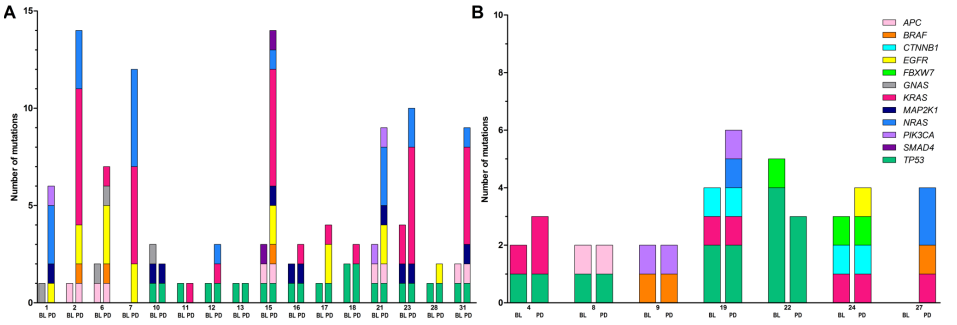


Figure 4 - Paired baseline and progressive disease ctDNA mutational analyses in patients with initial clinical benefit (A) and patients without clinical benefit (B). Mutations are depicted per gene, each gene having a separate color. Higher bars indicate polyclonal mutations. For example patient 2 gained seven different KRAS hotspot mutations at disease progression. Patient 20 with clinical benefit and patient 30 without clinical benefit were not included in the graph because of absence of mutations at baseline as well as progressive disease.

Left- versus right-sided mCRC

Based on tissue-tested mutation analyses, six out of nine patients with right-sided mCRC had a *RAS* or *BRAF* mutation. Incorporating the ctDNA mutation analyses, eight out of nine patients with right-sided mCRC had a *RAS* or *BRAF* mutation ($P < 0.001$). The one patient with right-sided mCRC without any *RAS* or *BRAF* mutation experienced treatment benefit, with disease control of almost 14 months and a censored OS of 23 months. Only one patient (1/25, 4%) with left-sided mCRC had a polyclonal *KRAS* mutation in ctDNA analysis and was free of progression for 6.2 months and died 8.7 months after start of cetuximab therapy.

Tumor load versus cfDNA concentration

The sum of diameters of all metastases per patient did correlate to baseline cfDNA concentration ($P = 0.033$) and also the number of metastases (median 5.5 lesions, range 1– 15) did correlate with cfDNA concentration ($P = 0.037$). Moreover, the MATV on [^{18}F]-FDG PET highly correlated with baseline concentration cfDNA (ng cfDNA/ mL plasma)

(Spearman ρ 0.67, $P < 0.001$; **Supplementary Fig. 3A**). In addition, the total number of hotspot mutant molecules per mL plasma as a surrogate for mutational load also correlated with MATV on [^{18}F] FDG PET (Spearman ρ 0.50, $P = 0.003$) (**Supplementary Fig. 3B**).

Discussion

The results of the current study indicate that a subset of patients with *RAS* wild-type tumors who have no clinical benefit on cetuximab monotherapy do have *KRAS* mutations in ctDNA. Our analysis of patients' baseline ctDNA revealed three additional patients who had *KRAS* mutations (*KRAS* p.G12A, p.G61H and a combination of the two) that had not been detected in tumor tissue. These discordant findings between tumor-tissue and ctDNA are in line with previous reports that have demonstrated that mutations can be heterogeneous within primary tumor lesions, between synchronous lesions and between metastases[32-36]. Apart from such tumor heterogeneity, the sensitivity of sequencing assays used in tumor tissue testing could also have led to false negative results since most of the clinically used assays have a limit of detection of MAF >5%[37]. This hypothesis has recently been supported by Khan *et al*. who showed that *RAS* mutations in ctDNA could be confirmed in tumor tissue at low frequencies by using deep sequencing[38]. The authors found that the MAFs of mutations detected in tumor tissue were indeed below the limit of detection of clinically used techniques. Furthermore, *KRAS* mutations detected in ctDNA at baseline were also detected at disease progression with higher MAFs, endorsing that *KRAS* is truly mutated in these cetuximab naive patients.

While most patients had known resistance-inducing mutations, one patient harbored a rare *KRAS* p.G60D mutation in both tissue and ctDNA. Since this mutation was not in one of the codons known to be resistance-inducing – and there has been anecdotal evidence of a patient with a p.G60D mutation having a partial response to cetuximab – this patient was allowed to participate in the study, but did not benefit from therapy[39].

As this study included only those patients who had *KRAS* and *NRAS* wild-type disease based on tumor tissue testing, a comparison of the mutational status in tissue versus ctDNA was not plausible for these genes. Since we included patients with *BRAF* mutations, a comparison of tissue versus ctDNA was possible in our cohort. We detected *BRAF* p.V600E mutations in ctDNA of three patients, in two patients by sequencing and in one by dPCR and these *BRAF* mutations were also present in tumor tissue. One

BRAF p.V600E mutation was present in tumor tissue of a fourth patient but was not detected in ctDNA with targeted NGS nor with an orthogonal technique as dPCR. We suggest three possible reasons for this. Firstly, the molecular coverage of *BRAF* in the NGS experiment for this patient (patient 14) was 709 molecules. This is far lower than the median molecular coverage of *BRAF* of 2191 molecules that we measured in 67 samples, which might explain why this variant was not detected. A second possible explanation is that following surgical removal of the primary tumor that provided tissue for the test, subsequent metastases originated from a different clone that did not carry the *BRAF* mutation. Thirdly, the cfDNA concentration of this patient was low, only 21.9 ng cfDNA/mL plasma, which is much lower than the median baseline cfDNA concentration in our cohort (49.4 ng/mL plasma). Since baseline cfDNA concentration was correlated with tumor load, low cfDNA concentrations could hypothetically lead to false negative results due to the fact the amount of tumor DNA carrying the mutation present in the circulation is simply too low. Nevertheless, for three out of four patients with the mutation in tumor tissue, the *BRAF* mutation was also detected in ctDNA. Although caution is warranted given the small number of patients, a detection rate of 75% is in line with that found in a previous study in non-small cell lung cancer patients: this study compared the detection of the *EGFR* p.T790M mutation in ctDNA with that in tumor and reported a sensitivity of 70%[40].

While almost all patients with additional *KRAS* or *BRAF* mutations were resistant to therapy, we also had one patient with clinical benefit who nevertheless had a polyclonal *KRAS* mutation (p.G61H and p.G12A) in ctDNA, for which we suggest three potential explanations. First, this patient received a cetuximab dose escalation from 500mg/m^2 to 1250mg/m^2, dosed every other week, based on the results of the [^{89}Zr]cetuximab PET scan, which showed no uptake after one cycle of cetuximab (in preparation van Helden *et al.*). Second, stable disease could also be a result of tumor heterogeneity, whereby only a small fraction of tumor cells harbor *KRAS* mutations and the majority are *RAS* wild-type[3,41]. A final possible explanation is that there were other reasons for an indolent disease course regardless of treatment with cetuximab.

Given that the *KRAS* and *BRAF* mutations detected in ctDNA indeed conferring resistance to cetuximab, we were interested to see whether these mutations would be present throughout disease course and whether new mutations would appear. When we analyzed the mutation status in ctDNA at progression, we found that in patients who had shown initial treatment benefit, 12/17 (71%) patients had new *RAS* and/or *BRAF* mutations that were not detected at the start of the study. The fact that nine of these patients (9/12, 75%) had multiple mutations in these genes and codons

suggests that the resistance to anti-EGFR treatment is caused by the emergence of various clones harboring different mutations. Our finding of a relatively high number of patients treated with cetuximab who harbor *RAS* mutations at disease progression is in line with that of a previous study[15,17,19,42]. They reported *RAS* mutations in tumor-tissue and ctDNA in 74% of patients who were mainly being treated with a combination of cetuximab and irinotecan. These mutations are most likely acquired by the tumor as a means of escape from the continuous pressure exerted by anti-EGFR MoAbs. But it is also possible that the mutations are due to tumor heterogeneity resulting in the selection and outgrowth of multiple resistant *RAS*/*BRAF*-mutated sub clones which are below the limit of detection at baseline.

Interestingly, at progression 8/17 patients (47%) with initial benefit had gained an *EGFR* mutation in ctDNA, and for six of these patients these mutations were also polyclonal. *EGFR* mutations in codon 465 were detected in seven patients, in codon 464 in six patients and in codon 492 in two patients. All of these *EGFR* mutations are located in domain III of the receptor and alter the epitope to which cetuximab binds, thereby inhibiting binding of cetuximab to EGFR[30,43-47]. Esposito *et al.* (2013) have suggested that these mutations only occur after treatment with cetuximab, as evidenced by their study of 505 patients, in which mutations in tumor tissue were detected after anti-EGFR therapy but not before. In our cohort, these *EGFR* mutations were also exclusively found at progression, rendering this mutation unsuitable for patient selection. It has been proposed that while these *EGFR* mutations occur after cetuximab therapy, they do not emerge after panitumumab therapy, leaving these tumor cells sensitive to panitumumab therapy[48]. However, given our observation that these mutations are almost always accompanied by other *RAS* or *BRAF* mutations, a treatment switch to panitumumab in *EGFR* mutated patients will probably not result in treatment benefit. Also, given the heterogeneity and convergence of the mutational pattern at progression, targeted blockage of the EGFR pathway will likely be difficult.

Finally, it is worth pointing out our finding of a correlation between the number of mutated molecules per mL plasma and the MATV measured by [^{18}F]FDG PET before treatment. A similar correlation has been described previously in patients with non-small cell lung cancer starting with erolitinib in a palliative setting[49]. To our knowledge, our study is the first to show a similar correlation between the number of mutated molecules and MATV measured by [^{18}F]FDG PET in patients with mCRC. Our study thereby supports the hypothesis that the total number of mutated molecules per mL plasma could serve as a surrogate for tumor load, which has also been described using CT to estimate tumor burden[50]. Important to note is that both techniques, [^{18}F]FDG PET

and ctDNA, are sensitive-limited technologies hampering both techniques to detect low tumor burden. Next to the correlation between mutant molecules and MATV, we also found a correlation between the cfDNA concentration and MATV. It should be noted that the correlation between cfDNA and MATV might be less tumor specific, since cfDNA is composed of a small fraction of tumor DNA while the majority is derived from normal apoptotic tissue and hematological cells[51,52].

There are several limitations of our study including the small sample size. Second, in our study, tumor tissues were sequenced with panels used in daily routine practice. Therefore, comparative analyses of ctDNA and tumor tissue were hampered by the use of different techniques.

Conclusions

NGS of ctDNA in patients with tissue-tested *RAS* wild-type mCRC — tested as part of routine clinical work-up — can identify additional *RAS* mutation-carriers. The majority of patients with initial clinical benefit from cetuximab therapy gain mutations in genes such as *RAS, BRAF* and *EGFR*, frequently occurring in multiple clones within individual patients. Hence, ctDNA analysis is a promising tool to optimize patient selection for anti-EGFR monoclonal antibodies and a minimally invasive method to gain more insight in mechanisms accounting for resistance.

Additional Information

Data accessibility: Raw data is available upon request. Please contact the corresponding author for details.

Acknowledgements

We would like to thank all the investigators and site staff, with special thanks to the patients and their families. We thank Ronald van Marion and Peggy Atmodimedjo for their help with the ctDNA-NGS experiments. We also wish to acknowledge the financial support from the Dutch Cancer Society (Grant IMPACT, RUG 2012-5565).

Funding

This study was financially supported by a grant from the Dutch Cancer Society (Grant IMPACT, RUG 2012-5565).

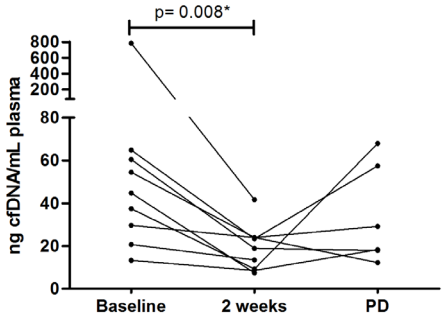
References

1. Sorich, M.J. *et al.* Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann Oncol* **26**, 13-21 (2015).
2. Van Cutsem, E. *et al.* Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol* **33**, 692-700 (2015).
3. Karapetis, C.S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med* **359**, 1757-65 (2008).
4. Lievre, A. *et al.* KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol* **26**, 374-9 (2008).
5. van Helden, E.J. *et al.* Optimal use of anti-EGFR monoclonal antibodies for patients with advanced colorectal cancer: a meta-analysis. *Cancer Metastasis Rev* **36**, 395-406 (2017).
6. van Helden, E.J. *et al.* Early 18F-FDG PET/CT Evaluation Shows Heterogeneous Metabolic Responses to Anti-EGFR Therapy in Patients with Metastatic Colorectal Cancer. *PLoS One* **11**, e0155178 (2016).
7. Menke-van der Houven van Oordt, C.W. *et al.* 89Zr-cetuximab PET imaging in patients with advanced colorectal cancer. *Oncotarget* **6**, 30384-93 (2015).
8. Rowland, A. *et al.* Meta-analysis of BRAF mutation as a predictive biomarker of benefit from anti-EGFR monoclonal antibody therapy for RAS wild-type metastatic colorectal cancer. *Br J Cancer* **112**, 1888-94 (2015).
9. Pietrantonio, F. *et al.* Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: a meta-analysis. *Eur J Cancer* **51**, 587-94 (2015).
10. Vermaat, J.S. *et al.* Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. *Clin Cancer Res* **18**, 688-99 (2012).
11. Artale, S. *et al.* Mutations of KRAS and BRAF in primary and matched metastatic sites of colorectal cancer. *J Clin Oncol* **26**, 4217-9 (2008).
12. Italiano, A. *et al.* KRAS and BRAF mutational status in primary colorectal tumors and related metastatic sites: biological and clinical implications. *Ann Surg Oncol* **17**, 1429-34 (2010).
13. He, Q. *et al.* Comparison of KRAS and PIK3CA gene status between primary tumors and paired metastases in colorectal cancer. *Onco Targets Ther* **9**, 2329-35 (2016).
14. Kim, M.J. *et al.* Different metastatic pattern according to the KRAS mutational status and site-specific discordance of KRAS status in patients with colorectal cancer. *BMC Cancer* **12**, 347 (2012).
15. Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532-6 (2012).
16. Thierry, A.R. *et al.* Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. *Nat Med* **20**, 430-5 (2014).
17. Diaz, L.A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-40 (2012).
18. Thierry, A.R. *et al.* Circulating DNA Demonstrates Convergent Evolution and Common Resistance Mechanisms during Treatment of Colorectal Cancer. *Clin Cancer Res* **23**, 4578-4591 (2017).
19. Van Emburgh, B.O. *et al.* Acquired RAS or EGFR mutations and duration of response to EGFR blockade in colorectal cancer. *Nat Commun* **7**, 13665 (2016).
20. Spindler, K.L. *et al.* Changes in mutational status during third-line treatment for metastatic colorectal cancer--results of consecutive measurement of cell free DNA, KRAS and BRAF in the plasma. *Int J Cancer* **135**, 2215-22 (2014).
21. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-47 (2009).
22. Sie, D. *et al.* Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. *Cell Oncol (Dordr)* **37**, 353-61 (2014).
23. Kramer, D. *et al.* A fast, sensitive and accurate high resolution melting (HRM) technology-based assay to

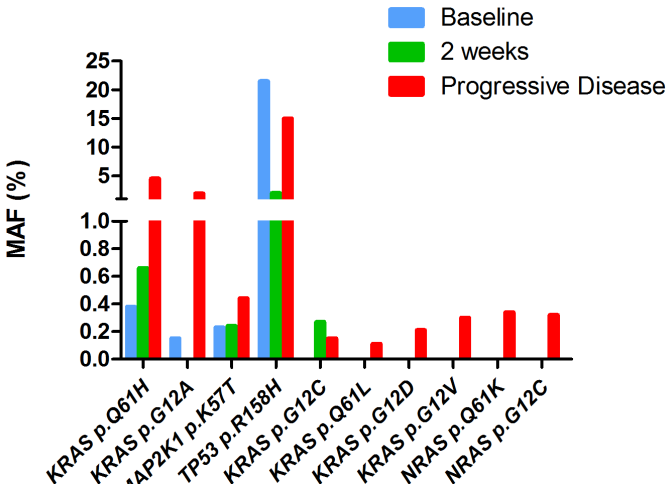
screen for common K-ras mutations. *Cell Oncol* **31**, 161-7 (2009).
24. Heideman, D.A. *et al.* KRAS and BRAF mutation analysis in routine molecular diagnostics: comparison of three testing methods on formalin-fixed, paraffin-embedded tumor-derived DNA. *J Mol Diagn* **14**, 247-55 (2012).
25. Boleij, A. *et al.* RAS testing in metastatic colorectal cancer: excellent reproducibility amongst 17 Dutch pathology centers. *Oncotarget* **6**, 15681-9 (2015).
26. Bijlsma, R.M. *et al.* Unsolicited findings of next-generation sequencing for tumor analysis within a Dutch consortium: clinical daily practice reconsidered. *Eur J Hum Genet* **24**, 1496-500 (2016).
27. Jansen, M.P. *et al.* Cell-free DNA mutations as biomarkers in breast cancer patients receiving tamoxifen. *Oncotarget* **7**, 43412-43418 (2016).
28. van Dessel, L.F. *et al.* Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions. *Mol Oncol* **11**, 295-304 (2017).
29. Boellaard, R. *et al.* FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* **37**, 181-200 (2010).
30. Sickmier, E.A. *et al.* The Panitumumab EGFR Complex Reveals a Binding Mechanism That Overcomes Cetuximab Induced Resistance. *PLoS One* **11**, e0163366 (2016).
31. Malapelle, U. *et al.* Less frequently mutated genes in colorectal cancer: evidences from next-generation sequencing of 653 routine cases. *J Clin Pathol* **69**, 767-71 (2016).
32. Jeantet, M. *et al.* High Intra- and Inter-Tumoral Heterogeneity of RAS Mutations in Colorectal Cancer. *Int J Mol Sci* **17**(2016).
33. Kosmidou, V. *et al.* Tumor heterogeneity revealed by KRAS, BRAF, and PIK3CA pyrosequencing: KRAS and PIK3CA intratumor mutation profile differences and their therapeutic implications. *Hum Mutat* **35**, 329-40 (2014).
34. Normanno, N. *et al.* Heterogeneity of KRAS, NRAS, BRAF and PIK3CA mutations in metastatic colorectal cancer and potential effects on therapy in the CAPRI GOIM trial. *Ann Oncol* **26**, 1710-4 (2015).
35. Oltedal, S. *et al.* Heterogeneous distribution of K-ras mutations in primary colon carcinomas: implications for EGFR-directed therapy. *Int J Colorectal Dis* **26**, 1271-7 (2011).
36. de Macedo, M.P. *et al.* RAS mutations vary between lesions in synchronous primary colorectal cancer: testing only one lesion is not sufficient to guide anti-EGFR treatment decisions. *Oncoscience* **2**, 125-30 (2015).
37. Shackelford, R.E. *et al.* KRAS Testing: A Tool for the Implementation of Personalized Medicine. *Genes Cancer* **3**, 459-66 (2012).
38. Khan, K.H. *et al.* Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov* **8**, 1270-1285 (2018).
39. Molinari, F. *et al.* Increased detection sensitivity for KRAS mutations enhances the prediction of anti-EGFR monoclonal antibody resistance in metastatic colorectal cancer. *Clin Cancer Res* **17**, 4901-14 (2011).
40. Oxnard, G.R. *et al.* Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol* **34**, 3375-82 (2016).
41. Benvenuti, S. *et al.* Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res* **67**, 2643-8 (2007).
42. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 795-801 (2015).
43. Voigt, M. *et al.* Functional dissection of the epidermal growth factor receptor epitopes targeted by panitumumab and cetuximab. *Neoplasia* **14**, 1023-31 (2012).
44. Bertotti, A. *et al.* The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526**, 263-7 (2015).
45. Esposito, C. *et al.* The S492R EGFR ectodomain mutation is never detected in KRAS wild-type colorectal carcinoma before exposure to EGFR monoclonal antibodies. *Cancer Biol Ther* **14**, 1143-6 (2013).

46. Arena, S. *et al.* Emergence of Multiple EGFR Extracellular Mutations during Cetuximab Treatment in Colorectal Cancer. *Clin Cancer Res* **21**, 2157-66 (2015).
47. Arena, S. *et al.* MM-151 overcomes acquired resistance to cetuximab and panitumumab in colorectal cancers harboring EGFR extracellular domain mutations. *Sci Transl Med* **8**, 324ra14 (2016).
48. Montagut, C. *et al.* Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nat Med* **18**, 221-3 (2012).
49. Winther-Larsen, A. *et al.* Correlation between circulating mutant DNA and metabolic tumour burden in advanced non-small cell lung cancer patients. *Br J Cancer* **117**, 704-709 (2017).
50. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-90 (2008).
51. Elshimali, Y.I. *et al.* The clinical utilization of circulating cell free DNA (CCFDNA) in blood of cancer patients. *Int J Mol Sci* **14**, 18925-58 (2013).
52. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res* **61**, 1659-65 (2001).

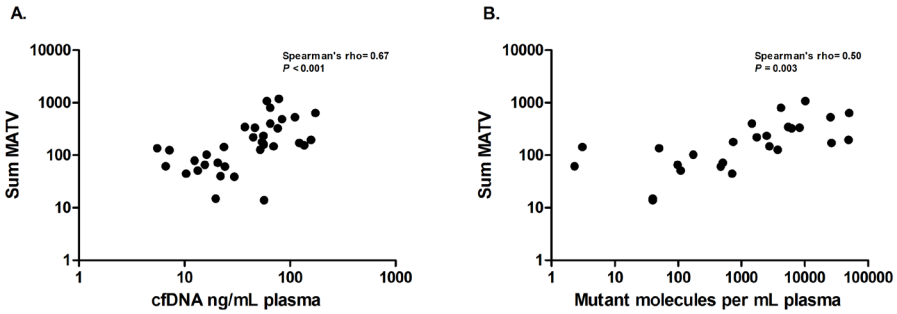
Supplemental Figures



Supplementary Figure S1 - cfDNA concentration measured in matched baseline, 2 weeks and PD samples.
Each line indicates one patient. cfDNA concentrations were available for 9 matched baseline and 2 week samples, and for 6 PD samples.* Related samples Wilcoxon signed rank test.



Supplementary Figure S2 - Patient 23 having a polyclonal *KRAS* mutation present at baseline, a marked decrease in the *TP53* p.R158H mutant allele frequency (MAF) after two weeks of treatment and an increase of the AMF at disease progression accompanied by emergence of four additional *KRAS* and two *NRAS* mutation.



Supplementary Figure S3 - Scatter plot of the concentration cfDNA (in ng per mL plasma) (A) and the number of mutant molecules per mL plasma (B) versus the sum of metabolically active tumor volume (MATV) on [18F] FDG PET scan per patient.

Supplemental Tables

Supplementary Table S1 - cfDNA concentrations and ng DNA input for targeted NGS

			Baseline		2 weeks		Progressive disease	
Patient	Responder	Left/ Right	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing
1	Yes	Left	23,65	23,1	Not tested	Not tested	51,79	20,0
2	Yes	Left	55,93	20,0	Not tested	Not tested	25,46	22,6
3	Yes	Left	69,76	20,0	Not tested	Not tested	Not tested	Not tested
4	No	Right	5,5	16,8	Not tested	Not tested	10,44	20,0
5	No	Left	12,48	20,0	Not tested	Not tested	Not tested	Not tested
6	Yes	Left	46,51	20,0	Not tested	Not tested	34,13	20,0
7	Yes	Left	84,07	20,0	Not tested	Not tested	217,8	20,0
8	No	Left	64,98	20,0	Not tested	Not tested	223,2	21,1
9	No	Right	15,55	18,9	Not tested	Not tested	9,37	12,4
10	Yes	Left	19,8	19,5	Not tested	Not tested	43,05	20,0
11	Yes	Left	6,62	11,5	Not tested	Not tested	4,91	13,1
12	Yes	Left	52,26	20,0	Not tested	Not tested	22,45	20,4
13	Yes	Left	56,8	20,0	Not tested	Not tested	163,36	20,0
14	No	Right	21,94	20,0	Not tested	Not tested	Not tested	Not tested
15	Yes	Left	111,72	20,0	Not tested	Not tested	57,46	20,0
16	Yes	Left	24,15	20,5	Not tested	Not tested	30,14	21,3
17	Yes	Left	54,48	20,0	23,97	21,5	29,15	20,0
18	Yes	Left	60,4	20,0	18,88	20,9	17,96	21,6
19	No	Right	158,06	20,0	Not tested	Not tested	126,37	20,0
20	Yes	Right	29,67	20,0	23,95	21,9	12,22	19,8
21	Yes	Left	37,43	20,0	9,27	21,0	67,96	20,0
22	No	Left	16,21	19,9	Not tested	Not tested	15,42	19,2
23	Yes	Left	64,86	20,0	23,45	21,0	57,44	20,0
24	No	Right	174,42	20,0	Not tested	Not tested	95,9	20,0
26	No	Right	7,21	11,5	Not tested	Not tested	31,37	21,3
27	No	Left	136,97	20,0	Not tested	Not tested	228,38	20,0
28	Yes	Left	13,34	17,4	8,61	20,8	18,27	22,5
29	No	Right	76,71	20,0	Not tested	Not tested	Not tested	Not tested
30	No	Left	56,42	20,0	Not tested	Not tested	17,51	17,8
31	Yes	Left	10,38	17,4	Not tested	Not tested	20,48	22,0
32	Yes	Left	20,71	21,5	13,51	22,7	Not tested	Not tested
33	No	Right	122,92	20,0	Not tested	Not tested	Not tested	Not tested
34	Yes	Left	44,72	20,0	7,42	33,6	Not tested	Not tested
35	Yes	Left	784	20,0	41,70	20,0	Not tested	Not tested

Supplementary Table S2 - Sequencing failures

		Baseline		2 weeks		Progressive disease	
Patient	Responder	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing	cfDNA concentration ng/mL plasma	ng input ionPGM sequencing
21	Yes	37,43	20,0	9,27	21,0	67,96	20,0
26	No	7,21	11,5	Not tested	Not tested	31,37	21,3
28	Yes	13,34	17,4	8,61	20,8	18,27	22,5

Samples in red failed during targeted NGS

Supplementary Table S3 - Double biopsies

Study No.	Date 1st cycle cetuximab	Method	Location	Date Biopsy	*KRAS* (MAF)	*BRAF* (MAF)	*APC* (MAF)	*ATM* (MAF)	*CTNNB1* (MAF)	*EGFR* (MAF)	*TP53* (MAF)	*VHL* (MAF)
2	Sep. 2014	HRM[a]	Primary tumor	2011			Not tested	Not tested	Not tested	Not tested	Not tested	Not tested
		TSACP	Primary tumor	2011			*APC* p.A1351fs*3 (36%)					
18	Dec. 2015	TSACP	Liver metastasis	Nov. 2015			*APC* p.P1427fs (52%)				*TP53* p.G245C (38%)	*VHL* p.R113L (34%)
		WGS	Primary tumor	Dec. 2015			*APC* p.P1427fs*46 (58%)			*EGFR* p.V536M (53%)	*TP53* p.Y163C (49%)	
22	Feb. 2016	TSACP	Liver metastasis	Jan. 2016								
		WGS	Liver metastasis	Jan. 2016							*TP53* p.Y234H (11%)	
29	Apr. 2016	Multiplex PCR[b]	Primary tumor	May. 2015		*BRAF* p.V600E (44%)	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested
		TSACP	Primary tumor	May. 2015		*BRAF* p.V600E (44%)					*TP53* p.R248Q (48%)	
24	Mar. 2016	TSACP	Liver metastasis	Feb. 2014					*CTNNB1* p.T41A (80%)		*TP53* p.S241F (62%)	
		TSACP	Liver metastasis	Mar. 2016	*KRAS* p.G60D (43%)				*CTNNB1* p.T41A (84%)		*TP53* p.S241F (79%)	
31	Jun. 2016	Multiplex PCR	Omentum	Sep. 2014			Not tested	Not tested	Not tested	Not tested	Not tested	Not tested
		TSACP	Omentum	Sep. 2014			*APC* p.E1309fs (33%)				*TP53* p.R273C (36%)	
33	Sep. 2016	HRM	Primary tumor	Mar. 2015	*KRAS* codon 89	not tested	Not tested	Not tested	Not tested	Not tested	Not tested	Not tested
		WGS	Peritoneal metastasis	Sep. 2016	*KRAS* p.S89P (44%)							
34	Nov. 2016	TSACP	Rib metastasis	Oct. 2016			*APC* p.E1306* (15%)				*TP53* p.R282W (54%)	
		WGS	Rib metastasis	Oct. 2016			*APC* p.E1306* (24%)	*ATM* p.P2665R (18%)			*TP53* p.R282W (63%)	

Abbreviations: WGS, whole genome sequencing; MAF, mutant allele frequency

[a] HRM panel, including *KRAS, NRAS, BRAF*

Supplementary Table 4 - Overview of all available mutation data.
Available at: https://febs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2F1878-0261.12550&file=mol212550-sup-0007-TableS4.xlsx

CHAPTER 7

Whole exome sequencing of cell-free DNA - a systematic review and Bayesian individual patient data meta-analysis

Cancer Treatment Reviews. *2020 Feb;83:101951.*

Manouk K. Bos*, **Lindsay Angus***, Kazem Nasserinejad, Agnes Jager, Maurice P.H.M. Jansen, John W.M. Martens, Stefan Sleijfer

* Both authors contributed equally to this work

Abstract

Molecular profiling of tumor derived cell free DNA (cfDNA) is gaining ground as a prognostic and predictive biomarker. However to what extent cfDNA reflects the full metastatic landscape as currently determined by tumor tissue analysis remains controversial. Though technically challenging, whole exome sequencing (WES) of cfDNA enables thorough evaluation of somatic alterations. Here, we review the feasibility of WES of cfDNA and determine the sensitivity of WES-detected single nucleotide variants (SNVs) in cfDNA on individual patient data level using paired tumor tissue as reference ($\frac{shared\ SNVs}{All\ tissue\ SNVs} \chi\ 100\%$). The pooled sensitivity was 50% (95% credible interval (CI): 29%-72%). The tissue mutant allele frequency (MAF) of variants exclusively identified in tissue was significantly lower (12.5%, range: 0.5-18%) than the tissue MAF of variants identified in both tissue and cfDNA (23.9%, range: 17-38%), p=0.004. The overall agreement ($\frac{shared\ SNVs}{All\ SNVs} \chi\ 100\%$) between SNVs in cfDNA and tumor tissue was 31% (95% CI: 15%-49%). The number of detected SNVs was positively correlated with circulating tumor DNA (ctDNA) fraction (p=0.016). A sub analysis of samples with ctDNA fractions ≥25% improved the sensitivity to 69% (95% CI: 46-89%) and agreement to 46% (95% CI: 36-59%), suggesting that WES is mainly feasible for patients with high ctDNA fractions. Pre- and post-analytical procedures were highly variable between studies rendering comparisons problematic. In conclusion, various aspects of WES of cfDNA are largely in its investigative phase, standardization of methodologies is highly needed to bring this promising technique to its clinical potential.

Introduction

Next generation sequencing of tumor tissue is increasingly being performed since more and more targeted treatments require presence of specific genomic alterations[1-3]. Although metastatic tissue can be obtained for this analysis, it is a cumbersome procedure for patients and repetitive sampling is frequently not feasible. Therefore, genomic profiling of plasma derived cell-free DNA (cfDNA) is considered as a minimally-invasive surrogate to predict outcome and predict or monitor treatment efficacy[4].

CfDNA consists of short fragments of DNA derived from normal- and tumor cells (ctDNA). Contrary to a single tumor tissue biopsy, ctDNA might give a more accurate representation of the entire mutational profile present across the different lesions within an individual cancer patient[5-7]. Although significant progress has been made for tracking previously detected tumor mutations using targeted gene panels or single gene assays[8], whole exome sequencing (WES) enables a more comprehensive analysis covering the complex landscape of somatic alterations. Hence, can be used as a tool to gain insight into tumor biology, for example by which genomic mechanisms tumor cells can confer resistance.

In addition, WES enables the identification of genomic signatures such as tumor mutational burden (TMB) and microsatellite instability (MSI), all being recognized as biomarkers for selected therapies such as immunotherapy[9,10]. So, compared to targeted panels comprising a relatively limited number of genes, WES analyses of ctDNA holds great promise to identify emerging genes that are of interest in treatment resistance and to capture DNA signatures important for treatment decision making. However, WES on cfDNA is technically challenging due to the often low tumor fractions in a high background of normal cfDNA.

The aims of this systematic review were to (1) describe to what extent WES of cfDNA in cancer patients is technically feasible and which approaches are being used, and to (2) analyze the sensitivity of WES-detected single nucleotide variants (SNVs) in cfDNA using tumor tissue as reference ($\frac{shared\ SNVs}{All\ tissue\ SNVs}\ \chi\ 100\%$) as well as the agreement between cfDNA and tumor tissue ($\frac{shared\ SNVs}{All\ SNVs}\ \chi\ 100\%$).

Methods

Literature search

PubMed was searched from May 2013 to July 2019 to find full publications. Search terms included *cell free DNA* and *whole exome sequencing*. Also synonyms of the terms

and MeSH terms were used (**Table A1**). For the technical feasibility analysis, studies were eligible if (1) they were written in English (2) WES was used for molecular profiling of cfDNA, and (3) patients had solid tumors. Exclusion criteria were: (1) solely focusing on bioinformatics pipeline not presenting unique data, (2) cfDNA derived from other liquids than blood, and (3) patients without cancer. Subsequently for sensitivity and agreement meta-analyses, studies that reported WES-detected SNVs in cfDNA and matched tumor tissue were included. Studies were excluded if: (1) time between collection of tumor tissue and cfDNA for individual cases exceeded 2 months, and if (2) SNVs in tumor tissue and cfDNA were not reported on individual patient level.

Data extraction

Two authors (M.B. and L.A.) independently performed the article selection and data extraction. For all studies the following data were extracted using a data-extraction form (**Table A2**): year of publication, sample size, cancer type, time between plasma and tissue collection, pre-analytical variables (amount of DNA input, ctDNA fraction), analytical conditions (sequencing methods and coverage), post-analytical conditions (variant calling and analysis), and the mutant allele frequency (MAF) of detected variants. An overview of used source files is available in **Table A3**. In case of discordances the authors reached agreement during a consensus meeting.

Pooled sensitivity and agreement analysis

To calculate a pooled sensitivity and agreement rate of WES-detected SNVs in paired cfDNA and tumor tissue (irrespective of primary- or metastatic lesion) we extracted individual patient data from each study. Per sample we collected the number of "shared SNVs" (SNVs detected in both tumor tissue and cfDNA), SNVs only found in tissue and SNVs only present in cfDNA. Also cfDNA input and sequencing coverage were extracted on individual sample level. Using SNVs detectable in tissue as reference, sensitivity was calculated as follows: $\frac{shared\ SNVs}{All\ tissue\ SNVs} \chi\ 100\%$. The agreement rate between WES-detected SNVs in tumor tissue and cfDNA was calculated as follows: $\frac{shared\ SNVs}{All\ SNVs} \chi\ 100\%$, in which "all SNVs" was defined as: SNVs only detected in tissue + SNVs only detected in cfDNA + shared SNVs. Patients without detectable SNVs in tumor tissue were excluded from the sensitivity and also from the agreement analysis to keep both groups comparable. We did not calculate specificity, since we were unable to calculate the numbers of true negatives (wild type genes).

Additional WES-detected SNVs in cfDNA

For all studies included in the meta-analysis, we extracted the number of additionally detected SNVs in cfDNA for each sample pair and calculated the fraction of uniquely

detected variants versus all variants in cfDNA: $\frac{ctDNA\ variants\ unique\ to\ plasma}{all\ ctDNA\ variants}\ x\ 100\%$. Per study we displayed the median of the individual sample data. To score the clinical potential of exclusively detected SNVs in plasma, we used the clinical annotation database OncoKB[11] (September 1st, 2019). Additionally, per study we scored whether variants detected exclusively in cfDNA had been described previously in the corresponding tumor type using cBioPortal for Cancer Genomics (September 4th, 2019)[12]. We reported SNVs with a MAF ≥2%.

Statistical analysis

An individual patient data (IPD) meta-analysis was used to estimate the overall sensitivity and agreement rates across all the studies. Taking into account the heterogeneity, the patient-specific effects and the study-specific effects were employed as random-effects in the (multilevel) model. For this purpose, Bayesian IPD *meta*-analyses were employed. Results of these analyses were shown using a forest plot, where the median and the 95% highest probability density (hpd) of credible intervals (CI) were reported for each study separately and pooled in an overall sensitivity and agreement rate. A sub-analysis was performed to estimate the sensitivity and agreement on a subset of cfDNA samples which contained an estimated tumor fraction ≥25%.

Computations and graphics were performed in R program language[13]. All Bayesian computations were performed using the Markov Chain Monte Carlo (MCMC) sampler through Jags[14] interface in R program language and relatively non-informative priors were used for the parameters in the model. The MCMC sampling was run for each analysis for 200k iterations after discarding the first 200k iterations (burn-in) to reach the convergence.

To assess the correlation between the total number of SNVs with ctDNA fraction, a Spearman's ρ was used. To compare the MAF in tumor tissue versus cfDNA a Mann-Whitney *U* test was performed.

Results

Data retrieval and study characteristics

In total, 20 studies were included in this review of which the individual patient data of 12 studies were included for the meta-analyses of SNV sensitivity and agreement (**Figure 1**)[5,15-33].

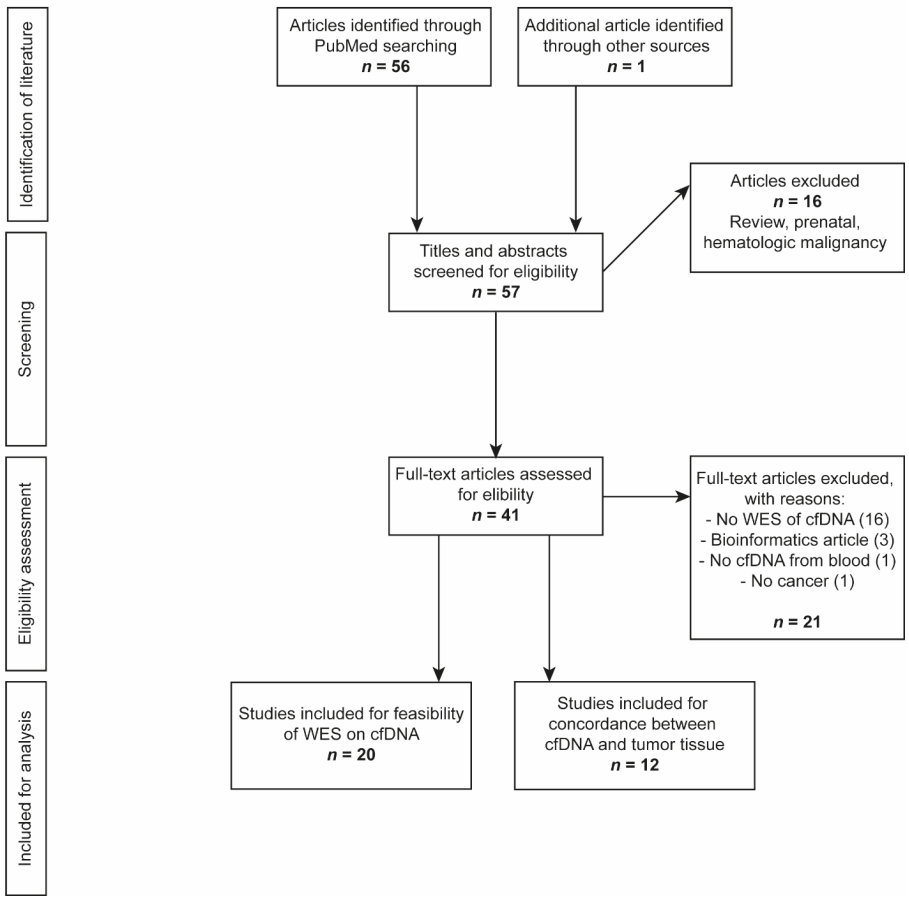


Figure 1 - Flowchart of the selection of publications included in the feasibility analysis and meta-analysis.

Feasibility of WES of cfDNA

To evaluate the technical feasibility of WES of cfDNA we summarized pre-analytical and analytical parameters of all studies performing WES of cfDNA (**Table 1**). In total, WES has been performed on 303 samples, with a median coverage of 137X (range: 43-500X) using a median of 15 ng cfDNA (range: 2-100 ng). Most studies (n= 7) extracted cfDNA from plasma collected in EDTA tubes or did not report the tube type used (n= 6). Four studies performed only WES on samples with a high tumor fraction, i.e. ≥10%[15,17] and ≥25%[33] or "high" was not further specified[26]. Overall, the median tumor fraction of all samples from which individual tumor fractions were available, was 37% based on estimation by different platforms such as ultra-low pass whole genome sequencing (ULP-WGS), Sequenza[34] or maximum MAF of variants.

Table 1 - Pre-analytical and analytical parameters of all studies which performed WES on cfDNA

First author	Year of publication	Pre-selection for WES analysis?	Sample size for WES (N)	Tumor type	Library Prep	Sequencing method	Median input cfDNA (ng)	Median coverage (X)	ctDNA fraction	Method for estimation of ctDNA fraction	Genetic alterations other than mutations identified in cfDNA
Adalsteinsson[15]	2017	Yes, samples with >10% ctDNA based on ULP-WGS	41	Breast and prostate	Kapa Hyper Prep kit with custom adapters (IDT and Broad Insitute)	HiSeq2500/ 4000	20	208	0.36	ULP-WGS	BRCA-like signature and APOBEC signature in tumor tissue and matched cfDNA were strongly correlated; Number of predicted neoantigens in tumor tissue and matched cfDNA were strongly correlated
Ahlborn[16]	2019	no	44	Pan-cancer	NEBNext Ultra II Library Prep Kit (New England Biolabs) and MedExome (Roche)/ SureSelect (Agilent)	NextSeq/ HiSeq	10	210	0.9	Maximum MAF of identified SCAA per patient	
Ahlborn[17]	2018	Yes, progression samples with known high MAF (>10%) for BRAF by dPCR	4	Colorectal- and bile duct	NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) and SeqCap EZ MEDExome (Roche NimbleGen)	Illumina	10	>50	0.17,0.12	Maximum MAF of identified SCAA per patient	
Butler[18]	2015	No	2	Sarcoma; Breast	SureSelectXT Human All Exon V4+UTRs (Agilent)	HiSeq 2000	100	416,5	0.227	Maximum MAF of identified SCAA per patient	
Chicard[19]	2018	no	36	Neuroblastoma	Kapa Library Preparation Kit (Kapa Biosystems) and SeqCap EZ Human Exome Kit v3.0 (Roche NimbleGen)	HiSeq 2500	100	NR	0.64	Sequenza[1]	CNAs were concordant between plasma and tumor tissue in 93% of all cases with an additional amplification of 15q harbording *IGF1R* in cfDNA

Table 1 - Continued

First author	Year of publication	Pre-selection for WES analysis?	Sample size for WES (N)	Tumor type	Library Prep	Sequencing method	Median input cfDNA (ng)	Median coverage (X)	ctDNA fraction	Method for estimation of ctDNA fraction	Genetic alterations other than mutations identified in cfDNA
Dietz[20]	2016	no	6	NSCLC	ThruPLEX-FD Prep Kit (Rubicon Genomics)	HiSeq2000	10	68,5	NR	No information on ctDNA fraction available	
Goodall[21]	2017	no	1	Prostate	Kapa Hyper Plus Library Prep (Kapa Biosystems) SureSelectXT V6 (Agilent)	NextSeq500	40	NR	0.21	Maximum MAF identified by WES per patient	
Gremel[22]	2016	no	1	Musocal melanoma	Accel-NGS 2S DNA Library Kits (Swift Biosciences)	NextSeq	10-25	NR	0.36	Maximum MAF identified by WES extracted from graph	
Huang[5]	2017	no	5	HCC	Agilent and in-house 50Mb system V4 (BGI Xome)	HiSeq4000	50-100	226,2	NR	No information on ctDNA fraction available	Mutational signatures with the largest contribution in tissue were found in plasma, some additional signatures in plasma were identified
Jiménez[23]	2019	no	18	Renal tumors	Kapa Library Preparation Kit (Kapa Biosystems) and SeqCap EZ Human Exome Kit (Roche NimbleGen)	HiSeq2500	60,9	NR	0.295	Sequenza[1]	CNAs were concordant between plasma and tumor tissue in 72% of all cases, profiles were closely related
Klevebring[24]	2014	no	9	Breast; Prostate	ThruPLEX-FD (Rubicon Genomics) and SeqCap EZ Exome Library Version 1 (Roche NimbleGen)	HiSeq2500	5	43	0	Maximum MAF identified by WES per patient	

Table 1 - Continued

First author	Year of publication	Pre-selection for WES analysis?	Sample size for WES (N)	Tumor type	Library Prep	Sequencing method	Median input cfDNA (ng)	Median coverage (X)	ctDNA fraction	Method for estimation of ctDNA fraction	Genetic alterations other than mutations identified in cfDNA
Koeppel[25]	2017	Samples with at least 6ng cfDNA remaining for analysis	32	Pan-cancer	NebtNext Ultra II DNA Library Prep (New England Biolabs) and SureSelect XT (Agilent)	HiSeq4000	25	140	0.19	Maximum MAF identified by WES per patient	TML in plasma and corresponding tissue was correlated for ctDNA positive patients with significant ctDNA fraction
Murtaza[26]	2013	Longitidinal samples with ' high' allele fraction	19	Breast- and ovarian	ThruPLEX-FD (Rubicon Genomics) and TruSeq Exome Enrichment Kit (Illumina)	HiSeq2500	10	110	0.52	Genomewide aggregated allelic loss (GAAL)[2]	CNAs were concordant between plasma and tumor tissue and more prominent in samples with ctDNA fraction >50%
Murtaza[27]	2015	Yes only sample T1, T2 and T9, reason not further specified	3	Breast	ThruPLEX-FD (Rubicon Genomics) and TruSeq Exome Enrichment Kit (Illumina)	HiSeq2500	2-18	79,6	0.215	Mean MAF of stem mutations[3]	
Olmedillas-Lopez[28]	2018	no	30	Colorectal	SeqCap EZ HGSC VCRome Kit (Roche NimbleGen)	NextSeq00	NR	40-80	NR	No information on ctDNA fraction available	
Song[29]	2018	no	1	Thyroid	SureSelect Human All ExonV5 (Agilent)	HiSeq2000	NR	133	NR	No information on ctDNA fraction available	
Sun[30]	2019	no	6	Glioma	Nebnext Ultra II DNA Library Prep Kit (New England Biolabs) and NEBN multiplex Oligos (New England Biolabs) kit	HISEQ Xten	61	500*	NR	No information on ctDNA fraction available	

Table 1 - Continued

First author	Year of publication	Pre-selection for WES analysis?	Sample size for WES (N)	Tumor type	Library Prep	Sequencing method	Median input cfDNA (ng)	Median coverage (X)	ctDNA fraction	Method for estimation of ctDNA fraction	Genetic alterations other than mutations identified in cfDNA
Tailor[31]	2019	no	33	Lung cancer; benign lung lesion	SureSelect All Exon V5+UTR (Agilent)	HiSeg2500	>10	49	NR	No information on ctDNA fraction available	
Toledo[32]	2018	no	1	Colorectal	ThruPLEX PlasmaSeq (Rubicon Genomics) SureSelectXT Target Enrichment System (Agilent)	HiSeq4000	15	183	0.47	Maximum MAF identified by WES per patient	
Vandekerkhove[33]	2017	Samples with ≥25% ctDNA fraction	11	Bladder cancer	SeqCap EZ MedExome kit (Roche NimbleGen)	MiSeqV3 600 or HiSeq2500	10-25	176	0.40	MAF of autosomal somatic mutations corrected for mutations with LOH[4], mutations with detectable amplification were excluded	Novel fusion genes (FGFR-ADD1) and rearrangements were detected and validated

Abbreviations: ULP-WGS: Ultra low-pass whole genome sequencing; NR: not reported; SCAA: selected cancer-associated alteration; MAF: Mutant Allele fraction; LOH: Loss of heterozygosity

[1]Sequenza: estimates ploidy in samples with a tumor content >30%

[2]Analysis of the allelic counts for SNPs exhibiting loss of heterozygosity in the tumor by the following equation: C = (Nnondel - Ndel) / Nnondel

C = plasma concentration of tumor-derived DNA, Nnondel = number of sequenced reads carrying the nondeleted alleles in the tumor tissues, Ndel = number of sequenced reads carrying the deleted alleles in the tumor tissues

[3]Stem mutations: common to all tumor biopsies

[4]ctDNA fraction = 2/((1/mutant allele fraction)+1)

*the original article mentions 'an average sequence depth of >500 bp and genome coverage >98%' which we interpreted as 500x depth

To focus on tumor-specific SNVs, all studies except for Dietz et al[20] sequenced germline DNA derived from leukocytes or normal tissue. In addition, most studies used a combination of databases such as dbSNP[35], 1000 Genomes Project[36] or Exome Sequencing Project[37] to filter out single nucleotide polymorphisms (SNPs) present in germline DNA. Final selection of variants based on MAF, coverage and sequencing quality was highly variable (**Table A4**). For example, some studies only called SNVs based on a minimum MAF ranging from 1[31,33] to 5%[16,17]. Sequencing quality scores involved either in-house developed algorithms or Phred scores with different cut-offs. Finally, not all studies performed an exome-wide final analysis and only reported data on cancer-associated genes[16] or genes involved in MAPK-pathway analysis[17] limiting the number of detected SNVs per patient. We further studied the correlation between total numbers of detected SNVs in cfDNA and ctDNA fraction on IPD which showed a positive correlation, p=0.016 (**Figure A1**). Since only the minority of studies provided IPD on cfDNA input or coverage, these variables were not individually tested.

In addition to SNV detection, WES of cfDNA also allowed analysis of copy number variation[19,23,26], mutational signatures[15], fusion genes[33], rearrangements[33], predicted neoantigens[15] and TMB[25]. Mutational signature fractions (i.e. APOBEC and BRCA-like)[15] and TMB[25] were identified and positively correlated between cfDNA and tumor tissue. The Spearman's correlation coefficient was 0.92 (no CI given) for mutational signature fractions and 0.85 (no CI given) for TMB. Sensitivity and agreement of the genomic signatures in cfDNA versus tumor tissue were not reported.

Pooled sensitivity and agreement rate of cell-free DNA versus tumor tissue

Out of the 303 cfDNA samples on which WES was performed, WES data of matched tumor tissue was available for only 71 unique sample pairs. To calculate a pooled sensitivity and agreement between WES-detected SNVs in cfDNA versus tumor tissue we performed a Bayesian random-effect meta-analysis on this subset (**Table 2**; **Table A5** for IPD). Most studies compared SNVs between metastatic tumor tissue and cfDNA[15-17,21,26,27,32], whilst three studies analyzed primary tumors[20,29,38] and two studies analyzed both[19,23]. The pooled sensitivity of WES-detected SNVs in cfDNA using tumor tissue as reference was 50% (95% CI: 29-72%) (**Figure 2**). The tissue MAF of variants exclusively identified in tissue was significantly lower (12.5%, range: 0.5-18%) than the tissue MAF of shared variants (23.9%, range: 17-38%), p=0.004. For cfDNA, the median MAF of variants detected in both tumor tissue and cfDNA (12.2%, range: 2.1-26.9%) was higher than the MAF of variants detected in cfDNA only (4.6%, range: 0.4-9.0%), although not statistically significant p=0.093. The pooled agreement was 31% (95% CI: 15-49%) (**Figure 3**). Since we found a positive correlation between total numbers of

SNVs and ctDNA fraction, we conducted a sub-analysis of 40 unique sample pairs with a ctDNA fraction ≥25%, which resulted in a sensitivity of 69% (95% CI: 46-89%) and agreement of 46% (95% CI: 36-59%).

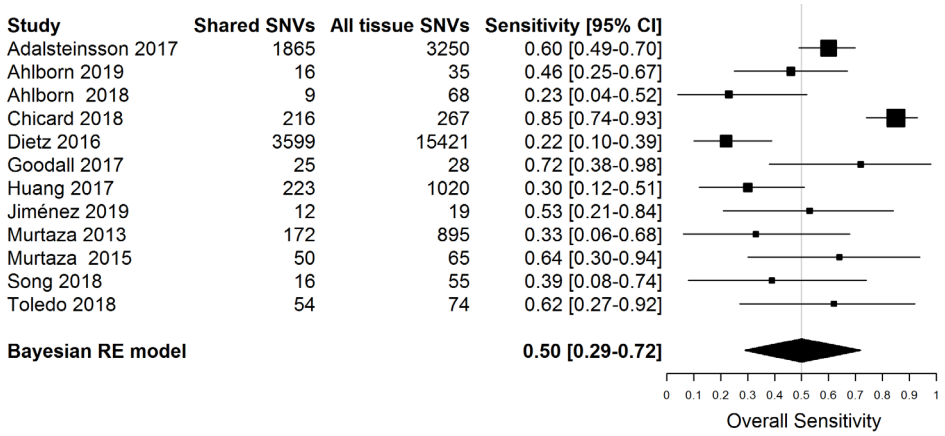


Figure 2 - Pooled sensitivity analysis of WES-detected SNVs in cfDNA versus tumor tissue.
The size of the black boxes that represent the point estimates indicate the precision of the point estimate based on sample size and heterogeneity of individual data.

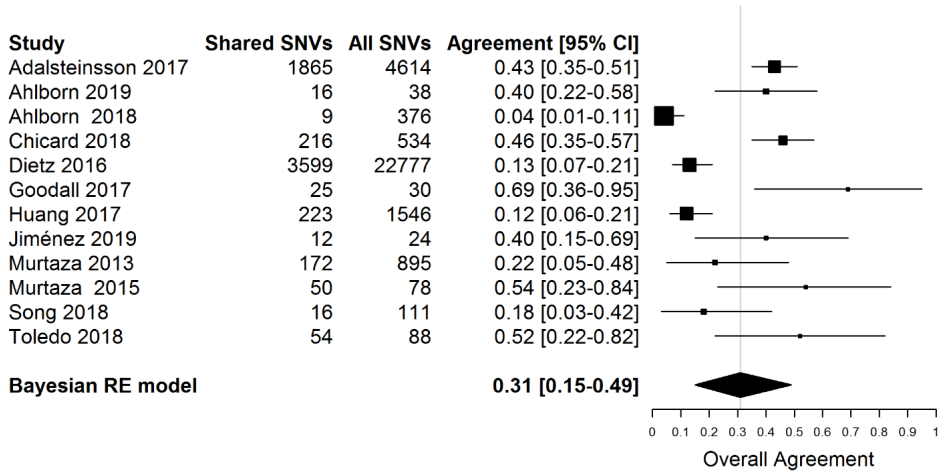


Figure 3 - Pooled agreement analysis of WES-detected SNVS in cfDNA versus tumor tissue.
The size of the black boxes that represent the point estimates indicate the precision of the point estimate based on sample size and heterogeneity of individual data.

Table 2 - Overview of studies investigating WES-detected SNVs in cfDNA and tissue

First author	Year of publication	Sample size for concordance analysis	Tumor type	Location of tissue biopsy	Input (ng)[1]		Coverage (X)[1]		Tumor purity (%)[1]	Analysis performed exome wide?
					cfDNA	Tissue DNA	cfDNA	Tissue DNA		
Adalsteinsson[15]	2017	24	Breast- and prostate	metastatic lesions	20	NR	208	173	58	yes
Ahlborn[16]	2019	13	Pan-cancer	metastatic lesions	10	10	210	>100	NR	yes but only selected cancer associated alterations were reported
Ahlborn[17]	2018	2	Bile duct	metastatic lesions	10	10	>50	NR	52	no, focused on MAPK-pathway genes
Chicard[19]	2018	14	Neuroblastoma	primary tumor and metastatic lesions	100	100	NR	NR	88	yes
Dietz[20]	2016	6	NSCLC	primary tumor	10	50	68,5	61	NR	no, 58 known most frequently mutated genes
Goodall[21]	2017	1	Prostate	metastatic lesions	40	NR	NR	NR	30	yes
Huang[5]	2017	5	HCC	primary tumor, multiple regions	50	500-2000	226	211	70	yes
Jiménez[23]	2019	2	Renal tumors	primary tumor and metastatic lesions	61	NR	NR	NR	35	yes
Murtaza[26]	2013	1	Ovarian	metastatic lesions	17.5	NR	110	NR	NR	yes
Murtaza[27]	2015	1	Breast	metastatic lesion	10	NR	77,3	16,6	70	yes
Toledo[32]	2018	1	Colorectal	metastatic lesion	15	70-110	183	88,5	NR	yes
Song[29]	2018	1	Follicular thyroid carcinoma	primary tumor	NR	NR	133	140	NR	yes

Abbreviations: NR: not reported; NSCLC: non-small cell lung cancer; HCC: hepatocellular carcinoma
[1]Values are calculated medians unless ranges are reported
Shared SNVs: Single nucleotide variants in both plasma and tissue
Tissue SNVs: All single nucleotide variants in tissue (independent of plasma)
All SNVs: All single nucleotide variants in plasma and/or tissue

Table 3 - Additional WES-detected SNVs in cfDNA only

First author	Year of publication	Tumor type	Ratio: ctDNA variants unique to plasma / all ctDNA variants	Median number of additional SNVs in cfDNA per patient	Mutated genes additionally identified in cfDNA	
					OncoKB actionable genes (level of evidence for this tumor type)	Genes with a somatic mutation prevalention ≥2% (cBioPortal)
Adalsteinsson[15]	2017	breast	0,39	27	ESR1(3A); PIK3CA(1); mTOR(4); NF1(4); ERBB2(3A)	MUC16; PIK3CA; SYNE1; AHNAK2; PTPRD; ESR1; TTN; HERC2; BIRC6; RYR2; LAMA2; TG; DNAH5; TM2TD; ATR; ERBB2; USH2A; AHNAK
		prostate	0,44	17	NF1(4)	KMT2D
Ahlborn[16]	2019	pan cancer	0,38	0		
Ahlborn[17]	2018	bile duct	0,96	154	mTOR(4); FGFR3(4)	ARID1A; POL;E DOT1L; ANKRD3C
Chicard[19]	2018	neuroblastoma	0,49	16		
Dietz[20]	2016	NSCLC	0,69	999	mTOR(4)	mTOR
Goodall[21]	2017	prostate	0,07	2		
Huang[5]	2017	HCC	0,73	135	PTEN (4)*	
Jiménez[23]	2019	clear cell sarcoma	0,42	3		
Murtaza[26]	2013	breast	0	0	NR	NR
Murtaza[27]	2015	breast	0,21	13		
Song[29]	2018	Thyroid	0,21	56		
Toledo[32]	2018	colorectal	0,21	14		MDC1; PABPC3; RTN4; CHD7

Abbreviations: NR: not reported; NSCLC: non-small cell lung cancer; HCC: hepatocellular carcinoma

Additional value of WES of cfDNA

To evaluate the additional value of WES of cfDNA to identify clinically useful SNVs that were not present in tissue, we calculated the number of additional SNVs unique to cfDNA and the ratio between ctDNA variants unique to plasma versus all variants detected in ctDNA (**Table 3**). Of all plasma-detected SNVs per sample a median of 43% (range: 0-96%) was exclusively detected in plasma. The median number of additionally detected SNVs per sample was 17 (range: 0-2,840). Of these additionally detected SNVs, 36 variants detected in 20 out of 53 patients (38%) were detected in cancer associated genes as reported in cBioportal. Matching IPD with targetable genes according to OncoKB, we identified in 11 variants in 9 out of 53 patients (17%). The targetability of these variants ranged from level 1 (FDA-approved) to level 4 (biological evidence) (**Table 3**).

Discussion

The increasing interest to capture the complex genomic landscape of individual cancer patients real time and in a minimally-invasive way, has initiated efforts on technical developments in the field of WES of cfDNA. The main purpose of this systematic review was to evaluate the technical feasibility of WES of cfDNA and to analyze the sensitivity and agreement of WES-detected SNVs in cfDNA using tumor tissue as reference.

It has become clear that there was significant variability between studies in the pre- and post-analytical conditions used (**Table 1; Table A4**) which severely impacted comparability of results. Differences between studies were observed regarding technical aspects of sequencing including sequencing coverage and amount of cfDNA used. Especially sequencing coverage was highly variable, ranging from 43-500X coverage, which theoretically results in lower limits of variant detection of 2.3% and 0.2% respectively, assuming that the variant is heterozygous and the genome is diploid. Although cfDNA input generally consisted of 10-20 ng, inputs ranged from 2 ng to 100 ng. In addition, the post-analytical part in which different bioinformatics pipelines were used also impacted the final variant calling since most studies used their own set of criteria to filter SNPs and to perform final variant calling (**Table A4**).

Clonal hematopoiesis has been identified as an important factor affecting accurate variant interpretation. During the process of aging different mutations accumulate in hematopoietic stem cells. This phenomenon occurs frequently in the elderly and its prevalence has been estimated at 31%[39]. The mutations resulting from clonal

hematopoiesis are often detected during cfDNA sequencing analysis, since the majority of cfDNA is derived from leukocytes. Recently, it has been demonstrated that 53.2% of all mutations detected by cfDNA sequencing analysis result from clonal hematopoiesis, indicating the need for collection and sequencing of leukocytes as a reference[39].

Taken into account that for only 71 out of 303 cfDNA samples WES data of matching tumor tissue was available, the merit of our comparison is that we performed the meta-analysis on IPD level which allowed us to adjust for patient-specific effects in our model. Compared to large studies describing agreement between cfDNA and tumor tissue based on targeted sequencing approaches covering pre-specified gene sets[40], the number of sample pairs which was analyzed by WES is considerably limited. Most studies lacked IPD on cfDNA input and sequencing coverage hampering analysis of the impact of those variables on sensitivity and agreement.

Some studies only performed WES on samples with a minimum tumor fraction[15,17,26,27,33]. Overall, the samples selected for WES consisted of a median tumor fraction of 37%, much higher than generally occurs in cancer patients[41]. By using techniques as ultralow-pass whole-genome sequencing with 10% tumor fraction as a cutoff value to pre-select samples, Adalsteinsson *et al*[15] showed that only 34% of cfDNA samples from metastatic breast- and prostate cancer patients were feasible for WES analysis, including samples from all treatment lines. This implies that the number of samples with a sufficient tumor fraction in earlier lines of treatment might be even lower. Notably, studies have not reported success rates of WES in correlation to tumor fraction.

Our meta-analysis of the sensitivity of WES-detected SNVs in cfDNA versus tumor tissue has shown that 50% of SNVs present in tumor tissue are also detected in cfDNA. The reason for the rather low sensitivity of WES of cfDNA is probably multifactorial including technical and biological aspects. A major technical issue is the generally low sequencing coverage used for WES resulting in false-negative results for cfDNA variants present below the limit of detection. This is supported by the comparison of WES (coverage 226X) with targeted deep sequencing (coverage 1,806X) on the same sample[5], demonstrating that some variants with low MAFs (<5%) were detected by targeted deep sequencing only[5]. However, with the introduction of unique molecular identifiers and Elimination of Recurrent Artefacts and Stochastic Errors (ERASE-Seq)[42] discrimination of sequencing artefacts from true variants can be improved, enabling detection DNA variants at ultralow frequency.

Another technical aspect possibly affecting WES sensitivity is the tube type used for

blood collection. Previous studies have demonstrated that cfDNA isolated from serum instead of plasma increased the background of normal DNA by the release of germline DNA (gDNA) due to lysis of leukocytes during coagulation[43,44]. This might partly explain the low sensitivity of Dietz *et al*[20]. Size selection of short fragments before sequencing might positively influence the ratio between gDNA and cfDNA as well[45]. Biological challenges are the amount of cfDNA available for sequencing, generally low overall tumor fraction present in the sample and the subclonal presence of clones bearing alterations associated with treatment resistance[19]. The importance of the ctDNA fraction for the sensitivity of WES is supported by our finding that sensitivity improves when analyzing samples with a ctDNA fraction ≥25%.

Our findings demonstrate that in addition to tumor-tissue detected SNVs, WES of cfDNA also discovers SNVs exclusively detected in cfDNA. When calculating the ratio of these SNVs versus all cfDNA variants, we observed that the fraction of variants unique to cfDNA was highly variable amongst studies. The variability might partly be explained by factors as sequencing coverage and cfDNA input. This assumption, however, could not be substantiated by our data as availability of sequencing coverage and cfDNA input on individual patient level was insufficient. Nevertheless, SNVs which are exclusively detected in cfDNA potentially reflect intra- and inter tumor heterogeneity. Whether these additionally detected SNVs in plasma are derived from clonal or subclonal fractions in tumor tissue remains a topic of interest. Adalsteinsson *et al*[15] estimated clonality and subclonality of SNVs detected in plasma and demonstrated that on average 88% of the clonal and 45% of the subclonal mutations were confirmed in the tumor. Assuming that all plasma detected SNVs are true variants, i.e. free of sequencing artefacts, these results imply that the majority of SNVs exclusively detected in cfDNA were subclonal in this study. Furthermore, SNVs indicated as subclonal in cfDNA might be of clonal origin in tumor tissue from other metastatic sites.

Studies comparing cfDNA to multiple tumor region sampling support this hypothesis[5,25]. Huang *et al.* found that nearly all exclusively detected SNVs in plasma by WES were also detected in tumor samples from different liver lesions using targeted deep sequencing (average 98.7%, range: 69.3-100%)[5]. Another study also showed that when two tissue biopsies were taken and compared to cfDNA, the number of overlapping alterations between cfDNA and tumor tissue increased[25]. Importantly, some of these exclusively identified mutations were previously associated with therapy resistance (*ESR1*, *ERBB2* and *NF1*)[46] and treatment outcome (*PIK3CA*)[47] in breast cancer. The clinical relevance of these findings and to what extent the MAF and its dynamics will impact outcome on certain therapies have thus far not been elucidated in prospective clinical studies. Secondly, the

number of additional identified targetable mutations in cfDNA is currently very limited, but might be improved by efforts unraveling new actionable targets or profiles.

The added value of WES currently thus mainly resides in the discovery of resistance mechanisms and genomic alterations for which a wide coverage of the genome is needed such as TMB and mutational signatures. Goodall *et al.*[21] demonstrated this discovery-capacity of WES by identifying frameshifts in germline and somatic DNA repair mutations as mechanisms of resistance to PARP inhibitors in prostate cancer. For estimation of TMB, large targeted sequencing panels can be used[48]. However when taking estimation of TMB by whole genome sequencing (WGS) as reference, 30% of patients were misclassified – either false negative or false positive – when targeted sequencing panels were used. Concordance improved by increasing number of megabases (Mb) covered by the targeted panel[49]. Furthermore, reported correlations between cfDNA and tissue TMB using targeted panels, Spearman's correlation coefficient of 0.64 and 0.6[50] are lower than correlations reported using WES, 0.85[25]. To this end, further design and validation of highly needed targeted sequencing panels for TMB estimation are currently ongoing.

Altogether WES is an attractive tool for identification of genomic signatures and discovery of resistance mechanisms. In this IPD meta-analysis we show that the sensitivity of WES of cfDNA is 50% and that the overall agreement is 31%. Furthermore we describe large variability in pre- and post-analytical conditions of WES of cfDNA. Moreover, our results underline that the applicability of WES mainly resides in a selected group of patients with high tumor fractions. We recognize that WES is still in its developmental phase and that implementation of methods such as unique molecular barcoding and ERASE-seq will further improve sensitivity of WES. However, standardization of methodologies is highly needed to further define the clinical utility of this promising approach.

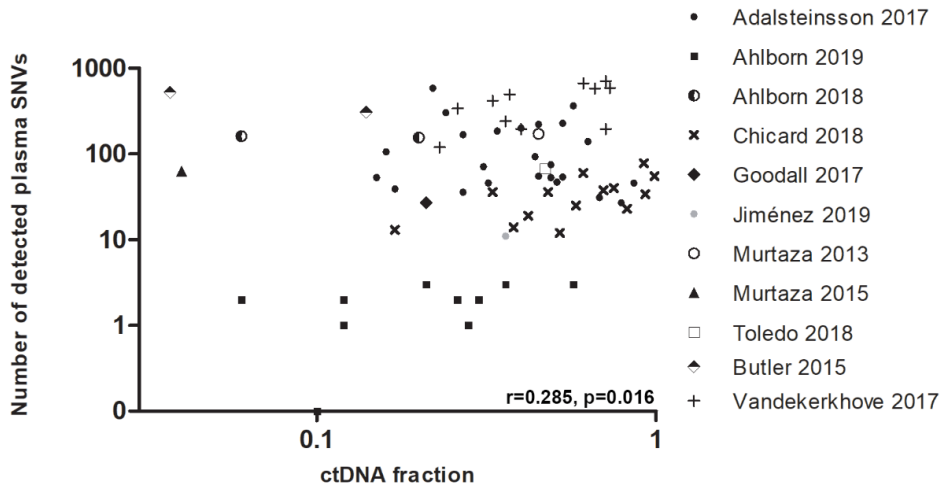
Acknowledgements
The authors would like to express their sincere gratitude to Marcel Smid for critically reading the manuscript.

References

1. Flaherty, K.T. *et al.* Improved survival with MEK inhibition in BRAF-mutated melanoma. *N Engl J Med* **367**, 107-14 (2012).
2. McArthur, G.A. *et al.* Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. *Lancet Oncol* **15**, 323-32 (2014).
3. Sorich, M.J. *et al.* Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials. *Ann Oncol* **26**, 13-21 (2015).
4. Frenel, J.S. *et al.* Serial Next-Generation Sequencing of Circulating Cell-Free DNA Evaluating Tumor Clone Response To Molecularly Targeted Drug Administration. *Clin Cancer Res* **21**, 4586-96 (2015).
5. Huang, A. *et al.* Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma. *J Hepatol* **67**, 293-301 (2017).
6. Merker, J.D. *et al.* Circulating Tumor DNA Analysis in Patients With Cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. *J Clin Oncol* **36**, 1631-1641 (2018).
7. De Mattos-Arruda, L. *et al.* Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol* **25**, 1729-35 (2014).
8. cobas EGFR Mutation Test v2. 2016.
9. Le, D.T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409-413 (2017).
10. Rizvi, N.A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-8 (2015).
11. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* **2017**(2017).
12. cbioportal. Vol. 2019.
13. R Core Team. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria 2019).
14. Plummer M. JAGS: Bayesian Graphical Models using MCMC. R package version 4-6. 2016.
15. Adalsteinsson, V.A. *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**, 1324 (2017).
16. Ahlborn, L.B. *et al.* Application of cell-free DNA for genomic tumor profiling: a feasibility study. *Oncotarget* **10**, 1388-1398 (2019).
17. Ahlborn, L.B. *et al.* Circulating tumor DNA as a marker of treatment response in BRAF V600E mutated non-melanoma solid tumors. *Oncotarget* **9**, 32570-32579 (2018).
18. Butler, T.M. *et al.* Exome Sequencing of Cell-Free DNA from Metastatic Cancer Patients Identifies Clinically Actionable Mutations Distinct from Primary Disease. *PLoS One* **10**, e0136407 (2015).
19. Chicard, M. *et al.* Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-spatial Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma. *Clin Cancer Res* **24**, 939-949 (2018).
20. Dietz, S. *et al.* Low Input Whole-Exome Sequencing to Determine the Representation of the Tumor Exome in Circulating DNA of Non-Small Cell Lung Cancer Patients. *PLoS One* **11**, e0161012 (2016).
21. Goodall, J. *et al.* Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition. *Cancer Discov* **7**, 1006-1017 (2017).
22. Gremel, G. *et al.* Distinct subclonal tumour responses to therapy revealed by circulating cell-free DNA. *Ann Oncol* **27**, 1959-65 (2016).
23. Jimenez, I. *et al.* Circulating tumor DNA analysis enables molecular characterization of pediatric renal tumors at diagnosis. *Int J Cancer* **144**, 68-79 (2019).
24. Klevebring, D. *et al.* Evaluation of exome sequencing to estimate tumor burden in plasma. *PLoS One* **9**, e104417 (2014).
25. Koeppel, F. *et al.* Whole exome sequencing for determination of tumor mutation load in liquid biopsy from

advanced cancer patients. *PLoS One* **12**, e0188174 (2017).
26. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108-12 (2013).
27. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* **6**, 8760 (2015).
28. Olmedillas-Lopez, S. *et al.* Liquid biopsy by NGS: differential presence of exons (DPE) in cell-free DNA reveals different patterns in metastatic and nonmetastatic colorectal cancer. *Cancer Med* **7**, 1706-1716 (2018).
29. Song, J. & Yang, Z. Case report: Whole exome sequencing of circulating cell-free tumor DNA in a follicular thyroid carcinoma patient with lung and bone metastases. *J Circ Biomark* **7**, 1849454418763725 (2018).
30. Sun, J. *et al.* Examination of Plasma Cell-Free DNA of Glioma Patients by Whole Exome Sequencing. *World Neurosurg* (2019).
31. Tailor, T.D. *et al.* Whole Exome Sequencing of Cell-Free DNA for Early Lung Cancer: A Pilot Study to Differentiate Benign From Malignant CT-Detected Pulmonary Lesions. *Front Oncol* **9**, 317 (2019).
32. Toledo, R.A. *et al.* Exome Sequencing of Plasma DNA Portrays the Mutation Landscape of Colorectal Cancer and Discovers Mutated VEGFR2 Receptors as Modulators of Antiangiogenic Therapies. *Clin Cancer Res* **24**, 3550-3559 (2018).
33. Vandekerkhove, G. *et al.* Circulating Tumor DNA Reveals Clinically Actionable Somatic Genome of Metastatic Bladder Cancer. *Clin Cancer Res* **23**, 6487-6497 (2017).
34. Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64-70 (2015).
35. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-11 (2001).
36. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
37. Exome Variant Server. (NHLBI Exome Sequencing Project (ESP), Seattle, WA).
38. Huang, M.N. *et al.* MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. *Sci Rep* **5**, 13321 (2015).
39. Razavi, P. *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med* (2019).
40. Jovelet, C. *et al.* Circulating Cell-Free Tumor DNA Analysis of 50 Genes by Next-Generation Sequencing in the Prospective MOSCATO Trial. *Clin Cancer Res* **22**, 2960-8 (2016).
41. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-90 (2008).
42. Kamps-Hughes, N. *et al.* ERASE-Seq: Leveraging replicate measurements to enhance ultralow frequency variant detection in NGS data. *PLoS One* **13**, e0195272 (2018).
43. Parpart-Li, S. *et al.* The Effect of Preservative and Temperature on the Analysis of Circulating Tumor DNA. *Clin Cancer Res* **23**, 2471-2477 (2017).
44. van Dessel, L.F. *et al.* Application of circulating tumor DNA in prospective clinical oncology trials - standardization of preanalytical conditions. *Mol Oncol* **11**, 295-304 (2017).
45. Mouliere, F. *et al.* Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* **10**(2018).
46. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438 e6 (2018).
47. Andre, F. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* **380**, 1929-1940 (2019).
48. Cabel, L. *et al.* Clinical potential of circulating tumour DNA in patients receiving anticancer immunotherapy. *Nat Rev Clin Oncol* **15**, 639-650 (2018).
49. Mankor, J.M. *et al.* Impact of panel design and cut-off on tumor mutational burden (TMB) assessment in metastatic solid tumor samples. *Submitted* (2019).
50. Snyder, A., Morrissey, M.P. & Hellmann, M.D. Use of Circulating Tumor DNA for Cancer Immunotherapy. *Clin Cancer Res* (2019).

Supplemental Figure



Supplementary Figure A.1 - Correlation between ctDNA fraction and the total number WES-detected SNVs in cfDNA. Individual patient data are shown.

Supplemental Tables

Supplementary Table A.1 - Overview of the literature search

Search	Search terms	Number of hits
1	Cell-free DNA [Tiab] OR plasma DNA [Tiab] OR cell-free nucleic acids [MeSH]	4584
2	Exome sequencing [Tiab] OR Whole-exome sequencing [MeSH]	13060
3	1 AND 2	56

Supplementary Table A.2 - Data-extraction form

Genera information				cfDNA - Tissue concordance
First author name				SNVs
Publication year				Tissue + -
Journal				Plasma + 0
				Plasma - 0
Sample size WES				0 0 0
Sample size for concordance				**Sensitivity**
Tumor type				**Agreement**
Tissue - cfDNA collected at same time?	If no, specify			
Patients included in concordance analysis				
Technical information				
cfDNA				
Sequencing method	(Library Prep)			
	(Sequencing)			
Barcoding?				
Variant annotation				
ctDNA samples used for concordance analysis		All ctDNA samples		
Input	ng	Input	ng	
Median coverage	X	Median coverage	X	
VAF	(Median)	VAF	(Median)	Genes included in analysis
	(Min)		(Min)	
	(Max)		(Max)	
Estimated ctDNA%	%	Estimated ctDNA%	%	
Method:				
Tissue				
Input	ng			
Coverage	X			
Sequencing method different than ctDNA?				
Tumor purity	% (median)			
Median VAF concordant alt	%			
Median VAF tDNA only alt	%			
(Pre-)selection				
Were samples pre-selected for WES?				
Was analysis performed exome wide?	If no, specifiy which genes were included			
Variant annotation				
Germline				
Allele frequency				
Quality score				
Read coverage				
Variant caller				
SNP filtering				

Supplementary Table A.3 - Overview of data extraction source

Study	Sample size for concordance	Patients included	Timepoints	Data	Source
Adalsteinsson 2017	24	CRPC_554; MBC_191; MBC_248; MBC_287; MBC_288; MBC_292; MBC_292; MBC_295; MBC_301; MBC_303; MBC_307; MBC_313; MBC_315; MBC_317; MBC_318; MBC_320; MBC_320; MBC_321; MBC_325; MBC_330; MBC_331; MBC_335; MBC_336; MBC_339; MBC_349; MBC_8	Baseline	SNVs	Supplementary Data 5, summarized in separate file
				Coverage cfDNA	Supplementary Data 2
				Time point collection	Supplementary Data 7
				tVAF dis-/concordant SNVs	Supplementary Data 5, summarized in separate file
				ctDNA%	Supplementary Data 6
				Tumor purity tissue	Supplementary Data 5
Ahlborn 2019	14	Only re-biopsies were included. P1; P2; P4; P6; P8; P9; P37; P38; P39; P40; P41; P42; P43; P44	Re-biopsy and archival FFPE tissue	SNVs	Supplemental table 5
				Coverage cfDNA	Supplemental table 2
				Time point collection	Supplemental table 1
				tVAF dis-/concordant SNVs	x
				ctDNA%	Supplemental table 3
				Tumor purity tissue	x
Ahlborn 2018	2	Pt5 One patient, two timepoints	Baseline and progression	SNVs	Supplementary file 2
				Coverage cfDNA	Paper text, methods section
				Time point collection	Supplementary file 2
				tVAF dis-/concordant SNVs	Supplementary file 2
				ctDNA%	Paper Figure 4B
				Tumor purity tissue	Paper Figure 4B
Dietz 2016	6	All		SNVs	Paper Table 2
				Coverage cfDNA	Paper Table 2
				Time point collection	x
				tVAF dis-/concordant SNVs	x
				ctDNA%	x
				Tumor purity tissue	x
Goodall 2017	1	WES performed and data available for one patient who respond to olaparib	Baseline	SNVs	Supplementary table 1
				Coverage cfDNA	x
				Time point collection	Supplementary table 1
				tVAF dis-/concordant SNVs	Supplementary table 1
				ctDNA%	Supplementary table 1
				Tumor purity tissue	x

Supplementary Table A.3 - Continued

Study	Sample size for concordance	Patients included	Timepoints	Data	Source
Jiménez 2019	2	Patient 1 and 2	At local and or metastatic relapse	SNVs	Supplementary table 3 and Supplementary figure 2
				Coverage cfDNA	x
				Time point collection	Paper, Table 1
				tVAF dis-/concordant SNVs	Supplementary table 3
				ctDNA%	Paper, Table 1
				Tumor purity tissue	Paper, Table 1
Koeppel 2017	32	All		SNVs	Paper text, page 9. VAFs: Supplementary table 2
				Coverage cfDNA	Paper text, page 6
				Time point collection	Paper text, page 3
				tVAF dis-/concordant SNVs	x
				ctDNA%	Supplementary table 2
				Tumor purity tissue	Paper text, page 3
Murtaza 2013	2	Case 1 and Case 4	Case 1 sample E1; Case 4 sample E2	SNVs	Supplementary information; paper text page 100
				Coverage cfDNA	Supplementary table S2
				Time point collection	Paper text, page 110
				tVAF dis-/concordant SNVs	x
				ctDNA%	Supplementary table S3
				Tumor purity tissue	x
Murtaza 2015	1	All	Plasma sample T1 and Tissue sample M2.1	SNVs	Supplementary data 1
				Coverage cfDNA	Supplementary table 3
				Time point collection	Paper, figure 1b
				tVAF dis-/concordant SNVs	Supplementary data 1
				ctDNA%	Supplementary data 1
				Tumor purity tissue	Supplementary table 1
Toledo 2018	1			SNVs	Supplementary table S1
				Coverage cfDNA	Supplementary table S1
				Time point collection	Paper figure 1A
				tVAF dis-/concordant SNVs	Supplementary table S1
				ctDNA%	Supplementary table S1
				Tumor purity tissue	x

Supplementary Table A.3 - Continued

Study	Sample size for concordance	Patients included	Timepoints	Data	Source
Chicard 2018	14	Case 1, Case 2, Case 3, Case 4, Case 5, Case 6, Case 7, Case 8, Case 9, Case 10, Case 15, Case 16, Case 17, Case 18, Case 19	Baseline	SNVs	Table S3
				Coverage cfDNA	Supplementary table S1
				Time point collection	Supplementary table S2
				tVAF dis-/concordant SNVs	x
				ctDNA%	Supplementary table S1
				Tumor purity tissue	Supplementary table S1

Supplementary Table A.4 - Variant calling and annotation of WES data

Author	Year	Germline	Quality score		Reference genome	Variant caller	SNP filtering	Variant selection	
			Sequencing quality	Mapping quality				Inclusion	Exclusion
Adalsteinsson	2017	yes			Hg19	MuTect	Removed sites where ≥20% of samples accross the panel of normals had reads supporting the mutation.	For all: 50% of all samples covering ≥8 reads and at most 30% of samples covering <8 reads. Additional for agreement analysis with tissue: ≥3 reads, or <3 reads but >0.9 power*	Artifactual OxoG mutations (OxoG3 filter)
Ahlborn	2019	yes			Hg19	Ingenuity Variant Analysis Software (Qiagen)	>1% in the ExAC database, 1000 Genomes Project or ESP	≥10X coverage; F:R balance ≥0.1; Tumor DNA MAF ≥5%;	
Ahlborn	2018	yes	Phred quality score ≥30		Hg19	Ingenuity Variant Analysis Software (Qiagen)	>1% in the ExAC database, 1000 Genomes Project or ESP	MAF≥5%	
Butler	2015	yes	Base quality sums (GATK BaseRecalibrator)		Hg19	MuTect v.1.1.4	dbSNP; except if the variant was also in COSMIC database.	≥30X coverage; MAF ≥1.5%; ≥2 reads supporting the variant	mutations present in matched normal >1 read and mutations only present in one strand
Chicard	2018	yes	Quality score ≥30			GATK(V3.5); Samtools-0.1.18	>1% in 1000 Genomes Project (10000gAprl_2012) or ESP6500	≥20X coverage; ≥2 reads supporting the variant; SNVs within exons or slice sites	Synonymous variants except those reported in COSMIC

Supplementary Table A.4 - Continued

Author	Year	Germline	Quality score		Reference genome	Variant caller	SNP filtering	Variant selection	
			Sequencing quality	Mapping quality				Inclusion	Exclusion
Dietz	2016	no	Base quality >50			GATK (V3.5)	dbSNP v138	Tumor DNA: MAF >20% or <80% and ≥20X coverage. cfDNA: ≥10X coverage	Variants in >2 patients to exclude technical artifacts
Goodall	2017	yes			Hg19	MuTect2			
Gremel	2016	yes			Hg19	VarScan v2.3.6	dbSNP v138		
Huang	2017	yes			Hg19	MuTect; Platypus			
Jiménez	2019	yes	Quality score ≥30		Hg19	GATK; Samtools-0.1.18; MuTect-1.1.7 MuTect2	>1% in 1000 Genomes Project (Aprl_2012) or ESP6500.	≥20X coverage; ≥2 reads supporting the variant; missense, splice site and non-synonymous variants	>5 reads in germline DNA
Klevebring	2014	yes	Base Alignment Quality (BAQ)>46; variants were filtered to remove positions residing in regions of the human genome with low uniqueness		Hg19	Mutect v1.1.5	1000 Genomes Project		
Koeppel	2017	yes	somatic score from 1-30 using in house pipeline		Hg19	CASAVA1.8.2	>1% in 1000 Genomes Project or ESP		Variants with mutated reads in the constitutional sample >5% were excluded
Murtaza	2013	yes	Phred quality ≥30	Mapping quality ≥60	Hg19	Samtools v0.1.17	1000 Genomes Project	≥10X coverage; ≥4 mutant reads with ≥1 read on each strand (foward and reverse); At loci with ,10-fold coverage in normal DNA and no mutant reads, mutations were called in plasma if a prior plasma sample showed no evidence of a mutation and was covered adequately (10 fold or more)	

Supplementary Table A.4 - Continued

Author	Year	Germline	Quality score		Reference genome	Variant caller	SNP filtering	Variant selection	
			Sequencing quality	Mapping quality				Inclusion	Exclusion
Murtaza	2015	yes	Phred quality ≥30	Mapping quality ≥60	Hg19	Samtools v0.1.17	1000 Genomes Project	≥10X coverage; ≥5 mutant reads with ≥1 read on each strand (foward and reverse); the binomial probability of observing the number of mutant reads given total depth at that locus was <0,001 assuming an error rate or 0,01	
Olmedillas-Lopez	2018	no	No phred threshold; median phred quality >28		Hg38	Not applicable; study focussed on differential presence of exons	Not applicable; study focussed on differential presence of exons	Not applicable; study focussed on differential presence of exons	Not applicable; study focussed on differential presence of exons
Song	2018	yes			Hg19	GATK	1000 Genomes Project and dbSNP		
Sun	2019	yes	No phred threshold; >85%, the guanine/ cytosine base pairs content of reads was qualified (the percentage of bases with a quality ≥ 20)			GATK	1000 Genomes Project	COSMIC; TCGA	
Tailor	2019	yes	Base quality >30 (FastQC)	Mapping quality ≥20	b37	VarScan2	dbSNP v138 (except for those observed in COSMIC), 1000 Genomes Project, ESP6500, CLINVAR Dabatase, RefGene, dbNSFP (v2.6)	MAF ≤1% for germline DNA; MAF >1.5% for plasma cfDNA; ≥20X coverage; ≥3 reads supporting the variant	
Toledo	2018	yes	Alelle score ≥12			NEXTGEN software (Softgenetics)	ExAC database, 1000 Genomes Project or ESP6500	≥20X coverage; ≥20 reads supporting the variant; MAF ≥3% and ≥35% of gDNA; F:R balance ≥0.1; F:R read percentage ≥0.45	

Supplementary Table A.4 - Continued

Author	Year	Germline	Quality score		Reference genome	Variant caller	SNP filtering	Variant selection	
			Sequencing quality	Mapping quality				Inclusion	Exclusion
Vandekerkhove	2017	yes	Base quality >30 (in-house algorithm)		Hg38		ExAC database, KAVIAR database	MAF ≥1%; ≥10 reads supporting the variant; MAF 25 times higher than the background error rate (average AF across all gDNA samples) and 3 times higher than the MAF in the paired gDNA sample	

Abbreviations: NR: Not reprted, gDNA: genomic DNA, MAF: Mutant Allele Fraction, ExAC: Exome Aggregation Consortium, ESP: Exome Sequencing Project, GATK: Genome Analysis Toolkit, F:R: Forward/Reverse

*based on power to observe ≥3 mutant allele reads with Cancer Cell Fraction = 1 and multiplicity = 1 using ABSOLUTE

Supplementary Table A.5 - Overview of WES detected SNVs and pre-analytical conditions per patient

Paper	Subject	Shared SNVs	All tissue SNVs	All SNVs	plasma only	All plasma SNVs	plasma only SNVs/All plasma SNVs	Sensitivity	Agreement rate	Input cfDNA	Coverage cfDNA	Tumor type	Estimated ctDNA fraction	Tumor purity
Adalsteinsson 2017	MBC_284	23	52	75	23	46	0,50	0,44	0,31	20	269,92	Breast	0,32	0,38
Adalsteinsson 2017	MBC_288	38	84	99	15	53	0,28	0,45	0,38	20	164,96	Breast	0,49	0,73
Adalsteinsson 2017	MBC_295	45	185	194	9	54	0,17	0,24	0,23	20	127,17	Breast	0,53	0,43
Adalsteinsson 2017	MBC_303	80	133	239	106	186	0,57	0,60	0,33	20	277,30	Breast	0,34	0,28
Adalsteinsson 2017	MBC_313	23	30	38	8	31	0,26	0,77	0,61	20	228,41	Breast	0,68	0,00
Adalsteinsson 2017	MBC_317	35	44	45	1	36	0,03	0,80	0,78	20	177,04	Breast	0,27	0,89
Adalsteinsson 2017	MBC_320	14	41	54	13	27	0,48	0,34	0,26	20	266,93	Breast	0,79	0,80
Adalsteinsson 2017	MBC_325	17	54	84	30	47	0,64	0,31	0,20	20	176,06	Breast	0,51	0,62
Adalsteinsson 2017	MBC_331	225	479	840	361	586	0,62	0,47	0,27	20	235,80	Breast	0,22	0,59
Adalsteinsson 2017	MBC_336	191	352	464	112	303	0,37	0,54	0,41	20	334,74	Breast	0,24	0,85
Adalsteinsson 2017	MBC_349	85	113	168	55	140	0,39	0,75	0,51	20	223,05	Breast	0,63	0,62
Adalsteinsson 2017	MBC_191	66	89	98	9	75	0,12	0,74	0,67	20	86,41	Breast	0,49	0,65
Adalsteinsson 2017	MBC_287	32	90	113	23	55	0,42	0,36	0,28	20	135,59	Breast	0,45	0,70
Adalsteinsson 2017	MBC_292	133	233	301	68	201	0,34	0,57	0,44	20	279,53	Breast	0,40	0,82
Adalsteinsson 2017	MBC_301	155	220	293	73	228	0,32	0,70	0,53	20	191,81	Breast	0,53	0,21
Adalsteinsson 2017	MBC_307	93	115	128	13	106	0,12	0,81	0,73	20	189,07	Breast	0,16	0,86

Supplementary Table A.5 - Continued

Paper	Subject	Shared SNVs	All tissue SNVs	All SNVs	plasma only	All plasma SNVs	plasma only SNVs/All plasma SNVs	Sensitivity	Agreement rate	Input cfDNA	Coverage cfDNA	Tumor type	Estimated ctDNA fraction	Tumor purity
Adalsteinsson 2017	MBC_315	28	40	58	18	46	0,39	0,70	0,48	20	210,17	Breast	0,86	0,70
Adalsteinsson 2017	MBC_318	89	110	189	79	168	0,47	0,81	0,47	20	202,33	Breast	0,27	0,00
Adalsteinsson 2017	MBC_321	125	173	269	96	221	0,43	0,72	0,46	20	220,73	Breast	0,45	0,16
Adalsteinsson 2017	MBC_330	26	40	67	27	53	0,51	0,65	0,39	20	206,06	Breast	0,15	0,46
Adalsteinsson 2017	MBC_335	60	93	126	33	93	0,35	0,65	0,48	20	269,06	Breast	0,44	0,40
Adalsteinsson 2017	MBC_339	191	326	499	173	364	0,48	0,59	0,38	20	211,15	Breast	0,57	0,58
Adalsteinsson 2017	MBC_8	69	124	126	2	71	0,03	0,56	0,55	20	149,42	Breast	0,31	0,42
Adalsteinsson 2017	CRPC_554	22	30	47	17	39	0,44	0,73	0,47	20	200,08	Prostate	0,17	0,50
Ahlborn 2019	P41	2	2	2	0	2	0,00	1,00	1,00	10	211,8	Bile duct	0,26	
Ahlborn 2019	P37	2	3	3	0	2	0,00	0,67	0,67	10	218,5	Breast	0,06	
Ahlborn 2019	P42	0	3	3	0	0		0,00	0,00	10	242,9	Breast	0,00	
Ahlborn 2019	P44	0	2	2	0	0		0,00	0,00	10	154,7	Breast	0,00	
Ahlborn 2019	P39	0	2	2	0	0		0,00	0,00	10	189,5	Colorectal	0,00	
Ahlborn 2019	P40	1	2	2	0	1	0,00	0,50	0,50	10	212,4	Colorectal	0,12	
Ahlborn 2019	P43	0	3	3	0	0		0,00	0,00	10	164	Colorectal	0,00	
Ahlborn 2019	P1	3	4	4	0	3	0,00	0,75	0,75	10	208,8	Endometrial	0,36	
Ahlborn 2019	P4	2	5	5	0	2	0,00	0,40	0,40	10	144,3	Head and Neck	0,12	
Ahlborn 2019	P2	1	1	2	1	2	0,50	1,00	0,50	10	202,6	Ovarian	0,30	
Ahlborn 2019	P6	2	4	5	1	3	0,33	0,50	0,40	10	404,6	Pancreatic	0,57	
Ahlborn 2019	P9	1	1	1	0	1	0,00	1,00	1,00	10	132	Prostate	0,28	
Ahlborn 2019	P8	2	3	4	1	3	0,33	0,67	0,50	10	366,9	Testicular	0,21	
Ahlborn 2018	Pt5 baseline	5	28	178	150	155	0,97	0,18	0,03	10	50	Bile duct	0,20	0,65
Ahlborn 2018	Pt5 progression	4	40	198	158	162	0,98	0,10	0,02	10	50	Bile duct	0,06	0,39
Chicard 2018	Case 1	14	16	21	5	19	0,26	0,88	0,67	100		NB	0,42	0,34
Chicard 2018	Case 3	12	27	53	26	38	0,68	0,44	0,23	100		NB	0,70	0,88
Chicard 2018	Case 4	23	25	62	37	60	0,62	0,92	0,37	20		NB	0,61	0,93
Chicard 2018	Case 5	20	24	59	35	55	0,64	0,83	0,34	100		NB	0,99	0,91
Chicard 2018	Case 6	20	20	40	20	40	0,50	1,00	0,50	100		NB	0,75	0,98
Chicard 2018	Case 7	15	17	27	10	25	0,40	0,88	0,56	200		NB	0,58	0,90
Chicard 2018	Case 8	12	12	34	22	34	0,65	1,00	0,35	100		NB	0,93	0,95
Chicard 2018	Case 9	10	10	78	68	78	0,87	1,00	0,13	100		NB	0,92	0,32
Chicard 2018	Case 10	12	17	28	11	23	0,48	0,71	0,43	100		NB	0,82	0,82

Supplementary Table A.5 - Continued

Paper	Subject	Shared SNVs	All tissue SNVs	All SNVs	plasma only	All plasma SNVs	plasma only SNVs/All plasma SNVs	Sensitivity	Agreement rate	Input cfDNA	Coverage cfDNA	Tumor type	Estimated ctDNA fraction	Tumor purity
Chicard 2018	Case 15	12	15	39	24	36	0,67	0,80	0,31	50		NB	0,33	0,18
Chicard 2018	Case 16	27	32	41	9	36	0,25	0,84	0,66	30		NB	0,48	0,86
Chicard 2018	Case 17	12	21	21	0	12	0,00	0,57	0,57	30		NB	0,52	0,15
Chicard 2018	Case 18	14	14	14	0	14	0,00	1,00	1,00	30		NB	0,38	0,77
Chicard 2018	Case 19	13	17	17	0	13	0,00	0,76	0,76	100		NB	0,17	0,94
Dietz 2016	P1	1090	3322	4892	1570	2660	0,59	0,33	0,22	10	68,5	NSCLC		
Dietz 2016	P2	234	1892	2731	839	1073	0,78	0,12	0,09	10	68,5	NSCLC		
Dietz 2016	P3	148	2861	3302	441	589	0,75	0,05	0,04	10	68,5	NSCLC		
Dietz 2016	P4	1265	2232	5072	2840	4105	0,69	0,57	0,25	10	68,5	NSCLC		
Dietz 2016	P5	621	2820	3958	1138	1759	0,65	0,22	0,16	10	68,5	NSCLC		
Dietz 2016	P6	241	2294	2822	528	769	0,69	0,11	0,09	10	68,5	NSCLC		
Goodall 2017		25	28	30	2	27	0,07	0,89	0,83	40		Prostate	0,21	
Huang 2017	HCC-07	5	76	91	15	20	0,75	0,07	0,05	50	226,2	HCC		0,70
Huang 2017	HCC-09	66	67	250	183	249	0,73	0,99	0,26	50	226,2	HCC		0,70
Huang 2017	HCC-10	108	378	513	135	243	0,56	0,29	0,21	50	226,2	HCC		0,70
Huang 2017	HCC-11	15	223	257	34	49	0,69	0,07	0,06	50	226,2	HCC		0,70
Huang 2017	HCC-12	29	276	435	159	188	0,85	0,11	0,07	50	226,2	HCC		0,70
Jiménez 2019	Patient 2	1	4	9	5	6	0,83	0,25	0,11	65		Clear Cell Sarcoma		0,32
Jiménez 2019	Patient 1	11	15	15	0	11	0,00	0,73	0,73	56,8		Renal Cell Carcinoma	0,36	0,38
Murtaza 2013	Case 4	172	895	895	0	172	0,00	0,19	0,19	12	73	Ovarian	0,45	
Murtaza 2015		50	65	78	13	63	0,21	0,77	0,64	10	77,3	Breast	0,04	0,70
Song 2018		16	55	111	56	72	0,78	0,29	0,14		140	Thyroid Carcinoma		
Toledo 2018		54	74	88	14	68	0,21	0,73	0,61	15	83	Colorectal	0,47	

Shared SNVs: Single nucleotide variants in both plasma and tissue; Tissue SNVs: All single nucleotide variants in tissue (independant of plasma); All SNVs: All single nucleotide variants in plasma and/or tissue

NB: neuroblastoma

CHAPTER 8

Novel methods to diagnose leptomeningeal metastases in breast cancer

Neuro Oncology. *2019 Mar 18;21(4):428-439.*

Lindsay Angus, John W.M. Martens, Martin J. van den Bent, Peter A.E. Sillevis Smitt, Stefan Sleijfer, Agnes Jager

Abstract

Leptomeningeal metastases (LM) in breast cancer patients are rare but often accompanied by devastating neurological symptoms and carries a very poor prognosis, even if treated. To date, two diagnostic methods are clinically used to diagnose LM: gadolinium MRI of the brain and/or spinal cord and cytological examination of cerebrospinal fluid (CSF). Both techniques are however hampered by limited sensitivities, often leading to a long diagnostic process requiring repeated lumbar punctures and MRI examinations. To improve the detection rate of LM, numerous studies have assessed new techniques. In this review, we present the current diagnostic work-up to diagnose LM, set out an overview of novel techniques to diagnose LM and give recommendations for future research.

Introduction

Leptomeningeal metastases (LM) are caused by tumor involvement of the leptomeninges and its occurrence is often accompanied by devastating symptomatology. Malignant cells can invade the leptomeninges through different ways including hematogenous spread or direct infiltration from (para)vertebral metastases or any other metastasis in close contact to the cerebrospinal fluid (CSF). Once shed into the CSF, cancer cells may float along CSF pathways to other areas of the nervous system where they may settle and grow[1]. Both solid as well as hematological malignancies can give rise to LM.

Breast cancer (BC) is one of the most common solid malignancies that metastasizes to the leptomeninges, accounting for 19 - 36% of all LM cases[2-5]. BC is a heterogeneous disease comprising several molecular subtypes, differing between each other in natural course, molecular background and sensitivity to anti-tumor treatments. Among BC patients with LM, 23 - 40% of the patients have a triple negative subtype, 35 - 46% ER-positive/ HER2-negative, and 22 - 28% of the patients have HER2-positive disease[6-8]. In addition, in postmortem case series, LM were found in 12 - 16% of the patients with lobular carcinoma compared to only 0.3% - 5% of patients with ductal carcinoma[9,10]. During the past decades the incidence of LM seems to increase, probably as a result of higher success rates of systemic treatments resulting in more patients achieving long-term survival allowing LM to develop.

The prognosis of BC patients with symptomatic untreated LM is dismal with a median survival time of 4 - 8 weeks[5]. Currently, there is no consensus for choice of treatment for these patients. Treatment options consist of radiotherapy of clinically symptomatic areas and systemic and/or intrathecal delivered chemotherapy, improving the median overall survival to 3 - 8 months[6,8,11,12]. An explanation for the dismal prognosis of BC patients with LM could be the delay in diagnosing LM. This is both due to the frequently discrete symptoms at presentation and the limited sensitivity of the currently available diagnostic techniques especially early on in the development of LM. Consequently, once diagnosed, patients often have a poor clinical condition resulting in an impaired tolerance of systemic treatment or even worse, the inability of starting treatment at all.

Currently, cytological identification of malignant cells in CSF is the gold standard for diagnosing LM[3]. Although this technique has a high specificity (>95%), the sensitivity is only 45 - 75% at initial CSF examination and increases to 64 - 84% after a second CSF examination[2,3,13,14]. Clearly, improvement of the diagnostic work-up for suspicion of LM is needed. In this review we summarize the current diagnostic work-up to diagnose

LM, and different methods that have been investigated over the years to detect LM with a particular focus on patients with BC. Of note, these data are to a great extent generalizable and relevant for LM from other solid malignancies.

Clinically available diagnostic techniques

CSF: general laboratory assessments

Nearly all patients with LM have some kind of abnormality in their CSF including elevated opening pressure (30 - 57%), elevated leukocyte counts (44 - 57%), increased protein concentration (74 - 86%) and decreased glucose concentration (31 - 56%)[2,13,14]. Nevertheless, none of these are pathognomonic for LM.

CSF: pathology

After obtaining CSF by lumbar puncture (LP), a cytospin is made and stained with May-Grünwald Giemsa. A positive cytology result is defined as the presence of tumor cells in CSF. For BC, stainings as pan-cytokeratine, estrogen receptor and progesterone receptor can be helpful to confirm LM.

The diagnostic value of a positive CSF cytology has been evaluated by comparing premortem CSF examinations of patients diagnosed with cancer with autopsy results. The presence of malignant cells as diagnosed by cytology in the CSF was in 96% confirmed by leptomeningeal involvement at autopsy. In this study false positives, defined as no pathological evidence of LM at autopsy, were rare: only 5 of 117 CSF examinations concerning four patients with a hematological malignancy and one with medulloblastoma[3]. Detection of malignant cells in CSF by cytopathological analysis therefore has a specificity of >95%. In patients with solid tumors, CSF examination showed no false-positive results, indicating an even higher specificity of nearly 100%. In addition, in 42 patients with parenchymal brain metastases only found at autopsy, no tumor cells were found in CSF. However, only 30 of 51 autopsy-proven LM patients had a premortem positive cytology, resulting in a relatively low sensitivity of 59%[3]. The number of analyzed CSF samples per patients was not indicated.

To improve the sensitivity of CSF cytology several recommendations have been made including[15]: analysis of a large volume (ideally >10.5mL) of CSF; sampling from a clinically or radiographically suspicious location (i.e. LP in case of spinal signs or symptoms and ventricular fluid in case of cranial signs or symptoms); process CSF immediately after collection and perform a second or even a third CSF sample if the

first examination remains negative. Repeated LPs have been shown to increase the sensitivity of CSF cytology by 30%[2,13,14].

Gadolinium-enhanced magnetic resonance imaging (Gd-MRI)

In addition to CSF cytology, the gadolinium-enhanced magnetic resonance imaging (Gd-MRI) of the neuraxis is used to detect LM. Until now, no MRI studies for LM have been compared to autopsy studies and due to a limited number of studies it remains difficult to appreciate the diagnostic accuracy of Gd-MRI. Reported sensitivities and specificities range from 53 - 80% and 77 - 93%, respectively[4,6,11,12,16]. Comparison of T1-weighted Gd-MRI with contrast enhanced computed tomography (CE-CT) in patients with cytologically confirmed LM showed a favorable detection rate for Gd-MRI: 70% versus 36% and all abnormalities detected by CE-CT were also detected by Gd-MRI[17]. As a consequence, CE-CT should only be considered as diagnostic tool when an MRI is contraindicated.

Current diagnostic work up for clinical suspicion of LM

As recommended in the first edition of the EANO-ESMO clinical practice guideline for LM, cancer patients with suspicion of LM should undergo Gd-MRI assessment as first choice evaluation[18]. In patients presenting with typical clinical signs and symptoms of LM, corresponding abnormalities on Gd-MRI are sufficient to diagnose LM without cytological confirmation. Whenever the Gd-MRI results are inconclusive, a LP for CSF cytology is recommended. In case the first CSF examination is negative, a second LP is advised[14].

The role of biomarkers in CSF

Due to the limited sensitivities of Gd-MRI and cytopathology, together with the urge for earlier diagnosis of LM, there is an unmet need for novel diagnostic tests to improve the detection rate of LM. To achieve that, numerous potential biomarkers have been studied in CSF of diverse tumor types, including BC. To compare and interpret the value of these diagnostic tests, test characteristics such as sensitivity and specificity are necessary, requiring the presence of a gold standard. These comparisons are challenging because diagnostic criteria are not standardized, resulting in different definitions for LM in various publications. Until now, the majority of clinical trials have evaluated diagnostic tests for detection of LM using either positive CSF cytology or a combination of positive

CSF analysis, MRI findings and clinical presentation consistent with LM as reference standard. Definitions used for the diverse studies for LM positive cases and control groups are amongst others presented in **Table 1** and an overview of normal values for these biomarkers is shown in **Table 2**.

Proangiogenic proteins: VEGF, uPA, tPA, TGF-beta

Angiogenesis, the process leading to the formation of new blood vessels from preexisting vasculature, plays an important role in tumor growth, invasion and metastasis formation[19]. Vascular endothelial growth factor (VEGF), urokinase-type plasminogen activator (uPA), tissue-type plasminogen activator (tPA) and transforming growth factor-beta (TGF-beta) are involved in angiogenesis and have been evaluated as biomarkers for detection of LM[20].

Stockhammer *et al*[21] measured VEGF levels using ELISA in CSF and matched serum from eleven patients with solid malignant tumors including four patients with BC and cytology- or MRI-proven LM. In this small study, all eleven patients with LM showed high VEGF levels in CSF (median 6,795 pg/mL, range 745 - 18,791 pg/mL), compared to matched serum (median 438 pg/mL, range 47 - 580 pg/mL). Patients with LM had significantly higher CSF VEGF concentrations than patients with bacterial meningitis (median 38 pg/ml, range <25 - 633pg/mL; p<0.001). In patients with brain metastasis without LM, VEGF levels were undetectable. To discriminate between intrathecal VEGF production and passive influx of VEGF from blood a VEGF-index was calculated. Higher VEGF indices were found in LM patients, suggesting local VEGF production or (less likely) active import. These data were supported by the study of Corsini *et al*[22] in which 15/18 (83%) patients with cytology proven LM had an increased VEGF index compared to only 3/26 patients with nonmalignant neurologic diseases resulting in a specificity of 88%.

Thereafter, four studies confirmed increased CSF VEGF levels or VEGF indices in patients with LM with three studies reporting sensitivities of 51 - 75%[20,23-25]. Reijneveld *et al*[24] found besides increased VEGF CSF levels also higher uPA CSF levels in patients with LM. In an another prospective study paired serum and CSF from patients with metastatic disease with and without LM were collected, from patients with bacterial and viral meningitis and a control group of patients with nonmalignant neurologic diseases[20]. VEGF, uPA, tPA, TGFβ1 indices were calculated. Although the VEGF concentration was significantly higher in CSF of patients with LM than in all other groups, the VEGF index was not significantly different between groups. In contrast, the tPA index was significantly decreased in LM compared with other groups (p <0.01) whereas uPA and TGFβ1 CSF indices showed no differences between groups.

Table 1 - Test characteristics for the detection of leptomeningeal metastases in breast cancer patients by using CSF biomarkers

Ref.	Method	Number of patients with LM from BC versus LM from all tumor types	Biomarker	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Definition of LM	Definition control group (no. of patients)
				Pro-angiogenic proteins					
Stockhammer *et al.* [21]	ELISA VEGF	4 out of 11	VEGF	NA	NA	NA	NA	Clinical signs, MRI findings, and positive CSF cytology	I: Pts with solid malignancies without LM (16) II: Pts with infectious meningitis (35) III: Pts with non-malignant and noninfectious diseases (100)
Corsini *et al.* [22]	ELISA VEGF	11 out of 18	VEGF index (cut-off 10)	83.3	88.4	NA	NA	Positive CSF cytology	Non-malignant neurological diseases: inflammatory (33), neurodegenerative (17)
Herrlinger *et al.*[23]	ELISA VEGF	5 out of 37	VEGF (cut-off 100 pg/mL)	73	93	77.1	91.5	Positive CSF cytology, or contrast- enhancing subarachnoid tumor cell deposits on MRI or both	Non-malignant neurological disease (50), multiple sclerosis (28), presumed CNS infectious disease (37)
			VEGF (cut-off 250 pg/mL)	51.4	98.3	90.5	86.3		
Reijneveld *et al.* [24]	ELISA VEGF	31 out of 53	VEGF	NA	NA	NA	NA	Clinical features compatible with LM and positive CSF cytology	I:Pts with malignancies without LM (negative CSF and no signs of LM or brain metastases on imaging) (18) II: No malignancy (no systemic malignancy, no CNS infection, no signs of LM or brain metastases on imaging) (25)
Langerijt, van de *et al.* [20]	ELISA VEGF, tPA	9 out of 19	VEGF index	54[a]/62[b]	72[a]/71[b]			Positive CSF cytology and/or enhancement of leptomeninges on MRI	I: Pts with solid malignancy (negative cytology and MRI) (54) =[a] II: Infectious meningitis: viral (16), bacterial (16), non-malignant and noninfectious neurologic disorders (27) =[b]
			tPA index	85[a]/85[b]	79[a]/96[b]				
			VEGF + tPA	100	73	54	100		
Groves *et al.* [25]	ELISA VEGF	8 out of 22	VEGF (all tumor types)	71	97	88	91	Positive CSF cytology	I: BC (33), lung (25), melanoma (9) pts suspected for LM with negative CSF cytology II: No disease at all (1)
			VEFG (BC)	75	97	86	94		

Table 1 - Continued

Ref.	Method	Number of patients with LM from BC versus LM from all tumor types	Biomarker	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Definition of LM	Definition control group (no. of patients)
				CK-BB, LDH, β2-microglobulin					
Bach *et al.* [27]	CK-BB bio-luminescence assay	12 out of 12	CK-BB (cut-off 0.2 U/l)	83	87	60	96	Positive CSF cytology or LM at autopsy less than 1 month after assayed CSF	I: BC pts with brain metastases (18) II: BC pts without CNS metastases (34)
			LDH (cut-off 80 U/l)	50	77	38	92		
Twijnstra *et al.* [30]	LDH with ACA method	15 out of 15	β2-microglobulin (cut-off >2.2 mg/l)	60	88	NA	NA	Positive CSF cytology (13) or LM at autopsy (2)	Neurological disorders without CNS metastases (16)
			LDH (cut-off>26 U/l)	60	93	NA	NA		
Twijnstra *et al.* [31]	LDH with ACA method	15 out of 34	LDH (cut-off >26 U/l)	79	83	33	83	Positive CSF cytology (32) or LM at autopsy (2)	I: Pts with solid tumors (66), hematological tumors (10), benign primary CNS tumors (7), malignant primary CNS tumors (10), others (112) II: Controls (110)
				CA15.3, CEA					
Le Rhun *et al.* [39]	CSF CA15.3 Automatized immuno-enzymatic technology	20 out of 20	CA 15.3 (cut-off 3 UI/mL)	80	70	NA	NA	Positive CSF cytology or clinical signs in combination with positive MRI	I: LM from other cancers than BC (20) (group 2) II: BC pts with brain metastases only (20) (group 3) III: No malignancy (20) (group 4)
Yap *et al.* [41]	Triple-isotope double antibody method	23 out of 23	CEA (cut-off 1.5 ng/mL)	70	100	NA	NA	Positive CSF cytology and absence of parenchymal brain metastases on CT	Stage IV BC pts and no evidence of metastatic CNS disease (10)
Corsini *et al.* [22]	CEA, CA15.3, CA125, CA19.9 (Modular Analytics SWA)	11 out of 18	Intrathecal synthesis of CEA/CA 15.3/ CA125/CA19.9	100	100	NA	NA	Positive CSF cytology	Non-malignant neurological diseases: inflammatory (33), neurodegenerative (17)

Table 1 - Continued

Ref.	Method	Number of patients with LM from BC versus LM from all tumor types	Biomarker	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Definition of LM	Definition control group (no. of patients)
Twijnstra *et al.* [30]	CEA	15 out of 15	CEA (cutoff >4 ng/mL)	60	93	-	-	Positive CSF cytology or LM at autopsy	Neurological disorders without CNS metastases (16)
					Proteomics				
Dekker *et al.* [42]	MALDI-TOF MS	54 out of 54		79	76	NA	NA	Positive CSF cytology or a compatible neurological syndrome and diagnostic MRI	I: BC pts with negative cytology and follow-up incompatible with LM (54) II: No cancer and no neurological disease
					EpCAM-based assays				
Nayak *et al.* [48]	CellSearch®	7 out of 15	Cut-off: presence of CSFTC	100	97.2	93.8	100	Positive CSF cytology or positive MRI findings observed within 1 month of the initial evaluation	I: Pts with clinical suspicion of LM (36) II: Non-malignant neurologic disease (9)
Lee *et al.* [49]	CellSearch®	18 out of 18	Cut-off >1.9 cells/mL	81	85	NA	NA	Positive CSF cytology	I: Non-malignant neurologic disease (14) II: BC pts with negative cytology (20)
Lin *et al.* [50]	CellSearch®	14 out of 30	Cut-off ≥1 cell/mL	93	95	90	97	Positive CSF cytology or unequivocal MRI findings performed within a month of the CSF analysis. Unequivocal MRI findings defined as: LM enhancement associated with subarchnoid nodules, basal cistern enhancement, or nerve root enhancement/clumping.	Cancer pts with clinical suspicion of LM without LM confirmation (65)
Kerklaan *et al.* [51]	EpCAM FCI	7 out of 13	Cut-off ≥2 CSFTC/ 5 mL CSF	100	100	NA	NA	Positive CSF cytology or a MRI with positive MRI or progressive neurological symptoms compatible with LM and exclusion of other causes	Cancer pts with clinical suspicion of LM without LM confirmation (16)
Subirá *et al.* [45]	EpCAM FCI	32 out of 49	Cut-off 10 clustered events	75.5	96.1	97.4	67.6	Positive CSF cytology or the combination of clinical signs and either MRI or biochemical CSF findings	Cancer pts with clinical suspicion of LM in whom LM was excluded (26)

Table 1 - Continued

Ref.	Method	Number of patients with LM from BC versus LM from all tumor types	Biomarker	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Definition of LM	Definition control group (no. of patients)
Subirá *et al.* [52]	EpCAM FCI	39 out of 94	Cut-off 16 clustered events	79.8	84	90.36	68.85	Positive CSF cytology and/or compatible clinical signs plus MRI findings with biochemical CSF abnormalities	Cancer pts with clinical suspicion of LM in whom LM was excluded (50)
					ctDNA				
Bougel *et al.* [60]	Methylation of h*TERT* with MS-HRM	8 out of 9	Methylation h*TERT*	92	100	100	95	Positive CSF cytology	No malignancy (21)

Abbreviations: MS-HRM= methylation-sensitive high resolution melting, NA= not available, pts= patients,
[a] = Sensitivity and specificity to discriminate LM from LM- negatives; [b]= Sensitivity and specificity to discriminate LM from control patients

Table 2 - Normal values of biomarkers in CSF

Biomarker	Normal range	Unit	Ref
VEGF	0 - 633	pg/mL	20,21,24
VEGF index	12 - 41		20
TGF-beta	65 - 115	pg/mL	20
uPA	41 - 282	pg/mL	20,24
tPA	122 - 222	pg/mL	20
tPA index	13 - 42		20
CK-BB	0.04 - 0.19	U/L	27
LDH	0 - 26	U/L	30
β2-microglobulin	0.65 - 2.2	mg/L	30
CA15-3	0 - 0.3	IU/mL	40
CEA	0.8 - 4.0	ng/mL	30
CSFTC (CellSearch)	range varies among studies, <1 - 2	cells/mL	49-51
CSFTC (FCI)	range varies among studies, <10 - 16	clustered events	46,53
ctDNA	no established cut-off. Since ctDNA is tumor specific it should be undetectable in healthy controls.		

Based on these studies we can conclude that VEGF CSF levels are increased in patients with LM, however the threshold for increased VEGF differs between the various studies. In addition, the sensitivity of VEGF in CSF does not improve the sensitivity of cytology and therefore is not promising enough to replace cytology.

Enzymes: CK-BB and LDH

Creatine kinase-BB (CK-BB) is one of three isoenzymes of creatine kinase that reversibly catalyzes the conversion of creatine in phosphocreatine, consuming ATP. Since tumor cells have increased cellular activities to meet the demand for their high energy consumption high cytosol concentrations of CK-BB have been measured[26]. These increased CK-BB levels in tumor cells lead to the hypothesis that CK-BB levels in CSF of patients with LM may also be elevated. Bach *et al*[27] measured CK-BB in CSF of BC patients suspected of having central nervous system (CNS) metastases. Elevated CK-BB levels (cut-off 0.20 U/l) in CSF were reported in 83% of BC patients with LM, however levels were also elevated in 39% of BC patients with parenchymal brain metastases but no LM. In a companion study, CK-BB levels in CSF of BC patients without LM, were significantly lower compared to patients with CNS metastases (median 0.12 U/l versus 0.42 U/l, $p<0.001$)[28]. From these studies can be concluded that elevated CK-BB could be an indicator of CNS metastases, however seems unsuitable to distinguish between LM from parenchymal metastases.

Lactate dehydrogenase (LDH) is an enzyme reversibly catalyzing lactate into pyruvate by converting NAD+ into NADH. The normal range of total CSF LDH level is 0 - 26 U/l[29]. Increased LDH levels in CSF have been reported in numerous conditions including cerebrovascular accidents, infectious meningitis, acute brain injury, primary CNS tumors, CNS metastases and in patients with LM of solid and hematological tumors[14,29-31]. LDH consists of five isoenzymes, expressed at different levels in various regions of the brain. In normal brain tissue, particularly aerobically active isoenzymes such as LDH-1 and LDH-2 are expressed[32]. Malignant cells, which are more dependent on anaerobic glycolysis, have a preponderance for anaerobically active LDH-4 and LDH-5 enzymes[33]. However, quantification of total LDH levels in CSF did not distinguish patients with LM from patients with bacterial meningitis[29]. In this latter study, determination of isoenzymes was not performed, because increased CSF LDH-4 and LDH-5 levels were already reported in infectious meningitis and therefore not specific for LM[32,34]. Thus, LDH levels in CSF, even when considering the isoenzymes LDH-4 and LDH-5, are not specific enough to detect LM.

β2-microglobulin

β2-microglobulin (B_2-m) is a protein, a small subunit of the Human Leucocyte Antigens, present on the surface of all nucleated cells, but particularly expressed on lymphocytes and macrophages. B_2-m is shed from cellular membranes and various (non)malignant conditions lead to detectable B_2-m levels in plasma, serum, urine, saliva, amniotic fluid and CSF[35,36]. Theoretically, high cell membrane turnover rates, as is the case in malignancies, would lead to increased B_2-m levels in surrounding fluids. However, only 60% of advanced BC patients with LM defined by either positive CSF cytology or autopsy had increased β2-microglobulin CSF levels, which is disappointing[30].

CA15.3 and CEA

Cancer antigen 15.3 (CA15.3), is a large transmembrane glycoprotein, produced by normal glandular breast epithelial cells. CA15.3 is often increased in BC but sometimes also in other malignancies as lung, pancreatic, colon, ovarian, prostate and benign conditions such as liver cirrhosis[37]. Nevertheless, elevated serum levels of CA15.3 are quite specific for BC and could serve as a useful marker for the diagnosis of LM[38] .

Interestingly, Le Rhun *et al*[39] compared CA15.3 levels in serum and CSF in four patient groups: BC patients with LM (1), patients with LM from other primary solid malignancies (2), BC patients with parenchymal brain metastases without LM (3) and women undergoing diagnostic LP for various nonmalignant neurological indications (4). LM was defined as a positive cytology and/or clinical signs and imaging. Significantly elevated CA15.3 levels in CSF were observed in BC patients with LM (median 8.7 IU/ml, range 0.1 - 251.0 IU/ml) compared to the other groups (median of patients with brain metastases 0.5 IU/ml, range 0.1 - 18.5 IU/ml). A cut-off CA 15.3 level of 3.0 IU/ml in CSF, resulted in a sensitivity of 80% and a specificity of 70% for detecting LM in BC.

Carcinoembryonic antigen (CEA) is a cell surface glycoprotein involved in cell adhesion. CEA is elevated in colon cancer but also in BC and some benign diseases of the gastro-intestinal tract and the liver[40]. Yap *et al*[41] studied CEA levels in CSF of 23 BC patients with cytologically proven LM who were treated with whole brain irradiation and intrathecal methotrexate. In 16 patients (70%) the CEA level before treatment was above the limit of detection of 1.5 ng/ml and decreased in patients with response and remained elevated in two patients without response. No correlation was found between CEA levels in CSF and in serum, suggesting local synthesis of CEA within the CSF of patients with LM[41]. Corsini *et al*[22] measured the well-known tumor markers CEA, CA15.3, CA125 and CA19.9 in serum and CSF of 18 patients with LM of solid tumors (11 BC patients) and 50 patients with other neurological diseases. Based on the Reiber's

formula intrathecal synthesis of the tumor markers was calculated[22]. All patients with LM had intrathecal synthesis for at least one tumor marker, while none of the controls had tumor marker production in CSF. Interestingly, intrathecal synthesis of CEA was observed in 17 of 18 (in 10/11 BC) patients. CA125 and CA19.9, were elevated and intrathecally synthesized in 6 (55%) of the BC patients. For now, limited data are available regarding the sensitivity and specificity of these markers and therefore short-term implementation into the clinic is not expected.

Proteomics

Multiplex immune-assays and mass spectrometry can obtain information on intra- and extracellular protein expression that could be relevant to the biology of LM. Dekker *et al*[42] developed a MALDI-TOF mass spectrometry assay, requiring only 20 µl of CSF, to investigate protein expression profiles in CSF of patients with advanced BC with and without LM. Patients were classified in three groups: BC patients with (group I, n=54) and without LM (group II, n=52) and control patients without any neurological disease (group III, n=52). 164 peptide peaks discriminated between the three patient groups (p-value <0.1). After bootstrap validation, a sensitivity of 79% and specificity of 76% to distinguish patients with LM from patients without LM was found. Using this method, it is not possible to identify specific peptides.

To detect the exact masses of the peptides, electrospray FITCR mass spectrometry was performed on a subset of samples of the study of Dekker *et al*[43] Using this method, 17 peptides corresponding to 9 proteins were identified. Proteins detected in the samples of patients with LM were mainly related to host-disease interaction, inflammation and immune defense (serotransferrin, alpha 1-antichymotrypsin, hemopexin, haptoglobin and transthyretin).

Based on previously obtained in vivo evidence that tumor cell adhesion is crucial for LM progression in mice and that leptomeningeal tumor growth elicits an intrathecal inflammatory response in the CSF, Brandsma *et al*[45,46] measured a profile of nine proteins, including adhesion molecules, cytokines, and chemokines by using a multiplex immunoassay. CSF of patients with cytologically proven LM (n= 57), patients with systemic malignancy without LM (n=20), patients with aseptic or viral meningitis (n=11) and patients with (non)-neurological diseases (n= 19) were analyzed. Median CSF levels of soluble Vascular Cell Adhesion Molecule-1 (sVCAM-1), soluble Intercellular Adhesion Molecule-1 (sICAM-1), Interleukin-8 (IL-8), Pulmonary and Activation Regulated Chemokine (PARC), Interleukin-18 (IL-18) and Interferon-γ inducible protein (IP-10) in patients with LM were significantly higher compared to the

control groups. Sensitivity and specificity for these markers were not calculated which makes it hard to appreciate the diagnostic value of these proteins.

In summary, proteomic studies give more insight into the biology of LM, for example showing that inflammatory proteins do play a role in LM. Despite the fact that unbiased mass spectrometry does elucidate more of the biology of LM, a more well-defined and specific for LM set of proteins is needed before clinical implementation can be considered. Nevertheless, if such a subset of proteins could be identified and determined by techniques that could be swiftly implemented in routine diagnostic work up, protein CSF analyses could be a promising diagnostic tool.

EpCAM-based circulating tumor cell detection

Solid tumors of epithelial origin like BC frequently express epithelial cell adhesion molecule (EpCAM) on the cell surface, allowing for detection and enumeration of these cells using anti-EpCAM antibodies. The most important methods for circulating tumor cell detection and enumeration in peripheral blood are the FDA-approved CellSearch® method and EpCAM based flowcytometry immunofenotyping (FCI) assays. In short, the CellSearch® assay uses immunomagnetic enrichment of circulating tumor cells after adding anti-EpCAM ferrofluid to 7.5mL peripheral blood. Subsequently, stainings for the nucleus (DAPI), cytokeratin (8, 18 and 19) and pan-leukocyte marker CD45 are added, the latter to stain leukocytes which need to be distinguished. Finally, a reviewer counts all cells meeting the criteria for circulating tumor cells[44].

The other frequently used technique, EpCAM-based fluorescent activated cell sorting (FACS), uses antibodies for anti-EpCAM, anti-CD45 for discrimination of leukocytes and markers for detection of nucleated cells (Hoechst or DRAQ5)[45].

Until now, four studies have used the CellSearch® method to detect tumor cells in CSF (CSFTC) of BC patients. Two small pilots studies with patients with BC and LM showed that detection and enumeration of CSFTC is feasible[46,47].

Nayak *et al*[48] detected CSFTC in 15/15 patients (8/8 BC patients) with LM defined as positive cytology or clear MRI findings and in one patient without LM, resulting in a sensitivity of 100% and specificity of 97.2%. One patient with a false-positive result, developed 6 months after the initial LP evidence of LM on MRI suggesting that CSFTC detection may have preceded findings on MRI being a very sensitive tool for LM detection.

In a prospective study, Lee *et al*[9] showed a high correlation (Pearson's r=0.94) between the CellSearch® technique and EpCAM-based FCI in CSF of 38 advanced BC patients suspected or known to have LM. To define the specificity of the CellSearch® assay, 14 patients with hematological malignancies were included as controls. Patients with either positive cytology in one of their CSF samples or unequivocal MRI signs were considered having LM. With a cut-off of ≥1.9 cell/mL CSF, a sensitivity of 81% and specificity of 85% was reached. Explorative analysis of serial CSFTC levels in seven patients receiving treatment for LM showed that patients with a decrease and at least one negative CSFTC measurement had longer survival times than patients who did not clear CSFTCs. Recently, the largest CellSearch®-based study so far involving 95 patients with solid tumors and clinical suspicion of LM, of whom 36 had BC, showed a high sensitivity of 93% and a high specificity of 95% using a cut-off of ≥1 CSFTC/mL CSF[50].

Using a cut-off of ≥2 CSFTC/5 mL CSF measured with EpCAM-based FCI, Kerklaan *et al*[51] demonstrated a sensitivity and specificity of 100% in 13 patients with LM from solid tumors.

Another EpCAM-based FCI study in 78 patients with carcinomas (44 BC), of whom 49 ultimately were diagnosed with LM based on positive cytology, or the combination of clinical signs and either MRI or biochemical CSF findings found a sensitivity of 75.5% and a specificity of 96.1%, using a cut-off of 10 clustered events[45]. In a subsequent study of 144 patients with carcinomas of whom 94 were diagnosed with LM, now using higher cut-off of 16 events, an even higher sensitivity of 79.8% but with a lower specificity of 84%[52].

In conclusion, EpCAM-based assays show promising sensitivities ranging from 76-100%, and specificities ranging from 85 - 100% and allow for the absolute quantification of cells present in a certain volume of CSF[45,48-52]. Hence, quantification of CSFTC could be used for disease monitoring and response assessments in addition to the new RANO response criteria for LM[53,54]. Moreover, the lower leukocyte background compared to peripheral blood also allows for more sensitive molecular characterization of the enriched CSFTC. In addition, recent whole exome sequencing efforts have revealed that the molecular profile of brain metastases differs from matched primary tumors[55]. Next, these brain metastases harbored clinically informative alterations, as a homozygous missense mutation in *BRCA2* and an activating *EGFR* (L858R) mutation in two patients with BC, which were not detected in their primary tumors[55]. Hence, isolation and molecular characterization of CSFTCs could potentially reveal why these cells give rise to LM and hopefully could lead to targets for therapy[56].

However, important to stress is that both techniques, CellSearch® and FCI, do not provide cytogenetic proof that EpCAM-positive cells are truly malignant. Another disadvantage of EpCAM-based assays, is that not all tumor cells do express EpCAM on their cell surface and subsets of EpCAM-negative tumor cells could be missed[57]. Future studies should focus on the ideal cut-off for CSFTC positivity.

CSF circulating tumor DNA (ctDNA)

Another potential diagnostic method for early diagnosis of LM is the analysis of circulating tumor DNA (ctDNA) in CSF. Solid malignant tumors, like BC, shed significant amounts of tumor specific DNA into the systemic circulation mainly through cellular necrosis or apoptosis. ctDNA contains tumor specific DNA alterations such as somatic mutations, copy number alterations and epigenetic modifications as methylation, but is present in a background of cell-free DNA derived from normal cells[58]. The challenge for sensitive variant detection in plasma ctDNA is the relative abundance of wild-type DNA derived from normal tissue and leukocytes. Although cell counts in CSF of LM patients are raised in 44-57% of the patients, the amount of leukocytes is still much lower compared to blood[59]. Therefore, the background of contaminating DNA derived from healthy cells may probably be less important, allowing for more sensitive detection of tumor derived alterations in CSF.

To date, only small studies have been performed focusing on detection of cfDNA in CSF of patients with LM of BC. In CSF, human telomerase reverse transcriptase (*hTERT*) methylation status has been studied in various cancer types including BC[60]. CSF *hTERT* methylation was detected in 11/12 samples from 9 patients (8 with BC) with a positive CSF according to cytopathology and *hTERT* methylation in the primary tumor, resulting in a sensitivity of 92%. *hTERT* methylation was not detected in control samples, consisting of inflammatory conditions or viral syndromes. In CSF samples of patients with a suspicious cytological result (insufficient intensity of cell atypia and/or insufficient number of atypical cells), with a corresponding *hTERT* methylated primary tumor, *hTERT* methylation was detected in 17/26 samples. In 10 patients without *hTERT* methylation in the primary tumor, no *hTERT* methylation could be detected in suspicious CSF samples, underscoring that it is essential to know the *hTERT* methylation status of primary tumor in order to report results of CSF analysis based on *hTERT* methylation. A panel targeting the most frequently mutated genes or epigenetic aberrations could overcome the problem of an unknown molecular status of the primary tumor.

In a cohort of metastatic patients, including 6 BC patients with CNS metastases, targeted capture massive parallel sequencing was performed on CNS metastases,

CSF and plasma ctDNA. Genomic alterations were detected in all CSF samples and confirmed in a matching CNS metastasis. In "warm" autopsy materials from a patient with HER2-positive metastatic BC three mutations (*PIK3CB* M819L, *PIK3CB* Q818H and *AHNAK2* L5292V) were exclusively present in a meningeal lesion and CSF but not in the extracranial metastases or plasma. This may indicate that some CNS derived genomic alterations are exclusively present in CSF and that the genetic landscape of CNS metastases should preferentially be examined in CSF[61]. CSF cytology from an advanced BC patient with clinical suspicion of LM was three times negative, however mutations in *ESR1, PTEN* and *MRPS33* were detected at mutant allelic frequencies ranging from 20 - 50% in CSF. LM was confirmed at autopsy, suggesting that CSF ctDNA assessed with next generation sequencing techniques could detect LM in a more sensitive way than CSF cytology.

Even though the number of currently available papers on CSF cfDNA is limited, CSF cfDNA seem a promising tool to diagnose CNS metastases. Recently, in patients with LM from non- small cell lung cancer driver genes were detected in all 26 CSF samples[62]. Future CSF cfDNA analyses should, based on the research question either use whole exome sequencing, targeted sequencing panels or digital PCR assays with or without including analyses of tumor tissue and germ-line DNA. For the investigation of the presence or absence of tumor DNA in CSF, tumor tissue containing a sufficient percentage of tumor cells should ideally be sequenced for detection of patient-specific genomic alterations. These patient-specific alterations then could be analyzed using a targeted sequencing approach or digital PCR in the suspected CSF sample. If the primary tumor or metastasis is not amenable for analysis or if unbiased genotyping of CSF is preferred, targeted sequencing panels covering the most frequently mutated genes in BC could be applied or less sensitive assays as whole exome sequencing and whole genome copy number analyses. With the advent of unique molecular identifiers in targeted sequencing panels, it is possible to quantify the number of mutated molecules in a certain volume of CSF. Nevertheless, it is important to emphasize that CSF ctDNA in patients with concurrent brain metastases could be derived from both brain metastases and LM, which diminishes the chance that ctDNA can be used to distinguish these conditions but can be of important value to determine actionable targets for therapy and could be used in response assessments[61,63].

Conclusions and recommendations

Many biomarkers have been evaluated in an attempt to improve the diagnostic sensitivity and accuracy for detection of LM. However, the majority of studied biomarkers have not reached a wide clinical use, due to limited specificity, sensitivity and/or lack of validation. A major problem of the investigation of new biomarkers for the detection LM is the use of a suboptimal 'gold standard' references such as cytopathology and MRI, as theoretically, autopsy to confirm LM is the ultimate proof of absence or presence of LM. In order to ease the comparison of different biomarker studies a uniform definition of LM should be used across studies. The EANO-ESMO guideline suggests to use the following definitions: confirmed for patients with cytologically or histologically proven LM; probable for patients with a history of cancer, without cytological proof but with typical clinical findings and neuroimaging findings as linear contrast enhancement, nodules or a combination of both MRI findings; possible for patients with a history of cancer, without cytological proof, without typical clinical findings but with linear and/or nodular MRI findings or typical clinical findings only; lack of evidence for patients without cytological proof, without MRI findings and without typical clinical signs. Although these criteria and the recently proposed RANO criteria have to be validated in prospective clinical trials, we believe that the use of uniform definitions will allow for better comparability of results from new trials[18].

In addition to optimizing the LM definition, trials should include control groups to determine the specificity of their assays. Selection of the appropriate control group could be done based on the clinical differential diagnosis. For example, patients with a history of cancer who are suspected of having LM are often not suspected of having bacterial meningitis because the latter can be easily distinguished from LM based on CSF protein, white blood cells and culture together with clinical signs and symptoms. In patients with an oncological history there is a need for markers which could distinguish patients with brain metastases only, brain metastases in combination with LM, LM only and no malignant CNS pathology at all. This discrimination is important because this will have clinical implications resulting in different treatment strategies and these four groups should therefore be included in future clinical trials.

New markers will only reach clinical use if they 1) are more sensitive than the currently established diagnostic methods: MRI and cytology or 2) do add significantly to the sensitivity of established methods or 3) can be used in a quantitative manner instead of only being qualitative as cytology and MRI enabling for response evaluation or 4) could predict outcome on therapies for LM and 5) are cost-effective. Applying these

criteria to the markers described in this paper, identification and enumeration of CSFTC by EpCAM-based assays and detection of CSF cfDNA, both seem the most promising tumor specific candidates to detect LM at an earlier stage.

To determine the value of CSFTC and CSF cfDNA in the diagnosis and management of patients with LM of BC, studies with larger number of patients have to be performed to validate and standardize these methods. As a consequence of the increasing use of DNA sequencing in the diagnostic field, the implementation of CSF cfDNA analyses will become feasible soon if standard operating procedures (SOPs) become available. To achieve this, we have to determine the optimal way of CSF collection and subsequent DNA isolation before sequencing and to subsequently evaluate the cost-effectiveness of these new tests in diagnostics. These SOPs will be even more important for CSFTC detection, since only some centers have a CellSearch system available and shipment of samples will therefore be required.

In addition, since brain metastases can harbor other genetic alterations than matched primary tumors[55], studying the genomic profile of CSFTC and ctDNA, could lead to a better insight why these cells metastasize to the leptomeninges and could potentially reveal actionable targets for therapy[62]. For now, cytopathology remains the 'golden standard', but it is important to gain additional proof for the value of new tumor specific markers in CSF.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. DeAngelis, L.M. & Posner, J.B. *Neurologic Complications of Cancer, 2nd ed. Contemporary Neurology Series.*, (Oxford University Press, 2008).
2. Olson, M.E. *et al.* Infiltration of the leptomeninges by systemic cancer. A clinical and pathologic study. *Arch Neurol* **30**, 122-37 (1974).
3. Glass, J.P. *et al.* Malignant cells in cerebrospinal fluid (CSF): the meaning of a positive CSF cytology. *Neurology* **29**, 1369-75 (1979).
4. Clarke, J.L. *et al.* Leptomeningeal metastases in the MRI era. *Neurology* **74**, 1449-54 (2010).
5. Little, J.R. *et al.* Meningeal carcinomatosis. Clinical manifestations. *Arch Neurol* **30**, 138-43 (1974).
6. Le Rhun, E. *et al.* A retrospective case series of 103 consecutive patients with leptomeningeal metastasis and breast cancer. *J Neurooncol* **113**, 83-92 (2013).
7. Lee, S. *et al.* Leptomeningeal metastases from breast cancer: intrinsic subtypes may affect unique clinical manifestations. *Breast Cancer Res Treat* **129**, 809-17 (2011).
8. Niwinska, A. *et al.* Breast cancer leptomeningeal metastasis: propensity of breast cancer subtypes for leptomeninges and the analysis of factors influencing survival. *Med Oncol* **30**, 408 (2013).
9. Harris, M. *et al.* A comparison of the metastatic pattern of infiltrating lobular carcinoma and infiltrating duct carcinoma of the breast. *Br J Cancer* **50**, 23-30 (1984).
10. Lamovec, J. & Bracko, M. Metastatic pattern of infiltrating lobular carcinoma of the breast: an autopsy study. *J Surg Oncol* **48**, 28-33 (1991).
11. Gauthier, H. *et al.* Survival of breast cancer patients with meningeal carcinomatosis. *Ann Oncol* **21**, 2183-7 (2010).
12. de Azevedo, C.R. *et al.* Meningeal carcinomatosis in breast cancer: prognostic factors and outcome. *J Neurooncol* **104**, 565-72 (2011).
13. Balm, M. & Hammack, J. Leptomeningeal carcinomatosis. Presenting features and prognostic factors. *Arch Neurol* **53**, 626-32 (1996).
14. Wasserstrom, W.R. *et al.* Diagnosis and treatment of leptomeningeal metastases from solid tumors: experience with 90 patients. *Cancer* **49**, 759-72 (1982).
15. Glantz, M.J. *et al.* Cerebrospinal fluid cytology in patients with cancer: minimizing false-negative results. *Cancer* **82**, 733-9 (1998).
16. Straathof, C.S. *et al.* The diagnostic accuracy of magnetic resonance imaging and cerebrospinal fluid cytology in leptomeningeal metastasis. *J Neurol* **246**, 810-4 (1999).
17. Chamberlain, M.C. *et al.* Leptomeningeal metastasis: a comparison of gadolinium-enhanced MR and contrast-enhanced CT of the brain. *Neurology* **40**, 435-8 (1990).
18. Le Rhun, E. *et al.* EANO-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up of patients with leptomeningeal metastasis from solid tumours. *Ann Oncol* **28**, iv84-iv99 (2017).
19. Folkman, J. What is the evidence that tumors are angiogenesis dependent? *J Natl Cancer Inst* **82**, 4-6 (1990).
20. van de Langerijt, B. *et al.* CSF levels of growth factors and plasminogen activators in leptomeningeal metastases. *Neurology* **67**, 114-9 (2006).
21. Stockhammer, G. *et al.* Vascular endothelial growth factor in CSF: a biological marker for carcinomatous meningitis. *Neurology* **54**, 1670-6 (2000).
22. Corsini, E. *et al.* Intrathecal synthesis of tumor markers is a highly sensitive test in the diagnosis of leptomeningeal metastasis from solid cancers. *Clin Chem Lab Med* **47**, 874-9 (2009).
23. Herrlinger, U. *et al.* Vascular endothelial growth factor (VEGF) in leptomeningeal metastasis: diagnostic and prognostic value. *Br J Cancer* **91**, 219-24 (2004).
24. Reijneveld, J.C. *et al.* CSF levels of angiogenesis-related proteins in patients with leptomeningeal metastases. *Neurology* **65**, 1120-2 (2005).
25. Groves, M.D. *et al.* Biomarkers of disease: cerebrospinal fluid vascular endothelial growth factor (VEGF)

and stromal cell derived factor (SDF)-1 levels in patients with neoplastic meningitis (NM) due to breast cancer, lung cancer and melanoma. *J Neurooncol* **94**, 229-34 (2009).
26. Zarghami, N. *et al.* Creatine kinase BB isoenzyme levels in tumour cytosols and survival of breast cancer patients. *Br J Cancer* **73**, 386-90 (1996).
27. Bach, F. *et al.* Creatine kinase-BB in the cerebrospinal fluid as a marker of CNS metastases and leptomeningeal carcinomatosis in patients with breast cancer. *Eur J Cancer Clin Oncol* **25**, 1703-9 (1989).
28. Bach, F. *et al.* Diagnostic value of cerebrospinal fluid cytology in comparison with tumor marker activity in central nervous system metastases secondary to breast cancer. *Cancer* **72**, 2376-82 (1993).
29. van Zanten, A.P. *et al.* Cerebrospinal fluid lactate dehydrogenase activities in patients with central nervous system metastases. *Clin Chim Acta* **161**, 259-68 (1986).
30. Twijnstra, A. *et al.* Sensitivity and specificity of single and combined tumour markers in the diagnosis of leptomeningeal metastasis from breast cancer. *J Neurol Neurosurg Psychiatry* **49**, 1246-50 (1986).
31. Twijnstra, A. *et al.* Serial lumbar and ventricle cerebrospinal fluid lactate dehydrogenase activities in patients with leptomeningeal metastases from solid and haematological tumours. *J Neurol Neurosurg Psychiatry* **50**, 313-20 (1987).
32. Fleisher, M. *et al.* Lactic dehydrogenase isoenzymes in the cerebrospinal fluid of patients with systemic cancer. *Cancer* **47**, 2654-9 (1981).
33. Gerhardt, W. *et al.* Changes of LDH-isozymes, esterases, acid phosphatases and proteins in malignant and benign human brain tumors. *Acta Neurol Scand* **39**, 85-111 (1963).
34. Nussinovitch, M. *et al.* Cerebrospinal fluid lactate dehydrogenase isoenzymes in children with bacterial and aseptic meningitis. *Transl Res* **154**, 214-8 (2009).
35. Twijnstra, A. *et al.* Cerebrospinal fluid beta 2-microglobulin: a study in controls and patients with metastatic and non-metastatic neurological diseases. *Eur J Cancer Clin Oncol* **22**, 387-91 (1986).
36. Svatonova, J. *et al.* Beta2-microglobulin as a diagnostic marker in cerebrospinal fluid: a follow-up study. *Dis Markers* **2014**, 495402 (2014).
37. Colomer, R. *et al.* Circulating CA 15-3 antigen levels in non-mammary malignancies. *Br J Cancer* **59**, 283-6 (1989).
38. Kokko, R. *et al.* Ca 15-3 in the follow-up of localised breast cancer: a prospective study. *Eur J Cancer* **38**, 1189-93 (2002).
39. Le Rhun, E. *et al.* CSF CA 15-3 in breast cancer-related leptomeningeal metastases. *J Neurooncol* **117**, 117-24 (2014).
40. Hammarstrom, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* **9**, 67-81 (1999).
41. Yap, B.S. *et al.* CSF carcinoembryonic antigen in meningeal carcinomatosis from breast cancer. *JAMA* **244**, 1601-3 (1980).
42. Dekker, L.J. *et al.* MALDI-TOF mass spectrometry analysis of cerebrospinal fluid tryptic peptide profiles to diagnose leptomeningeal metastases in patients with breast cancer. *Mol Cell Proteomics* **4**, 1341-9 (2005).
43. Rompp, A. *et al.* Identification of leptomeningeal metastasis-related proteins in cerebrospinal fluid of patients with breast cancer by a combination of MALDI-TOF, MALDI-FTICR and nanoLC-FTICR MS. *Proteomics* **7**, 474-81 (2007).
44. Medical devices; immunology and microbiology devices; classification of the immunomagnetic circulating cancer cell selection and enumeration system. Final rule. *Fed Regist* **69**, 26036-8 (2004).
45. Subira, D. *et al.* Role of flow cytometry immunophenotyping in the diagnosis of leptomeningeal carcinomatosis. *Neuro Oncol* **14**, 43-52 (2012).
46. Le Rhun, E. *et al.* Development of a new method for identification and quantification in cerebrospinal fluid of malignant cells from breast carcinoma leptomeningeal metastasis. *BMC Clin Pathol* **12**, 21 (2012).
47. Patel, A.S. *et al.* Identification and enumeration of circulating tumor cells in the cerebrospinal fluid of breast cancer patients with central nervous system metastases. *Oncotarget* **2**, 752-60 (2011).
48. Nayak, L. *et al.* Rare cell capture technology for the diagnosis of leptomeningeal metastasis in solid tumors. *Neurology* **80**, 1598-605; discussion 1603 (2013).

49. Lee, J.S. *et al.* Detection of cerebrospinal fluid tumor cells and its clinical relevance in leptomeningeal metastasis of breast cancer. *Breast Cancer Res Treat* **154**, 339-49 (2015).
50. Lin, X. *et al.* Cerebrospinal fluid circulating tumor cells: a novel tool to diagnose leptomeningeal metastases from epithelial tumors. *Neuro Oncol* **19**, 1248-1254 (2017).
51. Milojkovic Kerklaan, B. *et al.* EpCAM-based flow cytometry in cerebrospinal fluid greatly improves diagnostic accuracy of leptomeningeal metastases from epithelial tumors. *Neuro Oncol* (2015).
52. Subira, D. *et al.* Diagnostic and prognostic significance of flow cytometry immunophenotyping in patients with leptomeningeal carcinomatosis. *Clin Exp Metastasis* **32**, 383-91 (2015).
53. Chamberlain, M. *et al.* Leptomeningeal metastases: a RANO proposal for response criteria. *Neuro Oncol* **19**, 484-492 (2017).
54. Chamberlain, M. *et al.* Leptomeningeal metastasis: a Response Assessment in Neuro-Oncology critical review of endpoints and response criteria of published randomized clinical trials. *Neuro Oncol* **16**, 1176-85 (2014).
55. Brastianos, P.K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* **5**, 1164-1177 (2015).
56. Li, X. *et al.* Clinical significance of detecting CSF-derived tumor cells in breast cancer patients with leptomeningeal metastasis. *Oncotarget* **9**, 2705-2714 (2018).
57. Mostert, B. *et al.* Detection of circulating tumor cells in breast cancer may improve through enrichment with anti-CD146. *Breast Cancer Res Treat* **127**, 33-41 (2011).
58. Diaz, L.A. *et al.* Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579-86 (2014).
59. Wright, B.L. *et al.* Cerebrospinal fluid and lumbar puncture: a practical review. *J Neurol* **259**, 1530-45 (2012).
60. Bougel, S. *et al.* Methylation of the hTERT promoter: a novel cancer biomarker for leptomeningeal metastasis detection in cerebrospinal fluids. *Clin Cancer Res* **19**, 2216-23 (2013).
61. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, 8839 (2015).
62. Li, Y.S. *et al.* Unique genetic profiles from cerebrospinal fluid cell-free DNA in leptomeningeal metastases of EGFR-mutant non-small cell lung cancer: a new medium of liquid biopsy. *Ann Oncol* (2018).
63. Pentsova, E.I. *et al.* Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. *J Clin Oncol* **34**, 2404-15 (2016).

CHAPTER 9

Detection of aneuploidy in cerebrospinal fluid from patients with breast cancer can improve diagnosis of leptomeningeal metastases

Clinical Cancer Research. *2021 May 15;27(10):2798-2806.*

Lindsay Angus, Teoman Deger, Agnes Jager, John W.M. Martens, Vanja de Weerd, Irene van Heuvel, Martin J. van den Bent, Peter A.E. Sillevis Smitt, Johan M. Kros, Eric M.J. Bindels, Ellen Heitzer, Stefan Sleijfer, Joost L.M. Jongen, Saskia M. Wilting

Abstract

Purpose

Detection of leptomeningeal metastasis (LM) is hampered by limited sensitivities of currently used techniques: MRI and cytology of cerebrospinal fluid (CSF). Detection of cell-free tumor DNA (ctDNA) in CSF has been proposed as a tumor specific candidate to detect LM at an earlier stage. The aim of this study was to investigate mutation and aneuploidy status in CSF-derived cfDNA of breast cancer patients with a clinical suspicion of LM.

Methods

Cell-free DNA was isolated from stored remnant CSF and analyzed by targeted next generation sequencing (NGS) (n=30) and the modified Fast Aneuploidy Screening Test-Sequencing System (mFAST-SeqS) (n=121). The latter method employs selective amplification of long interspaced nuclear elements (LINE-1)-sequences which are present throughout the genome and allows for fast and cheap detection of aneuploidy. We compared these results with the gold standard to diagnose LM: cytology.

Results

LM was cytology-proven in 13 of 121 patients. Low DNA yields resulted in insufficient molecular coverage of NGS for the majority of samples (success rate 8/30). The mFAST-SeqS method, successful in 112/121 (93%) samples, detected genome-wide aneuploidy in 24 patients. Ten of these patients had cytology-proven LM, 8 additional patients were either concurrently diagnosed with CNS metastases by radiological means or developed these soon after the LP. The remaining 6 cases were suspected of LM but could not be confirmed by cytology or imaging. Aneuploidy was associated with development of LM and significantly worse overall survival.

Conclusions

Aneuploidy in CSF-derived cfDNA may provide a promising biomarker to improve timely detection of LM.

Introduction

The incidence of leptomeningeal metastasis (LM) in patients with breast cancer is estimated to be around 5%[1,2]. Although the incidence is relatively low, once patients become symptomatic the symptoms can be devastating and the prognosis deteriorates with a median overall survival (OS) of 4-8 weeks[3], which increases to 3 to 8 months when treated[4-7]. The detection of LM in breast cancer patients is hampered by limited sensitivities of routinely used techniques. According to the EANO-ESMO guidelines, patients presenting with typical symptoms of LM and characteristic abnormalities on gadolinium-enhanced (Gd)-MRI, MRI may be diagnosed with (probable) LM without cytological confirmation[8]. However, Gd-MRI has a sensitivity of 53-80% and specificity of 77-93%[4,6,7,9,10]. A lumbar puncture (LP) for cerebrospinal fluid (CSF) cytology is recommended, which has a sensitivity of 45-75% at first examination and increases to 64-84% when a second LP is performed[11-14]. These limited sensitivities inevitably lead to delayed or missed diagnoses, thus improvement of diagnostic methods is urgently needed to allow for more timely treatment which may result in improved survival and quality of life.

During the past decades multiple biomarkers have been interrogated for their ability to improve the detection rate of LM specifically in the CSF[15]. Promising sources for the future application of biomarkers which are directly derived from LM are tumor cells in CSF, detected by EpCAM based methods[16-22], and the tumor derived cell-free DNA (ctDNA) fraction within the total cell-free DNA (cfDNA) pool present in the CSF[23]. Solid tumors such as breast cancer release tumor DNA by apoptosis and necrosis in all bodily fluids[24]. To date, only small studies have been performed focusing on detection of ctDNA in CSF of patients with LM originating from breast cancer. For instance, in a patient with clinical suspicion of LM of whom the CSF cytology result was three times negative mutations in *ESR1, PTEN* and *MRPS33* in CSF-derived cfDNA could be detected[23]. In this patient, LM was confirmed at autopsy, suggesting that assessment of tumor derived cfDNA in CSF could detect LM more sensitively than CSF cytology[23]. In lung cancer patients, *EGFR* mutations have been detected in CSF of patients with LM[25]. Similarly, in patients with brain metastases derived from solid tumors as well as in patients with primary brain tumors somatic alterations in CSF-derived cfDNA have been detected in 63% and 50%, respectively[26]. However, not all tumors carry hotspot mutations and therefore the use of mutations for detection of disease in CSF requires prior knowledge on the tumor's genetic make-up. Genome-wide untargeted approaches have the great advantage that upfront knowledge about the genetic alterations to be detected is not required. For example, the modified Fast Aneuploidy Screening Test-

Sequencing System (mFAST-SeqS) method employs selective amplification of long interspaced nuclear elements (LINE-1-sequences) which are present throughout the genome[27]. This method allows for the detection of somatic copy number alterations at a chromosome-arm resolution, representing a fast and affordable assessment of tumor fractions requiring only low amounts of DNA input (~1 ng)[27]. Belic *et al.* showed that copy number alteration patterns observed with either the mFAST-SeqS method or genome-wide shallow sequencing of plasma-derived cfDNA from metastatic breast cancer patients were highly correlated, whereas known chromosomal profiles from cell line DNA could be captured with mFAST-SeqS as well[27]. Considering virtually all breast cancers harbor copy number alterations (CNA), analysis of CNA represents an attractive alternative in CSF-derived cfDNA as well[28].

In this retrospective proof-of-concept study, we assessed the value of mutational analyses performing targeted next generation sequencing (NGS) and aneuploidy analyses on archival CSF samples using the mFAST-SeqS method in a large cohort of breast cancer patients who underwent an LP for the clinical suspicion of LM. Furthermore, the prognostic value of CSF-derived cfDNA analyses were, together with other routinely collected clinical parameters, CSF cytology, and CSF chemistry, associated with OS.

Patients and Methods

Study design and patients

Adult patients (≥18 years old) with a history of breast cancer who underwent an LP for clinical suspicion of LM (clinical signs and symptoms e.g. headache, nausea, mental changes, gait difficulties, meningeal rigidity, cranial nerve palsies, spinal symptoms, abnormalities at neurological examination) and from whom stored remnant CSF was available were included in this retrospective analysis. CSF samples used in this study were collected and stored as part of standard diagnostic work-up. Remaining CSF has been stored at -80˚C at the department of neuro-oncology at the Erasmus MC, Rotterdam, The Netherlands. The following clinical data were collected: medical history, information of the LP (date, volume of CSF used for cytology), routine CSF chemistry results (including leukocyte count, protein and glucose concentration), cytology result as reported by the pathologist (positive, equivocal (suspicious or atypical cells) or negative), MRI results as reported by the radiologist and neurological signs and symptoms as derived from the medical record prior to or at time of CSF collection, and follow-up (final diagnosis; start of systemic therapy/radiotherapy after LP; date of death). LM was defined as either malignant cells at cytology ("cytology+") and/or when characteristic MRI

abnormalities were observed (enhancing leptomeninges, leptomeningeal nodules or linear and/or radicular enhancement; "radiology+")[8]. To evaluate whether patients with CNS metastases at CSF collection developed additional CNS localizations over time or whether patients without CNS metastasis at CSF collection did develop these during follow-up, consecutive scans were evaluated for development of LM and/or brain metastases (scored as "final CNS diagnosis"). For the comparison with mutation analysis and mFAST-SeqS, the cytology results were used as reference. This study has been performed according the "Code of conduct for responsible use (2011)"[29] and the study design was approved by the Medical Research Ethics Committee of the Erasmus MC (MEC-2019-0504). Non-oncologic female patients with a clinical indication for a diagnostic LP for the following diagnoses were included as control group (n=12): idiopathic intracranial hypertension (2x), impaired consciousness, headache with spontaneous recovery, acute headache (no subarachnoidal bleeding), Alzheimer's disease, frontotemporal dementia, vertigo, suspected demyelinating disease (2x), neurosarcoidosis, and idiopathic facial nerve paresis.

CSF sample collection, cfDNA isolation and quantification

CSF samples were centrifuged for 10 minutes 2,000 g at 4˚C. After centrifugation, supernatant was stored at -80˚C until further handling. For cfDNA isolation, CSF samples were thawed at room temperature and 0.5-4.1 mL was used. cfDNA was isolated and eluted in 20 µL buffer using the QIAamp Circulating Nucleic Acid Kit (Qiagen) as per the manufacturer's instructions and stored at -20˚C. cfDNA concentrations were quantified using the Quant-iT dsDNA high-sensitivity assay (Invitrogen, Life Technologies, Carlsbad, CA) according to the manufacturer's instructions, and the Qubit fluorometer (Invitrogen) was used as read out.

Mutation analysis

A targeted NGS approach with molecular barcoding using Oncomine™ Breast cfDNA Assay v2 (Thermo Fisher Scientific) was applied for low limit somatic variant detection according the manufacturer's instructions. This assay consists of ten genes, frequently affected in breast cancer, covering 157 hotspots in genes including *AKT1*, *EGFR*, *ERBB2*, *ERBB3*, *ESR1*, *FBXW7*, *KRAS, PIK3CA*, *SF3B1,* and *TP53.* The amount of cfDNA used for sequencing ranged from 3.3-22.3 ng. Analyses were done as previously reported, using Ion S5 XL sequencing system and 540 chips, and evaluated with a standard variant calling pipeline[30]. First, raw Ion S5 sequencing results with the Oncomine cfDNA assays were loaded into the TorrentSuite variant caller 5.10. Applying additional filtering, hotspot variants were called when 1) at least 500 unique molecules for that particular position were sequenced resulting in a limit of detection of 0.2%, and 2) if the mutant

sequence was covered by 3 unique molecules with at least 3 reads per unique molecule.

mFAST-SeqS, sequencing and data analysis

We used the recently described mFAST-SeqS method, which has initially been established as a minimally-invasive screening method for fetal aneuploidy from maternal blood [31], but has been adapted by Belic *et al*. to estimate tumor fractions in cfDNA[27]. LINE-1 (L1) amplicon libraries were prepared as described by Belic *et al*. Briefly, using target-specific L1 primers and Phusion Hot Start II Polymerase II, a primary PCR step was performed to amplify L1 sequences throughout the genome using a single primer pair. This is followed by a secondary PCR step which amplifies all molecules from the first PCR step and adds adaptors and sample-specific index sequences. After both PCR steps, the PCR products are purified using AMPure XP beads (1.4 x, Beckman Coulter). The resulting libraries were quantified using the NEBNext Library Kit for Illumina (New England Biolabs), pooled equimolarly for 20 samples (2nM), supplemented with 25% of a PhiX control library, and sequenced on a MiSeq-system (Illumina) generating 150 base pair single reads aiming for at least 100,000 reads[32].

We trimmed the primers of the first PCR of the sequenced reads using Trimmomatic (v0.38). The trimmed reads were mapped on human reference genome hg19 using Burrows-Wheeler alignment (v0.7.17) and the read counts per chromosome arm were determined. Reads with a mapping quality >15 were counted and read counts were normalized to the total read count per sample. Subsequent computational random down-sampling in steps of 5,000 reads for 24 samples (12 cases and 12 controls) at 100 iterations showed that reliable results were obtained down to 90,000 reads per sample (**Supplementary Figure S1**). In short, we determined the delta in genome-wide z-score for every iteration compared to the genome-wide z-score obtained at 100,000 reads and observed an increase in the average delta genome-wide z-score with lower total numbers of read counts for both cases and controls. The average delta z-scores became divergent between cases and controls whereas the associated standard deviation showed a clear increase below 90,000 reads. Importantly, no false positive chromosome arms were observed in any of the control samples at 90,000 reads. Based on the foregoing we included samples with at least 90,000 mapped reads in the analysis. To test for over- and underrepresentation of each chromosome arm, we calculated z-scores by subtracting the mean and dividing by the standard deviation of normalized read-counts for the respective chromosome arm from a panel of 12 CSF controls. Because little to no reads align to the short arms of the acrocentric chromosomes 13p, 14p, 15p, 21p, 22p, and Y, these have been excluded from analysis. To get a general overview of aneuploidy, we squared and summed z-scores per chromosome arm into a genome-wide z-score.

Following the original threshold set by Belic *et al.*[27], we considered samples with a genome-wide z-score ≥5 as aneuploid.

Statistical methods

Descriptive statistics were calculated for variables of interest. Mann-Whitney *U* test was performed for univariate analyses of continuous variables and a Fisher's Exact test was used for categorical variables. To calculate the correlation between DNA input and molecular coverage we calculated the Spearman's rho. OS was calculated from time of CSF collection until death (event) or last follow-up (censored). Models associating variables of interest and OS, time to developing LM or brain metastases, were constructed using Cox proportional hazards methodology (enter method). The Kaplan-Meier method was used to graphically represent OS. Two-sided *P*-values below 0.05 were considered significant. All statistical analyses were performed using IBM SPSS version 25. Figures were constructed with GraphPad Prism.

Results

Patients and CSF characteristics

From January 2002 to April 2016, 121 breast cancer patients underwent an LP for suspected LM of whom left over CSF was available for cfDNA analyses (**Figure 1**). At CSF sampling the median age was 55 (interquartile range (IQR): 45-63 years). Thirteen patients (10.7%) had a positive cytology, whereas 2 (1.7%) and 106 (87.6%) patients had an equivocal or negative cytology result, respectively (**Table 1**). The median total amount of cfDNA, isolated from a median of 1.8 mL of CSF, was 8.72 ng. The median cfDNA concentration was 5.17 ng/mL CSF (IQR: 3.62-10.75 ng/mL CSF).

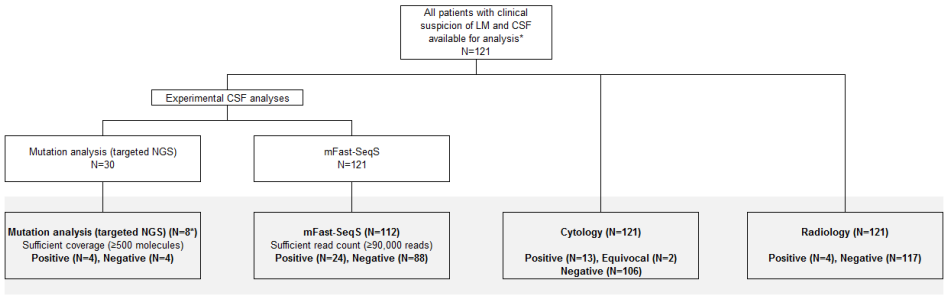


Figure 1 - Flowchart of the number of patients included per analysis.
*All eight patients with sufficient molecular coverage in the mutation analysis had also a sufficient read count in the mFAST-SeqS analysis

Table 1 - Patient characteristics

	N=121	
	N	**%**
Age at LP†	55 (45-63)	
Gender		
Female	121	100
CSF cytology		
Positive	13	10.7
Equivocal	2	1.7
Negative	106	87.6
CSF chemistry†		
Leukocytes (x 10^6/L, normal = 0-4 x 10^6/L)	2.0 (2.0-3.0)	
Protein (g/L, normal = 0.18-0.58 g/L)	0.32 (0.23-0.46)	
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	3.5 (3.3-4.0)	
MRI brain*	78	64.5
Normal	53	43.8
LM only	1	0.8
Suspicion of LM (leptomeningeal enhancement)	7	5.8
LM and brain metastases	2	1.7
Brain metastases only	7	5.8
Brain metastases and status after RT or resection	4	3.3
Dural metastases	5	4.1
Suspicion of brain metastases	1	0.8
Status after resection of brain metastases	3	2.5
Spine MRI*	63	52.1
Normal	53	43.8
LM and bone metastases	1	0.9
Suspicion of LM (leptomeningeal enhancement or nodules)	5	4.1
Breast cancer subtype		
ER-positive/HER2-negative	77	63.6
ER-positive/HER2-positive	10	8.3
Triple negative	14	11.6
ER-negative/HER2-positive	8	6.1
Unknown	12	9.9
Prior systemic therapy		
Yes	106	87.6
Endocrine therapy only	16	15.1
Chemotherapy only	15	14.2
Endocrine and chemotherapy	62	58.5
Endocrine, chemo and targeted therapy	8	7.6
Chemo and targeted therapy	5	4.7
No	15	12.4
Metastatic disease at time of LP		
Yes	81	66.9
No	40	33.1
Started Radiotherapy after LP		
Yes	26	21.5
Whole brain	17	14.0
Up to and including vertebra C2	15	13.4
Localized	7	5.8
Stereotactic	2	1.7
No	95	78.5
Started systemic therapy after LP <6 months		
Yes	58	47.9
No	58	47.9
Unknown	5	4.1
Median OS in years (IQR)	1.78 (0.42-11.7)	

* The number of findings exceeds the number of patients who underwent an MRI because some patients have multiple findings on their MRI

† Values are median (Inter quartile range)

$ Metastatic disease was defined as either extra-cranial and/or brain metastases at time of CSF collection

Abbreviations: cerebrospinal fluid (CSF); estrogen receptor (ER); human epidermal growth factor receptor 2 (HER2); lumbar puncture (LP); magnetic resonance imaging (MRI)

As CSF cytology is the gold standard to diagnose LM, we compared clinico-pathological variables between patients with a positive and negative CSF cytology result (**Supplementary Table S1a**). For the comparison between positive versus negative cytology the two equivocal samples were excluded. The only clinical sign which was more frequently observed in the positive cytology group was meningeal rigidity. All 13 patients with a positive cytology were already diagnosed with metastatic disease at the time of CSF collection and 6 of these patients were previously diagnosed with brain metastases. Regarding routinely performed CSF analyses, the CSF leukocyte concentration was significantly higher in patients with positive cytology. The glucose concentration was significantly lower in patients with positive cytology. No differences in CSF protein concentration and total cfDNA concentration were observed between cytology positive and negative samples. There was no difference in breast cancer subtype between the cytology positive versus negative group.

Mutation and mFAST-SeqS analyses

To study the concordance between CSF cytology and mutations in CSF, we performed targeted NGS on a subset of 30 patient samples, of which 9 had a positive, 1 an equivocal and 20 a negative cytology result. We found that the molecular coverage (i.e., the number of uniquely sequenced molecules) was significantly correlated with the amount of DNA available for sequencing (Spearman's rho: 0.68, $P<0.001$). When <10 ng was used, the molecular coverage was below 500 molecules in 73% of samples (**Supplementary Figure S2**). As the median total CSF-derived cfDNA yield in our cohort was only 8.72 ng, the majority of samples (69/121) had too low DNA yield for reliable NGS analysis with our currently used method. Sufficient molecular coverage was obtained for 8 out of 30 samples, in 4 of which hotspot mutations were detected (**Supplementary Table S2**). All 4 samples in which mutations were detected were reported as cytology positive.

From the total cohort of 121 patients, the mFAST-SeqS method yielded sufficient number of reads (>90,000) for 114 samples from 112 patients allowing for reliable determination of the aneuploidy status. For two patients, 2 sequential CSF samples were available of which the first was included for the cohort-wide analyses (**Supplementary Table S3**). Aneuploidy (mFAST-SeqS z-score ≥5) was observed in 10 out of 13 (76.9%) samples that were cytology positive, which was significantly more often than in the cytology negative group (9%) ($P<0.001$) (**Figure 2, Supplementary Table S1b**). Three patients with a positive cytology did not show genome-wide aneuploidy according to the threshold of ≥5 we employed, indicating a false-negative rate of 23.1%. Patients with aneuploidy had more frequently gait difficulties, cranial nerve palsies, lymph node metastasis and a higher CSF protein concentration (**Supplementary Table S1b**). The

total cfDNA concentration was not different between either patients with and without aneuploidy or patients with and without abnormal cytology (**Supplementary Table S1a/b**), suggesting that specific detection of a tumor-specific signal within the total pool of cfDNA is more informative.

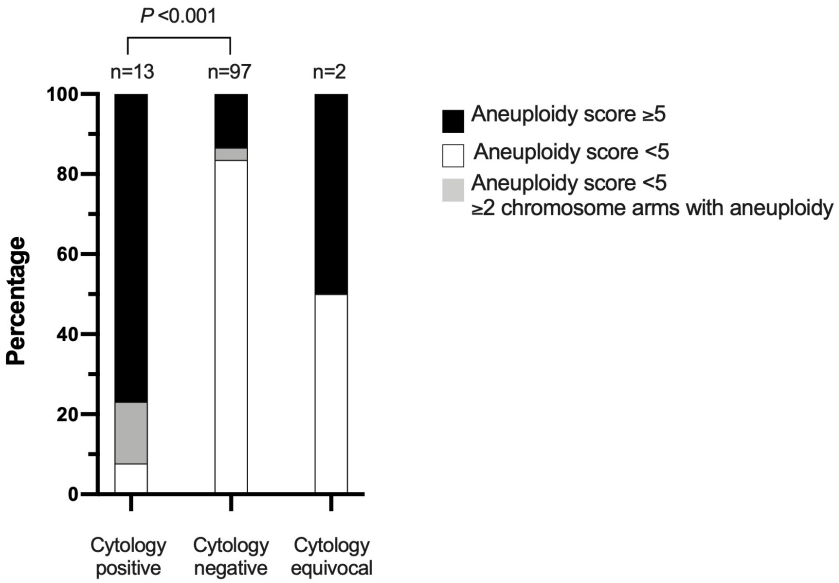


Figure 2 - Aneuploidy-score versus cytology result. The aneuploidy-score (<5 *versus* ≥5) was significantly more often ≥5 in patients with a cytology positive results of the same CSF sample.

Besides a genome-wide aneuploidy score, the mFAST-SeqS also provides an aneuploidy score per chromosome arm. From all 88 patients without genome-wide aneuploidy, 6 patients did show alterations in two or more individual chromosome arms. In patients with aneuploidy of ≥2 chromosome arms, 2 patients were cytology-positive, 2 patients had intracranial metastases (dural/brain metastasis) and 2 others were not diagnosed with CNS metastases (**Supplementary Table S4**). Hence, using a threshold for aneuploidy when ≥2 chromosome arms are aneuploid decreases the false-negative rate to 7.7%, whereas none of the healthy control samples (n=12) showed ≥2 aneuploid chromosome arms in a leave-one-out analysis.

Fourteen cytology-negative patients were scored as having aneuploidy. Of these 14 patients, four patients were diagnosed with LM at time of CSF collection (n=1) or immediately afterwards (n=3) either on MRI or at second LP. Three patients were not diagnosed with LM on imaging but had dural metastases at CSF collection, of which

one patient developed cytology-proven LM 619 days after initial LP. Finally, one patient developed parenchymal brain metastases after 188 days (**Table 2**). The remaining 6 patients had a high aneuploidy score but no final diagnosis of CNS metastases, and therefore represent potential false-positives (5.4% of all patients; 25% of all patients with aneuploidy). However, four patients died soon after CSF collection (after 43, 48, 51 and 83 days) whereas the other two patients were still alive at time of analyses (4.8 years and 14.5 years of follow-up, respectively). Unfortunately no autopsies were performed to determine the cause of death and exclude CNS involvement in these patients, but based on the clinical data case 119 had near certain LM but a negative cytology. An overview of clinical symptoms, CSF chemistry, imaging results, extra-cranial metastatic localizations and OS of the 14 patients with genome-wide aneuploidy but negative CSF cytology is provided in **Supplementary Table S5**. If we were to use the cut off of ≥2 aneuploid chromosome arms instead of using the cut off of a genome-wide Z-score ≥5, 2 additional patients without a final diagnosis of CNS metastases become positive, which would increase the potential false positive rate to 7.1%.

Table 2 - mFAST-SeqS z-score versus CNS metastasis

	z-score ≥5 (n=24)		z-score <5 (n=88)	
	N	%	N	%
Diagnosis of CNS metastasis at time of LP				
LM (Cytology+) only	8	33.3	3	2.3
LM (Cytology+) and brain metastasis	2	8.3	0	0.0
LM (Radiology+) only	1	4.2	0	0.0
LM and brain metastasis (radiology)	0	0.0	1	1.4
Brain metastases	1	4.2	9*	10.2
Dural metastases	3	12.5	2	2.3
Status after resection or RT of brain metastasis at LP	0	0.0	2	6.8
Final CNS diagnosis				
LM only	9	29.2	6	6.8
LM and brain metastases	5	20.8	7	8.0
LM and dural metastases	1	4.2	0	0.0
Brain metastasis	1	4.2	11	12.5
Dural metastases	2	8.3	2	2.3
Status after resection/ RT of brain metastases at LP	0	0.0	2	4.5
No CNS metastasis at all	6	25	60	68.2
Median time to LM (days)[$]	0 (0-5)		69 (0-554)	
Median time to brain metastases (days)[$]	0 (0-38)		0 (0-832)	

* 4 patients underwent resection and/or RT and still had brain metastases

[$] Median (inter quartile range)

Abbreviations: central nerve system (CNS); leptomeningeal metastases (LM); lumbar puncture (LP); radiotherapy (RT)

Association with clinical outcome

With a median follow-up of 10.7 years and 30 patients still alive at the time of data analysis, the median OS of the entire cohort was 1.78 years (IQR: 0.42-11.7 years).

Next, we associated routine CSF chemistry results with OS. CSF leukocyte count was available for 119 patients and a leukocyte count above the upper limit of normal (>4 x 10^6/L) was associated with a greater hazard of death (HR: 1.78, 95% CI: 1.07-2.98, *P*=0.027). CSF protein levels were available for 120 patients; a CSF protein level above the upper limit of normal (>0.58 g/L) was associated with a greater hazard of death (HR: 2.43, 95% CI: 1.34-4.40, *P*=0.003). Glucose levels were available for all 121 patients, a glucose level below the lower limit of normal (<2.5 mmol/L) was associated with a greater hazard of death (HR: 10.34, 95% CI 3.49-30.62, *P*<0.001). Of note, only four patients had a glucose concentration <2.5mmol/L. CSF cytology results were available for all patients at time of CSF collection. To test whether the cytology result was associated with OS, we used the definition of CSF cytology positive as those samples in which malignant cells were reported by the pathologist, equivocal results were analyzed as cytology negative. A positive cytology at the time of CSF collection was associated with increased risk of death (HR: 5.38, 95% CI: 2.92-9.91, *P* <0.001). In addition to the above described routine CSF chemistry results the presence of aneuploidy (Z-score ≥5) and the presence of any metastases (extra-cranial and/or brain metastases) at time of CSF collection was associated with a greater hazard of death (HR: 3.41, 95% CI 2.07-5.61, *P* <0.001 and HR:12.86, 95% CI: 6.63-24.95, *P* <0.001, respectively). Other clinical parameters including age at CSF collection, ER and HER2 status of the primary tumor were not associated with OS.

In subsequent multivariable analysis, including all univariable significant variables, only CSF aneuploidy (**Figure 3A**) and the presence of any metastatic localization at time of CSF collection (**Figure 3B**) were significantly associated with a greater risk of death (HR 2.24, 95% CI: 1.13-4.43, *P*=0.021 and HR 12.79, 95% CI: 6.29-26.02, *P* <0.001, respectively) (**Table 3A; Supplementary Figure S3**).

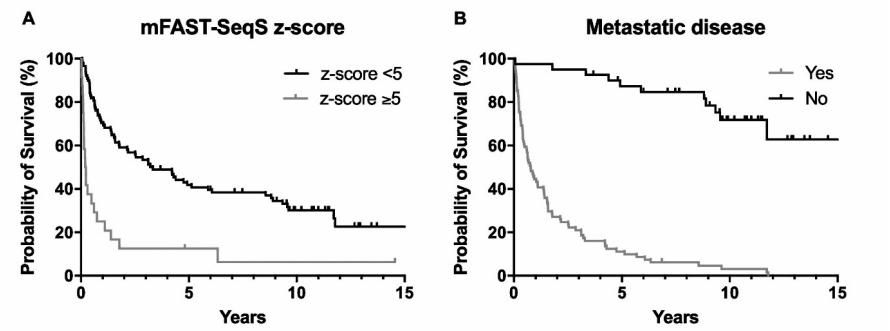


Figure 3 - Variables that were significantly associated with OS in the multivariable analysis. There was a significantly increased risk of death for patients with (A) mFAST-SeqS z-score ≥5, and (B) metastatic disease at time of CSF collection.

To investigate whether the presence of aneuploidy was associated with time to LM development and time to brain metastases development, we also performed Cox regression for only those patients that were not diagnosed with LM (n=97) or brain metastases (n=99) at the time of CSF collection, respectively. In univariable analysis for development of brain metastases only HER2 status and metastatic disease at CSF collection were significantly associated with a higher likelihood for brain metastases development. In the multivariable analyses, both variables still showed a significantly greater hazard of developing brain metastases (**Table 3B**).

Similarly, in univariable Cox regression with time to LM as dependent variable, metastatic disease at time of the LP (no vs. yes) and the aneuploidy-score (<5 vs. ≥5) were significantly associated with a higher likelihood of the development of LM. In multivariable analysis, both variables still showed a significantly greater hazard of developing LM (**Table 3C**).

Table 3 - Cox regression

A) Cox regression for OS

	Univariable analysis			Multivariable analysis		
Variable	**HR**	**95%CI**	***P*-value**	**HR**	**95%CI**	***P*-value**
CSF cfDNA concentration ng/mL CSF (low vs. High)	0.81	0.54-1.23	0.321			
CSF leukocyte count (normal vs. high)	1.78	1.07-2.98	**0.027**	1.41	0.76-2.62	0.276
CSF protein concentration (normal vs. high)	2.43	1.34-4.40	**0.003**	0.65	0.30-1.44	0,288
CSF glucose concentration (normal vs. low)	10.34	3.49-30.62	**<0.001**	3.38	0.96-11.89	0.057
CSF cytology (negative vs. positive)	5.38	2.92-9.91	**<0.001**	1.15	0.43-3.10	0.777
mFAST-SeqS z-score (<5 vs. >5)	3.41	2.07-5.61	**<0.001**	2.24	1.13-4.43	**0.021**
ER status (ER neg vs. ER pos)	1.18	0.67-2.21	0.566			
HER2-status (HER2 neg vs. HER2 pos)	0.73	0.36-1.46	0.369			
Age	1.00	0.99-1.02	0.703			
Metastatic disease at LP (no vs. yes)*	12.86	6.63-24.95	**<0.001**	12.79	6.29-26.02	**<0.001**

B) Only those cases that developed **brain metastases** after CSF collection (n=11 events)

	Univariable analysis			Multivariable analysis		
Variable	**HR**	**95%CI**	***P*-value**	**HR**	**95%CI**	***P*-value**
CSF cfDNA concentration ng/mL CSF (low vs. High)	0.40	0.11-1.49	0.170			
CSF leukocyte count (normal vs. high)	0.04	0-75.43	0.402			
CSF protein concentration (normal vs. high)	1.24	0.16-9.75	0.838			
mFAST-SeqS Z-score (<5 vs. >5)	3.76	0.96-14.75	0.058			
ER-status (ER neg vs. ER pos)	1.17	0.25-5.41	0.843			
HER2-status (HER2 neg vs. HER2 pos)	3.52	1.03-12.04	**0.045**	5.92	1.63-21.48	**0.007**
Age	0.99	0.94-1.04	0.705			
Metastatic disease at LP (no vs. yes)*	20.78	2.55-169.50	**0.005**	28.27	3.19-250.74	**0.003**

C) Only those cases that developed **LM** after CSF collection (n=13 events)

	Univariable analysis			Multivariable analysis		
Variable	**HR**	**95%CI**	***P*-value**	**HR**	**95%CI**	***P*-value**
CSF cfDNA concentration ng/mL CSF (low vs. high)	0.70	0.23-2.15	0.534			
CSF leukocyte count (normal vs. high)	1.93	0.53-7.10	0.321			
CSF protein concentration (normal vs. high)	1.36	0.18-10.48	0.771			
mFAST-SeqS Z-score (<5 vs. >5)	5.26	1.57-17.63	**0.007**	4.88	1.38-17.19	**0.014**
ER-status (ER neg vs. ER pos)	0.78	0.21-2.88	0.709			
HER2-status (HER2 neg vs. HER2 pos)	1.73	0.47-6.40	0.411			
Age	0.96	0.91-1.00	0.072			
Metastatic disease at LP (no vs. yes)*	8.52	1.75-41.55	**0.008**	8.51	1.69-42.87	**0.009**

* Metastatic disease was defined as either extra-cranial and/or brain metastases at time of CSF collection
Abbreviations: cerebrospinal fluid (CSF); cell-free DNA (cfDNA); confidence interval (CI); estrogen receptor (ER); hazard ratio (HR); human epidermal growth factor receptor 2 (HER2); lumbar puncture (LP)

Discussion

Plasma cfDNA analyses are increasingly implemented in routine diagnostics in metastatic cancer patients and more recently cfDNA analyses of CSF have sparked interest to characterize molecular aberrations of primary brain tumors and brain metastases. Our study comprises the largest breast cancer cohort in which CSF-derived cfDNA analyses have been performed. Here, we show that aneuploidy, measured by the mFAST-SeqS method 1) identifies 77% of patients with cytologically proven LM, 2) identifies a subgroup of patients with CNS metastases prior to routine diagnostics and 3) has prognostic value.

Specifically, we established for the first time the relation between aneuploidy detection in CSF-derived cfDNA and OS and showed that the previously established genome-wide z-score of ≥5 – developed to select plasma samples with high tumor fractions (>5-10%)[27] – yields prognostic value for patients with breast cancer. Importantly, as tumor DNA can be derived from brain metastases and LM, CSF-derived cfDNA analyses will not necessarily discriminate between these two conditions. Notably, a mFAST-SeqS z-score of ≥5 was associated with developing LM but not associated with developing brain metastases. Although the number of events in both Cox models was limited, it is likely that tumor DNA derived from LM is more abundantly present in CSF than tumor DNA derived from parenchymal brain metastases. More data, especially from negative control samples and EANO-ESMO confirmed LM cases, is necessary to determine the optimal diagnostic cut off for aneuploidy in this specific setting. Using the current threshold of genome-wide z-score ≥5 for the definition of 'aberrant', we missed three cases which were cytology positive. Two of the three missed cases had alterations on multiple chromosome arms, which in our cohort only occurred in a total of six patients, suggesting a lower threshold or other way of scoring aneuploidy could decrease the false

negativity rate without greatly affecting the false positivity rate. The third patient with a false negative mFAST-SeqS result had no alterations on any of the single chromosome arms. The initial cytology report of this CSF sample mentioned "no malignant cells" but at second examination "low cellular CSF with two atypical cells compatible with adenocarcinoma", which might be below the limit of detection of the mFAST-SeqS method. An alternative explanation for this false negative sample, might be a relatively copy number neutral breast cancer, which will also be missed by this method.

On the other hand, 14 cytology-negative patients were positive for aneuploidy at the current cut-off. Four of these cytology-negative patients were actually diagnosed with LM at time of CSF collection or immediately after the initial LP by imaging or a second CSF assessment. This indicates that these initial cytology results should be considered as false-negatives and demonstrates the potential additive value of mFAST-SeqS to conventional cytology. We identified only 6 patients with CSF aneuploidy without a final diagnosis of LM or brain metastases. It is possible that due to wide-spread metastatic disease in four of these patients tumor DNA from the blood has diffused over the blood-brain barrier. Hence, comparative studies between plasma and CSF obtained at the same time from the same patient are needed to elucidate whether the same chromosomal alterations are detected in blood and CSF.

Although targeted UMI-based NGS approaches are known to enable much more sensitive ctDNA detection compared to mFAST-SeqS (down to 0.2% *versus* 5-10%[27]), due to low cfDNA amounts the majority of our samples yielded a molecular coverage that was too low for reliable detection of mutations. Based on our findings, we believe that currently panel-based sequencing with Oncomine™ Breast cfDNA Assay v2 only provides an option for those samples with sufficient DNA yield, which only can be determined after cfDNA isolation. For low cfDNA yielding CSF samples, singleplex digital PCR (dPCR) assays can be performed, which was recently shown by van Bussel *et al*.[16] for *EGFR* mutations in patients with non-small cell lung cancer. However, in contrast to melanoma and non-small cell lung cancer, breast cancer mutation analyses of the primary tumor or metastatic lesion are not routinely performed because until recently no targeted treatments were available requiring knowledge of the tumors' mutational profile. More importantly, the mutational profile of breast cancer is quite heterogeneous[33], requiring a broad targeted panel or ideally whole exome or whole genome sequencing of tumor tissue followed by a patient specific dPCR for CSF analyses. Hence, a more general approach aiming at detection of virtually universal cancerous alterations, such as provided by the mFAST-SeqS method, seems a more attractive option, requiring low amounts of DNA input and no upfront knowledge of the genetic make-up of the tumor.

However, for samples with sufficient cfDNA yield, i.e., above 10 ng, targeted NGS is feasible as shown by the four cytology positive samples in which we detected hotspot mutations and is more sensitive than mFAST-SeqS for the detection of tumor-derived cfDNA.

Although our study represents one of the largest cohorts of CSF-derived cfDNA analyses, due to the retrospective nature of the study the clinical data collection is suboptimal and dependent on the extent of information that has been captured in the medical record by the treating physician. Moreover, in this proof-of-concept study, we used CSF that has been stored at -80°C for many years, so we cannot exclude that long-time storage might have influenced DNA quality and quantity compared to freshly obtained CSF.

Future studies employing the mFAST-SeqS method in patients suspected of LM, should ideally be prospectively conducted and focus on determination of the optimal cutoff for aneuploidy in a sample based on a larger series of negative control samples and true positive cases. Moreover, standardized clinical, pathological and radiological assessments, according to the EANO guidelines, should be performed to investigate to what extent the mFAST-SeqS method complements the current diagnostic armamentarium to diagnose LM in a patient's CSF.

In conclusion, aneuploidy as measured by the mFAST-SeqS method provides a robust and affordable technique to detect tumor-derived DNA in the CSF of patients with CNS metastases from breast cancer. The detected aneuploidy is associated specifically with development of LM and OS, but not with development of brain metastases. Future prospective trials investigating LM should employ this method in combination with other promising techniques such as EpCAM-based tumor cell detection assays to improve detection of LM in CSF of advanced cancer patients. Ultimately, the combination of standardized clinical symptoms and neuro-imaging scoring together with sensitive detection of tumor-derived material in the CSF will improve LM diagnosis.

Acknowledgements
We thank Charlotte C.J. Rimmelzwaan and Laura Pasquet for their help with the CSF cfDNA isolations and cfDNA mFAST-SeqS experiments.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

E.H. has an unrelated sponsored research agreement with Servier within CANCER-ID, a project funded by the Innovative Medicines Joint Undertaking (IMI JU), E.H. receives funding from Freenome, South San Francisco, CA and PreAnalytiX, Hombrechtikon, Switzerland. E.H. received honoraria from Roche for advisory boards.

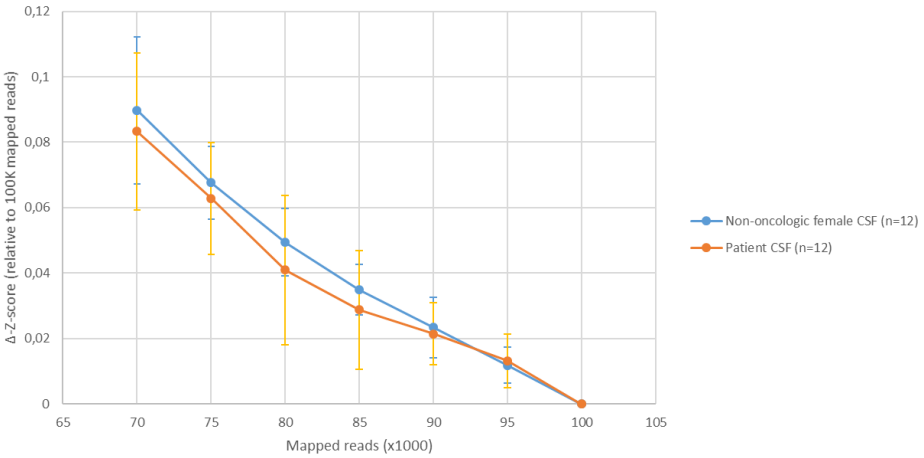
The other authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

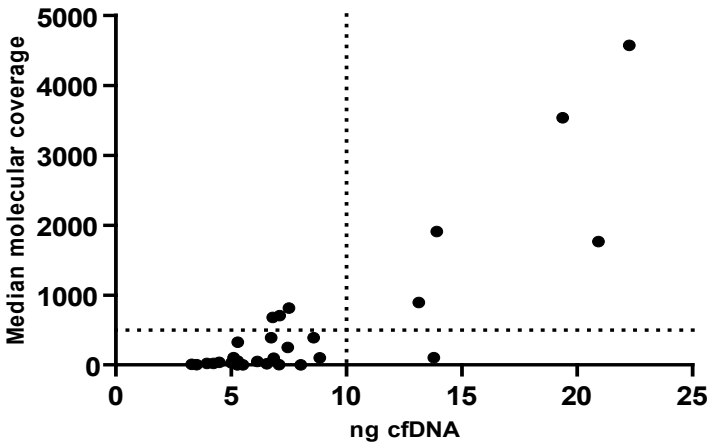
1. Harris, M. *et al.* A comparison of the metastatic pattern of infiltrating lobular carcinoma and infiltrating duct carcinoma of the breast. *Br J Cancer* **50**, 23-30 (1984).
2. Lamovec, J. & Bracko, M. Metastatic pattern of infiltrating lobular carcinoma of the breast: an autopsy study. *J Surg Oncol* **48**, 28-33 (1991).
3. Little, J.R. *et al.* Meningeal carcinomatosis. Clinical manifestations. *Arch Neurol* **30**, 138-43 (1974).
4. Le Rhun, E. *et al.* A retrospective case series of 103 consecutive patients with leptomeningeal metastasis and breast cancer. *J Neurooncol* **113**, 83-92 (2013).
5. Niwinska, A. *et al.* Breast cancer leptomeningeal metastasis: propensity of breast cancer subtypes for leptomeninges and the analysis of factors influencing survival. *Med Oncol* **30**, 408 (2013).
6. Gauthier, H. *et al.* Survival of breast cancer patients with meningeal carcinomatosis. *Ann Oncol* **21**, 2183-7 (2010).
7. de Azevedo, C.R. *et al.* Meningeal carcinomatosis in breast cancer: prognostic factors and outcome. *J Neurooncol* **104**, 565-72 (2011).
8. Le Rhun, E. *et al.* EANO-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up of patients with leptomeningeal metastasis from solid tumours. *Ann Oncol* **28**, iv84-iv99 (2017).
9. Clarke, J.L. *et al.* Leptomeningeal metastases in the MRI era. *Neurology* **74**, 1449-54 (2010).
10. Straathof, C.S. *et al.* The diagnostic accuracy of magnetic resonance imaging and cerebrospinal fluid cytology in leptomeningeal metastasis. *J Neurol* **246**, 810-4 (1999).
11. Olson, M.E. *et al.* Infiltration of the leptomeninges by systemic cancer. A clinical and pathologic study. *Arch Neurol* **30**, 122-37 (1974).
12. Glass, J.P. *et al.* Malignant cells in cerebrospinal fluid (CSF): the meaning of a positive CSF cytology. *Neurology* **29**, 1369-75 (1979).
13. Balm, M. & Hammack, J. Leptomeningeal carcinomatosis. Presenting features and prognostic factors. *Arch Neurol* **53**, 626-32 (1996).
14. Wasserstrom, W.R. *et al.* Diagnosis and treatment of leptomeningeal metastases from solid tumors: experience with 90 patients. *Cancer* **49**, 759-72 (1982).
15. Angus, L. *et al.* Novel methods to diagnose leptomeningeal metastases in breast cancer. *Neuro Oncol* **21**, 428-439 (2019).
16. van Bussel, M.T.J. *et al.* Circulating epithelial tumor cell analysis in CSF in patients with leptomeningeal metastases. *Neurology* **94**, e521-e528 (2020).
17. Lee, J.S. *et al.* Detection of cerebrospinal fluid tumor cells and its clinical relevance in leptomeningeal metastasis of breast cancer. *Breast Cancer Res Treat* **154**, 339-49 (2015).
18. Nayak, L. *et al.* Rare cell capture technology for the diagnosis of leptomeningeal metastasis in solid tumors. *Neurology* **80**, 1598-605; discussion 1603 (2013).
19. Milojkovic Kerklaan, B. *et al.* EpCAM-based flow cytometry in cerebrospinal fluid greatly improves diagnostic accuracy of leptomeningeal metastases from epithelial tumors. *Neuro Oncol* (2015).
20. Subira, D. *et al.* Role of flow cytometry immunophenotyping in the diagnosis of leptomeningeal carcinomatosis. *Neuro Oncol* **14**, 43-52 (2012).
21. Subira, D. *et al.* Diagnostic and prognostic significance of flow cytometry immunophenotyping in patients with leptomeningeal carcinomatosis. *Clin Exp Metastasis* **32**, 383-91 (2015).
22. Lin, X. *et al.* Cerebrospinal fluid circulating tumor cells: a novel tool to diagnose leptomeningeal metastases from epithelial tumors. *Neuro Oncol* **19**, 1248-1254 (2017).
23. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, 8839 (2015).
24. Diaz, L.A. *et al.* Liquid biopsies: genotyping circulating tumor DNA. *J Clin Oncol* **32**, 579-86 (2014).
25. Li, Y.S. *et al.* Unique genetic profiles from cerebrospinal fluid cell-free DNA in leptomeningeal metastases of EGFR-mutant non-small cell lung cancer: a new medium of liquid biopsy. *Ann Oncol* (2018).

26. Pentsova, E.I. *et al.* Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. *J Clin Oncol* **34**, 2404-15 (2016).
27. Belic, J. *et al.* Rapid identification of plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. *Clin Chem* **61**, 838-49 (2015).
28. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
29. FEDERA. Human Tissue and Medical Research: Code of Conduct for responsible use (2011).
30. Jansen, M.P. *et al.* Cell-free DNA mutations as biomarkers in breast cancer patients receiving tamoxifen. *Oncotarget* **7**, 43412-43418 (2016).
31. Kinde, I. *et al.* FAST-SeqS: a simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLoS One* **7**, e41162 (2012).
32. Suppan, C. *et al.* Untargeted assessment of tumor fractions in plasma for monitoring and prognostication from metastatic breast cancer patients undergoing systemic treatment. *Cancers (Basel)* **11**(2019).
33. Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* (2019).

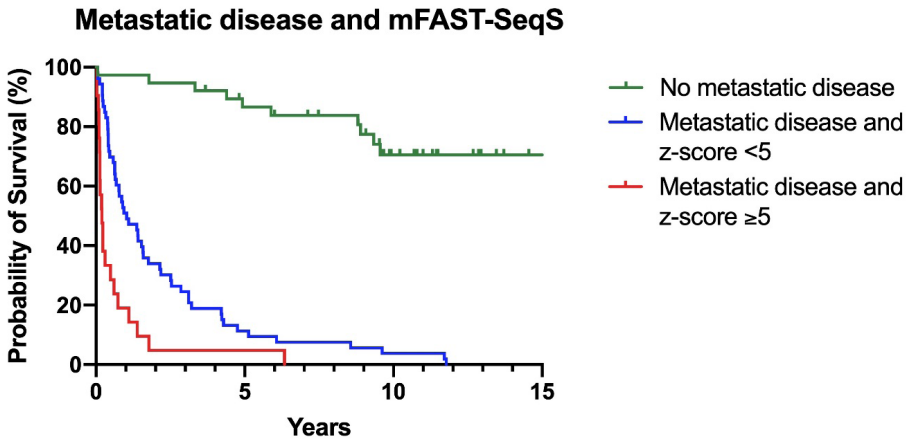
Supplemental Figures



Supplementary Figure 1 - The delta genome-wide Z score increases with lower numbers of mapped reads and diverges for controls (healthy donor liquor) and cases (patient liquor) below 90,000 reads. In addition, a higher standard deviation over the samples is observed below 90,000 reads.



Supplementary Figure S2 - Correlation between input ng cfDNA in the Oncomine Breast V2 breast panel and the median molecular coverage (i.e., the median number of uniquely sequenced molecules per sample). Spearman's rho correlation coefficient: 0.684, *P* <0.001 (two-tailed). Dashed lines indicate thresholds of median of 500 sequenced molecules and cfDNA input of 10 ng.



Supplementary Figure S3 - Combination of mFAST-SeqS z-score and metastatic disease at time of CSF collection.

Supplemental Tables

Supplementary Table S1

A) Cytology positive versus negative

	Cytology positive (n=13)		Cytology negative (n=106)		
	N	%	N	%	*P*-value
Age (median)	56 (42.5-77.0)		54.5 (45.0-63.0)		0.399
Breast cancer subtype					0.494
ER-positive/HER2-negative	10	76.9	58	59.8	
ER-positive/HER2-positive	0	0.00	10	10.3	
Triple negative	2	15.4	10	10,3	
ER-negative/HER2-positive	1	7.7	7	7.2	
Unknown	0	0	12	12.4	
Clinical signs					
Headache	7	53.8	41	38.7	0.506
Nausea/vomiting	5	38.5	31	29.2	0.631
Mental changes	3	23.1	13	12.3	0.567
Gait difficulties	4	30.8	18	17.0	0.483
Meningeal rigidity	2	15.4	1	0.9	**0.015**
Cranial nerve palsies	5	38.5	22	20.8	0.187
Spinal symptoms	10	76.9	64	60.4	0.107
Metastatic disease at time of LP					
Yes	13	100.0	66	62.3	**0.004**
Bone	10	76.9	54	50.9	0.232
Liver	4	30.8	22	20.8	0.157
Lymph node	4	30.8	33	31.1	0.294
Lung/pleural	2	15.4	22	20.8	1.000
(sub)cutaneous	2	15.4	9	8.5	0.419
Gastrointestinal	1	7.7	2	1.9	0.374
Gynaecological	0	0.0	1	0.9	1.000
Brain (or status after resection/radiotherapy)	6	46.2	8	7.5	**0.001**
Kidney/Adrenal	1	7.7	1	0.9	0.374
No	0	0.0	40	37.7	
CSF Chemistry†					
Leukocytes (x 10^6/L, normal = 0-4 x 106/L)	2.00 (1.50-16.15)		1.30 (1.00-3.00)		**0.044**
Protein (g/L, normal = 0.18-0.58 g/L)	0.48 (0.26-0.79)		0.32 (0.23-0.43)		0.114
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	3.30 (2.05-3.75)		3.50 (3.30-4.00)		**0.050**
Volume for cytology (mL)	5.5 (4.25-7.0)		5.0 (3.0-9.0)		0.639
CSF cfDNA†					
cfDNA concentration (ng/mL CSF)	3.92 (2.57-12.51)		5.32 (3.70-10.47)		0.263

B) mFAST-SeqS positive versus negative

	FastSeq ≥5 (n=24)		FastSeq <5 (n=88)		
	N	%	N	%	*P*-value
Age (median)	57.5 (42.0-64.8)		54.5 (44.3 - 64.5)		0.798
Breast cancer subtype					0.864
ER-positive/HER2-negative	14	58.3	55	62.5	
ER-positive/HER2-positive	2	8.3	8	9.1	
Triple negative	4	16.7	9	10.2	
ER-negative/HER2-positive	1	4.2	7	8	
Unknown	3	12.5	9	10.2	

B) - Continued

	FastSeq ≥5 (n=24)		FastSeq <5 (n=88)		
	N	%	N	%	*P*-value
Clinical signs					
Headache	9	37.5	34	38.6	1.000
Nausea/vomiting	10	41.7	22	25.0	0.230
Mental changes	3	12,5	12	13.6	1.000
Gait difficulties	9	37.5	13	14.7	**0.035**
Meningeal rigidity	2	8.3	1	1.1	0.098
Cranial nerve palsies	10	41.7	15	17.0	**0.034**
Spinal symptoms	20	83.3	52	59.1	**0.030**
Metastatic disease at time of LP					
Yes	21	87.5	53	60.2	**0.009**
Bone	17	70.8	43	48.9	0.129
Liver	6	25.0	17	19.3	0.331
Lymph node	13	54.1	18	20.5	**0.001**
Lung/pleural	7	29.1	15	17.0	0.387
(sub)cutaneous	3	12.5	9	10.2	0.780
Gastrointestinal	1	4.2	2	2.3	0.625
Gynecological	0	0.0	1	1.1	1.000
Kidney	1	4.2	1	1.1	0.625
Brain	6	25	8	9.1	0.074
No	3	12.5	35	39.8	
CSF Cytology					**<0.001**
Positive	10	41.7	3	3.4	
Equivocal	1	4.2	1	1.1	
Negative	13	54.2	84	95.5	
CSF Chemistry					
Leukocytes (x 10^6/L, normal = 0-4 x 106/L)	1.7 (1.0-6.5)		2.0 (1.0-3.0)		0.789
Protein (g/L, normal = 0.18-0.58 g/L)	0.46 (0.24-0.63)		0.32 (0.23-0.42)		**0.042**
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	3.5 (3.3-4.3)		3.4 (3.2-3.8)		0.736
CSF					
cfDNA concentration (ng/mL CSF)	5.5 (3.3-30.5)		5.0 (3.3-8.6)		0.239

† Values are median (Inter quartile range)

Abbreviations: cerebrospinal fluid (CSF); cell-free DNA (cfDNA); estrogen receptor (ER); human epidermal growth factor receptor 2 (HER2); lumbar puncture (LP); magnetic resonance imaging (MRI)

Supplementary Table S2 - Overview of mutation analyses of samples with sufficient (median molecular coverage ≥500) coverage. This table will be provided as separate excel file.

https://clincancerres.aacrjournals.org/highwire/filestream/188713/field_highwire_adjunct_files/0/253342_2_supp_6854620_qn5d44.xlsx

Supplementary Table S3 - Patient characteristics of patients with successful mFAST-SeqS result

	N=112	
	N	**%**
Age at LP		
Median	55 (45.0-64.8)	
Gender		
Female	112	100
CSF cytology		
Positive	13	13.0
Equivocal	2	2.2
Negative	97	84.8
CSF chemistry		
Leukocytes (x 10^6/L, normal = 0-4 x 10^6/L)	2.0 (1.0-3.1)	
Protein (g/L, normal = 0.18-0.58 g/L)	0.33 (0.23-0.47)	
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	3.5 (3.2-3.8)	
MRI brain*	70	62.5
Normal	36	32.1
LM only	1	0.9
Suspicion of LM (leptomeningeal enhancement)	7	6.3
LM and brain metastases	2	1.8
Brain metastases	10	8.9
Dural metastases	5	4.5
Brain metastases and status after RT or resection	2	1.8
Suspicion of brain metastases	1	0.9
Bone metastases	10	8.9
Spine MRI*	61	54.5
Normal	25	22.3
LM and bone metastases	1	0.9
Suspicion of LM (leptomeningeal enhancement or nodules)	5	4.5
Bone metastases	30	26.8
Breast cancer subtype		
ER-positive/HER2-negative	69	61.6
ER-positive/HER2-positive	10	8.9
Triple negative	13	11.6
ER-negative/HER2-positive	8	7.1
Unknown	12	10.7
Prior systemic therapy		
Yes	97	86.6
Endocrine therapy only	14	12.5
Chemotherapy only	14	12.5
Endocrine and chemotherapy	56	50.0
Endocrine, chemo and targeted therapy	8	7.1
Chemo and targeted therapy	5	4.5
No	15	13.4
Metastatic disease at time of LP		
Yes	74	66.1
No	38	33.9
Started Radiotherapy after LP		
Yes	26	23.2
Whole brain	17	15.2
Up to and including vertebra C2	15	13.4
Localized	7	6.3
Stereotactic	2	1.8
No	86	76.8
Started systemic therapy after LP <6 months		
Yes	53	47.3
No	54	48.2
Unknown	5	4.5
Median OS in years (IQR)	1.78 (0.41-11.7)	

Supplementary Table S4 - Samples with mFAST-SeqS z-score <5 having alterations on single chromosome arms

Patient no.	Genome-wide z-score	Loss	Gain	Final CNS diagnosis	Cytology
3	0.21		8q	No CNS diagnosis	Negative
13	1.42	5q		Radiological suspicion of LM	Negative
14	2.10		1q and 8q	LM (positive cytology) and status after resection/RT	Positive
24	2.09	11q	5q	No CNS diagnosis	Negative
56	2.43		1q and 5q	No CNS diagnosis	Negative
61	3.86	4p	12q and 3p	Dural metastases	Negative
65	2.15	2q	8q	Brain and dural metastasis at LP	Negative
66	2.81	5q	10q and 7q	LM (positive cytology)	Positive
85	2.12	5q		No CNS diagnosis	Negative
116	2.46		5q	No CNS diagnosis	Negative

Supplementary Table S5 - CSF samples with mFAST-SeqS z-score ≥5 but with negative CSF cytology

	No final diagnosis of CNS metastasis					
Patient no.	**54**	**59**	**60**	**119**	**9**	**82**
z-score	22.72	40.97	7.13	16.98	8.28	7.55
Clinical signs						
Headache	No	Yes	No	Yes	No	Yes
Nausea/vomiting	No	Yes	Yes	Yes	No	No
Mental changes	No	No	Yes	Yes	No	No
Walking difficulties	No	No	Yes	No	Yes	No
Meningeal rigidity	No	No	No	No	No	No
Cranial nerve palsies	No	Yes	Yes	No	No	No
	-	Trigeminal nerve dysfunction; Hypoglossal palsy	Dysarthria	-	-	-
Spinal symptoms	Yes	No	Yes	Yes	Yes	Yes
	Limb weakness; neck/ back pain; sensory symptoms (limbs)		Limb weakness; bladder/bowel dysfunction; sensory symptoms (limbs)	Radiating limb pain; neck/back pain	Limb weakness; radiating limb pain; neck/back pain	Radiating limb pain; neck/ back pain
OS (days)	48	43	51	83	1756 (still alive)	5313 (still alive)
Cytology	Negative	Equivocal	Negative	Negative	Negative	Negative
CSF chemistry						
Leukocytes (x 10^6/L, normal = 0-4 x 106/L)	0	7	0.3	23	1	1
Protein (g/L, normal = 0.18-0.58 g/L)	0.13	0.18	0.49	1.18	0.22	0.49
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	4.4	4.4	3.3	4.4	3.3	3.4
Imaging	MRI brain: normal	CT brain: normal	MRI brain: normal	MRI brain: ossal metastases skull with enhancement of dura (probably in continuated with bone metastases)	MRI spine: normal	MRI Spine: Linear enhancement suggestive for LM but not conclusive

Supplementary Table S5 - Continued

	No final diagnosis of CNS metastasis					
Patient no.	**54**	**59**	**60**	**119**	**9**	**82**
Extra-cranial metastatic sites	Lymph nodes; lung; lymphangitis carcinomatosa	Bone; lymph nodes	Bone; liver; lymph node; lung	Bone; lymph nodes; lung	No	No
Other remarks / Final CNS diagnosis	*/No CSN disease at all	SPECT: enhanced uptake skull base; received radiation of the skull base (1x8Gy)/Clinical suspicion of LM high	Clinically high suspicion of LM; but no second LP at request of patient; later MRI showed metastases of the skull base/Clinical suspicion of LM high	Radicular syndroom S1 and developed polyradiculopathy. Clinical suspicion of LM high for which whole brain radiotherapy up to and including C2/ Clinical suspicion of LM high	Five years after initial LP progressive radiating backpain; MRI spine shows enhacement of cauda equina but complicated by movement artefacts. Conclusion: no sure LM diagnosis/ Radiographic suspicion of LM	*/No CNS disease at all

	Final diagnosis of CNS metastasis							
Patient no.	**52**	**69**	**106**	**107**	**109**	**111**	**113**	**117**
z-score	31.01	16.27	37.53	7.50	12.84	6.76	5.79	51.61
Clinical signs								
Headache	Yes	No	Yes	No	No	No	No	No
Nausea/vomiting	Yes	No	Yes	Yes	No	No	No	Yes
Mental changes	No	No	No	No	No	Yes	No	No
Walking difficulties	Yes	No	No	No	No	Yes	Yes	Yes
Meningeal rigidity	No	No	No	No	No	No	No	No
Cranial nerve palsies	Yes	Yes	No	No	Yes	No	No	No
	Visual disturbances	Trigeminal nerve dysfunction	-	-	Trigeminal nerve dysfunction	-	-	-
Spinal symptoms	No	No	Yes	Yes	Yes	Unknown	Yes	Yes
	-	-	Limb weakness	Limb weakness; radiating limb pain	Sensory symptoms (limbs)	-	Back pain; sensory symptoms (limbs)	Limb weakness; bladder/bowel dysfunction; sensory symptomes (limbs)

Supplementary Table S5 - Continued

	Final diagnosis of CNS metastasis							
Patient no.	**52**	**69**	**106**	**107**	**109**	**111**	**113**	**117**
OS (days)	221	649	39	178	269	19	505	28
Cytology	Negative	Negative	Negative	Negative	Negative	Negative	Negative	Negative
CSF chemistry								
Leukocytes (x 10^6/L, normal = 0-4 x 106/L)	19	1	1.7	1	0.7	0	1	1.7
Protein (g/L, normal = 0.18-0.58 g/L)	0.56	0.37	0.44	0.36	0.37	0.65	0.26	0.71
Glucose (mmol/L, normal = 2.5-3.7 mmol/L)	3.5	3.7	4.6	4.3	3.5	4.4	3.4	3.9
Imaging	CT brain: only brain metastasis	MRI brain: bone and dural metastasis	MRI brain: normal	MRI brain: right frontal dural enhancement; MRI spine: bone metastasis	MRI brain: dural enhancement alongside the left hemisphere	MRI spine: pathological leptomeningeal enhancement along the ventral and dorsal spinal cord (linear + nodular), mainly thoracally located	MRI spine: bone metastasis	MRI spine: bone metastasis
Extra-cranial metastatic sites	No	Bone; lymh node; subcutaneous	Bone; liver; lymph node; lung	Bone; liver; lymph node	Bone	No	Bone; lymph node; lung	Bone; lymph node
Other remarks / Final CNS diagnosis	MRI performed 15 days after LP showed leptomeningeal enhancement besides metastasis in the vermis/ Brain metastasis at time of LP; LM diagnosed after LP	*/Dural metastasis at time of LP; LM diagnosed after LP	*/LM diagnosed at second LP (619 days after first LP)	* Dural metastasis only	*/Dural metastasis only	*/LM at LP (radiology proven)	*/Brain metastasis diagnosed after LP	*/Brain metastasis and LM diagnosed after LP

CHAPTER 10

Expression levels of TSPAN13, CD44, MAGEA3 and PLAU in circulating tumor cells are associated with brain metastases development in patients with metastatic breast cancer

In preparation

Lindsay Angus, Agnes Jager, Anieta M. Sieuwerts†, John W.M. Martens, Stefan Sleijfer, Saskia M. Wilting

† Deceased August 2019

Abstract

Background

Brain metastases represent a devastating complication in metastatic breast cancer patients. Given the huge clinical consequences, tools for the identification of breast cancer patients at risk for developing brain metastases are urgently needed to select those patients that will benefit from routine screening to reduce the morbidity associated with brain metastases by earlier detection. Here, we aimed to identify a gene expression profile established in circulating tumor cells (CTCs) associated with the occurrence of brain metastases.

Methods

132 breast cancer patients starting with first line systemic therapy for metastatic disease were included. Prior to treatment blood was drawn for CTC enumeration and gene expression profiling. Time to brain metastasis was defined as time between blood sampling and the date of the first CT or MRI on which brain metastases were detected. Cox proportional hazards analyses were performed to investigate the potential associations between time to brain metastases, clinico-pathological parameters, and CTC gene expression profiles which were previously generated from CTC-enriched fractions. Subsequently, a prediction model was built based on the significant variables in the multivariable Cox regression to predict the development of brain metastases within 2 years after blood draw.

Results

Multivariable Cox analysis showed that ER-negative status, younger age, lower gene expression of *TSPAN13* and *CD44,* and higher gene expression of *MAGEA3* and *PLAU* were associated with a significantly greater hazard of developing brain metastases. A logistic ridge regression model for brain metastases development within 2 years after blood draw containing both the significant clinico-pathological and gene expression variables, resulted in an AUC of 0.77 (95% CI 0.67-0.88), which was significantly better than using only the significant clinico-pathological variables (AUC of 0.68, P=0.021).

Conclusions

Gene expression profiles in CTCs and clinico-pathological variables are associated with the development of brain metastases in breast cancer patients, irrespective of CTC count. Our encouraging results warrant further research into the most optimal predictive marker panel for future selection of metastatic breast cancer patients at risk to develop for brain metastases.

Introduction

Brain metastases represent a devastating complication of solid tumors and, after lung cancer, breast cancer is the second most common cancer associated with brain metastases[1]. During the past decades the incidence of brain metastases in breast cancer patients has increased[2,3]. This is likely the result of higher success rates of systemic treatments resulting in more patients achieving long-term survival allowing brain metastases to develop. Once brain metastases appear, quality of life often declines due to progressive neurologic impairments[4,5]. The risk of developing brain metastases is subtype-dependent[6,7]. Among patients who develop metastatic disease, patients with estrogen receptor (ER)-negative/HER2-positive disease develop brain metastases most frequently (11.5-35%)[6-8], followed by ER-positive/HER2-positive (8-15.4%)[6-8], triple negative (7.2-11.4%)[6,7], and ER-positive/HER2-negative disease (5.5-7.6%)[6,7].

The median overall survival of patients who are diagnosed with brain metastases varies widely, ranging from 4.6 to 22 months, and is also subtype dependent, in which patients with triple negative breast cancer have the worst prognosis (median overall survival (OS); range 4.6 - 6 months[6,9,10]) and patients with ER-positive/HER2-positive patients harbor a more favorable prognosis (median OS; range 16.5 - 22 months[6,9,10]).

According to the ESMO guidelines for advanced breast cancer[11] and ASCO guidelines for patients with advanced HER2-positive breast cancer and brain metastases[12], brain imaging is currently not recommended for asymptomatic patients. This approach differs from recommendations for other tumor types such as non-small cell lung cancer (NSCLC) (depending on the guideline; stage II[13] or III-IV[14]), small cell lung cancer (all stages)[15] and melanoma (stage III-IV)[16]. Hence, most brain metastases in patients with breast cancer are detected because of neurological symptoms, often requiring local interventions such as neurosurgical resection and/or radiotherapy (whole brain or stereotactic approaches)[17]. Notably, in a comparative study between patients with brain metastases from breast cancer and NSCLC, breast cancer patients had larger and more numerous metastases, which were also more likely to be symptomatic[17]. In addition, neurological deaths were more commonly observed in breast cancer patients with brain metastases (37.3%) than in NSCLC patients with brain metastases (19.9%)[17]. Since identification of patients at risk could lead to preventive approaches and detection of brain metastases in a more limited stage of disease that might allow for more localized treatments such as stereotactic radiotherapy, there are several prospective trials underway to evaluate the value of brain metastases screening in stage IV breast cancer (clinicaltrial.gov identifiers: NCT04030507; NCT03881605).

Risk factors that have been associated with the development of breast cancer brain metastases include a younger age (<50 years), high T-stage (≥T2), lymph node-positivity, grade 3 tumors, a higher number of extra-cranial metastatic sites, hormone receptor negativity and HER2-positivity[6,18]. Ultimately, markers for reliable prediction of the risk for brain metastases development will enable selection of high risk patients who might eventually benefit from screening and treatment with drugs capable of crossing the blood-brain-barrier (BBB).

Circulating tumor cells (CTCs) represent a minimally invasive means to characterize tumor cells and CTCs capable of initiating brain metastases might harbor different molecular characteristics than those that are not. To date, limited data are available about the molecular mechanisms that render CTCs capable of extravasation through the BBB. Bos *et al*. showed years ago that a set of 17 genes, measured in primary breast tumors, was associated with brain relapse, including *COX2* and *EGFR* ligands to prime the breast cancer cells for extravasation into the brain[19]. Recently, Klotz *et al.* established cell lines from CTCs obtained from four patients with breast cancer and showed that these cell lines exhibited tissue tropism by partly recapitulating human metastatic disease in mice[20]. More specifically, *ex-vivo* expanded CTCs that were isolated from a patient in which brain metastases appeared one year after CTC isolation showed the highest tropism for the brain in mice, providing proof of concept that intrinsic molecular features of cancer cells are at least partly associated with the development of brain metastases. More insight into these molecular features will aid in the identification of patients at high risk for developing these metastases.

In this study we used our previously described breast cancer expression panel[21] to evaluate the molecular characteristics of CTCs obtained from breast cancer patients starting their first line of systemic treatment for metastatic disease. The associations between the clinico-pathological features, gene expression profiles, and the development of brain metastases were determined to investigate whether these parameters are predictive for the development of brain metastases within two years and could be used for selection of high-risk patients.

Materials and Methods

Patients

Patients with metastatic breast cancer either starting with first line endocrine treatment or first line chemotherapy were selected from a prospective multicenter CTC trial (METC

2006-248[22]) (**Supplementary Figure 1**). In this study, 10 mL of blood was drawn for CTC enumeration and another 10 mL of blood for CTC gene expression profiling. According to standard of care brain imaging was performed when CNS metastases were suspected and the diagnosis of brain metastases was made using radiological imaging approaches such as computed tomography (CT) or magnetic resonance imaging (MRI). Since the subarachnoid space and brain parenchyma are recognized as different soils, it is likely that different mechanisms are involved in the development of either brain metastases or leptomeningeal metastases. Therefore, patients who only developed leptomeningeal metastases were excluded from our analysis. Patients with both parenchymal brain metastases and leptomeningeal metastases were included in the analysis. Three groups of patients were distinguished: 1) patients with brain metastases at the time of blood draw prior to start of first line therapy (baseline) (n=11); 2) patients who developed brain metastases within two years after the blood draw (n=22) and 3) patients who did not develop brain metastases or leptomeningeal metastases during at least two years of follow-up (n=99) (**Supplementary Figure 1**).

CTC enumeration, mRNA isolation from CTCs and gene expression analysis

A detailed description of CTC enumeration and CTC characterization has been reported previously[21,23]. In brief, for CTC enumeration 10 mL blood was drawn in CellSave tubes (Menarini-Silicon Biosystems, Huntington Valley, USA) of which 7.5mL was used for CTC enumeration within 96 hours after blood draw using the CellSearch System (Menarini-Silicon Biosystems). For CTC characterization, 10 mL of blood was drawn in Vacutainer® EDTA tubes (BD, Franklin Lakes, USA). Subsequently, CTCs were isolated from 7.5 mL within 24 hours after blood draw using the CellSearch profile kit (Menarini-Silicon Biosystems). After CTC enrichment, RNA was isolated from the CTC-enriched fraction using the AllPrep DNA/RNA Microkit (Qiagen, Germantown, MD, USA), cDNA generated, pre-amplified for the 93 targets of interest (**Supplementary Table S1**) and 3 reference genes (*GUSB, HMBS* and *HPRT1*), and real-time amplified using Taqman Gene Expression Assays (Applied Biosystems, Carlsbad, CA). All patients with CTC samples with sufficient PCR signal (average Cq<26.5 for reference genes *GUSB, HMBS* and *HPRT1*), were included. Expression levels were expressed as delta Cq (Cq target gene – average Cq reference genes).

Statistical Analysis

The primary objective of this study was to investigate whether gene expression measured in the CTC-enriched fraction was associated with the development of brain metastases within two years after the initial blood draw. Time to brain metastasis was defined as time between blood draw for CTC enumeration and gene expression

profiling and the date of the first CT or MRI on which brain metastases were detected. For patients with brain metastasis at baseline, time to brain metastases was set to zero. Univariable and multivariable Cox proportional hazards analyses were performed to determine the associations between clinico-pathological parameters, CTC gene expression profiles and time to brain metastases. For all significant variables in the multivariable Cox proportional hazard analyses, Kaplan-Meier plots were generated for visualization. Continuous variables were split at the median or in case no gene expression was measured in more than 50% of the samples (*MAGEA3*), gene expression was scored as present or absent for plotting purposes.

Furthermore, a prediction model was built using all significant variables in the multivariable Cox proportional hazard analysis. Logistic ridge regression was carried out to predict 1) all patients that developed brain metastasis (baseline + within 2 years of follow-up) and 2) only patients who developed brain metastasis during 2 years of follow-up. The performance of the ridge model was visualized by a receiver operating curve (ROC), obtained by 10-fold cross-validation, and quantified by the area under the curve (AUC). To compare the ROCs between models including both clinico-pathological and gene expression versus clinico-pathological variables only, we used the DeLong's test[24] for two correlated ROC curves. In addition, we calculated Youden's index[25] to determine the most optimal combination of sensitivity and specificity for our model.

All statistical analyses were performed in R statistical program version 3.6.2. The survival package (v 3.1.11) for Cox proportional hazards analyses, and the grridge package (release 3.10) for Ridge logistic regression analysis were used[26]. All statistical tests were two-sided and considered statistically significant at $P < 0.05$.

Results

Patient characteristics and brain metastases incidence

At baseline, 11 (8.3%) patients were diagnosed with brain metastases. During two years of follow up, 22 (16.7%) additional patients developed brain metastases, while 99 (75%) patients were not diagnosed with brain metastases within this timeframe (**Table 1**). Among the 132 patients that were included from December 2008 until August 2014, 56.8% were ER-positive/HER2-negative, 25% ER-positive/HER2-positive, 6.8% ER-negative/HER2-positive and 11.4% triple negative. Sixty-two patients (47%) started with first line chemotherapy after the blood draw, while 70 (53%) patients started with first line endocrine therapy. Before start of first line treatment, 97 (73%) patients

had ≥1 CTC/7.5mL blood, with a median of 4 CTCs/7.5mL (interquartile range (IQR) 0-28). Among the 11 patients who had brain metastases at baseline, 3 (27.3%) were ER-positive/HER2-negative, 4 (36.4%) ER-positive/HER2-positive and 4 (36.4%) triple negative. Among the 22 patients that developed brain metastases during two years of follow up, 6 (27.3%) were ER-positive/HER2-negative, 6 (27.3%) ER-positive/HER2-positive, 4 (18.2%) ER-negative/HER2-positive and 6 (27.3%) triple negative.

Table 1 - Patient Characteristics

Parameter	Description *	Patients with BM at baseline (*n*= 11)	Patients who developed BM within 2 years of FU (*n* = 22)	Patients who did not develop BM within 2 years of FU (*n* = 99)
Age at sample draw	Median age†	51.2 (46.0 - 61.7)	61.9 (46.1 - 74.6)	62.4 (53.3 - 69.6)
Receptor status primary tumor				
ER	Positive	7 (63.6)	12 (54.5)	89 (89.9)
	Negative	4 (36.4)	10 (45.5)	10 (10.1)
HER2	Positive	4 (36.4)	10 (45.5)	28 (28.3)
	Negative	7 (63.6)	12 (54.5)	71 (71.7)
Adjuvant endocrine therapy (%)	No	8 (72.7)	16 (72.7)	61 (61.6)
	Yes, tamoxifen only	2 (18.2)	4 (18.2)	24 (24.2)
	Yes, tamoxifen + AI	1 (9.1)	1 (4.5)	9 (9.1)
	Yes, AI only	0 (0)	1 (4.5)	5 (5.1)
Adjuvant chemotherapy (%)	No	9 (81.8)	14 (63.6)	70 (70.7)
	Yes	2 (18.2)	8 (36.4)	29 (29.3)
Adjuvant anti-HER2 therapy (%)	No	10 (90.9)	19 (86.4)	99 (100)
	Yes	1 (9.1)	3 (13.6)	0 (0)
Therapy after start of inclusion	First line endocrine	0 (0)	5 (22.7)	65 (65.7)
	First line chemo	11 (100)	17 (77.3)	34 (34.3)
CTC count	Median count†	7 (2 - 28)	10 (1 - 43)	3 (0 - 20)
Number of extra-cranial metastatic sites	Median number	3 (1 - 4)	2 (1 - 3)	2 (1 - 2)

*Values in parentheses are percentages unless indicated otherwise;
†Values in parentheses are interquartile ranges

The median overall survival (OS) of patients with brain metastases at diagnosis was 12.7 months (IQR: 6.9 – 24 months). In patients who developed brain metastases after inclusion, the median time to brain metastases diagnosis was 11.4 months (inter quartile range (IQR): 7.9 – 15.2 months) and the OS was 17.4 months (IQR: 12.1 – 27.4 months). The OS in patients who did not develop brain metastases within two years of FU was 57.3 months (IQR: 32.7 – 91.4 months).

Variables associated with the time to brain metastases development

In univariable Cox regression, ER-negative status, younger age, and a higher number of extra-cranial metastatic sites were associated with a higher likelihood of the

development of brain metastases (**Table 2**). In addition to these clinico-pathological parameters, high expression of *KRT17*, *CD133, MAGEA3, PLAU* and low expression of *TSPAN13 (*previously known as *TM4SF13*)*, CD29, CD44,* and *CAV1* were significantly associated with the development of brain metastases. Neither HER2 status nor baseline CTC count were significantly associated with the risk of brain metastases in the univariable model. In multivariable analysis, ER-negative status, younger age, higher gene expression of *MAGEA3* and *PLAU*, and lower gene expression of *TSPAN13* and *CD44* still showed a significantly greater hazard of developing brain metastases. From all significant variables in the multivariable model, a Kaplan-Meier plot was generated (**Figure 1**).

Table 2 - Cox regression analysis (N=132)

	Univariable analysis			Multivariable analysis		
Variable	**HR**	**95%CI**	**P-value**	**HR**	**95%CI**	**P-value**
ER-status	0.23	0.11-0.46	0.003	0.34	0.15-0.78	0.010
Age	0.97	0.95-1.00	0.042	0.96	0.93-0.99	0.012
Number of extra-cranial metastatic sites	1.83	1.33-2.51	<0.001	1.49	0.98-2.25	0.061
KRT17	1.13	1.06-1.20	<0.001	1.04	0.96-1.13	0.322
CD133	1.31	1.09-1.58	0.004	1.07	0.83-1.36	0.611
MAGEA3	1.11	1.03-1.20	0.009	1.12	1.02-1.23	0.017
CD44	0.55	0.35-0.89	0.014	0.43	0.24-0.77	0.004
CAV1	0.86	0.76-0.97	0.015	0.94	0.81-1.09	0.386
TSPAN13	0.84	0.73-0.98	0.022	0.79	0.65-0.96	0.015
PLAU	1.13	1.00-1.27	0.048	1.16	1.01-1.33	0.041

Predicting the development of brain metastases within two years after the initial blood draw

Next, we set out to determine whether the significant variables from the multivariable Cox regression analyses (i.e. ER-negative status, younger age, higher gene expression of *MAGEA3*, *PLAU*, and lower gene expression of *TSPAN13* and *CD44*) could be used to predict the development of brain metastases within 2 years after blood sampling. Using logistic ridge regression, we built a prediction model which resulted in an area under the curve (AUC) for brain metastases of 0.77 (95% CI 0.67-0.88) (**Figure 2A**). At the highest Youden's index a sensitivity of 67%, a specificity of 85%, an accuracy of 80%, positive predictive value (PPV) of 60%, and a negative predictive value (NPV) of 88% were obtained. The combination of both clinico-pathological variables and gene expression performed better than clinico-pathological variables alone (AUC of 0.77 versus 0.68, *P*=0.021).

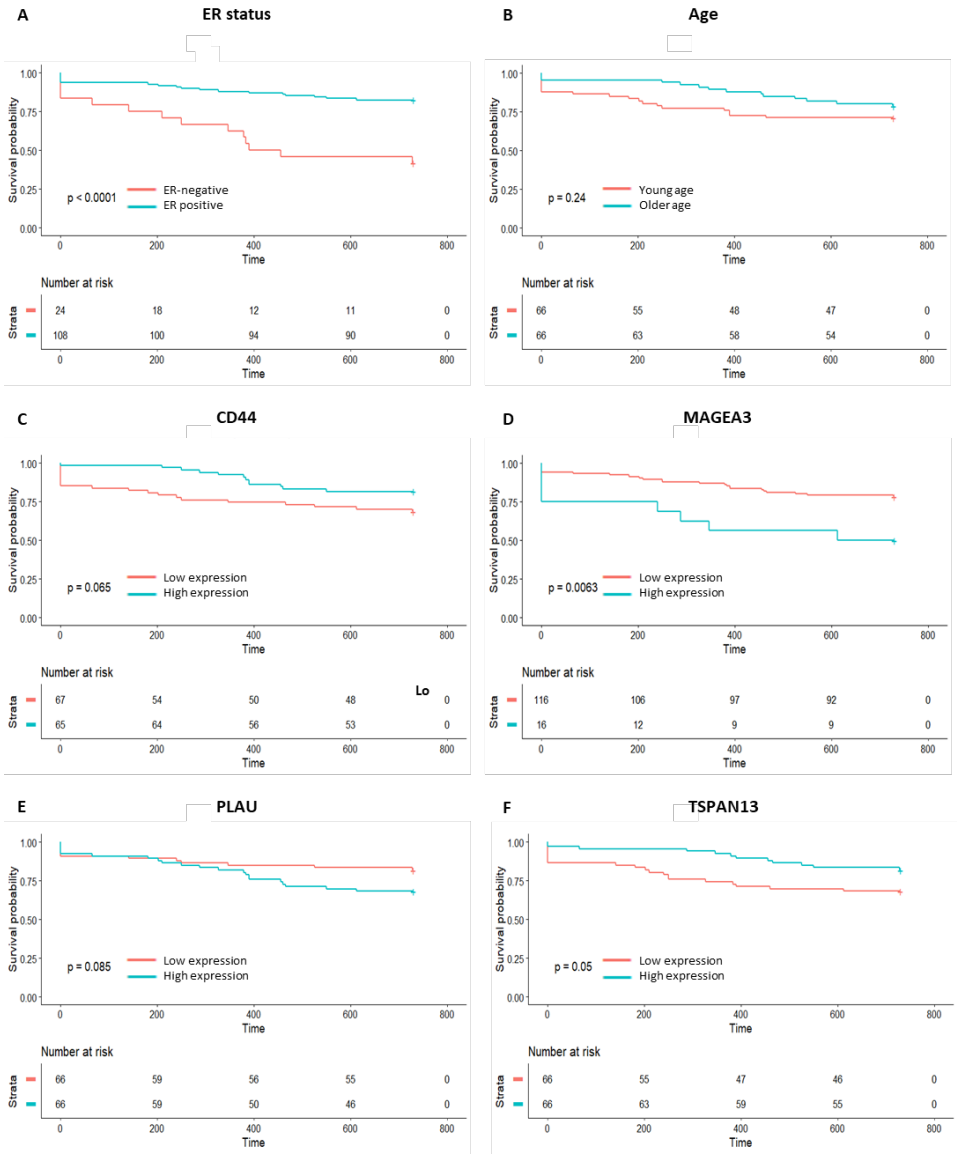
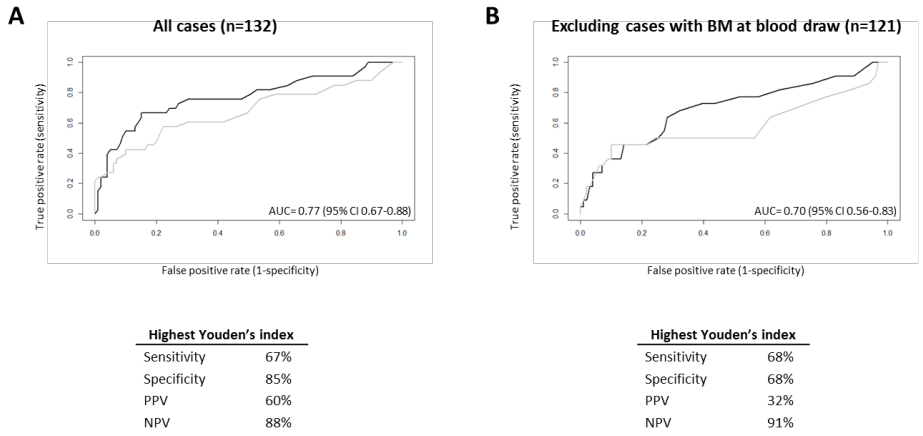


Figure 1 - Kaplan Meier plots of each variable significant in the multivariable Cox regression.

When all cases who already had brain metastases at baseline (n=11) were excluded, we obtained an AUC of 0.70 (95% CI 0.56-0.83, **Figure 2B**) for brain metastases development within two years. At the highest Youden's index for this second model a sensitivity of 68%, a specificity of 68%, an accuracy of 68%, positive predictive value (PPV) of 32%, and a negative predictive value (NPV) of 91% were obtained. Again, the combination

of clinico-pathological variables and gene expression performed better than clinico-pathological variables alone (AUC of 0.70 versus 0.59, *P*=0.028). Test performance and test characteristics of both models to predict brain metastases development within 2 years are depicted in **Figure 2**.



Highest Youden's index	
Sensitivity	67%
Specificity	85%
PPV	60%
NPV	88%

Highest Youden's index	
Sensitivity	68%
Specificity	68%
PPV	32%
NPV	91%

Figure 2 - Receiver operating characteristic (ROC) curves to predict brain metastases at baseline and during two years after blood sampling (n=132)(A) and only those cases that developed brain metastases during two years of follow up after blood sampling (n=121, patients with brain metastases at baseline were excluded)(B). The black line indicates the ROC curve for all variables (clinico-pathological plus genes), the gray line indicates the ROC curve for the clinico-pathological variables only.

Discussion

In the current study, encompassing a large cohort globally representing the real-life distribution of breast cancer subtypes, we found that expression of *MAGEA3*, *PLAU*, *TSPAN13* and *CD44* measured in CTCs was associated with the development of brain metastases in breast cancer patients and that this association was independent from additional clinico-pathological factors. A combined prediction model containing both genes and clinico-pathological variables resulted in a 85% specificity and 67% sensitivity for the development of brain metastases within 2 years. Although the NPV (88%) of this model would allow for the selection of patients that are at low-risk for brain metastases development, the poor PPV of 60% will result in many patients undergoing unnecessary imaging with the associated costs and anxiety. However, when evaluating our model performance, it is important to consider that brain imaging was only performed in our cohort when symptoms of brain metastases occurred. Since we

did not routinely perform imaging of the brain, presence of brain metastasis was not excluded in the control group of patients who did not have clinical signs and symptoms of brain metastases. As a result, our control group could contain some patients with subclinical brain metastasis.

In concordance with previous results[6,18], we showed that a younger age and ER-negative receptor status were significantly associated with brain metastases development in our multivariable Cox regression, whereas the number of extra-cranial metastatic sites was borderline significant in multivariable analyses. Of note, in our cohort HER2-status was not associated with an increased risk of brain metastases development. Grade and lymph node status of the primary tumor, both of which were associated with the development of brain metastases in patients with breast cancer previously, were not recorded in our database, and could therefore not be assessed in our cohort.

Importantly, the 93-gene expression profile applied to our cohort was not specifically designed for the purpose of predicting the development of brain metastasis[21]. Incorporation of other genes that have previously been linked to brain metastases development – such as *MYC*[20], *SEMA4D*[20], *STAT3*[27] and the 17 gene panel associated with development of breast cancer brain metastases identified by Bos *et al.*[19] – might have enhanced the classifier. However, as enriched CTCs are analyzed in a background of contaminating leukocytes, our 93-gene panel has been carefully selected for genes that have a higher expression in breast cancer cells than in leukocytes[23]. *SEMA4D* has a higher expression in whole blood than in breast tissue and *STAT3* is similarly expressed in whole blood and breast tissue (https://gtexportal.org/), which render these genes unsuitable for the current methodology in which only few CTCs are measured in a relatively large background of leukocytes.

Among the genes that were significantly associated with brain metastases development in our study, two have previously been linked with breast cancer brain metastases: the tumor suppressor gene caveolin-1 (*CAV1*)[19,27] and stemness gene *CD44*[28]. In our dataset, a lower expression of *CAV1* was associated with brain metastasis in the univariable Cox regression analysis. Chiu *et al.* have shown that caveolin-1 negatively regulates the activation of STAT3, of which the expression was higher in breast cancer brain metastases than in ductal carcinoma in situ and invasive ductal carcinoma. Conversely, activation of STAT3 in brain-metastatic cancer led to lower expression of caveolin-1, which might be one of the mechanisms responsible for sustained STAT3 activation in brain metastases which promotes breast cancer invasion and brain metastases[27]. In addition to *CAV1*, CD133 – a stem cell marker – has been linked to a poor prognosis in

several cancer types[29]. In medulloblastoma, a primary brain tumor, higher expression of CD133 has been linked with higher proliferation capacity, and compared to CD133 negative cells CD133 positive cells formed much larger tumors in xenograft models[30]. Also, a higher CD133 expression was associated with a higher expression of c-MYC via STAT3[30]. Since both *CAV1* and *CD133* are related to the expression of *STAT3,* this represents an interesting marker for future research on CTC-gene expression and development of brain metastases. Matched analyses of CTCs of patients who develop brain metastases and brain metastatic tissue might yield additional genes of interest.

In contrast with the study of Sirkisoon *et al*.[28] which showed that truncated glioma-associated oncogene homolog 1 (*TGLI1*) led to transcriptional activation of stemness gene *CD44*, thereby promoting breast cancer metastasis to the brain, we found that a lower *CD44* gene expression was associated with brain metastases development. The fact that we measured this gene in a CTC-enriched fraction also containing leukocytes which can express *CD44,* could have influenced this result. So, assuming that only a subset of CTCs is capable to extravasate through the BBB, measurement of gene expression profiles in single CTCs rather than studying pools of enriched CTCs in a high background of contaminating leukocytes[23], might lead to improved identification of patients with CTCs capable of crossing the BBB.

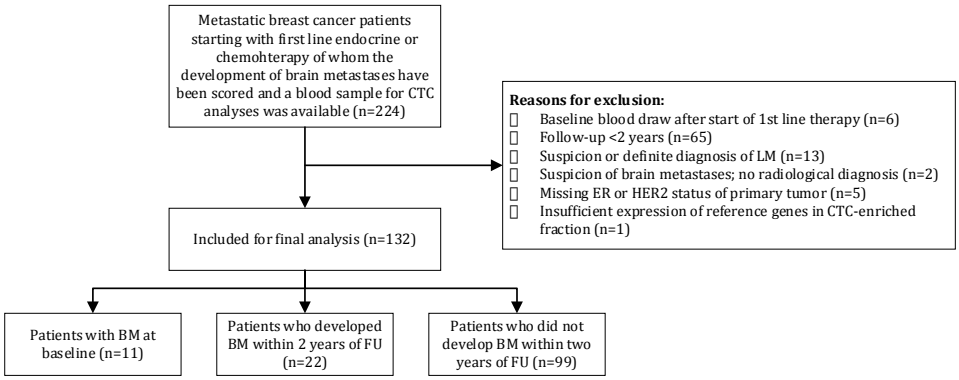
In conclusion, we show that clinico-pathological variables and gene expression of some genes (*MAGEA3*, *PLAU*, *TSPAN13* and *CD44*) in the CTC-enriched fraction are associated with the development of brain metastases, irrespective of CTC count. Although our predictor for developing brain metastasis within two years of follow-up has a high NPV, to be of clinical value future research should focus on optimization of the CTC gene-panel (ideally measured in pools of pure CTCs) – in which at least *MYC*, *STAT3* and *SEMA4D* should be incorporated – in order to enhance the PPV of the classifier.

References

1. Nieder, C. *et al.* Presentation, patterns of care, and survival in patients with brain metastases: what has changed in the last 20 years? *Cancer* **117**, 2505-12 (2011).
2. Thulin, A. *et al.* Clinical outcome of patients with brain metastases from breast cancer - A population based study over 21 years. *Breast* **50**, 113-124 (2020).
3. Frisk, G. *et al.* Incidence and time trends of brain metastases admissions among breast cancer patients in Sweden. *Br J Cancer* **106**, 1850-3 (2012).
4. Klos, K.J. & O'Neill, B.P. Brain metastases. *Neurologist* **10**, 31-46 (2004).
5. Meyers, C.A. *et al.* Neurocognitive function and progression in patients with brain metastases treated with whole-brain radiation and motexafin gadolinium: results of a randomized phase III trial. *J Clin Oncol* **22**, 157-65 (2004).
6. Martin, A.M. *et al.* Brain Metastases in Newly Diagnosed Breast Cancer: A Population-Based Study. *JAMA Oncol* **3**, 1069-1077 (2017).
7. Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *J Clin Oncol* **28**, 3271-7 (2010).
8. Clayton, A.J. *et al.* Incidence of cerebral metastases in patients treated with trastuzumab for metastatic breast cancer. *Br J Cancer* **91**, 639-43 (2004).
9. Leone, J.P. *et al.* Prognostic factors and survival according to tumour subtype in women presenting with breast cancer brain metastases at initial diagnosis. *Eur J Cancer* **74**, 17-25 (2017).
10. Niikura, N. *et al.* Treatment outcomes and prognostic factors for patients with brain metastases from breast cancer of each subtype: a multicenter retrospective analysis. *Breast Cancer Res Treat* **147**, 103-12 (2014).
11. Cardoso, F. *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4). *Ann Oncol* **29**, 1634-57 (2018).
12. Ramakrishna, N. *et al.* Recommendations on Disease Management for Patients With Advanced Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer and Brain Metastases: ASCO Clinical Practice Guideline Update. *J Clin Oncol* **36**, 2804-2807 (2018).
13. Ettinger, D.S. *et al.* NCCN clinical practice guidelines in oncology: non-small cell lung cancer version 3.2020 - February 11, 2020. https://www.nccn.org/professionals/physician_gls/PDF/nscl.pdf. (20 February 2020, date last accessed).
14. Postmus, P.E. *et al.* Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **28**, iv1-iv21 (2017).
15. Fruh, M. *et al.* Small-cell lung cancer (SCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **24 Suppl 6**, vi99-105 (2013).
16. Michielin, O. *et al.* Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **30**, 1884-1901 (2019).
17. Cagney, D.N. *et al.* Implications of Screening for Brain Metastases in Patients With Breast Cancer and Non-Small Cell Lung Cancer. *JAMA Oncol* **4**, 1001-1003 (2018).
18. Arvold, N.D. *et al.* Brain metastases after breast-conserving therapy and systemic therapy: incidence and characteristics by biologic subtype. *Breast Cancer Res Treat* **136**, 153-60 (2012).
19. Bos, P.D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-9 (2009).
20. Klotz, R. *et al.* Circulating Tumor Cells Exhibit Metastatic Tropism and Reveal Brain Metastasis Drivers. *Cancer Discov* **10**, 86-103 (2020).
21. Sieuwerts, A.M. *et al.* mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. *Clin Cancer Res* **17**, 3600-18 (2011).
22. Mostert, B. *et al.* Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients. *Ann Oncol* **26**, 510-6 (2015).
23. Sieuwerts, A.M. *et al.* Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR. *Breast Cancer Res Treat* **118**, 455-68 (2009).

24. DeLong, E.R. *et al.* Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837-45 (1988).
25. Youden, W.J. Index for rating diagnostic tests. *Cancer* **3**, 32-5 (1950).
26. van de Wiel, M.A. *et al.* Better prediction by use of co-data: adaptive group-regularized ridge regression. *Stat Med* **35**, 368-81 (2016).
27. Chiu, W.T. *et al.* Caveolin-1 upregulation mediates suppression of primary breast tumor growth and brain metastases by stat3 inhibition. *Cancer Res* **71**, 4932-43 (2011).
28. Sirkisoon, S.R. *et al.* TGLI1 transcription factor mediates breast cancer brain metastasis via activating metastasis-initiating cancer stem cells and astrocytes in the tumor microenvironment. *Oncogene* **39**, 64-78 (2020).
29. Burke, W.M. *et al.* Inhibition of constitutively active Stat3 suppresses growth of human ovarian and breast cancer cells. *Oncogene* **20**, 7925-34 (2001).
30. Garg, N. *et al.* CD133(+) brain tumor-initiating cells are dependent on STAT3 signaling to drive medulloblastoma recurrence. *Oncogene* **36**, 606-617 (2017).

Supplemental Figure



Supplementary Figure 1 - Overview of included patients for analyses.

Supplementary Table S1 - Genes included in CTC gene panel

ACTA1	*ESR1*	*MSMB*
AGR2	*FEN1*	*MUC1*
AURKA	*FGFR2*	*MYL3*
BST1	*FGFR3*	*NME1*
CAV1	*FGFR4*	*NOTCH3*
CCNB1	*FKBP10*	*PIP*
CCND1	*FOXA1*	*PKP3*
CCNE1	*GALGT*	*PLAU*
CCNE2	*GATA3*	*PLK1*
CD133	*GUSB*	*PLOD2*
CD24	*HMBS*	*PSMD10*
CD29	*HPRT1*	*PTPRC*
CD44	*IGFBP2*	*PTPRK*
CDH1	*IGFBP3*	*PTRF*
CDH5	*IGFBP4*	*S100A16*
CEACAM5	*IGFBP5*	*S100A7*
CEP55	*IL17BR3*	*SBEM*
CLDN3	*ITGA6*	*SCGB1D2*
COL1A1	*KIF11*	*SCGB2A2*
COL2A1	*KPNA2*	*SELE*
CRABP2	*KRT17*	*SEPP1*
CTTN	*KRT18*	*SNAPC2*
CXCL14	*KRT19*	*SPDEF*
DTL	*KRT7*	*TFF1*
DTX3	*KRT81*	*TFF3*
DUSP4	*LAD1*	*TIMP3*
EEF1A2	*LOXL2*	*TNRC9*
EGFR	*MAGEA3*	*TOP2A*
EPCAM	*MCAM*	*TSPAN13 (previously known as TM4SF13)*
ERBB2	*MELK*	*TWIST*
ERBB3	*MET*	*VEGFR2*
ERBB4	*MKI67*	*VWF*

PART III

Radiomics

CHAPTER 11

The BRAF p.V600E mutation status of melanoma lung metastases cannot be discriminated on computed tomography by LIDC criteria nor radiomics using machine learning

Journal of personalized medicine. *2021 Apr 1;11(4):257.*

Lindsay Angus*, Martijn P.A. Starmans*, Ana Rajicic, Arlette E. Odink, Mathilde Jalving, Wiro J. Niessen, Jacob J. Visser, Stefan Sleijfer, Stefan Klein‡ and Astrid A.M. van der Veldt‡

* Both authors contributed equally to this work
‡ Both authors contributed equally to this work

Abstract

Objectives

Patients with *BRAF* mutated (*BRAF*-mt) metastatic melanoma benefit significantly from treatment with BRAF inhibitors, therefore determination of the *BRAF* status in these patients is strongly recommended. Currently, this is determined either on archival tumor tissue or on fresh tumor tissue from an invasive biopsy. The aim of this study was to evaluate whether radiomics can predict the *BRAF* mutation status in a non-invasive manner in melanoma lung metastases.

Methods

Patients with melanoma lung metastases, known *BRAF* status, and a pretreatment computed tomography (CT) scan were included between January 2012 and February 2018. A maximum of two lung metastases per patient were included. After semi-automatic annotation of the lesions, 540 radiomics features were extracted. A chest radiologist scored all segmented lung lesions according to the Lung Image Database Consortium (LIDC) criteria. Univariate analysis was performed to assess the predictive value of each feature for *BRAF* mutation status. A combination of various machine learning methods was used to develop *BRAF* decision models based on the radiomics features and LIDC criteria.

Results

169 lung lesions from 103 patients (51 *BRAF*-mt; 52 *BRAF* wild type) were included. There were no features with a significant discriminative value in the univariate analysis. Models based on radiomics features and LIDC criteria both performed as poorly as guessing with a mean area under the curve (AUC) of 0.49 (95% CI: 0.38-0.59) and 0.46 (95% CI: 0.38-0.55), respectively.

Conclusions

The *BRAF* mutation status in melanoma lung metastases cannot be predicted using radiomics features or visually scored LIDC criteria.

Introduction

Cutaneous melanoma is an aggressive skin cancer most commonly occurring on the ultra-violet light exposed skin of Caucasians[1,2]. In Europe, it is the 8th most common malignancy in men and the 5th most common in women, with an annual incidence of 144,200 new cases and 27,100 deaths[3]. In the coming years, the incidence of melanoma is expected to increase rapidly, resulting in an increased melanoma-associated mortality[4].

The introduction of new systemic treatment modalities, including immunotherapy and BRAF inhibitors, has significantly improved the prognosis of patients with metastatic melanoma[5]. Approximately 50% of melanomas harbor a mutation in the *BRAF* gene, with p.V600E being the most common variant[6-8]. Patients with *BRAF*-mutant (*BRAF*-mt) melanoma benefit significantly from treatment with BRAF inhibitors and onset of response is often rapid[9]. To enhance response rates and duration of response, patients are usually treated with a combination of a BRAF and a MEK inhibitor[10-13]. Due to the therapeutic consequences, determination of the *BRAF* mutation status in patients with metastatic melanoma is mandatory according to the European Society of Medical Oncology guidelines[14].

Currently, the *BRAF* mutation status is usually determined by molecular analysis of a metastatic lesion[15]. However, tissue biopsies are invasive, thereby exposing patients to potential risks including bleeding, infection and in case a lung biopsy is taken the risk of pneumothorax. In addition, molecular analyses can be time-consuming, especially when the tumor specimen has been archived at another hospital. Since patients with metastatic melanoma can experience rapidly progressive disease with life-threatening symptoms and an urgent medical need for systemic therapy, faster and less invasive diagnostics to determine the BRAF mutation status may significantly improve patient management.

Recently, various tumor characteristics have been predicted non-invasively using quantitative imaging features– also referred to as 'radiomics'. In non-small cell lung cancer, radiomics on computed tomography (CT) can predict tumor stage and epidermal growth factor receptor (EGFR) mutation status[16-24]. In patients with primary colorectal cancer, a CT radiomics signature that was associated with *BRAF/NRAS/KRAS* mutation status[25]. Although CT-based radiomics has been successfully applied to predict response to immunotherapy in melanoma lymph node metastases[26], the value of radiomics for predicting *BRAF* mutation status has

not been investigated. If CT-based radiomics could predict *BRAF* mutation status with a high positive predictive value, this may provide a faster and more patient-friendly alternative to determine the *BRAF* mutation status in metastatic melanoma.

The aim of this study was to evaluate the utility of CT-based radiomics to predict BRAF mutation status (mutant versus wild type) in metastatic melanoma. In metastatic melanoma, lung metastases are relatively easy to annotate on CT as compared to other metastases since they can be clearly distinguished from healthy lung tissue. Therefore, the aim of this study was to evaluate the utility of CT-based radiomics to predict *BRAF* mutation status (mutant versus wild type) in melanoma lung metastases.

Materials and Methods

Data collection

This study was approved by the Erasmus MC institutional research board (MEC-2019-0693). Anonymized patient data was used and therefore need for written informed consent was waived by the Institutional Review Board. All patients diagnosed with metastatic melanoma at Erasmus MC between January 2012 and February 2018 were included retrospectively if they met the following pre-specified criteria: known tumor *BRAF* mutation, diagnostic contrast-enhanced thoracic CT scan prior to commencement of any systemic therapy, and at least one lung metastasis of ≥10 mm evaluable according to Response Evaluation Criteria In Solid Tumors (RECIST) v1.1[27]. Patients with *BRAF* mutations other than p.V600E were excluded from the analysis, since BRAF inhibitors may be less effective in patients with other *BRAF* mutations[28]. Formalin-fixed paraffin-embedded material of the primary tumor and/ or metastasis is tested for *BRAF* (exon 15) using a polymerase chain reaction (PCR) based assay or next generation sequencing as part of standard care.

Radiomics

Lung metastases were measured according to RECISTv1.1[27]. For 3D segmentation, up to two lung lesions ≥10 mm were selected by a clinician supervised by an experienced chest radiologist. In patients with >2 lung metastases of ≥10 mm, either the two largest or the two most easily distinguishable lesions were segmented (i.e. two separate lesions were preferred over two adjacent lesions). Using in-house developed software[29], selected lung metastases were segmented semi-automatically using a lung window for visualization. The result was visually inspected and manually corrected when necessary by an experienced chest radiologist to ensure that the semi-automatic

segmentation resembled the manual segmentation. The clinician and chest radiologist were both blinded for *BRAF* mutation status. From each segmented lesion, 540 radiomics features were extracted to quantify intensity, shape, and texture. Details are described in **Supplementary Materials 1**. To create a decision model using these features, the Workflow for Optimal Radiomics Classification (WORC) toolbox was used (**Figure 1**)[30-32]. Details are described in **Supplementary Materials 2**. In brief, the creation of a decision model in WORC consists of several steps, including selection of relevant features, resampling, and machine learning techniques to identify patterns to distinguish *BRAF*-mt from *BRAF* wild type (*BRAF*-wt) lesions. WORC performs an automated search including a variety of algorithms for each step and determines which combination of algorithms maximizes the predictive performance on the training set. The open-source code for the feature extraction and model optimization has been published[33].

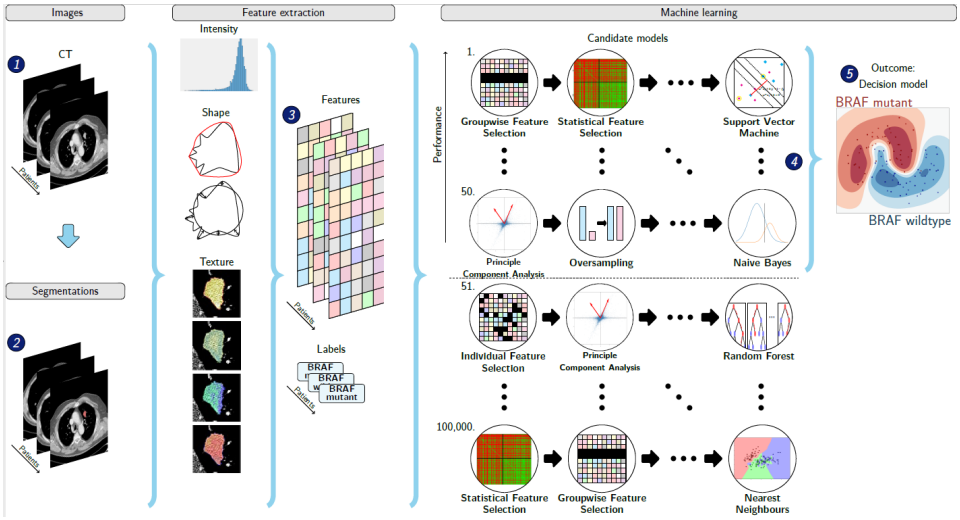


Figure 1 - Schematic overview of the radiomics approach: adapted from Vos and Starmans *et al.*[32]. Inputs to the algorithm are (1) contrast-enhanced thoracic CT images of patients with *BRAF* mutated or *BRAF* wild type metastatic melanoma and (2) a segmentation of the lung metastasis. Processing steps include (3) feature extraction and (5) the creation of a machine learning decision model, using (4) an ensemble of the best 50 workflows from 100,000 candidate workflows, which are different combinations of the different processing and analysis steps (e.g. the classifier used).

Scoring by radiologist

A chest radiologist (certified for 8 years) scored the segmented lung lesions. There are no guidelines to differentiate histologic subtypes in lung metastases, therefore the Lung Image Database Consortium (LIDC) criteria were used. These criteria were developed to standardize the description of radiological features of lung abnormalities in clinical practice[34]. The following LIDC features were rated: subtlety, calcification, internal structure, lobulation, likelihood of malignancy, margin, sphericity, spiculation, and texture (see **Supplementary Table S1** for the rating system). The radiologist was blinded for the *BRAF* status, but not to the diagnosis of metastatic melanoma and had access to the CT scan, age, and sex of the patient.

Experimental setup

To assess the predictive value of quantitative imaging features (i.e. radiomics features) and LIDC features, four models were trained and tested using WORC based on: 1) automatically extracted radiomics features only 2) similar to model 1, but only including the largest lesion per patient; 3) similar to model 1, but only including patients with *NRAS* and *BRAF* wild type melanoma for the comparison with *BRAF*-mt; 4) manually scored LIDC features only; and 5) a simple benchmark model. Model 2 was applied to assess a potential bias for patients with multiple lesions. Model 3 was included because activating *NRAS* mutations could potentially result in a similar phenotype as *BRAF*-mt, since mutations in both genes lead to activation of the mitogen-activated protein kinase (MAPK) pathway. The simple benchmark model was evaluated in a similar way as model 1, i.e. using all lesions and automatically extracted radiomics features. Model 5 was applied to compare the performance of WORC to a simple benchmark machine learning model, which uses binary logistic regression with LASSO feature selection (i.e. ElasticNet).

Statistics

To assess the predictive value of the individual features, the Mann-Whitney U test was performed for univariate analyses of continuous variables and Pearson's chi-squared test was used for categorical variables. For radiomics, P-values were corrected for multiple testing using the Bonferroni correction according to the default in WORC. A P-value of <0.05 was considered to be statistically significant.

Evaluation of the radiomics models was performed using a 100x random-split cross-validation. In each iteration, the data was randomly split into 80% for training and 20% for testing in a stratified manner to guarantee a similar distribution of the classes in the training and test set as compared to the original set. Metastases from the same patients were always grouped together in either the training or test set. To eliminate

the risk of overfitting, in each iteration, all model optimization was performed strictly within the training set by using a second internal 5x random-split cross-validation (see **Supplementary Figure S1**). The final model consists of an ensemble of the 50 best workflows, i.e. combination of methods and parameters, each defined by a specific set of hyperparameters. This final model may be different in each of the 100x random-split cross-validation iterations. For each of the five models described in the experimental setup, these sets hyperparameters are included with the code[33]. Details are described in **Supplementary Materials 2**.

The performance of all four models was described by the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). The positive class was defined as *BRAF*-mt. For each metric, the average over the 100 cross-validation iterations and a 95% confidence interval (CI) were reported. The 95% CIs were constructed using the corrected resampled t-test based on the results from all 100 cross-validation iterations, thereby taking into account that the samples in the cross-validation splits are not statistically independent[35]. ROC confidence bands were constructed using fixed-width bands[36].

Results

Study population

In total, 103 patients were included, see **Supplementary Figure S2** for a flowchart of patient inclusion. Characteristics of these patients and their CT scans are summarized in **Table 1**. The median age was 65 years (interquartile range (IQR) 52 – 72), and 50.5% of the patients were men. *BRAF* mutation status was either determined on the primary tumor (*N*=20), local recurrence (*N*=3), or metastasis (*N*=79). In these lesions, *BRAF* p.V600E was detected in 51 patients, whereas 52 patients had *BRAF*-wt melanomas. In total, 103 CT scans were acquired from 10 different CT scanners, resulting in in the inclusion of data acquired with different acquisition protocols (**Table 1**). Although for all acquisition parameters the difference between *BRAF*-mt and *BRAF*-wt was not statistically significant, the difference in tube current reached almost statistical significance (p=0.05).

Radiomics and LIDC features and models

In total, 169 lung metastases in 103 patients were segmented. **Figure 2** illustrates randomly selected segmentations of lung metastases from patients with *BRAF*-mt and

BRAF-wt metastatic melanoma. Median volume of segmented lung lesions was 18.3 ml (IQR: 7.3-48.6 ml). None of the radiomics or LIDC features were significantly different between *BRAF*-mt and *BRAF*-wt lung metastases, as none of the features had a *P*-value <0.05 after Bonferroni correction. LIDC criteria scores are shown in **Supplementary Table S2**. Using all 169 lung metastases, the radiomics model (model 1) resulted in a mean AUC of 0.49, sensitivity of 0.61, and specificity of 0.37 (**Figure 3A**; **Table 2**). Model 2, i.e. only inclusion of the largest lesion per patient, slightly improved the performance (AUC of 0.65), whereas model 3, i.e. only inclusion of *BRAF*-wt melanoma who were also *NRAS* wild type, still had a poor performance (AUC of 0.49) (**Figure 3B** and **3C**; **Table 2**). In addition, model 4, i.e. based on the LIDC features scored by a radiologist, resulted in an AUC of 0.46 (**Figure 3D**). The simple benchmark (model 5) resulted in a similar performance (AUC of 0.50).

Table 1 - Patient and imaging characteristics

Patient	***BRAF*-mt (*N*=51)**	***BRAF*-wt (*N*=52)**	***P*-value**
Age (years)[†]	59 (50-69)	66 (57-74)	0.048
Sex			0.768
Male	25 (49)	27 (52)	
Female	26 (51)	25 (48)	
Primary tumor localization			0.027
Skin	49 (96)	42 (81)	
Mucosal	0 (0)	6 (11)	
Unknown	2 (4)	4 (8)	
Determination of *BRAF-mutation* status			0.851
Primary tumor	9 (18)	11 (21)	
Local recurrence	1 (2)	2 (4)	
Metastasis	40 (78)	39 (75)	
Unknown	1 (2)	0 (0)	
***NRAS* mutation status**[$]			Not determined
Mutant	-	22 (42)	
Wild type	-	23 (44)	
Unknown	-	7 (2)	
Imaging			
Acquisition protocol			
Slice thickness (mm) [†1]	1.5 (1.5, 1.5)	1.5 (1.5, 1.5)	0.23
Pixel spacing (mm) [†]	0.68 (0.64, 0.74)	0.67 (0.61, 0.73)	0.16
Tube current (mA) [†]	405 (278, 553)	333 (210, 490)	0.05
Peak kilovoltage[†1]	120 (120, 120)	120 (118, 120)	0.44
Contrast Agent			0.84
Visipaque 320	35	37	
Ultravist	1	0	
Omnipaque	1	1	
Optiray	0	1	
Unknown	14	13	
Number of segmented lesions per patient			0.54
One	20 (39)	17 (33)	
Two	31 (61)	35 (67)	

Values in parentheses are percentages unless stated otherwise. [†] Values are median (Inter quartile range)
[$] *NRAS* and *BRAF* mutations are mutually exclusively occurring, hence we did not test for significance between *BRAF* wild type versus mutant cases
[1] Other values than those given in the median and inter quartile range do occur

Table 2 - Performance of the models for *BRAF* mutation prediction based on different sets of features and lesions

	Model 1 Radiomics all lesions - WORC	**Model 2** Radiomics largest lesion	**Model 3** Radiomics *NRAS* wild type	**Model 4** LIDC all lesions	**Model 5** Radiomics all lesions - benchmark
AUC	0.49 [0.38, 0.59]	0.65 [0.51, 0.79]	0.49 [0.37, 0.61]	0.46 [0.38, 0.55]	0.50 [0.42, 0.58]
Accuracy	0.48 [0.39, 0.57]	0.61 [0.50, 0.72]	0.65 [0.58, 0.71]	0.49 [0.42, 0.56]	0.50 [0.43, 0.57]
Sensitivity	0.61 [0.44, 0.77]	0.61 [0.42, 0.80]	0.94 [0.87, 1.00]	0.29 [0.11, 0.48]	0.56 [0.32, 0.80]
Specificity	0.37 [0.22, 0.52]	0.60 [0.38, 0.82]	0.08 [0.00, 0.17]	0.66 [0.46, 0.86]	0.44 [0.20, 0.69]
NPV	0.53 [0.39, 0.66]	0.61 [0.46, 0.76]	0.35 [0.00, 0.75]	0.52 [0.42, 0.61]	0.43 [0.21, 0.66]
PPV	0.45 [0.37, 0.53]	0.63 [0.48, 0.77]	0.67 [0.62, 0.72]	0.44 [0.30, 0.58]	0.47 [0.37, 0.56]

* Abbreviations: AUC: area under the receiver operating characteristic curve; PPV: positive predictive value; NPV: negative predictive value.

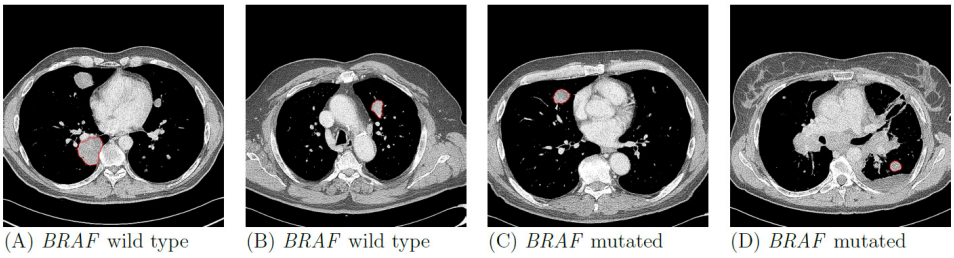


Figure 2 - Examples of *BRAF* mutant and *BRAF* wild type lung metastases of four patients with metastatic melanoma.

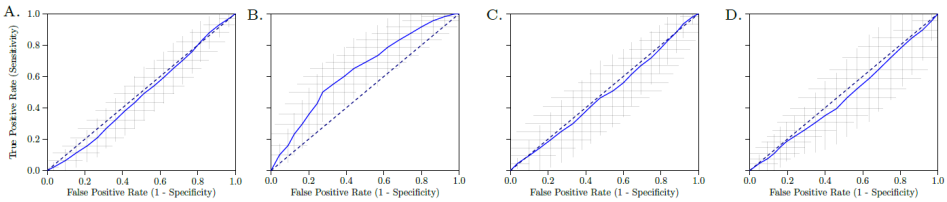


Figure 3 - Receiver operating characteristic (ROC) curve of the radiomics model of all lesions (A), only the largest lesion (B), only BRAF wild type lesions with NRAS wild type (C), and LIDC features (D). The crosses identify the 95% confidence intervals of the 100x random-split cross-validation; the blue curve is fit through their means.

Discussion

The results of this study show that there is no association between radiomics features of lung metastases and the *BRAF* mutation status in patients with metastatic melanoma. Our model using only the largest lesion per patient performed best with a moderate mean AUC, but still none of the features had any individual discriminative value. Also, the performance confidence intervals (e.g. the sensitivity and specificity) still included

many values below the performance of guessing. The LIDC criteria as scored by a thorax radiologist also failed to discriminate the *BRAF* mutation status in melanoma lung metastases.

Despite the remarkable success of BRAF inhibitors and immunotherapy in patients with metastatic melanoma, only a subset of patients benefits from these therapies[11,37]. Tools to select the patients most likely to benefit are of great interest and this has resulted in several radiomics studies aiming to predict tumor response. Similar to our study, previous radiomics models, either to predict therapy response or survival, had a low to moderate performance in metastatic melanoma[26,38,39]. In the largest radiomics study in melanoma thus far, 483 lesions from 80 melanoma patients were included and a greater morphological heterogeneity of lymph nodes determined by CT was associated with immunotherapy response, resulting in a moderate AUC of 0.64[26]. However, the model performed poorly in lung and liver lesions (AUC of 0.55). Comparable to our CT-based findings, a recent study showed that radiomics features derived from ^{18}F-FDG PET to determine the *BRAF* p.V600E mutation status also had a moderate performance (AUC of 0.62). They studied 176 lesions, including 18 lung lesions from 70 patients with melanoma (35 *BRAF*-mt and 35 *BRAF*-wt)[40]. To the best of our knowledge, this PET study[40] and our CT study are the first melanoma studies aiming to predict *BRAF* p.V600E mutation status, showing that neither PET nor CT radiomics features can discriminate between patients with *BRAF*-mt and *BRAF*-wt melanomas. We therefore believe that our comprehensive study provides insight into the potential of radiomics in this area, which can guide future research[41].

The lack of discrimination between *BRAF*-mt and *BRAF*-wt melanoma could potentially be explained by activating mutations in the *NRAS* gene in *BRAF*-wt melanoma. Since *NRAS* and *BRAF* are involved in the same pathway, i.e. the MAPK pathway, activating *NRAS* and *BRAF* mutations could result in a similar phenotype. Therefore, we evaluated an additional model which only included *NRAS* wild type lesions in patients with *BRAF*-wt melanoma (model 3). In our cohort of patients with *BRAF*-wt melanoma, 22 out of 45 (49%) patients - with known *NRAS* mutation status - had a *NRAS* mutation. Exclusion of all patients with *NRAS* mutation or unknown *NRAS* mutation status resulted in an AUC of 0.54 (95% CI 0.44-0.64). Based on these findings, it is very unlikely that inclusion of *NRAS* mutant melanomas negatively impacted our results. In addition, our findings are supported by the low predictive value of PET radiomics in the same setting in which patients with NRAS mutations were also excluded[40].

Our study was designed for a comprehensive evaluation of the relationship between

CT imaging features and the *BRAF* mutation status in melanoma lung metastases. To our knowledge this is currently the largest CT-based radiomics study on the *BRAF* mutation status in patients with metastatic melanoma. It is unlikely that, treatment-related resistance mechanisms influenced the outcome, since the study population was treatment-naïve, thereby reflecting the appearance of untreated melanoma lung metastases. The investigated patient population only included melanoma patients for whom correct determination of the *BRAF* status is of utmost importance for rapid treatment stratification. The WORC radiomics method applied has been previously validated to predict mutation status of several genes in other tumor types, such as lipoma and liposarcoma[32], desmoids[42], gastrointestinal stromal tumors[43], liver cancer[29,44], prostate cancer[45] and mesenteric fibrosis[46]. In these previous studies, the radiomics models had a much better performance (mean AUCs between 0.71 - 0.89) and multiple features were statistically significant in univariate statistical testing. In the current study, none of the radiomics features had any discriminative value, therefore it can be concluded that radiomics features of melanoma lung metastases are not related to the *BRAF* mutation status. WORC includes a wide variety of radiomics approaches and automatically optimizes the combination, thereby evaluating many different approaches. Hence, it is unlikely that a different radiomics approach will lead to a positive result. In addition to the radiomics analysis, a radiologist visually evaluated the lesions. Similar to radiomics results, the radiologist could not discriminate between *BRAF*-wt and *BRAF*-mt lesions by applying the LIDC criteria. Although radiomics can potentially correlate imaging features with clinical outcome even in cases where a radiologist cannot, the relation between quantitative imaging features and clinical outcome is considered stronger when clinical outcomes can be discriminated visually by a radiologist. This was not evident in the current study and this can be considered additional evidence that a CT-based radiomics signature probably does not exist for BRAF in melanoma lung metastases.

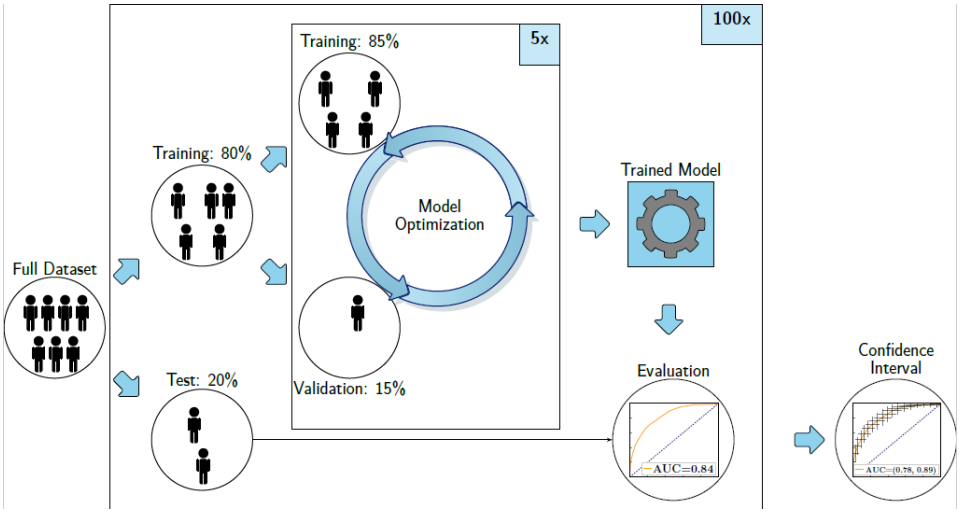
Our study has several limitations. Firstly, the *BRAF* mutation status was often determined on other tumor tissue than the segmented lung metastases. The *BRAF* status was determined on biopsy material from a lung metastasis, which did not necessarily match the segmented lung lesion, in only 12 patients. Although the concordance rate of the *BRAF* mutation status between primary melanoma and metastases is quite high[8,47,48], a recent meta-analysis showed a pooled discrepancy rate of 13.4% between primary melanomas and metastases, and a 7.3% discrepancy rate between metastatic sites[49]. Hence, tumor heterogeneity might have caused misclassification of *BRAF* mutation status, thereby negatively affecting the results. Secondly, the segmentation of regions of interest (ROI) was performed semi-automatically. Automatic segmentation methods

may improve the consistency of the segmentations and thus affect the radiomics model. However, due to the clear distinction of lung lesions and their surroundings, it is not expected that automatic segmentation will substantially alter the results. Thirdly, the heterogeneity in the acquisition protocols may have negatively affected the performance or our radiomics model. These variations may have led to variations in the imaging features, which complicate the recognition of patterns. Using a single acquisition protocol would give an estimate of the performance unaffected by such variations. However, the variations in the acquisition protocols were small, making it unlikely this significantly affected the results of the current study. The difference in tube current between *BRAF*-mt and *BRAF*-wt almost reached statistical significance and could have been implicitly used by the model to distinguish these lesions. However, our results show that, despite this difference, the performance of the model was similar to guessing. Lastly, although a rigorous cross-validation was used, strictly separating training from testing data, we did not validate our findings on an independent, external dataset.

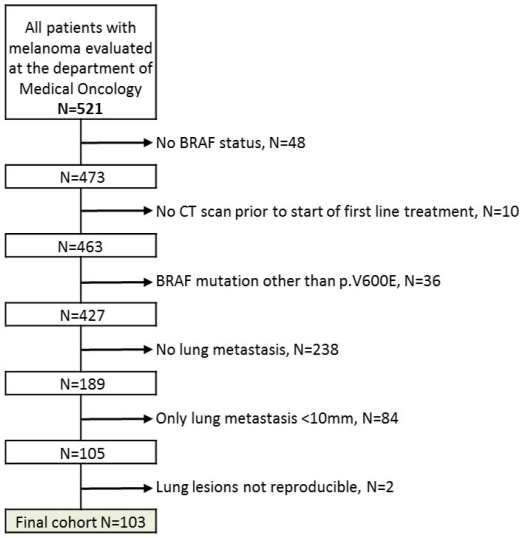
Conclusion

In summary, our study demonstrates that neither CT-based radiomics features, nor CT-derived LIDC features scored by a radiologist can discriminate between *BRAF* mutant and *BRAF* wild type lung metastases in patients with metastatic melanoma. Therefore, CT-based parameters cannot replace determination of *BRAF* mutation status on tumor tissue.

Supplemental Figures



Supplementary Figure S1 - Visualization of the 100x random split-cross validation, including a second cross validation within the training set for the model optimization.



Supplementary Figure S2 - Flowchart of patient inclusion.

Supplemental Tables

Supplementary Table S1 - LIDC Nodule Characteristics, Definitions, and Ratings [1]

Characteristic	Ratings		Description
Calcification	1	Popcorn	Calcification appearance in the nodule - the smaller the nodule, the more likely it must contain calcium in order to be visualized. Benignity is highly associated with central, non-central, laminated, and popcorn calcification
(categorical)	2	Laminated	
	3	Solid	
	4	Non-central	
	5	Central	
	6	Absent	
Internal structure	1	Soft tissue	Expected internal composition of the nodule
(categorical)	2	Fluid	
	3	Fat	
	4	Air	
Lobulation	1	Marked	Whether a lobular shape is apparent from the margin or not - lobulated margin is an indication for benignity
(ordinal)	2	.	
	3	.	
	4	.	
	5	None	
Malignancy	1	Highly unlikely	Likelihood of malignancy of the nodule - malignancy is associated with large nodule size while small nodules are more likely to be benign. Most malignant nodules are non-calcified and have speculated margins.
(ordinal)	2	Moderately unlikely	
	3	Indeterminate	
	4	Moderately suspicious	
	5	Highly suspicious	
Margin	1	Poorly defined	How well defined the margins of the nodules are
(ordinal)	2	.	
	3	.	
	4	.	
	5	Sharp	
Sphericity	1	Linear	Dimensional shape of nodule in terms of roundness
(ordinal)	2	.	
	3	Ovoid	
	4	.	
	5	Round	
Spiculation	1	Marked	Degree to which the nodule exhibits spicules, spike-like structures, along its border - spiculated margin is an indication of malignancy
(ordinal)	2	.	
	3	.	
	4	.	
	5	None	
Subtlety	1	Extremely subtle	Difficulty in detection - refers to the contrast between the lung and its surroundings
(ordinal)	2	Moderately subtle	
	3	.	
	4	Fairly subtle	
	5	Obvious	
Texture	1	Nonsolid	Internal density of a nodule - texture plays an important role when attempting to segment a nodule, since part-solid and nonsolid texture can increase the difficulty of defining the nodule boundary
(ordinal)	2	.	
	3	Part-solid/mixed	
	4	.	
	5	Solid	

Supplementary Table S2 - LIDC criteria scored by a thorax radiologist

	BRAF Mutant (N=82 lesions)	*BRAF* wild type (N=87 lesions)
Calcification		
Popcorn		
Yes	0	0
No	82	87
Laminated		
Yes	0	0
No	82	87
Solid		
Yes	0	1
No	82	86
Non-central		
Yes	0	0
No	82	87
Central		
Yes	1	0
No	82	87
Absent		
Yes	75	80
No	7	7
Internal structure		
Soft tissue		
Yes	75	81
No	7	6
Fluid		
Yes	0	0
No	82	87
Fat		
Yes	0	0
No	82	87
Air		
Yes	1	1
No	81	86
Lobulation (ordinal)		
1 Marked	10	7
2	1	0
3	4	5
4	20	26
5 None	47	49
Malignancy		
Highly unlikely	8	5
Moderate unlikely	2	0
Indeterminate	0	1
Moderately suspicious	1	1
Highly suspicious	71	80
Margin (ordinal)		
1 Poorly defined	8	5
2	3	1
3	12	11
4	4	12

Supplementary Table S2 - Continued

	***BRAF* Mutant (N=82 lesions)**	***BRAF* wild type (N=87 lesions)**
5 Sharp	55	58
Sphericity (ordinal)		
1 Linear	9	7
2	3	2
3 Ovoid	33	28
4	20	25
5 Round	17	25
Spiculation (ordinal)		
1 Marked	8	6
2	2	1
3	1	2
4	6	6
5 None	65	72
Subtlety		
1 Extremely subtle	7	5
2 Moderately subtle	0	0
3	0	0
4 Fairly subtle	0	1
5 Obvious	75	81
Texture		
1 Nonsolid	10	5
2	0	0
3 Part-solid/mixed	0	0
4	0	0
5 Solid	72	82

Supplementary Table S3. Overview of the 540 features used in this study. Gray Level Co-occurrence Matrix (GLCM) features were calculated in four different directions (0, 45, 90, 135 degrees) using 16 gray levels and pixel distances of 1 and 3. Local Binary Pattern (LBP) features were calculated using the following three parameter combinations: 1 pixel radius and 8 neighbours, 2 pixel radius and 12 neighbours, and 3 pixel radius and 16 neighbours. Gabor features were calculated using three different frequencies (0.05, 0.2, 0.5) and four different angles (0, 45, 90, 135 degrees). Laplacian of Gaussian (LoG) features were calculated using three different widths of the Gaussian (1, 5 and 10 pixels). Vessel features were calculated using the full mask, the edge, and the inner region. Local phase features were calculated on the monogenic phase, phase congruency and phase symmetry.

Supplementary Table S3 - Overview of the 540 features used in this study

Histogram (13 features)	LoG (12*3=36 features)	Vessel (12*3=36 features)	GLCM (MS) (6*3*4*2=144 features)	Gabor (12*4*3=144 features)	NGTDM (5 features)	LBP (12*3=36 features)
min	min	min	contrast (normal, MS mean + std)	min	busyness	min
max	max	max	dissimilarity (normal, MS mean + std)	max	coarseness	max
mean	mean	mean	homogeneity(normal, MS mean + std)	mean	complexity	mean
median	median	median	angular second moment (ASM) (normal, MS mean + std)	median	contrast	median
std	std	std		std	strength	std
skewness	skewness	skewness	energy (normal, MS mean + std)	skewness		skewness
kurtosis	kurtosis	kurtosis	correlation (normal, MS mean + std)	kurtosis		kurtosis
peak	peak	peak		peak		peak
peak position	range	range		range		range
range	energy	energy		energy		energy
energy	quartile	quartile		quartile range		quartile range
quartile range	entropy	entropy		entropy		entropy
entropy						

GLSZM (16 features)	GLRM (16 features)	GLDM (14 features)	Shape (35 features)	Orientation (9 features)	Local phase (12*3=36 features)
Gray Level Non Uniformity	Gray Level Non Uniformity	Dependence Entropy	compactness (mean + std)	theta_x	min
Gray Level Non Uniformity Normalized	Gray Level Non Uniformity Normalized	Dependence Non-Uniformity	radial distance (mean + std)	theta_y	max
Gray Level Variance	Gray Level Variance	Dependence Non-Uniformity Normalized	roughness (mean + std)	theta_z	mean
High Gray Level Zone Emphasis	High Gray Level Run Emphasis	Dependence Variance	convexity (mean + std)	COM index x	median
Large Area Emphasis	Long Run Emphasis	Gray Level Non-Uniformity	circular variance (mean + std)	COM index y	std
Large Area High Gray Level Emphasis	Long Run High Gray Level Emphasis	Gray Level Variance	principal axes ratio (mean + std)	COM index z	skewness
Large Area Low Gray Level Emphasis	Long Run Low Gray Level Emphasis	High Gray Level Emphasis	elliptic variance (mean + std)	COM x	kurtosis
Low Gray Level Zone Emphasis	Low Gray Level Run Emphasis	Large Dependence Emphasis	solidity (mean + std)	COM y	peak
SizeZoneNonUniformity	RunEntropy	Large Dependence High Gray Level Emphasis	area (mean, std, min + max	COM z	range
SizeZoneNonUniformityNormalized	RunLengthNonUniformity	Large Dependence Low Gray Level Emphasis	volume (total, mesh, volume)		energy
SmallAreaEmphasis	RunLengthNonUniformityNormalized	Low Gray Level Emphasis	elongation		quartile
SmallAreaHighGrayLevelEmphasis	RunPercentage	Small Dependence Emphasis	flatness		entropy
SmallAreaLowGrayLevelEmphasis	RunVariance	Small Dependence High Gray Level Emphasis	least axis length		
ZoneEntropy	ShortRunEmphasis	Small Dependence Low Gray Level Emphasis	major axis length		
ZonePercentage	ShortRunHighGrayLevelEmphasis		minor axis length		
ZoneVariance	ShortRunLowGrayLevelEmphasis		maximum diameter 3D		
			maximum diameter 2D (rows, columns, slices)		
			sphericity		
			surface area		
			surface volume ratio		

*Abbreviations: COM: center of mass; GLCM: gray level co-occurrence matrix; MS: multi slice; NGTDM: neighborhood gray tone difference matrix; GLSZM: gray level size zone matrix; GLRLM: gray level run length matrix; LBP: local binary patterns; LoG: Laplacian of Gaussian; std: standard deviation.

References

1. MacKie, R.M. *et al.* Epidemiology of invasive cutaneous melanoma. *Ann Oncol* **20 Suppl 6**, vi1-7 (2009).
2. Shain, A.H. & Bastian, B.C. From melanocytes to melanomas. *Nat Rev Cancer* **16**, 345-58 (2016).
3. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *European Journal of Cancer* **103**, 356-387 (2018).
4. Whiteman, D.C. *et al.* The Growing Burden of Invasive Melanoma: Projections of Incidence Rates and Numbers of New Cases in Six Susceptible Populations through 2031. *J Invest Dermatol* **136**, 1161-1171 (2016).
5. Luke, J.J. *et al.* Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* **14**, 463-482 (2017).
6. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-54 (2002).
7. Curtin, J.A. *et al.* Distinct sets of genetic alterations in melanoma. *N Engl J Med* **353**, 2135-47 (2005).
8. Colombino, M. *et al.* BRAF/NRAS mutation frequencies among primary tumors and metastases in patients with melanoma. *J Clin Oncol* **30**, 2522-9 (2012).
9. Flaherty, K.T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* **363**, 809-19 (2010).
10. Long, G.V. *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med* **371**, 1877-88 (2014).
11. Long, G.V. *et al.* Dabrafenib plus trametinib versus dabrafenib monotherapy in patients with metastatic BRAF V600E/K-mutant melanoma: long-term survival and safety analysis of a phase 3 study. *Ann Oncol* **28**, 1631-1639 (2017).
12. Larkin, J. *et al.* Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med* **371**, 1867-76 (2014).
13. Ascierto, P.A. *et al.* Cobimetinib combined with vemurafenib in advanced BRAF(V600)-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *Lancet Oncol* **17**, 1248-60 (2016).
14. Michielin, O. *et al.* Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* **30**, 1884-1901 (2019).
15. Cheng, L. *et al.* Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine. *Mod Pathol* **31**, 24-38 (2018).
16. Rios Velazquez, E. *et al.* Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* **77**, 3922-3930 (2017).
17. Aerts, H.J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006 (2014).
18. Gevaert, O. *et al.* Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. *Radiology* **264**, 387-96 (2012).
19. Huang, Y. *et al.* Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* **281**, 947-957 (2016).
20. Nair, V.S. *et al.* Prognostic PET 18F-FDG uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer. *Cancer Res* **72**, 3725-34 (2012).
21. Parmar, C. *et al.* Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer. *Sci Rep* **5**, 11044 (2015).
22. Scrivener, M. *et al.* Radiomics applied to lung cancer: a review. *Translational Cancer Research* **5**, 398-409 (2016).

23. van Timmeren, J.E. *et al.* Longitudinal radiomics of cone-beam CT images from non-small cell lung cancer patients: Evaluation of the added prognostic value for overall survival and locoregional recurrence. *Radiother Oncol* **136**, 78-85 (2019).
24. Yamamoto, S. *et al.* ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. *Radiology* **272**, 568-76 (2014).
25. Yang, L. *et al.* Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? *Eur Radiol* **28**, 2058-2067 (2018).
26. Trebeschi, S. *et al.* Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* **30**, 998-1004 (2019).
27. Eisenhauer, E.A. *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-47 (2009).
28. Menzer, C. *et al.* Targeted Therapy in Advanced Melanoma With Rare BRAF Mutations. *J Clin Oncol* **37**, 3142-3151 (2019).
29. Starmans, M.P. *et al. Classification of malignant and benign liver tumors using a radiomics approach*, (SPIE, 2018).
30. Starmans, M. Workflow for Optimal Radiomics Classification (WORC). Available via https://github.com/MStarmans91/WORC. Accessed 21-4-2020. DOI: 10.5281/zenodo.3840534. (2018).
31. Starmans, M. *et al.* Fully automatic construction of optimal radiomics workflows. . *ECR 2019: Book of Abstracts Insights into Imaging 10:22* (2019).
32. Vos, M. *et al.* Radiomics approach to distinguish between well differentiated liposarcomas and lipomas on MRI. *Br J Surg* **106**, 1800-1809 (2019).
33. Github URL omitted for anonymization purposes. Published 2020. Accessed 2020-12-29. DOI will be added on final release
34. Opulencia, P. *et al.* Mapping LIDC, RadLex, and lung nodule image features. *J Digit Imaging* **24**, 256-70 (2011).
35. Nadeau, C. & Bengio, Y. Inference for the Generalization Error. *Machine Learning* **52**, 239-281 (2003).
36. Macskassy, S.A. *et al.* ROC confidence bands: an empirical evaluation. in *Proceedings of the 22nd international conference on Machine learning* (Association for Computing Machinery, Bonn, Germany, 2005).
37. Robert, C. *et al.* Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol* **20**, 1239-1251 (2019).
38. Durot, C. *et al.* Metastatic melanoma: pretreatment contrast-enhanced CT texture parameters as predictive biomarkers of survival in patients treated with pembrolizumab. *Eur Radiol* **29**, 3183-3191 (2019).
39. Sun, R. *et al.* A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* **19**, 1180-1191 (2018).
40. Saadani, H. *et al.* Metabolic Biomarker-Based BRAFV600 Mutation Association and Prediction in Melanoma. *J Nucl Med* **60**, 1545-1552 (2019).
41. Buvat, I. & Orlhac, F. The Dark Side of Radiomics: On the Paramount Importance of Publishing Negative Results. *J Nucl Med* **60**, 1543-1544 (2019).
42. Timbergen, M.J.M. *et al.* Differential diagnosis and mutation stratification of desmoid-type fibromatosis on MRI using radiomics. *European Journal of Radiology*, 109266 (2020).
43. Timbergen, M.J. *et al.* Annals of Oncology (2019) 30 (suppl_5): v683-v709. 10.1093/annonc/mdz283.
44. Starmans, M.P.A. *et al.* Prediction of histopathological growth patterns by radiomics and CT-imaging in

patients with operable colorectal liver metastases: a proof-of-concept study. ECR 2020 Book of Abstracts. *Insights into Imaging* **2020;11(1):34**(2020).

45. Castillo, T.J. *et al. Classification of prostate cancer: high grade versus low grade using a radiomics approach. IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) 1319-1322*, (2019).
46. Starmans, M.P.A. *et al.* Prediction of surgery requirement in mesenteric fibrosis on CT using a radiomics approach. European Conference of Radiology (ECR) Book of Abstracts (SS 1401b), B-1279, Vienna, Austria. . (2019).
47. Manca, A. *et al.* Mutational concordance between primary and metastatic melanoma: a next-generation sequencing approach. *J Transl Med* **17**, 289 (2019).
48. Riveiro-Falkenbach, E. *et al.* Intra- and Inter-Tumoral Homogeneity of BRAF(V600E) Mutations in Melanoma Tumors. *J Invest Dermatol* **135**, 3078-3085 (2015).
49. Valachis, A. & Ullenhag, G.J. Discrepancy in BRAF status among patients with metastatic malignant melanoma: A meta-analysis. *Eur J Cancer* **81**, 106-115 (2017).

Supplementary Materials

Supplementary Materials 1: Radiomics feature extraction

This supplementary material is similar to[1], but details relevant for the current study are highlighted.

A total of 540 radiomics features were used in this study. All features were extracted using the defaults for CT scans from the Workflow for Optimal Radiomics Classification (WORC) toolbox[2], which internally uses the PREDICT[3] and PyRadiomics[4] feature extraction toolboxes. For CT scans, the images are not normalized as the scans already have a fixed unit and scale (i.e. Hounsfield), contrary to MRI. The code to extract the features for this specific study has been published open-source[5]. An overview of all features is depicted in **Supplementary Table S3**. For details on the mathematical formulation of the features, we refer the reader to Zwanenburg *et al.* (2020)[6]. More details on the extracted features can be found in the documentation of the PREDICT, PyRadiomics, and mainly the WORC documentation[7].

The features can be divided in several groups. Thirteen intensity features were extracted using the histogram of all intensity values within the ROIs and included several first-order statistics such as the mean, standard deviation and kurtosis. These describe the distribution of Hounsfield units within the lesion. Thirty-five shape features were extracted based only on the ROI, i.e. not using the image, and included shape descriptions such as the volume, compactness and circular variance. These describe the morphological properties of the lesion. Nine orientation features were used, describing the orientation of the ROI, i.e. not using the image. Lastly, 483 texture features were extracted using Gabor filters (144 features), Laplacian of Gaussian filters (36 features), vessel (i.e. tubular structures) filters (36 features)[8], the Gray Level Co-occurrence Matrix (144 features)[6], the Gray Level Size Zone Matrix (16 features)[6], the Gray Level Run Length Matrix (16 features)[6], the Gray Level Dependence Matrix (14 features)[6], the Neighbourhood Grey Tone Difference Matrix (5 features)[6], Local Binary Patterns (18 features)[9], and local phase filters (36 features)[10]. These features describe more complex patterns within the lesion, such as heterogeneity, occurrence of blob-like structures, and presence of line patterns.

Supplementary Materials 2: Model optimization

This appendix is similar to[1], but details relevant for the current study are highlighted. The Workflow for Optimal Radiomics Classification (WORC) toolbox[2] makes use of adaptive algorithm optimization to create the optimal performing workflow from a

variety of methods. WORC defines a workflow as a sequential combination of algorithms and their respective parameters. To create a workflow, WORC includes algorithms to perform feature scaling, feature imputation, feature selection, oversampling, and machine learning. If used, as some of these steps are optional as described below, these methods are performed in the same order as described in this appendix. More details can be found in the WORC documentation[7]. The code to use WORC for creating the *BRAF* decision models in this specific study has been published open-source[5].

When a feature could not be computed, e.g. the lesion is too small or a division by zero occurs, feature imputation was used to estimate replacement values for the missing values. Strategies for imputation included 1) the mean; 2) the median; 3) the most frequent value; and 4) a nearest neighbor approach.
Feature scaling was performed to make all features have the same scale, as otherwise the machine learning methods may focus only on those features with large values. This was done through z-scoring, i.e. subtracting the mean value followed by division by the standard deviation, for each individual feature. In this way, all features had a mean of zero and a variance of one. A robust version of z-scoring was used, in which outliers, i.e. values below the 5^{th} percentile or above the 95^{th} percentile, are excluded from computing the mean and variance.

Feature selection was performed to eliminate features which were not useful to distinguish between the classes, i.e. *BRAF* mutant vs. *BRAF* wild-type. These included; 1) a variance threshold, in which features with a low variance (<0.01) are removed. This method was always used, as this serves as a feature sanity check with almost zero risk of removing relevant features; 2) optionally, a group-wise search, in which specific groups of features (i.e. intensity, shape, and the subgroups of texture features as defined in Supplementary Materials 1) are selected or deleted. To this end, each feature group has an on/off variable which is randomly activated or deactivated, which were all included as hyperparameters in the optimization; 3) optionally, individual feature selection through univariate testing. To this end, for each feature, a Mann-Whitney U test is performed to test for significant differences in distribution between the labels (e.g. *BRAF* mutant vs *BRAF* wild-type). Afterwards, only features with a p-value above a certain threshold are selected. A Mann-Whitney U test was chosen as features may not be normally distributed and the samples (i.e. lesions) were independent; and 4) optionally, principal component analysis (PCA), in which either only those linear combinations of features were kept which explained 95% of the variance in the features or a limited amount of components (between 10 – 50). These feature selection methods may be combined by WORC, but only in the mentioned order.

Oversampling was used to make sure the classes were balanced in the training dataset. These included; 1) random oversampling, which randomly repeats patients of the minority class; and 2) the synthetic minority oversampling technique (SMOTE)[11], which creates new synthetic "lesions" using a combination of the features in the minority class. Randomly, either one of these methods or no oversampling method was used.
Lastly, machine learning methods were used to determine a decision rule to distinguish the classes. These included; 1) logistic regression; 2) support vector machines; 3) random forests; 4) naive Bayes; and 5) linear and quadratic discriminant analysis.

Most of the included methods require specific settings or parameters to be set, which may have a large impact on the performance. As these parameters have to be determined before executing the workflow, these are so-called "hyperparameters". In WORC, all parameters of all mentioned methods are treated as hyperparameters, since they may all influence the decision model creation. WORC simultaneously estimates which combination of algorithms and hyperparameters performs best. A comprehensive overview of all parameters is provided in the WORC documentation[7].

By default in WORC, the performance is evaluated in a 100x random-split train-test cross-validation. In the training phase, a total of 100,000 pseudo-randomly generated workflows is created. These workflows are evaluated in a 5x random-split cross-validation on the training dataset, using 85% of the data for actual training and 15% for validation of the performance. All described methods were fit on the training datasets, and only tested on the validation datasets. The workflows are ranked from best to worst based on their mean performance on the validation sets using the F1-score, which is the harmonic average of precision and recall. Due to the large number of workflows executed, there is a chance that the best performing workflow is overfitting, i.e. looking at too much detail or even noise in the training dataset. Hence, to create a more robust model and boost performance, WORC combines the 50 best performing workflows into a single decision model, which is known as ensembling. These 50 best performing workflows are re-trained using the entire training dataset, and only tested on the test dataset. The ensemble is created through averaging of the probabilities, i.e. the chance of a lesion being *BRAF* mutant or *BRAF* wild-type, of these 50 workflows.
The code for the model creation, including more details, has been published open-source[5].

Supplementary References

1. Anonymized. (2019).
2. Starmans, M.P.A. Workflow for Optimal Radiomics Classification (WORC). (2018).
3. van der Voort, S.R. & Starmans, M.P.A. Predict a Radiomics Extensive Differentiable Interchangable Classification Toolkit (PREDICT). (2018).
4. van Griethuysen, J.J.M. *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research* **77**, e104-e107 (2017).
5. Anonymized. MelaRadiomics. (2020).
6. Zwanenburg, A. *et al.* The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **295**, 191145 (2020).
7. Starmans, M.P.A. Workflow for Optimal Radiomics Classification (WORC) Documentation. (2018).
8. Frangi, A.F. *et al.* Multiscale vessel enhancement filtering. in *International conference on medical image computing and computer-assisted intervention* 130-137 (Springer, 1998).
9. Ojala, T. *et al.* Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, 971-987 (2002).
10. Kovesi, P. Phase congruency detects corners and edges. in *The Australian pattern recognition society conference: DICTA* (2003).
11. Han, H. *et al.* Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. 878-887 (2005).

PART **IV**

Integration of modalities

CHAPTER 12

Non-invasive estrogen receptor assessment by [^{18}F]-fluorestradiol (^{18}F-FES) PET/CT or circulating tumor cells predicts estrogen receptor status in patients with metastatic breast cancer

In Preparation

Bertha Eisses*, **Lindsay Angus***, Sjoerd G. Elias, Carla M.L. van Herpen, C. Willemien Menke-van der Houven van Oordt, Bert van der Vegt, Andor W.J.M. Glaudemans, Adrienne H. Brouwers, Otto S. Hoekstra, Wim J.G. Oyen, Jasper Emmering, Sarah Verhoeff, Sophie L. Gerritse, Anieta M. Sieuwerts†, Agnes Jager, Jaco Kraan, John W.M. Martens, Stefan Sleijfer, Elisabeth G.E. de Vries, Carolina P. Schröder

* Both authors contributed equally to this work

† Deceased August 2019

Abstract

Purpose

Estrogen receptor (ER) expression largely determines the choice of therapy for patients with metastatic breast cancer (MBC). ER is assessed using immunohistochemistry (IHC) on metastatic tissue. Since a biopsy is cumbersome and sometimes not feasible, [^{18}F]-fluorestradiol (^{18}F-FES) PET/CT or circulating tumor cells (CTCs) might be alternatives. We hypothesized that ^{18}F-FES PET/CT and/or CTCs can determine ER-status in patients with MBC in a more patient-friendly way than a biopsy. Therefore, the purpose of this prospective multicenter study was to assess the predictive value of baseline ^{18}F-FES PET/CT or CTCs using ER expression of the biopsied metastatic lesion as a reference.

Patients and methods

Patients with newly diagnosed non-rapidly progressive MBC, regardless of subtype, underwent a metastasis biopsy, a ^{18}F-FES PET/CT and CTC assessment before start of first-line systemic therapy (IMPACT breast NCT01832051). ER expression in ≥10% of tumor cells was considered positive. The ^{18}F-FES PET/CT was defined positive when ≥ 1 lesion had a maximum standardized uptake value of ≥1.5. CTC ER RNA expression was determined with *ESR1* gene expression analysis of CellSearch-enriched CTCs. Reverse transcription polymerase chain reaction served to quantify the *ESR1* expression in CTCs. ER positivity was defined as an *ESR1* mRNA ΔCq level ≥-7.86, corrected for background healthy donor blood signal. CTC data was analyzed with and without a threshold of 5 CTCs. Positive and negative predictive values were calculated (PPV and NPV) using the ER-status of the biopsy as reference. Patient-reported outcomes (PROs) were measured directly after baseline biopsy and ^{18}F-FES PET/CT on a visual analogous scale. A higher score was considered as less patient-friendly.

Results

In 180 out of 201 patients, an evaluable baseline biopsy, blood sample for CTC analysis, and ^{18}F-FES PET/CT were available. ER on ^{18}F-FES PET/CT and biopsy were concordant in 159/180 (88.3%) patients with a PPV of 91.7% and a NPV of 79.2% of the ^{18}F-FES PET/CT to predict biopsy ER. In 93/180 patients (52%) CTC *ESR1* RNA expression could be assessed. In this subgroup, CTCs and biopsy were concordant in 71 cases (76.3% of 93; 39.4% of total 180), with a PPV and NPV of CTCs to predict biopsy ER of 81.1% and 57.9%, respectively. In patients with ≥5 CTCs (56/180 patients total), ER in CTCs and biopsy were concordant in 46 cases (82.1% of 56; 25.6% of total 180), with a PPV and NPV to predict biopsy ER status of 87.2% and 55.6%, respectively. The PROs, evaluable in 152 patients, of the ^{18}F-FES PET/CT were significantly lower than of the baseline

biopsy (P <0.0001).

Conclusions

In patients with non-rapidly progressive MBC, baseline ^{18}F-FES PET/CT could predict metastasis ER-status, which is considered more patient friendly than a baseline biopsy. In only 52% of patients, CTCs could also predict metastasis ER expression.

Introduction

Breast cancer is the most common cancer in women worldwide and the breast cancer incidence is still increasing[1]. Despite advances in neoadjuvant and adjuvant therapies for primary breast cancer, metastatic disease still develops in 20-30% of the patients with early breast cancer, while 5% of the patients presents with metastatic breast cancer (MBC) at initial diagnosis[2]. Once breast cancer has metastasized, molecular characteristics such as estrogen receptor (ER) status, progesterone receptor (PR) status, and human epidermal growth factor receptor (HER2) are essential parameters for treatment decision-making. Importantly, receptor conversion and heterogeneity can occur during the course of disease[3,4]. Hence, a metastasis biopsy is considered the gold standard to obtain current information on the receptor status[3,5,6].

Although receptor conversion is a well-known phenomenon, clinicians may refrain from taking a biopsy because of the invasive nature of this procedure and taking multiple biopsies over time is often not feasible. Even if a biopsy of a metastasis is feasible, the tissue only reflects a small piece of a single lesion and will not show the potential heterogeneity of the ER and HER2 status between[3,7] and within metastatic lesions[8]. Therefore, techniques to obtain accurate whole-body information are currently under investigation. To determine the ER status in metastases throughout the body, the 16α-[^{18}F]-fluoro-17β-estradiol (^{18}F-FES) tracer has been used to visualize the ER expression by positron emission tomography (PET). This is potentially a more patient-friendly way to visualize the ER status than by taking a biopsy[9,10]. Moreover, the ^{18}F-FES PET/CT allows for visualization of the ER status of all metastatic sites whereas a tissue biopsy only reflects a part of single lesion. Next to the ^{18}F-FES PET/CT, another easily accessible way to determine the expression of ER is the characterization of circulating tumor cells (CTCs)[11,12]. CTC count at diagnosis and changes during treatment have prognostic value[13,14]. In addition to CTC enumeration, these CTCs can be enriched from whole blood for molecular analysis such as *ESR1* RNA expression. Since CTCs likely originate from all metastases, they might provide a comprehensive view on tumor characteristics such as ER status.

We hypothesized that the ^{18}F-FES PET/CT and *ESR1* RNA expression measured in CTCs can determine the ER status in patients with MBC in a more patient-friendly way than a biopsy. Therefore, we investigated in this prospective multicenter study the predictive value of baseline ^{18}F-FES PET/CT or CTCs for ER immunohistochemistry (IHC) of the metastasis biopsy. Moreover, we assessed the concordance between the ER status measured by ^{18}F-FES PET/CT, *ESR1* RNA expression in CTCs and ER IHC of a metastasis biopsy in patients with newly diagnosed MBC before the start of first-line systemic therapy.

Patients and methods

Study design and patients

This is a Dutch multicenter prospective observational cohort study in patients with non-rapidly progressive MBC who were eligible for first-line systemic therapy (IMPACT-MBC trial; NCT01832051). Non-rapidly progressive MBC was defined as not requiring urgent initiation of chemotherapy, based on the discretion of the oncologist which included no recent worsening of MBC related signs and symptoms and in case of liver metastases no significant increase in plasma liver enzymes (ASAT and ALAT) 2 weeks prior to the screenings visit. Patients with MBC at first presentation, regardless of breast cancer subtype, underwent extensive molecular imaging including ^{18}F-FES PET/CT, [^{89}Zr]-trastuzumab PET/CT and serial [^{18}F]-fluorodeoxyglucose (FDG) PET/CT, a metastasis biopsy and blood sampling to obtain a whole body molecular profile of their disease (**Fig. 1**). This multicenter study, conducted at the University Medical Center Groningen (UMCG), Amsterdam University Medical Center (UMC), location VU Medical Center (VUMC), Radboud University Medical Center (Radboud UMC), and Erasmus Medical Center (Erasmus MC), was performed in accordance with the Declaration of Helsinki and centrally approved by the Medical Research Ethics Committee of the UMCG, Groningen, The Netherlands. All patients gave written informed consent prior to study procedures.

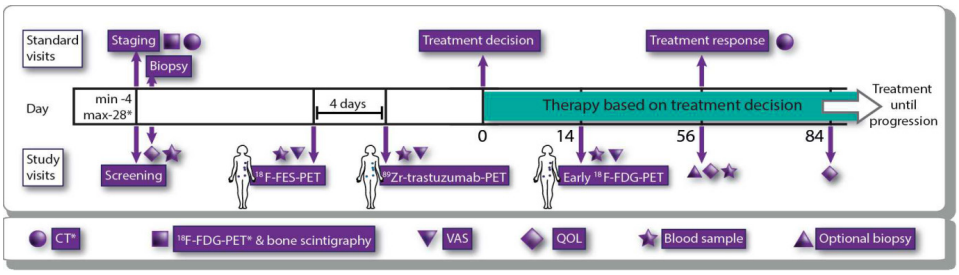


Figure 1 - Study flowchart of part of the measurements in the IMPACT-MBC study. During the screening period patients underwent a ^{18}F-FES PET/CT, a ^{89}Zr-trastuzumab-PET,a biopsy of a metastatic lesion and blood sampling for CTC enumeration and characterization.

Tumor histology

All patients underwent a biopsy of a metastatic lesion at baseline before molecular imaging. This biopsy was subsequently paraffin-embedded and stained for standard histology and ER, PR and HER2 as part of standard clinical work-up. A pathologist centrally reviewed all biopsies at the UMCG. The ER status was evaluated by IHC

(CONFIRM anti-Estrogen Receptor (ER) (SP1) Rabbit Monoclonal Primary Antibody, Ventana / Roche, Illkirch, France). The antibody was pre-diluted by the supplier and stained on an automated staining platform (Benchmark Ultra, Ventana / Roche, Illkirch, France) and considered positive if at least 10% of the tumor cells showed ER expression of any intensity[15]. HER2 (SP3, Thermo Fisher Scientific, pre-diluted by supplier, stained on Benchmark Ultra, Ventana / Roche, Illkirch, France) was scored according to the guidelines of the ASCO/CAP[16].

Molecular imaging: ^{18}F-FES PET/CT

All patients underwent a ^{18}F-FES PET/CT at baseline. ^{18}F-FES has been produced as described previously[17]. Patients received a single intravenous bolus injection of 200 MBq (± 10%) ^{18}F-FES before PET/CT[18]. Whole-body imaging was performed 60 minutes (± 5 minutes) after tracer injection. Scanning time was dependent on body weight of the patient and administered dose of ^{18}F-FES[19]. In general, the same guidelines as for ^{18}F-FDG PET/CT imaging have been followed[20]. To prevent false-negative results, patients had to stop with adjuvant ER antagonists (such as tamoxifen) at least 5 weeks before the ^{18}F-FES PET/CT. Aromatase inhibitors could be continued. During the waiting period between ^{18}F-FES injection and PET scan, patients were allowed to move and talk. Also, fasting before ^{18}F-FES PET/CT imaging was not necessary. The ^{18}F-FES PET/CT was considered positive if at least one tumor area showed ER expression with a maximum standard uptake (SUV_{max}) ≥1.5. PET scans were acquired using a Biograph mCT 64-slice PET/CT camera or a Biograph mCT 40-slice PET/CT camera (Siemens) at the UMCG, Radboud UMC and Erasmus MC. PET scans at the VUMC were acquired using an Ingenuity TF or Gemini TF PET/CT system (both Philips). All scans were reconstructed according to EANM/EARL protocols for SUV calculations[20]. All scans were combined with a low-dose CT for attenuation correction and anatomic reference.

Isolation of RNA from enriched CTCs and determination of ESR1 expression

In addition to molecular imaging, 20 mL of blood was drawn; 10mL in CellSave tubes (Menarini-Silicon Biosystems, Huntingdon Valley, PA, USA) for CTC enumeration and 10 mL in Vacutainer® EDTA tubes (BD, Franklin Lakes, NJ, USA) for CTC characterization. Blood tubes were subsequently transported by a courier to the laboratory of Translational Cancer Genomics and Proteomics, Erasmus MC Cancer Institute, Rotterdam, The Netherlands. A detailed description of the CTC enumeration and RNA isolation from CTCs has been reported previously[21-24]. CTC enumeration was performed on 7.5 mL CellSave blood within 96 hours after blood draw, using the CellSearch System (Menarini-Silicon Biosystems). For ER characterization, CTCs were isolated from 7.5 mL EDTA blood within 24 hours after blood draw using the CellSearch system

with the CellSearch profile kit (Menarini-Silicon Biosystems). After CTC enrichment, RNA was isolated from isolated CTCs using the AllPrep DNA/RNA Microkit (Qiagen, Germantown, MD, USA). Subsequently, cDNA was generated, pre-amplified for targets of interest, amongst which was *ESR1*, and real-time amplified (RT-qPCR) using Taqman Gene Expression Assays (Applied Biosystems, Carlsbad, CA, USA).

For *ESR1* RNA expression analysis we firstly analyzed those samples with ≥5 CTCs in the CTC enumeration tube, sufficient mRNA signal (average Cq<26.5 of reference genes glucuronidase beta (*GUSB*), hydroxymethylbilane synthase (*HMBS*) and hypoxanthine phosphoribosyltransferase 1 (*HPRT1*)) and sufficient epithelial signal (keratin 19 (*KRT19*)/ epithelial cell adhesion molecule (*EPCAM*) average Cq<26.5), as described previously[12,22]. ER-positivity was defined as an *ESR1* mRNA ΔCq ≥ -7.86, based on previous results[22]. Secondly, we analyzed all samples with sufficient mRNA signal (reference genes average ΔCq<26.5) and epithelial signal (*KRT19/EPCAM* average ΔCq<26.5), regardless of the CTC count in the enumeration tube. The cut-off for ER-positivity was similar in both analyses.

Patient-reported outcome

Directly after the baseline biopsy and ^{18}F-FES PET/CT procedure, patients were asked to evaluate the impact of the procedures on a visual analogous scale (scale 0 – 100). A higher score was considered as less patient friendly.

Statistical analyses

Concordance between ^{18}F-FES PET/CT and ER IHC and CTC *ESR1* RNA expression and ER IHC was calculated by the following formula: percentage concordance= $\frac{\textit{Number of concordant pairs}}{\textit{Total number of pairs}}$. For example, the concordance of ^{18}F-FES PET/CT and ER IHC was calculated as follows: $\frac{\text{(IHC ER and FES–PET positive)+(IHC ER and FES–PET negative)}}{\textit{All patients}}$. We used the kappa statistic to test the agreement between the ER measurements by IHC, ^{18}F-FES PET/CT and CTC *ESR1* RNA expression. Since ER expression measured by IHC is clinically used as gold standard, we calculated the sensitivity, specificity, positive (PPV) and negative predictive values (NPV) for ^{18}F-FES PET/CT and CTC *ESR1* RNA expression using ER IHC as reference. Additionally, to compare the patient reported outcomes directly obtained after biopsy and after ^{18}F-FES PET/CT we used a non-parametric test for related samples. All statistical tests were two-sided and considered statistically significant at P <0.05.

Results

Patient population

Of the 217 patients included between August 2013 and May 2018, 180 patients had an evaluable biopsy, ^{18}F-FES PET/CT and blood sample for CTC analysis (**Fig. 2**). Detailed baseline clinical characteristics for the evaluable 180 patients are presented in **Table 1**.

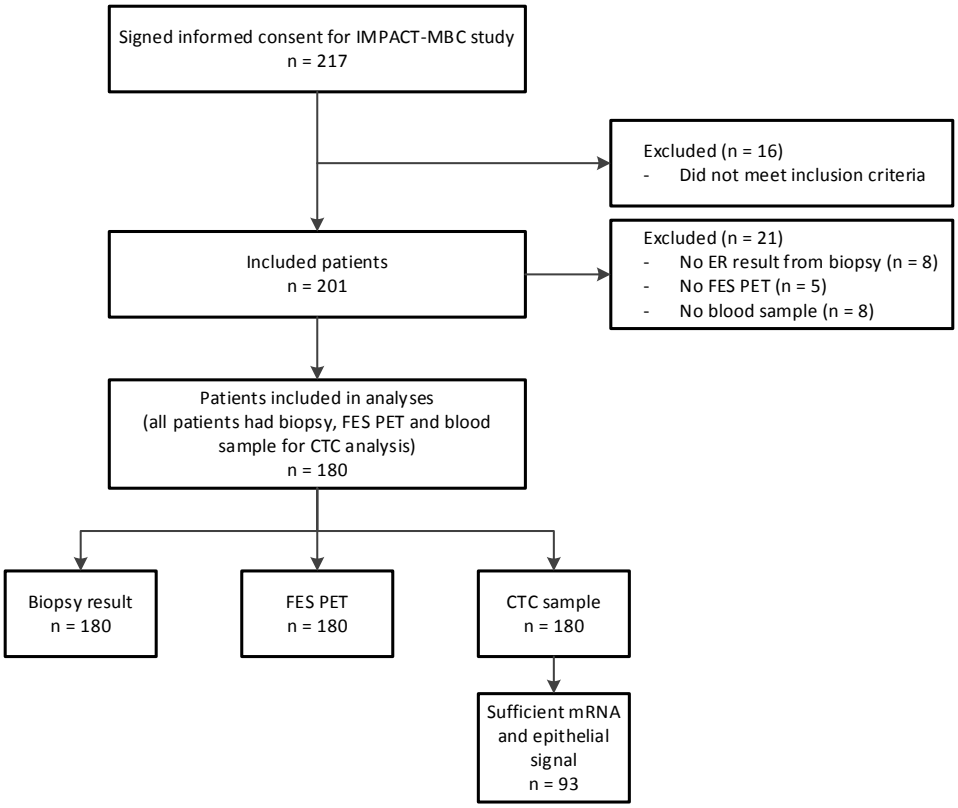


Figure 2 - Study flowchart showing all included and evaluable patients.

Table 1 - Clinical data of eligible patients (n=180)

	N	%
Median age (range)	61.5 (32 - 84)	-
Menopausal status		
Premenopausal	21	11.7
Surgically sterilized	9	5.0
Postmenopausal	124	68.9
Not applicable (male)	2	1.1
Unknown	24	13.3
Receptor status primary tumor		
ER-positive/HER2-negative	125	69.4
ER-positive/HER2-positive	18	10.0
ER-positive/HER2-unknown	3	1.7
ER-negative/HER2-negative	17	9.4
ER-negative/HER2-positive	16	8.9
Unknown	1	0.6
(Neo)adjuvant therapy		
None	52	28.9
Endocrine only	23	12.8
Chemotherapy only	14	7.8
Chemotherapy and endocrine therapy	67	37.2
Chemotherapy and targeted therapy	8	4.4
Chemo, endocrine and targeted therapy	16	8.9
Biopsy site metastasis		
Lymph node	43	23.9
Bone	86	47.8
Liver	24	13.3
Lung	9	5.0
Soft tissue	4	2.2
Other	14	7.8
Receptor status metastatic biopsy		
ER-positive/HER2-negative	117	65.0
ER-positive/HER2-positive	13	7.2
ER-positive/HER2-unknown	1	0.6
ER-negative/HER2-negative	30	16.7
ER-negative/HER2-positive	18	10.0
ER-negative/HER2-unknown	1	0.6
Median CTC count (range)	2 (0 - 1377)	-

Predictive value of ^{18}F-FES PET/CT for ER IHC and concordance between ^{18}F-FES PET/CT and ER IHC

From 180 patients both ER IHC of the biopsy and ^{18}F-FES PET/CT could be assessed. The PPV and NPV of ^{18}F-FES PET/CT were 91.7% (95% CI: 85.6 – 95.8) and 79.2% (95% CI: 65.0 – 89.5), respectively (**Table 2a**, **Fig. 3a**). ER IHC of the biopsy was positive in 131 (72.8%) patients, the ^{18}F-FES PET/CT in 132 patients (73.3%) out of 180 patients. ER IHC and ^{18}F-FES PET/CT results were concordant in 159 (88.3%) patients (kappa 0.70, 95% CI: 0.59 – 0.82, P <0.001). From the 21 discordant pairs, 11 patients had a positive ^{18}F-FES PET/CT while the IHC was negative, whereas 10 patients had a positive IHC while the ^{18}F-FES PET/CT was negative.

Predictive value and concordance between ^{18}F-FES PET/CT, ESR1 RNA expression in CTCs and ER IHC

For the *ESR1* RNA expression analyses, *ESR1* RNA expression was first analyzed in those samples with ≥5 CTCs. Fifty-six of the 180 blood samples contained ≥5 CTCs and sufficient mRNA signal for both reference and epithelial genes to analyze the *ESR1* RNA expression. This allowed for 56 patients the combined ER assessment by IHC, ^{18}F-FES PET/CT and CTCs. ER-positive CTCs were present in 47 of 56 patients (83.9%). In this subgroup of 56 patients, the PPV and NPV for CTC ER status were 87.2% (95% CI: 74.3 – 95.2) and 55.6% (95% CI: 21.2 – 86.3), respectively. In 46 (82.1%) patients the ER status of CTCs and IHC showed concordant results (kappa 0.39; 95% CI 0.087 – 0.7, P = 0.003). Overall, the results of all three modalities were concordant in 45 of the 56 (80%) patients with ≥5 CTCs (**Table 2b**; **Fig. 3b**).

In addition, we analyzed all samples with sufficient epithelial signal, regardless of CTC count measured in the CTC enumeration tube. Thus, 93 patients could be evaluated for all three modalities. We identified 74 (79.6%) patients with *ESR1* RNA expression in the CTC-enriched fraction. In these 93 patients, the PPV and NPV for CTC ER status were 81.1% (95% CI: 70.3 – 89.3) and 57.9% (95% CI: 33.5 – 79.7) respectively. The ER status measured by IHC and in the CTC-enriched fraction was concordant in 71 (76.3%) patients (kappa 0.35, 95% CI 0.11 – 0.57, P = 0.001) (**Table 2c**). The concordance of all three modalities was 73.1% (**Fig. 3c**). An overview of discrepant results is provided in **Supplementary Table S1**.

Patient-reported outcome

From 152 patients, patient-reported outcomes (PROs) were available for baseline biopsy and ^{18}F-FES PET/CT. A lower PRO, measured on a scale from 0 – 100, indicated a more patient-friendly method to assess the ER status. The PROs of the ^{18}F-FES PET/CT (median 10, interquartile range (IQR: 2.3-20) was significantly lower than the PROs of the metastasis biopsy (median 20, IQR: 5-50), suggesting that the ^{18}F-FES PET/CT is a more patient friendly way to assess ER status than taking a biopsy (**Fig. 4**).

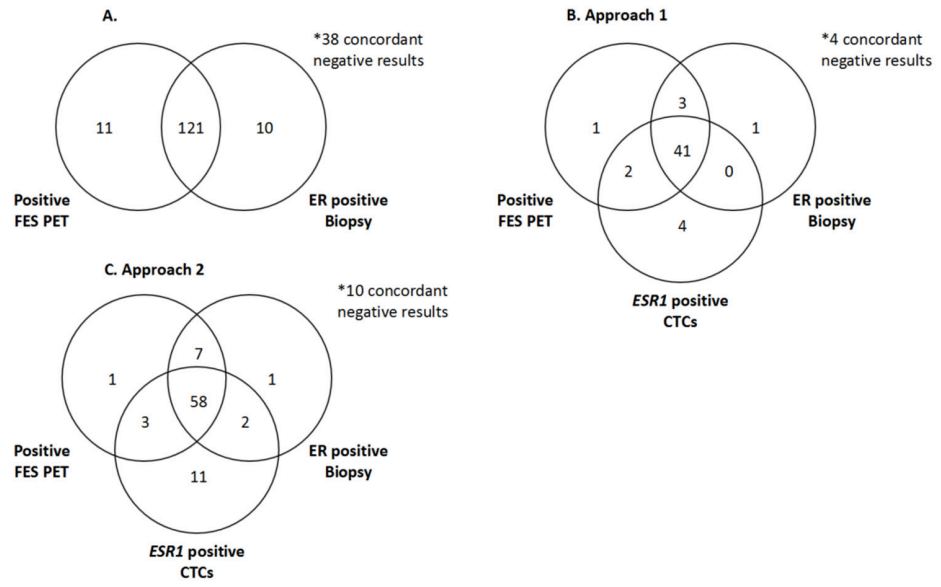


Figure 3 - Concordance between ER positivity measured by ^{18}F-FES PET/CT and IHC (A, n=180 patients), and between CTC *ESR1* gene expression, IHC and ^{18}F-FES PET/CT for approach 1 (B, n= 56 patients) and approach 2 (C, n= 93 patients).

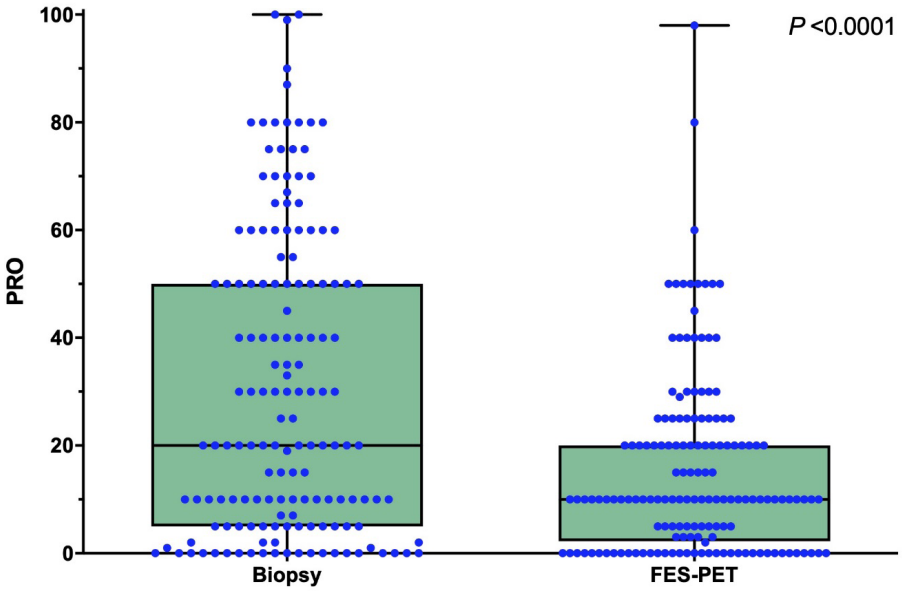


Figure 4 - Boxplots of patient reported outcomes (n= 152) directly measured after ^{18}F-FES PET/CT and biopsy. Each blue dot reflects one individual score; the green bar represents the 50% quartiles.

Table 2A - ^{18}F-FES PET/CT versus ER biopsy, N = 180

	Metastasis ER positive	ER negative	Sensitivity (95%CI)	PPV (95%CI)	Specificity (95%CI)	NPV (95%CI)	Kappa	SE	95% CI lower	95% CI upper	P-value
FES PET positive	121	11	92.4 (86.4 - 96,3)	91.7 (85.6-95.8)			0.70	0.06	0.59	0.82	<0.001
FES PET negative	10	38			77.6 (63.4-88.2)	79.2 (65.0-89.5)					

Table 2B - ER biopsy versus FES PET and CTC *ESR1* - Approach 1, N=56

	Metastasis ER positive	ER negative	Sensitivity (95%CI)	PPV (95%CI)	Specificity (95%CI)	NPV (95%CI)	Kappa	SE	95% CI lower	95% CI upper	P-value
FES PET positive	44	3	97.8 (88.2-99.9)	93.6 (82.5-98.7)			0.757	0.115	0.532	0.982	<0.001
FES PET negative	1	8			72.7 (39.0-94.0)	88.9 (51.8-99.7)					
CTC ER positive	41	6	91.1 (78.8-97.5)	87.2 (74.3-95.2)			0.393	0.156	0.087	0.699	0.003
CTC ER negative	4	5			45.5 (16.7-76,6)	55.6 (21.2-86.3)					

	^{18}F-FES PET/CT Positive	Negative	Kappa	SE	95% CI lower	95% CI upper	P-value
CTC ER positive	43	4	0.47	0.16	0.1564	0.7836	<0.001
CTC ER negative	4	5					

Table 2C - ER biopsy versus ^{18}F-FES PET/CT and CTC *ESR1* - Approach 2, N=93

	Metastasis ER positive	ER negative	Sensitivity (95%CI)	PPV (95%CI)	Specificity (95%CI)	NPV (95%CI)	Kappa	SE	95% CI lower	95% CI upper	P-value
FES PET positive	65	4	95.6 (87.6-99.1)	94.2 (85.8-98.4)			0.806	0.070	0.669	0.943	<0.001
FES PET negative	3	21			84.0 (63.9-95.5)	87.5 (67.6-97.3)					
CTC ER positive	60	14	88.2 (78.1-94.8)	81.1 (70.3-89.3)			0.349	0.110	0.133	0.565	0.001
CTC ER negative	8	11			44.0 (24.4-65.1)	57.9 (33.5-79.7)					

	^{18}F-FES PET/CT Positive	Negative	Kappa	SE	95% CI lower	95% CI upper	P-value
CTC ER positive	61	13	0.367	0.111	0.14944	0.58456	<0.001
CTC ER negative	8	11					

Discussion

This prospective cohort study shows that the metastasis ER-status could be predicted with baseline ^{18}F-FES PET/CT, with a PPV of 91.7% and NPV of 79.2%. Importantly, based on PROs, the ^{18}F-FES PET/CT is considered more patient friendly than a baseline biopsy. Moreover, in a subset of patients, *ESR1* RNA expression could be measured in CTCs to predict ER expression of the metastasis biopsy in a minimally-invasive manner. To our knowledge, this is the first study comprehensively assessing the predictive value of the ^{18}F-FES PET/CT and *ESR1* RNA expression in CTCs for ER status measured with IHC of a single metastasis per patient in a large cohort of patients with newly-diagnosed MBC.

With the cutoff for positivity of the ^{18}F-FES PET/CT, when at least one tumor area with a $SUV_{max} \geq 1.5$ was present, we showed a high sensitivity of 92.4% and moderate specificity of 77.6%. The sensitivity in our study is higher than the previously reported 76.6%[25]. This may be explained by the fact that we scored the ^{18}F-FES PET/CT positive when at least one tumor area showed uptake (not necessarily the biopsied lesion), whereas Chae *et al*[25] compared ^{18}F-FES uptake of the biopsied lesion with the ER expression of the biopsy. Moreover, the biopsy sites between both studies differed. In our study, all evaluable metastases were included for the ^{18}F-FES PET/CT analyses, including bone metastasis, whereas in the study of Chae *et al*.[25] bone and liver metastases were excluded. As patients with MBC frequently suffer from bone-only metastatic disease at first relapse[26], the inclusion of bone metastases in our analyses will provide a better reflection of the natural distribution of metastases in patients with MBC. Also, our approach will provide information from a 'disease' point of view, which was the initial goal of IMPACT-MBC. Nonetheless, uptake quantification of all metastatic lesions per patient is currently ongoing which will allow a more precise comparison of the two studies.

Although the concordance rate for ER IHC and ^{18}F-FES PET/CT is high, there were 21 patients with discordant results. Regarding false-negative ^{18}F-FES PET/CT results, 10 patients had an ER-positive biopsy but a negative ^{18}F-FES PET/CT. There are three possible explanations for this observation. Firstly, a false-negative ^{18}F-FES PET/CT might be a result of a low ER expression below the detection limit of the ^{18}F-FES PET/CT. Secondly, as in all nuclear imaging techniques, the detection of lesions depends on the number of ER+ tumor cells per unit of volume and the degree of ER expression on these cells, thus partially correlating with lesion size. Hence, false-negativity might occur in patients with small lesions with moderate ER expression. Thirdly, due to physiological

hepatobiliary uptake of the FES tracer and excretion into the gastrointestinal tract, liver metastases cannot be evaluated by ^{18}F-FES PET/CT[19]. In one of the patients with a false-negative ^{18}F-FES PET/CT, a liver biopsy indeed did show ER-positive disease.

Next to false-negative ^{18}F-FES PET/CTs, there were 11 patients with a positive ^{18}F-FES PET/CT but with an ER-negative biopsy. Tumor heterogeneity might have led to false-negative tissue biopsies, whereby different metastatic lesions or even single lesions do express ER heterogeneously.

Importantly, CTC results were only available for a subset of patients. However, if ER-positive CTCs are detected, ER-positive disease in at least a part of the metastases is very likely. Hence, if no CTCs or ER- negative CTCs are detected, MBC ER status needs to be further assessed by means of tissue confirmation or ^{18}F-FES PET/CT. Since CTC analysis can also be performed in a minimally invasive manner and at relatively low cost, this technique could be valuable for a subgroup of patients.

Discrepant results between *ESR1* RNA expression in CTCs and IHC of the biopsy were observed in 10 patients. For patients with *ESR1* RNA expression in CTCs but with an ER-negative biopsy, again tumor heterogeneity could be an explanation. Given the fact that the predictive value for response to endocrine therapy using ER IHC is around 50%[27], it will be a valid question to study whether patients with *ESR1* expressing CTCs but an ER-negative biopsy will benefit from endocrine therapies.

Regarding false-negative CTC results, two patients had expression levels just below the threshold used for positivity, suggesting that the threshold for positivity might be too strict. The other two false-negative results might also be explained by previously mentioned tumor heterogeneity or by the fact that the RNA expression measured in the CTC-enriched fraction is derived from CTCs as well as from healthy leukocytes[28], the remainder diluting the CTC specific signal.

Future analysis of the transcriptome of single or pools of pure CTCs might give a more accurate reflection of the *ESR1* expression in CTCs. In addition, single CTC analysis might give insight into the heterogeneity of *ESR1* expression within the pool of CTCs. An approach to increase the number of CTCs for analysis is the use of diagnostic leukapheresis (DLA)[29]. A higher number of CTCs will probably allow for a more robust gene expression analysis for a larger number of patients.

Ongoing analyses

The full quantitative analysis of the ^{18}F-FES PET/CT will enable correlation analyses between SUV_{max} and the other quantitative ER measures, namely the percentage of ER-positive tumor cells in the biopsy and CTC *ESR1* RNA expression levels. Moreover, integration of test characteristics of the ^{18}F-FES PET/CT and CTCs, PROs and analyses on cost-effectiveness could reveal which technique has potentially the highest clinical utility.

Conclusion

^{18}F-FES PET/CT, given its high PPV, is a reliable and a more patient-friendly alternative to assess the ER expression in patients with MBC compared to a tumor metastasis biopsy. In only 52% of patients, CTCs can also predict ER expression on metastases.

Acknowledgments

We would like to thank all the investigators and site staff, with special thanks to the patients and their families. We also like to acknowledge the financial support from the Dutch Cancer Society (Grant IMPACT, RUG 2012-5565).

Financial support

This study was financially supported by a grant from the Dutch Cancer Society (Grant IMPACT, RUG 2012-5565 and project EMCR 16-8196).

References

1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30 (2020).
2. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **365**, 1687-717 (2005).
3. Lindstrom, L.S. *et al.* Clinically used breast cancer markers such as estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 are unstable throughout tumor progression. *J Clin Oncol* **30**, 2601-8 (2012).
4. Schrijver, W. *et al.* Mutation profiling of key cancer genes in primary breast cancers and their distant metastases. *Cancer Res* **78**, 3112-21 (2018).
5. Cardoso, F. *et al.* 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4). *Ann Oncol* **29**, 1634-57 (2018).
6. Curigliano, G. *et al.* Should liver metastases of breast cancer be biopsied to improve treatment choice? *Ann Oncol* **22**, 2227-33 (2011).
7. Hoefnagel, L.D. *et al.* Discordance in ERα, PR and HER2 receptor status across different distant breast cancer metastases within the same patient. *Ann Oncol* **24**, 3017-23 (2013).
8. Brown, D. *et al.* Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun* **8**, 14944 (2017).
9. Kurland, B.F. *et al.* Between-patient and within-patient (site-to-site) variability in estrogen receptor binding, measured in vivo by ^{18}F-fluoroestradiol PET. *J Nucl Med* **52**, 1541-9 (2011).
10. Peterson, L.M. *et al.* Quantitative imaging of estrogen receptor expression in breast cancer with PET and ^{18}F-fluoroestradiol. *J Nucl Med* **49**, 367-74 (2008).
11. Beije, N. *et al.* Prognostic impact of HER2 and ER status of circulating tumor cells in metastatic breast cancer patients with a HER2-negative primary tumor. *Neoplasia* **18**, 647-53 (2016).
12. Beije, N. *et al.* Estrogen receptor mutations and splice variants determined in liquid biopsies from metastatic breast cancer patients. *Mol Oncol* **12**, 48-57 (2018).
13. Cristofanilli, M. *et al.* The clinical use of circulating tumor cells (CTCs) enumeration for staging of metastatic breast cancer (MBC): International expert consensus paper. *Crit Rev Oncol Hematol* **134**, 39-45 (2019).
14. Janni, W.J. *et al.* Clinical utility of repeated circulating tumor cell (CTC) enumeration as early treatment monitoring tool in metastatic breast cancer (MBC) - a global pooled analysis with individual patient data. in *SABCS 2020, GS4-08* (2020).
15. Yi, M. *et al.* Which threshold for ER positivity? a retrospective study based on 9639 patients. *Ann Oncol* **25**, 1004-11 (2014).
16. Wolff, A.C. *et al.* Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol* **36**, 2105-2122 (2018).
17. Romer, J. *et al.* Automated synthesis of 16alpha-[^{18}F]fluoroestradiol-3,17beta-disulphamate. *Appl Radiat Isot* **55**, 631-9 (2001).
18. van Kruchten, M. *et al.* PET imaging of oestrogen receptors in patients with breast cancer. *Lancet Oncol* **14**, e465-e475 (2013).
19. Venema, C.M. *et al.* Recommendations and technical aspects of 16α-[^{18}F]Fluoro-17β-estradiol PET to image the estrogen receptor in vivo: The Groningen experience. *Clin Nucl Med* **41**, 844-51 (2016).
20. Boellaard, R. *et al.* FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* **37**, 181-200 (2010).
21. Sieuwerts, A.M. *et al.* mRNA and microRNA expression profiles in circulating tumor cells and primary tumors of metastatic breast cancer patients. *Clin Cancer Res* **17**, 3600-18 (2011).
22. Onstenk, W. *et al.* Gene expression profiles of circulating tumor cells versus primary tumors in metastatic

breast cancer. *Cancer Lett* **362**, 36-44 (2015).
23. Reijm, E.A. *et al.* An 8-gene mRNA expression profile in circulating tumor cells predicts response to aromatase inhibitors in metastatic breast cancer patients. *BMC Cancer* **16**, 123 (2016).
24. Mostert, B. *et al.* Gene expression profiles in circulating tumor cells to predict prognosis in metastatic breast cancer patients. *Ann Oncol* **26**, 510-6 (2015).
25. Chae, S.Y. *et al.* Diagnostic accuracy and safety of 16α-[^{18}F]fluoro-17β-oestradiol PET-CT for the assessment of oestrogen receptor status in recurrent or metastatic lesions in patients with breast cancer: a prospective cohort study. *Lancet Oncol* **20**, 546-55 (2019).
26. Coleman, R.E. & Rubens, R.D. The clinical course of bone metastases from breast cancer. *Br J Cancer* **55**, 61-6 (1987).
27. Selli, C. *et al.* Accurate prediction of response to endocrine therapy in breast cancer patients: current and future biomarkers. *Breast Cancer Res* **18**, 118 (2016).
28. Sieuwerts, A.M. *et al.* Molecular characterization of circulating tumor cells in large quantities of contaminating leukocytes by a multiplex real-time PCR. *Breast Cancer Res Treat* **118**, 455-68 (2009).
29. Fischer, J.C. *et al.* Diagnostic leukapheresis enables reliable detection of circulating tumor cells of nonmetastatic cancer patients. *Proc Natl Acad Sci U S A* **110**, 16580-5 (2013).

Supplemental Table

Supplementary Table S1 - Discrepant results between ^{18}F-FES PET/CT and ER IHC and CTC-ER status and ER IHC

Patient number	ER IHC biopsy (%)	^{18}F-FES PET/CT	CTC first approach	CTC second approach	CTC ESR1 expression (ΔCq)	CTC count	ER IHC primary (%)	
1	Negative (0)	Positive	No	No	-	0	Negative (0)	
2	Negative (0)	Positive	No	No	-	2	Positive (100)	
3	Negative (0)	Positive	No	Yes	Positive (-6.98)	1	Positive (80)	
4	Negative (0)	Positive	No	No	-	12	Positive (60)	
5	Negative (0)	Positive	No	No	-	14	Positive (60)	
6	Negative (0)	Positive	No	No	-	0	Negative (0)	
7	Negative (0)	Positive	No	No	-	2	Negative (0)	
8	Negative (0)	Positive	No	No	-	0	Negative (0)	
9	Negative (0)	Positive	Yes	Yes	Positive (-6.18)	47	Negative (0)	
10	Negative (0)	Positive	Yes	Yes	Positive (-7.69)	135	Positive (100)	
11	Negative (0)	Positive	Yes	Yes	Negative (-8.26)	28	Positive (80)	
12	Positive (100)	Negative	No	Yes	Positive (-6.84)	2	Positive (90)	
13	Positive (20)	Negative	Yes	Yes	Negative (-10.45)	168	Positive (20)	
14	Positive (100)	Negative	No	No	-	0	Positive (100)	
15	Positive (10)	Negative	No	No	-	0	Positive (80)	
16	Positive (80)	Negative	No	No	-	5	Positive (100)	
17	Positive (100)	Negative	No	No	-	0	Positive (100)	
18	Positive (100)	Negative	No	No	-	0	Positive (100)	
19	Positive (60)	Negative	No	Yes	Positive (-3.064)	0	Positive (100)	
20	Positive (60)	Negative	No	No	-	10	Positive (20)	
21	Positive (10)	Negative	No	No	-	0	Positive (100)	
22	Negative (0)	Negative	Yes	Yes	Positive (-7.07)	48	Negative (0)	
23	Negative (0)	Negative	Yes	Yes	Positive (-6.55)	39	Positive (100)	
24	Negative (0)	Negative	Yes	Yes	Positive (-3.69)	6	Negative (0)	
9	Negative (0)	Positive	Yes	Yes	Positive (-6.18)	47	Negative (0)	
25	Negative (0)	Negative	Yes	Yes	Positive (-6.89)	8	Positive (80)	
10	Negative (0)	Positive	Yes	Yes	Positive (-7.69)	135	Positive (100)	
26	Positive (80)	Positive	Yes	Yes	Negative (-8.97)	5	Positive (100)	
13	Positive (20)	Negative	Yes	Yes	Negative (-10.45)	168	Positive (20)	
27	Positive (100)	Positive	Yes	Yes	Negative (-7.91)	5	Positive (100)	
28	Positive (100)	Positive	Yes	Yes	Negative (-7.87)	12	Positive (100)	
29	Negative (0)	Negative	No	Yes	Positive (-6.66)	3	Negative (0)	
3	Negative (1)	Positive	No	Yes	Positive (-6.98)	1	Positive (80)	
30	Negative (0)	Negative	No	Yes	Positive (-7.32)	0	Negative (0)	
31	Negative (0)	Negative	No	Yes	Positive (-6.38)	2	Negative (0)	
32	Negative (0)	Negative	No	Yes	Positive (-5.52)	4	Negative (0)	
33	Negative (0)	Negative	No	Yes	Positive (-5.52)	2	Unknown	
34	Negative (5)	Negative	No	Yes	Positive (-7.19)	2	Negative (0)	
35	Negative (0)	Negative	No	Yes	Positive (-7.74)	2	Negative (0)	
36	Positive (100)	Positive	No	Yes	Negative (-10.45)	3	Positive (100)	
37	Positive (100)	Positive	No	Yes	Negative (-8.97)	0	Positive (100)	
38	Positive (100)	Positive	No	Yes	Negative (-7.99)	4	Positive (100)	
39	Positive (100)	Positive	No	Yes	Negative (-10.18)	3	Positive (100)	

Discrepant ^{18}F-FES PET/CT /IHC results

Discrepant CTC-ER/ IHC results

CHAPTER 13

Summary & General Discussion

Introduction

Genomics-guided approaches swiftly become of greater importance in the treatment of cancer patients. The work described in this thesis focusses on the improvement of genomics-guided personalized treatment of patients with metastatic breast cancer, colorectal cancer, and melanoma. As the genomic features of cancer cells constantly change over time and under treatment pressure, there is consensus that molecular characteristics should be preferably determined on metastatic cells before start of a new line of treatment. Taking (repeated) tumor biopsies is however cumbersome and sometimes even not feasible due to the location of the metastatic site and tumor biopsies do not reflect the entire genomic landscape due to spatial heterogeneity. Hence, minimally invasive alternatives for genomic profiling have been extensively studied of which some already have entered the diagnostic field[1] over the past years. The studies described in thesis show how the application of three different modalities to interrogate the tumors' molecular make-up – tumor tissue biopsies, liquid biopsies and radiomics – to improve our insight into tumor biology and to help us to come to more genomics-guided treatment for cancer patients. These modalities all have their pros and cons, which are described in **Table 1**.

Table 1 – Pros and cons of the different modalities used for genomic profiling

Modality	Pros	Cons
Tumor tissue biopsies	Allows for assessment of tumor-related factors other than DNA alterations such as histology, proteins, and tumor micro-environment Compared to liquid biopsies, tumor tissue often results in higher DNA yield allowing for extensive genomic characterization by whole exome or whole genome sequencing Allows for wide screening of actionable mutations, mutational signatures, rearrangement signatures, copy number analysis, and tumor mutational burden	Cumbersome procedure hampering repetitive sampling Does not reflect intra- and inter-lesional heterogeneity
Liquid biopsies *General*	Minimally invasive procedure (no inherent risk of complications of tissue biopsies) and allows for repeated sampling over time Theoretically derived from all tumor sites thereby reflecting intra- and inter-lesional heterogeneity	Until now no reliable assessment is available to determine signatures such as microsatellite instability and tumor mutational burden on ctDNA and CTCs
ctDNA	Reflects the real time genomic status of the tumor due to its short half-life time (minutes – hours)	Cell-free DNA is highly fragmented (fragments of 140-175 base pairs) Amounts of cfDNA and ctDNA are relatively low; ctDNA fractions often <1% Extensive genomic characterization by whole exome or whole genome sequencing only possible in a subset of patients with sufficient ctDNA fractions Technical issues High error-rates of currently used sequencing methods Optimal unit of measure to monitor ctDNA unknown: variant allele frequency or mutant copies/mL plasma

Table 1 – Continued

Modality	Pros	Cons
CTCs	Allows for assessment of proteins, RNA expression and RNA splice-variants in addition to DNA analysis Allows for single tumor cell characterization and functional assays	CTCs are rare events requiring special enrichment techniques for detection and isolation No high-throughput procedures available No tumor-specific marker available, mostly enrichment of epithelial cells (using EpCAM) Single-cell analyses prone to artefacts
Radiomics	Non-invasive Enables analysis of the entire tumor volume overcoming sampling bias, allowing the quantitative analysis of intra- and inter-lesional heterogeneity during the course of disease and under treatment	Not expected that radiomics can replace or reflect all molecular details present in tumor tissue

Part I: Tumor tissue biopsies

During the past years implementation of sequencing techniques into clinical trials and routine practice has led to an increase in genomic-matched therapies for patients with cancer[2]. In the USA, it has been estimated that 15.4% percent of patients with advanced or metastatic cancer are currently eligible for genome-informed therapy and that of these patients only 6.6% will benefit[2]. Given the fact that the molecular characteristics of tumors constantly change over time, it is of utmost importance to characterize metastatic lesions and find new leads for targeted treatment and improve the numbers of patients that could benefit from genomics-guided therapies.

Getting insight into the genomic landscape of breast cancer by WGS of tumor biopsies

One of the strengths of examining metastatic tumor cells by taking tumor biopsies is that a substantial amount of tissue can be acquired resulting in high DNA and RNA yields, which allows in-depth investigation of the tumors' genome. An example of this is provided in **Chapter 2**, in which we analyzed the genomic profiles as assessed by whole genome sequencing (WGS) of metastatic tissue biopsies of 442 patients with metastatic breast cancer in order to get better insight into the DNA alterations occurring in metastatic disease.

We found that most driver genes that were detected in metastatic breast cancer had already been described in primary breast cancer[3,4] but there were some driver genes enriched in the metastatic biopsies. In estrogen receptor (ER)-positive/human epidermal growth factor receptor 2 (HER2)-negative breast cancer *TP53*, *ESR1*, *PTEN, NF1*, *KMT2C* and *AKT1* were more frequently mutated in metastatic than in primary tumors. Our results are in line with the findings of two other large cohorts of metastatic

breast tumors described by Bertucci[5] and Razavi *et al*.[6]. Next to mutations in single genes, we further assessed the contribution of mutational signatures and observed an increase in the relative contribution of COSMIC mutational signatures 2 and 13 which have been linked to APOBEC mutagenesis. Furthermore, an increase in COSMIC mutational signature 17 has been observed, which has currently an unknown etiology. Shifts in mutational signatures were also reported by the French group[5]. They reported next to an increase in signatures 3 (linked to homologous recombination deficiency) and signature 10 (POLE-associated) also increases of signatures 2, 13 and 17. Studies, such as the AURORA trial (NCT02102165) in which matched primary tumor and metastatic tissue are collected, could reveal whether primary tumors harboring these mutational signatures have higher metastatic potential or that these signatures are enriched over time and under (adjuvant) treatment pressure. In addition to the observation of changes in mutational signature contributions, we observed that some of the mutational signatures were associated with specific prior treatments. De novo signature I, which was characterized by CC>AA mutations, was more frequently observed in patients pre-treated with platinum-based chemotherapy. In addition, mutational signature 17 was more frequently observed in patients who were pretreated with fluorouracil, taxanes, platinum-containing chemotherapy and/or eribulin. The characteristic T>G ad T>C mutations in a CTT context implicated that 5-FU was the most likely drug contributing to this pattern. Recently, the association between 5-FU and signature 17 has been confirmed by organoid experiments and pre- and post-5-FU-treatment biopsies, which showed that the number of signature 17 mutations increased after exposure to 5-FU[7]. Finally, a large proportion of patients, 42%, had genomic alterations in their biopsy qualifying for genomics-guided therapy.

Although our findings provide a better understanding of the genomic landscape of metastatic breast cancer, an important weakness of our study is the large heterogeneity of treatments that have been administered resulting in small subgroups of patients treated with the same drug. Efforts to merge large datasets such as CPCT-02, MOSCATO-01[8], SHIVA[9], SAFIR-01[10], SAFIR-02 (NCT02299999) and AURORA (still recruiting, NCT02102165) will increase the number of patients treated with similar drugs and will have more power to detect genomic features associated with response. Importantly, future sequencing efforts that aim to reveal genomic predictors for treatment response should focus on the inclusion of patients in homogeneous cohorts starting with the same drug after taking the tumor biopsy. Furthermore, to get more insight into resistance mechanisms, one should try to take pre- and post-progression tumor biopsies and/or plasma circulating tumor DNA. Finally, more prospective studies – such as the drug rediscovery protocol (DRUP)[11] and MOSCATO-1 trial[8] – treating

patients based on the genomic profile of the tumor are eagerly awaited and will learn us whether genome-informed therapy will lead to improved survival and quality of life. Two examples of studies currently recruiting are The MATCH Screening Trial (NCT02465060) and DRUP (NCT02925234) trials, both matching patients based on genomic alterations of the tumor to FDA and/or EMA approved drugs.

Genomic analysis in relation to specific anti-cancer treatments

As personalized medicine focusses on giving the right drug, to the right patient, at the right moment, identification of genomic features associated with response to therapy is highly warranted. This requires broad-scale investigation of the genomic features of metastatic tumor cells. Also for this purpose, WGS analyses of biopsy material from metastatic lesions is currently likely the best method. **Chapter 3** highlights the potential of WGS analyses in breast cancer patients treated with capecitabine monotherapy. Here, we found clinical (ER status) and genomic predictors for response in which mutations (*TP53*, *PTPRS*, *HMCN1, CEP350,* and *ADGRG4*), copy number regions (amplification of 17q23.1 and loss of 4p16.3) and signatures (COSMIC mutational signature 16 and rearrangement signature 1) were associated with response. Our study is one of the first studies that has associated WGS data with response to therapy. However, our study is limited by the relatively small number of patients included and limited by the differences between patients regarding the number and type of prior treatments. Hence, we are planning to validate our results in an independent cohort of patients treated with capecitabine monotherapy of whom whole exome sequencing of a metastatic lesion and clinical data is available. Efforts associating genomic data with response to treatment will provide more insight into genomic alterations underlying sensitivity and intrinsic resistance to specific drugs.

Next to the discovery of intrinsic resistance mechanisms, studies investigating acquired resistance are important because understanding these mechanisms is the first step in overcoming these mechanisms. During the past years, activating mutations in the ligand binding domain of the estrogen receptor (*ESR1*) have been linked to acquired endocrine resistance[12-15]. **Chapter 4**, provides an overview of the pre-clinical and clinical studies on these activating *ESR1* mutations. Overall, *ESR1* mutations lead to constitutive ER activation and reduce the efficacy of aromatase inhibitors[16]. There are currently several studies ongoing investigating if this acquired resistance can be overcome by next-generation selective estrogen receptor modulators and selective estrogen receptor degraders that target both mutant and wild type ER[17]. Once these new endocrine treatments become available determination of the *ESR1* mutation status will probably become clinically relevant to select the most optimal endocrine treatment.

Integration of DNA and RNA sequencing

Several mechanisms of acquired endocrine resistance have been elucidated. Next to the abovementioned *ESR1* mutations, clinical associations with endocrine resistance have been shown for amplifications of *ERBB2*[18,19], *EGFR*[6], *FGFR1*[20,21], and *MYC*[6,22], and mutations in *ERBB2*[6,23,24], *NF1*[6,25-27], *PIK3CA*[28], *KRAS*[6,29], *BRAF*[6], *MAP2K1*[6], *FOXA1*[6], *CTCF*[6], *ARID1A*[6,30], *ARID2*[6].

Although whole genome and whole exome sequencing (WGS and WES, respectively) efforts have revealed enriched gene alterations in metastatic tumors compared to primary breast cancer, integration with gene expression is necessary to enable subsequent analysis of the downstream effects of these alterations. In **Chapter 5**, we showed a comprehensive integration of WGS and RNA sequencing data of biopsies of patients with ER-positive/HER2-negative breast cancer. Here, we observed a gene expression signature that was characterized by increased RNA expression of *ESR1* and its target genes. This "hyper-activated ER" signature was significantly more often present in biopsies with an *ESR1* mutation than in samples with wild type *ESR1*. Next to samples with *ESR1* mutations, there was a subgroup of 33 *ESR1* wild type samples expressing the ER-target genes at a similar levels as samples with an *ESR1* mutation amongst which were biopsies harboring *FGFR1* amplification. These results indicate that the ER-pathway remains an important driver for tumors with this "hyper-activated ER" signature and probably highlights the importance of blocking this activated pathway regardless of *ESR1* mutations.

Tissue analyses beyond sequencing

In addition to genomic and transcriptomic information that can be derived from tumor tissue biopsies, tissue provides an important basis to study other clinically relevant factors, such as the tumor micro-environment including tumor infiltrating lymphocytes and functional aspects such as the capacity of homologous DNA repair. For instance, breast cancer patients with germline *BRCA1* and *BRCA2* mutations show homologous recombination deficiency resulting in sensitivity to PARP inhibitors[31]. Functional assays such as the recombination REpair CAPacitiy (RECAP) test could identify tumors with homologous recombination deficiency beyond germline mutation carriers[32]. So depending on the clinical question one should select the most appropriate diagnostic method.

Part II: Liquid Biopsies

Although the genomic analysis of single tumor biopsies already have provided a better insight into the genomic alterations present in metastatic disease[5,6,33] there are still challenges with tissue biopsies. These challenges are mainly dominated by the invasiveness of taking biopsies and by intra- and inter lesional heterogeneity which is not entirely captured by a single tissue biopsy. Hence, markers that provide a real-time reflection of the genomic landscape and deal with plasticity and heterogeneity in a minimally invasive manner are needed.

Liquid biopsies can be used for several purposes and from the different analytes distinct information can be derived. Focusing on ctDNA and CTCs, **Table 2** provides per analyte a summary of the origin, the components that can be analyzed, tumor specific alterations that can be derived, and the downstream applications.

Table 2 – Information that can be derived from CTCs and ctDNA. Adapted from Heitzer *et al*.[34]

	Characteristics	CTCs	ctDNA
Origin	Viable cells	√	X
	Apoptotic cells	√	√
Components	DNA	√	√
	RNA	√	X
	Protein	√	X
Analyzable parameters	Mutations	√	√
	Epigenetic alterations	√	√
	Copy number alterations	√	√
	Fusion genes	√	√
	Splice variants	√	X
	Single cells	√	X
	Functional assays	√	X
Application	Prognostic information	√	√
	Predictive information	√	√
	Early detection of cancer	√	√
	Risk stratification	√	√
	Detection of minimal residual disease	√	√
	Disease monitoring	√	√
	Detection of resistance mechanisms	√	√
	Insights into the genomic landscape	√	√

In the following paragraphs, the various applications of liquid biopsies will be set out in relation to the work described in this thesis, currently available literature, and future perspectives.

Detection of minimal residual disease after initial curative treatment

Although this thesis focusses on genomics in metastatic cancer, curation of cancer

is mostly achieved in early stages of the disease by treatment of localized disease or minimal residual disease (MRD) in the (neo)adjuvant setting. Studies in early-stage colorectal cancer (CRC) and breast cancer have shown that patients who develop detectable ctDNA[35-40] – measured by mutations identified in the primary tumor and subsequently tracked in plasma – or CTCs[41,42] during follow-up have a higher risk of future relapse. It is likely that these patients might benefit mostly from adjuvant treatments or need intensified treatment regimens in order to prevent relapse. Intervention studies for colorectal cancer are now on the way including patients with high risk (stage II/III) CRC in which patients who have detectable ctDNA (ctDNA+) post-surgery will receive an intensified adjuvant regimen (PEGAGUS trial; NCT04259944). Similarly, a phase II study in patients with triple negative breast cancer is currently conducted. Those patients who develop detectable ctDNA during follow-up will be randomized to an immune checkpoint inhibitor (pembrolizumab) or observation (c-TRAK-TN trial; NCT03145961). One of the primary objectives is the proportion of patients without either detectable ctDNA or disease recurrence 6 months after commencing pembrolizumab.

Although ctDNA and CTCs are promising means to detect MRD, not all patients that suffer from recurrent disease are identified using liquid biopsies. In patients with non-metastatic cancer the detection rate of CTCs is low and, if detected, only at a median of 1 CTC per tube of blood[43] . So it is likely that one of the reasons for under detection of CTCs is the limited blood volume that is analyzed (7.5mL[41] - 30mL[42]). A technique to increase the analyzable blood volume is diagnostic leukapheresis (DLA)[44,45]. Using this technique, a large volume of blood is passed through a centrifuge by which peripheral blood mononuclear cells (PBMCs) as well as CTCs are isolated. The PBMC fraction can subsequently be examined for the presence of CTCs by the CellSearch technique. Using DLA in a pre-operative setting, the percentage of patients with detectable CTCs could be increased from 28% to 72% and the number of detected CTCs increased accordingly[44]. Given these numbers, it is likely that the use of DLA in the adjuvant setting will enhance the detection rate of patients with MRD. Currently, this approach is studied in a prospective study at our institute (MEC20-0384). Patients with ER-positive lymph node positive primary breast cancer who have received five years of adjuvant endocrine therapy and are still free of recurrence will undergo DLA. The goal is to detect CTCs and to demonstrate that the DLA-based method will increase the number of CTC-positive patients compared to screening of only 7.5 mL of blood. If the detection rate of patients with MRD will be increased, this justifies subsequent DLA-studies focusing on categorization of patients for risk of recurrence and escalate treatment in patients with MRD and de-escalate in patients without MRD in a randomized fashion.

In the adjuvant setting, "trackable" mutations in the primary tumor are identified for either 59-78%[35,40] of patients using pre-specified gene panels or 100% of patients using whole exome sequencing (WES)[38,39]. Mutations identified in the primary tumor are subsequently used to design custom MRD test to detect MRD in cfDNA during follow-up. Using custom MRD tests, the sensitivity for MRD detection is driven by the number of mutations per patient available to track[38]. To increase the number of trackable mutations per patient, WGS instead of WES could be considered. Although designing custom MRD tests is technically feasible, such an approach will be costly and time-consuming. Besides mutations, untargeted alternatives for MRD detection are techniques detecting cancer-associated alterations such as somatic copy number alterations[46], methylation patterns[47] or DNA fragmentation patterns[48,49].

Importantly, although the detection of CTCs and ctDNA seem promising means to detect MRD, it is key to study whether early identification of MRD and subsequent intensification or switching of adjuvant treatment regimens will indeed result in lower recurrence rates and not only result in a longer lead time. So currently, outside of well-designed clinical trials there is no role for ordering routine ctDNA or CTC analyses during follow-up for monitoring of recurrence in early-stage breast cancer and colorectal cancer patients. Furthermore, studies focusing on in-depth characterization of CTCs could elucidate affected pathways which might uncover the tumor cells' vulnerabilities and help to select for targeted therapies.

Earlier detection of cancer

In addition to detection of MRD in patients who have been treated with curative intent for their primary tumor, studies investigating cfDNA as screening tool for cancer have been performed as well[49-53]. Several studies focused on the detection of somatic mutations yielding increasing sensitivities by ascending stages of the disease[51,52]. Of note, there are two major disadvantages of using mutations in the screening setting: first, in the screening setting there is no prior knowledge of the presence of a specific mutation requiring large gene panels to yield an appropriate sensitivity. Second, recently was found that also non-malignant hematopoietic cells can acquire somatic mutations resulting from clonal hematopoiesis[54,55], which complicates the use of mutations even more in this setting. Therefore analyses of more general cancer-related alterations have been performed such as genome-wide fragmentation patterns of cfDNA[49], analyses of methylation profiles[53], or a combination of frequently occurring mutations and protein analyses[51]. These types of analyses have two advantages. First, no prior knowledge of the presence of a specific mutation is needed, and second, the tissue of origin could be traced which is difficult using mutations since the same mutations can drive multiple

tumor types. Identification of the affected organ system is needed to further determine what diagnostic test should be performed next.

Currently, the results from studies using multi-analyte screening test to detect cancer at an earlier stage are now awaited. For example, AI-Emerge (NCT03688906) and ECLIPSE (NCT04136002) are two studies focusing on earlier detection of CRC. AI-Emerge uses a machine learning platform to discover tumor- and immune- derived cfDNA signals in combination with epigenetic and protein biomarkers to develop a blood test for detection of early-stage CRC[50]. The ECLIPSE study combines genomic with epigenomic alterations for early detection of CRC. One should, however, be aware that there is a delicate balance between increasing the number of tested analytes which will lead to improved sensitivities and a potential decrease of specificity leading to unwanted false positive test results. Importantly, before implementation of liquid biopsy based screening methods into screening programs, we need to prove that implementation of these assays will improve disease-specific or overall survival compared to no screening or accepted screening methods such as colonoscopy or mammography.

Risk stratification

Within the field of oncology, risk stratification is routinely applied to select which patients should receive (neo)adjuvant systemic therapies. In the metastatic setting, for example in patients with metastatic breast cancer risk stratification is used to define whether a patient needs chemotherapy or can be treated with less toxic endocrine therapies. These strategies are usually driven by the extent of the disease – such as tumor grade, number of affected lymph nodes and in the metastatic setting involvement of visceral organs – and are less focused on the molecular characteristics of the disease.

The prognostic value of CTCs has been acknowledged for many years but treatment switches based on CTC counts in breast cancer patients who started with first line chemotherapy have not resulted in better outcomes in terms of progression free and overall survival[56]. This observation might be explained by the fact that CTC enumeration alone does not provide sufficient information about the molecular status of the disease and so cannot be used to direct treatment decisions. Characterization of CTCs and single CTCs might elucidate which pathways are affected in the pool of CTCs and could provide leads to further personalize cancer treatment and hopefully increase the survival of patients.

From gene expression analysis of primary tumors, we learned that specific profiles are predictive for the site of relapse – including brain metastases[57]. Once metastasized, CTCs present in the bloodstream might provide a better reflection of cells capable of

crossing the blood brain barrier than the use of the primary tumor. In **Chapter 10**, we aimed to predict the development of brain metastases based on gene expression profiles measured in the CTC-enriched fraction in patients with metastatic breast cancer starting with first line endocrine therapy or chemotherapy. Next to clinico-pathological features such as ER-negative status and a younger age, RNA expression of *MAGEA3*, *PLAU*, *TSPAN13*, and *CD44* in the CTC-enriched fraction were also associated with the development of brain metastases. Using the significant clinico-pathological variables (age and ER status) and gene expression of the four genes in a model to predict the development of brain metastases within two years after the initial blood draw, this resulted in a modest sensitivity of 68% and specificity of 68%. Importantly to note, the gene panel used for these analyses was not designed with the goal to detect brain metastases-initiating cells and future studies in this field should therefore include genes associated with the development of brain metastases, including *MYC*[58], *SEMA4D*[58], *STAT3*[59].

Single cell analyses

Enhancements of single-cell technologies allow for the in-depth analysis of CTCs at single-cell resolution providing insights into tumor heterogeneity[60]. In the coming years implementation of these single-cell techniques on multiple CTCs of a patient collected at multiple time points throughout different therapy lines, will gain information on heterogeneity at single time points but will also further our understanding of therapy resistance. Especially by enlarging the number of analyzable CTCs by using DLA techniques, more robust estimates of the proportion of cells harboring specific (targetable/resistance) alterations could be provided. Also, insights in the degree of CTC heterogeneity could yield prognostic information.

Detection of variants for targeted treatment

In lung oncology, detection of EGFR mutations in cfDNA to select for tyrosine kinase inhibitors has already entered the diagnostic arena. For the other tumor types, mutation testing in cfDNA is currently not routinely performed outside of clinical trials and still need clinical validation before its use will be implemented in the clinic.

To test the feasibility of using cfDNA to identify clinically actionable mutations for patients that are eligible for early phase clinical trials is the TARGET study[61] (Tumor chAracterization to Guide Experimental Targeted Therapy). In the first part of this study, one-hundred patients were included to test the real world feasibility for implementation of cfDNA profiling to increase the chance of matching patients with advanced cancers to a phase I trial. Using a variant allele frequency (VAF) threshold of

2.5%, actionable mutations were identified in 41 patients of whom 11 (27%) received a matched targeted treatment. Although this percentage seems low, depending on the number of drugs and trials available , the number of patients matched to a genome-guided therapy using tumor tissue lies between 25-50%[8,9,62]. In part B of the TARGET study, another 450 patients will be included which will provide a more robust overview of the number of patients that qualify for targeted treatment based on their cfDNA mutational profile. In addition, the plasmaMATCH study is one of the landmark studies that matches genomic alterations detected in cfDNA of ~1000 patients with advanced breast cancer with targeted treatments. Using a multi-cohort platform 142 entered one of four cohorts: A) *ESR1* mutation – extended dose fulvestrant 500mg every two weeks; B) *HER2* mutation – neratinib +/- fulvestrant (standard dosing); C) *AKT1* mutation in ER-positive breast cancer – capivasertib + fulvestrant (standard dosing) and D) *AKT1* mutation in ER-negative breast cancer or *PTEN* inactivating mutation – capivasertib. The preliminary results of this trial have been presented at the San Antonio Breast Cancer Symposium 2019[63]. Here, Turner *et al.* showed that the predefined efficacy criteria were met in cohorts B (neratinib for *HER2* mutations) and C (capivasertib for *AKT1* mutations), but that the efficacy criteria for cohort A (extended-dose fulvestrant for *ESR1* mutations) were not met. These large studies, is are good examples of studies we need to determine the efficacy of cfDNA informed treatments.

One of the strategies to improve patient inclusion for genomics-guided clinical trials, could include a two-step approach. First, perform panel-based sequencing or WES of cfDNA to evaluate the presence of alterations for which targeted therapies are available. If no targetable alteration has been identified – which can be due to a low tumor burden or low cfDNA yield – a tissue biopsy for WGS can be considered to screen for potential targets.

Detection of resistance mechanisms

In the search to improve upfront selection of patients receiving targeted treatments, not only the detection of actionable mutations is importance but equally important is the detection of variants associated with resistance. In **Chapter 6** we observed that a subset of patients with metastatic CRC who had tissue-tested *RAS* wild-type disease, did have *RAS* mutations in cfDNA. Patients that harbored a *RAS* or *BRAF* mutation in either tumor tissue and/or cfDNA had a worse outcome on cetuximab monotherapy. At disease progression, the majority of patients acquired mutations in *RAS, BRAF* and/or *EGFR.* The presence of *RAS* mutations – not detected in tumor tissue – and the emergence of *RAS* mutations at progression on anti-EGFR therapies have been shown by others[64-66]. The decision, however, to refrain from giving targeted therapy based on liquid biopsies or to discontinue treatment based on a rising number of mutant molecules associated

with resistance should be preceded by answering some questions:

1. If mutations occur during treatment, what change in mutant molecules or VAF is due to assay variability and from what change does is reflect real tumor biology (i.e. response or progression)?
2. How many mutant molecular per mL plasma or what threshold of VAF leads to non-response?
3. When an increase in the number of mutant molecules is observed, does early treatment switching, lead to better outcomes?
4. Does a decline in mutant molecules after withdrawal of targeted therapies provide ground for a re-challenge with targeted therapies?

First, in order to learn what change in mutant molecules or VAF is due to assay variability, standardization of pre-analytical conditions is needed[67,68] and subsequently studies investigating assay variability are of utmost importance to provide reference ranges for the number of mutant molecules/VAF[69].

Second, to learn which number of mutant molecules/mL plasma or what height of VAF leads to non-response, patients should ideally be treated with targeted therapies and the outcome should be reported stratified by the number of mutant molecules or height of the VAF. One could imagine that the presence of a sub-clonal *ESR1* mutation, only present in a small subset of cells, could lead to response in the non-mutated cancer cells.

Third, tracking resistance during treatment using liquid biopsies becomes clinically relevant when switching of treatments based on a rising number of mutant molecules, leads to better outcomes for patients. This question will be addressed in the PADA-1 trial (NCT03079011). In this study, patients with ER-positive/HER2-negative breast cancer are treated with the combination of an aromatase inhibitor and palbociclib as first line treatment for metastatic disease[70]. After the onset of rising *ESR1* mutation(s) in blood, patients without RECIST tumor progression will be randomized between continuation of the same treatment regimen or switch to fulvestrant/palbociclib. The primary objectives of this trial are treatment safety and progression free survival in the two treatment arms.

Fourth, it has been shown that *RAS* and *EGFR* mutant clones, which have emerged during anti-EGFR monoclonal antibody therapy, decline exponentially – with a half-life of 4.3 months – upon withdrawal of this treatment[66,71]. Hence, the question whether re-

challenge with anti-EGFR monoclonal antibodies after a drug holiday or after receiving other therapies, could lead to response has been raised. This question has been studied in the prospective phase II CRICKET study[72], in which patients that had experienced initial response and then progression on first-line combination therapy with irinotecan and cetuximab, were exposed again to this combination in the third line, after receiving second-line chemotherapy plus bevacizumab. In twelve patients harboring *RAS* mutations in their liquid biopsy before re-challenge, no objective responses were observed. Of thirteen patients who were *RAS* wild type in their liquid biopsy, six experienced a partial response. Also patients with *RAS* wild type liquid biopsies experienced a longer progression free survival than those with a *RAS* mutations (4.0 versus 1.9 months). Although the CRICKET trial provides preliminary proof-of-concept evidence about the potential role of liquid biopsy for selecting patients for re-challenge therapy, robust confirmation in larger clinical trials is needed. Currently, several studies are being performed to define whether there is a *RAS* dynamic threshold that predicts clinical failure/success[73]. The CHRONOS trial (NCT03227926) is currently running in which patients are only eligible for a re-challenge with panitumumab if patients show a >50% drop in *RAS* mutational load at the time of re-challenge compared to the time of progression on the first-line anti-EGFR containing therapy.

In conclusion, the fact that the abundance of mutant clones is a dynamic process, marks the need for trials investigating the efficacy re-challenges. Interventional trials that randomize between arms that repeatedly perform cfDNA testing during the course of the disease to tailor treatments real-time versus versus treating on the discretion of the treating oncologist will learn whether liquid-biopsy informed treatment strategies will lead to better outcomes.

Providing insight into the genomic landscape

Besides the significant progress that has been made to track mutations using targeted gene panels and single gene assays, WES of cfDNA enables a more comprehensive analysis providing a larger picture of the landscape of somatic alterations. In **Chapter 7**, we described to what extent WES of cfDNA is technically feasible. Using individual patient data, we showed that the pooled sensitivity of WES-detected single nucleotide variants in cfDNA, using tumor tissue as reference, was 50% (95% CI:29%-72%). In addition, the number of detected SNVs was positively correlated with the tumor fraction in cfDNA. Hence the sub-analysis of samples with a tumor fraction ≥25% improved the sensitivity to 69% (95% CI: 46%-89%). Regarding technical feasibility, the pre- and post-analytical procedures were highly variable, rendering comparison between studies problematic.

Although efforts have to be made before these techniques will enter the diagnostic arena, we already can learn a lot from studies that have been applying WES on paired samples to reveal resistance mechanisms. For example, in a small subset of patients from the PALOMA-3 study[74], in which the efficacy of fulvestrant plus palbociclib was compared with fulvestrant plus placebo, WES was performed on cfDNA samples with ≥10% tumor fraction collected at baseline and at the end of treatment. Subclonal changes in driver genes at end of treatment were observed. *RB1* and *FGFR2* mutations were observed in end of treatment samples which were not detected at baseline. Interestingly, the subclones harboring these specific mutations at the end of treatment were predominantly characterized by APOBEC mutations. Based on the observations in the subset of patients for which WES was possible, targeted sequencing panels – including *RB1* and *FGFR2* amongst other genes – were developed to interrogate baseline and end of treatment samples for a larger set of patients.

The discovery capacity of WES to identify resistance mechanisms have also been shown by Goodall *et al.*[75]. WES on cfDNA samples collected at disease progression, following initial response on the PARP inhibitor olaparib, revealed subclonal frameshifts in genes carrying germline or somatic mutations in *BRCA2/PALB2*, reverting these genes back in frame as mechanism of resistance. Altogether, the applicability of WES currently mainly resides in a subset of patients with high tumor fractions in cfDNA. In this population it is an attractive tool for identification of genomic signatures and discovery of resistance mechanisms.

Going beyond blood-based analyses: cerebrospinal fluid

Challenges in obtaining tumor tissue from primary brain tumors, brain metastases and leptomeningeal metastases stipulates the need for alternative diagnostic tools. DNA analyses of cerebrospinal fluid (CSF) has been considered as an attractive means to diagnose and characterize these tumors at the genomic level[76-78]. In **Chapter 8** we reviewed alternative diagnostic tools to detect leptomeningeal metastases in patients with breast cancer, where we made a case for the use of cfDNA and CTC analyses to detect LM. Subsequently, in **Chapter 9**, we demonstrated that the majority of cytology positive CSF samples harbored aneuploidy. In addition, we detected aneuploidy in CSF samples prior to the clinical diagnosis of CNS metastases. In these patients, the aneuploidy could have detected CNS metastases at an earlier stage. Furthermore, we demonstrated that a high aneuploidy score was associated with worse overall survival and development of LM. The results of this retrospective analysis are encouraging and need prospective validation.

Recently, van Bussel *et al*. compared the epithelial cell adhesion molecule (EpCAM) immunoflow cytometry technique to detect CTCs in CSF with CSF cytology and showed a high sensitivity and specificity of 94% and 100%, respectively[79]. Currently, a prospective study (MEC15-419) is running at our institute in which patients with breast cancer suspected of LM in whom the MRI scan is inconclusive and who will undergo a LP for regular cytological analysis are included. Additional CSF will be collected to enumerate and characterize CTCs in CSF by using the FDA-approved CellSearch technique. The primary aim of this study is to compare the detection rate of LM by identifying CTCs in CSF using CellSearch with standard care using CSF cytology. Secondary objectives are to determine the molecular profile of these CTCs in CSF and compare the molecular profiles with matched CTCs in peripheral blood. Furthermore, CSF supernatant is being collected for copy number and mutational profiling. Prospective studies like this, will answer the question whether CTC and ctDNA analysis of CSF will improve the detection rate of LM. Furthermore, molecular analyses will enhance our understanding of the pathophysiologic mechanism of cells metastasizing to the leptomeninges, and could ultimately guide future therapies.

Part III: Radiomics

Next to taking tumor biopsies or using liquid biopsies, radiomics is another method to characterize tumor lesions in cancer patients. This field of research is swiftly gaining ground as a patient-friendly alternative for tissue biopsies[80]. Until now, radiomics models have been built for several cancer types including breast[81], lung[82], and melanoma[83]. In **Chapter 11**, we investigated the discriminative power of quantitative imaging features to distinguish patients with *BRAF* mutated and wild type melanomas. This study showed that neither radiomics features nor scoring by a thorax radiologist according to the Lung Image Database Consortium (LIDC) criteria[84] could discriminate between *BRAF* wild type and *BRAF* mutant lesions. These results show that although there is significant promise in radiomics models, it is not expected that radiomics can replace or reflect all molecular details which can be derived from tumor tissue analyses. Like all new diagnostic fields, standardization of pre-analytical parameters such as data collection (including image acquisition protocols and segmentation procedures) and analysis pipelines are required[85]. Furthermore, integration of radiomics into prospective clinical trials will be necessary to link imaging biomarkers with biological and clinical parameters.

Part IV: Integration of modalities

As already summarized in **Table 1**, each method to characterize tumor lesions has its own pros and cons. Though it is likely that none of these methods is able to completely replace any of the others, comparison, and/or integration of these methods to get the most complete overview is important. In **Chapter 12**, we studied the ER status using multimodal ER data and compared the gold standard tissue biopsy with CTC *ESR1* RNA expression and ER expression measured by the FES-PET. Here, we showed that CTC *ESR1* RNA expression and FES-PET analyses are promising alternatives for ER immunohistochemistry on tumor tissue. We do recognize, that tumor tissue analyses yield additional information to ER status such as HER2 status. In routine work-up, tissue biopsies shall therefore – especially in the setting of newly diagnosed breast cancer – remain the gold standard until accurate HER2 amplification liquid biopsy techniques or imaging techniques could provide the complete overview of clinically used biomarkers. One should recognize, that the FES-PET does provide a more complete overview of all metastatic sites expressing ER and could be of use when some lesions are responding but others are not. Comprehensive studies such as IMPACT Breast (NCT01832051), CPCT-02 (NCT01855477) and SONIA (NCT03425838), in which information from tumor tissue, liquid biopsies and imaging biomarkers is collected, show the potential of Dutch collaboration. These studies will contribute to a better understanding of metastatic disease and will provide answers about which technique(s) will yield clinically relevant information.

Conclusion

This thesis describes several studies in the field of genomics-guided personalized cancer treatment using data originating from three modalities: tissue biopsies, liquid biopsies and radiomics. Based on the studies described, we can conclude that every technique has its own pros and cons and that one should match the research and/or clinical question with the appropriate technique(s) taking these pros and cons into account. Large scale clinical implementation of WGS on metastatic biopsies, liquid biopsy analyses and radiomics still depends on studies that show clinical utility in terms of survival benefit or improved quality of life.

To ultimately come to clinical utility, we need studies that not only show test results in relation to outcome but intervention studies that starts with genomic testing that subsequently guide downstream clinical management. Studies such as the PADA-1 and

The MATCH Screening Trial are nice examples of studies using upfront genomic testing, using either liquid biopsies or tissue testing, followed by a clinical consequence and a primary endpoint focused on efficacy of the intervention based on the genomic test used.

Moreover, our WGS data of metastatic biopsies of patients with metastatic breast cancer exemplifies the need of interventional studies treating patients based on potentially clinically relevant molecular features such as TMB, HRD, MSI and specific targetable mutations. Determination of these features will only yield clinical utility when studies performing WGS will subsequently guide treatment according to the tumors' molecular profile and show clinical benefit.

In the coming years, results of such interventional studies investigating the clinical utility of genomics-guided therapy using (a combination of) liquid biopsies, tumor tissue or radiomics to select patients, are eagerly awaited.

References

1. cobas EGFR Mutation Test v2. 2016.
2. Marquart, J. *et al.* Estimation of the Percentage of US Patients With Cancer Who Benefit From Genome-Driven Oncology. *JAMA Oncol* **4**, 1093-1098 (2018).
3. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54 (2016).
4. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
5. Bertucci, F. *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564 (2019).
6. Razavi, P. *et al.* The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* **34**, 427-438 e6 (2018).
7. Christensen, S. *et al.* 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat Commun* **10**, 4571 (2019).
8. Massard, C. *et al.* High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial. *Cancer Discov* **7**, 586-595 (2017).
9. Le Tourneau, C. *et al.* Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet Oncol* **16**, 1324-34 (2015).
10. Andre, F. *et al.* Comparative genomic hybridisation array and DNA sequencing to direct treatment of metastatic breast cancer: a multicentre, prospective trial (SAFIR01/UNICANCER). *Lancet Oncol* **15**, 267-74 (2014).
11. van der Velden, D.L. *et al.* The Drug Rediscovery protocol facilitates the expanded use of existing anticancer drugs. *Nature* **574**, 127-131 (2019).
12. Jeselsohn, R. *et al.* Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res* **20**, 1757-67 (2014).
13. Merenbakh-Lamin, K. *et al.* D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res* **73**, 6856-64 (2013).
14. Robinson, D.R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet* **45**, 1446-51 (2013).
15. Toy, W. *et al.* ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet* **45**, 1439-45 (2013).
16. Hanker, A.B. *et al.* Overcoming Endocrine Resistance in Breast Cancer. *Cancer Cell* **37**, 496-513 (2020).
17. Fanning, S.W. & Greene, G.L. Next-Generation ERalpha Inhibitors for Endocrine-Resistant ER+ Breast Cancer. *Endocrinology* **160**, 759-769 (2019).
18. Johnston, S. *et al.* Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor-positive metastatic breast cancer. *J Clin Oncol* **27**, 5538-46 (2009).
19. Kaufman, B. *et al.* Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized phase III TAnDEM study. *J Clin Oncol* **27**, 5529-37 (2009).
20. Giltnane, J.M. *et al.* Genomic profiling of ER(+) breast cancers after short-term estrogen suppression reveals alterations associated with endocrine resistance. *Sci Transl Med* **9**(2017).
21. Turner, N. *et al.* FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* **70**, 2085-94 (2010).

22. Miller, T.W. *et al.* A gene expression signature from human breast cancer cells with acquired hormone independence identifies MYC as a mediator of antiestrogen resistance. *Clin Cancer Res* **17**, 2024-34 (2011).
23. Nayar, U. *et al.* Acquired HER2 mutations in ER(+) metastatic breast cancer confer resistance to estrogen receptor-directed therapies. *Nat Genet* **51**, 207-216 (2019).
24. Smyth, L.M. *et al.* Efficacy and Determinants of Response to HER Kinase Inhibition in HER2-Mutant Metastatic Breast Cancer. *Cancer Discov* **10**, 198-213 (2020).
25. Pearson, A. *et al.* Inactivating NF1 Mutations Are Enriched in Advanced Breast Cancer and Contribute to Endocrine Therapy Resistance. *Clin Cancer Res* **26**, 608-622 (2020).
26. Sokol, E.S. *et al.* Loss of function of NF1 is a mechanism of acquired resistance to endocrine therapy in lobular breast cancer. *Ann Oncol* **30**, 115-123 (2019).
27. Griffith, O.L. *et al.* The prognostic effects of somatic mutations in ER-positive breast cancer. *Nat Commun* **9**, 3476 (2018).
28. Andre, F. *et al.* Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer. *N Engl J Med* **380**, 1929-1940 (2019).
29. Fribbens, C. *et al.* Tracking evolution of aromatase inhibitor resistance with circulating tumour DNA analysis in metastatic breast cancer. *Ann Oncol* **29**, 145-153 (2018).
30. Xu, G. *et al.* ARID1A determines luminal identity and therapeutic response in estrogen-receptor-positive breast cancer. *Nat Genet* **52**, 198-207 (2020).
31. Lord, C.J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152-1158 (2017).
32. Meijer, T.G. *et al.* Functional Ex Vivo Assay Reveals Homologous Recombination Deficiency in Breast Cancer Beyond BRCA Gene Defects. *Clin Cancer Res* **24**, 6277-6287 (2018).
33. Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* (2019).
34. Heitzer, E. Circulating Tumor DNA for Modern Cancer Management. *Clin Chem* (2019).
35. Garcia-Murillas, I. *et al.* Assessment of Molecular Relapse Detection in Early-Stage Breast Cancer. *JAMA Oncol* (2019).
36. Tie, J. *et al.* Circulating Tumor DNA Analyses as Markers of Recurrence Risk and Benefit of Adjuvant Therapy for Stage III Colon Cancer. *JAMA Oncol* (2019).
37. Reinert, T. *et al.* Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA Oncol* (2019).
38. Parsons, H.A. *et al.* Sensitive detection of minimal residual disease in patients treated for early-stage breast cancer. *Clin Cancer Res* (2020).
39. Coombes, R.C. *et al.* Personalized Detection of Circulating Tumor DNA Antedates Breast Cancer Metastatic Recurrence. *Clin Cancer Res* **25**, 4255-4263 (2019).
40. Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133 (2015).
41. Sparano, J. *et al.* Association of Circulating Tumor Cells With Late Recurrence of Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncol* **4**, 1700-1706 (2018).
42. Trapp, E. *et al.* Presence of Circulating Tumor Cells in High-Risk Early Breast Cancer During Follow-Up and Prognosis. *J Natl Cancer Inst* **111**, 380-387 (2019).
43. Lucci, A. *et al.* Circulating tumour cells in non-metastatic breast cancer: a prospective study. *Lancet Oncol* **13**, 688-95 (2012).
44. Fischer, J.C. *et al.* Diagnostic leukapheresis enables reliable detection of circulating tumor cells of nonmetastatic cancer patients. *Proc Natl Acad Sci U S A* **110**, 16580-5 (2013).

45. Bidard, F.C. *et al.* Circulating Tumor Cells in Breast Cancer Patients Treated by Neoadjuvant Chemotherapy: A Meta-analysis. *J Natl Cancer Inst* **110**, 560-567 (2018).
46. Belic, J. *et al.* Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clin Chem* **61**, 838-49 (2015).
47. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* **9**, 5068 (2018).
48. Snyder, M.W. *et al.* Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68 (2016).
49. Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389 (2019).
50. Putcha, G. *et al.* Blood-based detection of early-stage colorectal cancer using multiomics and machine learning. *J Clin Oncol* **38**, 66-66 (2020).
51. Cohen, J.D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926-930 (2018).
52. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra24 (2014).
53. Liu, M.C. *et al.* Genome-wide cell-free DNA (cfDNA) methylation signatures and effect on tissue of origin (TOO) performance. *J Clin Oncol* **37**, 3049-3049 (2019).
54. Razavi, P. *et al.* High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med* (2019).
55. Hu, Y. *et al.* False-Positive Plasma Genotyping Due to Clonal Hematopoiesis. *Clin Cancer Res* **24**, 4437-4443 (2018).
56. Smerage, J.B. *et al.* Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J Clin Oncol* **32**, 3483-9 (2014).
57. Smid, M. *et al.* Subtypes of breast cancer show preferential site of relapse. *Cancer Res* **68**, 3108-14 (2008).
58. Klotz, R. *et al.* Circulating Tumor Cells Exhibit Metastatic Tropism and Reveal Brain Metastasis Drivers. *Cancer Discov* **10**, 86-103 (2020).
59. Chiu, W.T. *et al.* Caveolin-1 upregulation mediates suppression of primary breast tumor growth and brain metastases by stat3 inhibition. *Cancer Res* **71**, 4932-43 (2011).
60. Keller, L. & Pantel, K. Unravelling tumour heterogeneity by single-cell profiling of circulating tumour cells. *Nat Rev Cancer* **19**, 553-567 (2019).
61. Rothwell, D.G. *et al.* Utility of ctDNA to support patient selection for early phase clinical trials: the TARGET study. *Nat Med* **25**, 738-743 (2019).
62. Stockley, T.L. *et al.* Molecular profiling of advanced solid tumors and patient outcomes with genotype-matched clinical trials: the Princess Margaret IMPACT/COMPACT trial. *Genome Med* **8**, 109 (2016).
63. Turner, N. *et al.* Abstract GS3-06: Results from the plasmaMATCH trial: A multiple parallel cohort, multi-centre clinical trial of circulating tumour DNA testing to direct targeted therapies in patients with advanced breast cancer (CRUK/15/010). *Cancer Research* **80**, GS3-06-GS3-06 (2020).
64. Vidal, J. *et al.* Plasma ctDNA RAS mutation analysis for the diagnosis and treatment monitoring of metastatic colorectal cancer patients. *Ann Oncol* **28**, 1325-1332 (2017).
65. Khan, K.H. *et al.* Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov* **8**, 1270-1285 (2018).
66. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med* **21**, 795-801 (2015).
67. van Dessel, L.F. *et al.* Application of circulating tumor DNA in prospective clinical oncology trials -

standardization of preanalytical conditions. *Mol Oncol* **11**, 295-304 (2017).
68. van Dessel, L.F. *et al.* High-throughput isolation of circulating tumor DNA: a comparison of automated platforms. *Mol Oncol* **13**, 392-402 (2019).
69. Bos, M.K. *et al.* Variant allele frequency versus number of mutant molecules: choosing the optimal unit of measurement for circulating tumor DNA. *Submitted* (2020).
70. Bidard, F.C. *et al.* Abstract PD2-06: Circulating *ESR1* mutation detection rate and early decrease under first line aromatase inhibitor and palbociclib in the PADA-1 trial (UCBG-GINECO). *Cancer Research* **79**, PD2-06-PD2-06 (2019).
71. Parseghian, C.M. *et al.* Anti-EGFR-resistant clones decay exponentially after progression: implications for anti-EGFR re-challenge. *Ann Oncol* **30**, 243-249 (2019).
72. Cremolini, C. *et al.* Rechallenge for Patients With RAS and BRAF Wild-Type Metastatic Colorectal Cancer With Acquired Resistance to First-line Cetuximab and Irinotecan: A Phase 2 Single-Arm Clinical Trial. *JAMA Oncol* **5**, 343-350 (2019).
73. Mauri, G. *et al.* Retreatment with anti-EGFR monoclonal antibodies in metastatic colorectal cancer: Systematic review of different strategies. *Cancer Treat Rev* **73**, 41-53 (2019).
74. O'Leary, B. *et al.* The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer Discov* **8**, 1390-1403 (2018).
75. Goodall, J. *et al.* Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition. *Cancer Discov* **7**, 1006-1017 (2017).
76. Seoane, J. *et al.* Cerebrospinal fluid cell-free tumour DNA as a liquid biopsy for primary brain tumours and central nervous system metastases. *Ann Oncol* **30**, 211-218 (2019).
77. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun* **6**, 8839 (2015).
78. Bougel, S. *et al.* Methylation of the hTERT promoter: a novel cancer biomarker for leptomeningeal metastasis detection in cerebrospinal fluids. *Clin Cancer Res* **19**, 2216-23 (2013).
79. van Bussel, M.T.J. *et al.* Circulating epithelial tumor cell analysis in CSF in patients with leptomeningeal metastases. *Neurology* **94**, e521-e528 (2020).
80. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* **48**, 441-6 (2012).
81. Zhu, Y. *et al.* Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. *Sci Rep* **5**, 17787 (2015).
82. Rios Velazquez, E. *et al.* Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* **77**, 3922-3930 (2017).
83. Trebeschi, S. *et al.* Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol* **30**, 998-1004 (2019).
84. Opulencia, P., Channin, D.S., Raicu, D.S. & Furst, J.D. Mapping LIDC, RadLex, and lung nodule image features. *J Digit Imaging* **24**, 256-70 (2011).
85. Miles, K. Radiomics for personalised medicine: the long road ahead. *Br J Cancer* **122**, 929-930 (2020).

APPENDICES

Nederlandse samenvatting
Curriculum Vitae
PhD Portfolio
List of Publications
Author Affiliations
Dankwoord

NEDERLANDSE SAMENVATTING

Kanker is een ziekte van het DNA. Door een reeks aan foutieve veranderingen in het DNA kan een gezonde cel veranderen in een kankercel. Een kankercel bezit eigenschappen die leiden tot ongebreidelde celdeling en het vermogen te verplaatsen door het lichaam. Doordat de kankercellen zich kunnen verplaatsen door het lichaam, kunnen ze op een andere plek in het lichaam opnieuw een tumor vormen.

Aangezien elk individu wat betreft genetische opmaak uniek is (het DNA verkregen van vader en moeder), eeneiige tweelingen uitgezonderd, kunt u zich voorstellen dat tumoren op zichzelf ook een unieke genetische opmaak hebben. Deze verschillen - ook wel tumor specifieke eigenschappen genoemd - maken dat de ene tumor agressiever is dan de andere en dat de ene tumor goed en de andere tumor slecht reageert op een specifieke anti-kanker behandeling.

Voor de behandeling van kanker is het belangrijk om te realiseren dat patiënten over het algemeen niet overlijden aan de **primaire tumor**, maar aan **metastasen** die zich op verschillende plekken in het lichaam gevormd hebben. Voor de behandeling van patiënten met uitgezaaide kanker is het dus van belang de behandeling te richten op de metastasen. Tot voor kort werden patiënten met gemetastaseerde ziekte behandeld op basis van het orgaan waarin de tumor zich ontwikkeld had. Ondanks dat deze benadering nog steeds een grote rol speelt in de hedendaagse oncologische zorg, wordt tegenwoordig steeds vaker onderzocht of behandeling op basis van de tumor specifieke eigenschappen beter werkt. Om te onderzoeken welke genetische afwijkingen er in de tumor aanwezig zijn, wordt vaak gebruikt gemaakt van een weefselbiopt van de primaire tumor of van een metastase. Omdat tumorcellen gedurende de tijd en onder druk van behandeling kunnen veranderen is men het er over eens dat de moleculaire eigenschappen van de tumor bij voorkeur bepaald zouden moeten worden op een metastase voor start van een nieuwe behandeling. Om de eigenschappen van een metastase te onderzoeken wordt vaak een biopt afgenomen. Echter, het nemen van (herhaaldelijke) tumor biopten is onprettig voor patiënten en niet altijd mogelijk door de locatie van de metastase. Daarnaast weerspiegelt één enkel tumor biopt slechts de genetische opmaak van een deel van één metastase en niet van de verschillende metastasen. Derhalve zijn minimaal invasieve manieren om de genetische opmaak van tumoren in kaart te brengen zeer gewenst.

Een **primaire tumor** ligt in het orgaan waar de kanker begonnen is met groeien. Voor borstkanker geldt dat de primaire tumor gelegen en ontstaan is in de borst.
Een **metastase** is een uitzaaiing die zich op elke plek van het lichaam kan vormen, doordat cellen van de primaire tumor zijn losgekomen en zich door de bloed- of lymfebanen hebben verspreid en op een andere plek in het lichaam zijn uitgegroeid tot een nieuwe tumor. Afhankelijk van waar deze uitzaaiing zich bevindt kan een uitzaaiing leiden tot specifieke klachten.

Voorbeelden van minimaal invasieve manieren om dit te bereiken zijn: vloeibare biopten en "radiomics". Belangrijk om te vermelden is, dat zowel liquid biopsies als radiomics technieken zijn die momenteel met name gebruikt worden in onderzoeksverband en nog niet of nauwelijks in de dagelijkse praktijk.

Vloeibare biopten, ook wel liquid biopsies genoemd, zijn biopten van lichaamsvocht, bijvoorbeeld bloed of hersenvocht, waarin DNA afkomstig van tumorcellen of hele tumorcellen aanwezig zijn die bestudeerd kunnen worden.
Radiomics. Deze techniek maakt gebruik van radiologiebeelden. In deze beelden zit namelijk veel informatie opgeslagen zoals de grootte, de vorm en heterogeniteit van de tumor. Door gebruik te maken van algoritmes kunnen combinaties van eigenschappen, die wij met het blote oog niet kunnen zien, gebruikt worden om onderscheid te maken tussen wel of geen kanker, maar ook om moleculaire eigenschappen van de tumor te voorspellen zonder dat daar een weefselbiopt voor nodig is.

De studies die in dit proefschrift beschreven worden, hebben als doel de moleculaire eigenschappen van de tumor beter in kaart te brengen. Hierdoor wordt ons inzicht in de tumorbiologie verbeterd. Het ultieme doel is om de meest effectieve behandeling voor het individu te kiezen waardoor enerzijds de duur van het leven verlengd wordt, maar belangrijker nog om de kwaliteit van het leven te verbeteren dan wel te behouden. In dit proefschrift worden verschillende technieken onderzocht om de moleculaire eigenschappen van de tumor te onderzoeken: analyses van weefselbiopten, vloeibare biopten ("liquid biopies") en radiomics data.

Deel I: Weefselbiopten

Eén van de voordelen van weefselbiopten is dat er relatief veel tumor materiaal verkregen kan worden, wat leidt tot een goede opbrengst van DNA en RNA. Dit maakt een nauwkeurige analyse van het genoom van de tumor mogelijk. In **hoofdstuk 2**, hebben we het DNA-profiel van 442 metastasen van patiënten met borstkanker in kaart gebracht met behulp van "whole genome sequencing (WGS)".

Wat is whole genome sequencing (WGS)?
Dit is een techniek waarmee de complete DNA-sequentie in kaart gebracht kan worden. Voor een tumor geldt dat door middel van WGS alle belangrijke fouten in het DNA van de tumor (het DNA-profiel) in kaart gebracht kunnen worden. Deze nieuwe techniek wordt gebruikt om de eigenschappen van de tumor beter in kaart te brengen, maar ook om nieuwe/betere behandelingen voor patiënten met kanker te ontwikkelen.

In deze studie werden twee belangrijke vragen beantwoord:

1. In hoeverre verschilt het DNA-profiel van metastasen van dat van primaire tumoren?
2. Zijn er groepen van patiënten te identificeren die op basis van hun DNA-profiel baat zouden kunnen hebben van specifieke therapieën?

Vergelijking primaire tumoren versus metastasen

Bij de vergelijking tussen de DNA afwijkingen van metastaten en primaire tumoren hebben we gebruik gemaakt van een cohort primaire tumoren dat ook geanalyseerd is middels WGS, het BASIS cohort. Dit is een cohort van primaire tumoren van andere vrouwen dan de vrouwen die deelnamen aan onze studie. In vergelijking met de primaire tumoren vertoonden gemetastaseerde tumoren een hoger percentage van mutaties in zogeheten drivergenen *TP53*, *ESR1*, *PTEN, NF1, KMT2C* en *AKT1.* De vraag blijft echter of de verrijking van mutaties in deze genen in de gemetastaseerde setting optreedt door eerdere behandeling of doordat tumoren met deze mutaties gemakkelijker metastaseren. Om deze vraag te kunnen beantwoorden zijn studies nodig waarin gepaarde biopten van de primaire tumor en metastase van dezelfde patiënt onderzocht worden.

Naast onderzoek naar enkele mutaties in genen, hebben we ook gekeken naar mutatieprofielen, ook wel bekend als "**mutational signatures**". We observeerden een toename van bepaalde signatures in de gemetastaseerde setting zoals signatures 2 en 13. Deze signatures zijn geassocieerd met een specifiek mutatiemechanisme APOBEC. Daarnaast vonden we dat signature 17 geassocieerd was met eerdere behandeling, meest waarschijnlijk 5-FU.

Mutational signatures
Mutational signatures zijn mutatiepatronen. Deze mutatiepatronen kunnen veroorzaakt worden door DNA schade en/of afwezigheid van reparatie, waardoor specifieke veranderingen optreden in het DNA die herkend kunnen worden als patroon. Voor een deel van de mutational signatures is duidelijk door welk mechanisme deze mutaties zijn opgetreden. Zo is er een specifiek signature dat veroorzaakt wordt door UV-licht en een ander signature dat wordt veroorzaakt door roken.

DNA-profielen en therapie op maat

Op basis van de WGS data hebben we drie groepen van patiënten kunnen identificeren die baat zouden kunnen hebben van doelgerichte therapie op basis van hun DNA-profiel:

1. Tumoren met veel mutaties (een hoge tumor mutational burden (>10 mutaties/Mb)) maakt deze in theorie gevoelig voor immunotherapie;
2. Tumoren met een defect in de homologe recombinatie, herkend door een specifiek mutatiepatroon, als gevolg van mutaties in *BRCA1, BRCA2* of andere genen die betrokken zijn bij reparatie van het DNA maakt deze tumoren gevoelig voor platinum-bevattende chemotherapie en/of PARP-remmers;
3. Tumoren met een specifieke genetische afwijking waarvoor reeds een doelgericht medicijn is ontwikkeld en goedgekeurd door de FDA.

Voor in totaal 42% van de patiënten kan op basis van bovenstaande afwijkingen een doelgerichte therapie voorgesteld worden. Of deze DNA-gerichte behandeling daadwerkelijk succesvol zal zijn, moet in prospectieve studies aangetoond worden.

DNA analyse en behandeling met specifieke therapie

Bij een geïndividualiseerde behandeling is het van belang om de juiste behandeling aan de juiste patiënt, op het juiste moment te geven. Bij de selectie van de juiste therapie voor de juiste patiënt kan DNA analyse van de tumor een belangrijke rol spelen. Het doel hiervan is om patiënten met tumoren die resistent zijn (= ongevoelig) tegen deze behandeling, de behandeling te onthouden en patiënten met tumoren die juist gevoelig zijn, te selecteren. In **hoofdstuk 3** hebben we met behulp van WGS onderzocht of bepaalde DNA afwijkingen die aanwezig zijn in het biopt voordat chemotherapie in de vorm van capecitabine wordt gegeven, geassocieerd zijn met respons op deze behandeling. Capecitabine is een chemotherapeuticum dat onder andere wordt voorgeschreven aan patiënten met gemetastaseerd mammacarcinoom. In deze studie waren de volgende DNA veranderingen geassocieerd met respons: mutaties in *TP53*, *PTPRS*, *HMCN1, CEP350,* en *ADGRG4*, amplificatie van 17q23.1 en verlies van 4p16.3 en twee signatures (COSMIC mutational signature 16 en rearrangement signature 1). Belangrijke beperkingen van deze studie zijn dat er slechts 73 patiënten geïncludeerd werden en dat patiënten op verschillende manieren voorbehandeld zijn geweest. Deze resultaten zullen gevalideerd worden in een onafhankelijk Frans cohort van patiënten die tevens behandeld zijn met capecitabine monotherapie en van wie voor start van deze behandeling DNA data beschikbaar is om dezelfde analyses te verrichten.

Verworven resistentie

Zoals hierboven beschreven zou de behandeling van patiënten gepersonaliseerd kunnen worden als voor start van de behandeling duidelijk is wie wel of niet resistent is voor de behandeling. Naast "intrinsieke resistentie", waarbij een tumor op voorhand al resistent is tegen een bepaald middel, kan een tumor ook gedurende behandeling veranderen waardoor de tumor die initieel wel gevoelig was, na een bepaalde tijd resistent wordt. Dit wordt ook wel verworven resistentie genoemd. Bij patiënten met borstkanker is recent een mutatie beschreven in de oestrogeen receptor, de *ESR1* mutatie, die er voor zorgt dat de oestrogeen receptor geactiveerd blijft ondanks behandeling met anti-hormonale therapie met aromataseremmers. In **hoofdstuk 4** wordt een overzicht gegeven van de pre-klinische en klinische studies met betrekking tot deze *ESR1* mutaties. Op dit moment zijn er geneesmiddelenstudies gaande die onderzoeken of deze verworven resistentie in *ESR1* teniet kan worden gedaan. Als deze nieuwe hormonale behandelingen beschikbaar komen, zal de bepaling van deze mutatie klinisch relevant worden om de keuze voor de meest optimale behandeling te begeleiden.

Integratie van DNA en RNA data

Om te evalueren wat het effect van bepaalde DNA veranderingen op RNA niveau is, hebben we in **hoofdstuk 5** voor een cohort van 101 patiënten met gemetastaseerd mammacarcinoom de RNA en DNA data geïntegreerd. Analyse van de RNA data toonde twee groepen van samples: een cluster van samples dat *ESR1* en de genen die door de oestrogeen receptor geactiveerd worden hoog tot expressie bracht en een cluster met samples met een veel lagere expressie van deze genen. In het cluster waarin *ESR1* hoog tot expressie werd gebracht waren ook meer *ESR1* mutaties aanwezig. Daarnaast kwamen er meer amplificaties van het *FGFR1* gen voor. Deze resultaten tonen dat de oestrogeen receptor pathway in deze tumoren waarschijnlijk nog steeds een belangrijke rol speelt en dat het blokkeren van deze oestrogeen receptor pathway voor deze tumoren een belangrijke hoeksteen van de behandeling blijft.

Deel II: Vloeibare biopten (liquid biopsies)

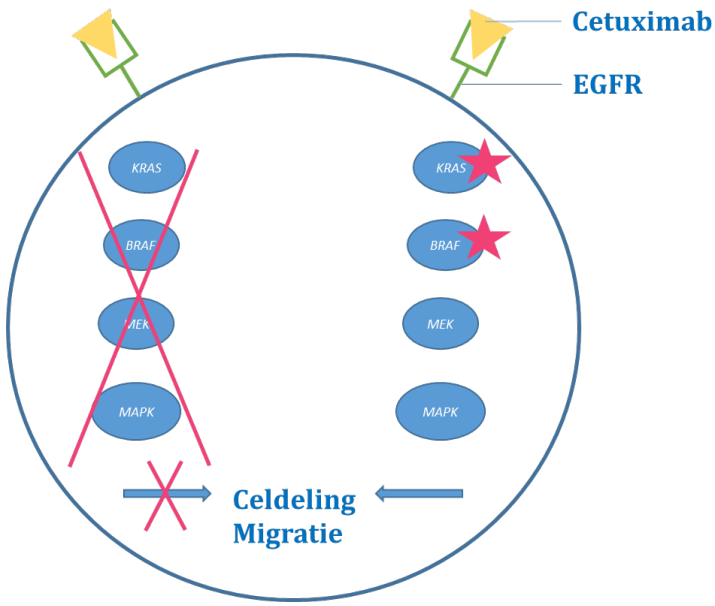
Ondanks dat de analyse van tumorbiopten reeds meer inzicht gegeven heeft over welke genomische afwijkingen er aanwezig zijn in metastasen blijft het nemen van biopten uitdagend. Dit heeft met name te maken met het feit dat het (herhaaldelijk) nemen van biopten niet zonder risico is en dat een weefselbiopt slechts een afspiegeling geeft van het ene kleine stukje gebiopteerde weefsel. Vloeibare biopten (hierna: liquid biopsies) zijn minder invasief. In eigenlijk alle vloeistoffen van het lichaam, zoals bloed, hersenvocht en urine kunnen tumorcellen (circulerende tumor cellen), delen van tumor cellen of celvrij tumor DNA voorkomen. Deze tumorcellen en/of tumor DNA kunnen vervolgens geanalyseerd worden. Een aantal voorbeelden waarvoor deze tumorcellen / tumor DNA ingezet zouden kunnen worden:

- Detectie van minimaal residuale ziekte
- Vroeg detectie van kanker
- Risicostratificatie
- Detectie van mutaties die doelgerichte behandeling mogelijk maken
- Detectie van resistentie mechanismen

Van de bovenstaande voorbeelden waar liquid biopsies voor gebruikt zouden kunnen worden, worden enkele hieronder uitgewerkt.

Detectie van resistentie mechanismen

In **hoofdstuk 6** wordt een studie beschreven waarbij van patiënten met gemetastaseerd darmkanker bloed werd afgenomen voor start van de behandeling met cetuximab. Cetuximab is een middel dat bindt aan de *epidermal growth factor receptor (EGFR)*. Hierdoor wordt de intracellulaire signaaltransductie onderdrukt hetgeen resulteert in onderdrukking van de celdeling. Voor deze behandeling zijn reeds resistentie mechanismen beschreven. Patiënten die in de tumor een mutatie in *RAS* of *BRAF* hebben reageren niet of nauwelijks op deze behandeling. Dit wordt veroorzaakt doordat mutaties in deze genen de signaal transductieroute onder het niveau van EGFR activeren, waardoor het effect van de EGFR blokkade teniet wordt gedaan (**Figuur 1**). Bij patiënten die in aanmerking komen voor behandeling met cetuximab wordt voor start van deze therapie onderzocht of deze mutaties aanwezig zijn en indien aanwezig, zal de patiënt niet met cetuximab behandeld worden. In **hoofdstuk 6** laten we zien dat er patiënten zijn met deze mutaties in het bloed, die niet aanwezig zijn in de primaire tumor. Bij patiënten die mutaties in de tumor en/of het bloed hadden, hadden een slechtere uitkomst op de cetuximab therapie. Daarnaast ontwikkelde het merendeel van de patiënten die progressie vertonen tijdens de behandeling met cetuximab *RAS, BRAF* en/of *EGFR* mutaties.



Figuur 1 - Schematische weergave van de werking van de werking van Cetuximab.

Naast de detectie van mutaties in een geselecteerde set aan genen in het bloed van patiënten, zijn er ook technieken ontwikkeld waarbij afwijkingen in alle genen gedetecteerd kunnen worden. Deze techniek heet *whole exome sequencing (WES).* **Hoofdstuk 7** is een meta-analyse waarin wordt beschreven dat het technisch haalbaar is om WES toe te passen op celvrij DNA. Deze meta-analyse toont dat 50% van de varianten die aanwezig zijn in tumorweefsel ook gedetecteerd worden met WES op celvrij DNA. Een hogere tumorfractie in het bloed was geassocieerd met een hoger aantal gedetecteerde varianten in het bloed. Een sub-analyse toonde ook dat als alleen de samples met een tumor fractie van ≥25% werden geselecteerd, de sensitiviteit van 50% naar 69% toenam. Ondanks dat deze techniek nog niet direct toepasbaar is in de klinische praktijk, zijn studies die WES gebruiken op gepaarde samples wel heel waardevol omdat analyse van bloedsamples van voor en na therapie inzicht kan geven in welke resistentie mechanismen optreden.

Risicostratificatie

Binnen de oncologie wordt risicostratificatie regelmatig toegepast om te selecteren welke patiënten baat zouden kunnen hebben van aanvullende (adjuvante) behandeling met medicijnen na bijvoorbeeld chirurgie om de primaire tumor te verwijderen, In **hoofdstuk 10**, wordt beschreven hoe circulerende tumorcellen gebruikt kunnen worden om het risico op het ontwikkelen van hersenmetastasen bij patiënten met gemetastaseerd borstkanker te voorspellen. In deze circulerende tumorcellen kan onderzocht worden welke genen er “aan” of “uit” staan. Op basis hiervan kan een voorspelling gemaakt worden welke patiënten er een hoger risico hebben op het ontwikkelen van hersenmetastasen en bij wie de clinicus hierop dus mogelijk meer bedacht moet zijn. In deze studie waren naast de klinische parameters zoals een ER-negatieve primaire tumor en een jongere leeftijd, ook de expressie van genen *MAGEA3*, *PLAU*, *TSPAN13*, en *CD44* geassocieerd met de ontwikkeling van hersenmetastasen.

Analyses van andere lichaamsvochten: hersenvocht

Het verkrijgen van tumorweefsel van primaire hersentumoren, hersenmetastasen of leptomeningeale metastasen (metastasen in de hersenvliezen) brengt een extra uitdaging met zich mee, omdat het risico op complicaties van de weefselafname nog groter is en in het geval van de hersenvliezen vaak ook niet mogelijk om een biopt te nemen. Hersenvocht (liquor) dat in contact staat met de tumorcellen kan derhalve een aantrekkelijke bron van DNA zijn waarop onderzoek gedaan kan worden. In **hoofdstuk 8**, zijn de diagnostische technieken om leptomeningeale metastasen te detecteren uiteengezet. In deze studie komt naar voren dat het gebruik van celvrij DNA en circulerende tumorcellen in hersenvocht aantrekkelijke technieken lijken om

leptomeningeale metastasen te detecteren. In **hoofdstuk 9**, is celvrij DNA geanalyseerd van liquoren van patiënten met borstkanker die een lumbaal punctie hebben ondergaan vanwege een klinische verdenking op leptomeningeale metastasen. In deze studie werd getoond dat de meerderheid van de liquoren waarin ook tumorcellen werden gezien door de patholoog, wat de gouden standaard is voor de diagnose leptomenigneale metastasen, ook tumor DNA aantoonbaar was. Dit tumor DNA werd aangetoond door de mFAST-SeqS methode. Dit is een methode die de aanwezigheid van een abnormaal aantal chromosoom(armen) analyseert. Normale cellen bevatten 46 chromosomen, die bij de celdeling verdubbelen en verdelen over de dochtercellen. Bij kankercellen kan deze verdeling fout gaan waardoor cellen ontstaan met meer of minder chromosomen. Dit wordt aneuploïdie genoemd. In de liquor samples van patiënten werd bij een deel van de samples aneuploïdie aangetoond voordat de klinische diagnose hersen- of leptomeningeale metastasen werd gesteld. Ook was een hoge aneuploïdie score geassocieerd met een kortere overleving en met de ontwikkeling van leptomeningeale metastasen. De resultaten van deze studie zijn veelbelovend en prospectieve validatie is nodig om deze techniek te introduceren in de diagnostiek naar leptomenigneale metastasen.

Deel III: Radiomics

Naast het nemen van biopten en analyseren van liquid biopsies, is er nog een derde methode om tumoren te karakteriseren: radiomics. Dit onderzoeksgebied wint snel terrein als patiëntvriendelijk alternatief voor weefselbiopten. In **hoofdstuk 11** is onderzocht of door het gebruik van CT-beelden mogelijk is om onderscheid te maken tussen in *BRAF* mutatiestatus bij patiënten met gemetastaseerd melanoom. Bij patiënten met gemetastaseerd melanoom is het van belang om onderscheid te maken tussen tumoren met of zonder *BRAF* mutatie, omdat patiënten met deze mutatie in aanmerking komen voor een behandeling waarbij BRAF geremd wordt. Aanwezigheid van deze mutatie biedt dus een extra behandeloptie. Normaal gesproken wordt de mutatiestatus bepaald op het primaire melanoom of er wordt een biopt genomen van een metastase waarop de analyse wordt uitgevoerd. Omdat het nemen van biopten risico's met zich meebrengt en analyse van de primaire tumor vaak enige dagen in beslag neemt, zou een minimaal invasieve methode zoals radiomics gewenst zijn. In **hoofdstuk 11** wordt getoond dat radiomics de *BRAF* mutatiestatus niet kan onderscheiden. Deze resultaten laten zien dat ondanks dat radiomics een veelbelovende onderzoekstechniek is, het niet waarschijnlijk is dat alle moleculaire details herkend kunnen worden door gebruik te maken van machine learning van radiomics data.

Deel IV: Integratie van verschillende technieken

Alle drie de technieken – weefselbiopten, liquid biopsies en radiomics – hebben voor- en nadelen. Waarschijnlijk zal geen van deze technieken een van de andere technieken compleet kunnen vervangen. Het is dus belangrijk dat de verschillende methoden te vergelijken en de data verkregen met de verschillende technieken te integreren. In **hoofdstuk 12**, is de ER status op drie verschillende manieren geanalyseerd: oestrogeen receptor kleuring op een metastase biopt (gouden standaard); FES-PET en *ESR1* expressie in circulerende tumorcellen. In deze studie zijn *ESR1* expressie in circulerende tumorcellen en FES-PET analyses veelbelovende alternatieven voor de ER kleuring op het biopt. We erkennen dat weefselanalyse aanvullende informatie oplevert zoals de HER2-status. Bij patiënten met nieuw gediagnosticeerde gemetastaseerde borstkanker zal het nemen van biopten vooralsnog de gouden standaard blijven totdat er beeldvormende of liquid biopsy technieken zijn die de HER2-status nauwkeurig en accuraat kunnen bepalen. De FES-PET kan echter wel een vollediger overzicht geven van alle metastatische locaties die ER tot expressie brengen en zou nuttig kunnen zijn wanneer sommige laesies reageren op behandeling, maar andere niet.

Conclusie

In dit proefschrift worden verschillende genomische analyses beschreven met het uiteindelijke doel om de behandeling voor patiënten met kanker te personaliseren. Hiervoor zijn gegevens afkomstig van drie modaliteiten gebruikt: weefselbiopsieën, vloeibare biopten en radiomics. Op basis van de beschreven studies kunnen we concluderen dat elke techniek zijn eigen voor- en nadelen heeft en dat de onderzoeks- en/of klinische vraag gematcht moet worden met de juiste techniek(en) en dat daarbij rekening gehouden moet worden met de voor- en nadelen. Grootschalige klinische implementatie van WGS op biopten, liquid biopsy analyses en radiomics hangt af van onderzoeken die aantonen dat het gebruik leidt tot een overlevingsvoordeel en/of een betere kwaliteit van leven.

CURRICULUM VITAE

Lindsay Angus werd op 14 juni 1991 geboren in Den Haag. Ze groeide op in Schiedam en doorliep de lagere school op Basisschool "De Regenboog". Op de middelbare school was zij reeds geïnteresseerd in geneeskunde. Gedurende 5 en 6 gymnasium volgde zij het *Junior* Med School programma van het Erasmus MC. Als onderdeel van dit programma deed zij onderzoek naar circulerende tumor cellen in bloed van patiënten met borstkanker, onder supervisie van Prof. dr. E.M.J.J. Berns, Prof. dr. J.W.M. Martens en Dr. B. Mostert. In 2009 behaalde zij haar Gymnasium diploma aan Scholen Gemeenschap Spieringshoek te Schiedam. In datzelfde jaar begon zij aan de studie geneeskunde aan de Erasmus Universiteit te Rotterdam. Tijdens haar studie geneeskunde deed zij onderzoekservaring op bij de afdeling Interne Geneeskunde. Onder supervisie van Dr. J.G. Langendonk deed zij onderzoek naar zwangerschappen van patiënten met erfelijke stofwisselingsziekten wat in 2012 resulteerde in een eerste wetenschappelijke publicatie. In 2012 behaalde zij cum laude de Bachelor of Science en in datzelfde jaar won zij de Baltimore-Beurs, waarmee zij haar afstudeeronderzoek mocht verrichten aan Johns Hopkins University, Baltimore, Marlyand, USA. In 2013 verbleef ze voor zes maanden in Baltimore en verrichtte haar afstudeeronderzoek onder supervisie van Prof. S. Sukumar en Prof. dr. E.M.J.J. Berns naar gemethyleerd celvrij DNA in serum van patiënten met stadium IV borstkanker. Na het afronden van haar afstudeeronderzoek begon zij in september 2013 met haar coschappen en behaalde na een keuze coschap Interne Oncologie in het Antoni van Leeuwenhoek ziekenhuis en een oudste coschap op de afdeling Interne Oncologie in het Erasmus MC in 2015 cum laude de Master of Science. Aansluitend startte zij als arts-onderzoeker een promotie-onderzoek op de afdeling Interne Oncologie van het Erasmus MC te Rotterdam, onder supervisie van Prof.dr. S. Sleijfer, Prof.dr. J.W.M. Martens en Dr. A. Jager, waarvan de resultaten in dit proefschrift beschreven staan. Tijdens haar promotie-onderzoek kreeg zij de mogelijkheid om op nationale en internationale congressen haar werk te presenteren, onder andere tijdens het San Antonio Breast Cancer Symposium in 2018. In mei 2020 is zij begonnen als ANIOS (arts-niet-in-opleiding-tot-specialist) Interne Geneeskunde in het IJsselland ziekenhuis te Capelle aan den IJssel (opleider dr. E.L.E. de Bruijne). In januari 2021 is zij begonnen met de opleiding Interne Geneeskunde van het Erasmus MC (opleider Dr. A.A.M. Zandbergen) in het Albert Schweitzer Ziekenhuis te Dordrecht (opleider Dr. P.J.H. Smak Gregoor). In de toekomst hoopt zij het werk als internist-oncoloog te combineren met uitvoeren van translationeel onderzoek.

PhD PORTFOLIO

1.PhD Training	Year	Workload (ECTS)
General courses		
Basic course regulations and organization for clinical researchers (BROK)	2016	1.5
Research Integrity	2016	0.3
Biomedical English Writing and Communication	2017	3.0
Specific courses		
Biomedical Research Techniques	2013	1.0
Basic introduction course on SPSS	2016	1.0
Biostatistical Methods I: Basic principles (CCO2)	2016	5.7
Annual course on Molecular Medicine (Solid tumors, infections & host response)	2016	0.3
Circulating tumor cell isolation and diagnostics (CellSearch system)	2016	1.0
Real Time PCR tour training, ThermoFisher	2016	0.3
Minisymposium: wetenschappelijk onderzoek met mensen	2016	0.2
Novel options for cancer imaging: focus on urological tumors	2016	0.1
Workshop: Omgaan met groepen	2016	0.2
Photoshop and Illustrator CS6 workshop	2017	0.3
Course on R	2017	1.8
Workshop: Training on coaching medical students	2018	0.2
OpenClinica database building	2018	0.3
Teach the Teacher I	2018	0.6
NGS in DNA Diagnostics Course	2018	1.0
The Workshop UCSC Genome Browser- display engine for NGS sequencing data	2018	0.6
Microsoft Access: Basic workshop	2019	0.3
Microsoft Access: Advanced workshop	2019	0.4
Female Talent Class	2019	1.0
(Inter)national conferences		
Molecular Medicine Day, Rotterdam	2016-2019	1.2
Scientific meeting Medical Oncology: Erasmus MC, Rotterdam	2016-2019	0.8
Young oncologist evening: Erasmus MC, Rotterdam	2016, 2017, 2019	0.6
EORTC meeting, Rotterdam	2016	0.6
Center for personalized Cancer Treatment (CPCT) Symposium, Utrecht	2016	0.3
Daniel den Hoed Day, Rotterdam	2016, 2017	0.6
CGC annual meeting, Utrecht	2017-2018	0.6
CGC annual meeting New Horizons in Cancer Research, Amsterdam	2017	0.6
CMBD themadag Cell Free DNA, Utrecht	2017	0.3
LKI Symposium – Liquid Biopsies & Cancer, Leuven, Belgium	2017	1.0
Borstkanker Behandeling Beter Symposium, Rotterdam	2017-2019	0.6
San Antonio Breast Cancer Symposium, San Antonio, USA	2017	1.0
Association of Molecular Pathology, Rotterdam	2018	0.3
European Society for Medical Oncology (ESMO), Munich, Germany	2018	1.0
San Antonio Breast Cancer Symposium, San Antonio, USA	2018	1.0
CGC-Oncode annual scientific meeting, Amsterdam	2019	0.3
European Society for Medical Oncology (ESMO), Barcelona, Spain	2019	1.0

1.PhD Training	Year	Workload (ECTS)
Poster presentations		
CGC annual meeting New Horizons in Cancer Research, Amsterdam	2017	1.0
MolMed day, Rotterdam	2018	0.6
Association of Molecular Pathology, Rotterdam	2018	1.0
European Society for Medical Oncology (ESMO), Munich, Germany	2018	1.0
San Antonio Breast Cancer Symposium, San Antonio, USA	2018	1.0
Oral presentations		
Medical Oncology Research Meeting, Rotterdam	2016	0.2
Borstkanker Behandeling Beter Symposium, Rotterdam	2017	0.2
Medical Oncology Research Meeting, Rotterdam	2017	0.2
CGC annual meeting, Utrecht	2018	0.2
Association of Molecular Pathology, Rotterdam	2018	0.2
Scientific meeting Medical Oncology: Erasmus MC, Rotterdam	2018-2019	0.4
Borstkanker Behandeling Beter Symposium, Rotterdam	2018	0.2
San Antonio Breast Cancer Symposium, San Antonio, USA	2018	1.0
CGC-Oncode annual conference, Amsterdam	2019	0.2
Hartwig Medical Foundation, Raad van Toezicht, Amsterdam	2019	0.2
Young oncologist evening: Erasmus MC, Rotterdam	2019	0.2
MolMed Day	2019	0.2
Internal Medicine Clinical Demonstration, Erasmus MC	2019	0.2
Borstkankersymposium, Doorn	2019	0.4
Borstkanker onderzoeksgroep, Utrecht	2019	0.2
DCC-NET, Nettetal, Duitsland	2019	0.2
Nederlandse Oncologiedagen, Arnhem	2019	0.3
Oncode Institute, Clinical Workshop: Breast Cancer, Utrecht	2020	0.5

2. Teaching	Year	Workload (ECTS)
Lecturing		
Junior Med School Oncology course, Erasmus MC, Rotterdam	2017, 2018	0.6
Medical school training- associated teaching and counselling		
Supervisor/coordinator 4-week Junior Med School Medical Oncology Research Program	2016, 2017	1.5
Tutorial class first-year medical students	2016- 2018	4.5
Supervisor “clinical orientation on the medical profession” for first-year medical students	2017, 2018	1.0
Medical school bachelor phase coaching program	2018-2020	1.5
Supervising students in extracurricular research		
Ana Rajicic	2017-2019	1.0
Laura Pasquet (France)	2019	1.0
Romy Klein-Kranenbarg	2017-2018	1.0
Frederique Meinsma	2018-2019	1.0

3. Other	Year	Workload (ECTS)
Grant allocation		
KWF (together with prof. dr. S. Sleijfer): €400,511.30	2016	
MRace (together with prof. J. Gribnau): €50,000	2016	
Peer review of manuscripts for international peer-reviewed journals		
BMC Cancer	2017	
Cancer Biomarkers	2018	
Molecular Oncology	2019	
Cancer Medicine	2019	
Organization meetings		
Medical Oncology Research Meeting	2017-2018	1.0

LIST OF PUBLICATIONS

L Angus*, MPA Starmans*, A Rajicic, AE Odink, M Jalving, WJ Niessen, Jacob J Visser, Stefan Sleijfer, Stefan Klein*, Astrid A M van der Veldt*
The BRAF P.V600E Mutation Status of Melanoma Lung Metastases Cannot Be Discriminated on Computed Tomography by LIDC Criteria nor Radiomics Using Machine Learning
Journal of personalized medicine. 2021 Apr 1;11(4):257. doi: 10.3390/jpm11040257.

L Angus, T Deger, A Jager, JWM Martens, V de Weerd, I van Heuvel, MJ van den Bent, PA Sillevis Smitt, JM Kros, EM Bindels, E Heitzer, S Sleijfer, JLM Jongen, SM Wilting
Detection of aneuploidy in cerebrospinal fluid from patient with breast cancer can improve diagnosis of leptomeningeal metastases
Clinical Cancer Research. 2021 May 15;27(10):2798-2806. doi: 10.1158/1078-0432.CCR-20-3954.

PAJ Mendelaar, M Smid, J van Riet, **L Angus**, M Labots, N Steeghs, MP Hendriks, GA Cirkel, JM van Rooijen, AJ ten Tije, MP Lolkema, E Cuppen, S Sleijfer, JWM Martens, SM Wilting
Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features
Nature Communications. 2021 Jan 25;12(1):574. doi: 10.1038/s41467-020-20887-6.

MK Bos, K Nasserinejad, MPHM Jansen, **L Angus**, PN Atmodimedjo, E de Jonge, WNM Dinjens, RHN van Schaik, M Del Re, HJ Dubbink, S Sleijfer, JWM Martens
Comparison of variant allele frequency and number of mutant molecules as units of measurement for circulating tumor DNA
Molecular Oncology. 2020 Feb; 83:101951. doi: 10.1016/j.ctrv.2019.101951. Epub 2019 Dec 13.

MK Bos*, **L Angus***, K Nasserinejad, A Jager, MPHM Jansen, JWM Martens, S Sleijfer
Whole exome sequencing of cell-free DNA – a systematic review and Bayesian individual patient data meta-analysis.
Cancer Treatment Reviews. 2019 Dec 13;83:101951. doi: 10.1016/j.ctrv.2019.101951.

L Angus, RP Peeters, AAM van der Veldt
Drug effects on the thyroid.
The New England Journal of Medicine. 2019 Nov 14;381(20):1980. doi: 10.1056/NEJMc1912672.

L Angus, M Smid, SM Wilting, J van Riet, A van Hoeck, L Nguyen, S Nik-Zainal, TG Steenbruggen, VCG Tjan-Heijnen, M Labots, JMGH van Riel, HJ Bloemendal, N Steeghs, MP Lolkema, EE Voest, HJG van de Werken, A Jager, E Cuppen, S Sleijfer, JWM Martens
The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies.
Nature Genetics. 2019 Oct; 51(10):1450-1458. doi: 10.1038/s41588-019-0507-7. Epub 2019 Sep 30.

EJ van Helden*, **L Angus***, CW Menke-van der Houven van Oordt, DAM Heideman, E Boon, SC van Es, SA Radema, CML van Herpen, DJA de Groot, EGE de Vries, MPHM Jansen, S Sleijfer, HMW Verheul
RAS and *BRAF* mutations in cell-free DNA are predictive for outcome of cetuximab monotherapy in patients with tissue-tested *RAS* wild-type advanced colorectal cancer.
Molecular Oncology. 2019 Nov; 13(11):2361-2374. doi: 10.1002/1878-0261.12550. Epub 2019 Sep 30.

SC van Es, B van der Vegt, F Bensch, S Gerritse, EJ van Helden, **L Angus**, J Overbosch, CW Menke-van der Houven van Oordt, HMW Verheul, CML van Herpen, A Jager, SF Oosting, EGE de Vries, CP Schröder
Decalcification of breast cancer bone metastasis with EDTA does not affect ER, PR, and HER2 results.
The American Journal of Surgical Pathology. 2019 Oct;43(10):1355-1360. doi: 10.1097/PAS.0000000000001321.

SR Verhoeff, SC van Es, E Boon, EJ van Helden, **L Angus**, SG Elias, SF Oosting, EH Aarntzen, AH Brouwers, TC Kwee, S Heskamp, OS Hoekstra, HMW Verheul, AAM van der Veldt, EGE de Vries, OC Boerman, WTA van der Graaf, WJG Oyen, CML van Herpen
Lesion detection by ^{89}Zr-girentuximab and ^{18}F-FDG PET/CT in patients with newly diagnosed metastatic renal cell carcinoma
European Journal of Nuclear Medicine and Molecular Imaging. 2019 Aug;46(9):1931-1939. doi: 10.1007/s00259-019-04358-9.

L Angus, JWM Martens, MJ van den Bent, PAE Sillevis Smitt, S Sleijfer, A Jager
Novel methods to diagnose leptomeningeal metastases in breast cancer.
Neuro Oncology. 2019 Mar 18;21(4):428-439. doi: 10.1093/neuonc/noy186.

SR Vitale, AM Sieuwerts, N Beije, J Kraan, **L Angus**, B Mostert, EA Reijm, NM Van, R van Marion, LY Dirix, P Hamberg, FE de Jongh, A Jager, JA Foekens, P Vigneri, S Sleijfer, MPHM Jansen, JWM Martens
An optimized workflow to evaluate estrogen receptor gene mutations in small amounts of cell-free DNA
The Journal of Molecular Diagnostics. 2019 Jan;21(1):123-137. doi: 10.1016/j.jmoldx.2018.08.010.

L Angus*, N Beije*, A Jager, JWM Martens, S Sleijfer
ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients
Cancer Treatment Reviews. 2017 Jan;52:33-40. doi: 10.1016/j.ctrv.2016.11.001.

JG Langendonk , JC Roos, **L Angus**, M Williams, FP Karstens, JB de Klerk, C Maritz, T Ben-Omran, C Williamson, RH Lachmann, E Murphy
A series of pregnancies in women with inherited metabolic disease.
Journal of Inherited Metabolic Disease. 2012 May;35(3):419-24. doi: 10.1007/s10545-011-9389-2.

** Authors contributed equally*

AUTHOR AFFILIATIONS

Laurens V. Beerepoot	Department of Internal Medicine, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands
Nick Beije	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Martin J. van den Bent	Department of Neurology, The Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Eric M.J. Bindels	Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Haiko J. Bloemendal	Department of Medical Oncology, Meander Medical Center, Amersfoort, The Netherlands
Eline Boon	Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands
Manouk K. Bos	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Adrienne H. Brouwers	University of Groningen, University Medical Center Groningen, Medical Imaging Center, Department of Nuclear Medicine and Molecular Imaging, Groningen, The Netherlands
Edwin Cuppen	Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht, The Netherlands Hartwig Medical Foundation, Amsterdam, The Netherlands
Teoman Deger	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Bertha Eisses	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Sjoerd G. Elias	Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
Jasper Emmering	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
Suzanne C. van Es	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Sophie L. Gerritse	Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands
Andor W.J.M. Glaudemans	University of Groningen, University Medical Center Groningen, Medical Imaging Center, Department of Nuclear Medicine and Molecular Imaging, Groningen, The Netherlands
Derk Jan A. de Groot	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Daniëlle A.M. Heideman	Department of Pathology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Joan B. Heijns	Department of Internal Medicine, Amphia Hospital, Breda, The Netherlands
Ellen Heitzer	Christian Doppler Laboratory for Liquid Biopsies for Early Detection of Cancer, Institute of Human Genetics, Diagnostic and Research Center for Molecular BioMedicine, Medical University of Graz, Graz, Austria
Erik J. van Helden	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Carla M.L. van Herpen	Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands

Irene van Heuvel	Department of Neurology, The Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Arne Van Hoeck	Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht, The Netherlands
Otto S. Hoekstra	Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Radiology and Nuclear Medicine, Cancer Center Amsterdam, Amsterdam, The Netherlands
Agnes Jager	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Mathilde Jalving	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Maurice P.H.M. Jansen	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Joost L.M. Jongen	Department of Neurology, The Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Stefan Klein	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
Inge R.H.M. Konings	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Jaco Kraan	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Johan M. Kros	Department of Pathology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands
Mariette Labots	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Martijn P. Lolkema	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
John W.M. Martens	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
C. Willemien Menke - van der Houven van Oordt	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Kazem Nasserinejad	HOVON Data Center, Department of Hematology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Luan Nguyen	Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht, The Netherlands
Wiro J. Niessen	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands
Arlette E. Odink	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
Wim J.G. Oyen	Department of Nuclear Medicine, Radboud Medical Center, Nijmegen, The Netherlands

Sandra A. Radema	Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands
Ana Rajicic	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Johanna M.G.H. van Riel	Department of Internal Medicine, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands
Job van Riet	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands Cancer Computational Biology Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands Department of Urology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Carolina P. Schröder	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands
Anieta M. Sieuwerts	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands † Deceased August 2019
Peter A.E. Sillevis Smitt	Department of Neurology, The Brain Tumor Center at Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Stefan Sleijfer	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Marcel Smid	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Martijn P.A. Starmans	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
Neeltje Steeghs	Department of Medical Oncology, the Netherlands Cancer Institute, Amsterdam, The Netherlands
Tessa G. Steenbruggen	Department of Medical Oncology, the Netherlands Cancer Institute, Amsterdam, The Netherlands
Vivianne C.G. Tjan-Heijnen	Department of Medical Oncology, GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands
Astrid A.M. van der Veldt	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Bert van der Vegt	University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, The Netherlands
Henk M.W. Verheul	Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Sarah Verhoeff	Department of Medical Oncology, Radboud University Medical Center, Nijmegen, The Netherlands
Jacob J. Visser	Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
Emile E. Voest	Department of Medical Oncology, the Netherlands Cancer Institute, Amsterdam, The Netherlands
Elisabeth G.E. de Vries	Department of Medical Oncology, University Medical Center Groningen, Groningen, The Netherlands

Vanja de Weerd	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Harmen J.G. van de Werken	Cancer Computational Biology Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands Department of Urology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Saskia M. Wilting	Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
Agnes J. van de Wouw	Department of Medical Oncology, VieCuri Medical Center, Venlo, The Netherlands
Serena Nik-Zainal	Department of Medical Genetics, The Clinical School, University of Cambridge, Cambridge, UK

DANKWOORD

Zoals u wellicht heeft gezien en gelezen, zijn de hoofdstukken in mijn proefschrift tot stand gekomen door nauwe samenwerkingsverbanden tussen verschillende afdelingen en instituten. Ik ben dan ook oprecht heel veel mensen dankbaar voor hun inzet en bereidheid om hun kennis en ervaring in te zetten voor de verschillende projecten. Mensen die mij goed kennen, weten dat ik gek ben op bloemen, het liefst in zoveel mogelijk verschillende kleuren. De verscheidenheid aan projecten maakte dat ik mocht samenwerken met velen en het zijn deze mensen die kleur hebben gegeven aan mijn promotietijd.

Allereerst wil ik grote waardering uitspreken voor alle patiënten en hun naasten die deelgenomen hebben aan de verschillende klinische studies beschreven in dit proefschrift. Zonder er zelf baat bij te hebben, hebben zij extra biopten, bloedafnames en scans ondergaan om de wetenschap een stap vooruit te brengen. Velen vertelden mij mee te willen doen om de behandeling voor toekomstige patiënten te verbeteren. Ik vind het groots dat zij het belang van toekomstige patiënten in ogenschouw konden nemen op het moment dat zij zelf door een verdrietige en moeilijke periode van hun leven gingen.

Een proefschrift komt er niet zonder een betrokken promotieteam.

Prof. dr. Sleijfer, beste Stefan, bedankt voor het vertrouwen dat je in mij gesteld hebt. De uitdagende projecten die ik de afgelopen jaren onder jouw hoede heb kunnen uitvoeren, hebben ertoe geleid dat ik zowel als onderzoeker maar zeker ook als persoon gegroeid ben. Waar ik in het begin vaak twijfelachtig was over de uitkomst van een project, wist jij mij meermaals te vertellen dat het wel goed zou komen en dat gebeurde dan ook. Ik heb bewondering voor de manier waarop jij in je drukke schema wekelijks tijd maakt voor de “Liquid Biopsy Helden” en ervoor zorgt dat alle promovendi voldoende projecten hebben en dat er zelfs aan het eind van een promotietraject nog ruimte is om nieuwe projecten op te starten zodat de volgende, volgens jouw “dakpanconstructie”, de onderzoekslijn kan voortzetten. Daarnaast is de snelheid waarop jij manuscripten beoordeelt ongekend (soms naar mijn idee iets te snel, omdat ik blij was dat het even niet meer op mijn bureau lag ;-)). Naast onderzoek was er zeker ook tijd voor gezelligheid, zo ben ik er op de skireizen achter gekomen dat jij ook talent hebt voor verkleedpartijtjes, beschikt over een paar hele goede dansmoves en zelfs een duet van Marco & Davine gaat jou niet boven de pet. Ik wens je heel veel succes met je nieuwe functie als decaan.

Prof. dr. ir. Martens, beste John, wij kennen elkaar al sinds het moment dat ik als 17-jarige *"prutser"* op jouw lab CTCs uit bloed kwam vissen. Gelukkig mocht ik na mijn geneeskundestudie op het vertrouwde nest terugkeren om het "liquid biopsy" onderzoek voort te zetten. Ik heb de afgelopen jaren heel veel van je geleerd op het gebied van DNA en RNA sequencing en natuurlijk jouw stokpaardje "APOBEC". Alhoewel jij nog steeds regelmatig binnenkomt met de gevleugelde woorden "hé prutsers", hoop ik dat we in de toekomst weer samen zullen werken om de liquid biopsies daadwerkelijk naar de kliniek te brengen. Ik heb goede herinneringen aan de sportieve uitjes die we gehad hebben ("Save the Boobies run, hardlopen in San Antonio en het recente fietstocht(je)). Als klap op de vuurpijl kijk ik met veel trots terug op ons avontuur in San Antonio waarbij we de dag hilarisch afsloten met een rondrit in een "echte Cinderella koets".

Dr. Jager, lieve Agnes, wij leerden elkaar kennen tijdens mijn coschap chirurgie in de Daniel den Hoed. Wat ben ik blij dat ik toen tegen dr. Koppert, Linetta, heb durven zeggen dat ik oncoloog wilde worden en graag van de gelegenheid gebruik wilde maken om een keer op de poli mee te lopen met een oncoloog. Zo geschiedde en ik raakte direct betrokken bij jouw onderzoeksprojecten. Toen je me belde of ik als promovendus wilde starten na het afronden van de studie heb ik dan ook geen moment getwijfeld. Door de jaren heen hebben wij een prettige modus gevonden waarbij ik jou altijd kon vinden voor input. De manier en snelheid waarop jij nieuwe informatie in de context van de huidige literatuur kan plaatsen en vervolgens binnen 'no-time' de volgende klinisch relevante onderzoeksvraag weet te bedenken vind ik bewonderenswaardig. Ik hoop dat we de komende jaren samen blijven werken en dat ik nog veel meer van je mag leren op het gebied van onderzoek, en de behandeling en begeleiding van patiënten met mammacarcinoom.

De leden van de leescommissie, Prof. dr. van der Wall, Prof. dr. Van Laere en Prof. dr. Dinjens, wil ik hartelijk bedanken voor tijd die zij gestoken hebben in het kritisch lezen en beoordelen van mijn proefschrift.

De leden van de grote commissie, Prof. dr. Linn, Prof. dr. Wessels, Prof. dr. van den Bent en dr. Wilting, wil ik bedanken voor de bereidheid om plaats te nemen in de oppositie. Ik kijk ernaar uit om met eenieder van gedachten te wisselen tijdens de verdediging.

Prof. dr. Berns, lieve Els, vanaf het moment dat wij elkaar hebben leren kennen tijdens de Junior Med School ben jij een rode draad in mijn wetenschappelijke

carrière. Jij denkt in mogelijkheden, bent oprecht trots en hebt altijd tijd voor een vraag of een bemoedigend woord. Je bent een voorbeeld voor me.

Een translationeel proefschrift komt alleen tot stand in een lab waar het gesmeerd loopt. De combinatie van biologen, analisten, bio-informatici en clinici maakt ons een divers gezelschap met verschillende kwaliteiten. Ondanks de verschillen stond bij iedereen de deur altijd open voor inhoudelijke vragen of een gezellig praatje. Ik kijk met een warm gevoel terug op de afgelopen 4,5 jaar waarbij we de successen samen gevierd hebben, maar ook verdrietige momenten gedeeld hebben.

Lieve Anieta, dat het op het lab als een geoliede machine liep is voor een groot deel jouw verdienste. Wat had ik je graag mijn proefschrift overhandigd, wetende dat jij hier ook ontzettend trots op zou zijn geweest. Je wordt gemist.

Lieve Joan, de rots in de branding van het lab, onze lab-mama. Ik denk dat ik met zekerheid kan zeggen dat jij echt op elke vraag een antwoord weet. Of het nu over bestellingen, het versturen van een pakketje, een experiment of een levensvraag gaat, jij weet er wel een antwoord op. Dankjewel voor de koffiemomentjes in de vroege uurtjes voordat de rest kwam binnendruppelen. Geniet van je welverdiende pensioen!

Lieve Mai, onze CTC-Queen, geen melding op het CTC-apparaat is jou vreemd. Je weet altijd wel weer een oplossing te vinden om dat ene belangrijke sample nog te redden. Je bent een topper.

Lieve Saskia, wat een geluk dat jij tijdens mijn promotietraject ons lab kwam versterken. Wij hebben de afgelopen jaren veel leuke en bijzondere momenten beleefd: samen naar San Antonio, koukleumen op 1 januari en vrolijk worden bij Mamma Mia. Jij bent mega efficiënt en een kei in het polijsten van teksten waardoor de stukken waar jij aan meegewerkt hebt nog beter werden! Ik ben er trots dat ons FastSeqS project zo goed gelukt is! Ik kijk ernaar uit dat jij plaatsneemt in mijn promotiecommissie.

Marcel, de keren dat ik aan je bureau heb gestaan met een “lijstje TO DO’tjes”, waardoor ik zelfs de bijnaam “Lindsay Lijstje” heb vergaard, zijn niet op twee handen te tellen. Bedankt voor je geduld als ik nog “even” een klein vraagje had en voor je bioinformatische expertise bij de CPCT-projecten. Je bent een fijn persoon!

Jaco, alhoewel het liquor CTC project uiteindelijk niet in dit proefschrift terecht is gekomen, heb ik door jouw oneindige CTC kennis wel heel veel over CTCs geleerd. Je bent een vaste waarde binnen de groep. Bedankt dat je me regelmatig van m'n bureau losweekte om mee te gaan lunchen en dat kopje koffie na de lunch te drinken.

Maurice, het was fijn om met je samen te werken op de cfDNA projecten. Met name het IMPACT CRC ctDNA stuk was een ontzettend leuk project om samen te doen, met een mooi eindresultaat. Wat me vooral bij blijft is jouw "structuur" met tientallen geeltjes op je bureau waarop je alle TO DO'tjes bij hield. Stiekem heb ik dit van je overgenomen ;-)!

Jean, bedankt voor de gezelligheid op het lab en je hulp bij het ctDNA project van de IMPACT CRC.

Vanja, bedankt voor het opzetten van de FastSeqS methode en je hulp bij het opwerken van de liquor samples.

Alle andere analisten en post-docs van het lab, bedankt voor de leerzame en gezellige tijd!

Lieve mede (arts)-onderzoekers, alhoewel Jaco toch wat beduusd stond te kijken toen hij na het weekend mijn lege bureau zag, ben ik blij dat jullie me zijn komen halen. Vanaf het moment dat ik in Be-414a introk voelde ik me op mijn plek.

Nick en Wendy, wat is het fijn om zulke goede voorgangers te hebben. Nick, bedankt voor het inwerken en het leuke ESR1 stuk dat we in het eerste jaar van mijn promotie geschreven hebben. Hopelijk gaan wij elkaar weer in het Erasmus MC tegenkomen. Wendy, bedankt voor de gezelligheid en bemoedigende woorden aan het begin van mijn promotietraject. Jouw verdiensten voor de CPCT studie mogen niet onbenoemd blijven, wat een geweldige klus! Ik vind het leuk om te zien hoe jij in een andere rol betrokken bent gebleven bij het liquid biopsy onderzoek.

Wat begon met een reisje naar Stockholm om Inge te bezoeken, resulteerde in een jaarlijks terugkerend weekendje weg waardoor ons clubje inmiddels een rijke geschiedenis aan bijnamen heeft (Texelse Beach Babes, Düsseldorfse Dirndels en Volendendamse Viswijven). Ik kijk uit naar onze nieuwe bijnaam, want dat betekent een nieuw weekendje weg! Het was fijn om lief en leed met jullie te kunnen delen. Lieve Marjolein, wij als vroege vogels hebben heel wat koffietjes gedronken voordat de rest binnenkwam druppelen. Het was fijn om een "niet-

cliniclowntje" op de kamer te hebben bij wie ik terecht kon voor alle technische vraagjes. Lieve Inge, jij bent de creatieveling! Jij draait je hand niet om voor het ontwerpen van de cover van je proefschrift of een prachtige word-art kop boven de werkbesprekingbriefjes (zeker geen afleidingsmanoeuvre om te verdoezelen dat we even niet zoveel te melden hadden ;-)). Ik vind het dapper dat je de keuze gemaakt hebt om buiten het ziekenhuis je geluk te zoeken. Het RIVM heeft een goede aan jou! Lieve Lisanne, het was fijn om jou als buurvrouw te hebben en te kunnen sparren over de cfDNA projecten en de analyses van CPCT-data. Ik ben zo trots op je dat jouw doorzettingsvermogen heeft geresulteerd in die felbegeerde opleidingsplek tot uroloog. Ik weet zeker dat je een hele goede zult worden!

Lieve Manouk, dat jij het creatieve brein bent in onze groep illustreert de prachtige omslag van mijn proefschrift. Jij hebt het digitaal bloemschikken uitgevonden en tot in detail geperfectioneerd. Wat een bijzonder proces om samen te doorlopen. Je bent een open boek en je gezichtsuitdrukkingen zeggen meer dan je in woorden kunt uitdrukken! Ik vind het leuk dat jij het vaak net even anders aanpakt en ik hoop dat we elkaar blijven zien of dat nu binnen of buiten het ziekenhuis is. PS: ik heb de afgelopen maanden mijn rijkunsten geoptimaliseerd, dus mocht je mee willen rijden naar het ASZ… ;-)

Lieve Pauline, jij hebt de Brabantse gezelligheid meegenomen naar Be-414a en me opgevoed in de carnavalskrakers. Al in december begon het bij jou te kriebelen en kregen we de "vrijdagmiddag carnavalshit" te horen. Zo is skireisorganisatie bij ons twee jaar op rij in goede handen geweest en draaiden wij onze hand niet om voor het naaien van schoudervullingen in onze Trump jasjes. Naast de mooie en hilarische momenten op deze reizen, waren er soms ook iets mindere momenten (lees: dokter L krijgt stress van dokter P die een B probleem heeft ;-)). Lieve P, heel veel succes met afronden van jouw proefschrift!

Anouk, jij werd na het vertrek van Lisanne mijn nieuwe buurvrouw. Bedankt voor de gezelligheid, ook tijdens ons reisje naar de ESMO in Barcelona.

Teoman, wat was het fijn dat er weer een man tussen het gekakel van de vrouwen kwam! Jij bent heel wijs niet bij ons ingetrokken in Be414a, maar desondanks hebben we mooie momenten beleefd. Bedankt voor de leuke samenwerking van het FastSeqS stuk! Ik wens je heel veel succes met het afronden van jouw proefschrift.

Lisa, wij hebben maar kort fysiek samengewerkt, maar ik ben onder de indruk van de snelheid waarmee je alle nieuwe informatie eigen maakt. Ik weet zeker dat het liquorproject bij jou in goede handen is. Ook wij gaan elkaar zeker nog vaak zien!

Noortje, Khrystany en Noor, heel veel succes met jullie promotietrajecten!

Bianca, gekscherend noemden we jou tijdens onze onderzoekstage van de Junior Med School Mega Mindy, maar daar zit zeker een kern van waarheid in! Je bent een inspirator voor me geweest en hebt me tijdens die weken enthousiast weten te maken voor onderzoek en ik vind het echt ontzettend leuk dat ik jaren later weer terug mocht komen en toen veel beter kon begrijpen wat jij allemaal hebt betekend voor de start van de CTC groep. Ik kijk er naar uit om over een aantal jaar als fellow oncologie ook weer van je te mogen leren!

Martijn Starmans wil ik bedanken voor de prettig samenwerking op het Radiomics paper. Veel succes met het afronden van jouw proefschrift.

Promovendi van de "overkant" wil ik bedanken voor de gezelligheid: Florence, Bodine, Femke, Koen, Sander, Ruben, Louwrens, Daan, Karlijn, Mirjam, Dora, Yarne en Maud.

Femke, wij hebben onbewust vrijwel hetzelfde traject doorlopen. Alhoewel we ons artikel van het cfDNA deel van de REGORA studie helaas niet gepubliceerd kregen, was het super leuk om dit project samen met jou te doen! Ik kijk er naar uit om over een aantal jaar samen met jou als fellow in het EMC samen te werken.

Titia, bedankt voor de gezelligheid en mental support in San Antonio.

Ook wil ik studenten Ana Rajicic, Romy Klein Kranenbarg, Frederique Meinsma en Laura Pasquet hartelijk danken voor hun inzet voor de verschillende studies.

De oncologen van het Erasmus MC wil ik bedanken voor hun hulp bij het opzetten van nieuwe studies en het includeren van patiënten in de verschillende klinische studies.

In het bijzonder, dr. van der Veldt, beste Astrid, wat was ik blij dat jij naar het Erasmus MC kwam! Samen hebben we de eerste studies tussen de Interne Oncologie en Nucleaire Geneeskunde opgezet. Wat een klus, maar het is gelukt en heeft een basis gevormd voor veel meer onderzoek tussen de beide afdelingen. Daarnaast hebben we nog een uitdagend radiomics project tot een goed einde gebracht. Ik vind het inspirerend dat jij niet altijd over de gebaande paden gaat.

Dr. Lolkema, beste Martijn, bedankt voor alle input tijdens de cfDNA meetings en het meeschrijven aan de verschillende CPCT-02 stukken.

De neurologen, Prof. dr. van den Bent, Prof. dr. Sillevis Smitt en dr. Jongen, wil ik hartelijk danken voor hun enthousiasme en prettige samenwerking op de verschillende liquor projecten. Irene van Heuvel, dank voor het bij elkaar zoeken

van alle ingevroren liquor samples, wat een klus!

Het CPCT-02 consortium wil ik bedanken voor de inclusie van patiënten, het kritisch lezen en reviseren van het whole genome sequencing artikel over gemetastaseerd borstkanker. Dit artikel is het voorbeeld dat wij als Nederlandse onderzoekers het verschil kunnen maken door de handen ineen te slaan. Wat een prachtig initiatief. In het bijzonder wil ik hier graag nog noemen: Job van Riet en Harmen van de Werken voor het schrijven van de scripts om de data te kunnen analyseren. Prof. dr. Cuppen, beste Edwin, bedankt voor de kritische input op de uitgevoerde analyses.

Het IMPACT consortium, alle oncologen, mede arts-onderzoekers, nucleair geneeskundigen, researchverpleegkundigen en data-managers uit het UMCG, Radboud UMC, VUmc en Erasmus MC, wil ik bedanken voor de samenwerking en leerzame meetings. Ik kijk uit naar de uitkomsten van de verschillende studies!

Erik van Helden, waar een treinreisje vanuit het UMCG naar Rotterdam goed voor kan zijn. Je enthousiasme werkt aanstekelijk! Bedankt voor de leuke samenwerking op het IMPACT CRC cfDNA project.

Sophie Gerritse, bedankt voor de gezellige treinreisjes van en naar de verschillende IMPACT meetings. Hopelijk tot snel in het Erasmus MC!

Bertha Eisses, bedankt voor de leuke samenwerking op het IMPACT MBC CTC/FES-PET paper!

De collega's en internisten uit het IJsselland ziekenhuis wil ik bedanken voor het warme welkom bij de terugkeer naar de kliniek in deze bijzondere COVID-tijd.

De collega's en internisten van het Albert Schweitzer ziekenhuis wil ik bedanken voor de prettige start van mijn opleiding tot internist. Ik kijk er naar uit om me bij jullie de komende jaren verder te ontwikkelen tot een goede internist.

En dan zijn we aangekomen bij mijn lieve vriendinnetjes "de Tites", lieve Jolijn, Kirsten, Lizzy, Floor, Tessa en Emma, wat is het fijn om onderdeel te mogen zijn van deze vriendinnengroep. Wij kennen elkaar al sinds de middelbare school en jullie vormen een belangrijke basis in mijn leven. Het is fijn om te voelen en te weten dat wij er altijd voor elkaar zijn. Bedankt voor alle gezellige weekendjes weg, etentjes, wandelingetjes en goede gesprekken door de jaren heen. Jullie zijn onmisbaar!

Lieve Melissa, van samen studeren en samenwonen als huisgenootjes, uiteindelijk promoveren bij dezelfde promotor. Wat fijn om af en toe even te kunnen sparren

tijdens een kopje koffie en ik ben blij dat we elkaar nog steeds weten te vinden als vriendinnen. Hopelijk krijg jij snel het goede nieuws dat je ook mag starten als AIOS.

Thijmen, wij kennen elkaar al sinds het 1e jaar geneeskunde en sindsdien trekken we met elkaar op. Alhoewel wij standaard elkaars verjaardagen vergeten en elkaar daar dan twee maanden later op wijzen, weten we elkaar een aantal keer per jaar te vinden om samen met Bibi lekker te eten en een (half) glaasje wijn te drinken. Nu jij in het Maastrichtse aangenomen bent tot AIOS oogheelkunde, hebben wij een goed excuus om het bourgondische leven daar te komen inspecteren en je zachte G weer af te leren;-).

Lieve Bibi, sinds 4 Gymnasium bewandelen wij hetzelfde pad. Samen CTC's "vissen" tijdens ons Junior Med School project, vervolgens als huisgenootjes ploeterend voor de geneeskunde tentamens en op de fiets naar het tentamen nog even de "moeilijke onderwerpen" doornemen, zodat we superscherp van start konden. Daarna tegelijk van start met een promotietraject en nu allebei verzekerd van een mooie opleidingsplaats, jij als neurochirurg in spe, ik als internist in spe. Voor mij maakt het de cirkel rond dat jij vandaag achter mij staat!

Lieve Pim, gedurende dit promotietraject stond jij voor het grootste deel aan mijn zijde. Hoewel het anders is gelopen dan we allebei hadden gehoopt, ben ik je dankbaar voor de fijne tijd.

Lieve familie, wat is fijn dat jullie er altijd zijn.

Allerliefste pap en mam, dat ik hier vandaag sta heb ik volledig aan jullie te danken. Jullie steunen mij onvoorwaardelijk en ik weet dat ik altijd bij jullie terecht kan. Ik ben trots op de vorm die jullie gevonden hebben om met elkaar om te gaan en ik ben blij dat jullie vandaag vol trots naast elkaar op de eerste rij kunnen zitten!

Lieve Vera, je bent attent en altijd oprecht geïnteresseerd, bedankt daarvoor!

Lieve Oma, een van de redenen om mijn promotie een jaartje uit te stellen was om jou op de eerste rij bij mijn verdediging te zien. Ik vind het bijzonder om deze gebeurtenis in jouw bijzijn te beleven.

Lennard en Denise, jullie zijn lieverds! Denise, wat fijn dat jij een aantal jaar geleden bij ons in de familie bent gekomen, je bent een lief mens! Lieve Lennie, jij riep als eerste dat als ik zou promoveren jij dan echt mijn paranimf wilde zijn en zo zal geschieden! Hoe verschillend wij ook zijn - jij houdt van koken, ik ben blij als ik iets fatsoenlijks op tafel weet te toveren; ik houd van het ziekenhuis, jij krijgt

al zweetdruppeltjes rond je neus als we het hebben over vaccineren - we weten elkaar altijd te vinden als het nodig is. Ik ben trots op je en vind het bijzonder dat jij vandaag achter mij staat.