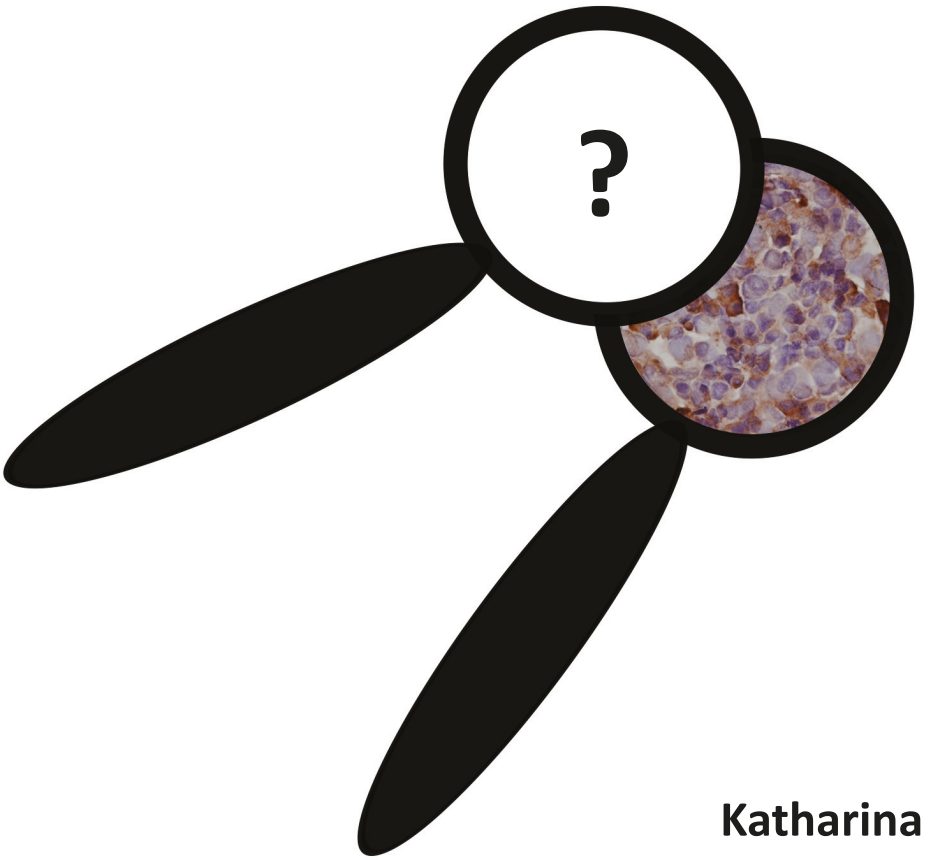
Non-coding RNAs as biomarkers in breast cancer



Katharina Uhr

Non-coding RNAs as Biomarkers in Breast Cancer

Katharina Uhr

The cover image includes the depiction of MCF-7 breast cancer cells stained for the protein MTOR.

Molecular Medicine
Postgraduate School

The research described in this thesis was performed within the framework of the Erasmus MC Molecular Medicine (MolMed) Graduate School at the department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands.

The funding of this research was provided by the Daniel den Hoed Stichting.

The research stay at the Rockefeller University, New York, USA was funded by a travel grant from the René Vogels Stichting.

Financial support for printing this thesis was generously provided by the Department of Medical Oncology of the Erasmus MC Cancer Institute and the Erasmus University

Cover: Katharina Uhr
Layout: Thomas van der Vlis, Persoonlijkproefschrift.nl
Printed by: Ridderprint BV | www.ridderprint.nl
ISBN: 978-94-6375-652-5

Non-coding RNAs as Biomarkers in Breast Cancer

Niet-coderende RNA's als biomerkers bij borstkanker

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday 8 January 2020 at 11:30 hrs
by

Katharina Uhr
born in Tönisvorst, Germany

Erasmus University Rotterdam

Promotors:	Prof.dr.ir. J.W.M. Martens
	Prof.dr. J.A. Foekens
Other members:	Prof.dr.ir. G. Jenster
	Prof.dr. C.G.J. Sweep
	Prof.dr. B. van de Water

Contents

Chapter 1

Introduction

Breast cancer – clinical perspective

In women breast cancer is the most common type of cancer, with about 2.1 million new cases expected to be diagnosed in 2018 worldwide and this disease accounts for most cancer deaths in women (Bray et al., 2018). Within the EU, the number of breast cancer deaths is declining from 17.9/100.000 in 2002 to 15.2/100.000 in 2012 and is expected to reach 13.4/100.000 by 2020 (Carioli et al., 2017). This decline is attributed to advances in therapy, management but also to early detection due to screening; additionally a decline in incidence in recent generations was observed aiding in the predicted decline in breast cancer deaths (Carioli et al., 2017). Furthermore, the discontinuation of hormonal use in postmenopausal women has aided as well (Bray et al., 2018). In other regions such as South America, Africa and Asia breast cancer incidence is, however, still rising (Bray et al., 2018). In the Netherlands incidence rates have increased since 1990 and are relatively stable since 2010 when accounting for age of the population, but do not show a decrease in incidence in recent years (Nederlandse Kankerregistratie (NKR), IKNL).

Breast cancer is a heterogeneous disease and as main distinction tumors can be divided into those which express the estrogen receptor (ER) and/or the progesterone receptor (PR) and those tumors which do not (Hammond et al., 2010). The difference of expression of mainly ER influences the tumors in almost every aspect, such as prognosis (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005; Hammond et al., 2010) and available therapeutic options for patients (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005; Hammond et al., 2010) but also at the molecular level affecting the expression of a multitude of genes (Carroll, 2016; Khan et al., 2012). ER and PR expression are highly correlated, only few tumors are ER-negative and PR-positive and this number has further declined due to better assaying techniques ruling out some of these tumors as false-negative for ER and warranting re-testing of tissue for ER-negative PR-positive samples (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011).

Besides this main distinction among tumors, one can further classify those breast tumors which express ERBB2, another important targetable receptor with a large impact on prognosis and tailored therapy options (Fitzgibbons et al., 2000; Hammond et al., 2010; Press et al., 2005; Slamon et al., 1987). At the molecular level, breast tumors can be grouped into several categories to optimize predictions on the behavior of the tumors and the course of disease (Sørlie et al., 2001, 2003). These classifications are called molecular subtypes and entail luminal A, luminal B, normal-like, basal-like and

ERBB2-overexpressing tumors (Perou et al., 2000; Polyak, 2007; Sørlie et al., 2001, 2003). The luminal subtypes consist of basically only ER-positive tumors and are called luminal because they are characterized by the expression of many genes that are also expressed by luminal epithelial cells of the healthy mammary gland (Reis-Filho and Pusztai, 2011). Luminal A and luminal B subtypes are differentiated through the expression of proliferation-related genes such as Ki67, and overall more ERBB2-positive breast cancers fall into the luminal B than the luminal A subtype (Reis-Filho and Pusztai, 2011). The ERBB2-overexpressing subtype is characterized by its high expression of ERBB2 predominantly as a result of *ERBB2* gene amplification (Perou et al., 2000) and tumors in this subgroup are often ER-negative (Reis-Filho and Pusztai, 2011). The basal-like subtype is often triple-negative (lack of expression of ER, PR and ERBB2) and expresses high levels of proliferation-related genes (Reis-Filho and Pusztai, 2011). This subtype is named based on the finding that it expresses many genes which are also expressed by basal and myoepithelial breast cells, such as the cytokeratins 5/6, 17 and the epidermal growth factor receptor (*EGFR*) (Reis-Filho and Pusztai, 2011). The normal-like breast cancer resembles in its expression profile the normal breast and fibroadenomas, however a large part of tumors falling into this classification have been found to be miscategorized due to high contents of non-tumor tissue in samples (Reis-Filho and Pusztai, 2011). Among these molecular subtypes luminal A tumors have the best prognosis, while normal-like tumors show an intermediate prognosis, luminal B tumors have an intermediate to poor prognosis and ERBB2-overexpressing as well as basal tumors have a poor prognosis (Reis-Filho and Pusztai, 2011).

Current treatments for breast cancer

When it comes to drug treatments the St. Gallen Consensus recommends chemotherapy for triple-negative tumors, anti-ERBB2 therapy plus chemotherapy for ERBB2-overexpressing tumors regardless of ER status and endocrine therapy for ER-positive breast cancer (Curigliano et al., 2018). For ER-positive low risk tumors tamoxifen is recommended as endocrine therapy in pre-menopausal women, for post-menopausal women aromatase inhibitors (AI) are the preferred option for endocrine therapy (Curigliano et al., 2018). For higher risk ER-positive tumors in pre-menopausal women ovarian function suppression (OFS) plus tamoxifen or OFS plus exemestane or OFS plus exemestane plus chemotherapy are recommended (Curigliano et al., 2018). In post-menopausal women with ER-positive disease and an intermediate or high risk, AI combined with chemotherapy are recommended, while for pre-menopausal women with high risk tumors with low ER expression chemotherapy and AI are recommended,

supplemented with OFS (Curigliano et al., 2018). In the metastatic or advanced setting of ER-positive, ERBB2-negative breast cancer CDK4/6 inhibitors such as palbociclib have recently shown great benefit in combination with AI as well (Sobhani et al., 2019). For (neo-)adjuvant chemotherapy anthracycline- or taxane-based therapies are preferred for triple-negative patients (Curigliano et al., 2018). In tumors with deficiencies in *BRCA1* or *BRCA2* alkylating therapy is recommended next to anthracycline- or taxane-based chemotherapy, while some experts consider platinum-based chemotherapy useful as well (Curigliano et al., 2018). Recently, PARP inhibitors have also been approved for treatment of BRCA-deficient breast cancers (Ashworth and Lord, 2018). For ERBB2-overexpressing tumors different treatments are recommended based on tumor grade; low grade should be treated with trastuzumab and chemotherapy e.g. paclitaxel but not anthracyclines, while higher grade should be treated with anthracyclines followed by taxanes with concurrent trastuzumab (Curigliano et al., 2018). Depending on the risk for relapse pertuzumab can be added next to trastuzumab in all settings independent of grade (Curigliano et al., 2018). For those ERBB2-overexpressing tumors that also express ER, chemotherapy plus endocrine therapy plus trastuzumab are recommended up to adding neratinib in some cases (Curigliano et al., 2018). While the aforementioned drugs are the most commonly prescribed treatments (Curigliano et al., 2018), other drugs are also used in certain circumstances and the National Cancer Institute in the US currently lists 64 drugs as approved for the treatment of breast cancer (www.cancer.gov, 2019).

Current biomarkers for breast cancer

Cancer biomarkers are measurable patient or tumor characteristics such as expression of a certain protein, a certain gene or levels of signaling molecules (Sawyers, 2008). Biomarkers can ideally be easily screened in a patient and enable predictions for prognosis to avoid over-treating patients which have a very good prognosis and might not need treatments with severe side effects (Duffy et al., 2016; Sawyers, 2008). Besides prognosis, biomarkers have also an important role in prediction, i.e. to predict which patients will respond best to a certain therapy or will likely be intrinsically resistant (i.e. meaning being resistant from the start of treatment in contrast to patients who develop resistance over time) (Duffy et al., 2016; Sawyers, 2008).

Gene signatures such as Oncotype DX® and MammaPrint® can aid in identifying patients who likely benefit from systemic therapy such as extended endocrine therapy or chemotherapy, however, the gain using these signatures does not replace clinical parameters such as tumor size and lymph node status but provides complementary

information and should be used in combination with the respective clinical parameters (Hayes, 2015; Weigelt et al., 2012). Furthermore, similar performance with regard to prognosis between Oncotype DX® and semiquantitative measurements of ER, PR, ERBB2 and Ki-67 by immunohistochemistry has been shown (Weigelt et al., 2012). However, the test of these four proteins, also called IHC4 carries the risk for low analytical validity as it has not been extensively validated and therefore, different outcomes depending on the lab determining the test might be observed (Hayes, 2015).

So far predictive power has overall been limited using these different molecular signatures, however, they do aid in fine-charting prognosis (Weigelt et al., 2012). For the gene signatures Oncotype DX®, EndoPredict®, PAM50-ROR (Prosigna®), Breast Cancer Index®, as well as the biomarker urokinase plasminogen activator combined with plasminogen activator inhibitor type 1 (uPA and PAI-1), sufficient evidence could be obtained that these biomarkers may be used in the clinic to aid in determining which patients should receive systemic adjuvant chemotherapy in early ER/PR-positive ERBB2-negative node-negative breast cancer (Harris et al., 2016). However, these biomarkers are not suitable for node-positive, ERBB2-positive or triple-negative breast cancers (Harris et al., 2016). In the scenario of advanced breast cancer the biomarker situation is not optimal either. In metastatic breast cancer it is recommended to analyze metastases also for ER, PR and ERBB2 expression if accessible (Van Poznak et al., 2015). However, when results are discordant with the primary tumor, considerations have to be made whether a switch in therapy based on the result of the metastasis seems useful as clinical evidence is currently lacking to support whether treatment based on primary measurements or treatment based on metastasis measurements generates a better outcome (Van Poznak et al., 2015). The current suggestion is to guide treatment based on the metastasis outcome though (Van Poznak et al., 2015). ER, PR and ERBB2 are currently the only biomarkers available (besides clinical parameters) for clear-cut recommendations for metastatic breast cancer in regard to e.g. systemic therapy (Van Poznak et al., 2015). The number of circulating tumor cells (CTCs) in metastatic breast cancer is prognostic, but CTC abundance does not predict benefit in switching to an alternate therapy (Van Poznak et al., 2015). Carcinoembryonic antigen (CEA), cancer antigen (CA) 15-3, and CA 27-29 may be used to provide complimentary information in metastatic breast cancer, as these markers provide indications, when properly interpreted, for disease progression under treatment, however, this is a suggestion based on clinical experience and not clinical studies (Van Poznak et al., 2015). For predictive biomarkers one challenge is that different mechanisms can achieve resistance against one drug, thereby complicating predictive studies in patients (Weigelt et al., 2012). Other issues are limited availability of archived tumor specimen for predictive

studies as well as drug resistance due to patient-specific factors such as blood supply of the tumor, tumor necrosis or drug metabolism (Weigelt et al., 2012). While different predictive biomarkers are available for tamoxifen, AIs, taxanes, anthracyclines and trastuzumab (in the latter case *PTEN* and soluble ERBB2 levels), these biomarkers do not provide the level of evidence required with regard to analytical validity, clinical validity or clinical utility (Harris et al., 2016). In breast cancer ER, PR and *ERBB2* are the only biomarkers for systemic therapy which meet the requirements in regard to analytical validity, clinical validity and clinical utility and should be used to guide treatment decisions (Van Poznak et al., 2015).

For breast cancer there are several classical prognosis/prediction markers such as tumor grade, tumor size, lymph node status, ER, PR and ERBB2 expression, as well as age, menopausal status and comorbidities which are taken into consideration in the decision-making process for systemic adjuvant therapy in early invasive disease (Henry et al., 2016). Chemotherapy is recommended for the following disease situations: positive lymph nodes (≥ 1 lymph node containing a tumor metastasis > 2mm), absence of ER expression & tumor size > 5 mm, ERBB2 expression, high risk lymph node negative tumors with a tumor size > 5 mm & an additional high risk feature, poor risk profile in the Adjuvant Online! risk stratification tool of > 10% risk of breast-cancer associated death within 10 years (Henry et al., 2016). High risk features in lymph-node negative tumors above 5 mm are: grade 3, triple negative tumors, lymphovascular invasion, ERBB2 expression and an Oncotype DX® risk score of ≥ 20% for distant recurrence within 10 years (Henry et al., 2016). The biomarker assay MammaPrint® may also be used for high clinical risk patients (as defined in the MINDACT trial) with hormone-receptor positive, ERBB2-negative tumors, who are node-negative or have up to three positive lymph nodes for aiding in the decision-making process if systemic adjuvant chemotherapy may be withheld, as MammaPrint® can identify good prognosis subgroups within these patient populations (Krop et al., 2017). However, in the case of one or more positive lymph nodes, systemic adjuvant chemotherapy should be discussed with a patient even if the MammaPrint® assay shows a good prognosis, as benefits through chemotherapy treatment cannot be excluded (Krop et al., 2017).

Patients with node negative small tumors below 5 mm and no further risk factors as well as strongly ER-positive, PR-positive, ERBB2-negative tumors less than 5 mm in size with micrometastatic nodal involvement (< 2mm) and Oncotype DX® assessed risk < 10% within 10 years, however, gain little benefit from chemotherapy and might be spared from this treatment (Henry et al., 2016). Furthermore, well-differentiated

tumors should be considered to be spared from chemotherapy especially if they show a luminal A gene expression signature (Henry et al., 2016).

Breast cancer – biological perspective

Breast cancer is a disease of the genome (Stephens et al., 2012; Stratton et al., 2009), caused by alterations in the DNA (Nik-Zainal et al., 2016; Stephens et al., 2009, 2012). These alterations include genomic instability as well as mutations and are together important hallmarks of cancer itself, which enable the cells to subsequently obtain further hallmarks/characteristics of cancer such as sustained proliferative signaling, evaded growth suppression, enhanced invasion and metastasis, escape from replicative senescence, enhanced angiogenesis, repressed cell death, evaded immune destruction, deregulated cellular energetics and tumor-promoting inflammation (Dai et al., 2016; Hanahan and Weinberg, 2011). These hallmarks or characteristics are general properties that cancer cells must gain or, if it concerns the environment, interplay with, to grow into a solid tumor (Hanahan and Weinberg, 2011). Depending on the type of cancer/tissue of origin each cancer type has specific traits such as reliance on certain growth signaling pathways or expression of certain cell surface markers which aid in determining the tissue of origin of a tumor as well as guide treatment options (Ke and Shen, 2017; Mohammed et al., 2017; Uhlén et al., 2005; Xu et al., 2016).

Mutations in breast cancer

The mutations that are found in cancer can be grouped in driver mutations and passenger mutations (Stratton et al., 2009). Driver mutations are mutations that provide a selective advantage to a cancer cell such as an enhanced growth rate over normal cells of the surrounding tissue; while passenger mutations do not carry an advantage for the cancer cell but have arisen e.g. in the normal progenitor of the cancer cell or the cancer cell itself by endogenous or exogenous DNA damage and/or by defective DNA repair (Stratton et al., 2009). Driver mutations are therefore enriched in tumors and more than 576 genes with driver mutations, which are linked to oncogenesis, have been identified in cancer to date (http://cancer.sanger.ac.uk/census, 2019). For breast cancer specifically 93 driver genes have been described with the five most commonly mutated genes being *TP53*, *PIK3CA*, *MYC*, *CCND1* and *PTEN* (Nik-Zainal et al., 2016). The number of mutated driver genes within a tumor varies between two to eight for most cancers (Vogelstein et al., 2013). Driver genes can be roughly grouped in three

functional categories: genes involved in cell fate, those involved in cell survival and finally genes playing a role in genome maintenance (Vogelstein et al., 2013). For breast cancer, Vogelstein *et al.* found a median of 3 driver genes within a dataset of 111 tumors taking mutations in oncogenes and tumor suppressors as well as driver gene amplifications and deletions into account (supplemental data) (Vogelstein et al., 2013).

Looking not only at the driver genes but all mutations present in a tumor, specific signatures can be detected, based on the prevalence of specific mutation types (Nik-Zainal and Morganella, 2017). These mutational signatures are characteristic for the type of mutagenic burden the DNA was exposed to, such as normal aging, deficient DNA damage repair (e.g. deficiency in homologous recombination or mismatch repair), carcinogens and APOBEC enzymatic activity (Nik-Zainal and Morganella, 2017). One tumor can have several mutational signatures in respective subclones representing the history of exposure to these mutagenic threats (Nik-Zainal and Morganella, 2017). So far 12 base substitution signatures have been discovered of which five are common, i.e. present in more than 20% of breast tumors (Nik-Zainal and Morganella, 2017). All of these signatures have also been detected in other tumor types representing wide-spread mechanisms of DNA damage (Nik-Zainal and Morganella, 2017). The mutational signatures in breast cancer have shown that within *ERBB2*-positive tumors the APOBEC-typical mutation pattern is more frequently observed (Ng et al., 2015). Furthermore, *TP53* mutations are associated with the mutational profile caused by APOBEC activity, potentially representing a consequence of the loss of function of *TP53* (Ng et al., 2015).

Furthermore, genomic rearrangements present in a cancer can also be classified into signatures and for breast cancer six have been described (Nik-Zainal and Morganella, 2017). Three of these signatures are present in tumors with deficiencies in the homologous recombination repair (Nik-Zainal and Morganella, 2017).

Prognostic and predictive genetic alterations in breast cancer

Mutations and genomic events (such as the gain or loss of a gene) can influence prediction and prognosis. E.g. in breast cancer the gene *ERBB2* is frequently amplified and serves as a biomarker for poor prognosis and response to different therapies but is itself a therapeutic target as well (Fitzgibbons et al., 2000; Incorvati et al., 2013; Mariani et al., 2009; Press et al., 2005; Slamon et al., 2011, 1987). High ERBB2 levels have been found to be associated with lower response to methotrexate-based therapies and tamoxifen-based therapies, while doxorubicin-based therapies were more successful in

this patient population (Carlomagno et al., 1996; Clark, 1998; Fitzgibbons et al., 2000; Gusterson et al., 1992; Leitzel et al., 1995; Muss et al., 1994; Paik et al., 1998; Pritchard et al., 2006; Thor et al., 1998; Wright et al., 1992). Besides, the protein product of *ERBB2* also presents a target for drug therapy reducing cancer-associated deaths (Incorvati et al., 2013; Slamon et al., 2011, 2001). Another example in breast but also ovarian cancer are germline mutations in the genes *BRCA1* and *BRCA2*, which do not only increase the risk of developing breast or ovarian cancer (Ashworth and Lord, 2018; Dziadkowiec et al., 2016; Fong et al., 2009) but have also predictive power to PARP inhibitor therapy, in the sense that only tumor cells deficient of functional BRCA genes are sensitive to PARP inhibitor therapy due to synthetic lethality (Ashworth and Lord, 2018; Dziadkowiec et al., 2016; Fong et al., 2009).

Breast tumors often consist of several subclones with different sets of mutations (Ng et al., 2015). Upon therapy these clones can gain a selective advantage e.g. due to intrinsic drug resistance (Ng et al., 2015). The acquisition of additional mutations conferring drug resistance can, however, also happen by chance during drug treatment or even be a by-product of mutagenic drug treatments (Ng et al., 2015). A rare event in breast cancer is, that 0.6% of luminal breast cancers show mutations within the ligand-binding domain of the *ESR1* gene (Ng et al., 2015). However, in metastases of patients receiving prior treatment with an aromatase inhibitor, the number of these mutations in *ESR1* is increased, arguing for enrichment of the *ESR1*-mutated subclones due to selective pressure induced by the drug therapy (Ng et al., 2015). Interestingly, activating mutations in the tyrosine kinase domain of the *ERBB2* gene have been identified in about 1.5% of breast cancers, likely affecting response to ERBB2-targeted therapies (Ng et al., 2015). Another mutation described in *ERBB2*, L755S, does not cause activation of the protein, however, it does cause resistance to lapatinib, a drug targeting the proteins ERBB2 and EGFR (Ng et al., 2015). The observed resistance is likely due to changed binding kinetics between the drug and the protein ERBB2 caused by this amino acid change (Ng et al., 2015).

Interestingly, not only single mutations but also the rate of mutations in a cancer can be associated with therapy outcome, e.g., AI-resistant tumors show a higher rate of mutations than AI-sensitive tumors, which might be due to tumor heterogeneity as tumors with high mutation rates are more heterogeneous (Ng et al., 2015).

To conclude, it is worthwhile to study single mutations as well as genomic aberrations on a more global scale as these might indicate sensitivity or resistance to specific drugs.

Non-coding RNAs

For a long time protein-coding genes have been the main focus of research into cancer (Futreal et al., 2004; Greenman et al., 2007; Stratton et al., 2009). In 2004 a detailed report of build 35 of the human genome project showed that only a small part of the human genome actually encodes proteins, and that the number of protein-coding genes was substantially lower than estimated (International Human Genome Sequencing Consortium, 2004). This sparked the question whether the remaining part had functional value (International Human Genome Sequencing Consortium, 2004; The ENCODE Project Consortium, 2004), a topic which had gained popularity due to the identification of new classes of functional "elements", such as new types of RNA molecules with a surprising range of functions, in the genomes of different organisms (Storz, 2002). Subsequently the ENCODE project began to study the human genome for its functional sequences (The ENCODE Project Consortium, 2004). To this day the ENCODE project has determined a large part of the human genome as functional and identified transcription factor binding sites and regulatory elements (e.g. enhancers), besides the protein-coding genes (The ENCODE Project Consortium, 2012). Additionally, another class of genes was established to be well-represented in the human genome, which is the class of non-coding RNAs (ncRNAs) (The ENCODE Project Consortium, 2012). ncRNA genes are defined as DNA regions which are transcribed into RNAs but not further translated into proteins and are neither a tRNA nor a rRNA (Storz, 2002). Among ncRNAs are, for example, microRNAs (miRNAs) (Hombach and Kretz, 2016) and circular RNAs (circRNAs) (Ashwal-Fluss et al., 2014), both of which will be discussed in the subsequent sections.

MiRNAs

MiRNAs are small oligonucleotides which consist of roughly 22 nucleotides (Fang et al., 2013) and were first discovered in the nematode *Caenorhabditis elegans* in 1993 (Lee et al., 1993). Subsequently this type of gene was discovered in many other animal species including humans but also in plants (Li et al., 2010; Tarver et al., 2012). MiRNAs bind mRNAs in a sequence-dependent fashion and in this way affect protein production post-transcriptionally (Calin and Croce, 2006) (see **Fig.1**). Via this mechanism of action they can affect up to several hundred transcripts (Hausser and Zavolan, 2014; Huntzinger and Izaurralde, 2011; Iorio and Croce, 2012; Krol et al., 2010; Lin and Gregory, 2015; Selbach et al., 2008; Varela et al., 2013). MiRNAs have been found to be dysregulated in many human diseases such as in viral infections, disorders of the nervous system,

cardiovascular diseases, muscular diseases, diabetes but also cancer (Wang et al., 2016). Using miRNA expression, it has been shown that the developmental lineage and differentiation stage of human cancers can be characterized (Lu et al., 2005) and miRNAs were shown to be able to predict prognosis in disease (Wang et al., 2016). A dysregulation of the miRNA expression profile has been found in multiple cancer types such as leukemia, liver cancer, ovarian cancer, pancreatic cancer, prostate cancer and also breast cancer (Wang et al., 2016). Specifically the oncogenic *hsa-miR-21* has been found to be upregulated in many cancer types (Wang et al., 2016). On the tumor suppressor side a few miRNAs have been found but while some apply to several cancer types, the roles of others depend on the cancer type and are less generally applicable (Wang et al., 2016).

With regard to ease of study, it has been found that miRNAs are less prone to get damaged than mRNAs in formalin-fixed paraffin embedded tissue (Lu et al., 2005) – which makes them a great candidate to be assessed in solid tumor samples. Others have found that miRNAs overall show a higher stability to stress e.g. heating of RNA samples than mRNA (Jung et al., 2010) and also show high stability in serum samples (Grasedieck et al., 2012). MiRNAs can be easily measured using different methods such as microarrays, qRT-PCR and next-generation sequencing (NGS) to name a few (Wang et al., 2016). MiRNAs hold therefore great promise for evaluation in different disease scenarios.

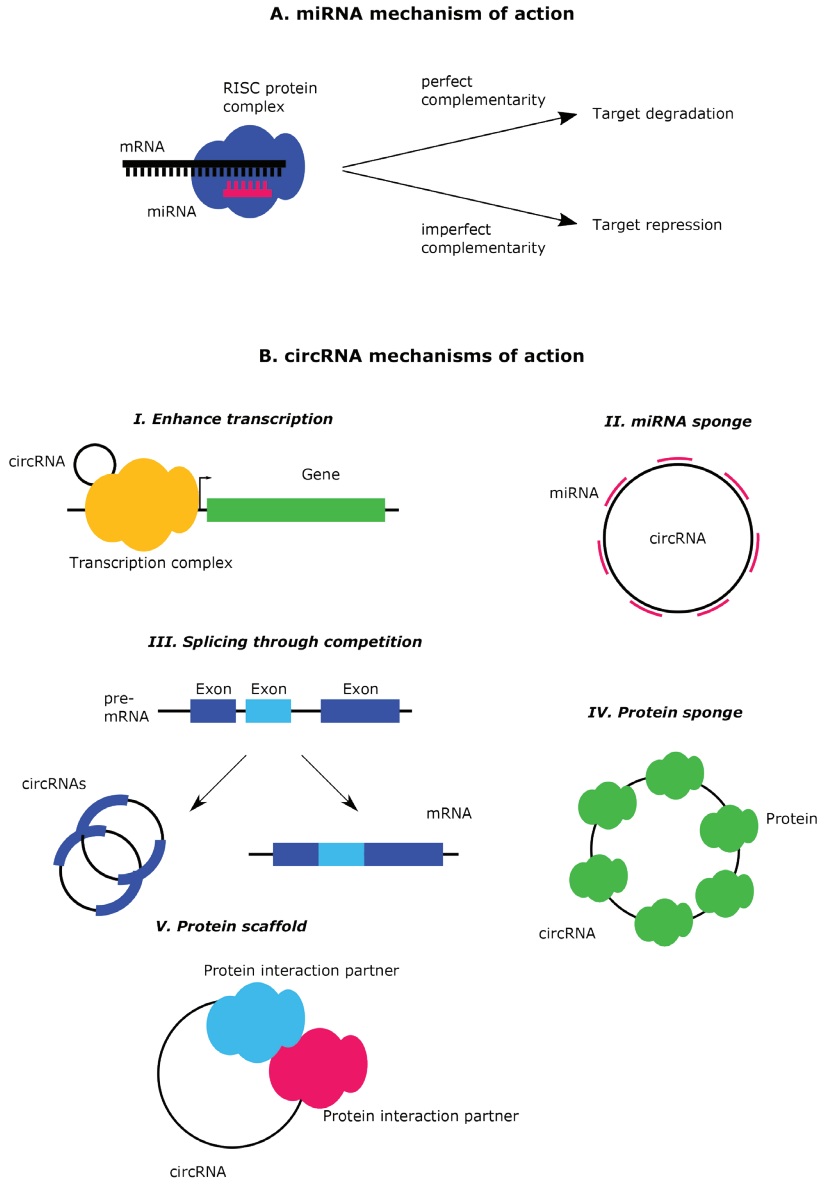


Figure 1. Mechanisms of action for miRNAs and circRNAs. A. miRNAs bind to their mRNA target genes based on sequence complementarity within the RNA-induced silencing complex (RISC). If sequence complementarity is perfect, the target mRNA is degraded, if sequence complementarity is imperfect the target mRNA is repressed from being translated. B. CircRNAs can have different roles. I. Through interaction with the transcription complex they can enhance the transcription of their host genes. II. CircRNAs can function as a miRNA sponge by providing multiple binding sites for a specific miRNA and in this way scavenge miRNAs. III. Furthermore, circRNAs can present competition in regard to splicing of their host gene and thus influence splicing. IV. Another function is providing protein binding sites and in this way sponging proteins. V. Finally, circRNAs have been observed to provide a scaffold for different proteins, bringing them into physical proximity.

CircRNAs

CircRNAs are a class of ncRNAs, which are characterized through their circular form (Wang et al., 2018). They were first described in 1979 (Wang et al., 2018) and have a wide variety of functions such as increasing transcription of their host genes by association with RNA polymerase II, influencing splicing of their linear host gene based on competition, providing a sponge for miRNAs, by reducing protein availability through sponging them and serving as scaffolds providing binding sites for interacting proteins (Kristensen et al., 2018) (see **Fig. 1**). Through these ways of action circRNAs have been shown to affect multiple cancer-relevant processes such as apoptosis, angiogenesis, migration or cell cycle progression and proliferation (Kristensen et al., 2018). So far dysregulation of circRNAs has been observed in a large number of human cancers such as malignancies of the hematological system, liver cancer, lung cancer, breast cancer, prostate cancer, bladder cancer, ovarian cancer, kidney cancer, gastric cancer and malignancies of the central nervous system (Kristensen et al., 2018). Besides this frequent dysregulation it has been noted that due to their circular structure circRNAs show a high transcript half-life probably due to exonuclease resistance (Jeck et al., 2013) and these two characteristics underline the potential of circRNAs as biomarkers.

Chapter 2

Scope of the thesis

The need for new therapies – the value of drug screenings

Tumors are fine-charted into subcategories to improve current predictions regarding prognosis and to help guide treatment decisions. Much success has been gained with drug treatments since the advance of the targeted agents tamoxifen, as well as AIs alone (or recently in combination with CDK4/6 inhibitors) against ER-positive breast cancers and anti-HER2 (ERBB2) therapy for ERBB2-overexpressing tumors, however, a significant proportion of patients with metastatic or advanced disease will eventually become resistant to therapy (Carroll, 2016; Curigliano et al., 2018; Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011; Incorvati et al., 2013; Slamon et al., 2011; Sobhani et al., 2019). Chemotherapy has been the longest available type of anti-cancer drug treatment, however, while it can potentially benefit all patients, this type of treatment is also associated with various types of side effects and patients will as well develop resistance (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005; Greene et al., 1994; Ke and Shen, 2017).

As targeted therapies hold great promise for mainly affecting the tumor but less so healthy tissues in the body, many investigations have been undertaken to study tumors for those proteins crucial for fueling the cancerous growth or enabling the cells to survive (Gerber, 2008; Sawyers, 2004; Widakowich et al., 2007). Many potential targets have been identified and targeted drugs have been developed (Gerber, 2008; Sawyers, 2004; Widakowich et al., 2007). Furthermore, some medications already prescribed for other conditions have shown indications for anti-cancer effects as well (Würth et al., 2016). It is therefore necessary to study these newly developed but also already available drugs for a potential use in the treatment of cancer. A first step to assess whether drugs might show benefit in a certain condition are drug screenings, i.e. monitoring cell line growth/behavior under drug treatment (Allen et al., 2005; Monks et al., 1991).

The need for new biomarkers to predict course of the disease and therapy response

Besides new available treatments, there is also a need for better biomarkers (Begley and Ellis, 2012; Hayes et al., 2013). For breast cancer the intrinsic molecular subtypes provided prognostic information but had only limited predictive value (Weigelt et al., 2012). Luminal A tumors, with a good prognosis have a neglectable benefit from

adjuvant chemotherapy, while the other intrinsic subtypes do show substantial benefit from adjuvant chemotherapy treatments (Weigelt et al., 2012).

Among the new biomarkers available in breast cancer, however, few are robust enough in their capabilities to be used in the clinics (see above for an overview of current clinical grade biomarkers) (Duffy et al., 2016; Győrffy et al., 2015; Harris et al., 2016).

To chart more potential biomarkers which might increase the capability for an optimal prediction forecast is therefore of great benefit (Duffy et al., 2016; Harris et al., 2016; Hayes et al., 2013). Furthermore, as new types of treatments evolve, it would be of great help to have also accompanying biomarkers in the near future available when such new therapies are migrating into the clinics.

A first step to improve patient management is to test drugs on a model system for their response profile (Begley and Ellis, 2012), preferably taking into account the biological diversity such as subtypes of the cancer. In this thesis, we have made use of our large collection of breast cancer cell lines to perform an extensive drug sensitivity screening using a wide array of newly developed targeted drugs, as well as chemotherapeutics for comparison of their response profiles. **Chapter 3** discusses the outcomes and conclusions from this study in further detail.

In the search for biomarkers, miRNAs hold great promise and are worth assessing in the search for suitable prognostic or predictive biomarkers. In **chapter 4** we studied the potential of miRNAs as predictive biomarkers for drug response in breast cancer cell lines.

As mutations can also influence drug response e.g. in the case of *BRCA1/2* (Ashworth and Lord, 2018), in **chapter 5** mutations and copy number aberrations (CNAs) were studied for their value as biomarkers in breast cancer cell lines for sensitivity to a wide array of drugs.

In the past *hsa-miR-7* was found to be a prognostic biomarker in breast cancer (Foekens et al., 2008); as subsequently the circRNA *CDR1-AS*, was identified to act as a miRNA sponge for *hsa-miR-7* (Hansen et al., 2013), we investigated whether *CDR1-AS* itself was also a biomarker in breast cancer, and this study is detailed in **chapter 6**.

In **chapter 7** we investigated the abundance of circRNAs in an unbiased manner in human primary breast tumors and could identify one of them (circCNOT2) as a predictive biomarker of progression-free survival for AI therapy in patients with advanced disease.

Overall, this thesis has provided further information on non-coding RNAs, as well as drug response in breast cancer.

References

Allen, D.D., Caviedes, R., Cárdenas, A.M., Shimahara, T., Segura-Aguilar, J., and Caviedes, P.A. (2005). Cell lines as in vitro models for drug screening and toxicity studies. Drug Dev. Ind. Pharm. *31*, 757–768.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA Biogenesis Competes with Pre-mRNA Splicing. Mol. Cell *56*, 55–66.

Ashworth, A., and Lord, C.J. (2018). Synthetic lethal therapies for cancer: what's next after PARP inhibitors? Nat. Rev. Clin. Oncol. *15*, 564.

Begley, C.G., and Ellis, L.M. (2012). Drug development: Raise standards for preclinical cancer research. Nature *483*, 531–533.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *68*, 394–424.

Calin, G.A., and Croce, C.M. (2006). MicroRNA-Cancer Connection: The Beginning of a New Tale. Cancer Res. *66*, 7390–7394.

Carioli, G., Malvezzi, M., Rodriguez, T., Bertuccio, P., Negri, E., and La Vecchia, C. (2017). Trends and predictions to 2020 in breast cancer mortality in Europe. The Breast *36*, 89–95.

Carlomagno, C., Perrone, F., Gallo, C., De Laurentiis, M., Lauria, R., Morabito, A., Pettinato, G., Panico, L., D'Antonio, A., Bianco, A.R., et al. (1996). c-erb B2 overexpression decreases the benefit of adjuvant tamoxifen in early-stage breast cancer without axillary lymph node metastases. J. Clin. Oncol. *14*, 2702–2708.

Carroll, J.S. (2016). Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. Eur. J. Endocrinol. *175*, R41–R49.

Clark, G.M. (1998). Should Selection of Adjuvant Chemotherapy for Patients With Breast Cancer Be Based on erbB-2 Status? JNCI J. Natl. Cancer Inst. *90*, 1320–1321.

Curigliano, G., Burstein, H.J., Winer, E.P., Gnant, M., Dubsky, P., Loibl, S., Colleoni, M., Regan, M.M., Piccart-Gebhart, M., Senn, H.-J., et al. (2018). De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. Ann. Oncol. *29*, 2153–2153.

Dai, X., Xiang, L., Li, T., and Bai, Z. (2016). Cancer Hallmarks, Biomarkers and Breast Cancer Molecular Subtypes. J. Cancer *7*, 1281–1294.

Duffy, M.J., O'Donovan, N., McDermott, E., and Crown, J. (2016). Validated biomarkers: The key to precision treatment in patients with breast cancer. The Breast *29*, 192–201.

Dziadkowiec, K.N., Gąsiorowska, E., Nowak-Markwitz, E., and Jankowska, A. (2016). PARP inhibitors: review of mechanisms of action and BRCA1/2 mutation targeting. Przegląd Menopauzalny Menopause Rev. *15*, 215–219.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. The Lancet *365*, 1687–1717.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. The Lancet *378*, 771–784.

Fang, Z., Du, R., Edwards, A., Flemington, E.K., and Zhang, K. (2013). The Sequence Structures of Human MicroRNA Molecules and Their Implications. PLOS ONE *8*, e54215.

Fitzgibbons, P.L., Page, D.L., Weaver, D., Thor, A.D., Allred, D.C., Clark, G.M., Ruby, S.G., O'Malley, F., Simpson, J.F., Connolly, J.L., et al. (2000). Prognostic Factors in Breast Cancer. Arch. Pathol. Lab. Med. *124*, 966–978.

Foekens, J.A., Sieuwerts, A.M., Smid, M., Look, M.P., de Weerd, V., Boersma, A.W.M., Klijn, J.G.M., Wiemer, E.A.C., and Martens, J.W.M. (2008). Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. Proc. Natl. Acad. Sci. U. S. A. *105*, 13021–13026.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. N. Engl. J. Med. *361*, 123–134.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Gerber, D.E. (2008). Targeted Therapies: A New Generation of Cancer Treatments. Am. Fam. Physician *77*, 311–319.

Grasedieck, S., Schöler, N., Bommer, M., Niess, J.H., Tumani, H., Rouhi, A., Bloehdorn, J., Liebisch, P., Mertens, D., Döhner, H., et al. (2012). Impact of serum storage conditions on microRNA stability. Leukemia *26*, 2414–2416.

Greene, D., Nail, L.M., Fieler, V.K., Dudgeon, D., and Jones, L.S. (1994). A comparison of patient-reported side effects among three chemotherapy regimens for breast cancer. Cancer Pract. *2*, 57–62.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., et al. (2007). Patterns of somatic mutation in human cancer genomes. Nature *446*, 153–158.

Gusterson, B.A., Gelber, R.D., Goldhirsch, A., Price, K.N., Säve-Söderborgh, J., Anbazhagan, R., Styles, J., Rudenstam, C.M., Golouh, R., and Reed, R. (1992). Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. J. Clin. Oncol. *10*, 1049–1056.

Győrffy, B., Hatzis, C., Sanft, T., Hofstatter, E., Aktas, B., and Pusztai, L. (2015). Multigene prognostic tests in breast cancer: past, present, future. Breast Cancer Res. *17*, 11.

Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., et al. (2010). American Society of Clinical Oncology/ College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Arch. Pathol. Lab. Med. *134*, e48-72.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Harris, L.N., Ismaila, N., McShane, L.M., and Hayes, D.F. (2016). Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Summary. J. Oncol. Pract. *12*, 384–389.

Hausser, J., and Zavolan, M. (2014). Identification and consequences of miRNA-target interactions--beyond repression of gene expression. Nat. Rev. Genet. *15*, 599–612.

Hayes, D.F. (2015). Clinical utility of genetic signatures in selecting adjuvant treatment: Risk stratification for early vs. late recurrences. The Breast *24*, S6–S10.

Hayes, D.F., Allen, J., Compton, C., Gustavsen, G., Leonard, D.G.B., McCormack, R., Newcomer, L., Pothier, K., Ransohoff, D., Schilsky, R.L., et al. (2013). Breaking a Vicious Cycle. Sci. Transl. Med. *5*, 196cm6-196cm6.

Henry, N.L., Somerfield, M.R., Abramson, V.G., Allison, K.H., Anders, C.K., Chingos, D.T., Hurria, A., Openshaw, T.H., and Krop, I.E. (2016). Role of Patient and Disease Factors in Adjuvant Systemic Therapy Decision Making for Early-Stage, Operable Breast Cancer: American Society of Clinical Oncology Endorsement of Cancer Care Ontario Guideline Recommendations. J. Clin. Oncol. *34*, 2303–2311.

Hombach, S., and Kretz, M. (2016). Non-coding RNAs: Classification, Biology and Functioning. Adv. Exp. Med. Biol. *937*, 3–17.

http://cancer.sanger.ac.uk/census (2019). Cancer Gene Census.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. *12*, 99–110.

Incorvati, J.A., Shah, S., Mu, Y., and Lu, J. (2013). Targeted therapy for HER2 positive breast cancer. J Hematol Oncol. *6*, 38.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

Iorio, M.V., and Croce, C.M. (2012). MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol. Med. *4*, 143–159.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141–157.

Jung, M., Schaefer, A., Steiner, I., Kempkensteffen, C., Stephan, C., Erbersdobler, A., and Jung, K. (2010). Robust microRNA stability in degraded RNA preparations from human tissue and cell samples. Clin. Chem. *56*, 998–1006.

Ke, X., and Shen, L. (2017). Molecular targeted therapy of cancer: The progress and future prospect. Front. Lab. Med. *1*, 69–75.

Khan, J.A., Bellance, C., Guiochon-Mantel, A., Lombès, M., and Loosfelt, H. (2012). Differential Regulation of Breast Cancer-Associated Genes by Progesterone Receptor Isoforms PRA and PRB in a New Bi-Inducible Breast Cancer Cell Line. PLOS ONE *7*, e45993.

Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018). Circular RNAs in cancer: opportunities and challenges in the field. Oncogene *37*, 555–565.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat. Rev. Genet. *11*, 597–610.

Krop, I., Ismaila, N., Andre, F., Bast, R.C., Barlow, W., Collyar, D.E., Hammond, M.E., Kuderer, N.M., Liu, M.C., Mennel, R.G., et al. (2017). Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. J. Clin. Oncol. *35*, 2838–2847.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Leitzel, K., Teramoto, Y., Konrad, K., Chinchilli, V.M., Volas, G., Grossberg, H., Harvey, H., Demers, L., and Lipton, A. (1995). Elevated serum c-erbB-2 antigen levels and decreased response to hormone therapy of breast cancer. J. Clin. Oncol. *13*, 1129–1135.

Li, S.C., Chan, W.C., Hu, L.Y., Lai, C.H., Hsu, C.N., and Lin, W. (2010). Identification of homologous microRNAs in 56 animal genomes. Genomics *96*, 1–9.

Lin, S., and Gregory, R.I. (2015). MicroRNA biogenesis pathways in cancer. Nat. Rev. Cancer *15*, 321–333.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., et al. (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834–838.

Mariani, G., Fasolo, A., De Benedictis, E., and Gianni, L. (2009). Trastuzumab as adjuvant systemic therapy for HER2-positive breast cancer. Nat. Rev. Clin. Oncol. *6*, 93–104.

Mohammed, A., Biegert, G., Adamec, J., and Helikar, T. (2017). Identification of potential tissue-specific cancer biomarkers and development of cancer versus normal genomic classifiers. Oncotarget *8*, 85692–85715.

Monks, A., Scudiero, D., Skehan, P., Shoemaker, R., Paull, K., Vistica, D., Hose, C., Langley, J., Cronise, P., Vaigro-Wolff, A., et al. (1991). Feasibility of a High-Flux Anticancer Drug Screen Using a Diverse Panel of Cultured Human Tumor Cell Lines. JNCI J. Natl. Cancer Inst. *83*, 757–766.

Muss, H.B., Thor, A.D., Berry, D.A., Kute, T., Liu, E.T., Koerner, F., Cirrincione, C.T., Budman, D.R., Wood, W.C., Barcos, M., et al. (1994). c-erbB-2 Expression and Response to Adjuvant Therapy in Women with Node-Positive Early Breast Cancer. N. Engl. J. Med. *330*, 1260–1266.

Ng, C.K.Y., Schultheis, A.M., Bidard, F.-C., Weigelt, B., and Reis-Filho, J.S. (2015). Breast Cancer Genomics From Microarrays to Massively Parallel Sequencing: Paradigms and New Insights. JNCI J. Natl. Cancer Inst. *107,* djv015.

Nik-Zainal, S., and Morganella, S. (2017). Mutational Signatures in Breast Cancer: The Problem at the DNA Level. Clin. Cancer Res. *23*, 2617–2629.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Paik, S., Bryant, J., Park, C., Fisher, B., Tan-Chiu, E., Hyams, D., Fisher, E.R., Lippman, M.E., Wickerham, D.L., and Wolmark, N. (1998). erbB-2 and Response to Doxorubicin in Patients With Axillary Lymph Node-Positive, Hormone Receptor- Negative Breast Cancer. JNCI J. Natl. Cancer Inst. *90*, 1361–1370.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Polyak, K. (2007). Breast cancer: origins and evolution. J. Clin. Invest. *117,* 3155–3163.

Press, M.F., Sauter, G., Bernstein, L., Villalobos, I.E., Mirlacher, M., Zhou, J.-Y., Wardeh, R., Li, Y.-T., Guzman, R., Ma, Y., et al. (2005). Diagnostic Evaluation of HER-2 as a Molecular Target: An Assessment of Accuracy and Reproducibility of Laboratory Testing in Large, Prospective, Randomized Clinical Trials. Clin. Cancer Res. *11*, 6598–6607.

Pritchard, K.I., Shepherd, L.E., O'Malley, F.P., Andrulis, I.L., Tu, D., Bramwell, V.H., and Levine, M.N. (2006). *HER2* and Responsiveness of Breast Cancer to Adjuvant Chemotherapy. N. Engl. J. Med. *354*, 2103–2111.

Reis-Filho, J.S., and Pusztai, L. (2011). Gene expression profiling in breast cancer: classification, prognostication, and prediction. The Lancet *378*, 1812–1823.

Sawyers, C. (2004). Targeted cancer therapy. Nature *432*, 294–297.

Sawyers, C.L. (2008). The cancer biomarker problem. Nature *452*, 548–552.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., et al. (2011). Adjuvant Trastuzumab in HER2-Positive Breast Cancer. N. Engl. J. Med. *365*, 1273–1283.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., and McGuire, W.L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science *235*, 177–182.

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., et al. (2001). Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. N. Engl. J. Med. *344*, 783–792.

Sobhani, N., D'Angelo, A., Pittacolo, M., Roviello, G., Miccoli, A., Corona, S.P., Bernocchi, O., Generali, D., and Otto, T. (2019). Updates on the CDK4/6 Inhibitory Strategy and Combinations in Breast Cancer. Cells *8*, E321.

Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U. S. A. *98*, 10869–10874.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. *100*, 8418–8423.

Stephens, P.J., McBride, D.J., Lin, M.-L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J., et al. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*, 1005–1010.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., et al. (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400–404.

Storz, G. (2002). An Expanding Universe of Noncoding RNAs. Science *296*, 1260–1263.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Tarver, J.E., Donoghue, P.C.J., and Peterson, K.J. (2012). Do miRNAs have a deep evolutionary history? BioEssays *34*, 857–866.

The ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.

The ENCODE Project Consortium (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature *489*, 57–74.

Thor, A.D., Berry, D.A., Budman, D.R., Muss, H.B., Kute, T., Henderson, I.C., Barcos, M., Cirrincione, C., Edgerton, S., Allred, C., et al. (1998). erbB-2, p53, and Efficacy of Adjuvant Therapy in Lymph Node-Positive Breast Cancer. JNCI J. Natl. Cancer Inst. *90*, 1346–1360.

Uhlén, M., Björling, E., Agaton, C., Szigyarto, C.A.-K., Amini, B., Andersen, E., Andersson, A.-C., Angelidou, P., Asplund, A., Asplund, C., et al. (2005). A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. Mol. Cell. Proteomics *4*, 1920–1932.

Van Poznak, C., Somerfield, M.R., Bast, R.C., Cristofanilli, M., Goetz, M.P., Gonzalez-Angulo, A.M., Hicks, D.G., Hill, E.G., Liu, M.C., Lucas, W., et al. (2015). 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*, 2695–2704.

Varela, M.A., Roberts, T.C., and Wood, M.J.A. (2013). Epigenetics and ncRNAs in brain function and disease: mechanisms and prospects for therapy. Neurother. J. Am. Soc. Exp. Neurother. *10*, 621–631.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wang, J., Chen, J., and Sen, S. (2016). MicroRNA as Biomarkers and Diagnostics. J. Cell. Physiol. *231*, 25–30.

Wang, M., Yang, Y., Xu, J., Bai, W., Ren, X., and Wu, H. (2018). CircRNAs as biomarkers of cancer: a meta-analysis. BMC Cancer *18*:303.

Weigelt, B., Pusztai, L., Ashworth, A., and Reis-Filho, J.S. (2012). Challenges translating breast cancer gene signatures into the clinic. Nat. Rev. Clin. Oncol. *9*, 58–64.

Widakowich, C., Castro, G. de, Azambuja, E. de, Dinh, P., and Awada, A. (2007). Review: Side Effects of Approved Molecular Targeted Therapies in Solid Cancers. The Oncologist *12*, 1443–1455.

Wright, C., Nicholson, S., Angus, B., Sainsbury, J., Farndon, J., Cairns, J., Harris, A., and Horne, C. (1992). Relationship between c-erbB-2 protein product expression and response to endocrine therapy in advanced breast cancer. Br. J. Cancer *65*, 118–121.

Würth, R., Thellung, S., Bajetto, A., Mazzanti, M., Florio, T., and Barbieri, F. (2016). Drug-repositioning opportunities for cancer therapy: novel molecular targets for known compounds. Drug Discov. Today *21*, 190–199.

www.cancer.gov (2019). Drugs Approved for Breast Cancer.

Xu, Q., Chen, J., Ni, S., Tan, C., Xu, M., Dong, L., Yuan, L., Wang, Q., and Du, X. (2016). Pan-cancer transcriptome analysis reveals a gene expression signature for the identification of tumor tissue origin. Mod. Pathol. *29*, 546–556.

Chapter 3

Understanding drugs in breast cancer through drug sensitivity screening

Katharina Uhr, Wendy J. C. Prager-van der Smissen, Anouk A. J. Heine, Bahar Ozturk, Marcel Smid, Hinrich W.H. Göhlmann, Agnes Jager, John A. Foekens, John W. M. Martens

Springerplus 2015;4:611

Abstract

With substantial numbers of breast tumors showing or acquiring treatment resistance, it is of utmost importance to develop new agents for the treatment of the disease, to know their effectiveness against breast cancer and to understand their relationships with other drugs to best assign the right drug to the right patient. To achieve this goal drug screenings on breast cancer cell lines are a promising approach. In this study a large-scale drug screening of 37 compounds was performed on a panel of 42 breast cancer cell lines representing the main breast cancer subtypes. Clustering, correlation and pathway analyses were used for data analysis. We found that compounds with a related mechanism of action had correlated IC_{50} values and thus grouped together when the cell lines were hierarchically clustered based on IC_{50} values. In total we found six clusters of drugs of which five consisted of drugs with related mode of action and one cluster with two drugs not previously connected. In total, 25 correlated and four anti-correlated drug sensitivities were revealed of which only one drug, Sirolimus, showed significantly lower IC_{50} values in the luminal/ERBB2 breast cancer subtype. We found expected interactions but also discovered new relationships between drugs which might have implications for cancer treatment regimens.

Keywords: Drugs, screening, cell line, subtype, pathway, breast cancer

Background

Life expectancy and survival of breast cancer patients have increased significantly over the last decades, due to – amongst other factors – an increasing number of effective drug therapies (Berry et al., 2005; Lichtenberg, 2009, 2011). Drug resistance remains a major issue (Gonzalez-Angulo et al., 2007) and since the discovery that expression of the protein markers ER, PR and her-2/neu determines response to a given targeted therapy (Bast et al., 2001), the assessment of their expression in breast cancer has become an important first step in selecting a patient's treatment (Bast et al., 2001). Subsequently, microarray studies have shown insight into molecular processes active in the tumor and linked those to diverse clinical outcomes (Sorlie et al., 2001; van 't Veer et al., 2002; Wang et al., 2005) including therapy failure (Jansen et al., 2005). In the last couple of years large scale next generation sequencing efforts have made a big contribution to our understanding of breast cancer by delivering precise information on cancer driver mutations (Desmedt et al., 2012; Kangaspeska et al., 2012; Previati et al., 2013; Radovich et al., 2013; The Cancer Genome Atlas Network, 2012). All these sources of

information combined have helped to elucidate how breast cancer evolves, progresses and metastasizes and some of them have led to the development of diagnostic tests to characterize breast cancer better (Kittaneh et al., 2013). Nevertheless, there is still significant room for improvement in regard to available drug therapies, as many patients do not respond to current treatments or become resistant during the course of treatment (Gonzalez-Angulo et al., 2007). New agents are therefore needed to target breast cancer, and screenings of multiple compounds for their activity against the various breast cancer subtypes are a good starting point. As a first step to test new compounds breast cancer cell lines are a good model, because they are easy to maintain, represent different subtypes of breast cancer, and the response to drug treatment can be easily assessed. For these reasons, we studied the activity of a wide variety of cytotoxic and targeted drugs in a large panel of breast cancer cell lines. The drugs were chosen based on current clinical utility e.g. for discrete cancer subtypes, potential clinical utility such as promising compounds in pre-clinical testing, aiming at molecular targets, and – for comparison – current state of the art drugs for the therapy of breast cancer. We investigated which drugs showed similar activity in the panel of breast cancer cell lines and could therefore potentially substitute or complement each other in the clinic, and, in addition, we aimed to identify shared biology in cell lines that are affected by highly correlated drugs.

Results

Relationships between drugs: clustering and correlation analysis

To investigate the relationships between different drugs the IC_{50} values of all 7 cytotoxic drugs and 30 targeted agents, measured in the 42 breast cancer cell lines, were correlated (**Fig. 1**). Capecitabine, cMet 605 and Cyclophosphamide exhibited no differential IC_{50} values and were consequently omitted from the clustering and further analyses. To express the relationships among drugs and cell lines hierarchical clustering was performed (**Fig. 2**). Clustering and correlation performed fairly similarly and are therefore discussed together.

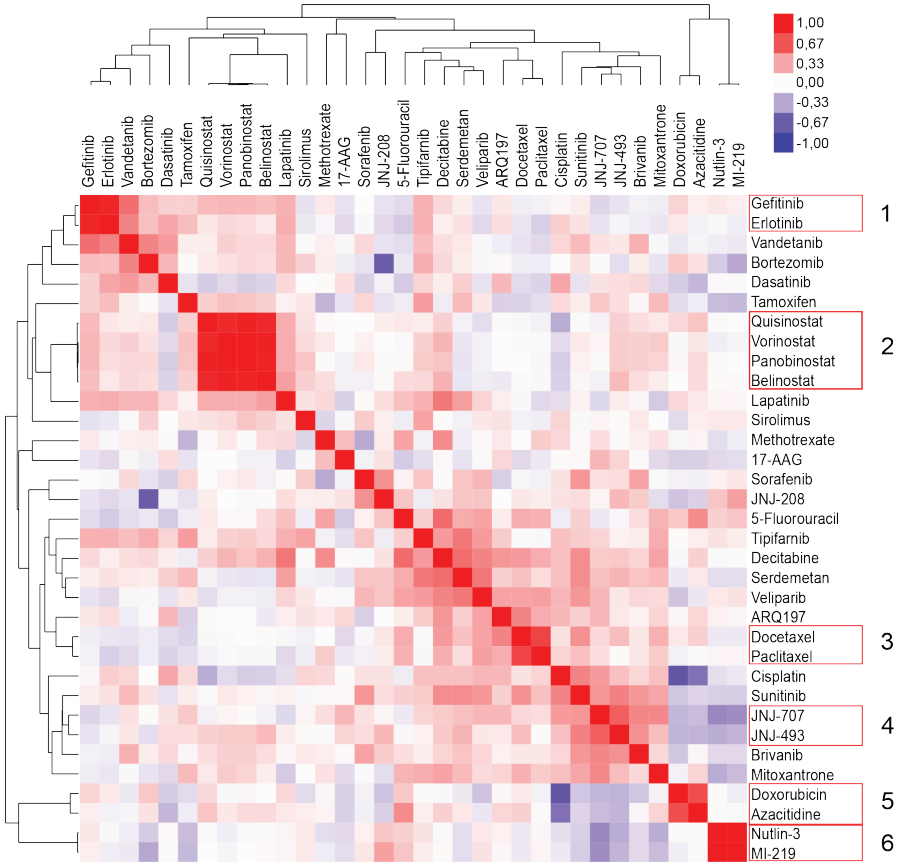


Fig. 1. Pearson correlation plot of absolute drug IC_{50} values. The *red color* indicates a positive correlation between the IC_{50} values of two drugs, and *blue* a negative correlation. The *color intensity* illustrates the correlation coefficient as shown in the legend at *top right*. Drugs are clustered on the basis of similarity; distances in the tree indicate the degree of difference between drugs.

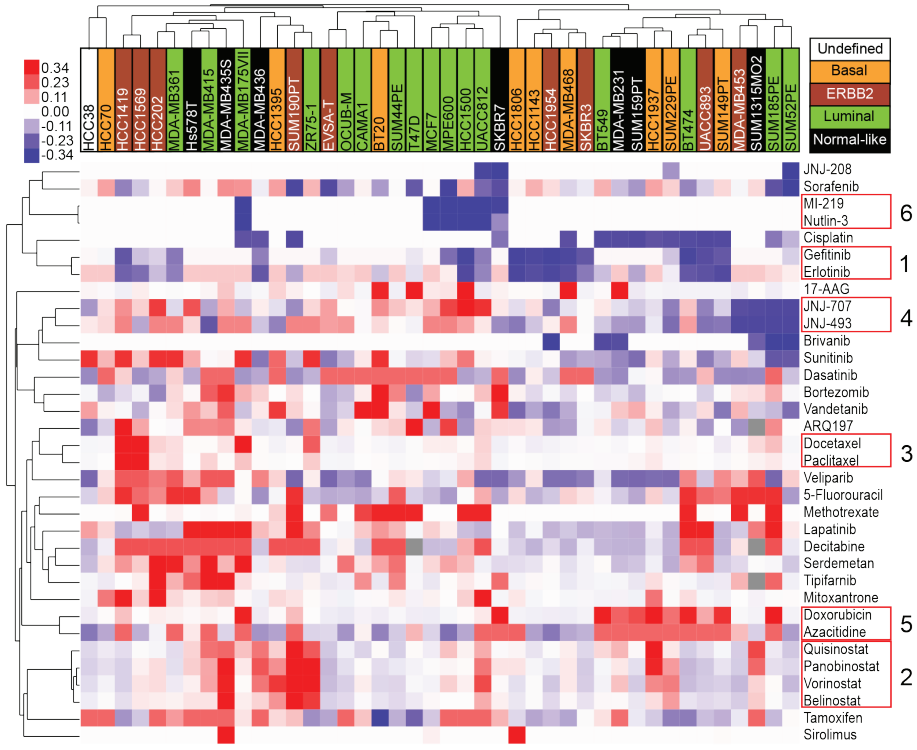


Fig. 2. Similar drugs cluster together. Depicted is a hierarchical unsupervised clustering of the analyzable drugs and cell lines. *Blue color* indicates low IC_{50} values (i.e. cells are drug-sensitive), and *red color* high IC_{50} values (i.e. cells are drug-resistant). *Color intensity* illustrates the degree of drug sensitivity or resistance; outliers exceeding the legend boundaries are set to the maxima colors of the legend to ensure visibility of small differences instead of few outliers. Breast-cancer subtypes are color-coded on the basis of the intrinsic subtypes of breast cancer cell lines as previously described (Riaz et al., 2013). The respective legend can be found on the *top right*. Tree distance is representative for similarity of drugs or cell lines. Drugs with similar response profiles among the cell lines are highlighted by *red boxes*.

Strong correlation and expected co-clustering was observed between Gefitinib and Erlotinib (cluster 1; r = 0.88), between Quisinostat, Panobinostat, Vorinostat and Belinostat (cluster 2; r = 0.85-0.96), between Docetaxel and Paclitaxel (cluster 3; r = 0.73), between JNJ-707 and JNJ-493 (cluster 4; r = 0.62) and between MI-219 and Nutlin-3 (cluster 6; r = 0.98); all correlations are listed additionally in **Table 1**. To illustrate the close relationship between related drugs, the IC_{50} values of MI-219 and Nutlin-3, the two drugs with the highest correlation, were ranked and plotted for all cell lines (**Fig. 3**). Interestingly, Serdemetan, a drug which acts on cholesterol transport but also targets MDM2 (Jones et al., 2013) – a mechanism shared with Nutlin-3 and MI-219 (Shangary and Wang, 2009) – showed no correlation with these two compounds.

Table 1. Correlated drugs

Drug 1	Drug 2	p-value	Pearson correlation coefficient
MI-219	Nutlin-3	1.77E-28	0.98
Panobinostat (Faridak®)	Vorinostat (Zolinza®)	2.14E-24	0.96
Panobinostat (Faridak®)	Quisinostat	1.66E-19	0.93
Belinostat	Vorinostat (Zolinza®)	2.05E-18	0.92
Belinostat	Panobinostat (Faridak®)	1.70E-16	0.91
Erlotinib (Tarceva®)	Gefitinib (Iressa®)	3.49E-14	0.88
Quisinostat	Vorinostat (Zolinza®)	1.05E-13	0.87
Belinostat	Quisinostat	8.08E-13	0.85
Paclitaxel (Taxol®, OnxalTM)	Docetaxel (Taxotere®)	4.61E-08	0.73
Azacitidine (Vidaza®)	Doxorubicin (Adriamycin®)	3.77E-07	0.7
JNJ-493	JNJ-707	1.39E-05	0.62
Decitabine (Dacogen®)	5-Fluorouracil	7.77E-05	0.58
Decitabine (Dacogen®)	Serdemetan	1.17E-04	0.56
Vandetanib (Zactima®)	Gefitinib (Iressa®)	1.52E-04	0.56
Serdemetan	Tipifarnib (Zarnestra®)	5.15E-04	0.52
Decitabine (Dacogen®)	Lapatinib	5.29E-04	0.52
Veliparib	Serdemetan	5.47E-04	0.51
JNJ-493	Sunitinib (Sutent®)	1.37E-03	0.48
Veliparib	Decitabine (Dacogen®)	1.63E-03	0.48
Vandetanib (Zactima®)	Erlotinib (Tarceva®)	1.78E-03	0.47
Bortezomib (Velcade®)	Vandetanib (Zactima®)	1.94E-03	0.47
ARQ197	Docetaxel (Taxotere®)	1.95E-03	0.47
Cisplatin	Sunitinib (Sutent®)	2.00E-03	0.47
JNJ-707	Brivanib	2.16E-03	0.46
Mitoxantrone (Novantrone®)	JNJ-707	2.98E-03	0.45
JNJ-707	Nutlin-3	2.87E-03	-0.45
Cisplatin	Azacitidine (Vidaza®)	2.16E-04	-0.55
JNJ-208	Bortezomib (Velcade®)	1.96E-06	-0.66
Cisplatin	Doxorubicin (Adriamycin®)	5.22E-08	-0.73

Correlation pairs were determined using IC_{50} values. Statistical thresholds for significance were defined as a p-value <0.01 and a Pearson correlation coefficient above 0.45 or below -0.45.

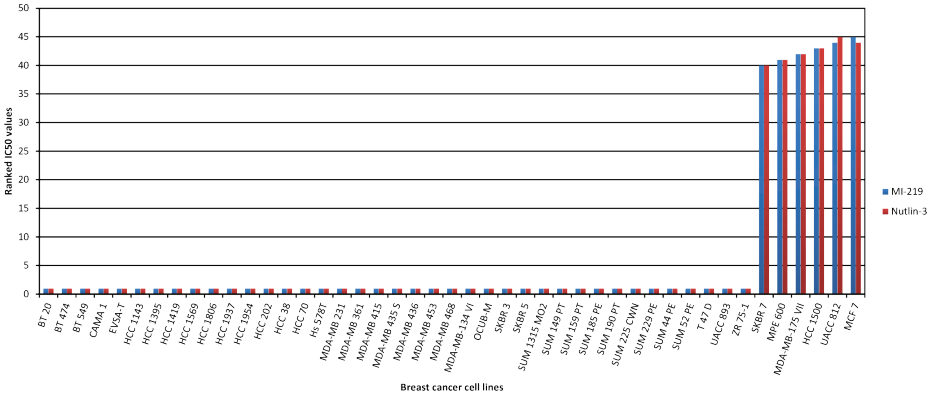


Fig. 3. Nutlin-3 and MI-219 have similar drug sensitivity profiles among the cell lines (relative IC_{50} values). The relative IC_{50} is inverted, with high numbers indicating sensitivity in this case and not resistance. Few cell lines are sensitive to these drugs, while the majority is resistant.

Unanticipated but highly significant correlations were observed between particularly Doxorubicin and Azacitidine (cluster 5; r = 0.70), between Decitabine and 5-Fluorouracil (r = 0.58) and Serdemetan (r = 0.56); and between Serdemetan and Tipifarnib (r = 0.52). Additional weaker, but expected correlations were found for Vandetanib with Erlotinib and Gefitinib (r = 0.47, r = 0.56). Decitabine was correlated with Lapatinib (r = 0.52) and Veliparib with Serdemetan and Decitabine (r = 0.51, r = 0.48). Furthermore, we also detected a remote relation between various tyrosine kinase inhibitors like JNJ-493 with the multi-receptor tyrosine kinase inhibitor Sunitinib (Keyvanjah et al., 2012) (r = 0.48), JNJ-707 with FGFR- and VEGFR-inhibitor Brivanib (Huynh et al., 2008) (r = 0.46) and between Docetaxel and ARQ197 (r = 0.47). The DNA targeting drug Cisplatin (Becker et al., 2014) showed surprisingly a correlation with Sunitinib (r = 0.47); Bortezomib was correlated with Vandetanib (r = 0.47) and the type II topoisomerase inhibitor Mitoxantrone (Hajihassan and Rabbani-Chadegani, 2009) was correlated with JNJ-707 (r = 0.45). In total, 25 pairs of positively correlated drugs were found.

Apart from positive correlations – and even more interesting – we also discovered significant negative correlations between certain drugs (**Table 1**). Particularly, Doxorubicin and the correlated drug Azacitidine had negative correlations with Cisplatin (r = -0.73, r = -0.55), the ERR1 targeting JNJ-208 with Bortezomib (r = -0.66), and the MDM2-targeting Nutlin-3 (Shangary and Wang, 2009) with the FGFR inhibitor JNJ-707 (r = -0.45).

Shared pathways between correlated drugs

To further understand the biology behind correlated drugs we used mRNA expression data of the untreated cell lines and the pathway information of the databases Biocarta and KEGG (Ogata et al., 1999) to characterize drug resistance in R (R_Core_Team, 2013). We identified significant pathways for each of the evaluable drugs, but focused on the pathways which were shared among correlated drugs, i.e. for the 23 positively and 3 negatively correlated remaining drug-drug pairs. Furthermore, we performed a pathway analysis where cell lines were grouped per subtype to identify subtype-related pathways. Subtype-specific pathways were excluded from further study in the pathway-drug resistance analysis. At a significance level of $p < 0.01$, only one of all 26 correlation pairs had pathways in common. This pair, Nutlin-3 and MI-219, had, after correction for subtype-specific pathways, only the DNA replication pathway in common. The Nutlin-3- and MI-219-associated genes of this pathway are displayed in **Fig. 4**.

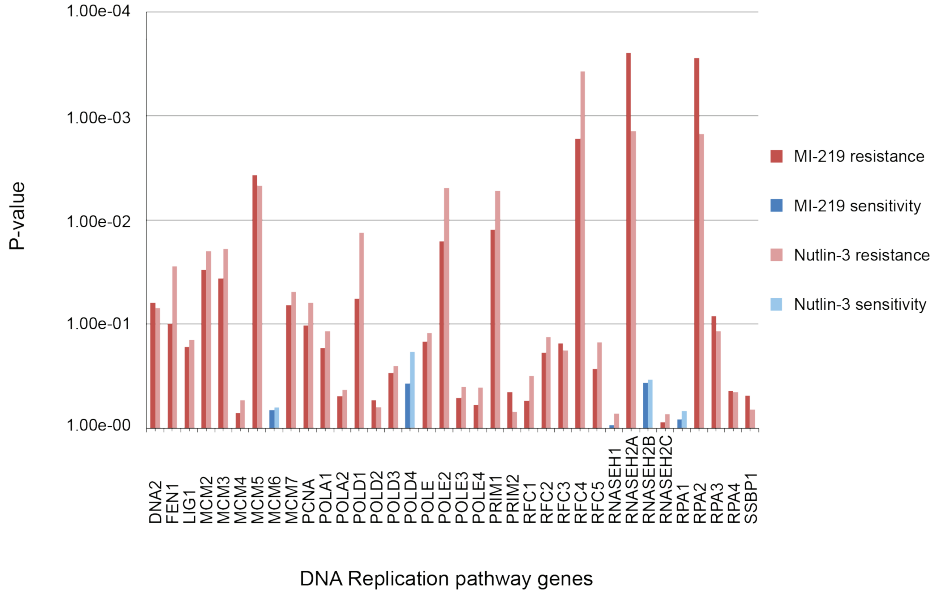


Fig. 4. Differentially expressed genes of the DNA replication pathway for Nutlin-3 and MI-219. *Bar graphs* display the differentially expressed genes of this pathway between resistant and sensitive cell lines for Nutlin-3 and MI-219. *Red shades* indicate an association with resistance, *blue shades* indicate an association with sensitivity.

Breast cancer subtype specific drugs

Earlier, several subtype-specific differences in drug sensitivity were observed (Heiser et al., 2012) and since breast cancer subtypes are biologically very different (Parker et al., 2009), we also explored whether drug response in our study was ER- or subtype-

related. Only one drug, Sirolimus, exhibited a significantly different subtype-specific effectiveness. Normal-like and basal cell lines were more resistant to this drug compared to luminal and ERBB2-overexpressing cell lines with a change in sensitivity of two orders of magnitude ($p = 0.005$). The expression of ER by the latter two subtypes was not the sole explanation though, as none of the screened drugs was associated with ER status (p-value > 0.01).

Discussion

Drug response to one drug indicates the response to another

To understand drug resistance in breast cancer, we compared drug sensitivity of a large set of drugs within a large panel of breast cancer cell lines. It became evident that some drugs target breast cancer cell lines similarly and thus may have unanticipated overlapping mechanisms while others display opposing effects indicating that vulnerability to a given drug is protective for another unrelated treatment.

The results of the overall clustering (**Fig. 2**) show that every breast cancer cell line had a unique drug response profile, which might be true for patients as well. This – first – observation underlines the personal factor in drug sensitivity, which we need to understand upfront to provide optimal patient care.

The second, expected, conclusion is that drugs with identical targets such as MDM2-antagonists (MI-219 and Nutlin-3) (Shangary and Wang, 2009), EGFR-inhibitors (Gefitinib and Erlotinib) (Cohen, 2003), FGFR-inhibitors (JNJ-707 and JNJ-493), HDAC inhibitors (Quisinostat, Panobinostat, Vorinostat, Belinostat) (Lemoine and Younes, 2010) and taxanes (Docetaxel and Paclitaxel) (Hagiwara and Sunada, 2004), showed correlated sensitivities and clustered together explaining five of the six observed clusters.

More interesting was the third observation that unrelated drugs showed co-clustering, which is best exemplified by the sixth cluster (**Figs. 1, 2**), made up of the positively correlated intercalating agent Doxorubicin (Frederick et al., 1990) and the DNA-methyltransferase-targeting Azacitidine (Creusot et al., 1982). Interestingly and remarkably, Decitabine, a derivative of Azacitidine (Lyko and Brown, 2005), which also targets a DNA-methyltransferase (Creusot et al., 1982), did not cluster with these two drugs. The reason for this might be that both Azacitidine and Doxorubicin have, next to their well-known properties, also the less known capability to interfere with RNA synthesis (Christman, 2002; Momparler et al., 1976), while Decitabine can only act on

DNA (Christman, 2002). Next to this most notable finding we also observed a less strong correlation of Decitabine sensitivity with sensitivity to various unrelated drugs, i.e. the thymidylate synthetase inhibitor 5-Fluorouracil (Longley et al., 2003), the cholesterol transport inhibitor and MDM2-antagonist Serdemetan (Jones et al., 2013; Lehman et al., 2013), the EGF-receptor- and HER2-inhibitor Lapatinib (Huang and Rizzo, 2012) and the PARP-inhibitor Veliparib (Glendenning and Tutt, 2011). Some of these drugs additionally correlated with each other. Although several of these compounds target DNA synthesis and/or repair, there is no real common denominator between them. While these drugs could be targeted by the same drug efflux pumps, we could not find any among the drug-associated genes (pre-treatment gene expression) and suspect another, unknown mechanism. The same holds true for the associations of Serdemetan with Veliparib and the farnesyltransferase inhibitor Tipifarnib (Armand et al., 2007).

Additionally, we also observed a rather surprising correlation between Docetaxel, which disorganizes microtubules (Hagiwara and Sunada, 2004), and the c-met kinase inhibiting agent ARQ197 (Scagliotti et al., 2013). However, supporting our findings, ARQ197 has also been linked to inhibition of microtubuli polymerization recently (Katayama et al., 2013).

Finally, sensitivity to Bortezomib, a proteasome inhibitor (Teicher et al., 1999), was also predictive of sensitivity to the multi-kinase inhibitor Vandetanib (Sathornsumetee and Rich, 2006) - the combination of these two drugs is currently in clinical trial testing (Gramza et al., 2011). Thus a protein or complex whose stability is proteasome-dependent, may affect sensitivity to this multi-kinase inhibitor. Furthermore, Cisplatin and JNJ-493 were somewhat correlated to Sunitinib sensitivity and JNJ-707 weakly correlated with response to Mitoxantrone, findings which remain to be understood.

The fourth finding was the absent or poor correlation of drugs acting on the same target, such as Serdemetan which was not correlated with the other two well-known and highly correlated MDM2-inhibitors Nutlin-3 and MI-219 (Shangary and Wang, 2009). Furthermore, Serdemetan lacked a correlation with TP53 mutation status (data not shown) highlighting that this putative MDM2-inhibitor acts differently from the other MDM2-inhibitors. This unexpected observation can be explained by the additional characteristics of Serdemetan, inhibition of the cholesterol transport and the increased degradation of ABCA1 (Jones et al., 2013). Clearly these additional properties dominate over the MDM2-inhibiting role. Moreover, another study confirms that sensitivity to Serdemetan is independent of TP53-mutation status (Jones et al., 2013).

Furthermore, we noted that Lapatinib and Vandetanib, two EGFR-antagonists (Nelson and Dolder, 2006; Sathornsumetee and Rich, 2006) neither clustered immediately next to Gefitinib or Erlotinib nor next to each other and the correlation coefficient was also lower than expected for Vandetanib, while Lapatinib did not correlate with the other EGFR-antagonists at all. In both cases this is less surprising as both Vandetanib and Lapatinib act on additional targets, like HER2 for Lapatinib (Nelson and Dolder, 2006), and the two proteins VEGFR2 and RET Kinase for Vandetanib (Sathornsumetee and Rich, 2006). We tried to support this hypothesis by correlating the drug response with mRNA expression data of EGFR, HER2, FGFR, VEGFR2, and RET Kinase for all those drugs. However, none of the correlations was significant, which might be different if pre-treatment protein expression data is used, as the proteins are the direct drug targets.

Similarly, Brivanib showed only a weak correlation with one of the FGFR-inhibitors, JNJ-707, in our panel, which might be due to its additional target VEGFR (Huynh et al., 2008).

Drug response to one drug indicates resistance to another

The fifth interesting observation was that for some drugs sensitivity predicted insensitivity to another drug. For instance – and most prominent – Cisplatin resistance correlated with Doxorubicin sensitivity. Cisplatin's mode of action involves DNA and RNA interstrand linkage (Stordal and Davey, 2007), while Doxorubicin blocks DNA unwinding (Fornari et al., 1994), this difference does however not explain the clearly opposing character in drug response. The clinical implication might, nevertheless, be, that a patient showing insensitivity to Doxorubicin upon treatment start, might be more likely to respond to a course of Cisplatin therapy (Perilongo et al., 2009). Another implication of this finding is that a mechanism responsible for resistance to Doxorubicin reveals a target that provides synthetic lethality to Cisplatin or vice versa.

Previously, Cisplatin resistance was found to correlate with Taxane sensitivity (Stordal et al., 2007), a finding, we could not confirm in the present study.

The FGFR inhibitor JNJ-707 had an inverse correlation with Nutlin-3 response. Therefore, we also investigated whether TP53 wild-type cell lines (Riaz et al., 2013), which are sensitive to Nutlin-3, have a different expression of FGFR genes in contrast to mutant cell lines, but found no significant difference (data not shown). While a true biological effect cannot be excluded, it has to be mentioned that only few cell lines were sensitive to Nutlin-3 and our observation might be due to the low numbers. Finally, the proteasome inhibitor Bortezomib is negatively correlated to JNJ-208. Thus, if this observation implies causality, a proteasome-dependent mechanism affects vulnerability to this drug.

Biology of correlated drugs

Next, we wanted to uncover shared biology of drug sensitivity in correlated drugs by performing a pathway analysis. The first thing we discovered was that there were hardly any shared pathways after excluding pathways with a strong subtype-association. This was rather surprising as we did not find many subtype-associated drugs, but can be explained by the analysis type since for the pathway analysis we used only the cell lines around the minimum and maximum drug response, while for the subtype analysis all cell lines were included. Hence the smaller number of cell lines might have introduced a bias, but generally, it seems that the biology driving the subtypes in breast cancer largely obscures the possible drug-related pathways. The two correlated drugs which had the DNA replication pathway in common were Nutlin-3 and MI-219. Nutlin-3 has been previously shown to downregulate proteins involved in DNA replication (Kumamoto et al., 2008), a process likely influenced by MI-219, as well. The cell lines which were sensitive to these drugs had intrinsically low expression of most pathway-associated genes pre-treatment and the drug-related shutdown of the remaining expression might be contributing to lethality.

From the results of the subtype-specific pathways it became obvious that normal-like cell lines are very different from luminal ones. However, when we did a global test to evaluate whether a certain breast cancer subtype showed an overall increased sensitivity to the tested compounds we found no differences. Therefore, we could not confirm a general drug resistance of normal-like cells which would be expected due to their mesenchymal and stem cell like properties (Al-Hajj et al., 2003; Ponti et al., 2006; Sieuwerts et al., 2009).

Subtype-specific drugs

Of all tested drugs in this screening only one drug, Sirolimus, was more active in the luminal and ERBB2-high subtypes, as was noted previously in a comparable study (Heiser et al., 2012). However, in contrast to this earlier study (Heiser et al., 2012), who discovered 23 subtype-related drugs, we did not find subtype-dependent sensitivity for the other eight drugs screened in both studies. This discrepancy is likely due to several differences in study design, e.g. drug incubation times, type of readout assay, the use of collagen-coated plates in our study to mimic cellular context better and differences in the assayed cell lines to name a few.

Conclusion

Through our cell line screening with new and well-known drugs, we found a number of interesting interactions between drugs of which several were not noticed earlier. Those findings have great potential for an application in the clinic as they might present opportunities when tumors show already resistance upon start of the treatment.

Next to expected similar sensitivity profiles for related drugs such as Gefitinib and Erlotinib, we also found opposing sensitivity profiles such as Cisplatin with Doxorubicin and confirmed one subtype-related drug, Sirolimus, which has been identified earlier. Further validation on the discovered positive and negative correlations and the subtype-specific drug are needed e.g. in the form of an animal study. In that aspect it would be very interesting to investigate whether animals with e.g. a Cisplatin-resistant tumor benefit from Doxorubicin treatment.

To conclude, our study provides new leads in the search for effective treatments especially in the context of inherent drug resistance.

Methods

Cell lines and drug screening

Forty-five breast cancer cell lines with confirmed identity (Riaz et al., 2013) and known origin (Hollestelle et al., 2010a) were cultured and screened in RPMI 1640 medium (Life Technologies, Paisley, UK) containing 10% FBS (Lonza, Walkersville, USA). Ninety-six well collagen I-coated plates (BD Biosciences, San Jose, USA) were used for drug screening. Each drug - cell line combination was assayed in triplicate. Cells were seeded at the density required to reach the end of the exponential growth phase at 120h of culture. Drug incubation was started 24h post-seeding and lasted 96h. For each drug 12 different dilutions were tested starting from 1.00^{-5} to 3.00^{-11} M (final concentration), except for Bortezomib (2.00^{-5} to 6.00^{-11} M), Sirolimus and 17-AAG (both: 2.00^{-6} to 6.00^{-12} M). Drug diluent was used as negative control treatment. DMSO was used as drug solvent and diluent for all drugs except for Methotrexate (drug solvent: 1M NaOH, drug diluent: 0.9% NaCl solution) and Cyclophosphamide (drug solvent: PCR-grade water, drug diluent: 0.9% NaCl solution).

Three cell lines, SUM225CWN, MDA-MB-134VI, and SKBR5, failed in our drug screening due to slow growth or half-suspension growth, which is incompatible with the SRB assay, resulting in 42 cell lines for analysis.

Assessment of drug sensitivity and IC_{50} calculation

Cell line growth was determined by measuring the total protein amount per well using the Sulforhodamine B Assay (SRB) (Voigt, 2005): After the medium was discarded the cells were incubated with 10% TCA for 60 min at 4°C for fixation. Then the cells were thoroughly washed 5x with distilled water, air-dried and incubated with 0.4% SRB solution for 2h for protein-staining. Additional washing steps followed using 1% acetic acid (4x) and cells were again air-dried. TRIS (10 mM) was added to the cells to dissolve the SRB overnight. Absorbances were measured at 570 nm in an Ascent MultiSkan (Thermo Electron Corporation, Waltham, USA). If required, samples were further diluted with TRIS to ensure optimal measurements. IC_{50} values were calculated using the respective absorbance values and are listed, besides all IC_{50} profiles, in the supplemental Excel file (**Additional File 1**).

Clustering of cell lines and drug-drug correlation analyses

Cell lines were clustered in a hierarchical fashion based on their IC_{50} values. For drug correlation analysis, IC_{50} values per compound of each cell line were correlated with each other in Excel 2007 (Microsoft, Redmond, USA) using Pearson correlation. Cell lines with missing data for several drugs were discarded from correlation and cluster analysis, as were drugs that did not show differential IC_{50} values. The programs Cluster 3.0 (Eisen Lab, Stanford University, Stanford, USA) and Java Treeview version 1.1.6r2 (Saldanha, 2004) were used to generate heatmaps of the correlation coefficients and the cluster analysis (**Figs. 1, 2**). Cluster 3.0 settings were as follows: normalize, median - center and average linkage using uncentered correlation as similarity metric. **Figure 3** was generated in Excel. To generate the figures Inkscape 0.91 (Free Software Foundation Inc., Boston, USA) was used, next to the aforementioned programs.

Association of signaling pathways with drug response

For pathway analysis we analyzed gene expression levels of cell lines with a high IC_{50}- versus cell lines with a low IC_{50}-value per drug. Cell lines were grouped in the high or low group by individual evaluation per drug instead of pre-selecting a fixed IC_{50} value for all drugs. This method was chosen to test the drug response extremes rather than testing values with little difference, which are present in gradual IC_{50} distributions. For this distribution type we used the cell lines at the distribution extremes and removed the intermediate values to reduce noise. A few drugs showed an IC_{50} profile that precluded

a sensible grouping; e.g. Paclitaxel had only two cell lines with a high IC_{50} value while the others had a very low IC_{50}; too few cell lines in a group renders the pathway analysis useless. For this reason, we excluded JNJ-208, Sirolimus, Docetaxel and Paclitaxel from the pathway analysis. For all other drugs, we were able to use meaningful group sizes of at least five samples each (**Additional File 1**).

For all cell lines, previously published mRNA expression data of our laboratory (Riaz et al., 2013) was used for pathway analysis (NCBI's Gene Expression Omnibus database, entry GSE41313). Pathway analysis was performed using the Global Test package (Goeman et al., 2004) in R (R_Core_Team, 2013) with information of the databases BioCarta LLC (San Diego, USA) and KEGG (Ogata et al., 1999). This R package was also used to generate **Figure 4**, next to Excel 2007 and Inkscape 0.91. Furthermore, we also tested the identified pathways for a stronger association with breast cancer subtypes and disregarded those subtype-associated pathways (**Additional File 2**). The pathways significantly associated with a drug, including the subtype-associated ones, are listed in **Additional File 2**.

Association of drug response with breast cancer subtype and ER status

All drugs were tested for association with breast cancer cell line subtypes. First, cell lines were grouped into luminal, basal, ERBB2-overexpressing and normal-like on the basis of intrinsic subtypes as previously reported (Hollestelle et al., 2010b). Statistical testing was performed either in BRB Array Tools version 4.2.1 Class Comparison using a T-test or in Analyse-it version 2.26 (Leeds, UK) for Chi-Square tests. To test whether ER protein expression was significantly associated with drug response, a Mann-Whitney test was used for linear IC_{50} profiles and a Chi-Square or Fisher's Exact test for two-group IC_{50} profiles. Previously published ER protein expression data (Riaz et al., 2013) was used as a categorical variable for ER status.

Author's contributions

KU, JWMM and JAF were involved in writing this article. WJCPvdS and BO performed the cell culture and drug screening. MS, AAJH, KU and JWMM interpreted the data and were involved with the statistical analysis. HWHG, JWMM and JAF were involved with the design of the study, AJ contributed to the design of the study. All authors have read and approved the final version of this manuscript.

Acknowledgments

Analyses were performed using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools Development Team. Janssen Pharmaceutica provided all compounds except for Veliparib and funded the experimental drug screening of those respective drugs. Janssen Pharmaceutica was strongly involved in study design and had minor involvements with the final version of the manuscript. Veliparib was supplied by AbbVie (North Chicago, USA). This study was furthermore financially supported by the Daniel den Hoed Foundation, Rotterdam, The Netherlands (KU, JAF), additional funding came from Cancer Genomics Netherlands - a grant from the Netherlands Organization of Scientific Research (NWO) (JWMM) and further, partial support came from the ERC Advanced Grant #322737 (JAF). These foundations did not exert influence on this study.

Competing interests

Janssen Pharmaceutica (Beerse, Belgium) initially funded this drug screening as contract research with the possibility to use the results later for academic research. HWHG is an employee of Janssen Pharmaceutica and was involved in this study as collaboration partner from Janssen Pharmaceutica. JWMM and JAF received funds from Janssen Pharmaceutica. AJ is involved as local PI in clinical trials on Veliparib and other compounds of AbbVie (North Chicago, USA). The authors KU, WJCPvdS, AAJH, BO, MS declare that they do not have competing interests.

Abbreviations

ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1
c-Met	MET proto-oncogene
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EGFR	epidermal growth factor receptor
ER	estrogen receptor
ERBB2	Erb-b2 receptor tyrosine kinase 2
ERR1	estrogen-related receptor alpha
FBS	fetal bovine serum
FGFR	fibroblast growth factor receptor
HDAC	histone deacetylase

HER2	human epidermal growth factor receptor 2, Erb-b2 receptor tyrosine kinase 2
Her2-/neu	Erb-b2 receptor tyrosine kinase 2
IC_{50}	inhibitory concentration 50, half maximal inhibitory concentration
KEGG	Kyoto Encyclopedia of Genes and Genomes
MDM2	MDM2 proto-oncogene, E3 ubiquitin protein ligase
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
PARP	poly ADP ribose polymerase
PR	progesterone receptor
RET	ret proto-oncogene
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute
SRB	sulforhodamine B
TCA	trichloroacetic acid
TP53	tumor protein p53
TRIS	tris(hydroxymethyl)aminomethane
VEGFR	vascular endothelial growth factor receptor, kinase insert domain receptor
VEGFR2	vascular endothelial growth factor receptor 2, kinase insert domain receptor

Additional files

Additional files can be accessed via https://doi.org/10.1186/s40064-015-1406-8.

Additional file 1. Detailed results of the drug screening and additional information regarding the data analysis. This table contains all calculated IC_{50} values for each drug and cell line. Furthermore the plotted IC_{50} values per drug are displayed and the categorization of the cell lines into resistant and sensitive per drug for the pathway analysis are listed.

Additional file 2. Pathways associated with drugs. This file lists all Biocarta and KEGG pathways that are significantly associated with at least one drug. Pathways that are more significantly associated with a breast cancer subtype are in italic. The p-values given are for the drug-pathway association.

References

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. *100*, 3983–3988.

Armand, J.-P., Burnett, A.K., Drach, J., Harousseau, J.-L., Löwenberg, B., and San Miguel, J. (2007). The emerging role of targeted therapy for hematologic malignancies: update on bortezomib and tipifarnib. The Oncologist *12*, 281–290.

Bast, R.C., Jr., Ravdin, P., Hayes, D.F., Bates, S., Fritsche, H., Jr., Jessup, J.M., Kemeny, N., Locker, G.Y., Mennel, R.G., and Somerfield, M.R. (2001). 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol *19*, 1865–1878.

Becker, J.P., Weiss, J., and Theile, D. (2014). Cisplatin, oxaliplatin, and carboplatin unequally inhibit in vitro mRNA translation. Toxicol. Lett. *225*, 43–47.

Berry, D.A., Cronin, K.A., Plevritis, S.K., Fryback, D.G., Clarke, L., Zelen, M., Mandelblatt, J.S., Yakovlev, A.Y., Habbema, J.D.F., and Feuer, E.J. (2005). Effect of Screening and Adjuvant Therapy on Mortality from Breast Cancer. N. Engl. J. Med. *353*, 1784–1792.

Christman, J.K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene *21*, 5483–5495.

Cohen, R.B. (2003). Epidermal Growth Factor Receptor as a Therapeutic Target in Colorectal Cancer. Clin. Colorectal Cancer *2*, 246–251.

Creusot, F., Acs, G., and Christman, J.K. (1982). Inhibition of DNA methyltransferase and induction of Friend erythroleukemia cell differentiation by 5-azacytidine and 5-aza-2'-deoxycytidine. J Biol Chem *257*, 2041–2048.

Desmedt, C., Voet, T., Sotiriou, C., and Campbell, P.J. (2012). Next-generation sequencing in breast cancer: first take home messages. Curr. Opin. Oncol. *24*, 597–604.

Fornari, F.A., Randolph, J.K., Yalowich, J.C., Ritke, M.K., and Gewirtz, D.A. (1994). Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. Mol Pharmacol *45*, 649–656.

Frederick, C.A., Williams, L.D., Ughetto, G., van der Marel, G.A., van Boom, J.H., Rich, A., and Wang, A.H. (1990). Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. Biochemistry *29*, 2538–2549.

Glendenning, J., and Tutt, A. (2011). PARP inhibitors--current status and the walk towards early breast cancer. Breast Edinb. Scotl. *20 Suppl 3*, S12-19.

Goeman, J.J., van de Geer, S.A., de Kort, F., and van Houwelingen, H.C. (2004). A global test for groups of genes: testing association with a clinical outcome. Bioinformatics *20*, 93–99.

Gonzalez-Angulo, A.M., Morales-Vasquez, F., and Hortobagyi, G.N. (2007). Overview of resistance to systemic therapy in patients with breast cancer. Adv Exp Med Biol *608*, 1–22.

Gramza, A.W., Wells, S.A., Balasubramaniam, S., and Fojo, A.T. (2011). Phase I/II trial of vandetanib and bortezomib in adults with locally advanced or metastatic medullary thyroid cancer: Phase I results. J. Clin. Oncol. *29*, 5565–5565.

Hagiwara, H., and Sunada, Y. (2004). Mechanism of taxane neurotoxicity. Breast Cancer *11*, 82–85.

Hajihassan, Z., and Rabbani-Chadegani, A. (2009). Studies on the binding affinity of anticancer drug mitoxantrone to chromatin, DNA and histone proteins. J. Biomed. Sci. *16*:31.

Heiser, L.M., Sadanandam, A., Kuo, W.L., Benz, S.C., Goldstein, T.C., Ng, S., Gibb, W.J., Wang, N.J., Ziyad, S., Tong, F., et al. (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc. Natl. Acad. Sci. U. S. A. *109*, 2724–2729.

Hollestelle, A., Elstrodt, F., Timmermans, M., Sieuwerts, A.M., Klijn, J.G.M., Foekens, J.A., den Bakker, M.A., and Schutte, M. (2010a). Four human breast cancer cell lines with biallelic inactivating α-catenin gene mutations. Breast Cancer Res. Treat. *122*, 125–133.

Hollestelle, A., Nagel, J.H.A., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J., et al. (2010b). Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res. Treat. *121*, 53–64.

Huang, Y., and Rizzo, R.C. (2012). A water-based mechanism of specificity and resistance for lapatinib with ErbB family kinases. Biochemistry *51*, 2390–2406.

Huynh, H., Ngo, V.C., Fargnoli, J., Ayers, M., Soo, K.C., Koong, H.N., Thng, C.H., Ong, H.S., Chung, A., Chow, P., et al. (2008). Brivanib alaninate, a dual inhibitor of vascular endothelial growth factor receptor and fibroblast growth factor receptor tyrosine kinases, induces growth inhibition in mouse models of human hepatocellular carcinoma. Clin Cancer Res *14*, 6146–6153.

Jansen, M.P.H.M., Foekens, J.A., van Staveren, I.L., Dirkzwager-Kiel, M.M., Ritstier, K., Look, M.P., Meijer-van Gelder, M.E., Sieuwerts, A.M., Portengen, H., Dorssers, L.C.J., et al. (2005). Molecular Classification of Tamoxifen-Resistant Breast Carcinomas by Gene Expression Profiling. J. Clin. Oncol. *23*, 732–740.

Jones, R.J., Gu, D., Bjorklund, C.C., Kuiatse, I., Remaley, A.T., Bashir, T., Vreys, V., and Orlowski, R.Z. (2013). The novel anticancer agent JNJ-26854165 induces cell death through inhibition of cholesterol transport and degradation of ABCA1. J Pharmacol Exp Ther *346*, 381–392.

Kangaspeska, S., Hultsch, S., Edgren, H., Nicorici, D., Murumagi, A., and Kallioniemi, O. (2012). Reanalysis of RNA-sequencing data reveals several additional fusion genes with multiple isoforms. PLoS One *7*, e48745.

Katayama, R., Aoyama, A., Yamori, T., Qi, J., Oh-hara, T., Song, Y., Engelman, J.A., and Fujita, N. (2013). Cytotoxic Activity of Tivantinib (ARQ 197) Is Not Due Solely to c-MET Inhibition. Cancer Res. *73*, 3087–3096.

Keyvanjah, K., DePrimo, S.E., Harmon, C.S., Huang, X., Kern, K.A., and Carley, W. (2012). Soluble KIT correlates with clinical outcome in patients with metastatic breast cancer treated with sunitinib. J. Transl. Med. *10*, 165.

Kittaneh, M., Montero, A.J., and Glück, S. (2013). Molecular Profiling for Breast Cancer: A Comprehensive Review. Biomark Cancer *5*, 61–70.

Kumamoto, K., Spillare, E.A., Fujita, K., Horikawa, I., Yamashita, T., Appella, E., Nagashima, M., Takenoshita, S., Yokota, J., and Harris, C.C. (2008). Nutlin-3a activates p53 to both down-regulate inhibitor of growth 2 and up-regulate mir-34a, mir-34b, and mir-34c expression, and induce senescence. Cancer Res *68*, 3193–3203.

Lehman, J.A., Hauck, P.M., Gendron, J.M., Batuello, C.N., Eitel, J.A., Albig, A., Kadakia, M.P., and Mayo, L.D. (2013). Serdemetan antagonizes the Mdm2-HIF1α axis leading to decreased levels of glycolytic enzymes. PloS One *8*, e74741.

Lemoine, M., and Younes, A. (2010). Histone deacetylase inhibitors in the treatment of lymphoma. Discov Med *10*, 462–470.

Lichtenberg, F.R. (2009). The effect of new cancer drug approvals on the life expectancy of American cancer patients, 1978–2004. Econ. Innov. New Technol. *18*, 407–428.

Lichtenberg, F.R. (2011). Despite steep costs, payments for new cancer drugs make economic sense. Nat Med *17*, 244.

Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. Nat. Rev. Cancer *3*, 330–338.

Lyko, F., and Brown, R. (2005). DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. J. Natl. Cancer Inst. *97*, 1498–1506.

Momparler, R.L., Karon, M., Siegel, S.E., and Avila, F. (1976). Effect of adriamycin on DNA, RNA, and protein synthesis in cell-free systems and intact cells. Cancer Res *36*, 2891–2895.

Nelson, M.H., and Dolder, C.R. (2006). Lapatinib: a novel dual tyrosine kinase inhibitor with activity in solid tumors. Ann Pharmacother *40*, 261–269.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *27*, 29–34.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol *27*, 1160–1167.

Perilongo, G., Maibach, R., Shafford, E., Brugieres, L., Brock, P., Morland, B., de Camargo, B., Zsiros, J., Roebuck, D., Zimmermann, A., et al. (2009). Cisplatin versus cisplatin plus doxorubicin for standard-risk hepatoblastoma. N. Engl. J. Med. *361*, 1662–1670.

Ponti, D., Zaffaroni, N., Capelli, C., and Daidone, M.G. (2006). Breast cancer stem cells: an overview. Eur. J. Cancer Oxf. Engl. 1990 *42*, 1219–1224.

Previati, M., Manfrini, M., Galasso, M., Zerbinati, C., Palatini, J., Gasparini, P., and Volinia, S. (2013). Next generation analysis of breast cancer genomes for precision medicine. Cancer Lett. *339*, 1–7.

Radovich, M., Clare, S.E., Atale, R., Pardo, I., Hancock, B.A., Solzak, J.P., Kassem, N., Mathieson, T., Storniolo, A.M., Rufenbarger, C., et al. (2013). Characterizing the heterogeneity of triple-negative breast cancers using microdissected normal ductal epithelium and RNA-sequencing. Breast Cancer Res Treat *143(1)*, 57-68.

R_Core_Team (2013). R: A Language and Environment for Statistical Computing (Vienna, Austria).

Riaz, M., van Jaarsveld, M.T., Hollestelle, A., Prager-van der Smissen, W.J., Heine, A.A., Boersma, A.W., Liu, J., Helmijr, J., Ozturk, B., Smid, M., et al. (2013). MicroRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. Breast Cancer Res *15*, R33.

Saldanha, A.J. (2004). Java Treeview-extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Sathornsumetee, S., and Rich, J.N. (2006). Vandetanib, a novel multitargeted kinase inhibitor, in cancer therapy. Drugs Today Barc *42*, 657–670.

Scagliotti, G.V., Novello, S., and von Pawel, J. (2013). The emerging role of MET/HGF inhibitors in oncology. Cancer Treat. Rev. *39*, 793–801.

Shangary, S., and Wang, S. (2009). Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol *49*, 223–241.

Sieuwerts, A.M., Kraan, J., Bolt, J., Spoel, P. van der, Elstrodt, F., Schutte, M., Martens, J.W.M., Gratama, J.-W., Sleijfer, S., and Foekens, J.A. (2009). Anti-Epithelial Cell Adhesion Molecule Antibodies and the Detection of Circulating Normal-Like Breast Tumor Cells. J. Natl. Cancer Inst. *101*, 61–66.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U A *98*, 10869–10874.

Stordal, B., and Davey, M. (2007). Understanding cisplatin resistance using cellular models. IUBMB Life *59*, 696–699.

Stordal, B., Pavlakis, N., and Davey, R. (2007). A systematic review of platinum and taxane resistance from bench to clinic: an inverse relationship. Cancer Treat. Rev. *33*, 688–703.

Teicher, B.A., Ara, G., Herbst, R., Palombella, V.J., and Adams, J. (1999). The proteasome inhibitor PS-341 in cancer therapy. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *5*, 2638–2645.

The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

van ’t Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Voigt, W. (2005). Sulforhodamine B assay and chemosensitivity. Methods Mol Med *110*, 39–48.

Wang, Y., Klijn, J.G.M., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet Lond. Engl. *365*, 671–679.

Chapter 4

MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines

Katharina Uhr, Wendy J. C. Prager-van der Smissen, Anouk A. J. Heine, Bahar Ozturk, Marijn T. M. van Jaarsveld, Antonius W. M. Boersma, Agnes Jager, Erik A. C. Wiemer, Marcel Smid, John A. Foekens, John W. M. Martens

PLoS One 2019;14(5):e0216400

Abstract

MicroRNAs (miRNAs) regulate gene expression post-transcriptionally. In this way they might influence whether a cell is sensitive or resistant to a certain drug. So far, only a limited number of relatively small scale studies comprising few cell lines and/or drugs have been performed. To obtain a broader view on miRNAs and their association with drug response, we investigated the expression levels of 411 miRNAs in relation to drug sensitivity in 36 breast cancer cell lines. For this purpose IC_{50} values of a drug screen involving 34 drugs were associated with miRNA expression data of the same breast cancer cell lines. Since molecular subtype of the breast cancer cell lines is considered a confounding factor in drug association studies, multivariate analysis taking subtype into account was performed on significant miRNA-drug associations which retained 13 associations. These associations consisted of 11 different miRNAs and eight different drugs (among which Paclitaxel, Docetaxel and Veliparib). The taxanes, Paclitaxel and Docetaxel, were the only drugs having miRNAs in common: *hsa-miR-187-5p* and *hsa-miR-106a-3p* indicative of drug resistance while Paclitaxel sensitivity alone associated with *hsa-miR-556-5p*. Tivantinib was associated with *hsa-let-7d-5p* and *hsa-miR-18a-5p* for sensitivity and *hsa-miR-637* for resistance. Drug sensitivity was associated with *hsa-let-7a-5p* for Bortezomib, *hsa-miR-135a-3p* for JNJ-707 and *hsa-miR-185-3p* for Panobinostat. Drug resistance was associated with *hsa-miR-182-5p* for Veliparib and *hsa-miR-629-5p* for Tipifarnib. Pathway analysis for significant miRNAs was performed to reveal biological roles, aiding to find a potential mechanistic link for the observed associations with drug response. By doing so *hsa-miR-187-5p* was linked to the cell cycle G2-M checkpoint in line with this checkpoint being the target of taxanes. In conclusion, our study shows that miRNAs could potentially serve as biomarkers for intrinsic drug resistance and that pathway analyses can provide additional information in this context.

Introduction

Biomarkers of drug sensitivity/resistance are of great interest for the clinic as they allow for patient stratification – thereby identifying the most effective therapy for patients faster, reducing overtreatment and toxicity burden as well as saving costs. In the search for highly suited biomarkers irrespective of research field or application, microRNAs (miRNAs) have become increasingly popular due to their stability and broad applicability, underlining their potential as biomarkers (Cortez et al., 2011; Kim, 2015).

MiRNAs are small oligonucleotides involved in multiple processes such as aging, tissue development and also cancer (Hammond, 2015; Lovat et al., 2011; Wiemer, 2007). They bind in a sequence-dependent fashion to mRNA 3' UTRs leading to inhibition of messenger RNA (mRNA) translation, endonucleolytic cleavage of the mRNA or mRNA destabilization. Through affecting mRNA translation and stability, miRNAs ultimately regulate protein expression (Guo et al., 2010; Lin and Gregory, 2015). As miRNA-recognition sequences are present in many genes, one miRNA can impact up to several hundred transcripts (Hausser and Zavolan, 2014; Huntzinger and Izaurralde, 2011; Iorio and Croce, 2012; Krol et al., 2010; Lin and Gregory, 2015; Varela et al., 2013).

Until now, several studies have been conducted to link miRNAs to treatment outcome (Fanini and Fabbri, 2016; Li and Yang, 2013). In these earlier studies, mainly drug-resistant cell lines were generated, which were then compared to their parental cell lines to reveal differentially expressed miRNAs, which were subsequently studied in functional experiments. While these experiments are instrumental for a mechanistic understanding of miRNAs in drug sensitivity/resistance, such experiments do not necessarily identify predictors of drug resistance or sensitivity. Furthermore, most of these studies involved only one drug and were performed with a limited number of cell lines. As breast tumors can be classified into distinct subtypes with different clinical outcome (Blows et al., 2010; Sørlie et al., 2003) this should also be reflected in study design by including larger series of cell lines, together better resembling the clinical variability. Furthermore, screening for drug resistance against many drugs may not only identify miRNAs associated with one compound but may also identify miRNAs linked to sensitivity/resistance of more than one drug.

Therefore, we analyzed 36 well-characterized breast cancer cell lines that we screened for sensitivity to 34 compounds to obtain more clinically meaningful results on miRNAs and their potential as biomarkers for drug response in breast cancer. Our study showed that this approach was feasible and led to the identification of several miRNA-drug associations not found earlier.

Materials and methods

Breast cancer cell lines

Cell lines were cultured and profiled for correct identity as described previously (Riaz et al., 2013). In summary: Cell lines were cultured for RNA isolation until 70-80% confluence on collagen-coated petri dishes in triplicate. RPMI 1640 medium with

10% heat-inactivated fetal bovine serum and antibiotic agents (80 µg/ml (0.08 kg/m^3) Streptomycin and 100 µg/ml (0.1 kg/m^3) Penicillin G) was used as medium.

To ensure cell line identity, DNA of the 36 breast cancer cell lines was isolated using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) and subsequently subjected to short-tandem repeat (STR) profiling using the PowerPlex16 system (Promega, Madison, WI, USA) following manufacturer's protocol. A 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) was used for detection and Genemarker 1.91 software from Softgenetics (State College, PA, USA) was employed for analysis. STR profiles were compared with those deposited at the American Type Culture Collection (Manassas, VA, USA) and the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany). For SUM cell lines, STR profiles were not available, so in-house profiles from the earliest passages were compared with those generated later. Cell lines had been obtained between 1997 and 2006 from the following sources:

American Type Culture Collection (BT-20 (Lasfargues and Ozzello, 1958), BT-474 (Lasfargues et al., 1978), BT-549 (Littlewood-Evans et al., 1997), CAMA-1 (Fogh et al., 1977), HCC1143 (Gazdar et al., 1998), HCC1395 (Gazdar et al., 1998), HCC1569 (Ahmadian et al., 1997), HCC1806 (Ahmadian et al., 1997), HCC-1937 (Tomlinson et al., 1998), HCC1954 (Gazdar et al., 1998), HCC202 (Gazdar et al., 1998), HCC38 (Sundaresan et al., 1998), HCC70 (Gazdar et al., 1998), Hs578T (Hackett et al., 1977), MCF-7 (Soule et al., 1973), MDA-MB-175VII (Cailleau et al., 1974), MDA-MB-231 (Cailleau et al., 1974), MDA-MB-361 (Cailleau et al., 1974), MDA-MB-415 (Cailleau et al., 1978), MDA-MB-436 (Cailleau et al., 1978), MDA-MB-468 (Cailleau et al., 1978), SK-BR-3 (Trempe, 1976), T47D (Keydar et al., 1979), UACC812 (Meltzer et al., 1991)),

Dr. N. de Vleesschouwer (Institut Jules Bordet, Brussels, Belgium) (EVSA-T (Lippman et al., 1976)),

Dr. H.S. Smith (California Pacific Medical Center, San Francisco, CA, USA) (MPE-600 (Smith et al., 1987)),

Dr. E. Stockert (Sloan-Kettering Institute for Cancer Research, New York, NY, USA) (SK-BR-7 (Davidson et al., 2000)),

Ethier laboratory (BioIVT, West Sussex, UK) (SUM1315M02 (Forozan et al., 1999), SUM149PT (Forozan et al., 1999), SUM159PT (Forozan et al., 1999), SUM185PE (Forozan

et al., 1999), SUM190PT (Forozan et al., 1999), SUM229PE (Forozan et al., 1999), SUM44PE (Ethier et al., 1993), SUM52PE (Ethier et al., 1996)),

Riken Gene Bank (Tsukuba, Japan) (OCUB-M (Sawada et al., 1994))

Drug screening

Drug screening data was previously published for an extended cell line panel and the detailed description of drug-drug relations as well as the entire drug response dataset is available there (Uhr et al., 2015). In brief: The breast cancer cell lines were cultured and plated in triplicate in 96-well collagen-coated plates, and then incubated for 96 hours with 12 serial dilutions of 37 drugs or vehicle per cell line. The drugs were supplied by Janssen Pharmaceutica (Beerse, Belgium) except for Veliparib, which was supplied by Abbott Laboratories (North Chicago, IL, USA). After 96 hours the Sulforhodamine B Assay was used to quantify total protein as a measure of cell number (Voigt, 2005) and IC_{50} values were computed subsequently (Uhr et al., 2015). Three drugs did not exhibit differential responses among the cell lines and were excluded from the final dataset (Uhr et al., 2015). Therefore, the present study included 34 drugs which were used for further analyses.

RNA isolation

RNA isolation was performed as described earlier (Riaz et al., 2013). To summarize: RNAzol-B reagent (Campro Scientific BV, Veenendaal, the Netherlands) was used for isolation of total RNA from all cell lines according to manufacturer's protocol. Quality of isolated RNA was verified using spectrophotometric assessments of the $A_{260nm}/A_{280\ nm}$ and the A_{260nm}/A_{230nm} ratios, with the first required to have a value of approximately 2 and the second to have a value of 2 or higher. For these measurements a NanoDrop ND-1000 (Isogen Life Science, De Meern, the Netherlands) was used. Additional quality checks were performed as described earlier (Sieuwerts et al., 2005).

Gene expression profiling

Gene expression profiling was performed as previously described (Riaz et al., 2013). Reverse transcription was performed on 200 ng (2E-10 kg) of total RNA, followed by cDNA synthesis and the generation of biotin-labeled cRNA using the 3' IVT express kit (Affymetrix, Santa Clara, CA, USA). Fragmentation of the labeled cRNA ensued and finally the fragmented and labeled cRNA was loaded on the Affymetrix GeneTitan. The hybridization mixture was loaded on Affymetrix Human_Genome_HT_HG-U133_Plus_PM GeneChip 96-well arrays. The subsequent steps (hybridization, washing and scanning) were performed within the GeneTitan. In the Affymetrix Expression Console

the generated ".CEL" files were subjected to normalization employing the default settings of the RMA method. The generated data have been uploaded to the Gene Expression Omnibus data repository under the following access code: [GEO:GSE41313].

Breast cancer subtypes were determined using hierarchical clustering. Clusters were obtained with average distance linkage hierarchical clustering with non-centered correlation as distance metric. The three resulting clusters were matched to previously established intrinsic subtypes (Hollestelle et al., 2010) of these cell lines (which were based on a different chip-type) and the clusters were labeled as 'basal', 'luminal' and 'normal-like'.

miRNA expression profiling

The procedure employed for miRNA expression profiling has been described previously in detail (Pothof et al., 2009; Riaz et al., 2013). MiRNA expression data of cell lines with confirmed identity were used. To generate the data set, one microgram (1E-09 kg) of total RNA was labeled with the dye Cy3 using the ULS aRNA labeling kit (Kreatech, Amsterdam, the Netherlands). Only RNA samples with a labeling efficiency higher than 15 pmol Cy3/µg RNA (1.5E-20 mol/kg) were used for hybridization. One sample was used per cell line. Hybridization was performed overnight. Normalization of the data was performed as previously described (Pothof et al., 2009). Labeled total RNA isolated from all cell lines was hybridized to home-made microarrays containing LNA modified capture probes from Exiqon (Vedbaek, Denmark) for the miRNA capture. This miRNA microarray design was based on miRBase version 10.0 (annotation version 13), contained 1344 probes and had the capacity to measure the expression levels of 725 human miRNAs.

Within this dataset two miRNAs (*hsa-miR-185* and *hsa-miR-620*) were represented twice on the microarray with differential probes. We removed the measurements with the older probes (*hsa-miR-185* probe ID 5560 and *hsa-miR-620* probe ID 32825) and kept the measurements of the newer probes (miRNA *hsa-miR-185* probe ID 42902 and miRNA *hsa-miR-620* probe ID 42994). We assessed whether the probe version influenced results, however, this was not the case.

The removal of the older probes resulted in a final number of 411 miRNAs expressed above background. MiRNA expression data has been published earlier (Riaz et al., 2013), except for the cell line OCUB-M, the miRNA expression data for this cell line can be found in **S1 Table**. The significant miRNAs are listed with the Exiqon oligonucleotide ID and the different miRBase (Griffiths-Jones et al., 2006) annotations, as well the corresponding MIMAT identifier in **S2 Table**.

Association of miRNA expression with IC_{50} values and co-expression analysis

By inspecting the results of the drug screening it became evident that the IC_{50} response curves for different drugs were very different across cell lines (see also Uhr *et al.* (Uhr et al., 2015)). While for some drugs most cell lines had different IC_{50} values, other drugs had a large number of cell lines with the same IC_{50} value up to almost all cell lines having identical IC_{50} values. Due to this it became clear that different analyses would be required to account for the different IC_{50} drug profiles. For most drugs, several cell lines had an IC_{50} value which was at the maximum of the tested drug concentration. Depending on the number of cell lines with values at maximum drug level, we chose to either use Spearman correlation (≤ 5 maximum IC_{50} values), a 2-step analysis (>5 IC_{50} values at maximum & ≥10 IC_{50} values not at the maximum) or a Mann-Whitney test (<10 IC_{50} values which are not at the maximum). MiRNA data were log10 – transformed for analyses, drug data were transformed to negative log 10 values.

The first step of the 2-step analysis consisted of Spearman correlation of those IC_{50} values that were not at the maximum drug concentration ("variable IC_{50} values") with miRNA expression values. Identified miRNA-drug associations were then tested in the second step, a Mann-Whitney test, classifying all IC_{50} values at maximum drug concentration as resistant and the remaining variable IC_{50} values as sensitive.

Over the results of each analysis group (Spearman correlation, 2-step analysis, Mann-Whitney test), q-values (Storey and Tibshirani, 2003) were calculated to account for multiple testing. A q-value of 0.3 or lower was determined significant. The analysis type used per drug is listed in **S3 Table**. Analyses were performed in R (versions 3.4.4 up to 3.5.1) (R Core Team, 2018) using the psych package (Revelle, 2018) and corr.test function for Spearman correlations, setting alpha at 0.05. The R function wilcox.test was used for Mann-Whitney testing. Q-values were calculated using the q value Bioconductor package (Dabney and Storey) in R.

Fig. 1 was created in Inkscape 0.92 (Free Software Foundation Inc., Boston, USA).

Multivariate analysis

MiRNAs significantly associated with drug response were evaluated in a multivariate analysis, including gene-expression-based breast cancer subtype as a co-variate. Linear regression was used, as only drugs with a linear IC_{50} profile had shown significant associations with miRNA expression. Subtypes were used as a binary variable: basal vs rest, luminal vs rest and normal-like vs rest. The IC_{50} values were used as the dependent variable, while the miRNA levels as well as one of the binary subtype groups were used

as independent variables. This regression model was run for each of the three binary subtype-groups separately and was calculated using Stata - version 13 (StataCorp, College Station, TX, USA).

Pathway analyses

Pathway analyses were performed on the mRNA data described above. The 11 significant miRNAs that were not related to subtype were analyzed by grouping the cell lines into top and bottom 25% based on the expression of the respective miRNA. This led to 9 cell lines per group. The mRNA expression of top and bottom cell lines was analyzed for differences in R (R Core Team, 2018) using the Global Test package (Goeman et al., 2004) and Biocarta (San Diego, CA, USA) and KEGG (Ogata et al., 1999) pathway information. P-values were subjected to multiple testing correction (method "BH", also known as "FDR") and a permutation test (1000 permutations) of the non-adjusted p-value was used as an additional selection criterion. Associations with BH-adjusted p-values below 0.1 and permutation p-values below 0.05 were deemed significant.

Results

Association of miRNAs with drug sensitivity

In the analysis we combined the data of 36 cell lines for which 34 drugs gave differential responses among the included cell lines (Uhr et al., 2015) and the expression data of 411 miRNAs (Riaz et al., 2013).

To associate miRNA expression with drug response, we first inspected the IC_{50} profiles of all 34 compounds among the breast cancer cell lines. Based on the number of cell lines with a maximum IC_{50} value we chose one of three different analysis approaches as defined in Materials and Methods.

Only the drugs with "linear" IC_{50} values showed significant associations with miRNA expression upon q-value correction, totaling to 27 miRNA-drug associations (see **Table 1 and Fig 1**).

Table 1. MiRNAs associated with drug response

Drug	MiRNA	Association type	R	p-value	q-value
Tivantinib	*hsa-let-7d-5p*	Sensitivity	0.57	3.96E-04	0.18
Tivantinib	*hsa-miR-18a-5p*	Sensitivity	0.57	3.88E-04	0.18
Tivantinib	*hsa-miR-637*	Resistance	-0.55	6.51E-04	0.24
Bortezomib	*hsa-let-7a-5p*	Sensitivity	0.57	2.94E-04	0.18
Docetaxel	*hsa-miR-187-5p*	Resistance	-0.65	1.89E-05	0.07
Docetaxel	*hsa-miR-586*	Resistance	-0.65	1.99E-05	0.07
Docetaxel	*hsa-miR-106a-3p*	Resistance	-0.58	2.26E-04	0.18
Docetaxel	*hsa-miR-338-3p*	Resistance	-0.58	1.85E-04	0.18
JNJ-707	*hsa-miR-135a-3p*	Sensitivity	0.55	5.15E-04	0.21
Mitoxantrone	*hsa-miR-497-3p*	Sensitivity	0.56	3.83E-04	0.18
Mitoxantrone	*hsa-miR-640*	Resistance	-0.58	1.95E-04	0.18
Mitoxantrone	*hsa-miR-101-3p*	Resistance	-0.53	8.85E-04	0.27
Mitoxantrone	*hsa-miR-522-3p*	Resistance	-0.53	9.67E-04	0.27
Mitoxantrone	*hsa-miR-125b-1-3p*	Sensitivity	0.52	1.07E-03	0.29
Paclitaxel	*hsa-miR-106a-3p*	Resistance	-0.56	4.17E-04	0.18
Paclitaxel	*hsa-miR-187-5p*	Resistance	-0.57	2.64E-04	0.18
Paclitaxel	*hsa-miR-522-3p*	Resistance	-0.56	3.44E-04	0.18
Paclitaxel	*hsa-miR-586*	Resistance	-0.56	3.40E-04	0.18
Paclitaxel	*hsa-miR-556-5p*	Sensitivity	0.54	7.65E-04	0.26
Panobinostat	*hsa-miR-185-3p*	Sensitivity	0.54	6.47E-04	0.24
Quisinostat	*hsa-miR-27a-3p*	Resistance	-0.53	9.06E-04	0.27
Serdemetan	*hsa-miR-182-5p*	Resistance	-0.58	2.08E-04	0.18
Sirolimus	*hsa-miR-23a-3p*	Resistance	-0.57	2.46E-04	0.18
Sunitinib	*hsa-miR-589-5p*	Sensitivity	0.54	7.35E-04	0.26
Tipifarnib	*hsa-miR-586*	Resistance	-0.57	3.54E-04	0.18
Tipifarnib	*hsa-miR-629-5p*	Resistance	-0.54	9.01E-04	0.27
Veliparib	*hsa-miR-182-5p*	Resistance	-0.63	4.25E-05	0.1

MiRNAs significantly associated with drug response. Associations were computed using Spearman's correlation and the correlation coefficient (R) is given, next to the unadjusted p-value and the q-value.

Fig 1. Overview of miRNAs associated with different drugs. The correlation coefficient is given on the y-axis. Associations with drug resistance are given in orange; associations with drug sensitivity are given in blue. Bor = Bortezomib, JNJ = JNJ-707, Mit = Mitoxantrone, Pan = Panobinostat, Qui = Quisinostat, Ser = Serdemetan, Sir = Sirolimus, Sun = Sunitinib, Tip = Tipifarnib, Vel = Veliparib.

Breast cancer subtype as potential confounder of miRNA drug associations

In a previous study (Uhr et al., 2015) we had found that certain breast cancer subtypes can respond differently to some drugs. As expression of certain miRNAs might be related to a specific subtype, we assessed our significant findings in a multivariate analysis. In this analysis we modeled the association between drug response and miRNA expression and subtype. All 27 associations were tested with subtype and the respective results are listed in **S4 Table**. Fourteen miRNA-drug associations were linked to one or more breast cancer subtypes. Basal subtype was associated stronger with the respective miRNA than the associated drug for the following associations: *hsa-miR-586* and Paclitaxel, *hsa-miR-589-5p* and Sunitinib and *hsa-miR-586* and Tipifarnib. Luminal subtype affected the association between *hsa-miR-27a-3p* and Quisinostat as well as *hsa-miR-23a-3p* and Sirolimus. Normal-like subtype confounded the associations between *hsa-miR-101-3p* and Mitoxantrone, *hsa-miR-125b-1-3p* and Mitoxantrone, *hsa-miR-497-3p* and Mitoxantrone, *hsa-miR-640* and Mitoxantrone and *hsa-miR-522-3p* and Paclitaxel. Several associations were related to more than one subtype: These were the subtypes basal and luminal for the two Docetaxel-associated miRNAs *hsa-miR-338-3p* and *hsa-miR-586*. Furthermore, the subtypes luminal and normal-like showed an

effect on the association between *hsa-miR-522-3p* and Mitoxantrone and the subtypes basal and normal-like were implicated in the association between *hsa-miR-182-5p* and Serdemetan.

Molecular subtype did not affect the 13 remaining miRNA-drug associations (8 drugs, 11 miRNAs), which are highlighted in more detail below.

Sensitivity to the c-Met inhibitor Tivantinib (Munshi et al., 2010) was associated with the expression of two miRNAs, i.e. *hsa-let-7d-5p* and *hsa-miR-18a-5p*, while drug resistance was associated with *hsa-miR-637.* Drug sensitivity to the proteasome inhibitor Bortezomib (Lightcap et al., 2000) was associated with *hsa-let-7a-5p* expression. Among the microtubule-targeting taxanes Docetaxel and Paclitaxel (de Weger et al., 2014), we found that *hsa-miR-106a-3p* and *hsa-miR-187-5p* expression were positively correlated with drug resistance for both drugs. *Hsa-miR-556-5p* expression, however, was solely associated with Paclitaxel sensitivity. *Hsa-miR-135a-3p* expression was related to sensitivity to JNJ-707, a drug targeting FGF receptors (Uhr et al., 2015). Panobinostat, one of the HDAC inhibitors (Lemoine and Younes, 2010) in our screening, showed a positive correlation with *hsa-miR-185-3p* expression for drug sensitivity. For Tipifarnib, a farnesyltransferase inhibitor targeting RAS signal transduction (Norman, 2002), expression of *hsa-miR-629-5p* was associated with resistance to the drug, while drug resistance to the PARP inhibitor Veliparib (Glendenning and Tutt, 2011) was associated with *hsa-miR-182-5p* expression.

Besides the confounding effect of the breast cancer subtypes on some miRNA-drug associations, we also noted that several drugs had more than one associated miRNA, which led to our next analysis whether miRNAs are co-expressed.

Co-expression among miRNAs associated with the same drug

To study whether miRNAs that significantly associated with the same drug in univariate analysis showed a correlated expression pattern, we performed a correlation analysis. Below we discuss Tivantinib, Docetaxel and Paclitaxel, as each of these three drugs had several associated miRNAs unrelated to one of the breast cancer subtypes (see **S4 Table**). Correlation coefficients and p-values are given in the **S5-S7 Tables**. In the case of Tivantinib the miRNAs *hsa-miR-18a-5p* and *hsa-let-7d-5p* showed a mild positive correlation with each other (R = 0.28), which was however not significant, while *hsa-miR-637* showed a significant negative association with the aforementioned miRNAs (R = -0.49 and -0.46, respectively). The two Docetaxel- and Paclitaxel-associated miRNAs *hsa-miR-187-5p* and *hsa-miR-106a-3p* showed an intermediate significant positive

correlation with each other (R = 0.40), while the other Paclitaxel-associated miRNA *hsa-miR-556-5p* was significantly negatively associated with *hsa-miR-187-5p* and *hsa-miR-106a-3p* (R = -0.50 and -0.44, respectively).

Pathway analyses on significant miRNAs

In the next step, we performed pathway analyses to learn more about the biology of the miRNAs associated with drug response in the breast cancer cell lines. For this purpose we used those 11 miRNAs which were significantly associated with drug response independent of subtypes (see **Table 1** and **S4 Table**). Messenger RNA expression data from the cell lines with the most differential expression of each miRNA (high versus low expression were compared; 9 per subgroup) were used as input for the pathway analysis. Pathways which were significantly different between samples with high or low expression of a tested miRNA are shown in **Table 2**.

Table 2. Pathways associated with miRNAs

MiRNA	Pathway	Permutation p-value	BH-adjusted p-value	Associated drugs
hsa-miR-106a-3p	TPO Signaling Pathway (Biocarta)	1.00E-03	7.48E-02	Docetaxel, Paclitaxel
	Nerve growth factor pathway NGF (Biocarta)	1.00E-03	7.48E-02	
	T Cell Receptor Signaling Pathway (Biocarta)	0.00E+00	9.74E-02	
	Insulin Signaling Pathway (Biocarta)	2.00E-03	9.74E-02	
	Signaling Pathway from GProtein Families (Biocarta)	3.00E-03	9.74E-02	
	Cadmium induces DNA synthesis and proliferation in macrophages (Biocarta)	3.00E-03	9.74E-02	
	EPO Signaling Pathway (Biocarta)	6.00E-03	9.74E-02	
hsa-miR-135a-3p	Pantothenate and CoA biosynthesis (KEGG)	0.00E+00	5.39E-02	JNJ-707

Table 2. Continued

MiRNA	Pathway	Permutation p-value	BH-adjusted p-value	Associated drugs
hsa-miR-182-5p	Intrinsic Prothrombin Activation Pathway (Biocarta)	4.00E-03	5.02E-02	Veliparib
	Role of Tob in Tcell activation (Biocarta)	8.00E-03	5.02E-02	
	Acute Myocardial Infarction (Biocarta)	1.20E-02	5.02E-02	
	Actions of Nitric Oxide in the Heart (Biocarta)	1.20E-02	5.02E-02	
	Erythrocyte Differentiation Pathway (Biocarta)	1.30E-02	5.02E-02	
	Signal transduction through IL1R (Biocarta)	1.70E-02	5.02E-02	
	ALK in cardiac myocytes (Biocarta)	1.70E-02	5.02E-02	
	Cytokines and Inflammatory Response (Biocarta)	1.90E-02	5.02E-02	
	Integrin Signaling Pathway (Biocarta)	2.30E-02	5.02E-02	
	Estrogenresponsive protein Efp controls cell cycle and breast tumors growth (Biocarta)	3.00E-02	5.02E-02	
	Selective expression of chemokine receptors during Tcell polarization (Biocarta)	3.30E-02	5.02E-02	
	Cell Cycle G1 S Check Point (Biocarta)	3.30E-02	5.02E-02	
	Cytokine Network (Biocarta)	3.30E-02	5.58E-02	
	Phosphoinositides and their downstream targets (Biocarta)	4.10E-02	5.02E-02	
	Role of ERBB2 in Signal Transduction and Oncology (Biocarta)	4.30E-02	5.02E-02	
	Hypoxia and p53 in the Cardiovascular system (Biocarta)	4.40E-02	5.36E-02	
	Trefoil Factors Initiate Mucosal Healing (Biocarta)	4.70E-02	6.71E-02	
	TGFbeta signaling pathway (KEGG)	8.00E-03	8.41E-02	
	Focal adhesion (KEGG)	1.30E-02	8.41E-02	

Table 2. Continued

MiRNA	Pathway	Permutation p-value	BH-adjusted p-value	Associated drugs
hsa-miR-187-5p	Selective expression of chemokine receptors during Tcell polarization (Biocarta)	3.00E-03	7.70E-02	Docetaxel, Paclitaxel
	Trefoil Factors Initiate Mucosal Healing (Biocarta)	3.00E-03	7.70E-02	
	CXCR4 Signaling Pathway (Biocarta)	1.60E-02	7.70E-02	
	Regulation of eIF4e and p70 S6 Kinase (Biocarta)	1.80E-02	7.70E-02	
	Skeletal muscle hypertrophy is regulated via AKT mTOR pathway (Biocarta)	2.00E-02	7.70E-02	
	CARM1 and Regulation of the Estrogen Receptor (Biocarta)	2.10E-02	7.70E-02	
	Neuropeptides VIP and PACAP inhibit the apoptosis of activated T cells (Biocarta)	2.30E-02	7.70E-02	
	Estrogenresponsive protein Efp controls cell cycle and breast tumors growth (Biocarta)	2.80E-02	7.70E-02	
	Role of BRCA1 BRCA2 and ATR in Cancer Susceptibility (Biocarta)	2.80E-02	7.70E-02	
	Role of ERBB2 in Signal Transduction and Oncology (Biocarta)	3.30E-02	7.70E-02	
	Role of Tob in Tcell activation (Biocarta)	3.30E-02	8.44E-02	
	Cell Cycle G2 M Checkpoint (Biocarta)	3.50E-02	7.70E-02	
	mTOR Signaling Pathway (Biocarta)	3.90E-02	7.70E-02	
	Effects of calcineurin in Keratinocyte Differentiation (Biocarta)	4.20E-02	7.70E-02	
	Nitrogen metabolism (KEGG)	3.00E-03	2.56E-02	
	AminoacyltRNA biosynthesis (KEGG)	3.00E-03	2.77E-02	
	Valine leucine and isoleucine degradation (KEGG)	1.90E-02	3.35E-02	
	Pentose phosphate pathway (KEGG)	6.00E-03	4.87E-02	
	Wnt signaling pathway (KEGG)	4.80E-02	8.28E-02	

Table 2. Continued

MiRNA	Pathway	Permutation p-value	BH-adjusted p-value	Associated drugs
hsa-miR-556-5p	Signaling Pathway from GProtein Families (Biocarta)	1.00E-03	4.86E-02	Paclitaxel
hsa-miR-637	Links between Pyk2 and Map Kinases (Biocarta)	1.00E-03	5.46E-02	Tivantinib

Pathways significantly associated with drug-associated miRNAs. Selection criteria for significant pathways were a permutation p-value below 0.05 and a Benjamini-Hochberg adjusted p-value below 0.1. The respective associated drugs are listed and highlighted in orange for an association of the miRNA with resistance to the respective drug and blue for an association with sensitivity.

For 6 of the 11 miRNAs we found significantly associated pathways. The miRNAs *hsa-miR-135a-3p, hsa-miR-556-5p and hsa-miR-637* each had only 1 pathway associated, while *hsa-miR-106a-3p* had 7 pathways associated and *hsa-miR-182-5p* and *hsa-miR-187-5p* each had 19 pathways significantly associated. For the 3 miRNAs associated with multiple pathways, these pathways typically included the same genes such as in the case of *hsa-miR-106a-3p* with the genes *JUN* and *FOS* upregulated in all significant pathways and *PIK3R1* being upregulated in 3 of the 7 pathways. In the case of *hsa-miR-182-5p* many of the associated pathways had genes of the COL4A family and the TGF-β family downregulated, while for *hsa-miR-187-5p* several of the pathways contained the genes *CXCR4*, *ESR1*, *PDK2*, *PTEN*; as well as the ER-regulated genes *TFF1* (Carroll et al., 2005) and *TFF3* (Tozlu et al., 2006).

Discussion

Our exploratory investigation showed that the expression of specific miRNAs can indicate drug sensitivity or resistance in breast cancer cell lines. In total, we were able to identify 21 different miRNAs associated with response to 13 different drugs, totaling up to 27 miRNA-drug associations. After excluding miRNAs that also associate with breast cancer subtype in multivariate analyses, 11 miRNAs involved in 13 miRNA-drug associations remained. Additional pathway analyses using available gene expression data gave insights into the associated biology for 6 of the significant miRNAs. To assess the applicability of these miRNAs as clinical biomarkers, further validation in independent sample series is needed to correct for possible additional parameters influencing drug response which were not considered here. Separately, the causal relation between the identified miRNA-drug associations in this hypothesis-generating study could be explored.

To try to understand the subtype independent miRNA-drug associations we discuss the possible mechanistic links for the most significant results based on current state-of-the-art knowledge.

The strongest miRNA-drug interaction was the association of *hsa-miR-187-5p* with Docetaxel resistance. This miRNA was additionally associated with Paclitaxel resistance. As taxanes have been known for a long time for their anti-tumorigenic activity and have been studied extensively (Band Horwitz, 1992), several resistance mechanisms are known: One of the main contributors to taxane resistance is the overexpression of multidrug transporters (e.g. P-gp; HGNC symbol: *ABCB1*) (Galletti et al., 2007). Besides, tubulin modifications (e.g. mutations) and alterations in the tubulin-microtubule system have been found such as altered expression of tubulin isoforms or associated proteins (e.g. MAP4 and Stathmin) (Galletti et al., 2007). While we did not find a clear link between *hsa-miR-187-5p* and one of these well-known resistance mechanisms, we found that this miRNA is associated with the pathway "Cell cycle G2-M checkpoint" and the "MTOR signaling" pathway (see **Table 2**). This is an interesting observation, as Paclitaxel is known to block cells between the G2 and the mitotic phase of the cell cycle (Band Horwitz, 1992). Furthermore, genes involved in the G2-M checkpoint and MTOR signaling were differentially expressed in residual breast tumors after Docetaxel treatment when compared to pre-treatment biopsies (Chang et al., 2005). In our study the genes in the G2-M checkpoint pathway are mainly downregulated in cell lines with high *hsa-miR-187-5p* expression, arguing that this miRNA might target several of the involved genes or an upstream regulator and in this way potentially increases cell cycle progression. Alternatively, *hsa-miR-187-5p* could be co-regulated with the genes in this pathway and as a result might affect events downstream.

The miRNA *hsa-miR-106a-3p* was also associated with Docetaxel and Paclitaxel resistance. In the pathway analysis, 7 pathways were significantly associated with this miRNA. Looking further into similarities between the pathways, it became clear that the genes *JUN* and *FOS* were part of all pathways and were clearly upregulated. Unfortunately only little is known about this miRNA and no link could be found between either the genes *JUN* and *FOS* or this miRNA with resistance to the two taxanes Docetaxel and Paclitaxel. As the expression of this miRNA is correlated to the expression of *hsa-miR-187-5p* in our study, this miRNA could act in concert with *hsa-miR-187-5p* or its associated pathways but based on current knowledge the role of this miRNA in taxane resistance remains to be clarified at this stage.

Within our drug screening, we had one PARP inhibitor, Veliparib (Bitler et al., 2017) and we observed that *hsa-miR-182-5p* was associated with resistance to this drug. This finding contrasts an earlier study in which *hsa-miR-182-5p* was found to target *BRCA1* and sensitized the breast cancer cell line MDA-MB-231 to PARP inhibitors (Moskwa et al., 2011). Neijenhuis *et al.* have studied miRNAs which sensitize cells to the PARP inhibitor Olaparib using a large miRNA mimic screening, but did not identify *hsa-miR-182-5p* as involved in drug sensitivity (Neijenhuis et al., 2013). It is tempting to speculate that different miRNAs might play a role in the action of different PARP inhibitors, but the genetic makeup of studied cell lines/models might contribute (Neijenhuis et al., 2013). Since our study contained a large number of cell lines we reduced the influence of the genetic makeup on the identified miRNA-drug association favoring a more general link with Veliparib response. In our study *hsa-miR-182-5p* was associated with a large number of pathways, including the pathway "Cell cycle G1-S check-point" (see **Table 2**), which showed overall a downregulation. It remains, however, unclear how the downregulation of this cell cycle checkpoint might exactly contribute to drug resistance, although cell cycle changes have been earlier found as a drug resistance mechanism for PARP inhibitors (Bitler et al., 2017).

The drug Tivantinib targets the c-MET kinase and prevents it from downstream signaling (Pievsky and Pyrsopoulos, 2016). In our screening we found 3 miRNAs significantly associated with Tivantinib drug response. *Hsa-let-7d-5p* and *hsa-miR-18a-5p* were associated with drug sensitivity and *hsa-miR-637* with drug resistance. Interestingly, others have found that *hsa-miR-637* expression leads to downregulation of *STAT3* activity in hepatocellular carcinoma cells (Zhang et al., 2011) and STAT3 is one of the downstream activated proteins of c-MET activity (Organ and Tsao, 2011). Furthermore, HGF-c-MET signaling also leads to activation of the PI3K/AKT pathway (Gherardi et al., 2012) and *AKT1* has been identified as a direct target of *hsa-miR-637* in pancreatic ductal adenocarcinoma cells (Xu et al., 2018). One can speculate that the inactivation of c-MET effector pathways (through *AKT1* and *STAT3*) through this miRNA characterizes cells which rely on different pathways for growth and survival and therefore identifies cell lines inherently resistant to c-MET inhibitors.

Regarding Tivantinib sensitivity, one of the identified miRNAs was *hsa-let-7d-5p*. This miRNA has been shown to be downregulated by *STAT3* in breast cancer cells (Guo et al., 2013). This contradicts the above hypothesis that Tivantinib-sensitive cells are characterized by active *STAT3* and high *hsa-let-7d-5p* expression and Tivantinib-resistant cells have inactive/low *STAT3* and high *hsa-miR-637* expression. Nevertheless, we also observed a negative correlation between *hsa-miR-637* and *hsa-let-7d-5p* in line with

their opposing association with drug sensitivity. Other factors might play an additional role here. The last Tivantinib-associated miRNA, *hsa-miR-18a-5p*, associated with drug sensitivity, did not show significantly correlated expression with *hsa-let-7d-5p*. Others have found that *hsa-miR-18a-5p* targets *PIAS3* directly and in this way causes an increase in *STAT3* transcriptional activity in gastric cancer cell lines (Wu et al., 2013) fitting with the hypothesis for *hsa-miR-637*.

For the proteasome inhibitor Bortezomib (Lightcap et al., 2000) several resistance mechanisms have been found. Thioredoxin reductase 1 (HGNC symbol: *TXNRD1*) upregulation has been linked to Bortezomib resistance via upregulation of NF-κB-regulated genes in myeloma cells (Raninga et al., 2016), as well as the upregulation of heat shock proteins and related genes in several cancer types (Farrell and Reagan, 2018). Further resistance mechanisms are upregulation of proteasome subunits or increased proteasome activity which were observed in mesothelioma and myeloma (Farrell and Reagan, 2018). Upregulation of the aggresome/autophagy pathway (both in myeloma), and constitutive NF-κB or AKT signaling in multiple myeloma have also been observed (Wallington-Beddoe et al., 2018). *Hsa-let-7a-5p* was associated with Bortezomib sensitivity in our study. This miRNA has been shown to target KBRAS2 (*NKIRAS2)*, an inhibitor of NF-κB signaling and seems to be itself upregulated upon NF-κB signaling in human macrophages (Murphy et al., 2010). *Hsa-let-7a-5p* targets furthermore a negative regulator of NF-κB signaling, *TNFAIP3*, in HEK293T cells (Liu et al., 2015). NF-κB pathway inhibition is one of the effects of Bortezomib treatment (Farrell and Reagan, 2018), but generally activation of NF-κB signaling has been implicated in Bortezomib resistance (Wallington-Beddoe et al., 2018). Interestingly, there has been one report stating that upregulation of NF-κB signaling due to treatment with Lapatinib sensitized cells to Bortezomib in triple-negative breast cancer (Chen et al., 2013). It remains therefore unclear whether upregulation of NF-κB signaling aids to sensitize cells to Bortezomib or increases drug resistance in breast cancer and whether *hsa-let-7a-5p* increases Bortezomib sensitivity via downregulation of *NKIRAS2* and *TNFAIP3*.

Sensitivity to the FGFR inhibitor JNJ-707 (Uhr et al., 2015) was associated with *hsa-miR-135a-3p*, however, so far very little is known about this miRNA and unfortunately no information supporting the association with this drug could be found.

For the histone deacetylase (HDAC) inhibitor Panobinostat (Lemoine and Younes, 2010) several resistance mechanisms have been observed such as overexpression of the anti-apoptotic protein Bcl-2 in cutaneous T-cell lymphoma patients (Lee et al., 2012). Of interest is also the observation that Panobinostat treatment led to activation of NF-κB

signaling in leukemic cells and the blockage of this signaling increased sensitivity to the drug (Rosato et al., 2010). In our study, we found that the expression of *hsa-miR-185-3p* was associated with sensitivity to Panobinostat. However, little is known so far about the functions of this miRNA and there are no known targets which could explain how this miRNA might sensitize cells to Panobinostat.

For the farnesyltransferase inhibitor Tipifarnib (Norman, 2002), there are a few reports on resistance mechanisms. In previously untreated AML patients a 2-gene-classifier has been found which can predict response to Tipifarnib (Raponi et al., 2008). Patients responding to treatment are characterized by high expression of *RASGRP1*, a guanine nucleotide exchange factor which can activate RAS and low expression of *APTX*, a protein involved in DNA excision repair (Raponi et al., 2008). One of the properties Tipifarnib has is the inhibition of RAS farnesylation (Raponi et al., 2007) and the upregulation of *RASGRP1* might characterize cancer cells which rely on RAS signaling. In our study *hsa-miR-629-5p* was associated with Tipifarnib resistance, however, based on the current knowledge about this miRNA, it is unclear if there is a direct mechanistic link between the observed drug resistance and the miRNA and how this miRNA might influence drug response.

Among the miRNA-drug associations where subtype played a role, *hsa-miR-23a-3p* was expressed lower in luminal cell lines. Others have found that this miRNA targets the progesterone receptor (Gilam et al., 2017) and seems to be downregulated by estradiol (Saumet et al., 2012), which could support preferential expression of this miRNA in triple-negative cell lines. *Hsa-miR-338-3p* was expressed lower in the basal subtype and higher in the luminal subtype. The basal subtype in cell lines is characterized by *EGFR* expression in contrast to the luminal subtype which typically lacks expression of this protein (Hollestelle et al., 2010). Interestingly it has been shown that *EGFR* expression downregulates *hsa-miR-338-3p* expression (Liang et al., 2017), matching our observations on the breast cancer cell lines.

Conclusions

We were able to identify several miRNAs associated with drug resistance or drug sensitivity in our large panel of breast cancer cell lines. Several of the miRNAs found in our screen have been linked to pathways targeted by the drugs or genes involved in drug-resistance mechanisms. Next to those associations, we also identified a number

of miRNAs, which have not been researched much in the context of drug sensitivity/ resistance so far but may hold great potential once more is known about their biology.

Besides identifying miRNAs associated with drugs already used in the clinic we also identified miRNAs associated with several new anticancer agents, which are expected to enter the clinic in the coming years. Since biomarkers can help in discriminating those patients which will respond better to therapy and miRNAs are well measurable, further research into this topic will be of great value and might potentially validate these miRNAs as biomarkers.

In conclusion, our hypothesis-generating study suggests that miRNAs could be used as predictors of drug response and once independently validated and/or experimentally confirmed holds great potential for an application in the clinic.

Supporting information

Supporting information can be accessed under: https://doi.org/10.1371/journal.pone.0216400

S1 Table. MiRNA expression data for the cell line OCUB-M. MiRNA expression data for the cell line OCUB-M is given including the respective probe ID for each miRNA assessed. (XLSX)

S2 Table. Annotation information for the miRNAs associated with drug response. The former and current miRNA nomenclature is given, together with the miRBase (Griffiths-Jones et al., 2006) MIMAT identifier and the Exiqon oligonucleotide probe ID. (XLSX)

S3 Table. Overview of all drugs and the employed analysis method. All drugs used in this study are listed with their respective type of analysis. The IC_{50} values and profiles of all drugs are available in the supplemental data of a previous publication (Uhr et al., 2015). (XLSX)

S4 Table. Multivariate analysis with molecular breast cancer subtype. MiRNAs significantly associated with molecular breast cancer subtypes. Breast cancer subtypes assessed include luminal, basal and normal-like subtype. MiRNAs which are not associated with any of the three subtypes are printed in bold. (XLSX)

S5 Table. MiRNAs associated with Tivantinib. Spearman correlation coefficients among the different miRNAs are listed and the significant associations are highlighted in orange or blue depending on the correlation type. Orange = significant negative association, blue = significant positive association. Furthermore, all p-values of the associations are given. (XLSX)

S6 Table. MiRNAs associated with Docetaxel. Spearman correlation coefficients among the different miRNAs are listed and the significant associations are highlighted in orange or blue depending on the correlation type. Orange = significant negative association, blue = significant positive association. Furthermore, all p-values of the associations are given. (XLSX)

S7 Table. MiRNAs associated with Paclitaxel. Spearman correlation coefficients among the different miRNAs are listed and the significant associations are highlighted in orange or blue depending on the correlation type. Orange = significant negative association, blue = significant positive association. Furthermore, all p-values of the associations are given. (XLSX)

S8 Table. MiRNA expression and IC_{50} values per cell line. MiRNA expression data for the significantly associated miRNAs, not influenced by subtype are given per cell line. The IC_{50} values of the respective associated drug are given as well per cell line. Cell lines are color-coded based on their subtype: green = luminal, orange = basal, black = normal-like. MiRNA expression values and IC_{50} values are colored in a red-green color range with highest values in red and lowest in green. (XLSX)

Acknowledgements

We would like to thank Maxime P. Look for her advice on statistical analyses.

References

Ahmadian, M., Wistuba, I.I., Fong, K.M., Behrens, C., Kodagoda, D.R., Saboorian, M.H., Shay, J., Tomlinson, G.E., Blum, J., Minna, J.D., et al. (1997). Analysis of the FHIT gene and FRA3B region in sporadic breast cancer, preneoplastic lesions, and familial breast cancer probands. Cancer Res *57*, 3664–3668.

Band Horwitz, S. (1992). Mechanism of action of taxol. Trends Pharmacol. Sci. *13*, 134–136.

Bitler, B.G., Watson, Z.L., Wheeler, L.J., and Behbakht, K. (2017). PARP inhibitors: Clinical utility and possibilities of overcoming resistance. Gynecol. Oncol. *147*, 695–704.

Blows, F.M., Driver, K.E., Schmidt, M.K., Broeks, A., van Leeuwen, F.E., Wesseling, J., Cheang, M.C., Gelmon, K., Nielsen, T.O., Blomqvist, C., et al. (2010). Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. PLoS Med *7*, e1000279.

Cailleau, R., Young, R., Olivé, M., and Reeves, W.J. (1974). Breast tumor cell lines from pleural effusions. J. Natl. Cancer Inst. *53*, 661–674.

Cailleau, R., Olivé, M., and Cruciger, Q.V. (1978). Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. In Vitro *14*, 911–915.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. Cell *122*, 33–43.

Chang, J.C., Wooten, E.C., Tsimelzon, A., Hilsenbeck, S.G., Gutierrez, M.C., Tham, Y.-L., Kalidas, M., Elledge, R., Mohsin, S., Osborne, C.K., et al. (2005). Patterns of Resistance and Incomplete Response to Docetaxel by Gene Expression Profiling in Breast Cancer Patients. J. Clin. Oncol. *23*, 1169–1177.

Chen, Y.-J., Yeh, M.-H., Yu, M.-C., Wei, Y.-L., Chen, W.-S., Chen, J.-Y., Shih, C.-Y., Tu, C.-Y., Chen, C.-H., Hsia, T.-C., et al. (2013). Lapatinib-induced NF-kappaB activation sensitizes triple-negative breast cancer cells to proteasome inhibitors. Breast Cancer Res. BCR *15*, R108.

Cortez, M.A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A.K., and Calin, G.A. (2011). MicroRNAs in body fluids—the mix of hormones and biomarkers. Nat. Rev. Clin. Oncol. *8*, 467–477.

Dabney, A., and Storey, J.D. Q-value estimation for false discovery rate control. (https://bioconductor.org/packages/release/bioc/html/qvalue.html).

Davidson, J.M., Gorringe, K.L., Chin, S.-F., Orsetti, B., Besret, C., Courtay-Cahen, C., Roberts, I., Theillet, C., Caldas, C., and Edwards, P. a. W. (2000). Molecular cytogenetic analysis of breast cancer cell lines. Br. J. Cancer *83*, 1309–1317.

Ethier, S.P., Mahacek, M.L., Gullick, W.J., Frank, T.S., and Weber, B.L. (1993). Differential isolation of normal luminal mammary epithelial cells and breast cancer cells from primary and metastatic sites using selective media. Cancer Res. *53*, 627–635.

Ethier, S.P., Kokeny, K.E., Ridings, J.W., and Dilts, C.A. (1996). erbB Family Receptor Expression and Growth Regulation in a Newly Isolated Human Breast Cancer Cell Line. Cancer Res. *56*, 899–907.

Fanini, F., and Fabbri, M. (2016). MicroRNAs and cancer resistance: A new molecular plot. Clin. Pharmacol. Ther. *99*, 485–493.

Farrell, M.L., and Reagan, M.R. (2018). Soluble and Cell–Cell-Mediated Drivers of Proteasome Inhibitor Resistance in Multiple Myeloma. Front. Endocrinol. *9*, 218.

Fogh, J., Wright, W.C., and Loveless, J.D. (1977). Absence of HeLa cell contamination in 169 cell lines derived from human tumors. J Natl Cancer Inst *58*, 209–214.

Forozan, F., Veldman, R., Ammerman, C.A., Parsa, N.Z., Kallioniemi, A., Kallioniemi, O.P., and Ethier, S.P. (1999). Molecular cytogenetic analysis of 11 new breast cancer cell lines. Br. J. Cancer *81*, 1328–1334.

Galletti, E., Magnani, M., Renzulli, M.L., and Botta, M. (2007). Paclitaxel and docetaxel resistance: molecular mechanisms and development of new generation taxanes. ChemMedChem *2*, 920–942.

Gazdar, A.F., Kurvari, V., Virmani, A., Gollahon, L., Sakaguchi, M., Westerfield, M., Kodagoda, D., Stasny, V., Cunningham, H.T., Wistuba, I.I., et al. (1998). Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer *78*, 766–774.

Gherardi, E., Birchmeier, W., Birchmeier, C., and Woude, G.V. (2012). Targeting MET in cancer: rationale and progress. Nat. Rev. Cancer *12*, 89–103.

Gilam, A., Shai, A., Ashkenazi, I., Sarid, L.A., Drobot, A., Bickel, A., and Shomron, N. (2017). MicroRNA regulation of progesterone receptor in breast cancer. Oncotarget *8*, 25963–25976.

Glendenning, J., and Tutt, A. (2011). PARP inhibitors--current status and the walk towards early breast cancer. Breast Edinb. Scotl. *20 Suppl 3*, S12-19.

Goeman, J.J., van de Geer, S.A., de Kort, F., and van Houwelingen, H.C. (2004). A global test for groups of genes: testing association with a clinical outcome. Bioinforma. Oxf. Engl. *20*, 93–99.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. *34*, D140-144.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, L., Chen, C., Shi, M., Wang, F., Chen, X., Diao, D., Hu, M., Yu, M., Qian, L., and Guo, N. (2013). Stat3-coordinated Lin-28–let-7–HMGA2 and miR-200–ZEB1 circuits initiate and maintain oncostatin M-driven epithelial–mesenchymal transition. Oncogene *32*, 5272–5282.

Hackett, A.J., Smith, H.S., Springer, E.L., Owens, R.B., Nelson-Rees, W.A., Riggs, J.L., and Gardner, M.B. (1977). Two syngeneic cell lines from human breast tissue: the aneuploid mammary epithelial (Hs578T) and the diploid myoepithelial (Hs578Bst) cell lines. J. Natl. Cancer Inst. *58*, 1795–1806.

Hammond, S.M. (2015). An overview of microRNAs. Adv. Drug Deliv. Rev. *87*, 3–14.

Hausser, J., and Zavolan, M. (2014). Identification and consequences of miRNA-target interactions — beyond repression of gene expression. Nat. Rev. Genet. *15*, 599–612.

Hollestelle, A., Nagel, J.H.A., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J., et al. (2010). Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res. Treat. *121*, 53–64.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. *12*, 99–110.

Iorio, M.V., and Croce, C.M. (2012). MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol Med *4*, 143–159.

Keydar, I., Chen, L., Karby, S., Weiss, F.R., Delarea, J., Radu, M., Chaitcik, S., and Brenner, H.J. (1979). Establishment and characterization of a cell line of human breast carcinoma origin. Eur. J. Cancer *15*, 659–670.

Kim, Y.-K. (2015). Extracellular microRNAs as Biomarkers in Human Disease. Chonnam Med. J. *51*, 51–57.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597–610.

Lasfargues, E.Y., and Ozzello, L. (1958). Cultivation of human breast carcinomas. J. Natl. Cancer Inst. *21*, 1131–1147.

Lasfargues, E.Y., Coutinho, W.G., and Redfield, E.S. (1978). Isolation of two human tumor epithelial cell lines from solid breast carcinomas. J. Natl. Cancer Inst. *61*, 967–978.

Lee, J.-H., Choy, M.L., and Marks, P.A. (2012). Chapter Two - Mechanisms of Resistance to Histone Deacetylase Inhibitors. In Advances in Cancer Research, S. Grant, ed. (Academic Press), pp. 39–86.

Lemoine, M., and Younes, A. (2010). Histone deacetylase inhibitors in the treatment of lymphoma. Discov. Med. *10*, 462–470.

Li, H., and Yang, B.B. (2013). Friend or foe: the role of microRNA in chemotherapy resistance. Acta Pharmacol. Sin. *34*, 870–879.

Liang, Y., Xu, X., Wang, T., Li, Y., You, W., Fu, J., Liu, Y., Jin, S., Ji, Q., Zhao, W., et al. (2017). The EGFR/miR-338-3p/EYA2 axis controls breast tumor growth and lung metastasis. Cell Death Dis. *8*, e2928.

Lightcap, E.S., McCormack, T.A., Pien, C.S., Chau, V., Adams, J., and Elliott, P.J. (2000). Proteasome Inhibition Measurements: Clinical Application. Clin. Chem. *46*, 673–683.

Lin, S., and Gregory, R.I. (2015). MicroRNA biogenesis pathways in cancer. Nat. Rev. Cancer *15*, 321–333.

Lippman, M., Bolan, G., and Huff, K. (1976). The effects of estrogens and antiestrogens on hormone-responsive human breast cancer in long-term tissue culture. Cancer Res. *36*, 4595–4601.

Littlewood-Evans, A.J., Bilbe, G., Bowler, W.B., Farley, D., Wlodarski, B., Kokubo, T., Inaoka, T., Sloane, J., Evans, D.B., and Gallagher, J.A. (1997). The osteoclast-associated protease cathepsin K is expressed in human breast carcinoma. Cancer Res. *57*, 5386–5390.

Liu, J., Zhu, L., Xie, G., Bao, J., and Yu, Q. (2015). Let-7 miRNAs Modulate the Activation of NF-κB by Targeting TNFAIP3 and Are Involved in the Pathogenesis of Lupus Nephritis. PLOS ONE *10*, e0121256.

Lovat, F., Valeri, N., and Croce, C.M. (2011). MicroRNAs in the pathogenesis of cancer. Semin Oncol *38*, 724–733.

Meltzer, P., Leibovitz, A., Dalton, W., Villar, H., Kute, T., Davis, J., Nagle, R., and Trent, J. (1991). Establishment of two new cell lines derived from human breast carcinomas with HER-2/neu amplification. Br. J. Cancer *63*, 727–735.

Moskwa, P., Buffa, F.M., Pan, Y., Panchakshari, R., Gottipati, P., Muschel, R.J., Beech, J., Kulshrestha, R., Abdelmohsen, K., Weinstock, D.M., et al. (2011). miR-182-Mediated Downregulation of BRCA1 Impacts DNA Repair and Sensitivity to PARP Inhibitors. Mol. Cell *41*, 210–220.

Munshi, N., Jeay, S., Li, Y., Chen, C.-R., France, D.S., Ashwell, M.A., Hill, J., Moussa, M.M., Leggett, D.S., and Li, C.J. (2010). ARQ 197, a novel and selective inhibitor of the human c-Met receptor tyrosine kinase with antitumor activity. Mol. Cancer Ther. *9*, 1544–1553.

Murphy, A.J., Guyre, P.M., and Pioli, P.A. (2010). Estradiol Suppresses NF-κB Activation through Coordinated Regulation of let-7a and miR-125b in Primary Human Macrophages. J. Immunol. *184*, 5029–5037.

Neijenhuis, S., Bajrami, I., Miller, R., Lord, C.J., and Ashworth, A. (2013). Identification of miRNA modulators to PARP inhibitor response. DNA Repair *12*, 394–402.

Norman, P. (2002). Tipifarnib (Janssen Pharmaceutica). Curr. Opin. Investig. Drugs Lond. Engl. 2000 *3*, 313–319.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *27*, 29–34.

Organ, S.L., and Tsao, M.-S. (2011). An overview of the c-MET signaling pathway. Ther. Adv. Med. Oncol. *3*, S7–S19.

Pievsky, D., and Pyrsopoulos, N. (2016). Profile of tivantinib and its potential in the treatment of hepatocellular carcinoma: the evidence to date. J. Hepatocell. Carcinoma *3*, 69–76.

Pothof, J., Verkaik, N.S., van IJcken, W., Wiemer, E.A.C., Ta, V.T.B., van der Horst, G.T.J., Jaspers, N.G.J., van Gent, D.C., Hoeijmakers, J.H.J., and Persengiev, S.P. (2009). MicroRNA-mediated gene silencing modulates the UV-induced DNA-damage response. EMBO J. *28*, 2090–2099.

R Core Team (2018). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Raninga, P.V., Di Trapani, G., Vuckovic, S., and Tonissen, K.F. (2016). TrxR1 inhibition overcomes both hypoxia-induced and acquired bortezomib resistance in multiple myeloma through NF-κβ inhibition. Cell Cycle Georget. Tex *15*, 559–572.

Raponi, M., Harousseau, J.-L., Lancet, J.E., Löwenberg, B., Stone, R., Zhang, Y., Rackoff, W., Wang, Y., and Atkins, D. (2007). Identification of Molecular Predictors of Response in a Study of Tipifarnib Treatment in Relapsed and Refractory Acute Myelogenous Leukemia. Clin. Cancer Res. *13*, 2254–2260.

Raponi, M., Lancet, J.E., Fan, H., Dossey, L., Lee, G., Gojo, I., Feldman, E.J., Gotlib, J., Morris, L.E., Greenberg, P.L., et al. (2008). A 2-gene classifier for predicting response to the farnesyltransferase inhibitor tipifarnib in acute myeloid leukemia. Blood *111*, 2589–2596.

Revelle, W. (2018). Psych: Procedures for psychological, psychometric, and personality research. (https://cran.r-project.org/package=psych) (Evanston, Illinois: Northwestern University).

Riaz, M., van Jaarsveld, M.T.M., Hollestelle, A., Prager-van der Smissen, W.J.C., Heine, A.A.J., Boersma, A.W.M., Liu, J., Helmijr, J., Ozturk, B., Smid, M., et al. (2013). miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. Breast Cancer Res. BCR *15*, R33.

Rosato, R.R., Kolla, S.S., Hock, S.K., Almenara, J.A., Patel, A., Amin, S., Atadja, P., Fisher, P.B., Dent, P., and Grant, S. (2010). Histone Deacetylase Inhibitors Activate NF-κB in Human Leukemia Cells through an ATM/NEMO-related Pathway. J. Biol. Chem. *285*, 10064–10077.

Saumet, A., Vetter, G., Bouttier, M., Antoine, E., Roubert, C., Orsetti, B., Theillet, C., and Lecellier, C.-H. (2012). Estrogen and retinoic acid antagonistically regulate several microRNA genes to control aerobic glycolysis in breast cancer cells. Mol. Biosyst. *8*, 3242–3253.

Sawada, T., Chung, Y.S., Nakata, B., Kubo, T., Kondo, Y., Sogabe, T., Onoda, N., Ogawa, Y., Yamada, N., and Sowa, M. (1994). [Establishment and characterization of a human breast cancer cell line, OCUB-1]. Hum. Cell *7*, 138–144.

Sieuwerts, A.M., Meijer-van Gelder, M.E., Timmermans, M., Trapman, A.M.A.C., Garcia, R.R., Arnold, M., Goedheer, A.J.W., Portengen, H., Klijn, J.G.M., and Foekens, J.A. (2005). How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *11*, 7311–7321.

Smith, H.S., Wolman, S.R., Dairkee, S.H., Hancock, M.C., Lippman, M., Left, A., and Hackett, A.J. (1987). Immortalization in Culture: Occurrence at a Late Stage in the Progression of Breast Cancer. JNCI J. Natl. Cancer Inst. *78*, 611–615.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc. Natl. Acad. Sci. *100*, 8418–8423.

Soule, H.D., Vazguez, J., Long, A., Albert, S., and Brennan, M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. J. Natl. Cancer Inst. *51*, 1409–1416.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. *100*, 9440–9445.

Sundaresan, V., Chung, G., Heppell-Parton, A., Xiong, J., Grundy, C., Roberts, I., James, L., Cahn, A., Bench, A., Douglas, J., et al. (1998). Homozygous deletions at 3p12 in breast and lung cancer. Oncogene *17*, 1723–1729.

Tomlinson, G.E., Chen, T.T., Stastny, V.A., Virmani, A.K., Spillman, M.A., Tonk, V., Blum, J.L., Schneider, N.R., Wistuba, I.I., Shay, J.W., et al. (1998). Characterization of a breast cancer cell line derived from a germ-line BRCA1 mutation carrier. Cancer Res. *58*, 3237–3242.

Tozlu, S., Girault, I., Vacher, S., Vendrell, J., Andrieu, C., Spyratos, F., Cohen, P., Lidereau, R., and Bieche, I. (2006). Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. Endocr. Relat. Cancer *13*, 1109–1120.

Trempe, G.L. (1976). Human breast cancer in culture. Recent Results Cancer Res , *57*, 33-41.

Uhr, K., Smissen, W.J.P. der, Heine, A.A., Ozturk, B., Smid, M., Göhlmann, H.W., Jager, A., Foekens, J.A., and Martens, J.W. (2015). Understanding drugs in breast cancer through drug sensitivity screening. SpringerPlus *4*, 611.

Varela, M.A., Roberts, T.C., and Wood, M.J. (2013). Epigenetics and ncRNAs in brain function and disease: mechanisms and prospects for therapy. Neurotherapeutics *10*, 621–631.

Voigt, W. (2005). Sulforhodamine B assay and chemosensitivity. Methods Mol Med *110*, 39–48.

Wallington-Beddoe, C.T., Sobieraj-Teague, M., Kuss, B.J., and Pitson, S.M. (2018). Resistance to proteasome inhibitors and other targeted therapies in myeloma. Br. J. Haematol. *182*, 11–28.

de Weger, V.A., Beijnen, J.H., and Schellens, J.H.M. (2014). Cellular and clinical pharmacology of the taxanes docetaxel and paclitaxel – a review: Anticancer. Drugs *25*, 488–494.

Wiemer, E.A. (2007). The role of microRNAs in cancer: no small matter. Eur J Cancer *43*, 1529–1544.

Wu, W., Takanashi, M., Borjigin, N., Ohno, S., Fujita, K., Hoshino, S., Osaka, Y., Tsuchida, A., and Kuroda, M. (2013). MicroRNA-18a modulates STAT3 activity through negative regulation of PIAS3 during gastric adenocarcinogenesis. Br. J. Cancer *108*, 653–661.

Xu, R., He, W., Tang, J., Guo, W., Zhuang, P., Wang, C., Fu, W., and Zhang, J. (2018). Primate-specific miRNA-637 inhibited tumorigenesis in human pancreatic ductal adenocarcinoma cells by suppressing Akt1 expression. Exp. Cell Res. *363*, 310–314.

Zhang, J., He, M., Fu, W., Wang, H., Chen, L., Zhu, X., Chen, Y., Xie, D., Lai, P., Chen, G., et al. (2011). Primate-specific microRNA-637 inhibits tumorigenesis in hepatocellular carcinoma by disrupting signal transducer and activator of transcription 3 signaling. Hepatol. Baltim. Md *54*, 2137–2148.

Chapter 5

Genomic events in breast cancer cell lines associated with drug response

Katharina Uhr, Marcel Smid, Anouk A. J. Heine, A. Hollestelle, John A. Foekens and John W. M. Martens

Manuscript in preparation

Abstract

Genomic events like mutations and copy number aberrations (CNAs) are the drivers of cancer, but can also be passenger events occurring during cancer progression. Driving changes in the genome affect cancer initiation, growth, invasion and metastasis, as well as the outcome of drug therapy.

To gain insight into the relationship between genomic events and drug response we analyzed drug sensitivity data of 34 drugs targeting various cellular pathways in 42 breast cancer cell lines originating from different breast cancer subtypes. Gene mutation status had been previously generated in our laboratory or, for 28 of the cell lines, was publicly available from the Cancer Cell Line Encyclopedia (CCLE). CNA information was obtained from Affymetrix SNP 6.0 microarrays.

Mutations were associated with drug response by employing either a Mann-Whitney or Fisher's Exact test depending on the response profile of the respective drug. A test for trend was used for associating CNAs with drug response.

Among our findings were a strong association of *RB1* pathway defects with 17-AAG resistance and the association of mutated *TP53* with Nutlin-3 and MI-219 resistance. Additional associations were between wild type *PTEN* and *ERBB2* amplification with Erlotinib sensitivity, mutated *PIK3CA* and sensitivity to Sirolimus and Vorinostat, mutated *CDH1* and sensitivity to Mitoxantrone and Quisinostat, mutated *MLL3* and Doxorubicin sensitivity and mutated *ATM* and Methotrexate resistance. All associations above concerned known tumor driver genes. The 89 significant CNA-drug associations included the following associations: 5-Fluorouracil resistance with a gain spanning *BIRC7* and *PTK6*, Cisplatin resistance with a gain including *KCNJ2*, Dasatinib resistance with a gain comprising *ZDHHC7*, Paclitaxel sensitivity with a gain spanning *DAXX* and Sunitinib resistance with a loss including the tumor driver *ARID1A*.

In summary, we have identified a large number of associations between mutated genes, CNAs and drug response including several known tumor driver genes.

Once independently validated, our findings could potentially serve as biomarkers aiding treatment decisions.

Introduction

Cancer is a genetic disease involving single nucleotide variants, insertions, deletions, translocations and copy number aberrations (CNAs), which not only influence tumor initiation and progression, but can also affect the response to therapy (Stratton et al., 2009). Such genomic events, particularly gains and losses, affect the gene and protein levels (Herman et al., 1995; Stratton et al., 2009). Mutations can generate truncated protein products due to frameshifts or stop codons (Miller and Pearce, 2014), but they can also modify the amino acid sequence of proteins, thereby affecting their characteristics and functional properties (Stratton et al., 2009). As a consequence, mutated proteins may lose their normal function, or exhibit increased activity (Stratton et al., 2009). Mutations can influence drug response in multiple ways by changing the activity of drug metabolizing enzymes e.g. certain cytochrome P 450 genes (Ingelman-Sundberg and Rodriguez-Antona, 2005) or a drug transporter such as *ABCC2* (Hulot et al., 2005). Additionally, mutations can create susceptibility to drugs such as the L858R missense mutation or the delL747-T751 deletion in the *EGFR* gene for sensitivity to Gefitinib (Bell et al., 2005). However, mutations can also disrupt the interaction between a drug and its target e.g. in the case of the T790M mutation in *EGFR* leading to resistance to Gefitinib and Erlotinib or the *BCR-ABL* T315I mutation and Imatinib resistance (Bell et al., 2005). Furthermore, they can also lead to the activation of another survival-promoting pathway, thereby bypassing the effect of the drug treatment (Holohan et al., 2013).

Genomic events can, however, also provide a new treatment opportunity, exemplified by the case of the *ERBB2* gene amplification in breast cancer (Slamon et al., 1987; Treish et al., 2000). This genetic alteration does not only enhance the growth of the tumor but it also appears to be its 'Achilles heel' because of its efficient druggability (Press et al., 2005; Slamon et al., 2011; Treish et al., 2000). Another example in breast cancer is the concept of synthetic lethality. Germline mutations in the homologous recombination repair proteins *BRCA1* and *BRCA2* confer a high risk to develop breast cancer (Fong et al., 2009), but due to impaired homologous recombination repair these tumors are very susceptible to treatment with an inhibitor of PARP1, a protein involved in single strand base excision repair (Fong et al., 2009). Treatment of *BRCA1* or *BRCA2* deficient cells with PARP inhibitors results in the accumulation of double-strand breaks, which affects DNA replication and ultimately leads to the death of tumor cells (Fong et al., 2009). PARP inhibitors are currently indicated for the treatment of *BRCA1/2* mutated breast and ovarian cancer (Ashworth and Lord, 2018).

While some information on drug response with regard to specific genomic alterations is known for breast cancer, most currently known genetic aberrations in breast cancer are not druggable. It is, therefore, of great interest to find drugs which target breast cancer based on the genomic alterations present in a tumor.

In the present study, we combined genomics data with drug screening data of 34 drugs in 42 breast-cancer cell lines. Significant genotype-drug associations observed here, may, after thorough independent and functional validation, be used to guide treatment of patients with breast cancer.

Materials & Methods

Culturing, drug incubation and IC_{50} determination

Data of a previously performed drug screening were used as described (Uhr et al., 2015). In brief, 45 breast cancer cell lines were screened using 37 drugs in different dilutions. Drugs were supplied by Janssen Pharmaceutica (Beerse, Belgium), except for Veliparib which was provided by AbbVie (North Chicago, USA). Cell proliferation was assessed with the Sulforhodamine B (SRB) assay (Voigt, 2005) and IC_{50} values were determined using the absorbance values. Finally, 42 cell lines and 34 drugs gave meaningful measurements (the SRB assay succeeded and showed differential drug response) and were used for further analyses. IC_{50} values, drug response profiles as well as omitted cell lines and drugs are detailed in Uhr *et al.* (Uhr et al., 2015).

Mutations and their association with drug sensitivity

Mutation data had been generated earlier as described (Hollestelle et al. 2010; Riaz et al. 2013) and was available for all 42 cell lines with successful drug screening data. Thirteen genes were mutated, amplified or methylated in at least four of the screened cell lines and subsequently used in our analyses. Mutation data has been published earlier (Hollestelle et al., 2010; Riaz et al., 2013) except for *ERBB2* and *KRAS* for which additional information was available; supplied in **supplemental table S1**. Cell lines with gene mutations of unclear oncogenic significance were excluded from the respective statistical tests. The Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) was used to acquire mutation data for genes not assessed in-house and contained 28 cell lines overlapping with our drug screening dataset. For the CCLE dataset the following variants were defined as mutations: de novo start in and out of frame, frameshift deletions and insertions, in-frame deletions and insertions, missense and nonsense mutations, splice-site deletions and insertions, splice-site single nucleotide variants

(SNVs) and stop-codon insertions. Variants in the 3' UTR, the 5' UTR, the 5' flank, in introns or variants which are silent were not counted as mutations. Limiting the number of genes to those which had an aberration in at least four of the cell lines resulted in 83 genes from the CCLE study for data analysis.

For statistical analysis gene mutation status was categorized as mutated or wildtype. For drugs with a linear profile of IC_{50} values, Mann-Whitney test was used to analyze IC_{50} values with mutation status, while a Fisher's Exact test was used to test non-linear drug-sensitivity profiles for their association with mutations. Linear drug sensitivity profiles were defined as drugs with less than four cell lines with IC_{50} values above the highest drug concentration used in the study. IC_{50} values of non-linear drug sensitivity profiles were categorized as resistant when the values were above the maximum tested drug concentration, while the remaining values were labeled as sensitive. Data analysis was performed in R versions 3.4.2 up to 3.5.0 and p-values were false discovery adjusted using the function p.adjust option "BH" within R. P-values below 0.05 and with a FDR below 0.3 were deemed significant. Subtype/EMT analyses were performed in R using Fisher's Exact or Kruskal Wallis test.

CNAs and their association with drug sensitivity

Copy number (CN) data were generated using SNP 6.0 chips from Affymetrix (Santa Clara, USA) as previously described (Riaz et al., 2013). Raw data was preprocessed with Nexus software (BioDiscovery, Hawthorne, USA) using the recommended settings for SNP calling, segmentation and copy number status provided by Nexus software. Chromosomal areas with an amplification or loss were collected for all cell lines, and genes present in these areas were labeled as Gain or Loss, accordingly. Next, the Gain, Loss or Neutral calls of the genes were associated with IC_{50} values of the cell lines. A test for trend across ordered groups was used to identify genes of which the copy number state was ordinally associated with IC_{50} values. Next, a selection was made for genes with at least 7 events in all three possible groups (Gain, Loss, Neutral). In total, 760 genes were included in the test for trend, associating the groups with the IC_{50} values of 34 drugs. For genes with too few observations in one of the groups, we analyzed the remaining two groups with sufficient observations (e.g. Loss versus Neutral and Gain versus Neutral). 3,692 genes were included for the Loss versus Neutral analysis, and 9,534 genes in the Gain versus Neutral analysis. All 93 breast cancer driver genes as defined by Nik-Zainal (Nik-Zainal et al., 2016), could be assessed for CNAs and 60 driver genes had sufficient events for statistical testing with drug response.

In the next step single genes were combined to regions: Those genes with a p-value < 0.01, and which were located next to each other on the genome, were used to subsequently define regions of copy number variation associated with drug response. When a gene had a p-value < 0.01 a region was started, if the next gene also had a p-value < 0.01 and showed an association with drug response in the same direction (based on the z-value) it was added to the region. This procedure was continued until the subsequent gene did not fulfill these criteria (i.e. the gene was not significant in the same direction). Then the region was ended and a new region was started with the next significant gene. In some cases, a gene that lies between regions was just above the p-value cutoff, thereby generating two separate regions that are more likely to be a single affected region. To circumvent this, sequential regions on the same chromosomal arm were evaluated if they actually showed the same patterns of copy number gains/losses in the same cell lines, using the Chi-square test. If this test showed a significant concordance, the regions were merged. The p-value of a region was based on the p-value of the gene in this region with the lowest p-value as this gene is the most likely candidate. Subsequently, Bonferroni correction was applied on all identified regions. Statistical tests were performed with STATA 13 (StataCorp LP, College Station, USA), and Bonferroni-corrected p-values < 0.05 were considered statistically significant. Figures were created in Excel 2010 (Microsoft, Redmond, USA), using the Genome Decoration Page of NCBI (https://www.ncbi.nlm.nih.gov/genome/tools/gdp) and Inkscape 0.92 (Free Software Foundation Inc., Boston, USA).

Results

Mutations and drug susceptibility

First, we associated in-house mutation data of 13 known breast cancer driver genes and methylation data of 2 genes in 42 cell lines with IC_{50} values of all 34 drugs, which led to 22 significant associations (**Table 1**, **Fig. 1**).

Table 1. Association of drug response with in-house mutation data

Drug	Gene	p-value	FDR-adjusted p-value	Mutation/amplification/ methylation is associated with
17-AAG	p16 ^{INK4a} (*CDKN2A*)	1.69E-02	2.53E-01	Resistance
Belinostat	*CDH1*	3.36E-03	5.04E-02	Sensitivity
Belinostat	*PIK3CA*	1.67E-02	1.25E-01	Sensitivity
Erlotinib*	*ERBB2* amplification	2.56E-03	3.84E-02	Sensitivity
Erlotinib*	*PTEN*	3.45E-02	2.59E-01	Resistance
JNJ-208*	*MDM2* amplification	1.38E-02	2.07E-01	Sensitivity
Methotrexate*	*PIK3CA*	1.06E-02	1.59E-01	Resistance
Methotrexate*	*CDH1* methylation	3.19E-02	2.40E-01	Sensitivity
MI-219*	*TP53*	1.04E-05	1.56E-04	Resistance
Mitoxantrone	*CDH1* methylation	1.84E-02	1.86E-01	Sensitivity
Mitoxantrone	*CDH1*	2.48E-02	1.86E-01	Sensitivity
Nutlin-3*	*TP53*	1.04E-05	1.56E-04	Resistance
Paclitaxel	p14ARF (*CDKN2A*)	3.94E-02	2.93E-01	Sensitivity
Panobinostat	*PIK3CA*	7.44E-03	1.12E-01	Sensitivity
Panobinostat	*CDH1* methylation	4.01E-02	2.72E-01	Resistance
Quisinostat	*CDH1*	3.29E-02	2.48E-01	Sensitivity
Quisinostat	*CDH1* methylation	3.29E-02	2.48E-01	Resistance
Serdemetan	*ERBB2* amplification	3.17E-02	2.65E-01	Resistance
Serdemetan	*CCND1* amplification	3.58E-02	2.65E-01	Resistance
Sirolimus	*PIK3CA*	6.21E-03	7.66E-02	Sensitivity
Sirolimus	*BRCA1*	1.53E-02	7.66E-02	Resistance
Vorinostat	*PIK3CA*	6.77E-03	1.01E-01	Sensitivity

Associations between drugs and mutated, methylated or amplified genes are listed, besides the type of association (resistance/sensitivity). P-values < 0.05 with a FDR < 0.3 were considered significant. *Drugs analyzed with Fisher's Exact test. The gene *CDKN2A* encodes the two proteins p16^{INK4A}, as well as, in another reading frame, the protein p14ARF. Due to the differing functions of these proteins and their differing amino acid sequence, they are listed separately here.

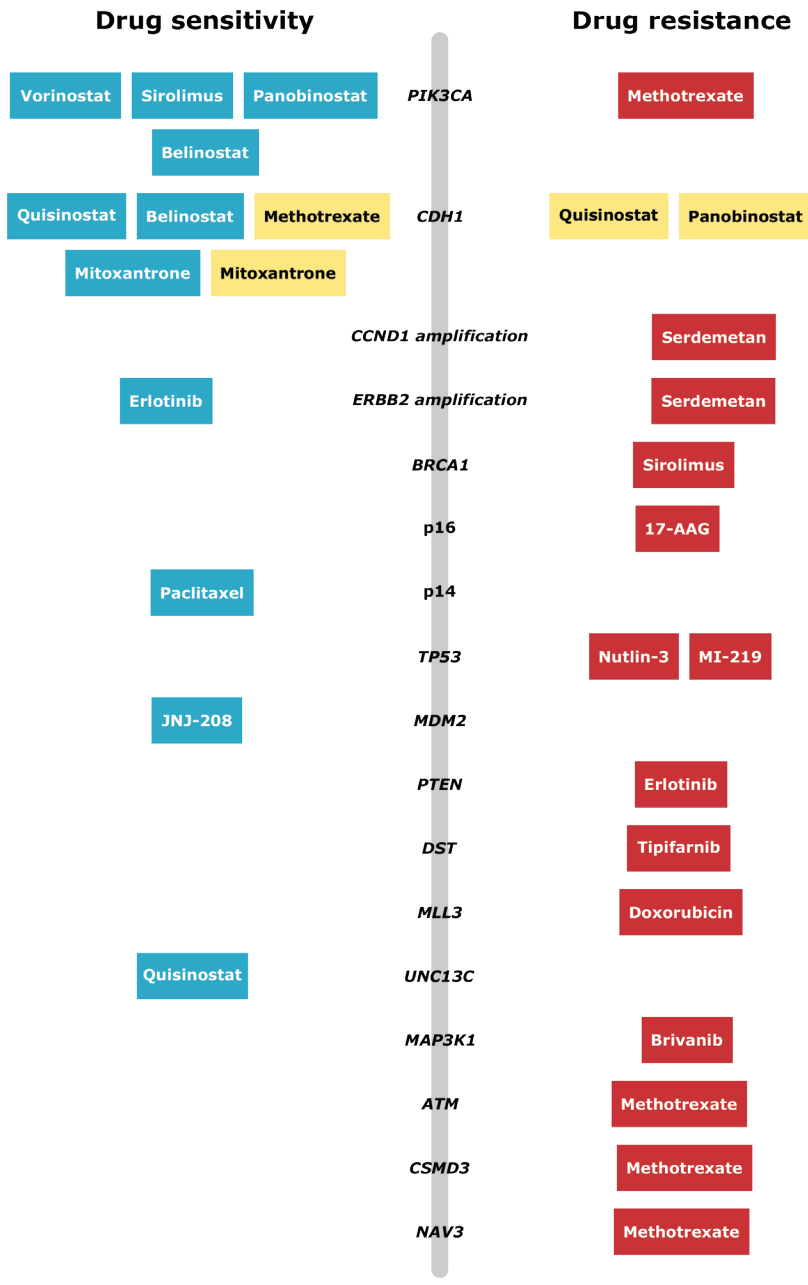


Fig. 1. Mutated, amplified and methylated genes associated with drug response. Associations with drug sensitivity are shown on the left, while associations with drug resistance are shown on the right. Blue = mutated/amplified gene associated with drug sensitivity, red = mutated/amplified gene associated with drug resistance, yellow = methylated gene associated with drug sensitivity or resistance depending on the position in the table

The most significant associations were between mutated *TP53* and resistance to Nutlin-3 as well as MI-219. Another highly significant finding was the association between sensitivity to Erlotinib and *ERBB2* amplification, and between resistance to Erlotinib and mutated *PTEN*. Sensitivity to the drug Mitoxantrone was associated with methylation of *CDH1* and, albeit to a lesser degree, with mutation of *CDH1*. For the drug Sirolimus, mutated *PIK3CA* was associated with sensitivity, while mutated *BRCA1* was associated with resistance. For the HDAC inhibitors, of which drug sensitivity among cell lines is highly correlated (Uhr et al., 2015), several drugs had associations with the same genes, however no gene was associated with all four HDAC inhibitors: Sensitivity to Belinostat was associated with mutated *CDH1* and mutated *PIK3CA*, while sensitivity to Vorinostat or Panobinostat was only associated with mutated *PIK3CA.* Sensitivity to Quisinostat, however, was only associated with mutated *CDH1*, but not *PIK3CA*. On the contrary, methylated *CDH1* was associated with resistance to Quisinostat and Panobinostat. Interestingly, mutated *PIK3CA* was associated with resistance to Methotrexate. Additionally, *CDH1* methylation also associated with sensitivity to Methotrexate. Furthermore, we noted an association between amplification of *MDM2* and sensitivity to JNJ-208 and mutated $p16^{INK4a}$ with resistance to 17-AAG. Amplification of *ERBB2* and *CCND1* was associated with resistance to Serdemetan. Finally, sensitivity to Paclitaxel was associated with mutated p14ARF.

Within our findings we noticed that Panobinostat and Methotrexate showed opposing associations with *PIK3CA* and *CDH1* methylation. None of the cell lines displayed methylated *CDH1* as well as mutations in *PIK3CA*, which suggests mutual exclusivity. We suspected that breast cancer subtype might influence the opposing drug-associations and indeed, *CDH1* methylation was exclusively present in the normal-like and basal subtype cell lines, with the majority of cell lines in both groups exhibiting *CDH1* methylation. However, neither of the two drugs showed an association with subtype, nor did *PIK3CA* associate with subtype in our cell line dataset ($p = 0.17$). As *CDH1* methylation has previously been described in cell lines that had undergone EMT (Hollestelle et al., 2013), we also tested whether EMT influenced the associations. While *CDH1* methylation was only present in cell lines with EMT, as published earlier (Hollestelle et al., 2013), neither Methotrexate, Panobinostat nor mutated *PIK3CA* were associated with EMT.

Therefore, subtype and EMT might be confounding factors in the opposing drug associations; however, they are not the only factors and do not fully explain our findings.

Next we used the CCLE dataset (Barretina et al., 2012) to test additional genes for an association between drug response and mutations. In total, we found seven associations between a compound and a mutated gene (**Table 2**, **Fig. 1**).

Table 2. Association of drug response with CCLE mutation data

Drug	Gene	p-value	FDR-adjusted p-value	Mutation is associated with
Brivanib*	*MAP3K1*	3.05E-03	2.53E-01	Resistance
Doxorubicin	*MLL3*	2.63E-03	2.18E-01	Sensitivity
Methotrexate*	*ATM*	7.69E-03	2.62E-01	Resistance
Methotrexate*	*NAV3*	7.69E-03	2.62E-01	Resistance
Methotrexate*	*CSMD3*	9.47E-03	2.62E-01	Resistance
Quisinostat	*UNC13C*	3.05E-03	2.53E-01	Sensitivity
Tipifarnib	*DST*	3.05E-03	2.53E-01	Resistance

Associations between drugs and mutated genes from the CCLE dataset are listed, besides the type of association (resistance/sensitivity). P-values < 0.05 with a FDR < 0.3 were considered significant. *Drugs analyzed with Fisher's Exact test.

These associations are: mutated *MLL3* with Doxorubicin sensitivity, mutated *MAP3K1* with Brivanib resistance, mutated *DST* with Tipifarnib resistance, mutated *UNC13C* with Quisinostat sensitivity, mutated *ATM, NAV3* and *CSMD3* with Methotrexate resistance. The three genes associated within the CCLE dataset with Methotrexate, were largely co-mutated in the same cell lines. Based on the p-value one can conclude that *CSMD3* is a less likely candidate, however, we were not able to distinguish between *ATM* and *NAV3* for the most likely candidate based on the outcome of the statistical tests and mutation profiles.

Correlation of copy number changes with drug response types

To find significant associations between copy number changes and IC_{50} values, we first correlated the copy number state of a gene with drug response data, then neighboring genes were combined to regions and Bonferroni-correction was applied to determine significant regions. All drugs but Lapatinib and JNJ-707 were significantly associated with copy number changes, ranging from one region (Cisplatin, Decitabine, Tamoxifen, Tivantinib, Veliparib) up to five regions (5-Fluorouracil, Vandetanib, Vorinostat) (**Fig. 2**).

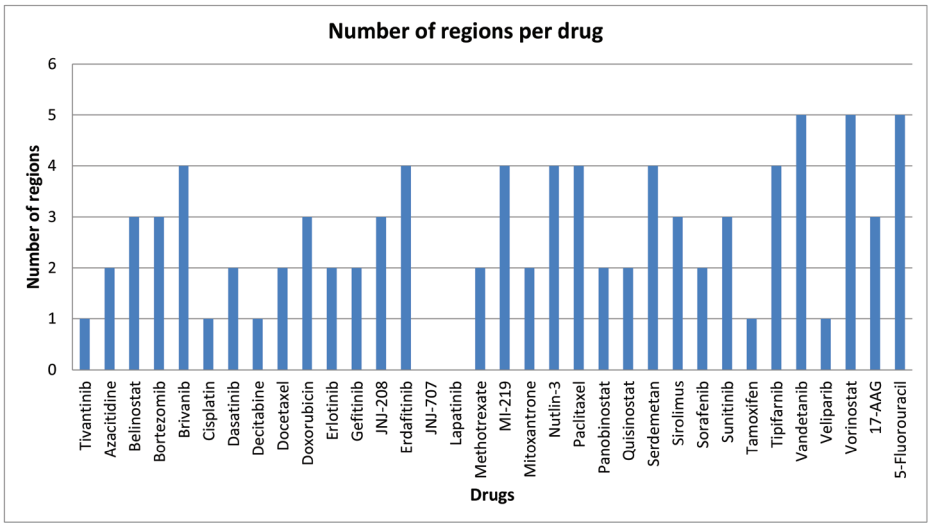


Fig. 2. Number of CNA regions associated with drug response. The number of CNA regions associated with a particular drug is depicted to illustrate differences among the drugs. JNJ-707 and Lapatinib were not associated with any CNA regions.

In total, there were 89 region-drug associations consisting of 1,438 individual gene-drug associations (**Supplemental Table S2**). These regions consisted of 1 up to 139 genes. The median number of genes per region was 8. In our cell line dataset, therefore medium-sized CNAs dominate in regard to drug response over large scale genomic events encompassing tens to hundreds of genes. We noticed that the loss of a region was in 18 cases (81.82%) associated with an increase in drug sensitivity, while a gain was associated in 45 cases (67.16%) with higher sensitivity to a drug (**Fig. 3**).

When looking at drug resistance, the loss of a region was associated with drug resistance in only 4 cases (18.18%), while in 22 instances (32.84%) a gain was associated with drug resistance (**Fig. 3**). Therefore, the majority of CNAs was associated with drug sensitivity but not resistance. Subsequently we were interested to investigate if regions overlapped between drugs, as this could argue that cell lines sensitive to one drug might also harbor sensitivity to another type of drug. Furthermore, if highly correlated drugs show different association patterns this could highlight small differences in response, potentially also seen in the clinic. Several regions or some genes of a region were associated with more than one drug (**Supplemental Table S2**; **Fig. 4**).

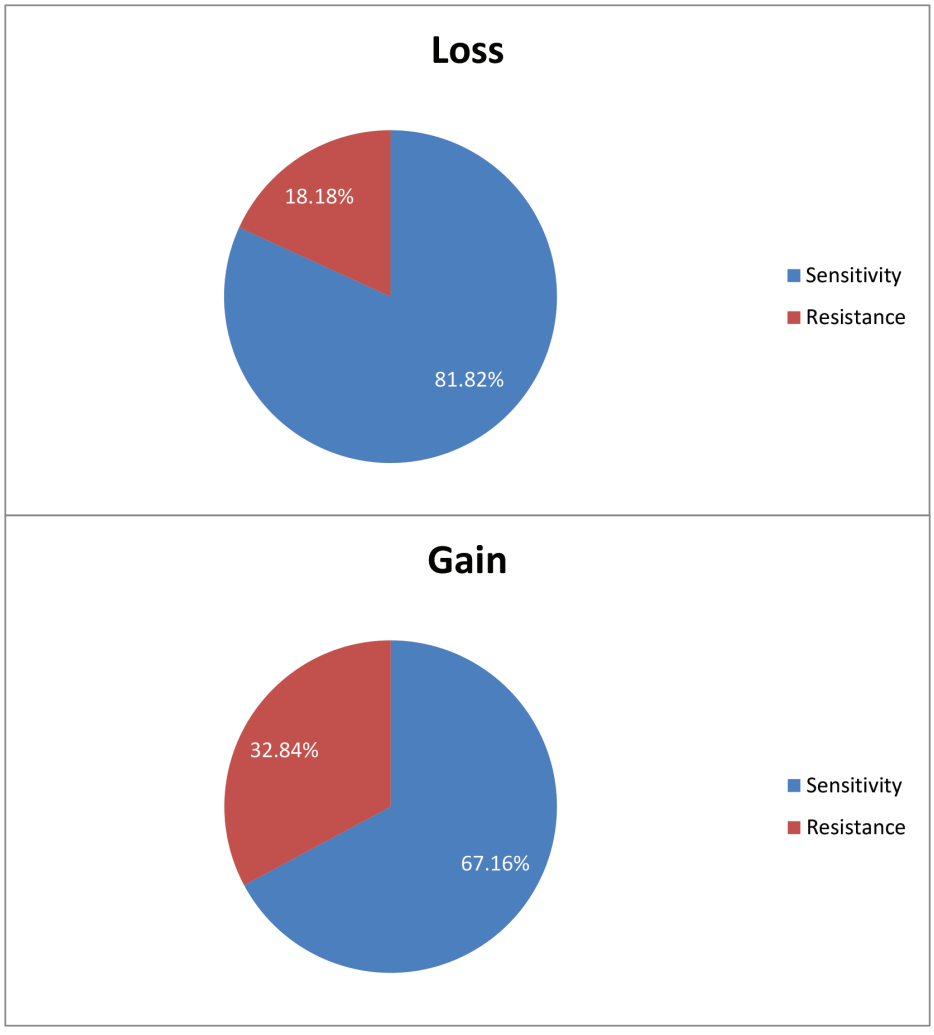


Fig. 3. CNA types and their association with drug response. The loss or gain of a region and the respective association with drug resistance or drug sensitivity are depicted in comparison to all associations.

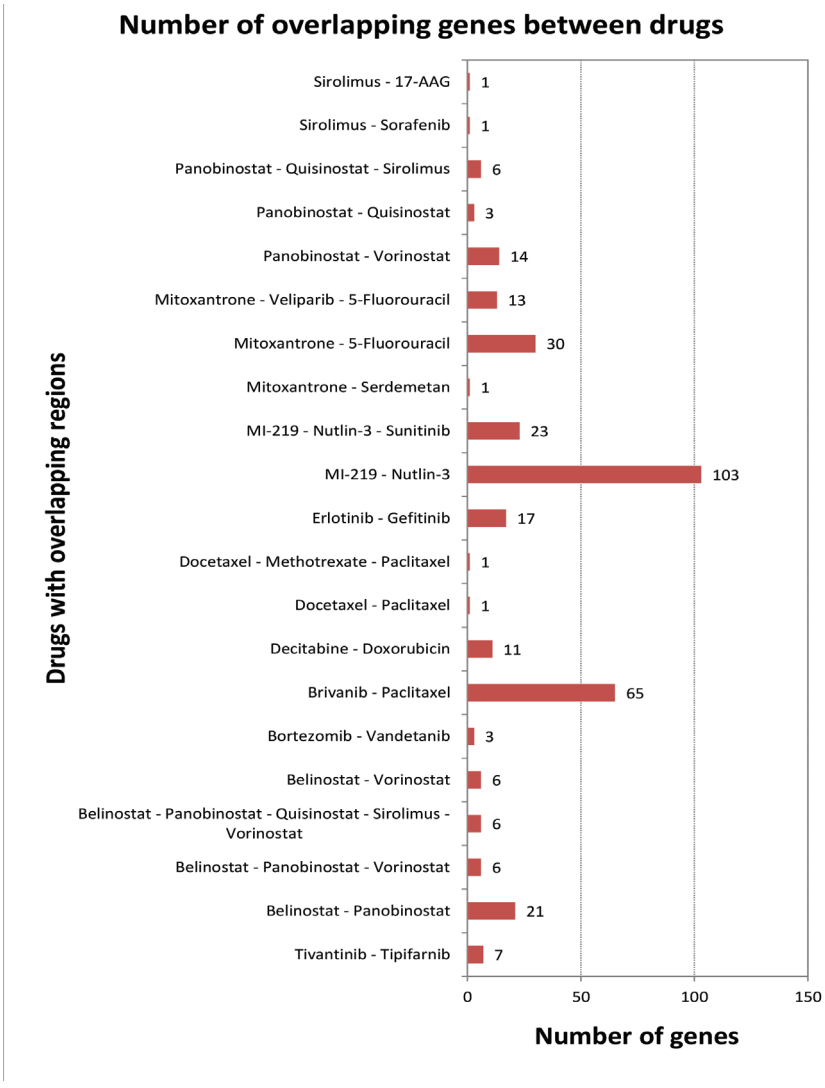


Fig. 4. CNA regions associated with more than one drug. The number of overlapping regions between different drugs is shown.

MI-219 and Nutlin-3 had the highest number of overlapping genes, namely 126 of which 103 were shared between MI-219 and Nutlin-3 while 23 genes were additionally associated with response to Sunitinib. Interestingly the association was in an opposing direction (loss of the respective overlapping region was associated with resistance to Sunitinib but sensitivity to Nutlin-3 and MI-219). Brivanib and Paclitaxel shared as well a high number of genes (n = 65), while surprisingly only few genes were shared between the highly correlated drugs Docetaxel and Paclitaxel (Uhr et al., 2015) (two

– of which one gene is additionally associated with response to Methotrexate). The *EGFR* targeting drugs Erlotinib and Gefitinib (Giovannetti et al., 2011) only had 17 genes in common among each other, while the different HDAC inhibitors (Uhr et al., 2015) shared several genes, but shared regions were different among different HDAC inhibitors. There was no gene affected by a CNA that was only associated with HDAC inhibitors, however, six genes were associated with all HDAC inhibitors in our screening (Panobinostat, Vorinostat, Quisinostat, Belinostat) and additionally with Sirolimus. CNAs associated with response to 5-Fluorouracil, Cisplatin, Dasatinib, Paclitaxel and Sunitinib are depicted in the **supplemental fig. 1-5** highlighting some of the affected genes within the CNA regions.

To make sure that the drug-associated CNAs were not cell-line specific we compared our gene list with gene copy number aberrations found in a large collection of breast tumors. For this purpose we used the discovery set comprised of roughly 1,000 breast tumors of an earlier study (Curtis et al., 2012). Of our 1,438 individual gene-drug associations, 827 comprised genes that had an aberrant copy number in breast tumors as well (Curtis et al., 2012) and all but nine regions included genes that had been identified by Curtis *et al.* (Curtis et al., 2012) (**Supplemental Table S2**). Interestingly, these nine cell-line specific regions consisted of only a single gene.

Discussion

Mutations and copy-number changes have a dominant role in cancer, such as initiating the tumorigenic process, fueling aggressiveness (Nik-Zainal et al., 2016; Stratton et al., 2009) as well as being involved in the effectiveness of therapy (Sharma and Settleman, 2007). We undertook a screening of a wide range of drugs in breast cancer cell lines to identify mutations and copy-number changes that show an association with drug sensitivity. In total, in our large panel of breast cancer cell lines we identified 29 significant associations between gene aberrations and drug response and 89 regions with CNAs showed an association with drug response. The observed associations included a fair amount of breast cancer driver genes as defined by Nik-Zainal *et al.* (Nik-Zainal et al., 2016). Besides the breast cancer driver genes screened in-house (**Table 1**), the CCLE-screened genes *MAP3K1*, *MLL3* and *ATM* which associated with drug response in our study (**Table 2**) were as well known breast cancer drivers. For the genes associated with drug response in our CNA analysis, three genes were known breast cancer drivers: *ARID1A*, *FOXA1* and *ZNF217* (Nik-Zainal et al., 2016) (**Supplemental Table S2**). While *ARID1A* and *FOXA1* have been previously found mutated in breast tumors

(Nik-Zainal et al., 2016), in our study they showed CNAs. However, these copy number changes were in line with their reported roles, as *ARID1A* is a recessive cancer gene (https://cancer.sanger.ac.uk/census; recessive = an event in both genomic copies is required for an effect), and in our study we found the most common aberration for this gene was a loss in copy number (9 cell lines with a copy number loss, 1 with a gain). The same was true for *FOXA1*, a dominant cancer gene (https://cancer.sanger.ac.uk/census; dominant = an event in one genomic copy causes an effect), which was mainly gained (12 cell lines with a copy number gain vs 1 loss). *ZNF217,* has been observed to be subjected to a copy number gain in breast cancer (Nik-Zainal et al., 2016), in line with our findings (24/42 cell lines have a gain). Our finding that these known breast cancer driver genes are associated with drug response is promising and might point to intrinsic vulnerabilities to specific drugs.

Mutations/methylations associated with drug response

Besides analyzing our in-house mutation data (Hollestelle et al., 2010; Riaz et al., 2013) on its own merit, we also sought to extend the dataset by using the CCLE dataset (Barretina et al., 2012).

Among the mutations that showed an association with drug response, we identified several known associations, such as mutations in *TP53* with resistance to Nutlin-3 and MI-219 (Aziz et al., 2011; Künkele et al., 2012; Long et al., 2010). Other previously reported associations were *ERBB2*-amplification with Erlotinib sensitivity (Emlet et al., 2007) and mutated *PTEN* with Erlotinib resistance (Mellinghoff et al., 2005; Sos et al., 2009; Yamasaki et al., 2007).

Within our study mutations affecting the reading frame for $p16^{INK4a}$, a protein involved in the *RB1* signaling pathway (Liggett and Sidransky, 1998), were associated with drug resistance to 17-AAG. Cell lines with defective *RB1* signaling have been observed to show a higher resistance to 17-AAG (Vilenchik et al., 2004; Yin et al., 2005) and ectopic expression of $p16^{INK4a}$ could successfully reverse the resistance phenotype in a cell line with methylated $p16^{INK4a}$ (Yin et al., 2005). As we had observed only an association between 17-AAG resistance with mutations in $p16^{INK4a}$, while Vilenchik *et al.* had shown increased resistance associated with mutations in *RB1* and Yin *et al.* with methylated $p16^{INK4a}$, we decided to re-analyze our data grouping all cell lines with mutated $p16^{INK4a}$, methylated $p16^{INK4a}$ and mutated *RB1* together as "inactivated *RB1* pathway". While in our original analysis only mutated $p16^{INK4a}$ was associated with resistance to 17-AAG (p-value: 1.68E-02), combining all cell lines with different pathway inactivations in one group resulted in a much stronger association with 17-AAG resistance (p-value:

6.59E-04). These data support the above-mentioned publications and underline the link between a dysfunctional *RB1* pathway and 17-AAG resistance.

Furthermore, we found that *PIK3CA* mutations were associated with Sirolimus sensitivity. A study on PI3K/AKT/mTOR inhibitors, including Sirolimus, in patients, showed that tumors with mutated *PIK3CA* responded better to therapy than tumors with wildtype *PIK3CA* (Janku et al., 2012). This finding argues for an increased sensitivity to this drug type due to the mutation. We also found that sensitivity to Vorinostat was associated with *PIK3CA* mutations. While there have been no findings supporting this association yet, e.g. due to limited case numbers (Park et al., 2016), Vorinostat has been shown to inhibit the protein PI3K directly (Kawamata et al., 2007). As *PIK3CA* mutations are typically activating mutations i.e. they increase the kinase activity of the protein (Kang et al., 2005; Samuels, 2004) and mutated *PIK3CA* is a breast cancer driver gene (Nik-Zainal et al., 2016), the mutated gene could increase cell vulnerability to Vorinostat due to reliance on this breast cancer driver gene for cancer growth (oncogene addiction) (Sharma and Settleman, 2007). Another finding in our study was the association of mutated *CDH1* with sensitivity to Quisinostat and Belinostat. In a study looking at drug sensitivity of the non-tumorigenic breast cell line MCF10A and its *CDH1*$^{-/-}$ derivative, a greater sensitivity to Paclitaxel, Quisinostat, Panobinostat and Vorinostat was discovered in the *CDH1*$^{-/-}$ cell line (Telford et al., 2015). While the association with Quisinostat was confirmed by this earlier study (Belinostat was not studied), we did, however, not observe a significant association between Vorinostat nor Panobinostat nor Paclitaxel and mutated *CDH1*. Potentially, Vorinostat, Panobinostat and Paclitaxel have additional targets in the breast cancer cell lines which have a stronger impact than mutation of the recessive cancer gene *CDH1* (https://cancer.sanger.ac.uk/census). Surprisingly methylation of *CDH1* was associated with resistance to Quisinostat and Panobinostat in our study. This finding could be influenced by breast cancer subtypes of the cell lines as *CDH1* methylation but not mutation was associated with subtype, however neither of these two HDAC inhibitors exhibited a subtype-dependent response. As *CDH1* methylation is strongly associated with EMT (Hollestelle et al., 2013), we analyzed whether Panobinostat and Quisinostat associated with EMT as well, however, this was not the case.

Another drug which was associated with mutated and methylated *CDH1* was Mitoxantrone. In a study on a patient-derived gastric cancer cell line with a loss of *CDH1*, Mitoxantrone was more effective than in gastric cancer cell lines without *CDH1* mutations (Chen et al., 2017), in line with our observation of increased sensitivity to Mitoxantrone among *CDH1* mutated or methylated breast cancer cell lines.

Using the CCLE dataset we found an association between mutations in *MLL3,* a histone methyltransferase, with Doxorubicin sensitivity. This finding is supported by a publication in which the authors observed an increase in Doxorubicin sensitivity upon *MLL3* knockout, furthermore noted MLL3's role in a tumor suppressor complex and mentioned activities of the MLL3 protein in the DNA damage response downstream of p53 (Lee et al., 2009). Another interesting finding was the association of mutated *ATM* with Methotrexate resistance. *ATM* is a gene involved in DNA repair and mutations in this gene are the cause of the cancer-predisposing disease Ataxia telangiectasia (Savitsky et al., 1995). Previous reports have noted a resistance to Methotrexate in cells of Ataxia telangiectasia patients by upregulation of the drug target DHFR (Lücke-Huhle, 1994; Lücke-Huhle et al., 1987), making the association of *ATM* mutations with Methotrexate resistance likely causative.

Drug response associated with copy number changes

Besides mutations, we also analyzed whether copy number changes associated with drug response and that was indeed the case. Overall, we noticed that CNAs were associated predominantly with drug sensitivity. This could imply that these genomic events create overall weaknesses that can be targeted by drugs and should warrant further research. A number of drugs showed associations with the same genomic regions. This was the case for highly correlated drugs such as Gefitinib and Erlotinib (Uhr et al., 2015), as well as the HDAC inhibitors albeit the genes that associated with all HDAC inhibitors were also associated with Sirolimus. Interestingly, the loss of several genes that associated with Nutlin-3 and MI-219 sensitivity was associated with Sunitinib resistance. However, the opposing finding for Sunitinib was not linked to *TP53* mutation status but only associated with the CNA. If validated this could imply intrinsic resistance to Sunitinib of these tumors and suggest e.g. Nutlin-3 and MI-219 as potential treatments.

The most significant association of a copy number change was with the drugs Nutlin-3 and MI-219. However, as these drugs are MDM2 inhibitors, which have a very narrow and specific function and seem to be heavily influenced by *TP53* mutation status (Aziz et al., 2011; Künkele et al., 2012; Long et al., 2010), the discovered associations might just be a coincidence because most breast cancer cell lines are *TP53* mutants and therefore the *TP53* wildtype cell lines are a minority in the analysis, decreasing the power of the analysis. Furthermore, none of these copy number changes spanned *TP53*, *MDM2* or was in the vicinity of these two genes.

Within the CNAs that were associated with 5-Fluorouracil resistance was a gain on chromosome 20 including the genes *BIRC7, PTK6* and *UCKL1*. *BIRC7* codes for an anti-

apoptotic protein, which increases the 5-Fluorouracil sensitivity when it is knocked down in renal carcinoma cells (Crnković-Mertens et al., 2007). *PTK6* has been found to increase sensitivity to 5-Fluorouracil when knocked down as well (Gierut et al., 2012). Therefore, it seems plausible that a gain of these genes increases 5-Fluorouracil resistance in breast cancer cells. The KEGG pathway 'Drug metabolism other enzymes' lists uridine kinases as involved in the 5-Fluorouracil metabolism, and *UCKL1* is among the three human uridine kinases listed. Within the pathway it acts in an intermediate step to 5-Fluorouridinetriphosphate (FUTP). An overexpression of UCKL1 could lead to a shift in favor of FUTP production over 5-Fluorodeoxyuridinetriphosphate, and therefore the RNA would be mainly affected through 5-Fluorouracil but less so the DNA of the treated cells. It is, however, also possible that the increased expression of one single enzyme in the pathway would lead to accumulation of intermediate metabolites. Based on the pathway, accumulation of metabolite at this point does not seem to enhance excretion as a potential resistance mechanism. It remains therefore unclear how increased uridine kinase expression would affect the 5-Fluorouracil metabolism. Besides, it has been noted that downregulation of *UCKL1* leads to increased sensitivity to apoptosis (Ambrose and Kornbluth, 2009), which in turn could mean that upregulation might be related to increased survival. This copy number aberration is also frequent in breast tumors (Curtis et al., 2012) and it could be interesting to investigate whether associations with response to 5-Fluorouracil are seen in patients as well.

Cisplatin resistance was found to be associated with a gain in 17q24.3, which included the frequently in breast cancer gained gene *KCNJ2* (Curtis et al., 2012), a gene which increases Cisplatin sensitivity when silenced (Li, 2013). Therefore, as the silencing of this gene leads to Cisplatin sensitivity a gain is likely to increase resistance.

Dasatinib resistance was found to be associated with a gain in 16q24.1. Within this region lies the gene *ZDHHC7,* which is frequently gained in breast cancers (Curtis et al., 2012) and has been shown to be a downstream target of Src (Knowlton et al., 2010), the protein kinase targeted by Dasatinib (Johnson et al., 2005). A gain of this palmitoylacyltransferase gene could therefore overcome the upstream inhibition of Src and let it fulfill its function in sex steroid receptor palmitoylation, a process required for membrane localization of the receptors (Pedram et al., 2012). Furthermore, a knockdown of *ZDHHC7* has been shown to influence cell proliferation of estrogen receptor positive breast cancer cells negatively (Pedram et al., 2012), a gain of this gene might therefore provide a survival benefit for breast cancer cells with or without Dasatinib treatment.

Paclitaxel sensitivity was found to be associated with a gain on chromosome 6, which affected the gene *DAXX*. A lack of DAXX protein has been linked earlier to Paclitaxel resistance in breast cancer cell lines and tumors (Giovinazzi et al., 2012) and fits a linear relationship between gene copy number and drug response.

We also looked whether the genes associated with drug response in the mutation analysis were also found in the CNA analysis. The genes *BRCA1, MAP3K1, NAV3* and *UNC13C* could not be tested due to too small group sizes, while the other genes did not show a significant association, which can likely be attributed to the lenient cutoff used for the mutation analysis, while the CNA analysis had a very stringent cutoff.

Within our CNA analyses we noticed that several breast cancer driver genes were affected by CNAs and were found to be associated with drug response. In detail: The loss of the breast cancer driver gene *ARID1A* was associated with resistance to Sunitinib and sensitivity to Nutlin-3 and MI-219. While *TP53* mutation status is the main determinant for Nutlin-3 and MI-219 activity (Aziz et al., 2011; Künkele et al., 2012; Long et al., 2010), in ovarian clear cell carcinoma *ARID1A* mutated tumors often have wildtype *TP53*, both genes have a role in the same pathway and stabilizing *TP53* through Nutlin-3 seems promising (Bitler et al., 2015). Furthermore, activity of Nutlin-3 in *ARID1A* mutant cells has been shown (Bitler et al., 2015). In renal clear cell carcinoma knockdown of *ARID1A* increased resistance to Sunitinib (Xiao et al., 2017), in line with our found association.

A gain of *FOXA1* was associated with sensitivity to all HDAC inhibitors screened in our study as well as Sirolimus and provides a promising lead for follow-up experiments. Gain of *ZNF217* was associated with resistance to Methotrexate, a chemotherapeutic drug frequently used in breast cancer (Kadakia et al., 2015).

To summarize: As breast cancer driver genes are required for breast cancer growth, our findings on association with drug response are very interesting as they could point towards vulnerabilities for targeting breast tumors. However, our presented results are associations and do not show functional validity and further research to validate these findings is required.

In conclusion: By employing CNA data as well as mutation data, we could identify large as well as smaller genomic events that showed an association with drug response. Our study provides a comprehensive collection of leads for functional validation experiments to determine causative or biomarker association. The literature indicates that several of our associations are likely causal; however, it is of course also possible

that the genomic events identified that associated with drug response here, function merely as biomarkers, and functional validation studies are warranted as subsequent step on the path to the clinic to ensure clinical validity of the results.

Acknowledgements

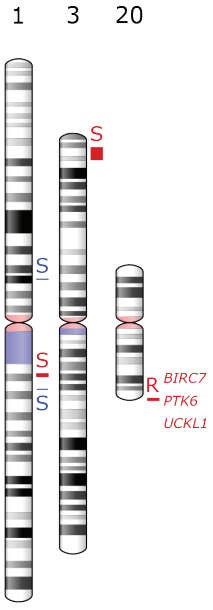
This study was funded by the Daniel den Hoed Stichting (K.U., J.A.F.), Cancer Genomics Netherlands - a grant from the Netherlands organization of Scientific Research (NWO) (J.W.M.M.) and the ERC Advanced Grant #322737 (J.A.F.). The drugs in the drug screening dataset were obtained from Janssen (Beerse, Belgium), except for Veliparib which was obtained from AbbVie (North Chicago, USA). The funders of this study had no influence in the design, data analysis, interpretation and writing of this manuscript.

Supplemental information

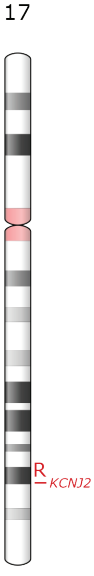
Supplemental Table S1. Previously unpublished mutation/amplification status for the genes *ERBB2* and *KRAS*.

Supplemental Table S2. CNAs associated with drug response

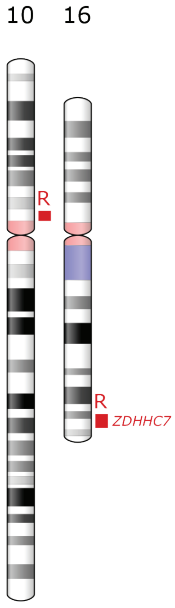
The supplemental tables can be accessed via: http://hdl.handle.net/1765/118526



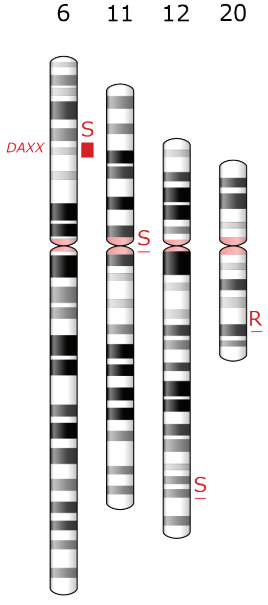
Supplemental fig. 1. CNAs associated with 5-Fluorouracil response. Losses are indicated in blue, gains in red and the type of association is indicated with S = sensitivity and R = resistance. Genes of interest in the affected regions are listed next to their location.



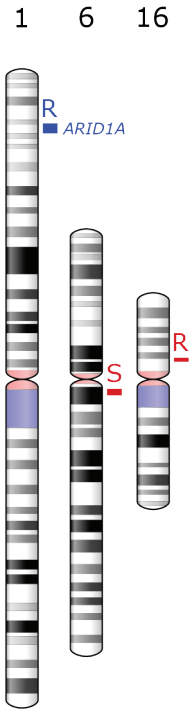
Supplemental fig. 2. CNAs associated with Cisplatin response. Losses are indicated in blue, gains in red and the type of association is indicated with S = sensitivity and R = resistance. Genes of interest in the affected regions are listed next to their location.



Supplemental fig. 3. CNAs associated with Dasatinib response. Losses are indicated in blue, gains in red and the type of association is indicated with S = sensitivity and R = resistance. Genes of interest in the affected regions are listed next to their location.



Supplemental fig. 4. CNAs associated with Paclitaxel response. Losses are indicated in blue, gains in red and the type of association is indicated with S = sensitivity and R = resistance. Genes of interest in the affected regions are listed next to their location.



Supplemental fig. 5. CNAs associated with Sunitinib response. Losses are indicated in blue, gains in red and the type of association is indicated with S = sensitivity and R = resistance. Genes of interest in the affected regions are listed next to their location.

References

Ambrose, E.C., and Kornbluth, J. (2009). Downregulation of uridine-cytidine kinase like-1 decreases proliferation and enhances tumor susceptibility to lysis by apoptotic agents and natural killer cells. Apoptosis Int. J. Program. Cell Death *14*, 1227–1236.

Ashworth, A., and Lord, C.J. (2018). Synthetic lethal therapies for cancer: what's next after PARP inhibitors? Nat. Rev. Clin. Oncol. *15*, 564-576

Aziz, M.H., Shen, H., and Maki, C.G. (2011). Acquisition of p53 mutations in response to the non-genotoxic p53 activator Nutlin-3. Oncogene *30*, 4678–4686.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity (vol 483, pg 603, 2012). Nature *492*, 290–290.

Bell, D.W., Gore, I., Okimoto, R.A., Godin-Heymann, N., Sordella, R., Mulloy, R., Sharma, S.V., Brannigan, B.W., Mohapatra, G., Settleman, J., et al. (2005). Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. Nat. Genet. *37*, 1315–1316.

Bitler, B.G., Fatkhutdinov, N., and Zhang, R. (2015). Potential therapeutic targets in *ARID1A* -mutated cancers. Expert Opin. Ther. Targets *19*, 1419–1422.

Chen, I., Mathews-Greiner, L., Li, D., Abisoye-Ogunniyan, A., Ray, S., Bian, Y., Shukla, V., Zhang, X., Guha, R., Thomas, C., et al. (2017). Transcriptomic profiling and quantitative high-throughput (qHTS) drug screening of CDH1 deficient hereditary diffuse gastric cancer (HDGC) cells identify treatment leads for familial gastric cancer. J. Transl. Med. *15*, 92.

Crnković-Mertens, I., Wagener, N., Semzow, J., Gröne, E.F., Haferkamp, A., Hohenfellner, M., Butz, K., and Hoppe-Seyler, F. (2007). Targeted inhibition of Livin resensitizes renal cancer cells towards apoptosis. Cell. Mol. Life Sci. *64*, 1137–1144.

Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Emlet, D.R., Brown, K.A., Kociban, D.L., Pollice, A.A., Smith, C.A., Ong, B.B.L., and Shackney, S.E. (2007). Response to trastuzumab, erlotinib, and bevacizumab, alone and in combination, is correlated with the level of human epidermal growth factor receptor-2 expression in human breast cancer cell lines. Mol. Cancer Ther. *6*, 2664–2674.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., et al. (2009). Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. N. Engl. J. Med. *361*, 123–134.

Gierut, J.J., Mathur, P.S., Bie, W., Han, J., and Tyner, A.L. (2012). Targeting protein tyrosine kinase 6 enhances apoptosis of colon cancer cells following DNA damage. Mol. Cancer Ther. *11*, 2311–2320.

Giovannetti, E., Zucali, P.A., Assaraf, Y.G., Leon, L.G., Smid, K., Alecci, C., Giancola, F., Destro, A., Gianoncelli, L., Lorenzi, E., et al. (2011). Preclinical emergence of vandetanib as a potent antitumour agent in mesothelioma: molecular mechanisms underlying its synergistic interaction with pemetrexed and carboplatin. Br. J. Cancer *105*, 1542–1553.

Giovinazzi, S., Lindsay, C.R., Morozov, V.M., Escobar-Cabrera, E., Summers, M.K., Han, H.S., McIntosh, L.P., and Ishov, A.M. (2012). Regulation of mitosis and taxane response by Daxx and Rassf1. Oncogene *31*, 13–26.

Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J.-P.J., Davidson, N.E., Sidransky, D., and Baylin, S.B. (1995). Inactivation of the CDKN2/p16/MTS1 Gene Is Frequently Associated with Aberrant DNA Methylation in All Common Human Cancers. Cancer Res. *55*, 4525–4530.

Hollestelle, A., Nagel, J.H.A., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J., et al. (2010). Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res. Treat. *121*, 53–64.

Hollestelle, A., Peeters, J.K., Smid, M., Timmermans, M., Verhoog, L.C., Westenend, P.J., Heine, A.A.J., Chan, A., Sieuwerts, A.M., Wiemer, E.A.C., et al. (2013). Loss of E-cadherin is not a necessity for epithelial to mesenchymal transition in human breast cancer. Breast Cancer Res. Treat. *138*, 47–57.

Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. Nat. Rev. Cancer *13*, 714–726.

Hulot, J.-S., Villard, E., Maguy, A., Morel, V., Mir, L., Tostivint, I., William-Faltaos, D., Fernandez, C., Hatem, S., Deray, G., et al. (2005). A mutation in the drug transporter gene ABCC2 associated with impaired methotrexate elimination. Pharmacogenet. Genomics *15*, 277–285.

Ingelman-Sundberg, M., and Rodriguez-Antona, C. (2005). Pharmacogenetics of drug-metabolizing enzymes: implications for a safer and more effective drug therapy. Philos. Trans. R. Soc. B Biol. Sci. *360*, 1563–1570.

Janku, F., Wheler, J.J., Westin, S.N., Moulder, S.L., Naing, A., Tsimberidou, A.M., Fu, S., Falchook, G.S., Hong, D.S., Garrido-Laguna, I., et al. (2012). PI3K/AKT/mTOR inhibitors in patients with breast and gynecologic malignancies harboring PIK3CA mutations. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *30*, 777–782.

Johnson, F.M., Saigal, B., Talpaz, M., and Donato, N.J. (2005). Dasatinib (BMS-354825) tyrosine kinase inhibitor suppresses invasion and induces cell cycle arrest and apoptosis of head and neck squamous cell carcinoma and non-small cell lung cancer cells. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *11*, 6924–6932.

Kadakia, A., Rajan, S.S., Abughosh, S., Du, X.L., and Johnson, M.L. (2015). CMF-regimen preferred as first-course chemotherapy for older and sicker women with breast cancer: Findings from a SEER-Medicare-based population study. Am. J. Clin. Oncol. *38*, 165–173.

Kang, S., Bader, A.G., and Vogt, P.K. (2005). Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic. Proc. Natl. Acad. Sci. *102*, 802–807.

Kawamata, N., Chen, J., and Koeffler, H.P. (2007). Suberoylanilide hydroxamic acid (SAHA; vorinostat) suppresses translation of cyclin D1 in mantle cell lymphoma cells. Blood *110*, 2667–2673.

Knowlton, M.L., Selfors, L.M., Wrobel, C.N., Gu, T.-L., Ballif, B.A., Gygi, S.P., Polakiewicz, R., and Brugge, J.S. (2010). Profiling Y561-dependent and -independent substrates of CSF-1R in epithelial cells. PloS One *5*, e13587.

Künkele, A., De Preter, K., Heukamp, L., Thor, T., Pajtler, K.W., Hartmann, W., Mittelbronn, M., Grotzer, M.A., Deubzer, H.E., Speleman, F., et al. (2012). Pharmacological activation of the p53 pathway by nutlin-3 exerts anti-tumoral effects in medulloblastomas. Neuro-Oncol. *14*, 859–869.

Lee, J., Kim, D.-H., Lee, S., Yang, Q.-H., Lee, D.K., Lee, S.-K., Roeder, R.G., and Lee, J.W. (2009). A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proc. Natl. Acad. Sci. U. S. A. *106*, 8513–8518.

Li, Y.L. (2013). Abstract 2119: Silencing of KCNJ2, a potassium influx channel, increases cisplatin-induced cell death in oral cancer. Cancer Res. *73*, 2119–2119.

Liggett, W.H., and Sidransky, D. (1998). Role of the p16 tumor suppressor gene in cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *16*, 1197–1206.

Long, J., Parkin, B., Ouillette, P., Bixby, D., Shedden, K., Erba, H., Wang, S., and Malek, S.N. (2010). Multiple distinct molecular mechanisms influence sensitivity and resistance to MDM2 inhibitors in adult acute myelogenous leukemia. Blood *116*, 71–80.

Lücke-Huhle, C. (1994). Similarities between human ataxia fibroblasts and murine SCID cells: high sensitivity to gamma rays and high frequency of methotrexate-induced DHFR gene amplification, but normal radiosensitivity to densely ionizing alpha particles. Radiat. Environ. Biophys. *33*, 201–210.

Lücke-Huhle, C., Hinrichs, S., and Speit, G. (1987). DHFR gene amplification in cultured skin fibroblasts of ataxia telangiectasia patients after methotrexate selection. Carcinogenesis *8*, 1801–1806.

Mellinghoff, I.K., Wang, M.Y., Vivanco, I., Haas-Kogan, D.A., Zhu, S., Dia, E.Q., Lu, K.V., Yoshimoto, K., Huang, J.H.Y., Chute, D.J., et al. (2005). Molecular Determinants of the Response of Glioblastomas to EGFR Kinase Inhibitors. N. Engl. J. Med. *353*, 2012–2024.

Miller, J.N., and Pearce, D.A. (2014). Nonsense-Mediated Decay in Genetic Disease: Friend or Foe? Mutat. Res. Rev. Mutat. Res. *762*, 52–64.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Park, H., Garrido-Laguna, I., Naing, A., Fu, S., Falchook, G.S., Piha-Paul, S.A., Wheler, J.J., Hong, D.S., Tsimberidou, A.M., Subbiah, V., et al. (2016). Phase I dose-escalation study of the mTOR inhibitor sirolimus and the HDAC inhibitor vorinostat in patients with advanced malignancy. Oncotarget *7*, 67521–67531.

Pedram, A., Razandi, M., Deschenes, R.J., and Levin, E.R. (2012). DHHC-7 and -21 are palmitoylacyltransferases for sex steroid receptors. Mol. Biol. Cell *23*, 188–199.

Press, M.F., Sauter, G., Bernstein, L., Villalobos, I.E., Mirlacher, M., Zhou, J.-Y., Wardeh, R., Li, Y.-T., Guzman, R., Ma, Y., et al. (2005). Diagnostic Evaluation of HER-2 as a Molecular Target: An Assessment of Accuracy and Reproducibility of Laboratory Testing in Large, Prospective, Randomized Clinical Trials. Clin. Cancer Res. *11*, 6598–6607.

Riaz, M., van Jaarsveld, M.T.M., Hollestelle, A., Prager-van der Smissen, W.J.C., Heine, A.A.J., Boersma, A.W.M., Liu, J., Helmijr, J., Ozturk, B., Smid, M., et al. (2013). miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. Breast Cancer Res. BCR *15*, R33.

Samuels, Y. (2004). High Frequency of Mutations of the PIK3CA Gene in Human Cancers. Science *304*, 554–554.

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D.A., Smith, S., Uziel, T., Sfez, S., et al. (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science *268*, 1749–1753.

Sharma, S.V., and Settleman, J. (2007). Oncogene addiction: setting the stage for molecularly targeted cancer therapy. Genes Dev. *21*, 3214–3231.

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., et al. (2011). Adjuvant Trastuzumab in HER2-Positive Breast Cancer. N. Engl. J. Med. *365*, 1273–1283.

Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A., and McGuire, W.L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science *235*, 177–182.

Sos, M.L., Koker, M., Weir, B.A., Heynck, S., Rabinovsky, R., Zander, T., Seeger, J.M., Weiss, J., Fischer, F., Frommolt, P., et al. (2009). PTEN Loss Contributes to Erlotinib Resistance in EGFR-Mutant Lung Cancer by Activation of Akt and EGFR. Cancer Res. *69*, 3256–3261.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719–724.

Telford, B.J., Chen, A., Beetham, H., Frick, J., Brew, T.P., Gould, C.M., Single, A., Godwin, T., Simpson, K.J., and Guilford, P. (2015). Synthetic Lethal Screens Identify Vulnerabilities in GPCR Signaling and Cytoskeletal Organization in E-Cadherin–Deficient Cells. Mol. Cancer Ther. *14*, 1213–1223.

Treish, I., Schwartz, R., and Lindley, C. (2000). Pharmacology and therapeutic use of trastuzumab in breast cancer. Am. J. Health. Syst. Pharm., *57*, 2063–2076; quiz 2077–2079.

Uhr, K., Smissen, W.J.P. der, Heine, A.A., Ozturk, B., Smid, M., Göhlmann, H.W., Jager, A., Foekens, J.A., and Martens, J.W. (2015). Understanding drugs in breast cancer through drug sensitivity screening. SpringerPlus *4*, 611.

Vilenchik, M., Solit, D., Basso, A., Huezo, H., Lucas, B., He, H., Rosen, N., Spampinato, C., Modrich, P., and Chiosis, G. (2004). Targeting Wide-Range Oncogenic Transformation via PU24FCl, a Specific Inhibitor of Tumor Hsp90. Chem. Biol. *11*, 787–797.

Voigt, W. (2005). Sulforhodamine B assay and chemosensitivity. Methods Mol Med *110*, 39–48.

Xiao, W., Lou, N., Ruan, H., Bao, L., Xiong, Z., Yuan, C., Tong, J., Xu, G., Zhou, Y., Qu, Y., et al. (2017). Mir-144-3p Promotes Cell Proliferation, Metastasis, Sunitinib Resistance in Clear Cell Renal Cell Carcinoma by Downregulating ARID1A. Cell. Physiol. Biochem. *43*, 2420–2433.

Yamasaki, F., Johansen, M.J., Zhang, D., Krishnamurthy, S., Felix, E., Bartholomeusz, C., Aguilar, R.J., Kurisu, K., Mills, G.B., Hortobagyi, G.N., et al. (2007). Acquired Resistance to Erlotinib in A-431 Epidermoid Cancer Cells Requires Down-regulation of MMAC1/PTEN and Up-regulation of Phosphorylated Akt. Cancer Res. *67*, 5779–5788.

Yin, X., Zhang, H., Burrows, F., Zhang, L., and Shores, C.G. (2005). Potent Activity of a Novel Dimeric Heat Shock Protein 90 Inhibitor against Head and Neck Squamous Cell Carcinoma In vitro and In vivo. Clin. Cancer Res. *11*, 3889–3896.

Chapter 6

Association of microRNA-7 and its binding partner *CDR1-AS* with the prognosis and prediction of 1st-line tamoxifen therapy in breast cancer

Katharina Uhr, Anieta M. Sieuwerts, Vanja de Weerd, Marcel Smid, Dora Hammerl, John A. Foekens, John W.M. Martens

Scientific Reports 2018;8(1):9657

Abstract

The large number of non-coding RNAs (ncRNAs) and their breadth of functionalities has fuelled many studies on their roles in cancer. We previously linked four microRNAs to breast cancer prognosis. One of these microRNAs, *hsa-miR-7*, was found to be regulated by another type of ncRNA, the circular non-coding RNA (circRNA) *CDR1-AS,* which contains multiple *hsa-miR-7* binding sites. Based on this finding, we studied the potential clinical value of this circRNA on breast cancer prognosis in a cohort based on a cohort that was previously analysed for *hsa-miR-7* and in an adjuvant hormone-naïve cohort for 1st-line tamoxifen treatment outcomes, in which we also analysed *hsa-miR-7*.

A negative correlation was observed between *hsa-miR-7* and *CDR1-AS* in both cohorts. Despite associations with various clinical metrics (e.g., tumour grade, tumour size, and relapse location), *CDR1-AS* was neither prognostic nor predictive of relevant outcomes in our cohorts. However, we did observe stromal *CDR1-AS* expression, suggesting a possible cell-type specific interaction. Next to the known association of *hsa-miR-7* expression with poor prognosis in primary breast cancer, we found that high *hsa-miR-7* expression was predictive of an adverse response to tamoxifen therapy and poor progression-free and post-relapse overall survival in patients with recurrent disease.

Introduction

Non-coding RNAs (ncRNAs) (The ENCODE Project Consortium, 2012) are defined as genes that are transcribed into RNAs, but not translated into proteins, and are not structural RNAs like tRNA or rRNA (Storz, 2002). MicroRNAs, which regulate mRNA translation (Esteller, 2011), are likely the best studied subgroup of ncRNAs; however, apart from this family is the recently recognised large group of long non-coding RNAs (lncRNAs) (Iyer et al., 2015). These lncRNAs are over 200 nt long (Ma et al., 2013) and have a multitude of mechanisms that affect cellular activity, e.g., (1) by influencing the accessibility of genes to the transcriptional machinery by interacting with chromatin modifiers (Delás and Hannon, 2017); (2) by supporting DNA looping, promoter binding and activator/transcription factor recruitment, which can increase gene expression (Delás and Hannon, 2017; Iyer et al., 2015); (3) by influencing mRNA splicing and mRNA stability and increasing translation (Delás and Hannon, 2017); and (4) by influencing protein phosphorylation, methylation and stability (Delás and Hannon, 2017).

Importantly, it has been recognised that specific lncRNAs can also interfere with microRNA functionality via several binding sites on their sequence, soaking up target microRNAs like a sponge (Hansen et al., 2013). Interestingly, microRNA binding does not necessarily lead to the degradation of a particular lncRNA (Hansen et al., 2013). One recently described lncRNA of this type is *CDR1-AS*, a circular RNA (circRNA) (Hansen et al., 2013), which is also known as *ciRS-7, CDR1as* or *CDR1NAT*, that specifically binds human microRNA-7 (*hsa-miR-7*) via 73 binding sites (Hansen et al., 2013). This interaction is conserved between human and mouse and probably other species as well (Hansen et al., 2013). *CDR1-AS* is generated from a linear transcript, following the general observation that circRNAs are spliced from longer transcripts and it shares a promoter with the lncRNA *LINC00632*, which is about 50x less abundant than the circRNA (Barrett et al., 2017).

CircRNAs are generally characterised as having high stability (transcript half-life), which is likely due to their resistance to exonucleases because of their circular conformation (Jeck et al., 2013). Furthermore, it has been shown that circRNAs show tissue-specific expression (Maass et al., 2017) and are involved in cellular differentiation (Kristensen et al., 2018a) and pluripotency (Yu et al., 2017). CircRNAs have a diverse range of functions, including influencing transcription, splicing and the translation of their host genes; serving as scaffolds for enzymes and substrates to enhance reaction kinetics and co-localization; functioning as protein sponges for RNA-binding proteins that can influence protein decoy; acting as microRNA sponges; and expressing peptides under rare circumstances (Kristensen et al., 2018b). At this point, there are several known mechanisms of circRNA biogenesis from linear transcripts, besides factors influencing their splicing (Kristensen et al., 2018b). It has been shown that reverse-complement ALU repeats are located to the right and left of the circularised sequence in some cases (Kristensen et al., 2018b). Furthermore, splicing factors have been shown to bind in the vicinity of encoded circRNAs, while in other cases, a lariat structure has been observed as a precursor molecule that includes the exon that was later identified to be a circRNA (Kristensen et al., 2018b). Finally, a less common mechanism involves RNA-binding proteins that are located further away from the circRNA, which can be classified as trans-acting factors (Kristensen et al., 2018b).

In regard to expression in malignancies, circRNAs have been found to show abnormal expression in haematological cancers as well as in several solid tumour types, such as colorectal cancer, lung cancer, kidney cancer, liver cancer, bladder cancer, gastric cancer, prostate cancer, CNS tumours, ovarian cancer and breast cancer (Kristensen et al., 2018b).

The multi-facetted microRNA *hsa-miR-7* is known to play roles in the differentiation of the intestinal epithelium (Nguyen et al., 2010) and regulation of β-cell proliferation (Wang et al., 2013b), as well as to influence toll-like receptor 9 growth signalling in lung cancer cells (Xu et al., 2013) and photoreceptor development in Drosophila (Li and Carthew, 2005). *Hsa-miR-7* has also been implied in several cancer types, including lung cancer (Chou et al., 2010), renal cancer (Yu et al., 2013) and colorectal cancer (Zhang et al., 2013).

In 2008, we discovered that *hsa-miR-7* was a prognostic marker in hormone receptor-positive breast cancer (Foekens et al., 2008). Considering the fact that *CDR1-AS* is now recognised to be an *hsa-miR-7* sponge (Hansen et al., 2013), we addressed whether *CDR1-AS* expression was related to *hsa-miR-7* expression and, if so, whether its expression was also associated with breast cancer prognosis. To determine this, we used a cohort of patients with lymph node-negative (LNN) disease who did not receive any type of systemic adjuvant treatment to study the link with pure prognosis, i.e., the natural course of the disease (hereafter also named prognostic cohort). Additionally, to determine whether *CDR1-AS* and/or *hsa-miR-7* expression may have predictive value, we studied their relationship with the response to tamoxifen therapy and the lengths of progression-free survival (PFS) and post-relapse overall survival (PR-OS) in an adjuvant hormone-naïve cohort of patients who received tamoxifen as a 1st-line treatment for recurrent disease (hereafter named predictive cohort).

Results

Cell lines

To determine whether *CDR1-AS* is variably expressed, we initially measured *CDR1-AS* expression in several breast cancer cell lines, a breast tumour pool, an endothelial cell line, a stromal fibroblast strain, and several sorted immune cells of healthy donors (for included specimens, see **Supplementary Fig. 1**). This analysis showed that *CDR1-AS* was differentially expressed among the breast cancer cell line subtypes (see **Supplementary Table S1**), with higher *CDR1-AS* levels in the luminal and normal-like subtypes (**Fig. 1**). *CDR1-AS* expression was not only confined to breast cancer cells but was also detectable in other cell types present in tumours, as shown by the *CDR1-AS* expression in the endothelial cell line, the T cells, the granulocytic NK cells and the fibroblast strain, with especially high expression levels in the fibroblast strain derived from a patient's stroma (**Fig. 1**). Furthermore, *CDR1-AS* was not only expressed by the cultured cell lines but was also present in a pooled sample containing cDNA from 100 different breast tumours.

Altogether, the variable *CDR1-AS* expression provided a rationale to proceed with our study on the clinical relevance of this circRNA in our two breast cancer cohorts.

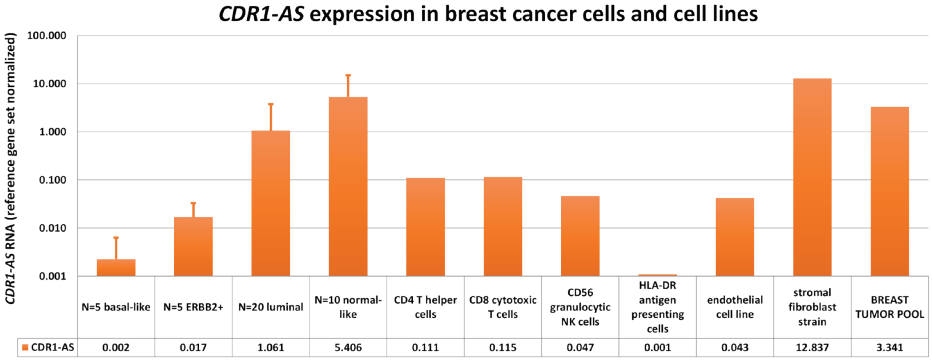


Figure 1. *CDR1-AS* expression in breast cancer cells and cell lines. Normalised *CDR1-AS* expression is shown in different breast cancer cell line subtypes (basal-like, ERBB2+, luminal and normal-like) on a logarithmic scale. Variations in expression levels within one subtype group are displayed as the standard deviation using error bars. N indicates the number of cell lines assessed per subtype group. Aside from the breast cancer cell lines, the expression is also shown for several immune cell types (CD4+ T helper cells, CD8+ cytotoxic T cells, CD56+ granulocytic NK cells and HLA-DR+ APCs), an endothelial cell line, a stromal fibroblast strain and a sample consisting of pooled material from different breast tumours. The HLA-DR+ APCs were from donor A, all other immune cells are from donor B. Of note is that the CD8+ T cells unfortunately showed a lower purity (61%) in contrast to the other immune cells (75%-96% purity). The measured expression value is listed below each sample type.

Hsa-miR-7 and *CDR1-AS* expression and prognosis

The first cohort evaluated for *CDR1-AS* expression consisted of RNA isolated from primary tumours from LNN breast cancer patients, including oestrogen receptor (ER, *ESR1*)-positive and ER-negative patients, who did not receive systemic adjuvant therapy (n = 345) (see the REMARK diagram in **Supplementary Fig. 1**) (prognostic cohort). This cohort included some of the ER-positive breast cancers in which we originally discovered that high *hsa-miR-7* expression was a marker for worse prognosis (Foekens et al., 2008) (see **Supplementary Table S2** and **Supplementary Fig. 1** for details). The patient overlap with our previous study was substantial (244 out of the n = 345 samples included in this study). The number of patients included in this study differed from our previous study because of the inclusion of additional available eligible patients (n = 101) as well as the exclusion of patients due to the lack of RNA or available qPCR data (n = 55). Furthermore, due to the differences in the ER expression measurement methods (mRNA- versus protein-based), the classification of some patients differed between the two studies (see **Supplementary Table S2**). As *CDR1-AS* has been shown to counteract *hsa-miR-7* (Hansen et al., 2013), we determined the association between *hsa-miR-7* and *CDR1-AS* in the 345 primary breast tumours from this prognostic cohort (**Table 1**). We observed an inverse correlation with *CDR1-AS* expressed at higher levels in the tumours with lower *hsa-miR-7* expression levels (Spearman rs -0.245, $P < 0.0001$).

Table 1. Associations of *hsa-miR-7* with *CDR1-AS* and clinical parameters in the prognostic patient cohort

		CDR1-AS		***hsa-mir-7***	
Parameters	**n**	**median expression [IQR]**	***P***	**median expression [IQR]**	***P***
All patients	345	4.97 [8.74]		0.034 [0.053]	
***hsa-miR-7* expression**					
1st quantile (low)	173	7.16 [10.02]			
2nd quantile (high)	172	3.52 [6.78]	***< 0.001***		
Age at time of surgery (years)					
≤40	38	3.37 [8.88]		0.039 [0.044]	
>40-≤ 55	134	5.09 [7.84]		0.032 [0.057]	
>55-≤ 70	117	5.65 [9.04]		0.033 [0.050]	
>70	56	4.46 [8.26]	*0.37*	0.047 [0.063]	*0.58*
Menopausal status					
Premenopausal	153	4.69 [8.31]		0.033 [0.057]	
Postmenopausal	192	5.19 [8.90]	*0.84*	0.035 [0.057]	*0.74*
Pathological tumour size					
pT1	161	6.95 [9.31]		0.029 [0.042]	
pT2 + unknown	171	3.97 [7.14]		0.037 [0.066]	
pT3+pT4	13	3.66 [13.7]	***0.013***	0.055 [0.032]	***0.024***
Tumour grade					
Poor	197	4.72 [8.07]		0.043 [0.062]	
Unknown	95	3.58 [7.87]		0.033 [0.051]	
Moderate/Good	53	8.63 [7.87]	***0.001***	0.023 [0.027]	***<0.001***
Tumour cell content					
30-70%	230	6.10 [9.38]		0.033 [0.053]	
>70%	115	3.48 [7.78]	***< 0.001***	0.042 [0.061]	*0.08*

Table 1. Continued

		CDR1-AS		*hsa-mir-7*	
Parameters	**n**	**median expression [IQR]**	***P***	**median expression [IQR]**	***P***
Hormone receptor/growth factor status (RT-qPCR)*					
ESR1-negative	120	4.21 [7.84]		0.046 [0.062]	
ESR1-positive	225	5.41 [8.92]	*0.10*	0.030 [0.047]	***0.007***
PGR-negative	158	4.21 [7.84]		0.046 [0.066]	
PGR-positive	187	5.41 [9.47]	*0.08*	0.028 [0.045]	***<0.001***
ERBB2-non-amplified	294	5.49 [9.07]		0.033 [0.054]	
ERBB2-amplified	51	3.97 [6.17]	*0.22*	0.039 [0.059]	*0.43*
EGFR quantile 1 (low)	115	2.84 [4.85]		0.042 [0.062]	
EGFR quantile 2 (in between)	116	7.61 [9.61]		0.031 [0.059]	
EGFR quantile 3 (high)	114	6.24 [10.12]	***<0.001***	0.032 [0.046]	***0.042***

For the analysis on the association between *hsa-miR-7* expression and *CDR1-AS* expression, patients were divided into two equally sized groups based on median *hsa-miR-7* expression level. Next to this analysis result, the associations of the clinical parameters with *hsa-miR-7* and *CDR1-AS* gene expression are listed. The median *CDR1-AS* and *hsa-miR-7* expression levels per subcategory, including the interquartile range (IQR) are given additionally. Pathological tumour size is defined as follows: pT1 <= 2 cm, pT2 > 2 cm and <= 5 cm, pT3 > 5 cm, and pT4 = tumour with direct extension to chest wall and/or skin. *Cut-offs for positive and negative hormone receptor/growth factor status established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). *P* = p-value and n = number of patients. Significant p-values are printed in bold.

Association of hsa-miR-7 and CDR1-AS with clinical variables

We next assessed whether *CDR1-AS* expression associated with the same traditional prognostic clinical parameters as *hsa-miR-7* did. No association was found for both markers with age, menopausal status or *ERBB2* status, while only *hsa-miR-7* was negatively associated with *ESR1* and progesterone receptor (*PGR*). By contrast, both markers were associated, although in opposite directions, with tumour size and tumour grade, indicating that high *CDR1-AS* levels were associated with good prognosis characteristics and high *hsa-miR-7* levels with poor prognosis characteristics. Additionally, *CDR1-AS* expression was found to be higher in tumours with a tumour cell content below 70%, i.e., tumours with a higher number of immune or stromal cells (**Table 1**).

A well-known target of *hsa-miR-7* is the epidermal growth factor receptor (*EGFR*) (Webster et al., 2009). We therefore also assessed whether *CDR1-AS* might show a positive association with *EGFR* mRNA expression, and this was the case. The results showed higher *CDR1-AS* (and lower *hsa-miR-7*) expression in *EGFR*-high expressing tumours (**Table 1**).

Hsa-miR-7 and CDR1-AS in metastasis-free survival (MFS) and overall survival (OS) in the prognostic patient cohort

We next studied whether the length of metastasis-free survival time (MFS) was correlated with *hsa-miR-7* and/or *CDR1-AS* expression. In the Cox univariate regression analysis, high *hsa-miR-7* levels were associated with poor prognosis in this cohort, including 225 ER-positive and 120 ER-negative patients, while *CDR1-AS* expression was not associated (**Table 2**).

Table 2. Univariate association of *CDR1-AS* and *hsa-miR-7* with MFS in the prognostic cohort

Clinical parameters	*CDR1-AS*				*hsa-miR-7*			
	n	HR	(95% CI)	*P*	n	HR	(95% CI)	*P*
All patients	345	1.02	(0.90-1.16)	*0.76*	345	1.14	(1.00-1.30)	***0.043***
Hormone receptor/growth factor expression (RT-qPCR)*								
ESR1-negative	120	1.10	(0.83-1.45)	*0.50*	120	1.03	(0.77-1.37)	*0.86*
ESR1-positive	225	0.98	(0.84-1.12)	*0.31*	225	1.21	(1.06-1.38)	***0.006***

The association of *CDR1-AS*/*hsa-miR-7* expression and MFS in all patients and the patient subgroups based on *ESR1* expression is displayed. *Cut-offs for positive and negative hormone receptor/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). n = number of patients, HR = hazard ratio, CI = confidence interval, and *P* = p-value. Significant p-values are printed in bold.

As ER-positive and ER-negative breast cancers can be considered two different diseases (due to different treatment requirements (Hammond et al., 2010) and profound differences in gene expression (Perou et al., 2000)), we also evaluated whether our genes of interest showed an association with MFS in the respective ER subgroups. High *hsa-miR-7* levels were strongly associated with a shorter MFS in ER-positive patients, but not in ER-negative patients, while *CDR1-AS* expression was not associated with the length of MFS in either ER subgroup (**Table 2**). Similar to our previous study (Foekens et al., 2008), *hsa-miR-7* (split at the median expression level) remained an independent biomarker for poor prognosis among ER-positive patients after correcting for the traditional prognostic factors: age, menopausal status, tumour size, grade, *ESR1*, *PGR* and *ERBB2* status (HR = 1.53, 95% CI 1.02-2.31, *P* = 0.04). Finally, we also related *hsa-miR-7* and *CDR1-AS* expression to the length of overall survival time (OS) in the complete prognostic cohort. Using Cox univariate regression analysis, among the two genes, only *hsa-miR-7* expression was associated with the OS length, with high *hsa-miR-7* levels being predictive of an earlier death (**Table 3**). However, this finding was not significant in the multivariate analysis (*P* > 0.05) (**Table 3**). In the ER-positive subset *hsa-miR-7* was associated with OS, a finding which remained significant in the multivariate analysis, including the traditional clinical factors (HR = 1.19, 95% CI 1.01-1.42, *P* = 0.041) (cohort size was n = 225 in the univariate analysis and n = 213 in the multivariate analysis due to the lack of some clinical data).

Table 3. Uni- and multivariate association of overall survival (OS) with *CDR1-AS*, *hsa-miR-7* and other clinical variables in the prognostic cohort

Parameters	univariate model n = [345]				multivariate model [n = 331]			
	n	HR	(95% CI)	*P*	n**	HR	(95% CI)	*P*
	345							
Age at time of surgery (years)								
≤40 years	45	1			43	1		
41-50 years	96	0.94	(0.57-1.54)	*0.80*	94	1.10	(0.64-1.87)	*0.74*
51-70 years	159	1.14	(0.72-1.80)	*0.58*	152	1.29	(0.60-2.75)	*0.51*
>70 years	45	0.93	(0.49-1.77)	*0.83*	42	0.97	(0.37-2.52)	*0.95*
Menopausal status								
Premenopausal	153	1			149	1		
Postmenopausal	192	1.11	(0.82-1.51)	*0.50*	182	0.93	(0.49-1.78)	*0.84*
Pathological tumor size**								
≤2 cm	155	1			155	1		
>2 cm	176	1.30	(0.95-1.78)	*0.10*	176	1.31	(0.94-1.83)	*0.11*
Tumour grade								
Poor	197	1			194	1		
Unknown	95	1.01	(0.71-1.44)	*0.94*	85	0.75	(0.64-1.38)	*0.75*
Moderate/ Good	53	0.74	(0.47-1.18)	*0.21*	52	0.58	(0.53-1.42)	*0.58*
Hormone/growth factor receptors*								
ER (log continuous *ESR1*)	345	0.99	(0.94-1.05)	*0.88*	331	1.08	(0.99-1.18)	*0.08*
PR (log continous *PGR*)	345	0.95	(0.89-1.00)	*0.07*	331	0.90	(0.82-0.98)	***0.020***
HER2 (*ERBB2*) status								
Not amplified	294	1			280	1		
Amplified	51	1.44	(0.97-2.14)	*0.07*	51	1.51	(1.00-2.27)	***0.048***
					***hsa-miR-7* separately added to the base model**			
***CDR1-AS* (log continuous)**	345	1.03	(0.90-1.17)	*0.69*				
***hsa-miR-7* (log continuous)**	345	1.16	(1.02-1.32)	***0.028***	331	1.12	(0.97-1.29)	*0.11*
≤median	173	1			165	1		
>median	172	1.42	(1.04-1.92)	***0.025***	166	1.37	(0.98-1.92)	*0.07*

The association of *CDR1-AS* and *hsa-miR-7* gene expression with overall survival is displayed for the prognostic cohort. The univariate model is on the left side of the table and the multivariate model is displayed on the right side. *Cut-offs for positive and negative hormone receptor/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). **In these analyses, clinical data are missing, and therefore, the included patients do not add up to 345. n = number of patients, HR = hazard ratio, CI = confidence interval, *P* = p-value. Significant p-values are highlighted in bold.

***Hsa-miR-7 and CDR1-AS* in the recurrent breast cancer cohort**

Based on the prognostic value of *hsa-miR-7* in the ER-positive patients, we were interested to assess whether the two ncRNAs could be predictive factors for the type of response to 1st-line tamoxifen therapy, PFS or PR-OS in recurrent disease. For these analyses, we had data available from 188 hormone-naïve breast cancer patients with ER-positive primary tumours who received tamoxifen as 1st-line therapy for recurrent disease (predictive cohort). Among this cohort, 39 patients had been included in our previous study (Foekens et al., 2008) (see also **Supplementary Fig. 1**).

Association of hsa-miR-7 and CDR1-AS with clinical parameters in the predictive cohort

As 91 patients in this cohort were also included in our prognostic cohort (discussed above; see also **Supplementary Fig. 1**) and the overlap is therefore substantial, we will only discuss associations with clinical parameters that have not been discussed for the prognostic cohort or show a different result (all associations assessed are listed in **Table 4**). The *hsa-miR-7* and *CDR1-AS* expression levels were once again inversely correlated with one another (Spearman rs -0.277, $P < 0.0001$). In the analysis with the clinical parameters, *hsa-miR-7* was significantly positively associated with tumour cell content. In the predictive cohort, the pathological tumour size was not significantly associated with the two markers; similar results were noted for the nodal status with *CDR1-AS* and *PGR* expression with *hsa-miR-7.* However, the variable "dominant site of relapse" showed an association with *CDR1-AS* expression with the lowest expression in the primary tumours metastasising to distant sites other than local regional or bone.

Table 4. Associations of *CDR1-AS* and *hsa-miR-7* with clinical parameters in the predictive patient cohort

Parameters	n	*CDR1-AS* median expression [IQR]	*P*	*hsa-mir-7* median expression [IQR]	*P*
All patients with *ESR1*-positive primary tumours	188	4.45 [6.42]		0.042 [0.077]	
Age in years at start of 1st-line treatment					
≤55	75	4.52 [5.82]		0.042 [0.083]	
56-70	67	3.84 [5.40]		0.040 [0.067]	
>70	46	4.35 [7.79]	*0.65*	0.045 [0.076]	*1.00*
Menopausal status at start of 1st-line treatment					
Premenopausal	53	4.69 [5.68]		0.046 [0.099]	
Postmenopausal	135	4.36 [6.91]	*0.4*	0.042 [0.069]	*0.66*
Adjuvant chemotherapy					
No	153	4.35 [6.87]		0.045 [0.086]	
CMF	21	4.82 [4.44]		0.033 [0.032]	
Anthracycline containing	14	4.36 [4.41]	*0.24*	0.047 [0.080]	*0.18*
Tumour cell content					
30-70%	111	4.98 [6.77]		0.033 [0.077]	
>70%	77	2.55 [5.22]	***<0.001***	0.049 [0.079]	***0.025***
Pathological tumour size					
pT1	53	4.85 [7.77]		0.038 [0.059]	
pT2 + unknown	116	4.39 [6.23]		0.040 [0.082]	
pT3+pT4	19	4.55 [4.79]	*0.63*	0.078 [0.225]	*0.16*
Tumour grade					
Poor	104	3.90 [5.35]		0.048 [0.084]	
Unknown	57	4.12 [6.71]		0.040 [0.067]	
Moderate/Good	27	6.07 [10.42]	*0.093*	0.034 [0.073]	*0.35*
Nodal status					
Negative	91	5.00 [7.49]		0.045 [0.068]	
Positive	67	3.73 [6.08]		0.034 [0.068]	
Positive (tumour outside lymph nodes)	30	3.52 [4.93]	*0.085*	0.077 [0.275]	***0.024***
Hormone/growth factor receptors*					
PGR-negative	35	3.31 [8.08]		0.060 [0.235]	
PGR-positive	153	4.47 [6.05]	*0.26*	0.038 [0.073]	*0.21*
ERBB2-non-amplified	165	4.42 [6.65]		0.044 [0.086]	
ERBB2-amplified	23	4.66 [5.48]	*0.54*	0.029 [0.047]	*0.14*

Table 4. Continued

Parameters	n	*CDR1-AS* median expression [IQR]	*P*	*hsa-mir-7* median expression [IQR]	*P*
Disease-free interval before start 1st-line tamoxifen					
≤1 year	76	4.67 [6.35]		0.046 [0.165]	
1 - 3 years	48	3.58 [7.16]		0.048 [0.074]	
>3 years	64	4.62 [6.33]	*0.96*	0.034 [0.060]	*0.61*
Dominant site of relapse					
Local regional	21	4.96 [10.53]		0.029 [0.021]	
Bone	101	4.90 [6.48]		0.044 [0.084]	
Other distant metastasis	66	3.43 [5.34]	***0.028***	0.050 [0.086]	*0.33*

Clinical parameters were analysed for their association with the gene expression of the two genes of interest; the median expression per gene including the interquartile range (IQR) and the number of patients per analysis (n) is shown. Pathological tumour size is defined as follows: pT1 <= 2 cm, pT2 > 2 cm and <= 5 cm, pT3 > 5 cm, and pT4 = tumour with direct extension to chest wall and/or skin. *Cut-offs for positive and negative hormone receptor/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). *P* = p-value. Significant p-values are printed in bold.

Hsa-miR-7, CDR1-AS and the response to 1st-line tamoxifen therapy in the predictive cohort

We also investigated whether *hsa-miR-7* or *CDR1-AS* expression levels in the primary tumours had an influence on the type of response to tamoxifen in the predictive cohort (**Table 5**). To this end, we evaluated the contribution of our two markers using logistic regression analysis in a univariate model as well as a multivariate model. The latter model contained age at start of the 1st-line tamoxifen therapy, disease-free interval, dominant site of relapse, *ESR1* and *PGR* expression levels and *ERBB2* status.

Table 5. Uni- and multivariate association of response with *CDR1-AS*, *hsa-miR-7* and other clinical variables in the predictive cohort

Parameters base model	univariate model				multivariate model		
	n	OR	(95% CI)	*P*	OR	(95% CI)	*P*
	188						
Age at start 1st-line tamoxifen							
≤55 years	75	1			1		
56-70 years	67	1.63	(0.82-3.25)	*0.16*	1.23	(0.55-2.72)	*0.61*
>70 years	46	3.06	(1.29-7.23)	***0.011***	2.05	(0.81-5.18)	*0.13*

Table 5. Continued

Parameters base model	univariate model				multivariate model		
	n	OR	(95% CI)	*P*	OR	(95% CI)	*P*
Disease-free interval							
≤1 year	76	1			1		
1- 3 years	48	1.52	(0.71-3.25)	*0.28*	1.44	(0.63-3.30)	*0.38*
>3 years	64	2.07	(0.99-4.28)	*0.05*	1.88	(0.86-4.14)	*0.12*
Dominant site of relapse							
Local regional	21	1			1		
Bone	101	0.25	(0.07-0.92)	***0.037***	0.20	(0.05-0.77)	***0.020***
Other distant metastasis	66	0.41	(0.11-1.56)	*0.19*	0.28	(0.07-1.19)	*0.08*
Hormone/growth factor receptors*							
ER (log continuous *ESR1*)	188	1.65	(1.22-2.23)	***0.001***	1.65	(1.15-2.36)	***0.006***
PR (log continuous *PGR*)	188	1.07	(0.92-1.25)	*0.39*	1.00	(0.84-1.20)	*0.99*
HER2 (*ERBB2*) status							
Not amplified	165	1			1		
Amplified	23	0.60	(0.25-1.45)	*0.26*	0.97	(0.33-2.85)	*0.96*
					***hsa-miR-7* separately added to the base model**		
***CDR1-AS* (log continuous)**	188	1.01	(0.78-1.31)	*0.96*			
***hsa-miR-7* (log continuous)**	188	0.76	(0.60-0.96)	***0.023***	0.73	(0.57-0.95)	***0.017***
≤median	94	1			1		
>median	94	0.56	(0.30-1.03)	*0.064*	0.59	(0.30-1.16)	*0.13*

The association of the different parameters, including *CDR1-AS* and *hsa-miR-7* expression, with the therapy response in the predictive cohort is displayed on the left side of the table. The hormone/growth factor receptor expression values used in this table were determined by qPCR. *Cut-offs for positive and negative hormone receptor/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). The same parameters are assessed with a multivariate model on the right side. n = number of patients, OR = odds ratio, CI = confidence interval, and *P* = p-value. Significant p-values are highlighted in bold.

We realised that the hazard ratios of the patients among the pre-menopausal patient age groups (≤55 years) were similar to one another, and the same was true for the hazard ratios for the post-menopausal patient age groups (>55 years). Due to this, the menopausal status was largely covered by the age groups, and due to the low patient numbers (n = 188) included in this cohort, we did not include menopausal status but rather the age categories in our analyses for the predictive cohort.

In the univariate analysis for response, only *hsa-miR-7* was a predictive biomarker, namely for a poor response to treatment, a finding which remained significant in the multivariate analysis (odds ratio 0.73, 95% CI 0.57-0.95, *P* = 0.017).

Hsa-miR-7, CDR1-AS and PFS in the predictive cohort

We subsequently investigated whether *hsa-miR-7* or *CDR1-AS* expression could predict the length of PFS (**Table 6**). In the Cox regression analysis, we evaluated the contribution of our two markers in a univariate model as well as in a multivariate model containing the same additional parameters as used for the response analysis.

Table 6. Uni- and multivariate association of *CDR1-AS*, *hsa-miR-7* and other clinical variables with PFS in the predictive cohort

Parameters base model	univariate model				multivariate model		
	n	HR	(95% CI)	*P*	HR	(95% CI)	*P*
	188						
Age at start 1st-line tamoxifen							
≤55 years	75	1			1		
56-70 years	67	0.76	(0.54-1.07)	*0.11*	0.78	(0.54-1.13)	*0.19*
>70 years	46	0.57	(0.38-0.84)	***0.005***	0.65	(0.43-0.98)	***0.039***
Disease-free interval							
≤1 year	76	1			1		
1- 3 years	48	0.79	(0.54-1.14)	*0.21*	0.83	(0.56-1.22)	*0.34*
>3 years	64	0.68	(0.48-0.97)	***0.032***	0.74	(0.51-1.05)	*0.09*
Dominant site of relapse							
Local regional	21	1			1		
Bone	101	1.60	(0.96-2.65)	*0.071*	1.99	(1.16-3.43)	***0.013***
Other distant metastasis	66	1.34	(0.78-2.27)	*0.28*	2.00	(1.10-3.63)	***0.023***
Hormone/growth factor receptors*							
ER (log continuous *ESR1*)	188	0.77	(0.67-0.89)	***<0.001***	0.82	(0.69-0.96)	***0.016***
PR (log continuous *PGR*)	188	0.89	(0.83-0.96)	***0.002***	0.92	(0.84-0.96)	***0.038***
HER2 (*ERBB2*) status							
Not amplified	165	1			1		
Amplified	23	1.86	(1.19-2.90)	***0.007***	1.49	(0.89-2.50)	*0.13*
					***hsa-miR-7* separately added to the base model**		
***CDR1-AS* (log continuous)**	188	0.97	(0.86-1.10)	*0.62*			
***hsa-miR-7* (log continuous)**	188	1.18	(1.04-1.33)	***0.008***	1.22	(1.08-1.38)	***0.002***
≤median	94	1			1		
>median	94	1.54	(1.14-2.09)	***0.005***	1.63	(1.18-2.25)	***0.003***

This table shows the associations of PFS with the cohort characteristics and *CDR1-AS* and *hsa-miR-7* expression. The different metrics were examined with a univariate analysis on the left side of the table, while they were analysed with a multivariate model on the right side. n = number of patients, HR = hazard ratio, CI = confidence interval, and *P* = p-value. Significant p-values are printed in bold. The hormone/growth factor receptor expression values used in this table were determined by qPCR. *Cut-offs for positive and negative hormone/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007).

In the Cox univariate regression analysis, *hsa-miR-7*, but not *CDR1-AS*, was associated with the length of PFS and remained an independent predictive marker for poor PFS in the multivariate analysis (log continuous variable: HR = 1.22, 95% CI 1.08-1.38, *P* = 0.002; split on median: HR = 1.63, 95%CI 1.18-2.25, *P* = 0.003). The association of *hsa-miR-7* with PFS is visualised by a Kaplan-Meier plot after dividing our cohort into *hsa-miR-7*-high and -low subgroups based on the median expression level (**Fig. 2**).

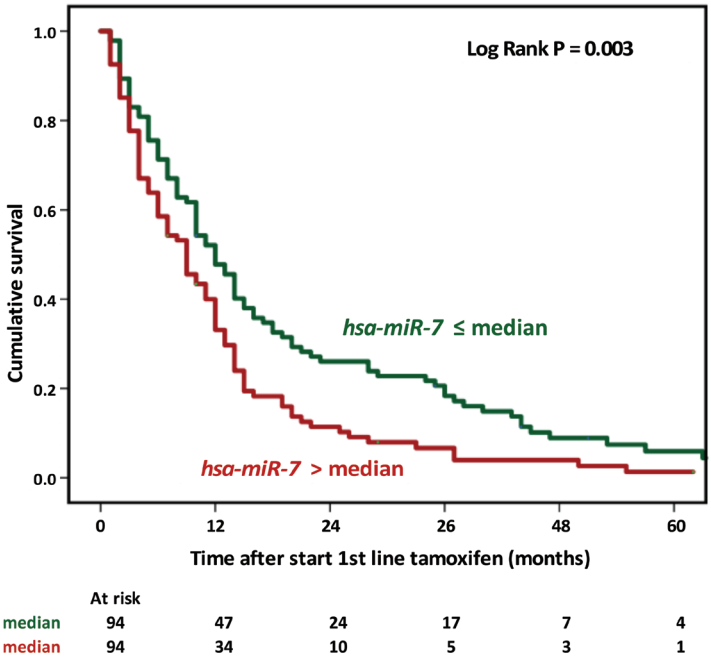


Figure 2. *Hsa-miR-7* as a function of progression-free survival. The figure shows *hsa-miR-7* expression split into two expression categories and its relationship with PFS in the predictive cohort in a Kaplan-Meier graph. The log rank test was applied to assess the significance and the obtained p-value is listed within the plot.

Hsa-miR-7 and CDR1-AS in regard to PR-OS in the predictive cohort

Finally, we related our markers to PR-OS after the beginning of tamoxifen therapy. While *CDR1-AS* did not show an association with PR-OS in the Cox univariate regression analysis, we did find an association between *hsa-miR-7* expression and PR-OS, which remained significant in the multivariate analysis (HR = 1.22, 95% CI 1.06-1.40, *P* = 0.004; split on median: HR = 1.64, 95% CI 1.17-2.29, *P* = 0.004 (**Table 7**).

Table 7. Uni- and multivariate association of post-relapse overall survival (PR-OS) with *CDR1-AS*, *hsa-miR-7* and other clinical variables in the predictive cohort

Parameters		univariate model			multivariate model		
	n	HR	(95% CI)	*P*	HR	(95% CI)	*P*
	188						
Age at start 1st-line tamoxifen							
≤55 years	75	1			1		
56-70 years	67	1.04	(0.72-1.48)	*0.85*	0.98	(0.65-1.47)	*0.93*
>70 years	46	0.94	(0.62-1.45)	*0.79*	1.10	(0.70-1.73)	*0.69*
Dominant site of relapse							
Local regional relapse	21	1			1		
Bone metastasis	101	1.63	(0.94-2.83)	*0.08*	1.61	(0.92-2.84)	*0.10*
Other distant metastasis	66	1.55	(0.87-2.75)	*0.14*	1.81	(0.99-3.31)	*0.06*
Disease-free interval							
≤1 year	76	1			1		
1-3 years	48	0.83	(0.56-1.22)	*0.34*	0.93	(0.62-1.41)	*0.75*
>3 years	64	0.56	(0.38-0.82)	***0.003***	0.56	(0.38-0.85)	***0.006***
Hormone/growth factor receptors*							
ER (log continuous *ESR1*)	188	0.80	(0.68-0.93)	***0.005***	0.79	(0.66-0.95)	***0.012***
PR (log continuous *PGR*)	188	0.88	(0.81-0.95)	***0.001***	0.90	(0.82-0.98)	***0.020***
HER2 (*ERBB2*) status							
Not amplified	165	1			1		
Amplified	23	1.45	(0.91-2.30)	*0.12*	1.15	(0.69-1.93)	*0.59*
					***hsa-miR-7* separately added to the base model**		
***CDR1-AS* (log continuous)**	188	0.97	(0.83-1.13)	*0.72*			
***hsa-miR-7* (log continuous)**	188	1.22	(1.07-1.39)	***0.004***	1.22	(1.06-1.40)	***0.004***
≤median	94	1			1		
>median	94	1.70	(1.23-2.35)	***0.001***	1.64	(1.17-2.29)	***0.004***

The association of *CDR1-AS* and *hsa-miR-7* expression and other clinical variables with PR-OS using the univariate model is shown on the left side and with the multivariate model on the right side. The hormone receptor/growth factor expression values used in this table were determined by qPCR. *Cut-offs for positive and negative hormone receptor/growth factor status were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007). n = number of patients, HR = hazard ratio, CI = confidence interval, and *P* = p-value. Significant p-values are highlighted in bold.

Discussion

Our study showed among the first results an inverse relationship between *CDR1-AS* and *hsa-miR-7* expression. This finding is supported by previous studies in different cancer patient cohorts (Weng et al., 2017; Xu et al., 2017). Besides, we observed an inverse relationship between *hsa-miR-7* and *EGFR*. This observation is in line with the known function of *hsa-miR-7* to target the *EGFR* mRNA 3' UTR and cause the downregulation of *EGFR* mRNA in this way (Webster et al., 2009). Due to the inverse relationship of *hsa-miR-7* with *CDR1-AS, CDR1-AS* can be expected to be expressed at higher levels in *EGFR*-high tumours, and this was the case in our dataset.

One point of caution is, however, that *EGFR* is not very highly expressed in most breast tumour cells (Kersting et al., 2006; Möller et al., 1989; Pilichowska et al., 1997). As tumours consist not only of tumour cells, but also of stroma and immune cells (Allinen et al., 2004), *CDR1-AS* might be mainly expressed by non-tumour cell types. Supporting this hypothesis, we found that *CDR1-AS* was associated with tumours with a lower tumour cell content, and in our cell line expression study, *CDR1-AS* expression was amongst the highest in the stromal fibroblast strain (**Fig. 1**). However, higher expression in immune cells infiltrating the tumours (tumour-infiltrating leukocytes; TILs) could also be a possible explanation. Therefore, we assessed several immune cells (CD4+ T helper cells, CD8+ T helper cells, CD56+ granulocytic NK cells and HLA-DR+ antigen-presenting cells (APC)) for *CDR1-AS* expression in a healthy donor and found low *CDR1-AS* expression. To further examine the relationship of *CDR1-AS* expression with stroma, we performed correlations for *hsa-miR-7* and *CDR1-AS* with several stromal markers (Allinen et al., 2004; Ma et al., 2009; Riaz et al., 2012; van Roozendaal et al., 1996), a newly developed stromal gene expression signature and our developed TIL signature, which is a previously published gene expression signature for tumour-infiltrating leukocytes (Massink et al., 2015; Smid et al., 2016) (see also **Supplementary Methods**). The TIL signature was established on breast tumours and was developed based on differential gene expression between breast tumours with pathologically high and low TIL scores and validated on an large independent cohort (data not shown) (Massink et al., 2015; Smid et al., 2016). The correlation analyses showed that tumour cell content was strongly negatively correlated with the TIL signature (Spearman rs = -0.428; *P* = 2.47E-12), but showed slightly lower negative correlations with the set of individual stromal markers and the stromal signature (**Supplementary Table S3**). This indicates that the majority of non-tumour cells in our tumour samples were mainly TILs and not stromal cells. That said, in our cohort, *CDR1-AS* correlated very strongly positively with the stromal signature (Spearman rs = 0.674, *P* = 8.30E-34) and the stromal markers (Spearman rs = 0.25 up

to 0.53; all *P* < 1E-05; see **Supplementary Table S3** for details), but not with the TIL signature (*P* = 0.16). Based on these results we conclude that *CDR1-AS* appears to be mainly expressed by stromal cells and not by TILs or tumour cells.

Hsa-miR-7, was also not associated with the TIL signature. Furthermore, *hsa-miR-7* was correlated negatively with the stromal signature as well as with the analysed stromal markers. In contrast *hsa-miR-7* showed a positive association with the tumour cell content in the primary tumours from our recurrent breast cancer cohort (predictive cohort). These associations support the expression of *hsa-miR-7* mainly by tumour cells.

Based on our correlative findings, it could be speculated that *CDR1-AS*' role is to prevent *hsa-miR-7* from being active in the stroma but future studies exploring this result in more detail are needed.

In our prognostic cohort, we observed that *hsa-miR-7* expression was associated with tumour size, tumour grade, ER status and PR status. Looking at the expression levels of this microRNA, it displayed higher expression in the prognostic-wise worse parameter categories, such as pT3 and pT4 size status, poor tumour grade, ER-negative tumours and PR-negative tumours. Based on these findings, *hsa-miR-7* appears to be associated with a generally more aggressive tumour type. Indeed, with respect to the prognosis of breast cancer patients with LNN disease who did not receive systemic adjuvant therapy, we confirmed a shorter MFS for patients with higher *hsa-miR-7* levels in ER-positive tumours, which is in line with our previous findings in a subset from this cohort (Foekens et al., 2008). Surprisingly, although overall *hsa-miR-7* expression was higher in ER-negative tumours, this microRNA was not associated with prognosis in this subgroup. However, the microRNA was related to prognosis in ER-positive tumours, indicating that these tumours, despite being ER-positive, have more aggressive features similar to ER-negative tumours.

While *CDR1-AS* did not appear to correlate with either ER or PR expression, we did observe several associations for *CDR1-AS* with clinical parameters in our LNN cohort (e.g., tumour size and grade) which were inverse to those found for *hsa-miR-7.* Nevertheless, our data showed that the expression of this circRNA is not significantly (*P* > 0.05) related to relevant parameters such as hormone receptor expression or clinical outcomes; therefore, it is likely that factors other than sequestering *hsa-miR-7* play a more prominent role here.

In our cohort of patients who received 1st-line tamoxifen treatment, we observed that high *hsa-miR-7* levels were associated with a shorter time to progression, which remained significant after adjustment for known predictive factors in breast cancer. This finding indicates that high *hsa-miR-7* levels are an independent predictor for shorter PFS. Furthermore, *hsa-miR-7* was also associated with a worse clinical response to 1st-line tamoxifen therapy. These observations not only confirm the general notion that *hsa-miR-7* is a marker of more aggressive tumours but also that *hsa-miR-7* is a predictor of poor tamoxifen therapy efficacy, irrespective of the intrinsically aggressive tumour type.

In addition to the associations of *hsa-miR-7* with therapy response and progression, we also found an association with survival. The association with OS was weak in our prognostic cohort and only present in the ER-positive subset (P = 0.041), while in the predictive cohort, *hsa-miR-7* was stronger associated with PR-OS (P = 0.004). These findings exemplify the potential of this microRNA as a biomarker in breast cancer.

Regarding the regulation of *CDR1-AS* and *hsa-miR-7* expression, it has been found that the upregulation of c-Myc causes *CDR1-AS* downregulation in B cells, although it remains unclear whether c-Myc can bind the *CDR1-AS* promoter region directly (Gou et al., 2017). Interestingly, it has been shown that c-Myc upregulates *hsa-miR-7* expression (Chang et al., 2008; Chou et al., 2010). This might indicate an additional level of regulation for both genes by c-Myc and the subsequent fine tuning of the RNA levels for these two genes through their interplay. One could hypothesise that *hsa-miR-671*, a negative regulator of *CDR1-AS* (Hansen et al., 2011), could be upregulated by c-Myc and in this way also downregulate *CDR1-AS* levels; however this does not seem to be the case as *hsa-miR-671* levels have been investigated upon c-Myc induction and no change in expression was found (Liao and Lu, 2011).

Recently, several studies showed an association of *CDR1-AS* with clinical outcomes in different cancer types. Two studies reported that the circRNA was associated with poor survival in colorectal cancer patients (Tang et al., 2017; Weng et al., 2017), a cancer-type in which low *hsa-miR-7* expression has previously been associated with poor survival (Suto et al., 2015). Interestingly there is a molecular subtype in colorectal cancer (CMS4) that is characterised by stromal invasion and has been linked to a worse overall survival, relapse-free survival and is frequently diagnosed at a higher stage (Guinney et al., 2015). This might explain the differential outcomes between our study and the previous colorectal cancer studies linking *CDR1-AS* to worse outcomes (Tang et al., 2017; Weng et al., 2017), if indeed subtype CMS4 is well-represented in those

cohorts, which would also support a link between *CDR1-AS* expression and stroma and, in colorectal cancer, poor outcome.

Another study on gastric cancer also found an association of increased *CDR1-AS* expression with poor survival (Pan et al., 2017). In hepatocellular carcinoma, the situation is less clear; while *CDR1-AS* was associated with several markers of poor prognosis, it was not significantly associated ($P > 0.05$) with recurrence (Xu et al., 2017).

Overall, there have been conflicting reports on *hsa-miR-7,* labelling it as an oncomir on the one hand and as a tumour suppressor microRNA on the other hand. Several cell line and animal studies support its tumour suppressor role with a specific focus on the target *EGFR* (Kefas et al., 2008; Wang et al., 2013a); associated proteins, such as *PAK1* (Puto et al., 2003; Reddy et al., 2008); as well as transcription factors, such as *YY1* (Zhang et al., 2013). By contrast, others have found oncogenic roles for *hsa-miR-7* by targeting a transcriptional repressor (Chou et al., 2010) or the tumour suppressor *KLF4* (Meza-Sosa et al., 2014). The associations with clinical outcomes in patients are conflicting for this microRNA as well. On the one hand, it has been found that lower *hsa-miR-7* expression is associated with poor survival in gastric cancer (n = 106) (Zhao et al., 2015) and with more metastases and a shorter OS in lung cancer patients (n = 108) (Cheng et al., 2017), indicating a tumour suppressor role of the microRNA. On the other hand, arguing for an oncogenic role, a study on colorectal cancer patients (n = 210) found an association of high *hsa-miR-7* levels with several clinical markers for poor prognosis, different relapse types and lower overall survival (Nagano et al., 2016). Furthermore, a study on prostate cancer patients (n = 45) also showed that increased expression of this microRNA is associated with a worse OS. In the latter study, *hsa-miR-7* expression was also linked to a shorter time span until the tumours became castration-resistant (Santos et al., 2014).

In the case of breast cancer, there has been one report associating lower *hsa-miR-7* expression levels with lymph node-positive breast cancer patients versus lymph node-negative patients (Kong et al., 2012). However, in this small study, neither the tumour ER status nor information on potential treatments before surgery or adjuvant treatment were reported (Kong et al., 2012). Those factors could introduce a bias, especially when the patient numbers are small, and this could explain the different results from our study.

In general, the contradictions in different studies regarding the role of *hsa-miR-7* might indicate that this microRNA is co-regulated with different causal factors making it a

biomarker of aggressive tumours in one cancer and a biomarker of less aggressive tumours in another. Further studies with large well-characterised patient cohorts as well as functional studies are needed to clarify the contradictions with regard to this microRNA and patient outcome.

Based on our findings in a 1st-line tamoxifen-treated patient cohort, we conclude that *hsa-miR-7* is a predictive biomarker for poor response and PFS. Our previous study showed that *hsa-miR-7* was a prognostic biomarker in a LNN breast cancer cohort (Foekens et al., 2008), and the present study confirmed this result in an, albeit not fully independent, extended LNN breast cancer cohort. Additionally, *hsa-miR-7* was found to be associated with short OS in the ER-positive subset in the prognostic cohort and PR-OS in the predictive cohort. By contrast, *CDR1-AS* did not show associations with any clinical outcomes in our breast cancer cohorts.

In conclusion, we showed that in addition to the known association between high *hsa-miR-7* expression and a worse course of disease (earlier development of metastases) in primary breast cancer, high *hsa-miR-7* levels are predictive of an adverse response to tamoxifen therapy and poor progression-free and post-relapse overall survival in patients with recurrent disease. These associations do not appear to be related to *CDR1-AS*. However, in light of recent reports describing the clinical utility of *CDR1-AS* expression in other cancer types (Pan et al., 2017; Tang et al., 2017; Weng et al., 2017; Xu et al., 2017), it will be interesting to study this gene further, including its stromal location, in large cohorts focused on those cancer types and to investigate its functional influence on cancer progression.

Materials and Methods

Patient samples

The protocol to study biological markers associated with disease outcome was approved in writing by the medical ethics committee of the Erasmus Medical Centre Rotterdam, The Netherlands (MEC 02.953) and was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in The Netherlands (https://www.federa.org/codes-conduct). The use of anonymous or coded left over material for scientific purposes is part of the standard treatment agreement with patients and therefore informed consent was not required according to Dutch law. The results of this study are reported based on the REMARK criteria for clinical reporting (McShane et al., 2005).

For our analyses we generated *CDR1-AS* and *hsa-miR-7* expression data using qPCR (see below) for two different clinical cohorts. These assays were performed blinded to the study endpoint.

In this retrospective study, tumours of female patients were included, who underwent surgery for invasive primary breast cancer between 1980 and 1995 in the Netherlands. A further selection criterion was no previously diagnosed cancers with the exception of basal cell carcinoma or stage Ia/Ib cervical cancer.

Within this study only data from sections of primary tumours with at least 30% invasive tumour cells, evaluated as described before (Sieuwerts et al., 2005), were included. The primary tumour specimens used in this study for RNA isolation consisted of fresh frozen tissue that had been stored in liquid nitrogen. At least 100 mg of tumour material per patient had to be available. The details of tissue processing have been described previously (Sieuwerts et al., 2005).

No stratification or matching was used for our cohorts. The tumour grade was assessed according to standard procedures at the time of inclusion.

For the classification of patients' RNA samples regarding ER expression, PR expression and HER2 amplification status, RT-qPCR was used with the following cut-offs: (1) ER (*ESR1* mRNA) 0.2, (2) PR (*PGR* mRNA) 0.1 and (3) HER2 (*ERBB2* mRNA) 18; the cut-offs were established as previously described (van Agthoven et al., 2009; Sieuwerts et al., 2007).

The first cohort (prognostic cohort) included primary tumours from LNN breast cancer patients who had not received any systemic (neo)adjuvant therapy. The different inclusion criteria discussed within this section led to a cohort consisting of 120 ER-negative and 225 ER-positive primary tumours. The median age of the patients in the study at time of surgery was 55 years (ranging from 27–88 years). The median follow-up time of living patients was 91 months (ranging from 8–319 months). MFS was defined as the time between surgery and the development of a distant metastasis. One-hundred-seventy (49%) patients developed a distant metastasis that counted as an event in the MFS analysis. Patients who died without evidence of disease were censored at the last follow-up in the MFS and OS analyses. In total, 178 patients were censored in the OS analysis and 175 were censored in the MFS analysis. Additional patient and tumour characteristics are presented in **Table 1**.

The second cohort (predictive cohort) consisted of 188 hormone-naïve breast cancer patients with ER-positive primary tumours who received tamoxifen as a 1st-line treatment for recurrent disease and fulfilled our inclusion criteria (discussed above). Thirty-five of the patients received adjuvant chemotherapy. The 188 patients [21 patients with local-regional relapse (LRR) and 167 patients with a distant metastasis] were treated with tamoxifen (40 mg daily) therapy. The median age of the patients at the start of tamoxifen therapy was 61 years (ranging from 29–90 years). We used the primary tumours of these patients for our analyses. The median time between primary surgery and the start of therapy was 27 months (ranging from 2-115 months). To evaluate PFS, the start of the 1st-line tamoxifen therapy was set as zero and the end point was at the time of progression or the last date of follow-up. The median follow-up time for living patients at the end of follow-up was 64 months (ranging from 8-272 months) after primary surgery and 10 months (ranging from 1-144 months) after the start of tamoxifen therapy. The disease-free interval was defined as the interval between the breast cancer diagnosis and the first recurrence of the disease. At the end of the follow-up period, 178 (95%) patients had developed tumour progression and 154 (82%) patients had died. In the PFS analysis, 10 patients were censored, while 34 patients were censored in the PR-OS analysis. Clinical response to tamoxifen therapy was defined by standard Unio Internationale Contra Cancrum criteria (Hayward et al., 1977) and described previously (Sieuwerts et al., 2005). Those 126 patients with an evident tumour reduction of 50% or more (partial and complete remission) or no change in tumour volume after more than 6 months were defined as responders to 1st-line tamoxifen therapy.

Non-responders included the 62 patients who showed tumour progression (progressive disease) or no change in tumour volume after ≤ 6 months. The median PFS times were as follows: (1) complete remission (n = 7), 16 months; (2) partial remission (n = 34), 15 months; (3) stable disease (n = 85), 14 months; (4) no change after ≤ 6 months (n = 9), 5 months; and (5) progressive disease (n = 53), 3 months. Additional patient and tumour characteristics for this cohort are shown in **Table 3**.

An overview of the specifics for the clinical cohorts and cell lines [see below] are depicted in a flow diagram in **Supplementary Fig. S1**.

Cell lines, immune cells and tumour pool used in the expression experiment

For our cell line experiments, we used total RNA isolated from breast cancer cell lines with different intrinsic subtype characteristics described previously (Hollestelle et al., 2010; Sieuwerts et al., 2009) (see **Supplementary Table S4** for cell lines, subtypes and

normalised *CDR1-AS* expression values) supplemented with total RNA isolated from the endothelial cell line EA.hy926 and a fibroblast strain (M92-19T) derived from a primary breast cancer tumour (Foekens et al., 2008). Furthermore we included MACS sorted immune cell subsets from PBMCs which were isolated from buffy coats by Ficoll gradient centrifugation from a healthy donor. All cultured cell lines were established to be of correct identity by performing STR profiling analyses using the PowerPlex® 16 system (Promega, Madison, WI, USA). The breast tumour pool used in this experiment consisted of cDNA from 100 primary breast tumours from lymph node-negative and -positive patients who did not receive neo-adjuvant treatment. Overall, the samples were *ESR1*-positive, *PGR*-positive, had no *ERBB2* amplification and had a high proliferative index [GGI; 1.1] (Toussaint et al., 2009).

MACS

Magnetic-activated cell sorting (MACS) was performed using the MACS technology with an anti-PE positive selection kit (Miltenyi Biotec, Leiden, the Netherlands) according to the manufacturer's protocol. Briefly, 5E06 PBMC in 100 µl MACS buffer were stained with 10µl PE-labelled primary antibody conjugate (CD4, CD8, CD56 or HLA-DR) and incubated for 10 minutes at 6-8°C, in the dark. Following washing with MACS buffer the cells were labelled with 20µl anti-PE magnetic beads for 15 minutes at 6-8°C in the dark. After additional washing, the cells were applied to MACS MS columns which were placed in a magnetic field. Finally, the positively selected cell fraction was eluted, cells were counted and the purity (percentage of PE labelled cells) was analysed by flow cytometry using a BD FACSCelesta (BD Biosciences, San Jose, CA, USA). Purities of the cell sorting per targeted immune cell population are given in **Supplementary Table S5**.

RNA isolation, cDNA synthesis and quantitative real time PCR (RT-qPCR)

Tissue processing, total RNA isolation and total RNA quality control checks have been described elsewhere (Sieuwerts et al., 2005, 2014). Briefly, total RNA was isolated with RNA Bee (Thermo Fisher Scientific, Waltham, MA, USA) from 30 µm tissue sections. MRNAs/circRNAs were reverse-transcribed with the RevertAid H Minus First Strand cDNA Synthesis Kit from Thermo Fisher Scientific, followed by an RNAse H step (Ambion, Thermo Fisher Scientific). QPCR reactions were performed in a 25 µL final volume using an Mx3000P™ Real-Time PCR System (Agilent Technologies, Amstelveen, the Netherlands). All qPCR assays were established to have an efficiency of at least 90% and did not generate a product within 35 cycles in the absence of reverse transcriptase. Three reference genes were selected based on the literature and tested for their signal intensity and stable expression level between both LNN and lymph-node positive (LNP) as well as ESR1-high and ESR1-low samples, using GeNorm (Vandesompele et

al., 2002) and NormFinder (Andersen et al., 2004) software packages available in GenEx 6.1.1.550 (MultiD Analyses AB, Gotenborg, Sweden). mRNA/circRNA levels were assessed relative to the average expression of *HMBS*, *HPRT1* and *TBP* using the delta Cq method (dCq = 2^(average Cq reference genes - Cq target gene)). *Hsa-miR-7* levels were measured and quantified relative to *hsa-miR-132* and *hsa-miR-374* as described previously (Foekens et al., 2008). For the cultured cell lines and clinical specimen we incorporated only samples in our study in which we could detect the reference gene *HMBS* within 25 qPCR cycles when using 10 ng of total RNA as input. This to ensure only RNA samples with sufficient quantity and of good quality are measured.

All of the primers and hydrolysis probe assays used for gene expression are shown in **Supplementary Table S6**.

Figure preparation and statistical analysis

Figures were generated using SPSS version 24 (IBM, Armonk, NY, USA), Excel (Microsoft Corporation, Redmond, WA, USA) and/or Inkscape (Free Software Foundation Inc., Boston, MA, USA). All statistical analyses were performed in SPSS (version 24). Cox univariate and/or multivariate regression analysis were used for the survival analyses (MFS, PFS, OS and PR-OS). MFS was defined as the time between the primary surgery and the first distant metastasis. PFS was defined as the time between the start of the first line therapy (tamoxifen) and the first sign of progression. OS was defined as the time between the primary surgery and death or the last follow up. PR-OS was defined as the time between relapse and death or the last follow-up. Mann-Whitney tests were used to analyse *CDR1-AS* and *hsa-miR-7* with clinical parameters (except for cases when more than 2 groups were analysed; then a Kruskal-Wallis test was performed); the clinical parameters were used as grouping variables. Logistic regression was used to assess whether treatment responses were associated with *CDR1-AS* and *hsa-miR-7* expression. The non-parametric Kaplan-Meier estimator was used to estimate and plot the survival functions, with the log-rank test to assess for differences. All statistical tests were two-sided and $P < 0.05$ was considered statistically significant.

Data availability

The full data sets used in this paper are available upon request.

Acknowledgements

We would like to thank Maxime P. Look for her advice on the statistical analyses, Marion Meijer van Gelder for her work with the clinical patient data and Corine Beaufort for technical assistance. This study was financially supported by the Daniel den Hoed

Foundation, Rotterdam, The Netherlands (K.U., J.A.F.). Additional funding came from Cancer Genomics Netherlands - a grant from the Netherlands organization of Scientific Research (NWO) (A.M.S., J.W.M.M.) and from the foundations Alpe d'HuZes and KWF (D.H.). Partial support also came from the ERC Advanced Grant #322737 (J.A.F.).

These foundations did not exert influence on this study.

Author contributions

K.U., A.M.S., J.W.M.M. and J.A.F. planned the study. A.M.S. and V.d.W. performed the laboratory work (RNA isolation and qPCR measurements). A.M.S. planned and performed the statistical analyses. K.U. and A.M.S. created the figures. M.S. contributed on the TIL and stromal signature analyses. D.H. performed the immune cell analyses. K.U. wrote the article, which was critically reviewed and approved by A.M.S., V.d.W., M.S., D.H., J.A.F. and J.W.M.M.

Additional information

Competing Interests: None of the authors has a competing interest.

Supplementary information accompanies this paper at:

https://doi.org/10.1038/s41598-018-27987-w

References

van Agthoven, T., Sieuwerts, A.M., Meijer-van Gelder, M.E., Look, M.P., Smid, M., Veldscholte, J., Sleijfer, S., Foekens, J.A., and Dorssers, L.C.J. (2009). Relevance of Breast Cancer Antiestrogen Resistance Genes in Human Breast Cancer Progression and Tamoxifen Resistance. J. Clin. Oncol. *27*, 542–549.

Allinen, M., Beroukhim, R., Cai, L., Brennan, C., Lahti-Domenici, J., Huang, H., Porter, D., Hu, M., Chin, L., Richardson, A., et al. (2004). Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell *6*, 17–32.

Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. *64*, 5245–5250.

Barrett, S.P., Parker, K.R., Horn, C., Mata, M., and Salzman, J. (2017). ciRS-7 exonic sequence is embedded in a long non-coding RNA locus. PLoS Genet. *13*, e1007114.

Chang, T.-C., Yu, D., Lee, Y.-S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., and Mendell, J.T. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. Nat. Genet. *40*, 43–50.

Cheng, M.-W., Shen, Z.-T., Hu, G.-Y., and Luo, L.-G. (2017). Prognostic Significance of microRNA-7 and its Roles in the Regulation of Cisplatin Resistance in Lung Adenocarcinoma. Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol. *42*, 660–672.

Chou, Y.-T., Lin, H.-H., Lien, Y.-C., Wang, Y.-H., Hong, C.-F., Kao, Y.-R., Lin, S.-C., Chang, Y.-C., Lin, S.-Y., Chen, S.-J., et al. (2010). EGFR promotes lung tumorigenesis by activating miR-7 through a Ras/ERK/Myc pathway that targets the Ets2 transcriptional repressor ERF. Cancer Res. *70*, 8822–8831.

Delás, M.J., and Hannon, G.J. (2017). lncRNAs in development and disease: from functions to mechanisms. Open Biol. *7*, 170121.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat. Rev. Genet. *12*, 861–874.

Foekens, J.A., Sieuwerts, A.M., Smid, M., Look, M.P., de Weerd, V., Boersma, A.W.M., Klijn, J.G.M., Wiemer, E.A.C., and Martens, J.W.M. (2008). Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. Proc. Natl. Acad. Sci. U. S. A. *105*, 13021–13026.

Gou, Q., Wu, K., Zhou, J.-K., Xie, Y., Liu, L., Peng, Y., Gou, Q., Wu, K., Zhou, J.-K., Xie, Y., et al. (2017). Profiling and bioinformatic analysis of circular RNA expression regulated by c-Myc. Oncotarget *5*, 71587–71596.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The Consensus Molecular Subtypes of Colorectal Cancer. Nat. Med. *21*, 1350–1356.

Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., et al. (2010). American Society of Clinical Oncology/ College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer (Unabridged Version). Arch. Pathol. Lab. Med. *134*, e48–e72.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. (2011). miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J. *30*, 4414–4422.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hayward, J.L., Carbone, P.P., Heuson, J.-C., Kumaoka, S., Segaloff, A., and Rubens, R.D. (1977). Assessment of response to therapy in advanced breast cancer: A project of the Programme on Clinical Oncology of the International Union against Cancer, Geneva, Switzerland. Eur. J. Cancer 1965 *13*, 89–94.

Hollestelle, A., Nagel, J.H.A., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J., et al. (2010). Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. Breast Cancer Res. Treat. *121*, 53–64.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141–157.

Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M., Lee, J., Fine, H., Chiocca, E.A., Lawler, S., et al. (2008). microRNA-7 Inhibits the Epidermal Growth Factor Receptor and the Akt Pathway and Is Down-regulated in Glioblastoma. Cancer Res. *68*, 3566–3572.

Kersting, C., Kuijper, A., Schmidt, H., Packeisen, J., Liedtke, C., Tidow, N., Gustmann, C., Hinrichs, B., Wülfing, P., Tio, J., et al. (2006). Amplifications of the epidermal growth factor receptor gene (*egfr*) are common in phyllodes tumors of the breast and are associated with tumor progression. Lab. Invest. *86*, 54–61.

Kong, X., Li, G., Yuan, Y., He, Y., Wu, X., Zhang, W., Wu, Z., Chen, T., Wu, W., Lobie, P.E., et al. (2012). MicroRNA-7 Inhibits Epithelial-to-Mesenchymal Transition and Metastasis of Breast Cancer Cells via Targeting FAK Expression. PLOS ONE *7*, e41523.

Kristensen, L.S., Okholm, T.L.H., Venø, M.T., and Kjems, J. (2018a). Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. RNA Biol. *15*, 280–291.

Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018b). Circular RNAs in cancer: opportunities and challenges in the field. Oncogene *37*, 555–565.

Li, X., and Carthew, R.W. (2005). A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. Cell *123*, 1267–1277.

Liao, J.-M., and Lu, H. (2011). Autoregulatory Suppression of c-Myc by miR-185-3p. J. Biol. Chem. *286*, 33901–33909.

Ma, L., Bajic, V.B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. RNA Biol. *10*, 924–933.

Ma, X.-J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D.C. (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. BCR *11*, R7.

Maass, P.G., Glažar, P., Memczak, S., Dittmar, G., Hollfinger, I., Schreyer, L., Sauer, A.V., Toka, O., Aiuti, A., Luft, F.C., et al. (2017). A map of human circular RNAs in clinically relevant tissues. J. Mol. Med. *95*, 1179–1189.

Massink, M.P.G., Kooi, I.E., Martens, J.W.M., Waisfisz, Q., and Meijers-Heijboer, H. (2015). Genomic profiling of CHEK2*1100delC-mutated breast carcinomas. BMC Cancer *15*, 877.

McShane, L.M., Altman, D.G., Sauerbrei, W., Taube, S.E., Gion, M., and Clark, G.M. (2005). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK). JNCI J. Natl. Cancer Inst. *97*, 1180–1184.

Meza-Sosa, K.F., Pérez-García, E.I., Camacho-Concha, N., López-Gutiérrez, O., Pedraza-Alva, G., and Pérez-Martínez, L. (2014). MiR-7 promotes epithelial cell transformation by targeting the tumor suppressor KLF4. PloS One *9*, e103987.

Möller, P., Mechtersheimer, G., Kaufmann, M., Moldenhauer, G., Momburg, F., Mattfeldt, T., and Otto, H.F. (1989). Expression of epidermal growth factor receptor in benign and malignant primary tumours of the breast. Virchows Arch. A *414*, 157–164.

Nagano, Y., Toiyama, Y., Okugawa, Y., Imaoka, H., Fujikawa, H., Yasuda, H., Yoshiyama, S., Hiro, J., Kobayashi, M., Ohi, M., et al. (2016). MicroRNA-7 Is Associated with Malignant Potential and Poor Prognosis in Human Colorectal Cancer. Anticancer Res. *36*, 6521–6526.

Nguyen, H.T.T., Dalmasso, G., Yan, Y., Laroui, H., Dahan, S., Mayer, L., Sitaraman, S.V., and Merlin, D. (2010). MicroRNA-7 modulates CD98 expression during intestinal epithelial cell differentiation. J. Biol. Chem. *285*, 1479–1489.

Pan, H., Li, T., Jiang, Y., Pan, C., Ding, Y., Huang, Z., Yu, H., and Kong, D. (2017). Overexpression of Circular RNA ciRS-7 Abrogates the Tumor Suppressive Effect of miR-7 on Gastric Cancer via PTEN/PI3K/AKT Signaling Pathway. J. Cell. Biochem. *119*, 440–446.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature *406*, 747–752.

Pilichowska, M., Kimura, N., Fujiwara, H., and Nagura, H. (1997). Immunohistochemical study of TGF-alpha, TGF-beta1, EGFR, and IGF-1 expression in human breast carcinoma. Mod. Pathol. *10*, 969–975.

Puto, L.A., Pestonjamasp, K., King, C.C., and Bokoch, G.M. (2003). p21-activated kinase 1 (PAK1) interacts with the Grb2 adapter protein to couple to growth factor signaling. J. Biol. Chem. *278*, 9388–9393.

Reddy, S.D.N., Ohshiro, K., Rayala, S.K., and Kumar, R. (2008). MicroRNA-7, a homeobox D10 target, inhibits p21-activated kinase 1 and regulates its functions. Cancer Res. *68*, 8195–8200.

Riaz, M., Sieuwerts, A.M., Look, M.P., Timmermans, M.A., Smid, M., Foekens, J.A., and Martens, J.W.M. (2012). High TWIST1 mRNA expression is associated with poor prognosis in lymph node-negative and estrogen receptor-positive human breast cancer and is co-expressed with stromal as well as ECM related genes. Breast Cancer Res. BCR *14*, R123.

van Roozendaal, K.E.P., Klijn, J.G.M., van Ooijen, B., Claassen, C., Eggermont, A.M.M., Henzen-Logmans, S.C., and Foekens, J.A. (1996). Differential regulation of breast tumor cell proliferation by stromal fibroblasts of various breast tissue sources. Int. J. Cancer *65*, 120–125.

Santos, J.I., Teixeira, A.L., Dias, F., Maurício, J., Lobo, F., Morais, A., and Medeiros, R. (2014). Influence of peripheral whole-blood microRNA-7 and microRNA-221 high expression levels on the acquisition of castration-resistant prostate cancer: evidences from in vitro and in vivo studies. Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med. *35*, 7105–7113.

Sieuwerts, A.M., Gelder, M.E.M., Timmermans, M., Trapman, A.M.A.C., Garcia, R.R., Arnold, M., Goedheer, A.J.W., Portengen, H., Klijn, J.G.M., and Foekens, J.A. (2005). How ADAM-9 and ADAM-11 Differentially From Estrogen Receptor Predict Response to Tamoxifen Treatment in Patients with Recurrent Breast Cancer: a Retrospective Study. Clin. Cancer Res. *11*, 7311–7321.

Sieuwerts, A.M., Usher, P.A., Gelder, M.E.M., Timmermans, M., Martens, J.W.M., Brünner, N., Klijn, J.G.M., Offenberg, H., and Foekens, J.A. (2007). Concentrations of TIMP1 mRNA Splice Variants and TIMP-1 Protein Are Differentially Associated with Prognosis in Primary Breast Cancer. Clin. Chem. *53*, 1280–1288.

Sieuwerts, A.M., Kraan, J., Bolt, J., Spoel, P. van der, Elstrodt, F., Schutte, M., Martens, J.W.M., Gratama, J.-W., Sleijfer, S., and Foekens, J.A. (2009). Anti-Epithelial Cell Adhesion Molecule Antibodies and the Detection of Circulating Normal-Like Breast Tumor Cells. J. Natl. Cancer Inst. *101*, 61–66.

Sieuwerts, A.M., Lyng, M.B., Meijer-van Gelder, M.E., de Weerd, V., Sweep, F.C.G.J., Foekens, J.A., Span, P.N., Martens, J.W.M., and Ditzel, H.J. (2014). Evaluation of the ability of adjuvant tamoxifen-benefit gene signatures to predict outcome of hormone-naive estrogen receptor-positive breast cancer patients treated with tamoxifen in the advanced setting. Mol. Oncol. *8*, 1679–1689.

Smid, M., Rodríguez-González, F.G., Sieuwerts, A.M., Salgado, R., Smissen, W.J.C.P.-V. der, Vlugt-Daane, M. van der, Galen, A. van, Nik-Zainal, S., Staaf, J., Brinkman, A.B., et al. (2016). Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nat. Commun. *7*, 12910.

Storz, G. (2002). An Expanding Universe of Noncoding RNAs. Science *296*, 1260–1263.

Suto, T., Yokobori, T., Yajima, R., Morita, H., Fujii, T., Yamaguchi, S., Altan, B., Tsutsumi, S., Asao, T., and Kuwano, H. (2015). MicroRNA-7 expression in colorectal cancer is associated with poor prognosis and regulates cetuximab sensitivity via EGFR regulation. Carcinogenesis *36*, 338–345.

Tang, W., Ji, M., He, G., Yang, L., Niu, Z., Jian, M., Wei, Y., Ren, L., and Xu, J. (2017). Silencing CDR1as inhibits colorectal cancer progression through regulating microRNA-7. OncoTargets Ther. *10*, 2045–2056.

The ENCODE Project Consortium (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature *489*, 57–74.

Toussaint, J., Sieuwerts, A.M., Haibe-Kains, B., Desmedt, C., Rouas, G., Harris, A.L., Larsimont, D., Piccart, M., Foekens, J.A., Durbecq, V., et al. (2009). Improvement of the clinical applicability of the Genomic Grade Index through a qRT-PCR test performed on frozen and formalin-fixed paraffin-embedded tissues. BMC Genomics *10*, 424.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*, RESEARCH0034.

Wang, W., Dai, L., Zhang, S., Yang, Y., Yan, N., Fan, P., Dai, L., Tian, H., Cheng, L., Zhang, X., et al. (2013a). Regulation of epidermal growth factor receptor signaling by plasmid-based microRNA-7 inhibits human malignant gliomas growth and metastasis in vivo. Neoplasma *60*, 274–283.

Wang, Y., Liu, J., Liu, C., Naji, A., and Stoffers, D.A. (2013b). MicroRNA-7 regulates the mTOR pathway and proliferation in adult pancreatic β-cells. Diabetes *62*, 887–895.

Webster, R.J., Giles, K.M., Price, K.J., Zhang, P.M., Mattick, J.S., and Leedman, P.J. (2009). Regulation of Epidermal Growth Factor Receptor Signaling in Human Cancer Cells by MicroRNA-7. J. Biol. Chem. *284*, 5731–5741.

Weng, W., Wei, Q., Toden, S., Yoshida, K., Nagasaka, T., Fujiwara, T., Cai, S., Qin, H., Ma, Y., and Goel, A. (2017). Circular RNA ciRS-7—A Promising Prognostic Biomarker and a Potential Therapeutic Target in Colorectal Cancer. Clin. Cancer Res. *23*, 3918–3928.

Xu, L., Wen, Z., Zhou, Y., Liu, Z., Li, Q., Fei, G., Luo, J., and Ren, T. (2013). MicroRNA-7–regulated TLR9 signaling–enhanced growth and metastatic potential of human lung cancer cells by altering the phosphoinositide-3-kinase, regulatory subunit 3/Akt pathway. Mol. Biol. Cell *24*, 42–55.

Xu, L., Zhang, M., Zheng, X., Yi, P., Lan, C., and Xu, M. (2017). The circular RNA ciRS-7 (Cdr1as) acts as a risk factor of hepatic microvascular invasion in hepatocellular carcinoma. J. Cancer Res. Clin. Oncol. *143*, 17–27.

Yu, C.-Y., Li, T.-C., Wu, Y.-Y., Yeh, C.-H., Chiang, W., Chuang, C.-Y., and Kuo, H.-C. (2017). The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. Nat. Commun. *8*, 1149.

Yu, Z., Ni, L., Chen, D., Zhang, Q., Su, Z., Wang, Y., Yu, W., Wu, X., Ye, J., Yang, S., et al. (2013). Identification of miR-7 as an oncogene in renal cell carcinoma. J. Mol. Histol. *44*, 669–677.

Zhang, N., Li, X., Wu, C.W., Dong, Y., Cai, M., Mok, M.T.S., Wang, H., Chen, J., Ng, S.S.M., Chen, M., et al. (2013). microRNA-7 is a novel inhibitor of YY1 contributing to colorectal tumorigenesis. Oncogene *32*, 5078–5088.

Zhao, X.-D., Lu, Y.-Y., Guo, H., Xie, H.-H., He, L.-J., Shen, G.-F., Zhou, J.-F., Li, T., Hu, S.-J., Zhou, L., et al. (2015). MicroRNA-7/NF-κB signaling regulatory feedback circuit regulates gastric carcinogenesis. J. Cell Biol. *210*, 613–627.

Chapter 7

The circular RNome of primary breast cancer

Marcel Smid, Saskia M. Wilting, Katharina Uhr, F. Germán Rodríguez-González, Vanja de Weerd, Wendy J.C. Prager-Van der Smissen, Michelle van der Vlugt-Daane, Anne van Galen, Serena Nik-Zainal, Adam Butler, Sancha Martin, Helen R. Davies, Johan Staaf, Marc J. van de Vijver, Andrea L. Richardson, Gaëten MacGrogan, Roberto Salgado, Gert G.G.M. van den Eynden, Colin A. Purdie, Alastair M. Thompson, Carlos Caldas, Paul N. Span, Fred C.G.J. Sweep, Peter T. Simpson, Sunil R. Lakhani, Steven Van Laere, Christine Desmedt, Angelo Paradiso, Jorunn Eyfjord, Annegien Broeks, Anne Vincent-Salomon, Andrew P. Futreal, Stian Knappskog, Tari King, Alain Viari, Anne-Lise Børresen-Dale, Hendrik G. Stunnenberg, Mike Stratton, John A. Foekens, Anieta M. Sieuwerts and John W.M. Martens.

Genome Research 2019;29(3):356-366

Abstract

Circular RNAs (circRNAs) are a class of RNAs that is under increasing scrutiny, although their functional roles are debated. We analyzed RNA-seq data of 348 primary breast cancers and developed a method to identify circRNAs that does not rely on unmapped reads or known splice-junctions. We identified 95,843 circRNAs, of which 20,441 were found recurrently. Of the circRNAs that match exon-boundaries of the same gene, 668 showed a poor or even negative ($R < 0.2$) correlation with the expression level of the linear gene. In silico analysis showed only a minority (8.5%) of circRNAs could be explained by known splicing events. Both these observations suggest that specific regulatory processes for circRNAs exist. We confirmed the presence of circRNAs of *CNOT2, CREBBP* and *RERE* in an independent pool of primary breast cancers. We identified circRNA profiles associated with subgroups of breast cancers and with biological and clinical features such as amount of tumor lymphocytic infiltrate and proliferation index. siRNA-mediated knockdown of *circCNOT2* was shown to significantly reduce viability of the breast cancer cell lines MCF-7 and BT-474, further underlining the biological relevance of circRNAs. Furthermore, we found that circular, and not linear, *CNOT2* levels are predictive for progression-free survival time to aromatase inhibitor (AI) therapy in advanced breast cancer patients, and found that *circCNOT2* is detectable in cell-free RNA from plasma. We showed that circRNAs are abundantly present, show characteristics of being specifically regulated, are associated with clinical and biological properties, and thus are relevant in breast cancer.

Keywords: circular RNA, RNA-sequencing, breast cancer, CNOT2, Aromatase Inhibitor

Introduction

It is a sign of the times that the ubiquitous use of massive parallel sequencing data has delivered a parade of new insights in the cancer field and has enriched our genomic vocabulary with events like chromothripsis, kataegis, and mutational and rearrangement signatures (Alexandrov et al., 2013; Maher and Wilson, 2012; Nik-Zainal et al., 2012, 2016; Stephens et al., 2011). Sequencing RNA has had less of an impact on this vocabulary, with many reports concerning traditional gene expression analysis. However, depending on the methodology of generating the sequencing library, RNA-seq has the potential to study the large variety of RNA species, including noncoding RNAs, fusion transcripts, known and novel isoforms, and, recently gaining attention, circular RNAs (circRNAs). This class of RNA was discovered many decades ago (Hsu and

Coca-Prados, 1979), and circRNAs were long considered idiosyncrasies of the splicing machinery processing precursor mRNA into mature mRNA. More recent studies showed an unanticipated abundance of circRNAs (Memczak et al., 2013; Salzman et al., 2012) in (normal and malignant) human cells and became particularly interesting for the cancer research field with the description (Hansen et al., 2013; Memczak et al., 2013) of a circRNA that functions as a highly potent miR-7 sponge. miR-7 has a well-described role in several malignancies, including breast cancer, and functions as a tumor suppressor in most cancers (for review, see (Zhao et al., 2015)) but has also been reported (Foekens et al., 2008) as a potential tumor promoter in breast cancer. Other circRNAs and additional regulatory transcriptional roles have subsequently been described in cancer (Guo et al., 2014; Kristensen et al., 2018; Li et al., 2015b; Salzman et al., 2013). Because circRNAs lack a free 5' or 3' end, such molecules escape exonucleic acid degrading enzymes, making them more stable (Memczak et al., 2013) than their linear counterparts. Therefore, circRNAs represent potentially useful biomarker candidates for diagnosis and therapy-monitoring; indeed, cell-free circRNAs are present in exosomes (Li et al., 2015a) and saliva (Bahn et al., 2015). In breast cancer, little has been described except for one study (Nair et al., 2016) using the The Cancer Genome Atlas (TCGA) data bank. However, this cohort has a huge limitation because the RNA-seq data were prepared using a poly(A) selection step, thereby omitting the majority of circRNAs (as these lack a poly(A) tail).

Here we describe the identification of an extensive catalog of circRNAs in a large cohort of 348 primary breast tumors, using RNA-seq data obtained via random-primed cDNA synthesis (Smid et al., 2016), likely preserving all the circRNAs. We developed a circRNA mapping algorithm that, in contrast to previous identification methods (Guo et al., 2014; Memczak et al., 2013; Nair et al., 2016; Salzman et al., 2012; Szabo and Salzman, 2016), does not rely on unmapped reads or on known splice junctions and that was applied directly on transcriptome sequence BAM files, thereby allowing the identification of circRNAs in a genome-wide and annotation-independent (Szabo and Salzman, 2016) fashion.

Results

Identification of a plethora of circRNAs in primary breast cancer

In total, 95,843 circRNAs were identified (**Fig. 1**), of which 27% (*n* = 25,783) had a start and end position exactly matching to an exon belonging to the same gene (**Fig. 2A**). The vast majority (79%) of all circRNAs were not recurrent (i.e., only found in one sample). The number of circRNAs per sample (**Fig. 2B**) ranged from 37 to 7105 (median, 966).

For recurrent circRNAs, found in at least two samples, the range per sample was 33 to 5269 (median, 834.5). **Figure 2B** also shows that the number of (recurrent) circRNAs is significantly higher in estrogen receptor (ER)-negative compared with ER-positive breast cancers (Mann-Whitney *U* test, $P < 1 \times 10^{-5}$ for both all and recurrent circRNAs). Because of the extraordinary abundance of candidate circRNAs, we focused on the – still sizeable – number of recurrently found circRNAs (n = 20,441) (total number in second bar of **Fig. 2A**). The most frequent recurring region in our cohort was the well-characterized (Hansen et al., 2013; Kristensen et al., 2018; Memczak et al., 2013) circRNA of *CDR1*, which was found in 339 out of 348 cases. Other previously reported and validated (Salzman et al., 2012) circRNAs such as *CAMSAP1, FBXW4, MAN1A2, RNF220, ZBTB44*, and *XIST* were also identified in our cohort. A full list of identified circRNAs is provided in **Supplemental Table S1**.

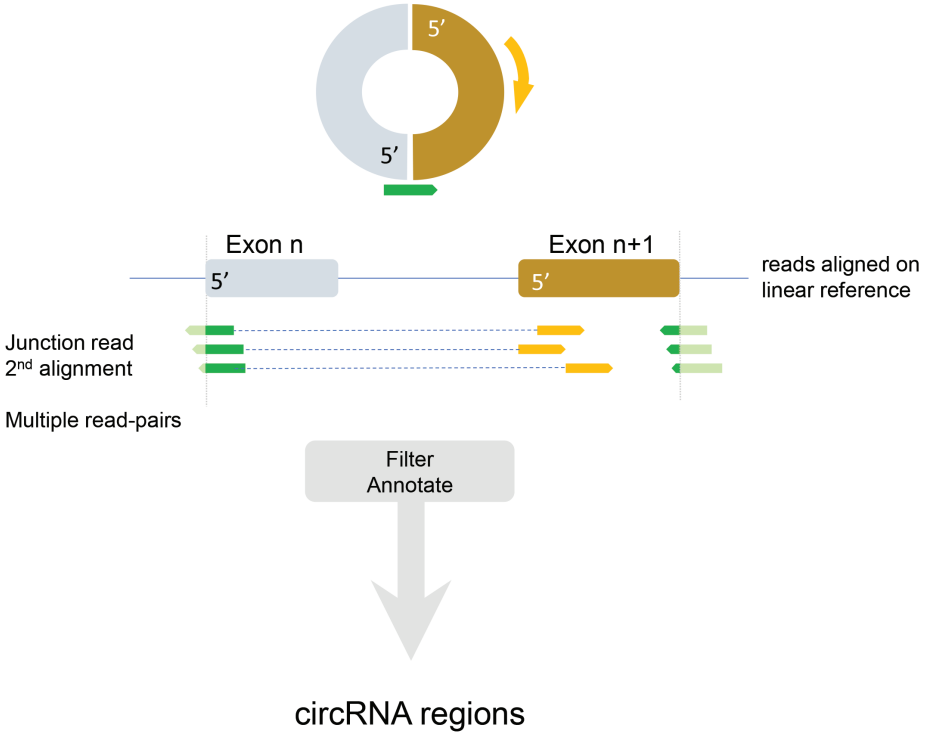


Figure 1. Schematic overview of identifying circular RNA (circRNA) regions. Assuming a circRNA molecule is present, a sequence read crossing the junction (green arrow) and its read-mate (gold arrow) would map to a linear reference in the manner depicted. The junction read would get multiple alignments, and the read-mate would be located in between the position of the junction read. Multiple read-pairs at the same junction strengthen the support for the circRNA. Subsequent additional filtering (details are in the Methods section) and annotation produced the list of circRNA regions.

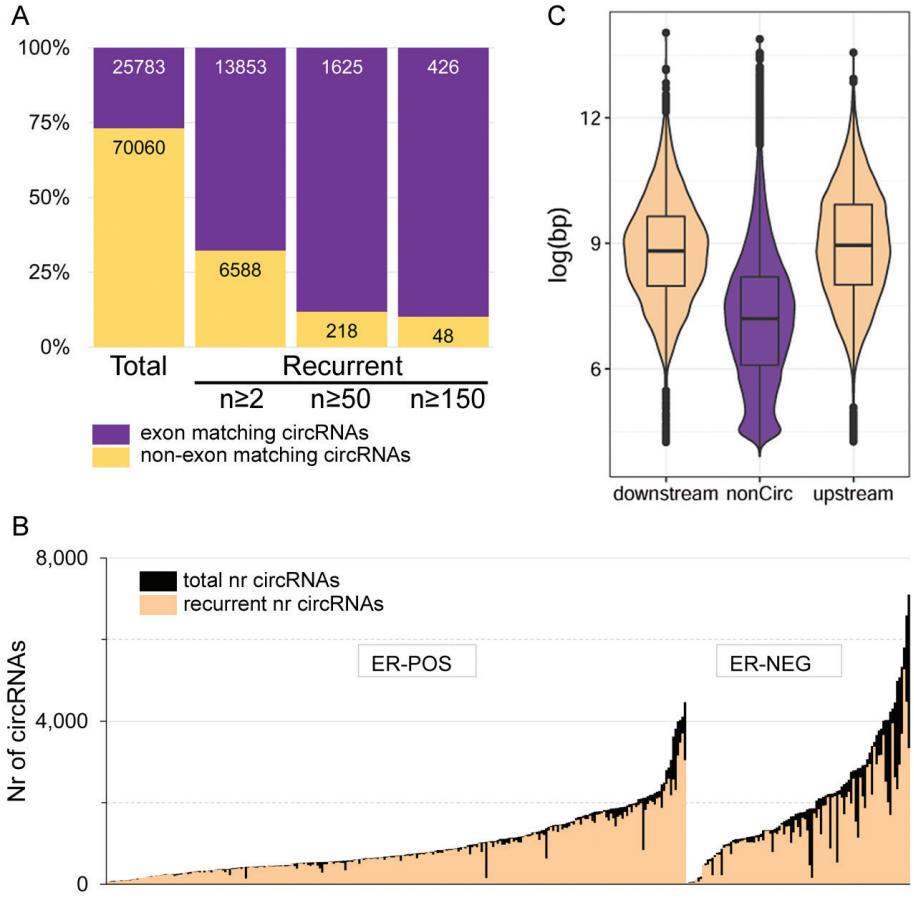


Figure 2. General characteristics of circRNAs in primary breast cancer. (*A*) Numbers of unique and recurrent circRNAs. Purple and gold indicate the number of circRNAs that, respectively, did or did not have a start and end position of a circRNA region exactly matching the start and end position of an exon of the same gene. (*B*) The number of circRNAs per sample, grouped by ER status. In black, the total number of circRNAs; in peach, the number of recurrent (identified in at least two samples) circRNAs. (*C*) Violin plots of the intron size (in log base pair) of noncircular regions and those located directly upstream of or downstream from a circRNA region.

General characteristics

Recurrent circRNAs were distributed across the genome, with one region on Chromosome 11 showing many closely spaced circRNAs (**Supplemental Fig. S1**). This region contains *MALAT1*, a highly abundant long noncoding RNA that is also frequently mutated in breast cancer (Nik-Zainal et al., 2016). Next, we evaluated the intron sizes upstream of and downstream from the circRNAs that match exon boundaries. Confirming previously reported results (Ivanov et al., 2015; Jeck et al., 2013; Zhang et al.,

2014), **Figure 2C** shows that introns next to circRNA regions are significantly larger than introns not adjacent to circRNAs: on average 2.33 and 2.27 times larger, respectively, for introns upstream of and downstream from the circRNA (Mann-Whitney *U* test, both $P < 1 \times 10^{-5}$).

Next, we correlated the number of circRNA reads per circRNA with the expression of the respective full, linear gene, because a previous report in a limited cell line panel reported no genome-wide correlation between the circular and linear counterpart (Salzman et al., 2013). To avoid spurious correlations, only those circRNAs found in at least 50 samples were considered ($n = 1625$), and data were first normalized using the trimmed mean of *M*-values (TMM) (Robinson and Oshlack, 2010). Correlations are listed in **Supplemental Table S1** and ranged between -0.34 and 0.97, with 210 circRNAs showing a negative correlation to the linear gene in which the circRNA is located. When considering the average and standard deviation of the distribution of all correlation coefficients, 30 circRNAs are at the low end of the distribution ($P < 0.05$), showing a correlation below R = -0.182 . Finally, by using GENCODE information, the position of annotated start codons was matched to the circRNA positions. circRNAs recurring in at least 50 samples showed an almost threefold higher than expected presence of a start codon compared with circRNAs that were not recurring (χ^2, $P < 0.0001$; 557 circRNAs; expected 207.3 circRNAs).

Are circRNAs distinct molecules, specifically regulated, or splicing residues?

Many of the circRNAs that are positively correlated with the expression level of the full-length linear transcript are thought to be a residue of splicing. **Figure 3A** shows an example; the circRNA of exons 3, 4 and 5 of the *WDR1* gene (Chr 4: 10,097,711-10,103,986, with a correlation coefficient of $R = 0.66$ to overall gene expression) exactly matches the difference between the known linear isoforms of this gene (Ensembl transcript ENST00000499869 includes exons 3, 4 and 5; ENST00000502702 lacks these exons). As current quantification methods do not take circRNAs into account, reads originating from the circular molecule are erroneously included in the read count of a linear isoform, resulting in an overestimation of the overall expression level. Another comprehensive example of a circRNA as splicing residue is shown in **Figure 3B** for *ESR1*. Full-length *ESR1* (ENST00000206249) has eight exons, whereas ENST00000406599 is a splice variant of *ESR1* that skips exons 2 to 5. circRNAs are found for several of these exons, indicating that they are likely splicing residues. We speculate that a single splice event from exon 1 to 6 generates an RNA molecule containing exons 2–5, from which a multitude of distinct circRNAs can be derived. In total, 23 patients show both a *circESR1* exon 2-3 and a *circESR1* of exon 4-5. If these circRNAs are derived from the same RNA

molecule, the linear transcript would be ENST00000406599. A sequential model, in which first exons 2 and 3 are spliced out would prohibit the formation of a circRNA molecule of exons 3 to 4. However, we observed *circESR1* exon 2-3 in 110 patients and *circESR1* exon 3-4 in 64 patients, with 29 patients showing both these circRNAs. These must be derived from separate RNA molecules.

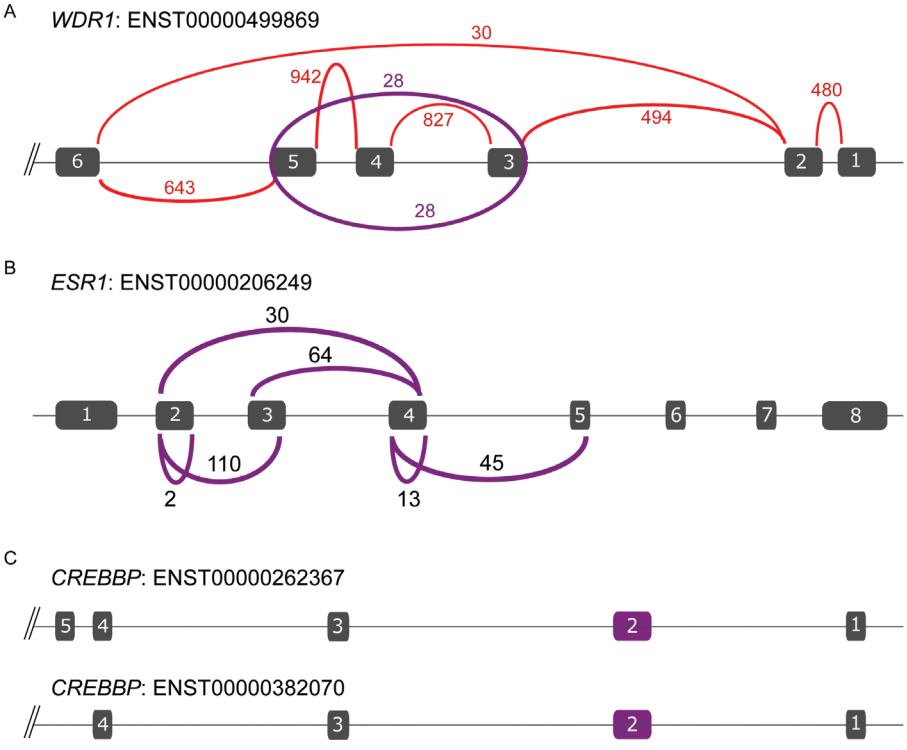


Figure 3. circRNAs are not just residues of splicing. (*A*) Sashimi plot of the number of reads that are aligned to *WDR1*, showing only the reads that span exons. In red are the normal exon-exon reads; in purple, the reads that span the circular junction. The line and boxes indicate the exons of the gene (the whole gene is not shown). (*B*) Isoform of *ESR1*. The arcs indicate the number of samples that have a particular circRNA. (*C*) Two isoforms of *CREBBP* that are known in the first five exons (other isoforms are described, but these start downstream from exon 5). Exon 2 (purple box) is an identified circRNA that is not a remainder of a splicing event.

To investigate whether or not in general circRNAs should be considered splicing residues, we systematically evaluated how many of the identified circRNAs exactly match those exons that make up the difference between known linear isoforms of a gene. By using the GENCODE annotation for each gene, every possible known combination

of spliced exons was matched to our circRNA catalog. Of the 25,783 circRNAs that matched to exons of the same gene, only 2193 (8.5%) exactly matched exons known to differ between described isoforms of a gene. This was 16.9% for the circRNAs with a correlation coefficient of $R > 0.5$ to linear gene expression. This suggests that the vast majority of circRNAs that matched to exons of the same gene are generated by yet-unknown splicing events of the gene.

Because the majority of circRNAs did not match to known spliced exons, we manually inspected several highly recurrent circRNAs in the UCSC Genome Browser. For example, two isoforms of *CREBBP* are described (ENST00000262367 and ENST00000382070) that differ in the presence of exon 5 (of note, there are 10 additional known transcripts, but all of these transcripts start downstream from exon 5). However, we observed exon 2 as circRNA (Chr 16: 3,850,297-3,851,009) that was present in 160 patients (**Fig. 3C**), indicating that this circRNA is either specifically generated or is a splicing residue of a yet-undescribed isoform of *CREBBP* that skips this exon. Visual inspection of 10 samples that had high levels of *circCREBBP* exon 2 (at least 30 circular junction reads) showed two samples that each had one read that crossed the junction from exon 1 to exon 3, whereas the other samples showed no evidence of an isoform that skipped exon 2. This favors the notion that the circRNA of exon 2 is not a byproduct of splicing at this location.

Finally, we matched publicly available circRNA lists to gather (indirect) evidence of functional roles for circRNAs. Rybak-Wolf and colleagues reported (Rybak-Wolf et al., 2015) 4522 circRNAs in the mammalian brain that were evolutionary conserved between human and mouse, which is considered an indication of function (Barbosa-Morais et al., 2012; Merkin et al., 2012). Of these, 2259 circRNAs (49.9%) were also present in our catalog. In addition, 3271 circRNAs were reported in an MCF-7 breast cancer cell line panel (Coscujuela Tarrero et al., 2018). In total, 922 circRNAs showed (increased) expression in estrogen-stimulated MCF-7 cells compared to cells cultured in hormone-deprived medium, of which 733 circRNAs (79.5%) were present in our list. A poor correlation ($R < 0.2$) with the linear transcript was observed in our data for 78 of these circRNAs, suggesting independent regulation from their linear gene instead of ER-induced overall higher expression of all transcripts from that gene.

Validation of circRNAs

Besides the fact that we detected several already published circRNAs, thereby in part validating our method, we performed RT-PCR on a previously established independent cDNA pool of 100 primary breast tumors to confirm expression of three circRNAs,

namely, *RERE* (circRNA Chr 1: 8,541,214-8,614,686), *CNOT2* (circRNA Chr 12: 70,278,132-70,311,017) and *CREBBP* (circRNA Chr 16: 3,850,297-3,851,009), all of which were poorly correlated with their linear counterpart. **Figure 4** shows the PCR fragment sizes; expected and observed sizes were 89 and 155 bp for the small and bigger *CNOT2* fragment, 100 bp for *RERE* and 91 bp for *CREBBP*. The primer pair for *CNOT2* was able to amplify the circRNA of exon 2 to exon 3 of *CNOT2* but also the circRNA of exons 2, 3, and 4 of this gene (a circRNA that was also identified in the RNA-seq cohort).

A different primer pair to PCR *circCNOT2* showed additional fragments in addition to the expected fragments of 140 and 206 bp (**Supplemental Fig. S2A**). Sanger sequence analysis confirmed the circular junction sequence from exon 3 to exon 2 of *CNOT2* (**Supplemental Fig. S2B**) and from exon 4 to exon 2 (**Supplemental Fig. S2C**). After using BLAST to identify the sequence, the largest excised PCR fragment was found to contain an additional exon of *CNOT2* (Chr 12: 70,294,237-70,294,293) located between exon 2 and exon 3 that is not present in most isoforms of *CNOT2* (**Supplemental Fig. S2D**). The sequence showed exon 3, across the circular junction to exon 2, but reading through the location where the reverse primer was located, continuing the whole of exon 2, the additional exon, and ending in exon 3 again (**Supplemental Fig. S2D**). A likely explanation for this observation is that during cDNA generation, the RT polymerase generated a linear cDNA molecule containing multiple copies of the circular transcript (**Supplemental Fig. S2E**). In summary, the investigated circRNAs were all confirmed to be truly present in primary breast cancer.

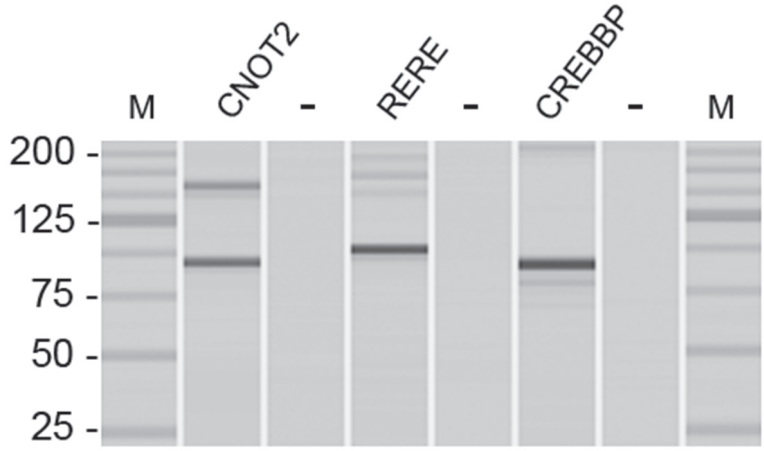


Figure 4. PCR products of circRNAs. PCR product sizes of circRNAs visualized using the MultiNA Microchip Electrophoresis System. (M) DNA size marker (25-bp fragment ladder); (–) the negative control (genomic DNA).

Functional relevance of circRNAs in breast cancer cells

The potential functional relevance of the validated *circCNOT2* and *circCREBBP* transcripts, which were both poorly correlated with their corresponding linear transcript, was evaluated in breast cancer cell lines. First, expression levels of *circCNOT2* and *circCREBBP* were established in a panel of 55 cell lines, showing variable levels (**Supplemental Fig. S3**). Next, an siRNA was designed to specifically target the circular junction of *circCNOT2* in both MCF-7 (moderate expression level) and BT-474 (high expression level). This siRNA reduced expression of *circCNOT2* by 76% in MCF-7 and 71% in BT-474 breast cancer cells, relative to cells transfected with a non-targeting control (NTC), which resulted in significantly reduced viability of both MCF-7 and BT-474 cells (Student's *t*-test $P < 1 \times 10^{-5}$ and $P = 4.94 \times 10^{-4}$, respectively) (**Fig. 5**).

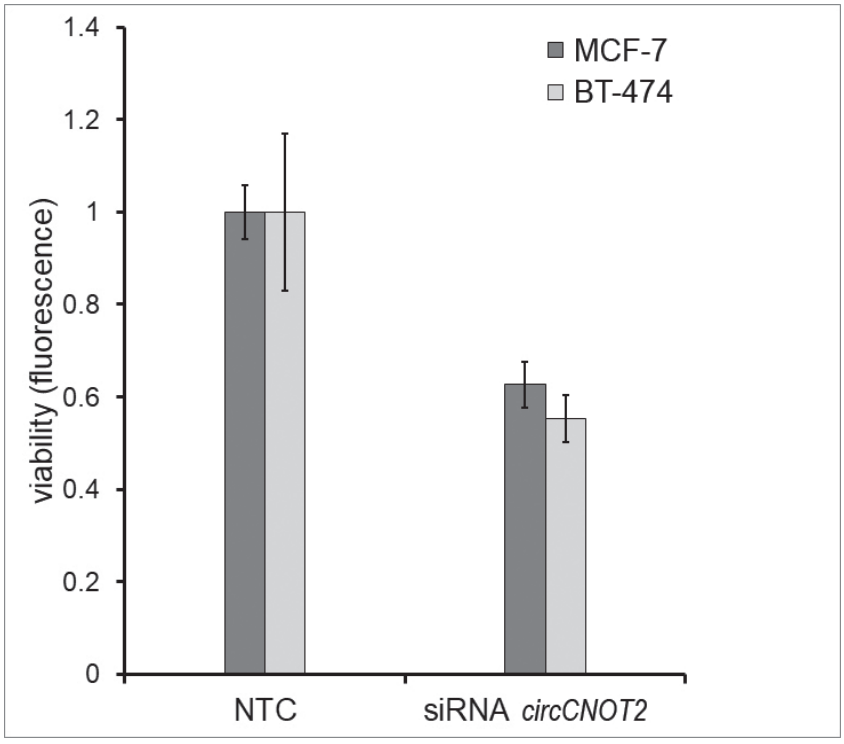


Figure 5. siRNA-mediated knockdown of *circCNOT2* affects viability in breast cancer cells. The effect of reduced *circCNOT2* expression on viability is shown in MCF-7 and BT-474 cells. Both cell lines show a significant decrease in viability ($P < 0.01$) following *circCNOT2* knockdown relative to cells transfected with nontargeting control (NTC). Error bars, SD of five wells.

circRNAs in driver genes

We matched our previously reported breast cancer driver gene list (Nik-Zainal et al., 2016) to our circRNA list. In total, 235 recurrent circRNAs were identified in 54 breast

cancer driver genes. To integrate the data and obtain sufficient observations for analysis, we selected samples for which we had both RNA and genomic DNA sequencing results available and selected the genes with somatically acquired genetic events (mutations, copy number variants and rearrangements) in at least 10 patients, yielding a list of 10 genes; *TP53, PIK3CA, PTEN, MAP3K1, CDH1, RB1, MAP2K4, ARID1B, ARID1A,* and *MLLT4*. For genes with multiple circRNAs, the circular region with the highest recurrence was chosen for analysis, with the exception of *TP53,* for which we only found two circRNAs in just one sample each (see **Table 1**). Only for *MAP2K4* was mutual exclusivity observed between the presence of a somatic mutation or a circRNA in this cohort, in which 20 samples had a somatic mutation, 77 samples had a circRNA (Chr 17: 12,054,889-12,113,360) and only two samples had both a mutation and a circRNA ($P = 0.025$, CoMEt exact test (Leiserson et al., 2015)). For *PIK3CA*, 25 patients showed a DNA event and a circRNA, three patients with a copy number aberration (amplification), and 22 patients with a base substitution in *PIK3CA*. These substitutions were located in four hotspots, p.H1047 (13 cases), p.E545 ($n = 4$), p.E542 ($n = 3$) and p.E726 ($n = 2$). None of these hotspots was located in the circRNA region that was found in these samples.

Table 1: Number of samples with a DNA event and/or circRNA

circRNA region	Gene Symbol	No. of samples		
		DNA events	circRNA	Both
Chr 17: 7,673,535-7,674,290	*TP53*	105	1	0
Chr 17: 7,674,859-7,676,622	*TP53*	105	1	1
Chr 3: 179,203,544-179,204,588	*PIK3CA*	84	77	25
Chr 10: 87,925,513-87,952,259	*PTEN*	47	13	4
Chr 5: 56,864,734-56,865,977	*MAP3K1*	30	210	22
Chr 16: 68,801,670-68,815,759	*CDH1*	22	7	0
Chr 13: 48,342,599-48,349,023	*RB1*	22	4	0
Chr 17: 12,054,889-12,113,360	*MAP2K4*	20	77	2
Chr 6: 156,829,227-156,935,576	*ARID1B*	14	28	1
Chr 1: 26,729,651-26,732,792	*ARID1A*	12	213	11
Chr 6: 167,870,383-167,889,326	*MLLT4*	10	28	0

Breast cancer relevance

To investigate common biology in the samples, we used multiple correspondence analysis (MCA) to find naturally occurring subgroups. MCA is a generalized principle component analysis, suitable for categorical data. We used recurrent circRNAs (at least 50 cases) with the junction annotated to exons of the same gene ($n = 1625$) and labeled these per sample as circular or not circular, based on the presence or absence of junction

reads in a sample. The main patient groups were, not unexpectedly, divided by ER status (**Fig. 6A**, left), whereas within circRNAs, the main division was whether or not the gene had a circRNA (**Fig. 6A**, right). Additional variation within the circRNAs was explained by the level of recurrence of the circRNAs (see **Supplemental Fig. S4**). Next, the presence/absence of circRNAs in a sample was used to cluster all samples into groups with distinct circRNA profiles. We used the gap statistic (Tibshirani et al., 2001) to determine the optimal number of sample groups, yielding six clusters (**Fig. 6B**; **Supplemental Fig. S5**). We evaluated these six sample groups on ER and tumor infiltrating lymphocyte (TIL) status (**Fig. 6C**, **D**), number of circRNAs (**Fig. 6E**), and outcome for the patients in the clusters (**Fig. 6F**). Samples in cluster 1 and 3 were predominantly ER-negative, whereas ER positivity was predominantly present in groups 2, 4, 5, and 6. TIL status was established using a previously reported gene expression signature (Massink et al., 2015; Smid et al., 2016), labeling samples as high-TIL if the average expression of the TIL signature genes fell into the top quartile (n = 87, 45 ER-negative and 42 ER-positive) (**Fig. 6D**, respectively, labeled as red and orange). High-TIL cases were significantly (χ^2 $P < 1 \times 10^{-5}$) more often present in clusters 1 (71% of cases) and 3 (45% of cases). Furthermore, the number of circRNAs per sample clearly distinguished the six clusters (**Fig. 6E**), showing a decreasing number of circRNAs from clusters 1 to 6. Finally, for a subset of 186 patients, relapse-free survival data were available; a survival plot for the six clusters (**Fig. 6F**) showed that the major difference was between clusters 1 and 3, that are both predominantly ER-negative. Although the number of events was low, direct comparison of cluster 1 with cluster 3 showed a significant difference in survival curves (log rank P = 0.04).

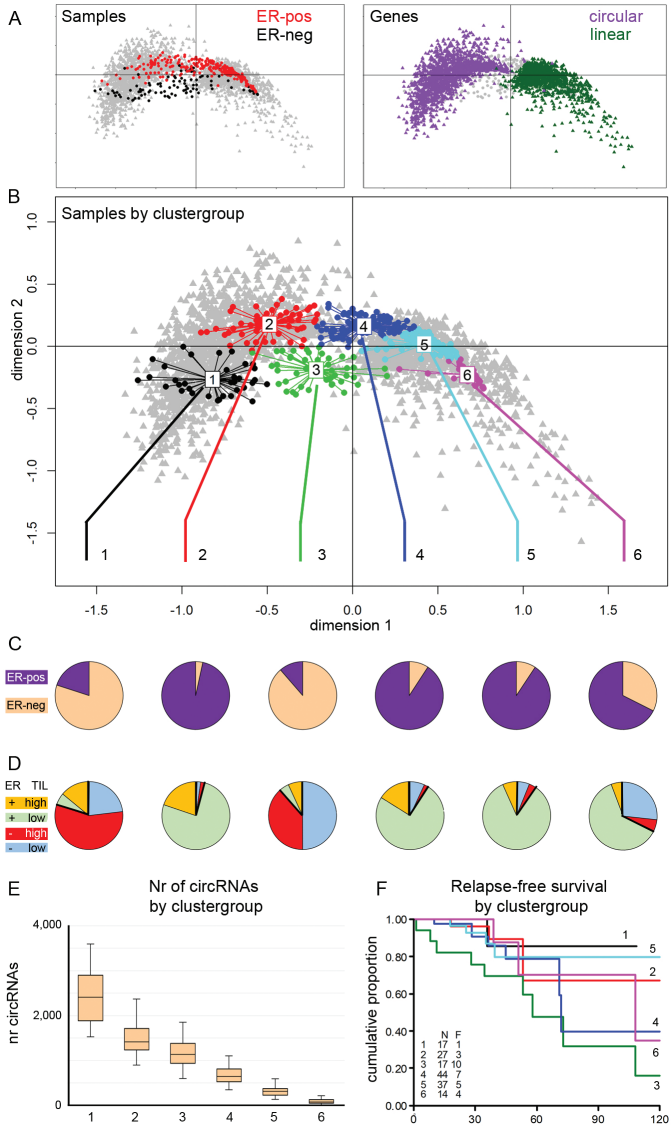


Figure 6. Analysis of sample groups according to circRNA presence. Multiple correspondence analysis (MCA) was used to find naturally occurring groups in the circRNA data. In an MCA plot, samples and circRNAs are projected onto the same plane, in which the relative distance to either the samples or the circRNAs is meaningful. The 0,0 point corresponds to a sample or circRNA with an average profile. (*A, left*) Samples are colored according to ER status: red, ER-positive; black, ER-negative. (*Right*) Purple and green indicate genes with or without circRNA expression, respectively. (*B*) Clustering identified samples with similar circRNA profiles; samples in the MCA plot are colored according to the cluster to which the sample belongs. (*C*) ER status (purple, ER-positive; peach, ER-negative) and (*D*) TIL status of the six sample groups: Red and orange are high-TIL cases and blue and green are low-TIL cases for ER-negative and ER-positive, respectively. (*E*) Number of circRNAs per sample group. (*F*) Relapse-free survival plot by sample group. (N) number of patients; (F) number of patients who relapse; (*x*-axis) months; (*y*-axis) the cumulative probability of relapse-free survival.

Differentially expressed circRNAs

We investigated if circRNA expression levels were associated with clinically relevant features of primary breast cancer, such as presence of TILs, the tumor's stromal content, proliferation, and hypoxia status. These features were inferred from generated (stroma; see Methods) or reported (Massink et al., 2015; Smid et al., 2016; Winter et al., 2007) gene expression signatures. By using these, we grouped our samples in a similar manner as explained earlier for the TIL status (**Fig. 6D**), labeling samples as high if the average expression of the signature genes fell into the top quartile. To identify significantly differentially expressed circRNAs, we compared the top quartile of samples to the remaining samples separately for the ER-positive and ER-negative cases. circRNAs with FDR-corrected *P*-values < 0.05 and a fold-change greater than two were selected. Of these, the circRNAs that had a negative correlation with the linear gene expression were considered of particular interest and are listed in **Table 2**. Several of these circRNAs may thus potentially play a role in, or are at least connected to, the tumors that show hypoxic characteristics (e.g., *circKMT2C*) or accumulate in highly proliferative cells (e.g., *circRERE*, *circATXN2*), whereas, for example, *circASH1L* and *circPCH3* may be generated by surrounding stromal cells or infiltrating cells.

Table 2: Fold-change of circRNAs in top quartile of samples versus remaining samples in ER-subgroups of tumors

Regions	Gene Symbol	Hypoxia		Proliferation		Stroma		TIL
		ER-negative	ER-positive	ER-negative	ER-positive	ER-negative	ER-positive	ER-negative
Chr 1: 155,438,327-155,459,898	*ASH1L*		2.0			2.1		2.0
Chr 1: 8,541,214-8,557,523	*RERE*	2.4		2.1				
Chr 1: 8,655,973-8,656,441	*RERE*				2.0			
Chr 2: 112,399,632-112,400,194	*RGPD8*	2.5	2.3		2.6			
Chr 3: 170,136,419-170,149,244	*PHC3*	3.4	2.1	2.6		2.8		2.5
Chr 5: 140,440,119-140,449,305	*ANKHD1*	2.2		2.2				
Chr 7: 152,309,966-152,315,338	*KMT2C*	2.1						
Chr 7: 155,672,867-155,680,908	*RBM33*	2.4	2.4	2.6	2.5		2.3	
Chr 7: 17,868,407-17,875,790	*SNX13*	2.4		2.3				
Chr 10: 32,543,300-32,584,304	*CCDC7*		2.2		2.4			
Chr 12: 111,554,158-111,555,919	*ATXN2*				2			
Chr 13: 75,560,753-75,569,507	*UCHL3*							2.0
Chr 15: 25,405,461-25,411,971	*UBE3A*	2.7	2.2	2.4	2.1			
Chr 15: 92,996,957-92,998,621	*CHD2*		2.4		2.3			
Chr 18: 76,849,526-76,851,939	*ZNF236*		2.1		2.0			
Chr 20: 35,716,740-35,725,155	*RBM39*	2.5		2.2				

No significant circRNAs were identified in the TIL ER-positive group.

Clinical relevance

One of the reasons *CNOT2* was selected for validation was because of the poor correlation with expression of the linear gene (*R* = -0.09, *P* = 0.34). This was more prominent in ER-positive (*R* = -0.14) compared with ER-negative cases (*R* = 0.097). We validated this finding by making use of in-house array data (Smid et al., 2008) for linear *CNOT2* expression and a quantitative RT-PCR assay to measure *circCNOT2* (exon 2-3 Chr 12: 70,278,132-70,311,017). The Spearman's correlation in ER-positive cases of *circCNOT2* with linear *CNOT2* was 0.079 (*n* = 187, *P* = 0.28) and in ER-negative cases Rs = 0.42 (*n* = 111, $P = 2.9 \times 10^{-6}$), again showing absence of correlation in ER-positive cases. Thus, the role of *circCNOT2* apparently differs between ER-positive and ER-negative cases, and in the ER-negative cases, the significance of *circCNOT2* cannot be easily segregated from the linear counterpart. Therefore, we evaluated two different ER-positive breast cancer cohorts (**Supplemental Table S3** shows clinical characteristics) for the potential clinical value of *circCNOT2*. We studied progression-free survival of aromatase inhibitor (AI) therapy in a multicenter cohort of 84 ER-positive patients who received this treatment for advanced disease. Quantitative reverse transcriptase-PCR (RT-qPCR) levels of *circCNOT2* showed a significant hazard ratio (HR) of 1.75 (95% confidence interval 1.32-2.33, $P = 1.06 \times 10^{-4}$), whereas RT-qPCR levels of linear *CNOT2* were not significant: HR 1.28, *P* = 0.187. **Figure 7A** shows a survival curve after grouping patients' *circCNOT2*-levels into three equally sized groups. A similar analysis in another cohort (Sieuwerts et al., 2005), which included patients receiving first-line tamoxifen treatment (*n* = 295 patients), did not show a significant HR (0.97, *P* = 0.57) for *circCNOT2* or for linear *CNOT2* (HR 1.16, *P* = 0.21).

As circular molecules are expected to be more stable than their linear counterparts, we explored whether *circCNOT2* is a potential candidate as minimally invasive biomarker. To this end, we used cell-free RNA (cfRNA) from plasma samples of four breast cancer patients and amplified *circCNOT2* by RT-qPCR. All samples showed detectable and variable levels of *circCNOT2* (**Fig. 7B**), indicating that detection of circRNAs in plasma seems attainable in this exploratory setting.

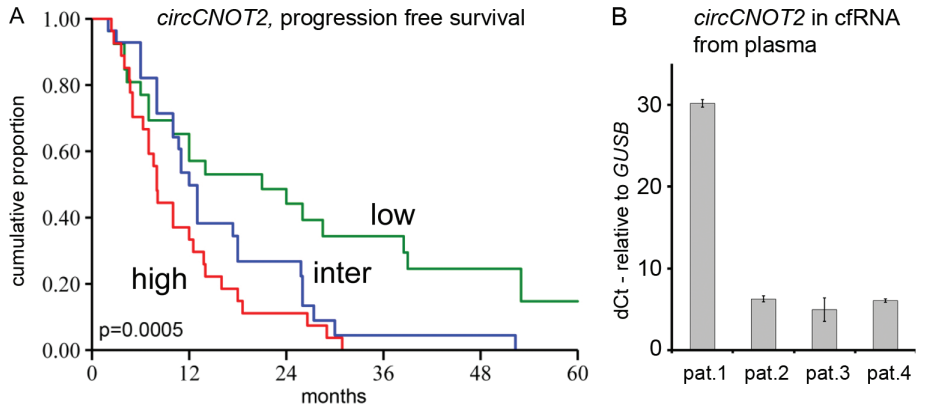


Figure 7. Clinical evaluation and presence in plasma samples. (*A*) Kaplan-Meier survival curve of progression-free survival for AI therapy in which patients were grouped in three equally sized groups based on their *circCNOT2* expression: red, blue and green indicate the samples with high, intermediate, and low expression. The *x*-axis is in months; *y*-axis depicts the cumulative probability of progression-free survival on AI therapy. The *P*-value is the logrank test for trend. (*B*) Expression levels of *circCNOT2* in plasma samples. Four metastatic breast cancer patients were evaluated. The *y*-axis depicts delta-C_t values of *circCNOT2* relative to *GUSB*. Error bars, SD of two measurements.

Discussion

To our knowledge, we are the first to analyze RNA sequencing data using random-primed cDNA libraries from a large primary breast cancer cohort for the presence of circRNAs, using a method that does not rely on unmapped reads. Previously, Nair and colleagues (Nair et al., 2016) analyzed TCGA RNA-seq data that were obtained using a poly(A)-based method. Although we identified 25,783 circRNAs that matched with an exon boundary of the same gene, Nair *et al.* (Nair et al., 2016) reported 2146 circRNAs when we applied the same selection criteria as for our data set, and after transferring the hg37 coordinates (Nair et al., 2016) to hg38 (our dataset), only 45 circRNAs were found that had the exact same start and end coordinates in both datasets. Thus, a random-primed method identifies many more circRNAs than when using poly(A)-selected material. On the other hand, there seem to be many uniquely identified circRNAs in these datasets. This could stem from differences in the methodology to detect or report the circRNAs but also could reflect the fact that many circRNAs are nonrecurrent.

We showed that circRNAs are found throughout the genome and have significantly larger-sized introns located directly adjacent to the region on the genome that borders the circRNA, as also reported previously (Ivanov et al., 2015; Jeck et al., 2013; Zhang et al., 2014). Based on the presence/absence of circRNAs, six groups of samples

were observed that differed in their ER status, TIL content, number of circRNAs, and prognosis. This indicates that there appears to be a functional biology associated with the biogenesis of circRNA molecules or at least a biology that cancerous cells can use to their advantage. Whether or not the circRNAs themselves serve that function or whether the process that generates differences in circRNA levels is the cause for the results presented here remains unknown at this time. The fact that several circRNAs were found differentially expressed in breast cancer subgroups, although these circRNAs are negatively correlated with the expression of the linear gene from which the circRNA is derived, corroborates the notion of functional circRNAs. Further experimentation is required to investigate the functional relationship of the differentially expressed circRNAs in the hypoxia, proliferation, stroma, and TIL phenotypes.

Although a synthetic circRNA construct including an internal ribosome entry site (IRES) can be translated (Wang and Wang, 2015), current literature (Guo et al., 2014; Jeck et al., 2013; Liang et al., 2017; Liu et al., 2018; Szabo and Salzman, 2016; Wang et al., 2018; Zeng et al., 2018) describes mostly noncoding functions for circRNAs, for example, as miRNA sponge *circCDR1* (Hansen et al., 2013; Kristensen et al., 2018; Memczak et al., 2013; Zhao et al., 2015) and *circZNF91* (Guo et al., 2014), whereas associative evidence was reported (Rybak-Wolf et al., 2015) of evolutionary conserved circRNAs between human and mouse. Further support that circRNAs may be specifically generated and regulated was derived from a study using (estrogen-stimulated) MCF-7 cells (Coscujuela Tarrero et al., 2018), showing higher levels of H3K36me3 (posttranscriptional histone modification) and a higher number of Ago binding sites in circularizing exons.

Here, we contribute to the search for relevant circRNAs in three ways; first, expression levels of circRNAs that are not, or even negatively correlated with the linear transcript of the gene may point to an intentional process. Although differences in degradation rates between the circular and linear isoforms may influence the correlation, we did not find systemic evidence for this (**Supplemental Fig. S6**). Furthermore, for genes that generate several distinct circRNAs, correlations can vary, indicating that degradation of the linear transcript cannot be the only explanation for the observed correlations. Second, circRNAs that include the start codon could potentially influence the expression of the linear gene, because the linear transcript from which the circRNA was spliced is now forced to use another start codon for its translation; if none is available, the transcript may be degraded. For example, *circCNOT2* (exon 2-3 Chr 12: 70,278,132-70,311,017) contains the start codon of the consensus (Pruitt et al., 2009) transcript (CCDS31857.1). Annotation shows that exon 1 is untranslated and both exon 2 and exon 4 start with the methionine codon (ATG). Thus, the linear transcript wherein exon 1

is ligated to exon 4 lacks the start codon from exon 2 and may use the ATG in exon 4 for its translation. Third, we observed that circRNAs matching exon boundaries of the same gene rarely (8.5%) overlap with known spliced exons. It could be that our analysis overlooked possible splice variants from the GENCODE annotation (both HAVANA and Ensembl exon annotations were included), but if the concordance is indeed this low, two scenarios may be applicable: Either circRNAs are still mostly a remnant of splicing, implying that many more transcript isoforms of genes exist, or otherwise circRNAs are specifically generated, implying that they do have a biological role. The observations of variable expression levels of *circCNOT2* and *circCREBBP* in cell lines and especially the effect of *circCNOT2* knockdown on cell viability corroborate a biological role for circRNAs. Future studies are needed to systematically evaluate if the correlation between a circRNA and its linear transcript, the presence of a start codon and/or known splice junctions are reliable criteria to prioritize circRNAs of interest.

Regardless, we were able to show clinical potential for *circCNOT2* by showing its association with the response to AI therapy. Knowing that circular molecules are not targeted by exonucleases, these molecules may be suitable candidates to be detected in cell-free environments (Li et al., 2015a), and in a pilot experiment, we showed that *circCNOT2* can indeed be detected in cfRNA from plasma samples of breast cancer patients. As such, *circCNOT2* could prove to be a useful biomarker to choose the right type of therapy or to monitor disease in a minimally invasive manner. Furthermore, we observed that very likely because of the strand displacement activity of the reverse transcriptase during cDNA generation, multiple concatemeric copies of a single circular molecule are made, contributing to the sensitive detection of circRNAs. In conclusion, we have demonstrated the abundance and potential roles of circRNAs in primary breast cancer. The methodology and selection criteria we used may help in making more sense of the seeming chaos and disorder existing in the flow from DNA to RNA to protein. circRNAs show the potential to function as relevant actors in the transcriptional regulation of RNA in addition to their promise as stable biomarkers that can be used for disease progression.

Methods

Sequencing

Internal review boards of each participating institution approved collection and use of samples of all patients in this study. RNA-seq data were generated by our consortium (Nik-Zainal et al., 2016; Smid et al., 2016) for 348 primary breast cancer tumors that are

available through the European Genome-phenome Archive under accession number EGAS00001001178. Sequence protocols of the samples were previously described in detail (Nik-Zainal et al., 2016); in short, total RNA after gDNA removal, clean-up, and depletion of ribosomal RNA using duplex-specific nuclease (DSN) treatment, was used as input for random-primed cDNA synthesis. Paired-end (75 bases) sequencing was performed on an Illumina HiSeq 2000. The resulting FASTQ files were mapped to GRCh38 using STAR (version 2.4.2a) (Dobin et al., 2013), and the resulting BAM files were sorted and indexed using Sambamba (version 0.6.6; https://github.com/lomereiter/sambamba/) (Tarasov et al., 2015). Gene annotation was derived from GENCODE Release 23 (https://www.gencodegenes.org/).

Identification of circular RNAs

A detailed explanation of the methodology to identify circRNA reads, including the Perl script, is stated in the **Supplemental Methods**. The script is also available at https://bitbucket.org/snippets/MSmid/Le949d/identify-circularrna-reads. In short, the method developed here uses sequence reads that have a "secondary alignment" (SA) tag. When using paired-end sequence data and assuming a circular RNA molecule is present (**Fig. 1**, top), the sequence read that aligns over the crossing of the junction (green arrows) would "point toward" its read-mate (orange arrow) somewhere in the circle. Aligning these reads to the linear reference (**Fig. 1**, middle), the junction read will get an SA tag and will be assigned to two locations if and only if this is the one and unique alignment configuration the STAR software can find. The read-mate aligns somewhere in between these two locations. Finding additional read-pairs showing this configuration, with a breakpoint at the exact same location, strengthens the evidence for circular transcripts. Only regions with at least five reads crossing the circular junction were included. After filtering (for details, see **Supplemental Methods**), GENCODE annotation was used to obtain the exon locations of genes that exactly matched to the circular region. For each sample, STAR also gives the raw read counts for all genes. These were normalized (TMM implemented in edgeR) (Robinson and Oshlack, 2010) and used to correlate with the number of junction reads of the circular transcripts.

Multiple correspondence analysis

Because many genes only show a linear transcript in many samples, standard cluster analysis to identify groups of samples with similar circRNA-related biology is problematic because of the many missing values. Thus, the circRNA data were considered categorical using "circular" or "not circular" if a circRNA was present or absent in a sample. These categorical data are suitable for an MCA, a generalization of a principle component analysis. The MCA generates a combined plot that shows both patients and circRNAs

such that patients/circRNAs that have similar patterns are closer together. R-packages "ade4", "canceracm", and "cluster" were used to perform the MCA and determine the optimum number of clusters. The latter was determined using the clusGap option (*k*-means to partition the samples) in the cluster package. R version 3.4.1 was used (R Core Team, 2017).

Reverse transcription, PCR, and Sanger sequencing

Candidate circRNAs were selected and primers were designed such that a PCR would only yield a product when the RNA was circular, whereas in the linear situation, the primers would be divergent. Primer sequences are listed in **Supplemental Table S2**.

First, total RNA, isolated with RNA-Bee according to the manufacturer's instructions (CS105B, TEL TEST) was reverse transcribed into cDNA with the RevertAid H Minus first-strand cDNA synthesis kit (K1632; Thermo Fisher Scientific), followed by a RNase H step (AM2293; Ambion). Next, circRNAs were real-time PCR amplified at 10 ng input in a final volume of 25 µL using 40 PCR cycles and an annealing temperature of 67°C with 330 nM of each primer and SensiFAST SYBR Lo-ROX mastermix (BIO-94050, Bioline), followed by a final 5-min extension at 72°C, in a MX3000P (Agilent Technologies). PCR products were visualized using a MultiNA microchip electrophoresis system (Shimadzu).

For sequencing, PCR fragments were separated on a standard agarose gel and were excised from gel using the QIAquick gel extraction kit from Qiagen according to manufacturer's protocol. The sequencing reaction contained 2 µL of gel-extracted PCR product, 1 µL BigDye terminator v3.1 reaction mix (Thermo Fisher Scientific), 1x BigDye terminator sequencing buffer (Thermo Fisher Scientific), and 0.16 µM of sequencing primer in a final volume of 10 µL and was performed using an ABI2720 thermal cycler according to the following protocol: one step for 2 min at 96 °C and 25 cycles of 30 sec at 96 °C , 30 sec at 58 °C, and 2 min at 72 °C. Subsequently, the sequencing product was precipitated with absolute ethanol and 3 M of NaAc, resuspended in 20 µL of Hi-Di formamide (Thermo Fisher Scientific), and ran on an ABI3130XL genetic analyzer (Thermo Fisher Scientific).

RT-qPCR

After RNA isolation and cDNA synthesis performed as described above, *circCNOT2* (Chr 12: 70,278,132-70,311,018) and *circCREBBP* (Chr 16: 3,850,297-3,851,009) transcripts were real-time PCR amplified at 10 ng input in a final volume of 25 µL in 40 PCR cycles and an annealing temperature of 60°C with 200 nM of each primer and 100 nM FAM-labeled TaqMan MGB probe that covers the circular junction (Thermo Fisher Scientific)

(**Supplemental Table S2**) in SensiFAST Probe Lo-ROX mastermix (BIO-84020, Bioline) using a MX3000P (Agilent Technologies). Levels were quantified relative to the average expression of three reference genes (*HPRT1*, *HMBS*, and *TBP*) (**Supplemental Table S2**) using the delta Cq method (dCq = $2^{(\text{average Cq reference genes - Cq target gene})}$) (Schmittgen and Livak, 2008). A serially diluted cDNA pool (Sieuwerts et al., 2014) of 100 independent breast tumor samples (containing both ER-positive/-negative and *ERBB2*-positive cases) was included in each experiment to evaluate the linear amplification and efficiencies for all genes and absence of amplification in the absence of reverse transcriptase. Samples in the cDNA pool were independent from the cases that were used for the RNAseq cohort.

Detection of circRNAs in plasma

cfRNA was isolated with the Maxwell RSC miRNA tissue kit (Promega) adapted for plasma according to the manufacturer's instructions. One milliliter of EDTA plasma of different metastatic breast cancer patients was used. These patients provided written informed consent. Six microliters of the resulting 50 µL cfRNA (3.8-7 ng RNA/µL) was used to generate 20 µL cDNA with the SuperScript IV VILO cDNA synthesis kit (Thermo Fisher Scientific). Next, 2 µL of the cDNA was preamplified in the presence of 0.50 nM of the reverse primers of the hydrolysis probe assays for *circCNOT2* and *GUSB* as a reference marker during 15 cycles with TaqMan preamp mastermix (Thermo Fisher Scientific). Finally, 0.5 µL of the preamplified product was measured real-time with the hydrolysis probe assays (200 nM forward primer, 200 nM reverse primer, and 100 nM FAM-labeled hydrolysis MGB probe) during 40 cycles with SensiFAST Probe Lo-ROX mastermix (BioLine) in a final qPCR volume of 25 µL in a MX3000P qPCR machine (Agilent Technologies).

siRNA-mediated knockdown of circRNAs and cell viability assay

All cell lines in this study were established to be genetically unique and monoclonal and of correct identity by performing STR profiling using the PowerPlex 16 system (Promega). MCF-7 and BT-474 were plated at 60-70% confluency in six-well plates and transfected with 50 nM ON-TARGETplus siRNA targeting *circCNOT2* (Horizon Discovery) using 4 µl (MCF-7) or 8 µl (BT-474) DharmaFECT 1 (Horizon Discovery) following the manufacturer's instructions. Used sequences (5'-3') were as follows: sense, AAAGAUAGGGAGACGUGGUUU; antisense, 5'-PACCACGUCUCCCUAUCUUUUU. The ON-TARGETplus nontargeting pool and On-TARGETplus human UBB smart pool were included in each experiment as negative and positive controls, respectively (Horizon Discovery). After 24 h of transfection, cells were trypsinized, counted, and seeded in quintuplicate at 20,000 cells per well in 96-well plates. Cell viability was determined using the CellTiter-Blue cell viability assay (Promega) at day 0 and day 3. Viability

measurements at day 3 were corrected for baseline viability values by subtracting the average measurement of day 0.

Gene expression signatures

We used several signatures: a TIL and proliferation signature (Smid et al., 2016), a hypoxia signature (Winter et al., 2007), and a stroma-specific signature using public data GSE5847 (Gene Expression Omnibus) (Boersma et al., 2008). We performed a paired *t*-test to obtain genes significantly higher expressed in microdissected stroma (FDR < 0.05 and fold-change > 1.7). For all signatures, genes that were up-regulated in the category of interest were matched to our data set, and the average expression of the signature genes was calculated per sample. Samples were labeled as high-TIL (or stroma, proliferation, hypoxia) if the average expression of the signature genes fell into the top quartile. To identify significantly differentially expressed circRNAs, we compared the top quartile of samples versus the rest, per ER-group. circRNAs were only included when detected in >50% of the samples and matched known exon locations of the same gene. Analyses were performed using BRB-ArrayTools developed by Dr. Richard Simon and the BRB-ArrayTools development team. circRNAs were considered significant when the FDR corrected *P*-value was below 0.05 and the fold-change greater than 2.

Breast cancer cohort treated with endocrine therapy

RT-qPCR was performed on a linear and circular isoform of *CNOT2* (Chr 12: 70,278,132-70,311,017) in a first-line TAM (Sieuwerts et al., 2005) and a first-line AI cohort to study the predictive value of *circCNOT2* on therapy response. The AI cohort was a multicenter cohort consisting of 30 patients from Erasmus MC, Rotterdam; 35 patients from The Netherlands Cancer Institute, Amsterdam; and 19 patients from the Translational Cancer Research Unit, Antwerp (Belgium). All 295 patients in the TAM cohort are patients of the Erasmus MC, Rotterdam. Patient characteristics are listed in **Supplemental Table S3**.

Statistical analyses

STATA version 14 was used to perform the statistical tests that are indicated in the text. *P*-values are two-sided, corrected for multiple testing when necessary, and considered significant below 0.05.

Supplemental files

Supplemental files can be accessed via: https://doi.org/10.1101/gr.238121.118

Acknowledgements

We thank the Erasmus MC Cancer Computational Biology Center for giving access to their IT infrastructure and the software that was used for the computations and data analysis in this study. We thank Sandra Albassam for her help with the first versions of the script to identify circular regions. We thank Maurice P.H.M. Jansen, Jean C. Helmijr, Inge de Kruijff, and Manouk K. Bos for their help in evaluating plasma samples that were gathered in the EU-FP7 CareMore (nr 601760) project. We thank for technical support Miriam Ragle Aure and Anita Langerød of the Oslo University Hospital, Norway; Ewan Birney of the European Bioinformatics Institute, UK; and Stefania Tommasi of the IRCCS Istituto Tumori "Giovanni Paolo II", Bari, Italy. We thank the Oslo Breast Cancer Research Consortium (OSBREAC), Norway (https://www.ous-research.no/home/kgjebsen/home/14105) for contributing patient samples and Sabine Linn and Marleen Kok of The Netherlands Cancer Institute for contributing samples for the AI cohort. Finally, we thank all members of the ICGC Breast Cancer Working Group. This work has been funded through the ICGC Breast Cancer Working group by the Breast Cancer Somatic Genetics Study (a European research project funded by the European Community's Seventh Framework Programme (FP7/2010-2014) under the grant agreement number 242006) and the Triple Negative project funded by the Wellcome Trust (grant reference 077012/Z/05/Z). F.G.R.-G. and S.M. were funded by BASIS. J.A.F. was funded through an ERC Advanced Grant (ERC-2012-AdG-322737) and ERC Proof-of-Concept Grant (ERC-2017-PoC-767854). K.U. was funded by the Daniel den Hoed Foundation. S.N.-Z. is a Wellcome Beit Fellow and personally funded by a Wellcome Trust Intermediate Fellowship (WT100183MA). A.L.R. is partially supported by the Dana-Farber/Harvard Cancer Center SPORE in Breast Cancer (NIH/NCI 5 P50 CA168504-02). A.M.S. was supported by Cancer Genomics Netherlands (CGC.nl) through a grant from the Netherlands Organization of Scientific research (NWO). M. Smid was supported by the EU-FP7-DDR response project. C.D. was supported by a grant from the Breast Cancer Research Foundation. J.E. was funded by The Icelandic Centre for Research (RANNIS).

Author contributions:

M. Smid, S.M.W., and J.W.M.M. wrote the main paper. M. Stratton, S.M., S.N.-Z., H.G.S., J.A.F., and J.W.M.M. were involved in the strategy and supervision of the project. Experiments were performed by S.M.W., F.G.R.-G., V.d.W., A.M.S., W.J.C.P.-v.d.S., M.v.d.V.-D., A.v.G., and J.S. M. Smid, A.M.S., K.U., S.M.W., S.N.-Z., J.S., M.J.v.d.V., A.L.R., A.B., H.R.D., F.C.G.J.S., A.V., A.B.-D., and J.W.M.M. analyzed data. Samples and/or clinical data were contributed by J.A.F., J.W.M.M., A.L.R., C.A.P., A.M.T., C.C., P.N.S., F.C.G.J.S., P.T.S., S.R.L., S.v.L., C.D., A.P., J.E., A.B., A.V.-S., A.P.F., S.K., T.K., A.V., A.B.-D., G.M., R.S., and G.G.G.M.v.d.E.

References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Bahn, J.H., Zhang, Q., Li, F., Chan, T.-M., Lin, X., Kim, Y., Wong, D.T.W., and Xiao, X. (2015). The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. Clin. Chem. *61*, 221–230.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. Science *338*, 1587–1593.

Boersma, B.J., Reimers, M., Yi, M., Ludwig, J.A., Luke, B.T., Stephens, R.M., Yfantis, H.G., Lee, D.H., Weinstein, J.N., and Ambs, S. (2008). A stromal gene signature associated with inflammatory breast cancer. Int. J. Cancer *122*, 1324–1332.

Coscujuela Tarrero, L., Ferrero, G., Miano, V., De Intinis, C., Ricci, L., Arigoni, M., Riccardo, F., Annaratone, L., Castellano, I., Calogero, R.A., et al. (2018). Luminal breast cancer-specific circular RNAs uncovered by a novel tool for data analysis. Oncotarget *9*, 14580–14596.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Foekens, J.A., Sieuwerts, A.M., Smid, M., Look, M.P., de Weerd, V., Boersma, A.W.M., Klijn, J.G.M., Wiemer, E.A.C., and Martens, J.W.M. (2008). Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. Proc. Natl. Acad. Sci. U. S. A. *105*, 13021–13026.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. Genome Biol. *15*, 409.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hsu, M.-T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. Nature *280*, 339–340.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., Dieterich, C., et al. (2015). Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. Cell Rep. *10*, 170–177.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141–157.

Kristensen, L.S., Hansen, T.B., Venø, M.T., and Kjems, J. (2018). Circular RNAs in cancer: opportunities and challenges in the field. Oncogene *37*, 555–565.

Leiserson, M.D.M., Wu, H.-T., Vandin, F., and Raphael, B.J. (2015). CoMEt: a statistical approach to identify combinations of mutually exclusive alterations in cancer. Genome Biol. *16*, 160.

Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., Chen, D., Gu, J., He, X., and Huang, S. (2015a). Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. Cell Res. *25*, 981–984.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., Dai, L., et al. (2015b). Exon-intron circular RNAs regulate transcription in the nucleus. Nat. Struct. Mol. Biol. *22*, 256–264.

Liang, H.-F., Zhang, X.-Z., Liu, B.-G., Jia, G.-T., and Li, W.-L. (2017). Circular RNA circ-ABCB10 promotes breast cancer proliferation and progression through sponging miR-1271. Am. J. Cancer Res. *7*, 1566–1576.

Liu, Y., Lu, C., Zhou, Y., Zhang, Z., and Sun, L. (2018). Circular RNA hsa_circ_0008039 promotes breast cancer cell proliferation and migration by regulating miR-432-5p/E2F3 axis. Biochem. Biophys. Res. Commun. *502*, 358–363.

Maher, C.A., and Wilson, R.K. (2012). Chromothripsis and Human Disease: Piecing Together the Shattering Process. Cell *148*, 29–32.

Massink, M.P.G., Kooi, I.E., Martens, J.W.M., Waisfisz, Q., and Meijers-Heijboer, H. (2015). Genomic profiling of CHEK2*1100delC-mutated breast carcinomas. BMC Cancer *15*, 877.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.

Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. Science *338*, 1593–1599.

Nair, A.A., Niu, N., Tang, X., Thompson, K.J., Wang, L., Kocher, J.-P., Subramanian, S., and Kalari, K.R. (2016). Circular RNAs and their associations with breast cancer subtypes. Oncotarget *7*, 80967–80979.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. Cell *149*, 979–993.

Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature *534*, 47–54.

Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J., et al. (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome Res. *19*, 1316–1323.

R Core Team (2017). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., et al. (2015). Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. Mol. Cell *58*, 870–885.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs Are the Predominant Transcript Isoform from Hundreds of Human Genes in Diverse Cell Types. Plos One *7*, e30733.

Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-Type Specific Features of Circular RNA Expression. Plos Genet. *9*, e1003777.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C-T method. Nat. Protoc. *3*, 1101–1108.

Sieuwerts, A.M., Meijer-van Gelder, M.E., Timmermans, M., Trapman, A.M.A.C., Garcia, R.R., Arnold, M., Goedheer, A.J.W., Portengen, H., Klijn, J.G.M., and Foekens, J.A. (2005). How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *11*, 7311–7321.

Sieuwerts, A.M., Lyng, M.B., Meijer-van Gelder, M.E., de Weerd, V., Sweep, F.C.G.J., Foekens, J.A., Span, P.N., Martens, J.W.M., and Ditzel, H.J. (2014). Evaluation of the ability of adjuvant tamoxifen-benefit gene signatures to predict outcome of hormone-naive estrogen receptor-positive breast cancer patients treated with tamoxifen in the advanced setting. Mol. Oncol. *8*, 1679–1689.

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A.M., Yu, J., Klijn, J.G.M., Foekens, J.A., and Martens, J.W.M. (2008). Subtypes of breast cancer show preferential site of relapse. Cancer Res. *68*, 3108–3114.

Smid, M., Rodríguez-González, F.G., Sieuwerts, A.M., Salgado, R., Smissen, W.J.C.P.-V. der, Vlugt-Daane, M. van der, Galen, A. van, Nik-Zainal, S., Staaf, J., Brinkman, A.B., et al. (2016). Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nat. Commun. *7*, 12910.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. Cell *144*, 27–40.

Szabo, L., and Salzman, J. (2016). Detecting circular RNAs: bioinformatic and experimental challenges. Nat. Rev. Genet. *17*, 679–692.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B-Stat. Methodol. *63*, 411–423.

Wang, Y., and Wang, Z. (2015). Efficient backsplicing produces translatable circular mRNAs. RNA-Publ. RNA Soc. *21*, 172–179.

Wang, H., Xiao, Y., Wu, L., and Ma, D. (2018). Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-000911/miR-449a pathway in breast carcinogenesis. Int. J. Oncol. *52*, 743–754.

Winter, S.C., Buffa, F.M., Silva, P., Miller, C., Valentine, H.R., Turley, H., Shah, K.A., Cox, G.J., Corbridge, R.J., Homer, J.J., et al. (2007). Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. Cancer Res. *67*, 3441–3449.

Zeng, K., He, B., Yang, B.B., Xu, T., Chen, X., Xu, M., Liu, X., Sun, H., Pan, Y., and Wang, S. (2018). The pro-metastasis effect of circANKS1B in breast cancer. Mol. Cancer *17*, 160.

Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary Sequence-Mediated Exon Circularization. Cell *159*, 134–147.

Zhao, J., Tao, Y., Zhou, Y., Qin, N., Chen, C., Tian, D., and Xu, L. (2015). MicroRNA-7: a promising new target in cancer therapy. Cancer Cell Int. *15*, 103.

Chapter 8

Discussion

The aim of this thesis was to investigate ncRNAs in breast cancer, focusing on their relationship with drug response and clinical outcome. With regard to the former, a large drug screening was performed in cell lines (**chapter 3**), which was subsequently used to study drug sensitivity in relation to miRNA expression (**chapter 4**), as well as genomic characteristics (**chapter 5**). With regard to the latter, the landscape of circRNAs, a relatively novel ncRNA species, is described in breast tumors (**chapter 7**) as well as the association of a subset with clinical parameters including response to treatment (**chapter 6, 7**) and survival (**chapter 6**).

Overview on the outcome of the drug association studies

An important part of this thesis involved a drug screening on breast cancer cell lines and the association of different genomic features with the respective drugs to identify markers for predicting drug response. While not all of the drugs studied in this thesis are currently used in clinical practice to treat breast cancer, those drugs might become of interest in the future for the treatment of breast cancer or other cancer types. By associating miRNA expression, mutational status and copy number aberration with drug response we provided an extensive characterization of drug sensitivity for 34 drugs with multiple genomic features. Furthermore, we investigated the relation with breast cancer subtype as well as undertook analyses to reveal the biology associated behind our observations. This rich data source aids in elucidating resistance mechanisms for these 34 drugs as well as confirms previously identified resistance mechanisms. For the drugs studied in this thesis we have summarized our findings per drug in **Table 1**. Twenty-eight of the drugs studied were correlated positively or negatively with another drug suggesting shared or opposing biology behind the observed drug sensitivity. Subtype, however, played a minor role in drug response since only one drug exhibited subtype-dependent sensitivity. Looking further at **Table 1**, several drugs were not only associated with miRNAs but also with mutations and CNAs. This raises the question whether the respective genomic features might be associated with each other. While an earlier study investigated whether mutations associated with miRNA expression (Riaz et al. 2013), not all mutations studied within this thesis were covered. Furthermore, Riaz et al. (Riaz et al. 2013) restricted their analysis to mutations in certain breast cancer subtypes. Therefore, it is useful to look within the restricted cell line dataset used for our miRNA study (Uhr et al. 2019) whether associations between miRNAs and the mutations studied in **chapter 5**, are present. Another point to add, is, that while miRNAs associating with a breast cancer subtype were excluded from the analyses in **chapter 4**, an association with subtype was not widely tested for mutations, and affected mutations were not excluded from the analyses in **chapter 5**.

Table 1. Genomic features found associated with drug response within this thesis

Drug	Correlated drugs	Genes overlapping with other drugs	Genomic features associated with sensitivity	Genomic features associated with resistance
5-Fluorouracil	Decitabine (R = 0.58)	13 genes (Mitoxantrone, Veliparib) 30 genes (Mitoxantrone)	2 gains and 2 losses	1 gain
Azacitidine	Doxorubicin (R = 0.7), Cisplatin (R = - 0.55)		2 gains	
Belinostat	Vorinostat (R = 0.92), Panobinostat (R = 0.91), Quisinostat (R = 0.85)	6 genes (Vorinostat) 6 genes (Panobinostat, Quisinostat, Sirolimus, Vorinostat) 6 genes (Panobinostat, Vorinostat) 21 genes (Panobinostat)	2 losses and 1 gain, *PIK3CA, CDH1*	
Bortezomib	Vandetanib (R = 0.47), JNJ-208 (R = -0.66)	3 genes (Vandetanib)	2 gains, *hsa-let-7a*	1 gain
Brivanib	JNJ-707 (R = 0.46)	65 genes (Paclitaxel)	4 gains	*MAP3K1*
Cisplatin	Doxorubicin (R = -0.73), Azacitidine (R = -0.55), Sunitinib (R = 0.47)			1 gain
Dasatinib				2 gains
Decitabine	5-Fluorouracil (R = 0.58), Serdemetan (R = 0.56), Lapatinib (R = 0.52), Veliparib (R = 0.48)	11 genes (Doxorubicin)	1 gain	
Docetaxel	Paclitaxel (R = 0.73), Tivantinib (R = 0.47)	1 gene (Methotrexate, Paclitaxel) 1 gene (Paclitaxel)	1 gain	1 gain, *hsa-miR-106a-3p, hsa-miR-187-5p*
Doxorubicin	Cisplatin (R = -0.73), Azacitidine (R = 0.7)	11 genes (Decitabine)	2 gains and 1 loss	*MLL3*

Table 1. Continued

Drug	Correlated drugs	Genes overlapping with other drugs	Genomic features associated with sensitivity	Genomic features associated with resistance
Erlotinib	Gefitinib (R = 0.88), Vandetanib (R = 0.47)	17 genes (Gefitinib)	2 gains, *ERBB2* amplification	*PTEN*
Gefitinib	Erlotinib (R = 0.88), Vandetanib (R = 0.56)	17 genes (Erlotinib)	2 gains	
JNJ-208	Bortezomib (R = -0.66)		2 gains and 1 loss, *MDM2* amplification	
JNJ-493 (Erdafitinib)	JNJ-707 (R = 0.62), Sunitinib (R = 0.48)		2 losses and 1 gain	1 gain
JNJ-707	Erdafitinib (R = 0.62), Brivanib (R = 0.46), Mitoxantrone (R = 0.45), Nutlin-3 (R = -0.45)		*hsa-mir-135a-3p*	
Lapatinib	Decitabine (R = 0.52)			
Methotrexate		1 gene (Docetaxel, Paclitaxel)	*CDH1*	2 gains, *PIK3CA*, *ATM*, *CSMD3, NAV3*
MI-219	Nutlin-3 (R = 0.98)	23 genes (Nutlin-3 and, in the opposing direction, with Sunitinib) 103 genes (Nutlin-3)	3 losses	1 gain, *TP53*
Mitoxantrone	JNJ-707 (R = 0.45)	13 genes (Veliparib, 5-Fluorouracil), 30 genes (5-Fluorouracil) 1 gene (Serdemetan)	1 gain, *CDH1*, *CDH1* methylation	1 gain
Nutlin-3	MI-219 (R = 0.98), JNJ-707 (R = -0.45)	23 genes (MI-219 and, albeit in the opposing direction, with Sunitinib) 103 genes (MI-219)	3 losses	1 gain, *TP53*
Paclitaxel	Docetaxel (R = 0.73)	1 gene (Docetaxel, Methotrexate) 1 gene (Docetaxel) 65 genes (Brivanib)	3 gains, *hsa-miR-556-5p*, p14ARF (*CDKN2A*)	1 gain, *hsa-miR-106a-3p*, *hsa-miR-187-5p*

Table 1. Continued

Drug	Correlated drugs	Genes overlapping with other drugs	Genomic features associated with sensitivity	Genomic features associated with resistance
Panobinostat	Vorinostat (R = 0.96), Quisinostat (R = 0.93), Belinostat (R = 0.91)	6 genes (Sirolimus, Quisinostat), 3 genes (Quisinostat) 14 (Vorinostat) 6 genes (Belinostat, Vorinostat, Quisinostat, Sirolimus) 6 genes (Belinostat, Vorinostat) 21 genes (Belinostat)	1 gain and 1 loss, *hsa-miR-185-3p*, *PIK3CA*	*CDH1* methylation
Quisinostat	Panobinostat (R = 0.93), Vorinostat (R = 0.87), Belinostat (R = 0.85)	6 genes (Sirolimus, Panobinostat), 3 genes (Panobinostat), 6 genes (Belinostat, Panobinostat, Sirolimus, Vorinostat)	1 gain and 1 loss, *CDH1, UNC13C*	*CDH1* methylation
Serdemetan	Decitabine (R = 0.56), Tipifarnib (R = 0.52), Veliparib (R = 0.51)	1 gene (Serdemetan, Mitoxantrone)	1 gain	1 gain and 2 losses, *ERBB2* amplification, *CCND1* amplification
Sirolimus		1 gene (17-AAG), 1 gene (Sorafenib), 6 genes (Panobinostat, Quisinostat), 6 genes (Belinostat, Panobinostat, Quisinostat, Vorinostat)	3 gains, *PIK3CA,* breast cancer subtype: luminal and ERBB2-overexpressing cell lines are more sensitive	*BRCA1*
Sorafenib		1 gene (Sirolimus)	1 gain	1 gain
Sunitinib	Erdafitinib (R = 0.48), Cisplatin (R = 0.47)	23 genes (Nutlin-3 and MI-219, however, in the opposing direction)	1 gain	1 gain and 1 loss
Tamoxifen			1 gain	

Table 1. Continued

Drug	Correlated drugs	Genes overlapping with other drugs	Genomic features associated with sensitivity	Genomic features associated with resistance
Tanespimycin (17-AAG)		1 gene (Sirolimus)	2 gains	1 gain, $p16^{INK4A}$ (*CDKN2A*), *RB1* signaling pathway ($p16^{INK4A}$ functions within this pathway)
Tipifarnib	Serdemetan (R = 0.52)	7 genes (Tivantinib)	2 gains	1 gain and 1 loss, *hsa-miR-629-5p, DST*
Tivantinib (ARQ197)	Docetaxel (R = 0.47)	7 genes (Tipifarnib)	1 gain, *hsa-let-7d-5p*, *hsa-miR-18a-5p*	*hsa-miR-637*
Vandetanib	Gefitinib (R = 0.56), Erlotinib (R = 0.47), Bortezomib (R = 0.47)	3 genes (Bortezomib)	3 gains	2 gains
Veliparib	Serdemetan (R = 0.51), Decitabine (R = 0.48)	13 genes (Mitoxantrone, 5-Fluorouracil)		1 gain
Vorinostat	Panobinostat (R = 0.96), Belinostat (R = 0.92), Quisinostat (R = 0.87)	14 genes (Panobinostat), 6 genes (Belinostat), 6 genes (Belinostat, Panobinostat, Quisinostat, Sirolimus) 6 genes (Belinostat, Panobinostat)	2 gains and 2 losses, *PIK3CA*	1 gain

For each drug the correlated drugs, including the correlation coefficient R, are displayed. The number of genes within the CNAs that also associated with another drug is listed. At the right side of the table all associated genomic features are listed in separate columns depending on their association with drug sensitivity or resistance. The gene *CDKN2A* encodes two structurally unrelated proteins within different reading frames ($p16^{INK4A}$ and p14ARF). Associations with both proteins were found and are listed for the respective protein.

Overall our studies showed that most drugs exhibited an association with CNAs (32 of 34) while associations with mutations (17 of 34) were less common and the least associations were seen with miRNAs (8 of 34). Among the drugs associating with a mutation, three showed an additional association with a miRNA (Paclitaxel, Tipifarnib and Panobinostat). While for Paclitaxel and Tipifarnib none of the associated miRNAs were associated with the respective aberration, this was not the case for Panobinostat. The Panobinostat-associated miRNA *hsa-miR-185-3p* did associate with *CDH1* methylation, with lower expression of this miRNA in *CDH1*-methylated cell lines. This

association is likely indirect as it cannot be explained by location of the two genes, given that *CDH1* lies on chromosome 16 and *hsa-miR-185-3p* on chromosome 22.

In the case of CNAs, several drug-associated miRNAs were associated with drug-associated CNAs. Two of the Bortezomib-associated CNAs showed higher *hsa-let-7a-5p* expression in samples with gains of these regions, however, all three genomic loci encoding *hsa-let-7a-5p* lie on different chromosomes than the respective gains. In the case of Docetaxel, a decrease in *hsa-miR-106a-3p* was seen in samples with a gain of a Docetaxel-associated CNA, while samples with a loss of this region showed the highest miRNA expression in line with the association of this gain with drug sensitivity while this miRNA was associated with resistance. *Hsa-miR-106a-3p* also associated with a Paclitaxel-associated CNA, again showing highest expression in samples with a loss and lowest in those with a gain – in line with the association of this gain with Paclitaxel sensitivity and *hsa-miR-106a-3p* with Paclitaxel resistance. Interestingly this Paclitaxel-associated CNA shares one gene, olfactory receptor family 4 subfamily C member 11 (*OR4C11*), with the aforementioned Docetaxel-associated CNA. However, besides a role as olfactory receptor, little is known about this gene and how it could potentially influence expression of this miRNA. The Tipifarnib-resistance associated miRNA *hsa-miR-629-5p* did show an association with a Tipifarnib-associated CNA, showing higher expression in those cell lines with a gain in the respective region albeit miRNA and CNA lie on different chromosomes. The respective gain, however, consists of eight zinc finger proteins and one zinc finger pseudogene. The positive association between this miRNA and the gain could potentially represent a regulatory role of one or of the several zinc finger proteins on expression of *hsa-miR-629-5p*.

A Veliparib-associated gain was furthermore associated with an increased expression of *hsa-miR-182-5p*, in line with the association of both with resistance to this drug. However, this miRNA and the respective CNA are located on different chromosomes and the association with upregulation of this miRNA is therefore not directly due to the gain but due to a different mechanism or coincidental.

Looking next whether drug-associated mutations were in addition associated with drug-associated CNAs, the following associations were significant: amplified *ERBB2* associated with an Erlotinib-associated gain. This gain encompasses only the gene *CDK12*, which lies a bit upstream of *ERBB2*. Not surprising the majority of the cell lines (37 of 42) showed the same copy number for *CDK12* as well as *ERBB2*, explaining the association found here. Amplification of *MDM2* was associated not only with JNJ-208 but also with a JNJ-208 associated gain. However, the respective CNA lies on a different

chromosome than *MDM2*. Mutated *CSMD3* was not only associated with Methotrexate resistance but as well with a Methotrexate-resistance associated gain. Four CNAs, that were associated with Nutlin-3 as well as MI-219, were associated with *TP53* mutation status, in line with their association with MI-219/Nutlin-3 resistance/sensitivity. *PIK3CA* mutation was associated with a Belinostat-associated gain, a Panobinostat-associated gain, two Sirolimus-associated gains and two Vorinostat-associated gains. The gains, as well as *PIK3CA* mutation associated with drug sensitivity and encompassed as minimum overlap the three genes *MIPOL1*, *FOXA1* and *SSTR1* (except for one Sirolimus-gain and one of the Vorinostat-gains). *CDH1* methylation showed an association with a Panobinostat-associated gain and a Quisinostat-associated gain. In both cases the gains were associated with sensitivity to the respective drug, while *CDH1* methylation was associated with drug resistance. These gains had 12 genes in common.

The loss of a CNA was associated with Serdemetan resistance and this CNA was additionally associated with the resistance-associated *CCND1* amplification. Interestingly two miRNA genes (*MIR34B* and *MIR34C*) lie within this CNA region. One could hypothesize that these miRNAs might target *CCND1* and loss of this region as well as amplification of *CCND1* could represent two ways that increase the CCND1 protein amount in cells. While there are two studies describing miRNAs from these genetic loci as targeting *CCND1* (Lee et al. 2011; Wang et al. 2016), the study of Lee et al. (Lee et al. 2011) shows this is only in certain scenarios the case (ER+ and *TP53* wt cells), while Wang et al. (Wang et al. 2016) do not show direct target-miRNA interaction. In our miRNA study we did not find an association between expression of miRNAs from the *MIR34B* and *MIR34C* genetic loci and Serdemetan response. This could have been influenced by the restricted number of cell lines in the miRNA study, by the fact that other genomic features play simply a greater role or there is a more complex regulation scenario which requires a more specialized data analysis to unravel. Of note is also that among the 36 cell lines used within the miRNA study (**chapter 4**) only one single cell line was ER+, *TP53* wt and without *CCND1* amplification.

A Brivanib-sensitivity associated gain was also associated with mutated *MAP3K1*, albeit in the opposing direction. Cell lines with mutated *MAP3K1* had predominantly no gain in this region, in line with the association of mutated *MAP3K1* with resistance while the gain was associated with drug sensitivity. A Doxorubicin-sensitivity associated gain associated also with mutated *MLL3*. None of the *MLL3* wildtype cell lines exhibited this gain, fitting the association of the gain, as well as mutated *MLL3* with drug sensitivity. Overall, several of the drug-associated genomic features also associate with each

other, albeit in most cases insufficient information is available to deduce whether this association presents different mechanisms to affect response to the same drug.

While our drug screening (**chapter 3**) and the association with mutations/CNAs (**chapter 5**) encompassed a large number of cell lines, drugs and genetic aberrations, there have been a number of similar studies (Heiser et al. 2012; Garnett et al. 2012; Barretina et al. 2012). Although there is some overlap between studies in regard to drugs and cell lines, and some findings were confirmed, there is less concordance between studies than one might expect. Extensive analyses have been performed comparing the two largest studies, the CCLE study (Barretina et al. 2012) and the GDSC (Garnett et al. 2012) in detail for agreements, discordance and potential reasons thereof (Haibe-Kains et al. 2013; The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015). These two analyses greatly illustrate common caveats in associating genomic features with drug response. Haibe-Kains and colleagues (Haibe-Kains et al. 2013) elaborate on differences between the studies and conclude that the main culprit for the differing outcomes lies within the drug screening, while measurements of genomic features show greater agreement between the studies by the GDSC (Garnett et al. 2012) and the CCLE (Barretina et al. 2012). A subsequent analysis by the CCLE and GDSC also comparing the two datasets (The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015), argues that the agreement between studies is better than presented by Haibe-Kains (Haibe-Kains et al. 2013), as e.g. analysis methods should take into account that typically only few cell lines respond versus an overwhelming majority of non-responders. Summarizing the lessons from these large drug screenings and their critical comparisons, one can state that when including different cancer types within one screening, it is useful to also account for this, as some cancer types are intrinsically resistant and common mutations in such a type represent an important confounding factor, leading to drug-mutation associations that cannot be reproduced in cell lines of a different cancer type (Garnett et al. 2012). Further challenges lie in the choice of analysis methods for drug screening data, due to very different drug response profiles. This includes the definition of resistance/sensitivity, as well as sufficient methodological sensitivity (data analysis statistics) in drug screening scenarios with only few responders; the last scenario and the impact of different analysis methods is well-illustrated by the CCLE/GDSC study (The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015). Finally, most pharmacogenomic studies differ in several points in regard to the actual drug screening, concentrations used and analysis methods used (Heiser et al. 2012; Garnett et al. 2012; Barretina et al. 2012). All these are further sources for differences in outcomes, well-illustrated also by Haibe-Kains *et al.* (Haibe-

Kains et al. 2013) and the shared CCLE/GDSC analysis (The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015). Of note is also, that known pharmacogenomic associations identified in one study but not in the other, often lie just below cut-off (The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015). Another issue is that different genomic features can lead to the same phenotypic drug sensitivity such as a mutation in one cell line, transcriptional regulation in another and gene expression in a third. This genomic redundancy greatly increases challenges in regard to data analysis to uncover such relationships, even more so, when only few cell lines are sensitive (Garnett et al. 2012; Barretina et al. 2012; The Cancer Cell Line Encyclopedia Consortium and The Genomics of Drug Sensitivity in Cancer Consortium 2015).

Based on the challenges observed with comparability of large pharmacogenomic studies, improvements in comparability could be made by greater standardization. A promising endeavor could be investments into the development as well as the implementation of gold-standard techniques/protocols to ensure lower variability within cell line screenings as well as in the subsequent analyses. Such an effort would improve comparability between studies and hopefully increase the success of such pre-clinical studies, decrease costs, speed up progress in the development of new cancer therapeutics as well as identification of sensitive patients.

circRNAs as cancer biomarkers

Besides the aforementioned drug screening, expression levels of circRNAs were in-depth characterized in 348 primary breast tumors (**chapter 7**). This study provided a detailed overview on circRNAs in regard to abundance, variance of expression as well as their expression specifics (e.g. size of adjacent introns, correlation with linear host transcript). Furthermore, associations of circRNAs with clinical and pathological features such as infiltration with lymphocytes, and prognosis were investigated. Finally, the potential of circRNAs as biomarkers was studied in **chapter 6** and **chapter 7**. While *circCNOT2* showed an association with clinical outcome, namely a worse progression-free survival in advanced breast cancer treated with aromatase inhibitors (AI) (**chapter 7**), *CDR1-AS* (*circCDR1*) was not associated with clinical outcome in ER-positive primary or advanced breast cancer (**chapter 6**). The fact that *circCNOT2* could be measured in plasma from breast cancer patients is promising (**chapter 7**) and could represent a clinical value of *circCNOT2* as a useful biomarker once validated in an independent cohort of AI-treated advanced breast cancer patients for sensitivity and specificity.

Next to these two circRNAs studied in more detail, a number of other circRNAs have been reported as biomarkers of patient outcome or other clinical parameters. Low expression of *circPVT1* is associated with decreased overall survival (OS) and disease-free survival (DFS) as well as T4 tumor stage and neural invasion in gastric cancer (Chen et al. 2017). In osteosarcoma, high *circUBAP2* expression was found to be prognostic for reduced OS and its expression was positively correlated to tumor stage; furthermore this circRNA exhibited oncogenic properties in functional studies (Zhang et al. 2017). In hepatocellular carcinoma low expression of *circMTO1* was associated with reduced survival, and functional experiments confirmed a tumor suppressor role of this circRNA (Han et al. 2017). Another circRNA, *circITCH*, showed an improved survival rate among hepatocellular carcinoma patients with high *circITCH* expression, as well indicative of a tumor suppressor role in this disease (Guo et al. 2017). While quite some studies have found circRNAs to be prognostic in different cancer types, only few have described a predictive role in patients. In osteosarcoma patients the circRNA *hsa_circ_0081001* was associated with chemoresistance, with higher expression levels in patients which exhibited chemoresistance, as well as showing higher expression in patients with a decreased OS and lung metastasis (Kun-Peng et al. 2018). In acute myeloid leukemia (AML), decreased *hsa_circ_0004277* expression was observed in newly diagnosed cases as well as refractory cases, while patients exhibiting complete remission showed the same levels of this circRNA as healthy controls (Li et al. 2017), indicating a role of this circRNA in monitoring chemotherapy success, albeit its predictive capability was not assessed in this study. Further research looking into the potential of circRNAs as predictive biomarkers in cancer are therefore warranted, especially in light of their stability (Jeck et al. 2013) as well as presence in plasma (Li et al. 2018), important characteristics for stable & easy-to-measure biomarkers.

Outlook – systems biology to improve therapy choices

This thesis showcases that resistance/sensitivity to a certain drug can be associated with – and potentially influenced by – multiple genomic features and that the type of feature associated with a drug varies substantially between drugs. It is therefore recommendable to perform broad investigations including miRNA expression, circRNA expression, mutations and CNAs to not overlook potential biomarkers for specific drugs. In the era of personalized medicine characterizing patients for biomarkers on the molecular biological level and combining these to a grand picture, next to traditional clinical information such as results from the pathological examination, will likely provide better guidance in regard to therapy choice and as a result optimize patient care.

In the future therefore, such a combination of several different genomic features might provide a more detailed insight into the tumor's biology as well as taking the individual genomic background of a patient into account when predicting aggressiveness of a tumor. Furthermore, it will help to determine which drugs would be the most promising and which ones will likely not bring benefit in defeating the cancer.

References

Barretina J, Caponigro G, Stransky N, et al (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*:603–607.

Carroll JS (2016) Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. Eur J Endocrinol *175*:R41–R49.

Chen J, Li Y, Zheng Q, et al (2017) Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. Cancer Lett *388*:208–219.

Garnett MJ, Edelman EJ, Heidorn SJ, et al (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*:570–575.

Guo W, Zhang J, Zhang D, et al (2017) Polymorphisms and expression pattern of circular RNA circ-ITCH contributes to the carcinogenesis of hepatocellular carcinoma. Oncotarget *8*:48169–48177.

Haibe-Kains B, El-Hachem N, Birkbak NJ, et al (2013) Inconsistency in large pharmacogenomic studies. Nature *504*:389–393.

Han D, Li J, Wang H, et al (2017) Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. Hepatology *66*:1151–1164.

Heiser LM, Sadanandam A, Kuo W-L, et al (2012) Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci *109*:2724–2729.

Incorvati JA, Shah S, Mu Y, Lu J (2013) Targeted therapy for HER2 positive breast cancer. J Hematol Oncol *6*:38.

Jeck WR, Sorrentino JA, Wang K, et al (2013) Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*:141–157.

Kun-Peng Z, Chun-Lin Z, Jian-Ping H, Lei Z (2018) A novel circulating hsa_circ_0081001 act as a potential biomarker for diagnosis and prognosis of osteosarcoma. Int J Biol Sci *14*:1513–1520.

Lee Y-M, Lee J-Y, Ho C-C, et al (2011) miRNA-34b as a tumor suppressor in estrogen-dependent growth of breast cancer cells. Breast Cancer Res *13*:R116.

Li T, Shao Y, Fu L, et al (2018) Plasma circular RNA profiling of patients with gastric cancer and their droplet digital RT-PCR detection. J Mol Med *96*:85–96.

Li W, Zhong C, Jiao J, et al (2017) Characterization of hsa_circ_0004277 as a New Biomarker for Acute Myeloid Leukemia via Circular RNA Profile and Bioinformatics Analysis. Int J Mol Sci *18*:E597.

Riaz M, van Jaarsveld MTM, Hollestelle A, et al (2013) miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. Breast Cancer Res BCR *15*:R33.

Rivera E, Gomez H (2010) Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone. Breast Cancer Res BCR *12*:S2.

The Cancer Cell Line Encyclopedia Consortium, The Genomics of Drug Sensitivity in Cancer Consortium (2015) Pharmacogenomic agreement between two cancer cell line data sets. Nature *528*:84–87.

Uhr K, Prager-van der Smissen WJC, Heine AAJ, et al (2019) MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines. PloS One *14*:e0216400.

Wang F, Lu J, Peng X, et al (2016) Integrated analysis of microRNA regulatory network in nasopharyngeal carcinoma with deep sequencing. J Exp Clin Cancer Res *35*:17.

Zhang H, Wang G, Ding C, et al (2017) Increased circular RNA UBAP2 acts as a sponge of miR-143 to promote osteosarcoma progression. Oncotarget *8*:61687–61697.

Chapter 9

Summary/Samenvatting

Summary

Significant progress has been made in breast cancer research over the last decades to reduce the number of cancer related deaths and to increase the quality of life by improving early detection as well as through the introduction of novel therapies to treat primary disease. On top of that, companion diagnostics research has been performed to identify patient populations which require additional treatment as well as those that are most receptive to a specific treatment. Current patient stratification is based on classical clinical and pathological markers as well as innovative genomic biomarkers such as multi-gene predictors (detailed in the introduction in **chapter 1**). Despite these improvements a significant portion of breast cancer patients develop resistance to currently available treatments. Also, current biomarkers for predicting a patient's response to known and novel drug treatments are sparse and there is an urgent need for development of such markers. This thesis focused on the clinical value of various types of non-coding RNAs as biomarkers with regard to drug response and clinical outcome (in the scope of the thesis, **chapter 2**). In **chapter 3**, within a significant drug screening effort currently clinically used drugs, newly developed drugs as well as drugs used in other disease scenarios were investigated for their sensitivity profiles in breast cancer cell lines. This study showed that various drugs exhibited positively or negatively correlated response patterns, indicating that sensitivity to one drug might be indicative of (in)sensitivity to another. Breast cancer subtype played only a minor role when looking at drug sensitivity with only one drug showing a subtype-dependent response. Interestingly, the majority of correlated drugs did not share biological pathways based on gene expression, except for one highly correlated drug pair having one pathway in common. In **chapter 4**, we studied miRNAs for their potential as biomarkers to predict response to the drugs screened in **chapter 3**. MiRNA expression was influenced by breast cancer subtype and upon exclusion of subtype-associated miRNAs, eleven different miRNAs were found to be associated with response to eight of the 34 investigated drugs. Highly-correlated drugs shared miRNAs, albeit none of the drugs shared all associated miRNAs. Pathway analyses led to the discovery of several miRNA-associated pathways known to be also affected by the respective drugs (e.g. the G2-M checkpoint is targeted by taxanes and associated with *hsa-miR-187-5p*). In **chapter 5**, somatically acquired mutations, as well as copy number aberrations, which can influence drug response as well, were investigated for their potential as biomarkers using the drug screening dataset generated in **chapter 3**. Mutation, methylation or amplification of 17 genes was associated with response to 17 drugs, including several earlier reported findings (e.g. mutated *TP53* with Nutlin-3 and MI-219 resistance or *ERBB2* amplification with Erlotinib sensitivity to name two examples). In one instance we discovered that the inactivation of different genes within

the *RB1* tumor-suppressor pathway (p16^{INK4A} mutation, p16^{INK4A} methylation and *RB1* mutation) showed an even stronger association with 17-AAG resistance than p16^{INK4A} mutation alone, which corroborated earlier published observations on single genes within this pathway and response to 17-AAG. Besides mutations, 89 associations between copy number aberrations and drug response were found. Several of the drug-associated regions spanned known breast cancer driver genes such as *ARID1A*, *FOXA1* or *ZNF217.*

With regard to the biomarker value of non-coding RNA species studied in this thesis, in **chapter 6** specific non-coding RNAs were investigated for their predictive and prognostic value. In this respect, the miRNA *hsa-miR-7* had been previously identified as a prognostic marker in ER-positive breast cancer for poor metastasis-free survival, and since the non-coding RNA *CDR1-AS* counteracts *hsa-miR-7,* the role of *CDR1-AS* as a prognostic marker in breast cancer was studied. Additionally, *hsa-miR-7* and *CDR1-AS* were examined for their potential as predictive biomarkers in tamoxifen-treated advanced breast cancer patients. *Hsa-miR-7* exhibited, next to the known prognostic capability, predictive value for poor tamoxifen response as well as shorter progression-free survival and overall survival in the advanced setting. *CDR1-AS*, however, apart from some weak associations with clinical parameters and mainly stromal cell expression, neither showed prognostic nor predictive biomarker capability. In **chapter 7**, the global circRNA expression landscape was described in primary breast tumors and the discovered circRNAs were characterized in regard to their genetic environment, their relation to the linear transcript of their host gene, and their relevance in breast cancer. Clustering of breast tumors based on circRNA expression identified subgroups with differences in biological and clinical characteristics. One of the discovered circRNAs, *circCNOT2*, showed an impact on breast cancer cell line viability in functional experiments. Furthermore, this circRNA could predict progression-free survival in aromatase inhibitor treated advanced breast cancer patients and was also found in cell-free RNA of patients' plasma.

Finally, in the discussion in **chapter 8**, the results of the studies linked to the cell line drug screening (**chapter 3**, **4**, **5**) were summarized and analyzed for potential connections between drug-associated miRNAs and drug-associated genetic changes. Furthermore, since drug screening studies show so far limited inter-study comparability, an overview of the important culprits (drug screening protocols, analyses methods) is described, and better standardization is suggested as a solution. Additional problems in biomarker identification for drug response are also discussed. As circRNAs have been gaining a lot of attention recently, and many studies investigated their role in cancer, several

circRNAs capable to predict prognosis were identified and an overview of these is given in **chapter 8**. So far, however, hardly any circRNAs predictive of drug response have been reported in cancer and the predictive circRNA identified in **chapter 7** is therefore thus far only one of few at this moment.

To conclude this thesis: The studies on cell lines provided a broad overview of drug sensitivity and associated genomic characteristics (miRNAs, mutations, CNAs) in breast cancer which are, when validated, potentially useful as biomarkers; the studies in breast tumors extensively describe the relatively new RNA species, circRNAs, and determine their predictive and prognostic potential in breast cancer.

Samenvatting

De afgelopen decennia is door vroeg-detectie en de introductie van nieuwe therapieën het aantal sterfgevallen aan borstkanker sterk verminderd en als mede de kwaliteit van leven met de ziekte sterk verbeterd. Aanvullend diagnostisch onderzoek heeft er bovendien voor gezorgd dat patiënten geïdentificeerd kunnen worden voor wie aanvullende systemische behandeling noodzakelijk is en wie het meest ontvankelijk is voor een specifieke behandeling. Zoals gedetailleerd beschreven in **de inleiding** (**hoofdstuk 1**) is de huidige patiëntenstratificatie gebaseerd op klassieke klinische en pathologische factoren aangevuld met innovatieve genomische biomerkers zoals multi-gen-gebaseerde voorspellers. Ondanks deze vooruitgang ontwikkelt een aanzienlijk deel van de borstkankerpatiënten resistentie tegen de momenteel beschikbare medicijnen. Ook zijn biomerkers die de respons van een patiënt op bekende en nieuwe medicamenteuze behandelingen kunnen voorspellen schaars. Kortom er is een dringende behoefte aan de ontwikkeling van dergelijke merkers en **het doel** van dit proefschrift zoals geschreven in **hoofdstuk 2** was om te onderzoeken of niet-coderende RNA's bruikbare biomerkers zouden kunnen leveren om de gevoeligheid van nieuwe en bestaande medicijnen of het ziektebeloop in algemene zin te voorspellen. In **hoofdstuk 3** hebben we in een uitgebreid aantal cellijnmodellen voor borstkanker de gevoeligheid voor 34 gangbare en nieuw-ontwikkelde geneesmiddelen in kaart gebracht. Deze studie toonde aan dat de gevoeligheid van cellijnen voor verschillende geneesmiddelen positief of negatief gecorreleerd gedrag vertoonde, hetgeen aangeeft dat gevoeligheid voor één geneesmiddel een indicatie kan zijn voor (on)gevoeligheid voor een ander. Bestaande sub-groepering van borstkanker (bijvoorbeeld hormoon gevoelig of niet) speelde hierbij een ondergeschikte rol. Opmerkelijk genoeg konden we in het onderzochte borstkankercellijnpanel met uitzondering van één sterk gecorreleerd geneesmiddelpaar, op basis van genexpressie geen biologische processen identificeren die gedeeld werden door de 29 geneesmiddelen die een gecorreleerd gedrag vertoonden. In **hoofdstuk 4** hebben we vervolgens bestudeerd of kleine RNAs, de zogenaamde microRNA's, als biomerkers kunnen dienen om de gevoeligheid voor de in **hoofdstuk 3** onderzochte geneesmiddelen te voorspellen. Om potentiële bias te voorkomen hebben we microRNAs die sterk met bestaande borstkankersubgroeperingen geassocieerd waren buiten het onderzoek gelaten. Hierna bleven er elf verschillende miRNA's over die geassocieerd waren met respons van de cellijnen op acht van de 34 onderzochte medicijnen. Sterk gecorreleerde geneesmiddelen deelden miRNA's, hoewel geen van de geneesmiddelen alle geassocieerde miRNA's met elkaar deelde. Biologische proces-analyses leidden tot de ontdekking van verschillende miRNA-geassocieerde processen waarvan reeds bekend is dat ze ook worden beïnvloed door de respectieve geneesmiddelen (bijv. het proces

dat de transitie van G2 naar M in celdeling controleert is een doelwit van taxanen en is eveneens geassocieerd met de expressie van *hsa-miR-187-5p*). In **hoofdstuk 5** werden met behulp van de in **hoofdstuk 3** gegenereerde geneesmiddelengevoeligheidsdata somatisch verworven mutaties en genkopieën onderzocht op hun potentieel als biomerkers. Mutatie of methylatie of amplificatie van 17 genen waren geassocieerd met gevoeligheid voor 17 van de geteste geneesmiddelen. Onder de gevonden associaties bevonden zich verschillende eerder gerapporteerde bevindingen. Om twee voorbeelden te noemen: *TP53* gemuteerde cellijnen vertoonden resistentie tegen Nutlin-3 en MI-219 en *ERBB2*-amplificatie voorspelde Erlotinib-gevoeligheid. Het onderzoek gaf ook nieuwe inzichten. Zo ontdekten we dat de inactivatie van verschillende genen uit het *RB1* gedreven signaalpad dat tumorgroei onderdrukt ($p16^{INK4A}$-mutatie, $p16^{INK4A}$-methylatie en *RB1*-mutatie) tezamen een sterkere associatie met 17-AAG-resistentie vertoonde dan $p16^{INK4A}$-mutatie alleen, hetgeen in overeenstemming is met eerder gepubliceerde waarnemingen dat deze genen elk individueel de 17-AAG gevoeligheid beïnvloeden. Naast mutaties vonden we 89 gebieden op het humane genoom, inclusief regio's waar bekende borstkanker-stimulerende genen zoals *ARID1A*, *FOXA1* of *ZNF217* liggen, waarvan het aantal kopieën de medicijngevoeligheid bepaalde.

In aanvulling op bovenstaand onderzoek in cellijnmodellen is in het tweede deel van het proefschrift in weefsels van borstkankerpatienten de klinische waarde van niet-coderende RNA's onderzocht met de focus op hun voorspellende waarde m.b.t. agressiviteit en/of gevoeligheid voor hormonale therapie. In eerder onderzoek was aangetoond dat *hsa-miR-7* in ER-positief borstkanker een voorspeller was voor een korte metastasevrije overleving, en aangezien het niet-coderende RNA *CDR1-AS hsa-miR-7* tegenwerkt, werd in **hoofdstuk 6** de rol van *CDR1-AS* als voorspeller van aggressiviteit bij borstkanker bestudeerd. Bovendien werden zowel *hsa-miR-7* als *CDR1-AS* onderzocht op hun potentieel als voorspellende biomerkers bij gevorderde borstkankerpatiënten die behandeld waren met het anti-hormoon tamoxifen. Naast de eerder aangetoonde maat voor agressiviteit, vertoonde *hsa-miR-7* in deze patientengroep eveneens voorspellende waarde voor een slechte respons op tamoxifen evenals een kortere progressievrije en totale overleving. *CDR1-AS* bleek daarentegen, afgezien van enkele zwakke associaties met klinische parameters en voornamelijk expressie in stroma van de tumor, noch een maat te zijn voor aggressiviteit noch voorspellende waarde m.b.t. gevoeligheid voor hormonale therapie te bezitten. In **hoofdstuk 7** is voor het eerst in een groot primair borstkankercohort het expressiepatroon van alle circulaire RNAs beschreven. Tevens werden de ontdekte circRNA's gerelateerd aan de lineaire transcripten waaruit ze zijn ontstaan, aan de genetische karakteristieken van de tumoren, en aan klinisch relevante factoren bij borstkanker. Groepering van borsttumoren op basis van de

expressie van deze circulaire RNAs identificeerde subgroepen met verschillen in biologische en klinische karakteristieken. Eén van de ontdekte circRNA's, *circCNOT2*, bleek in functionele experimenten de vitaliteit van borstkankercellijnen negatief te beïnvloeden. Bovendien kon dit circRNA de progressievrije overleving voorspellen bij gevorderde borstkankerpatiënten die behandeld waren met een aromataseremmer en was het detecteerbaar in het plasma van borstkankerpatiënten.

Tot slot zijn in **hoofdstuk 8** de resultaten van het proefschrift bediscussieerd. Het onderzoek naar geneesmiddelgevoeligheid van borstkankercellijnen (**hoofdstuk 3**, **4**, **5**) en geassocieerde miRNA's en genetische veranderingen is geïntegreerd waaruit bleek dat de relatie tussen medicijngevoeligheid gerelateerde miRNAs en gevonden genetische factoren met name indirect was. Daarnaast, aangezien onderzoeken naar medicijngevoeligheid in cellijnen een beperkte vergelijkbaarheid tonen, wordt een overzicht van de belangrijkste oorzaken beschreven (gebruikte protocollen en analysemethoden) en wordt een standaardisatie daarvan als oplossing voorgesteld. Bijkomende problemen met betrekking tot de identificatie van biomerkers met voorspellende waarde m.b.t. geneesmiddelgevoeligheid worden ook besproken. Omdat circRNA's de laatste tijd veel aandacht hebben gekregen, en diverse studies hun rol in kanker onderzochten en hun waarde als prognostische merkers beschreven, werd in **hoofdstuk 8** hiervan een overzicht gegeven. CircRNA's met voorspellende waarde m.b.t. geneesmiddelgevoeligheid zijn er nog nauwelijks en het voorspellende circRNA dat geïdentificeerd werd in **hoofdstuk 7** is daarom tot nu toe één van eerste.

Dit proefschrift samenvattend kunnen we stellen dat de studies op cellijnen inzicht hebben gegeven in de relatie tussen genomische kenmerken (miRNA's, mutaties, kopieënaantal van gebieden op het humane genoom) en de gevoeligheid van borstkanker voor een breed scala aan geneesmiddelen, die, indien onafhankelijk gevalideerd, mogelijk gebruikt kunnen worden als predictieve biomerkers; de studies in borsttumoren van patiënten beschrijven uitgebreid het relatief nieuwe RNA-soort, circRNA's, en stellen hun voorspellende potentieel bij borstkanker vast.

Appendices

Curriculum vitae

Katharina Uhr was born on 20th of September 1984 in Tönisvorst, Germany. In October 2004 she started her biology studies at the Ernst-Moritz-Arndt University in Greifswald, Germany. Within her studies she followed in 2006 a 10-week internship at the Institute of Human Genetics at the GSF Helmholtz research center in Neuherberg, next to Munich, Germany, investigating mutations in heritable mitochondrial diseases. During her university studies she worked from November 2007 until March 2008 as a student assistant in the junior research group Transcriptomics/Functional Genome Research of the Interfaculty Institute for Genetics and Functional Genome Research, Department Functional Genomics of the Ernst Moritz Arndt University aiding the search for genetic susceptibility markers of aggressive periodontitis. Subsequently she performed the research for her diploma thesis (topic: approaches to identify genetic susceptibility loci for Aggressive Periodontitis (AgP) by genome-wide association and candidate gene analyses) in the same research group. In August 2009 she completed her biology studies, majoring in genetics with the two minors immunology and molecular biology & microbiology. Katharina's work on aggressive periodontitis culminated in one replication study confirming previous genetic susceptibility loci for the disease (2nd author) as well as a technical study on the feasibility of allelotyping studies using pooled samples (4th author).

In September 2009 she started a PhD program investigating microRNAs in breast cancer at the Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus Medical Center, Rotterdam, The Netherlands. In April 2011 she received a travel grant from the René Vogels Stichting and spent three months at the laboratory of Robert B. Darnell, Rockefeller University, New York City, New York, USA, to learn the Ago HITS-CLIP technique.

Katharina's PhD studies led to four publications (three 1st author and one 3rd author) and one so far unpublished manuscript linking different non-coding genes, mutations and CNAs to drug response in breast cancer.

From May 2014 until July 2017 Katharina was working as Genomics Application Support Scientist (tier 3) at Agilent Technologies providing remote customer-facing support for QPCR, microarrays, next-generation sequencing products and classical molecular biological products.

PhD portfolio

Name PhD student:	Katharina Uhr
Erasmus MC department:	Medical Oncology
Research school:	Molecular Medicine
Promotor:	Prof.dr.ir. John W. M. Martens
Co-Promotor:	Prof.dr. John A. Foekens

General academic skills	**Year**	**ECTS**
CGC Valorisation Course „Intellectual Property and Commercialization“	2009	0.3
Integrity in Medical Research	2010	0.08
Research Management for PhD students and Postdocs	2010	0.75
Photoshop CS3 Workshop for PhD students and other researchers	2010	0.25
Biobusiness Summer School	2011	1.75
Workshop Scientific Career and Beyond	2012	0.05
Biomedical English Writing and Communication	2012-2013	4
In-depth courses (e.g. research school)		
Biomedical Research Techniques VIII	2009	1.5
Browsing Genes and Genomes with Ensembl	2009	0.35
Course Basic and Translational Oncology	2009	1.2
Course Molecular Medicine	2010	0.75
Basic Introduction Course on SPSS	2010	0.3
Systems Biology Applied Bioinformatics, Sequences and Variations	2010	0.5
Introduction to Bioinformatics for Molecular Biologists (Utrecht)	2010	3
Classical Methods for Data-Analysis	2010	5.75
SNPs and Human Diseases	2010	2
Basic Linux Course (Leiden)	2010	0.25
Next Generation Sequencing (NGS) Data Analysis (Leiden)	2010	0.75
Basic Course on R	2010	1
The Advanced Course Molecular Immunology	2011	0.45
International conferences		
AACR Translational Cancer Medicine 2010 Europe (Amsterdam)	2010	1
International Workshop on Small RNA in Cancer, Inflammation and Aging (Kopenhagen)	2012	1.6
104th AACR Annual Meeting (Washington D.C.)	2013	1.8

Seminars and workshops		
Journal Club	2009-2013	5
MiRNA Seminar	2009-2011	0.3
Bridge Meeting Molecular Medicine Bioinformatics	2009-2011	0.4
Flow Day (FACS)	2009	0.3
JNI Seminar	2009-2013	2.7
Cancer Workshop Woudestein	2009	0.15
Lecture Series „Frontiers in Science in the Low Countries“	2009	0.08
Breast Cancer Meeting	2009-2013	2.5
JNI Symposium „The Stem of Cancer“	2009	0.3
Autumn Symposium „Cost-Effective Interventions in Health Care: from Evaluation to Application“	2009	0.15
Molecular Medicine Day	2010, 2011, 2013	0.9
High-Throughput Sequencing of Pathogen Transcriptomes: the Campylobacter Example	2010	0.04
Oncology Lecture	2010-2013	0.2
Next Generation Sequencing Seminar	2010-2013	0.3
Matthias Selbach: Investigating Proteome Dynamics with Pulsed SILAC	2010	0.04
CGC Annual Scientific Meeting	2010-2012	0.9
JNI New Years Symposium	2010	0.1
PhD Day Erasmus MC	2010	0.14
Get out of your Lab Days	2011	0.44
Seminar „GATA3 and Breast Cancer“	2012	0.04
NVvO (Nederlandse Vereniging voor Oncologie) Oncologiedag „Longkanker“	2012	0.25
Teaching/student supervision		
Supervising HBO student	2011	9.5
Giving an internal training on transfection of miRNA mimics	2012	0.14
Supervising HBO student	2012-2013	13.6
Presentations (ECTS are included above)		
JNI seminar presentation	2012	
Poster presentation Molecular Medicine Day	2013	
Poster presentation International Workshop on Small RNA in Cancer, Inflammation and Aging (Kopenhagen)	2012	
Journal Club presentations	2009, 2010, 2012	
Breast Cancer meeting presentations	2012, 2013	
Total:		**68.85**

List of publications

Uhr K, Prager-van der Smissen WJC, Heine AAJ, Ozturk B, van Jaarsveld MTM, Boersma AWM, Jager A, Wiemer EAC, Smid M, Foekens JA, Martens JWM. (2019) MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines. PloS One 14:e0216400. doi: 10.1371/journal.pone.0216400

Smid M, Wilting SM, **Uhr K**, Rodríguez-González FG, de Weerd V, Prager-Van der Smissen WJC, van der Vlugt-Daane M, van Galen A, Nik-Zainal S, Butler A, Martin S, Davies HR, Staaf J, van de Vijver MJ, Richardson AL, MacGrogan G, Salgado R, van den Eynden GGGM, Purdie CA, Thompson AM, Caldas C, Span PN, Sweep FCGJ, Simpson PT, Lakhani SR, Van Laere S, Desmedt C, Paradiso A, Eyfjord J, Broeks A, Vincent-Salomon A, Futreal AP, Knappskog S, King T, Viari A, Børresen-Dale AL, Stunnenberg HG, Stratton M, Foekens JA, Sieuwerts AM, Martens JWM. (2019) The circular RNome of primary breast cancer. Genome Res 29:356–366. doi: 10.1101/gr.238121.118

Uhr K, Sieuwerts AM, de Weerd V, Smid M, Hammerl D, Foekens JA, Martens JWM. (2018) Association of microRNA-7 and its binding partner CDR1-AS with the prognosis and prediction of 1st-line tamoxifen therapy in breast cancer. Sci Rep 8:9657. doi: 10.1038/s41598-018-27987-w

Uhr K, Prager-van der Smissen WJ, Heine AA, Ozturk B, Smid M, Göhlmann HW, Jager A, Foekens JA, Martens JWM. (2015) Understanding drugs in breast cancer through drug sensitivity screening. SpringerPlus 4:611. doi: 10.1186/s40064-015-1406-8

Teumer A, Ernst FD, Wiechert A, **Uhr K**, Nauck M, Petersmann A, Völzke H, Völker U, Homuth G. (2013) Comparison of genotyping using pooled DNA samples (allelotyping) and individual genotyping using the affymetrix genome-wide human SNP array 6.0. BMC Genomics 14:506. doi: 10.1186/1471-2164-14-506

Ernst FD, **Uhr K**, Teumer A, Fanghänel J, Schulz S, Noack B, Gonzales J, Reichert S, Eickholz P, Holtfreter B, Meisel P, Linden GJ, Homuth G, Kocher T. (2010) Replication of the association of chromosomal region 9p21.3 with generalized aggressive periodontitis (gAgP) using an independent case-control cohort. BMC Med Genet 11:119. doi: 10.1186/1471-2350-11-119

Acknowledgement

Upon starting this thesis a lot of challenges were lying ahead. First, obtaining the required permits took much longer than anticipated and then the original research project, the functional characterization of specific miRNAs, did not lead to useable results albeit following a multitude of different approaches.

Within my PhD period there were more than ten different projects aiming to unravel the biology behind certain miRNAs. Although most did not succeed in publishable results, this breadth of approaches gave me the opportunity to work with many different people, learn a wide variety of techniques and broaden my knowledge in molecular biological research.

First, I would like to thank my current promotor and previous co-promotor, Prof. Dr. John Martens, for his enthusiasm about research, the opportunity to get in touch with many different research techniques as well as our regular project discussions. His learning by doing approach taught me to approach challenges without hesitation and with confidence. Furthermore, I would like to thank him for pointing me to the Ago HITS-CLIP technique and furthering a research stay at the laboratory of Prof. Dr. Robert B. Darnell at the Rockefeller University in New York to learn this technique. Due to the large and diverse number of projects, the freedom in building up projects from scratch and executing all stages as well as writing some small grant proposals to obtain project-based reagent funding from a company, I was in the lucky position to enrich my knowledge in many different areas of research within my PhD trajectory. Upon the end of my contract John Martens kept in touch and I would like to thank him for making time for phone meetings – be it in the evening, the weekend or during a very busy conference. Thank you!

Second, I would like to thank my former promoter and current co-promoter, Prof. Dr. John Foekens, for his precise and critical scientific evaluations of laboratory results and his well-thought suggestions on subsequent steps. Additionally, I would like to thank him for making sure all regulations regarding publications for the PhD were met, his suggestions for eligible journals and his rigor in assessing manuscripts for errors and completeness. Furthermore, I would like to acknowledge that John Foekens kept being involved in meetings to further my thesis even after his retirement. Thank you!

The committee members for my PhD defense I would like to thank for their thorough examination of my thesis and their feedback. Thank you for investing your time and efforts in this last stage of my PhD trajectory.

Next, I would like to thank the René Vogels Stichting for enabling me to learn the Ago HITS-CLIP technique at the laboratory of Prof. Dr. Robert B. Darnell at the Rockefeller University in New York and for funding this endeavor. This has been an exceptional chance and I am very grateful for having had this opportunity.

Within the Ago HITS-CLIP project, I would like to thank Prof. Dr. Robert B. Darnell for giving me the opportunity to work in his laboratory. Christina Marney and the rest of the lab I would like to thank for teaching me the Ago HITS-CLIP technique and aiding me with different parts of the protocol.

For her friendly and supportive talks and actions, I would like to thank Prof. Dr. Els Berns.

When I just started my PhD position I had only worked with few techniques and my thanks goes to Dr. Francisco German Rodriguez Gonzalez, who accompanied me in the beginning, taught me about cell culture as well as cell assays, cloning and Western Blot. Sharing his experience proved very helpful in starting off my own experiments.

For teaching me RNA isolation, qPCR and its analysis as well as aiding me with pri-microRNA profiling I would like to thank Dr. Anieta Sieuwerts. Especially her elaborate Excel spreadsheets for streamlined data analysis and quality control are a memorable masterpiece and inspired me to create my own (though much simpler ones). Huge thanks go out to her for contributing massively on the study of *CDR1-AS* in breast tumors. It was a true pleasure to work with her not only on this article but over the course of the years. I appreciated discussing with her the best experimental approaches in such a quick, goal-oriented and thoughtful manner as well as sharing a coffee/drink at the end of these discussions. Her passing has come as a great shock and I still cannot imagine not meeting her in the department, swiftly coming around the corner with a big smile and the obligatory coffee mug in her hand.

With Dr. Elena Martens I shared many interesting research discussions on deciphering interactions among non-coding RNAs. Furthermore, she helped me as well with understanding CLC Bio and I would like to thank her for her help as well as for sharing her immense knowledge, clever experimental approaches and ideas with me. I always found our research discussions very inspiring.

For her help on the mutation chapter, I would like to thank Dr. Antoinette Hollestelle, as well as for the many interesting research discussions we had over the years. I appreciate finally having you as a co-author on one of my chapters/publications.

Within my Western Blotting projects, Cynthia Sitaram and Dr. Arzu Umar, shared their knowledge with me and I would like to thank them for their help to get my experiments of the ground.

Next, I would like to thank Marcel Smid for running countless bioinformatic analyses for me, helping me to generate figures for different publications (especially my first publication) as well as being a sparring partner for in-depth data analysis discussions. Not to forget the countless questions you provided answers for and showing me some of the programs you used – enabling me to learn more about data display options, speeding up the selection of relevant results as well as pursuing additional types of analyses.

The first friend I made in the Netherlands, Anouk Heine, I would like to thank for her help on my projects by writing me customized scripts and running multiple bioinformatic analyses. Secondly, I would like to thank her for teaching me Dutch by forcing me to speak Dutch for one hour every time we met. Through all these years we remained friends and I have to say no course has taught me as much Dutch as Anouk. You also helped me integrate and understand Dutch society better, which I appreciate a lot.

Furthermore, I would like to thank Joan Bolt for her help with and advice on cell culture, administrative issues as well as being always there and helping out with every type of problem, be it ordering reagents or dealing with Dutch forms. She really is the heart of the lab, always friendly and helpful on all the issues we PhD students came across.

For helping me with different cell culture projects, I would like to thank Wendy Prager-v.d. Smissen, as well as for teaching me how to work with lentiviruses, cell assays and sharing her great knowledge on experiments with me. It was always a pleasure to work with you!

I would like to thank Mieke Timmermans and Anita Trapman for teaching me about Cytospins and IHC. Mieke I would furthermore like to thank for helping me choose the right antibodies for my experiments as well as sharing her knowledge on IHC, antibodies and reagents for improving antibody stainings.

For her help on the statistical analyses of different projects I would like to thank Maxime Look.

One of my office mates and fellow PhD student, Muhammad, I would like to thank for the interesting research discussions as well as for introducing me to the mouse facilities and the associated microscope, which enabled me to take luminescent images of living cells.

My desk neighbor and friend, Diana, I would like to thank for the exchange of ideas, your inspiring enthusiasm, unlimited optimism, your motivating talks and your support. Your different viewpoints were enriching and gave me new insights.

My fellow PhD students, Mila, Rene, Marjolein and Tommaso, I would like to thank for the friendly conversations, discussions and lunches together.

Louet, my student for about half a year, I would like to thank you for the immense help you have been in optimizing miRNA mimic/knockdown transfection in multiple cell lines. It was nice to share the workload with somebody so dedicated and precise. It was a pleasure to work with you!

Jean, Maurice, Renee as well as the rest of the lab not named here: I would like to thank you for many interesting conversations, viewpoints as well as hearing about your experimental approaches/methods you applied in your daily work.

My fellow PhD students Raju, Farhad, Esther, Ingrid and Paul, I would like to thank you for interesting discussions as well as being part of different PhD student group meet-ups.

My fellow miRNA PhD students and friends Marijn and Caroline, I would like to thank for the exchange on miRNA research techniques, culprits and possible solutions and of course the much beloved ice cream breaks.

For their support, discussions on research and information on practical PhD matters over a coffee or a drink, I would like to thank my friends Marcia, Lale, Ines and Ronak.

During my time in Rotterdam, I spent a lot of time with my friend Charlotte and I would like to thank her for her support, the many dinners and good times we shared.

Furthermore, I would like to thank Marie-Evelyn for supporting me and spending time outside work.

Andrée and Vanessa I would like to thank for encouraging me to continue working on my PhD after starting a new job as well as for motivating and supporting me.

Further, I would like to thank all the other people I interacted with from other departments, workgroups, the JNI or the different courses for nice conversations, new input and some distraction at social gatherings.

My friend Ute I would like to thank for pushing each other forward in our PhD projects. Furthermore, I would like to thank her for her support as well as for the time spent outside Erasmus MC.

For motivating me and providing some much needed distractions, I would like to thank my friends Rachel, Karolina, Ivana and Alex. And additionally I would like to acknowledge the help from Karolina with my questions on statistics.

My friends Janna, Anne-Elise, Amber and Leonie, I would like to thank for all the dinners and drinks. At the many gatherings, you have given me a lot of support, advice, motivation to keep going but also great distractions from work. Thank you!

Rebecca, Carina, Karola und Mia: Ich möchte euch dafür danken, dass ihr es geschafft habt, dass ich mich in Amsterdam zuhause fühle und für die motivierenden Gespräche.

Schließlich möchte ich meiner Mutter für ihre grenzenlose Unterstützung in den schwierigen Zeiten danken; sowie dafür, dass sie mich immer motiviert hat, trotz aller Gegenschläge weiter an meiner Doktorarbeit zu arbeiten. Ich möchte auch meinem Vater für seine Unterstützung und seine Ratschläge danken, sowie meiner Schwester für ihre kritischen Beurteilungen, ihre Unterstützung und ihre Fähigkeit mich zu motivieren. Meiner Tante Li möchte ich für ihre Unterstützung danken, die es mir ermöglicht hat meine Doktorarbeit abzuschließen.