Genetic Factors

for Breast Cancer Susceptibility and Clinical Outcome

Jingjing Liu

Genetic Factors for Breast Cancer Susceptibility and Clinical Outcome

Jingjing Liu

The studies described in this thesis were performed at the Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. The PhD program was supported by Chinese Scholarship Council. The research described here was performed within the framework of the Erasmus Postgraduate School of Molecular Medicine.

The printing of this thesis has been supported by the Erasmus University and the Department of Medical Oncology of the Erasmus MC Cancer Institute.

ISBN: 978-94-6299-674-8

Cover design, printing and binding by Ridderprint BV.

Genetic Factors for Breast Cancer Susceptibility and Clinical Outcome

Genetische factoren betrokken bij de gevoeligheid en klinische uitkomst van borstkanker

Thesis

to obtain in the degree of Doctor from Erasmus University Rotterdam

by the command of the rector magnificus

Prof. dr. H. A. P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Wednesday 1st of November 2017 at 11:30 hours

by

Jingjing Liu

Born in People's Republic of China

Erasmus University Rotterdam

Doctoral committee:

Promotor	Prof. dr. J. A. Foekens
Other members	Prof. dr. G. W. Jenster
	Prof. dr. P. Devilee
	Dr. P. N. Span
Copromotor	Dr. A. Hollestelle

CONTENTS

Chapter 1

General introduction

General introduction

1.1 Epidemiology of breast cancer

Breast cancer is one of the most frequently diagnosed cancers and the second leading cause of death in Western women. In 2012, this disease was diagnosed in nearly 1.7 million women and was responsible for more than 521,900 deaths Worldwide[1]. In the United States, the estimated number of new cases was 231,840 in 2015, accounting for 29% of all new cancers in women[2]. The incidence of breast cancer is rising in most countries and is expected to rise even further in the next 20 years[3,4]. However, survival of patients with breast cancer has increased, due to earlier detection and more effective treatment modalities amongst others[5]. International variation exists in breast cancer incidence rates. Rates are high in developed countries in Northern America and Northern and Western Europe, whereas rates are low in most of Africa and Asia. However, breast cancer incidence rates have been rising in countries with low rates in recent years as well due to altered reproductive behavior, increased obesity and decreased physical activity[1]. The risk of developing breast cancer is related to many factors such as age, gender, lifestyle, family history of breast cancer and so on. The reproductive risk factors including estrogen exposure are associated with breast cancer risk[6].

1.2 Breast cancer subtypes

Breast cancer is a hormonal regulated disease. The majority of breast cancers express the receptors for the hormones estrogen and progesterone. Estrogens have been demonstrated to be breast carcinogens and they stimulate (tumor) cell growth and thus play an important role in the development of breast cancer. Furthermore, estrogen exposure is associated with the risk to develop breast cancer[6,7]. The effects of estrogen are mediated by binding to its nuclear receptors. There are two different genes coding for two estrogen receptors ER-α and ER-β. ER-α (ER) is required for normal mammary gland development, and is of important clinical relevance for breast cancer; the role of ER-β is less well defined if relevant for breast cancer biology at all. Progesterone binds to the nuclear progesterone receptors (PR) which has two different isoforms, PRA and PRB. The expression of PR is regulated by estrogens via ER and

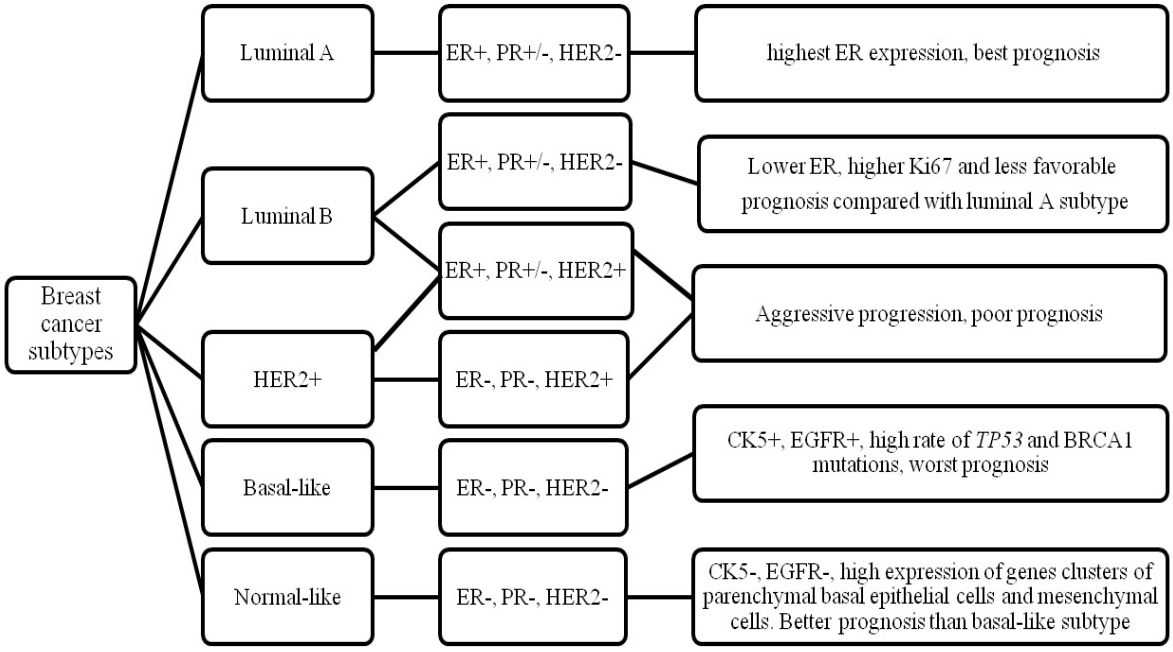
PR expression is considered an indicator of an intact estrogen-ER pathway in breast cancer[8]. Importantly, breast cancers behave differently depending on ER and PR expression.

Since human breast cancer is a heterogeneous disease and breast cancer patients are diverse in their prognosis and responsiveness to treatments, it is important to classify breast cancers in order to provide patients with personalized treatment. Based on gene expression profiling, breast cancers are classified into five molecular/intrinsic subtypes: luminal A, luminal B, HER2+, basal and normal-like breast cancer (Figure 1.1)[9]. ER is the major classifier since tumors of the luminal A and luminal B subtypes are predominantly ER positive, while basal and normal-like subtypes are mostly ER negative. Luminal subtype tumors are characterized by the expression of ER, PR, Bcl-2, CK8/18 and GATA3. Luminal A tumors, however, express higher levels of ER and GATA3 proteins and most ER regulated genes, whereas luminal B subtype tumors express relatively lower levels of ER and its target genes[10]. Moreover, Ki67, a proliferation marker, is important in the distinction between the luminal A and luminal B subtypes[11] with luminal B subtype tumors expressing higher levels of the Ki67 protein. Tumors of the HER2+ subtype are characterized by overexpressing HER2 and can be subdivided into two distinct subtypes based on the expression of ER. Tumors of the basal-like subtype usually lack ER, PR and HER2 expression and express CK5/6 and/or EGFR[12]. Basal-like breast carcinoma has a high expression of the basal gene cluster and a high rate of *TP53* mutations[13]. Tumors of the normal-like subtype generally lack ER, PR, HER2, CK5 and EGFR expression, and have a high expression of genes characteristic of parenchymal basal epithelial cells and mesenchymal cells[14]. Claudin-low tumors represent a subset of the basal- and normal-like subtypes[15]. They are characterized by the low gene expression of claudin-3, -4 and -7, E-cadherin and luminal genes. Moreover, Claudin-low tumors are enriched in epithelial-to-mesenchymal transition (EMT) genes and cancer stem cell-like markers[16]. Basal and normal-like cancers, often lack ER, PR and HER2 expression, the latter being referred to as triple-negative breast cancers (TNBCs), and account for 10 to 20% of all breast cancers. Since TNBC represents a heterogeneous group of breast cancers, gene expression analysis identified six molecular subtypes within TNBCs: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal

stem-like (MSL), and luminal androgen receptor (LAR) subtypes[17].

The different molecular subtypes were shown to be associated with different prognosis[14]. For example, within the luminal subgroup patients with tumors of the luminal A subtype have the best prognosis, whereas luminal B tumors present with a more aggressive phenotype and have a worse prognosis compared with luminal A tumors. Furthermore, the HER2+ and basal-like subtypes are less differentiated and have a poor prognosis[18].

Figure 1.1 Schematic drawing of the molecular/intrinsic breast cancer subtypes.



1.3 Genetics of breast cancer

Breast cancer arises as a consequence of genetic mutations in a subset of genes including somatic and germline mutations.

1.3.1 Sporadic breast cancer

All cancers carry somatic mutations, however, sporadic breast cancer develops only from somatic mutations and accounts for the majority of breast cancers (around 85%)[19]. Moreover,

somatic mutations occur in the genomes of a single cell accumulate over the lifetime of a multicellular organism and may be the cumulative result of DNA damage and repair processes. A large number of somatic mutations have been identified and global sequencing projects have generated catalogs of somatic mutations from these[20]. Around 5 million mutations have been identified from 7,042 tumor samples of 30 different classes[21]. There are two different types of somatic mutations: 'driver' and 'passenger' mutations. The driver mutations are implicated in oncogenesis, provide the growth advantage and have been positively selected during the development of tumors. The passenger mutations do not confer the selective growth advantage and also do not contribute to cancer development[22].

The passenger mutations account for the major part of somatic mutations while the number of driver mutations is limited. Approximately 140 genes that contain driver mutations have been revealed[23,24]. Most tumors contain two to six driver mutations, while the number of somatic driver mutations varies among different cancers. Lung cancers have a relatively high number of driver mutations while ovarian and breast cancers have less driver mutations[25]. Some driver mutations are commonly present among multiple tumor types such as mutations in *TP53*, *PI3KCA*, and *RAS* genes. In addition, there are also tumor type-specific somatic driver mutations identified. For example, mutations in *GATA3* and *MAP3K1* are breast cancer-specific, whereas *KEAP1* mutations are lung cancer-specific[25]. Somatic recurrent driver mutations relatively frequently identified in breast cancers are in genes such as *AKT1*, *BRCA1*, *CDH1*, *CYCLIND1*, *FGFRs*, *GATA3*, *MYC*, *MAP3K1*, *HER2*, *PIK3CA*, *PTEN*, *RB1*, *TBX3* and *TP53*[22,26]. For *HER2*, *GATA3*, *MYC*, *PIK3CA* and *TP53*, the frequency of somatic mutations is larger than 10% across all breast cancers[27]. Some of these genes are found mutated more frequently in specific subtypes of breast cancer. For example, somatic mutations in *GATA3* have been identified in 14% of ER+ breast cancers while the mutations are virtually absent in ER- breast cancers. Furthermore, the *TP53* mutation frequency in luminal subtype cancers (12% in luminal A subtype and 32% in luminal B subtype) is lower compared with the frequency in basal-like (84%) and HER2+ subtype tumors (75%)[27].

Catalogues of somatic driver and passenger mutations have been established and have provided us more insight into the mutational processes that take place during tumorigenesis[28].

Mutational processes evolve across the lifetime of cancers and cause the somatic mutations that arise during the development of tumors. Each process generates classes of mutations and these patterns, designated mutational signatures, can be recognized. The mutational signatures have illustrated several mutational processes in the mutational catalogs of cancer cells[29]. For example, the mutation signatures found in lung and skin tumors match those generated by tobacco exposures and ultraviolet light, respectively. Specifically, C:G>A:T substitutions are common in smoking-associated lung cancer, whereas CC:GG>TT:AA nucleotide substitutions are prevalent in UV light-associated skin cancers[29]. Five distinct mutational signatures have been defined in breast cancer named signature 1b, 2, 3, 8 and 13[21]. Signature 1b is characterized by the prominence of C>T substitutions at NpCpG trinucleotides and shows correlation with age of diagnosis. Signature 2 and 13 are characterized by C>T and C>G substitutions in the context of TpCpN nucleotides that are attributed to the AID/APOBEC family of cytidine deaminases. APOBEC3B was identified as one of the underlying enzymatic sources of these types of mutations in breast cancer[30]. In addition, APOBEC1[31], APOBEC3A[32] and AOPBEC3H[33] were also suggested to be involved in these mutational signatures. Moreover, this mutational process associates with regional somatic hypermutation termed kataegis in breast cancer[21,34,35]. Signature 3 arises as a consequence of BRCA1 or BRCA2 deficiency caused by germline mutations or somatic inactivation of these genes, which are implicated in homologous-recombination-based DNA double-strand break repair. By sequencing a total of 560 whole breast cancer genomes, Nik-Zainal *et al.*[36] have recently revealed a total of 12 base substitution mutational signatures including the five mutational signatures previously observed in breast cancer, but also six rearrangement signatures. Base substitution signature 8, which contributes CC>AA double nucleotide substitutions, as well as rearrangement signature 3 and 5, appear to be also associated with BRCA1 and BRCA2 deficiency[36].

1.3.2 Hereditary breast cancer

A family history of breast cancer is a major risk factor for developing breast cancer. Up to 15%

of breast cancer patients have at least one first-degree relative with breast cancer. This familial clustering of breast cancer suggests a hereditary component of the disease. In addition to somatic mutations, hereditary breast cancers are thus also caused by germline mutations. Familial breast cancer has been associated with mutations in different susceptibility genes, which may be classified as high-, moderate-, and low-risk susceptibility genes (Figure 1.2).

High-risk genes include *BRCA1*[37], *BRCA2*[38], *CDH1*[39], *PTEN*[40], *STK11*[41], and *TP53*[42]. The two high-risk genes *BRCA1* or *BRCA2* were discovered in the 1990s and pathogenic mutations in *BRCA1* or *BRCA2* are known to increase the breast cancer risk by 10- to 20-fold[43]. Mutations in these two high-risk genes are responsible for approximately 25% of the familial breast cancer risk. Other high-risk genes (*CDH1*, *PTEN*, *STK11* and *TP53*) cause 5% of the familial breast cancer risk[44].

Moderate-risk genes include *ATM*, *CHEK2*, *NBS1* and *RAD50* and mutations in these genes confer 2- to 4-fold increased breast cancer risks[45]. *PALB2* was originally also identified as a moderate-risk gene[46], but recent evidence suggests that the breast cancer risk conferred by *PALB2* mutations may be higher. The risk was shown to be eight to nine times as high among the *PALB2* mutation carriers at the age of 40 and six to eight times as high among those at the age of 40 to 60[47,48]. *BRIP1*, which codes for BRCA1-binding helicase was considered as a candidate breast cancer moderate-risk gene[49]. However, truncating variants in *BRIP1* were proven to be not associated with breast cancer risk[50]. Moderate-risk genes explain about 5% of the familial breast cancer risk.

Until now, more than 90 common low-risk alleles have been identified through genome wide association studies (GWAS). These alleles are common in the population and display low effect sizes (up to 1.3)[51]. The more than 90 identified low-risk susceptibility loci explain around 16% of the familial breast cancer risk (Table 1.1) and are scattered throughout the genome. Some variants locate near genes with known functions, while others are in gene-poor regions. Most of these variants map to non-coding regions of the genome and are enriched for FOXA1 and ER binding sites and H3K4me1 epigenome modifications[52]. Although the individual risks of these low-risk loci are not very substantial, their combined

effects are able to increase breast cancer risk significantly[53]. By stratifying the female population using 77 low-risk loci, the top 1% of women would have a 3.36-fold higher breast cancer risk than the population in the middle quintile, while the odds ratio for women in the highest compared with the middle quintile was 1.82[54]. It has also been observed that some low-risk loci significantly modify breast cancer risk in *BRCA1* and *BRCA2* mutation carriers[55]. The addition of modifier loci to risk models will improve risk assessment and influence the approach to genetic counseling, prevention and routine care of breast cancer patients with a *BRCA1* or *BRCA2* mutation[55]. Thus, polygenic risk scores can be used to stratify breast cancer risk and thereby improve the targeting of breast cancer screening programs.

Taken together, all the high-, moderate-, and low-risk susceptibility genes explain around 51% of the familial breast cancer. It is estimated that another 1,000 loci may still remain unidentified and could contribute an additional 14% of familial breast cancer risk[56]. GWAS and genome-wide sequencing projects are needed to continue identifying new breast cancer variants.

Figure 1.2 Familial breast cancer risk genes and alleles.

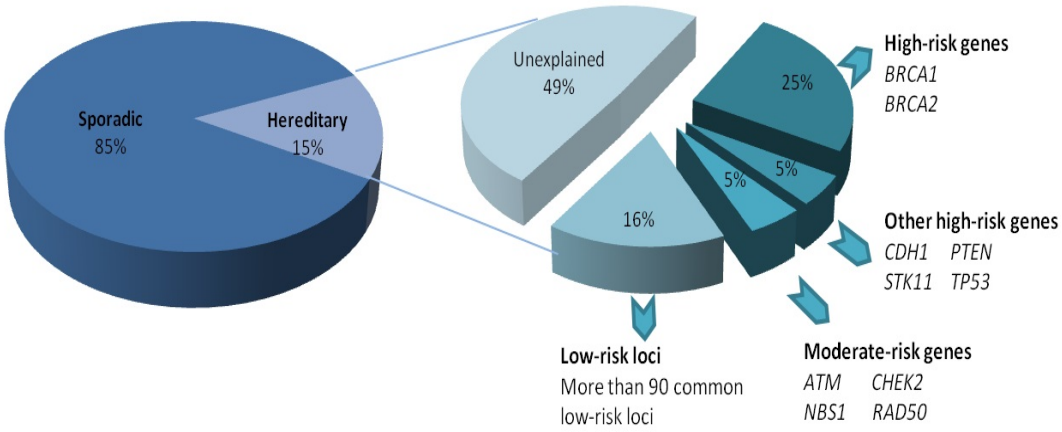


Table 1.1 Over 90 low-risk breast cancer loci identified from genome-wide association studies.

SNP	LOCUS	GENE	OR	MAF	Study
rs2981582	10q26	FGFR2	1.26	0.36	Easton, *et al.*[57], 2007
rs3803662	16q12	*TNRC9/LOC643714*	1.20	0.25	
rs889312	5q11	*MAP3K1*	1.13	0.28	
rs13281615	8q24		1.08	0.40	
rs3817198	11p15	*LSP1*	1.07	0.30	
rs1045485	2q	*CASP8*	0.88	0.13	COX, *et al.*[58], 2007
rs13387042	2q35	*TNP*	1.12	0.49	Stacey, *et al.*[59], 2007
rs10941679	5p12	*MRPS30*	1.19	0.25	Stacey, *et al.*[60], 2008
rs4973768	3p24	*SLC4A7*	1.11	0.47	Ahmed, *et al.*[61], 2009
rs6504950	17q22	*STXBP4* /COX11	0.95	0.28	
rs2046210	6q25	*ER*	1.29	0.36	Zheng, *et al.*[62], 2009
rs11249433	1p11		1.16	0.42	Thomas, *et al.*[63], 2009
rs999737	14q24	*RAD51L1*	0.94		
rs3757318	6q25	*CCDC170*	1.30	0.07	Turnbull, *et al.*[64], 2010
rs1011970	9p21	*CDKN2B*	1.20	0.16	
rs2380205	10p15		0.94	0.44	
rs614367	11q13		1.15	0.15	
rs10995190	10q21	*ZNF365*	0.86	0.14	
rs704010	10q22	*ZMIZ1*	1.07	0.39	
rs8170	19p13	*BABAM1*	1.26	0.19	Antoniou, *et al.*[65], 2010
rs2363956	19p13	*ANKLE1*	0.84	0.51	
rs865686	9q31		0.89	0.37	Fletcher, *et al.*[66], 2011
rs10069690	5p15	*TERT*	1.18	0.26	Haimen, *et al.*[67], 2011
rs10771399	12p11		0.85	0.12	Ghoussaini, *et al.*[68] 2012
rs1292011	12q24		0.92	0.41	
rs2823093	21q21		0.94	0.27	
rs9485372	6q25	*TAB2*	0.90	0.45	Long, *et al.*[69], 2012
rs17530068	6q14		1.12	0.18	Siddiq,, *et al.*[70], 2012
rs2284378	20q11	*RALY*	1.08	0.20	
rs616488	1p36	*PEX14*	0.94	0.33	Michailidou, *et al.*[56], 2013
rs11552449	1p13	*DCLRE1B*	1.08	0.17	
rs4849887	2q14		0.90	0.10	
rs2016394	2q31	*METAP1D-DLX1-DLX2*	0.95	0.48	
s1550623	2q31	*CDCA7*	0.91	0.16	
rs16857609	2q35	*DIRC3*	1.09	0.26	
rs6762644	3p26	*ITPR1-EGOT*	1.06	0.40	
rs12493607	3p24	*TGFBR2*	1.04	0.35	
rs9790517	4q24	*TET2*	1.09	0.23	
rs6828523	4q34	*ADAM29*	0.89	0.13	
rs10472076	5q11	*RAB3C*	1.06	0.38	

rs1353747	5q11	*PDE4D*	0.90	0.10	
rs1432679	5q33	*EBF1*	1.06	0.43	
rs11242675	6p25	*FOXQ1*	0.97	0.39	
rs204247	6p23	*RANBP9*	1.06	0.43	
rs720475	7q35	*ARHGEF5-NOBOX*	0.93	0.25	
rs9693444	8p12		1.07	0.32	
rs6472903	8q21	*CASC9*	0.88	0.18	
rs2943559	8q21	*HNF4G*	1.17	0.07	
rs11780156	8q24	*MIR1208*	1.13	0.16	
rs10759243	9q31		1.07	0.39	
rs7072776	10p12	*MLLT10-DNAJC1*	1.11	0.29	
rs11814448	10p12	*DNAJC1*	1.35	0.02	
rs7904519	10q25	*TCF7L2*	1.06	0.46	
rs11199914	10q26		0.94	0.32	
rs3903072	11q13	*DKFZp761E198*	0.92	0.47	
rs11820646	11q24		0.93	0.41	
rs12422552	12p13		1.11	0.26	
rs17356907	12q22	*NTN4*	0.89	0.30	
rs11571833	13q13	*BRCA2*	1.39	0.01	
rs2236007	14q13	*PAX9*	0.88	0.21	
rs2588809	14q24	*RAD51B*	1.07	0.16	
rs941764	14q32	*CCDC88C*	1.05	0.34	
rs17817449	16q12	*FTO*	0.95	0.40	
rs13329835	16q23	*CDYL2*	1.14	0.22	
rs527616	18q11		0.91	0.38	
rs1436904	18q11	*CHST9*	0.93	0.40	
rs4808801	19p13	*ELL*	0.94	0.35	
rs3760982	19q13	*C19orf61-KCNN4-LYPD5-ZNF283*	1.06	0.46	
rs132390	22q12	*EMID1-RHBDD3-EWSR1*	1.36	0.04	
rs6001930	22q13	*MKL1*	1.17	0.11	
rs554219	11q13	*CCND1*	1.33	0.12	French, *et al.*[71], 2013
rs75915166	11q13	*CCND1*	1.38	0.06	
rs78540526	11q13	*CCND1*	1.42	0.08	
rs4245739	1q32	*MDM4*	1.14	0.26	Garcia-Closas, *et al.*[72], 2013
rs6678914	1q32	*LGR6*	1.10	0.59	
rs11075995	16q12	*FTO*	1.11	0.24	
rs12710696	2p24		1.10	0.36	
rs10069690	5p15		1.06	0.26	Bojesen, *et al.*[73], 2013
rs7726159	5p15	*TERT*	0.05	0.34	
rs1053338	3p21	*ATXN7*	1.07	0.13	Milne, *et al.*[74], 2014
rs6964587	7q21	*AKAP9*	1.05	0.39	
rs2290203	15q26	*PRC1*	1.08	0.50	Cai, *et al.*[75], 2014
rs4951011	1q32	*ZC3H11A*	1.09	0.28	
rs10474352	5q14	*ARRDC3*	1.09	0.48	
rs12405132	1q21	*NBPF10*	0.96	0.36	Michailidou, *et al.*[51], 2015
rs12048493	1q21		1.04	0.34	
rs72755295	1q43	*EXO1*	1.19	0.03	

rs6796502	3p21		0.92	0.09	
rs13162653	5p15		0.92	0.45	
rs2012709	5p13		1.06	0.46	
rs7707921	5q14	*ATG10*	0.94	0.23	
rs9257408	6p22		1.05	0.38	
rs4593472	7q32	*FLJ43663*	0.92	0.35	
rs13365225	8p11		0.89	0.17	
rs13267382	8q23	*LINC00536*	1.07	0.36	
rs11627032	14q32	*RIN3*	0.94	0.26	
chr17:29230520	17q11	*ATAD5*	0.94	0.20	
rs745570	17q25		0.94	0.50	
rs6507583	18q12	*SETBP1*	0.91	0.07	

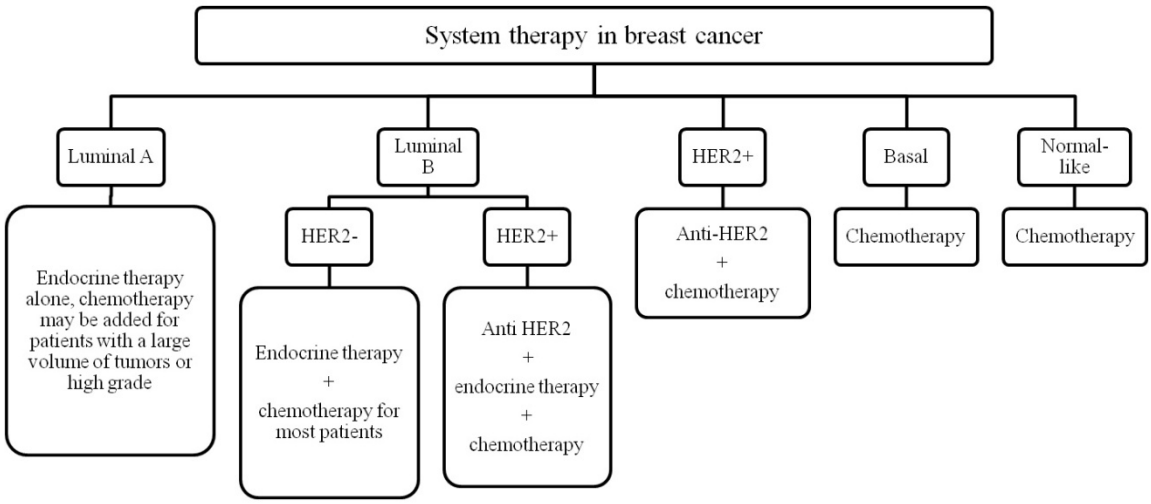
1.4 Clinical outcome of breast cancer

The types of treatment for breast cancer are surgery, radiation therapy, and systemic therapy, which can be administered in the neoadjuvant, adjuvant and recurrent settings. Surgery is usually the first step to treat localized breast cancer, which can be preceded by neoadjuvant systemic therapy to improve operability and reduce the extent of surgery, and followed by adjuvant systemic therapy to aim for cure and reduce the risk of metastases. Hormonal, anti-HER2 targeted therapies and chemotherapy are applied as systemic therapy. The decision on the type of systemic adjuvant therapy given to a patient is based on many factors. According to the 2013[76] and 2015[77] St. Gallen guidelines, this decision should be based on the intrinsic subtypes of breast cancer (Figure 1.3). Since luminal A and luminal B subtypes are ER positive, hormonal therapy should be applied for these subtypes. Whereas tamoxifen can be used for premenopausal patients, aromatase inhibitors should be used for postmenopausal patients[77]. In addition, for patients with luminal A tumors with a large tumor volume and patients with luminal B tumors, chemotherapy could be added to hormonal therapy[78]. The antibody trastuzumab is an HER2-targetted medication that can also be combined with chemotherapy (e.g. taxanes) for patients with HER2+ breast cancer[79-81]. Newer HER2-targeting agents (e.g. lapatinib, pertuzumab, T-DM1, neratinib) have been developed in recent years that may or may not be combined with trastuzumab[82]. For patients with basal-like and normal-like breast cancers, which usually have triple negative phenotypes, chemotherapy (anthracyclines and taxanes) is recommended.

Following adjuvant treatment, around 40% of all patients with breast cancer suffer a recurrence locally, regionally, or at distant metastatic sites. Ten to 20% of all recurrences are locoregional, while others are distant metastases[83]. As is the clinical practice in the adjuvant setting, treatment of patients with metastatic breast cancer is still based on the phenotype of the primary tumor. However, it has been shown that discordance in ER, PR and HER2 status between primary and recurrent breast cancer occurs frequently[84,85]. For therapeutic decision making, ER, PR, and HER2 status should, if possible, therefore be reevaluated in the recurrent tumor(s).

In order to individualize the treatment of breast cancer, prognostic and predictive biomarkers have been studied extensively. Prognostic markers intend to predict aggressiveness of a patient's tumor independent of treatment, while predictive markers evaluate the response of a patient to a specific cancer therapy.

Figure 1.3 The systemic treatment based on the intrinsic subtypes of breast cancer according St. Gallen criteria.



1.4.1 Traditional clinical prognostic and predictive factors

Tumor stage is a powerful predictor of breast cancer prognosis and the TNM system is the most common system used to describe the tumor stage[86]. This system is based on three

clinical characteristics: tumor size (T), axillary lymph node status (N) and metastases (M). Based on this system, breast cancers are classified into five stages (0-IV), with stage 0 indicating an excellent prognosis and stage IV a very poor prognosis.

Tumor grade is a strong prognostic factor and is determined by the Scarff-Bloom-Richardson grading system[87]. Grading is derived from an assessment of three morphological features: tubule formation, nuclear pleomorphism and mitotic count. Each of these three features is scored on a scale of 1-3 and the scores of all three features are added together. Based on the final score, breast cancers are classified into three grades. Low grade tumors are well differentiated and have a better prognosis than higher grade tumors. The risk of relapse at 10 years is around 95% for patients with grade 1 tumors compared with only 40% for grade 3 tumors[88]. Patients with grade 1 tumors of less than 2cm in size even have an excellent prognosis with a 5-year survival of 99%[89,90].

1.4.2 Traditional molecular prognostic and predictive markers

ER expression is one of the most important biomarkers in breast cancer since ER is the target of endocrine therapies. ER-positive breast cancers are associated with a slower tumor growth, lower histological grade and a better overall prognosis. ER currently is the best predictor for response to hormonal treatment. Patients with ER-positive disease have benefit from aromatase inhibitor or tamoxifen treatment.

PR is also an important biomarker in breast cancer. The expression of PR is regulated by ER and high tumor PR levels are associated with a good prognosis in primary breast cancer. Furthermore, high tumor PR levels are associated with a favorable clinical outcome in patients with primary and metastatic disease treated with endocrine therapy[91-95].

HER2 is overexpressed in 15% of all primary breast cancers. HER2 amplification was found to be a strong prognostic maker in primary breast cancer[96,97]. Furthermore, the overexpression of HER2 is a strong marker of favorable prognosis for trastuzumab therapy in primary breast cancer[98,99], but a marker of unfavorable prognosis in patients receiving endocrine therapy[100]. HER2-targeted therapy prolongs overall survival in HER2+ metastatic breast cancer[101,102].

Finally, Ki67 is a nuclear non-histone protein that is associated with cellular proliferation and has been shown to be of prognostic value in primary breast cancer[103,104]. Ki67 has also been shown to be an important predictive marker of unfavorable outcome for anti-hormonal therapy[105] and chemotherapy[106] in primary breast cancer and for unfavorable response to neoadjuvant chemotherapy in locally advanced breast cancer[107].

ER, PR, HER2 and Ki67 are traditional molecular prognostic and predictive markers, but many other biomarkers have been identified, including various gene expression signatures[108].

1.4.2 Prognostic and predictive gene signatures

Various prognostic gene expression signatures for patients with early breast cancer have been reported. A few of the most important signatures are described below. A 70-gene signature (MammaPrint) has been developed on a group of 78 patients with node-negative breast cancer who did not receive adjuvant chemotherapy and were younger than 55 years[109]. The top 70 genes were significantly correlated with clinical outcome and this signature was an independent predictor of metastasis-free survival. In addition, a 76-gene signature was identified in 286 lymph-node-negative patients who had not received adjuvant systemic treatment. This 76-gene profile represented a strong prognostic factor for the development of distant metastasis in untreated patients[110,111] and in patients who received adjuvant tamoxifen therapy[112]. A 21-gene recurrence score signature (Oncotype DX) was defined to predict the risk of distant recurrence in patients with ER-positive breast cancer who received adjuvant tamoxifen therapy[113]. Furthermore, a 97-gene signature has been developed to examine the association of gene expression profiles and histologic grade[114]. It discriminates grade 2 tumors into low and high risk of recurrence subgroups.

Several predictive gene expression signatures have also been developed to predict the patient's type of response to breast cancer treatment. For example, the 21-gene recurrence score was not only able to predict the prognosis in patients who receive adjuvant tamoxifen in primary breast cancer, but was also able to predict chemotherapy benefit in node-negative ER+ patients[113]. Furthermore, a two-gene expression ratio of *HOXB13* and *IL17BR* was identified to have predictive utility[115]. A high *HOXB13*:*IL17BR* ratio was associated with a

high risk of recurrence and predicted poor outcome on tamoxifen therapy in patients with primary breast cancer. The high *HOXB13*:*IL17BR* ratio was also associated with poor progression-free survival in patients with recurrent breast cancer received tamoxifen first-line therapy[116]. In addition, a 44-gene signature has been identified that predicts anti-estrogen therapy outcome in ER-positive recurrent breast cancer patients[117]. The 44-gene signature predicted a significantly longer progression-free survival time and tamoxifen resistance with an accuracy of 80%. Thus, gene expression profiling has been performed to aid clinicians to predict prognosis and guide the use of anti-cancer therapy in breast cancer[118,119].

Moreover, a mutiprotein signature urokinase plasminogen activator (uPA) and its inhibitor, PAI-1 have been shown to be prognostic and predictive biomarkers for breast cancer[120]. High levels of uPA/PAI-1 predicted adverse prognosis in breast cancer patients[121,122] and good response to adjuvant chemotherapy in patients with early breast cancer[123,124].

The continuing identification of relevant predictive and prognostic biomarkers and signatures facilitates the promise for individualized treatment. Gene expression profiling has been instrumental in this and provided novel predictive and prognostic tools. With respect to gaining biological insight, studying all the relevant genes present in large multigene gene signatures is less straight forward. With respect to the 2-gene *HOXB13*:*IL17BR* ratio this should not be problematic and it has become clear that HOXB13 is involved in the modulation of ER signaling pathways[125], which supports its clinical correlation with tamoxifen resistance.

1.5 Summary

Breast cancer is the most common cancer in women worldwide and incidence rates vary widely. Similar to countries with high incidence, rates have been rising in countries with lower rates as well in recent years. Many factors such as age, gender, life style, family history of breast cancer and estrogen exposure are associated with breast cancer. Importantly, breast cancer is a genetic disease that arises from somatic and/or germline mutations. Familial breast cancer has been associated with mutations in three different risk classes of susceptibility

genes/alleles, which account for approximately 51% of familial breast cancer risk. To explain the remaining 49% of familial breast cancer risk, larger GWAS and genome-wide sequencing projects need to be performed to identify new breast cancer genes/variants. Furthermore, human breast cancer is a heterogeneous disease and can be classified into various molecular subtypes. The different subtypes associate with different clinical outcomes. Breast cancer is the pioneer of treatment individualization. ER, PR and HER2 status as well as the intrinsic tumor subtypes have been used for many years to select appropriate systemic therapy. In recent years, additional molecular signatures such as Oncotype DX, MammaPrint and the *HOXB13*:*IL17BR* ratio have been developed to improve the prediction of prognosis and guide the use of anti-cancer therapy in breast cancer.

References

1 Torre, L. A., Bray, F., Siegel, R. L. *et al.* Global cancer statistics, 2012. *CA Cancer J Clin* 65, 87-108 (2015).

2 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2015. *CA Cancer J Clin* 65, 5-29 (2015).

3 Eccles, S. A., Aboagye, E. O., Ali, S. *et al.* Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. *Breast Cancer Res* 15, R92 (2013).

4 Arnold, M., Karim-Kos, H. E., Coebergh, J. W. *et al.* Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. *Eur J Cancer* 51, 1164-1187 (2015).

5 Eifel, P., Axelson, J. A., Costa, J. *et al.* National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 93, 979-989 (2001).

6 Anderson, K. N., Schwab, R. B. & Martinez, M. E. Reproductive risk factors and breast cancer subtypes: a review of the literature. *Breast Cancer Res Treat* 144, 1-10 (2014).

7 Dorgan, J. F., Longcope, C., Stephenson, H. E., Jr. *et al.* Relation of prediagnostic serum estrogen and androgen levels to breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 5, 533-539 (1996).

8 Yip, C. H. & Rhodes, A. Estrogen and progesterone receptors in breast cancer. *Future Oncol* 10, 2293-2301 (2014).

9 Perou, C. M., Sorlie, T., Eisen, M. B. *et al.* Molecular portraits of human breast tumours. *Nature* 406, 747-752 (2000).

10 Laakso, M., Tanner, M., Nilsson, J. *et al.* Basoluminal carcinoma: a new biologically and prognostically distinct entity between basal and luminal breast cancer. *Clin Cancer Res* 12, 4185-4191 (2006).

11 Goldhirsch, A., Wood, W. C., Coates, A. S. *et al.* Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 22, 1736-1747 (2011).

12 Rakha, E. A., El-Sayed, M. E., Green, A. R. *et al.* Breast carcinoma with basal differentiation: a proposal for pathology definition based on basal cytokeratin expression. *Histopathology* 50, 434-438 (2007).

13 O'Brien, K. M., Cole, S. R., Tse, C. K. *et al.* Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* 16, 6100-6110 (2010).

14 Sorlie, T., Perou, C. M., Tibshirani, R. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98, 10869-10874 (2001).

15 Herschkowitz, J. I., Simin, K., Weigman, V. J. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8, R76 (2007).

16 Prat, A., Parker, J. S., Karginova, O. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12, R68 (2010).

17 Lehmann, B. D., Bauer, J. A., Chen, X. *et al.* Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 121, 2750-2767 (2011).

18 Dai, X., Li, T., Bai, Z. *et al.* Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res* 5, 2929-2943 (2015).

19 Zimeri, A. M., Robb, S. W., Hassan, S. M. *et al.* Assessing Heavy Metal and PCB Exposure from Tap Water by Measuring Levels in Plasma from Sporadic Breast Cancer Patients, a Pilot Study. *Int J Environ Res Public Health* 12, 15683-15691 (2015).

20 Hudson, T. J., Anderson, W., Artez, A. *et al.* International network of cancer genome projects. *Nature* 464, 993-998 (2010).

21 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C. *et al.* Signatures of mutational processes in human cancer. *Nature* 500, 415-421 (2013).

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

23 Vogelstein, B., Papadopoulos, N., Velculescu, V. E. *et al.* Cancer genome landscapes. *Science* 339, 1546-1558 (2013).

24 Chen, J. C., Alvarez, M. J., Talos, F. *et al.* Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. *Cell* 166, 1055 (2016).

25 Kandoth, C., McLellan, M. D., Vandin, F. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333-339 (2013).

26 Stephens, P. J., Tarpey, P. S., Davies, H. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486, 400-404 (2012).

27 Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70 (2012).

28 Nik-Zainal, S., Van Loo, P., Wedge, D. C. *et al.* The life history of 21 breast cancers. *Cell* 149, 994-1007 (2012).

29 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C. *et al.* Deciphering signatures of

mutational processes operative in human cancer. *Cell Rep* 3, 246-259 (2013).

30 Burns, M. B., Lackey, L., Carpenter, M. A. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366-370 (2013).

31 Saraconi, G., Severi, F., Sala, C. *et al.* The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol* 15, 417 (2014).

32 Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B. *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 46, 487-491 (2014).

33 Starrett, G. J., Luengas, E. M., McCann, J. L. *et al.* The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* 7, 12918 (2016).

34 Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979-993 (2012).

35 Taylor, B. J., Nik-Zainal, S., Wu, Y. L. *et al.* DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife* 2, e00534 (2013).

36 Nik-Zainal, S., Davies, H., Staaf, J. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47-54 (2016).

37 Miki, Y., Swensen, J., Shattuck-Eidens, D. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66-71 (1994).

38 Wooster, R., Bignell, G., Lancaster, J. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789-792 (1995).

39 Kaurah, P., MacMillan, A., Boyd, N. *et al.* Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA* 297, 2360-2372 (2007).

40 Lynch, E. D., Ostermeyer, E. A., Lee, M. K. *et al.* Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* 61, 1254-1260 (1997).

41 Leggett, B. A., Young, J. P. & Barker, M. Peutz-Jeghers syndrome: genetic screening.

Expert Rev Anticancer Ther 3, 518-524 (2003).

42 Birch, J. M., Alston, R. D., McNally, R. J. *et al.* Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* 20, 4621-4628 (2001).

43 Stratton, M. R. & Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat Genet* 40, 17-22 (2008).

44 Melchor, L. & Benitez, J. The complex genetic landscape of familial breast cancer. *Hum Genet* 132, 845-863 (2013).

45 Hollestelle, A., Wasielewski, M., Martens, J. W. *et al.* Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* 20, 268-276 (2010).

46 Rahman, N., Seal, S., Thompson, D. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 39, 165-167 (2007).

47 Cybulski, C., Kluzniak, W., Huzarski, T. *et al.* Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. *Lancet Oncol* 16, 638-644 (2015).

48 Antoniou, A. C., Casadei, S., Heikkinen, T. *et al.* Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* 371, 497-506 (2014).

49 Seal, S., Thompson, D., Renwick, A. *et al.* Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet* 38, 1239-1241 (2006).

50 Easton, D. F., Lesueur, F., Decker, B. *et al.* No evidence that protein truncating variants in BRIP1 are associated with breast cancer risk: implications for gene panel testing. *J Med Genet* 53, 298-309 (2016).

51 Michailidou, K., Beesley, J., Lindstrom, S. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47, 373-380 (2015).

52 Cowper-Sal lari, R., Zhang, X., Wright, J. B. *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44, 1191-1198 (2012).

53 Pashayan, N., Duffy, S. W., Chowdhury, S. *et al.* Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer* 104, 1656-1663 (2011).

54 Mavaddat, N., Pharoah, P. D., Michailidou, K. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107 (2015).

55 Wang, X., Pankratz, V. S., Fredericksen, Z. *et al.* Common variants associated with breast cancer in genome-wide association studies are modifiers of breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* 19, 2886-2897 (2010).

56 Michailidou, K., Hall, P., Gonzalez-Neira, A. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45, 353-361, 361e351-352 (2013).

57 Easton, D. F., Pooley, K. A., Dunning, A. M. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087-1093 (2007).

58 Cox, A., Dunning, A. M., Garcia-Closas, M. *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 39, 352-358 (2007).

59 Stacey, S. N., Manolescu, A., Sulem, P. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39, 865-869 (2007).

60 Stacey, S. N., Manolescu, A., Sulem, P. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 40, 703-706 (2008).

61 Ahmed, S., Thomas, G., Ghoussaini, M. *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 41, 585-590 (2009).

62 Zheng, W., Long, J., Gao, Y. T. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 41, 324-328 (2009).

63 Thomas, G., Jacobs, K. B., Kraft, P. *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 41, 579-584 (2009).

64 Turnbull, C., Ahmed, S., Morrison, J. *et al.* Genome-wide association study identifies

five new breast cancer susceptibility loci. *Nat Genet* 42, 504-507 (2010).

65 Antoniou, A. C., Wang, X., Fredericksen, Z. S. *et al.* A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* 42, 885-892 (2010).

66 Fletcher, O., Johnson, N., Orr, N. *et al.* Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* 103, 425-435 (2011).

67 Haiman, C. A., Chen, G. K., Vachon, C. M. *et al.* A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet* 43, 1210-1214 (2011).

68 Ghoussaini, M., Fletcher, O., Michailidou, K. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* 44, 312-318 (2012).

69 Long, J., Cai, Q., Sung, H. *et al.* Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet* 8, e1002532 (2012).

70 Siddiq, A., Couch, F. J., Chen, G. K. *et al.* A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* 21, 5373-5384 (2012).

71 French, J. D., Ghoussaini, M., Edwards, S. L. *et al.* Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. *Am J Hum Genet* 92, 489-503 (2013).

72 Garcia-Closas, M., Couch, F. J., Lindstrom, S. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* 45, 392-398, 398e391-392 (2013).

73 Bojesen, S. E., Pooley, K. A., Johnatty, S. E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* 45, 371-384, 384e371-372 (2013).

74 Milne, R. L., Burwinkel, B., Michailidou, K. *et al.* Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer

Association Consortium. *Hum Mol Genet* 23, 6096-6111 (2014).

75 Cai, Q., Zhang, B., Sung, H. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* 46, 886-890 (2014).

76 Goldhirsch, A., Winer, E. P., Coates, A. S. *et al.* Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 24, 2206-2223 (2013).

77 Gnant, M., Thomssen, C. & Harbeck, N. St. Gallen/Vienna 2015: A Brief Summary of the Consensus Discussion. *Breast Care (Basel)* 10, 124-130 (2015).

78 Senkus, E., Kyriakides, S., Penault-Llorca, F. *et al.* Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 24 Suppl 6, vi7-23 (2013).

79 Callahan, R. & Hurvitz, S. Human epidermal growth factor receptor-2-positive breast cancer: Current management of early, advanced, and recurrent disease. *Curr Opin Obstet Gynecol* 23, 37-43 (2011).

80 Tsang, R. Y. & Finn, R. S. Beyond trastuzumab: novel therapeutic strategies in HER2-positive metastatic breast cancer. *Br J Cancer* 106, 6-13 (2012).

81 Swain, S. M., Im, Y. H., Im, S. A. *et al.* Safety profile of Pertuzumab with Trastuzumab and Docetaxel in patients from Asia with human epidermal growth factor receptor 2-positive metastatic breast cancer: results from the phase III trial CLEOPATRA. *Oncologist* 19, 693-701 (2014).

82 Mates, M., Fletcher, G. G., Freedman, O. C. *et al.* Systemic targeted therapy for her2-positive early female breast cancer: a systematic review of the evidence for the 2014 Cancer Care Ontario systemic therapy guideline. *Curr Oncol* 22, S114-122 (2015).

83 Gerber, B., Freund, M. & Reimer, T. Recurrent breast cancer: treatment strategies for maintaining and prolonging good quality of life. *Dtsch Arztebl Int* 107, 85-91 (2010).

84 Rossi, S., Basso, M., Strippoli, A. *et al.* Hormone Receptor Status and HER2

Expression in Primary Breast Cancer Compared With Synchronous Axillary Metastases or Recurrent Metastatic Disease. *Clin Breast Cancer* 15, 307-312 (2015).

85 Amir, E., Miller, N., Geddie, W. *et al.* Prospective study evaluating the impact of tissue confirmation of metastatic disease in patients with breast cancer. *J Clin Oncol* 30, 587-592 (2012).

86 Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 17, 1471-1474 (2010).

87 Elston, C. W. & Ellis, I. O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19, 403-410 (1991).

88 Rakha, E. A., Reis-Filho, J. S., Baehner, F. *et al.* Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 12, 207 (2010).

89 Henson, D. E., Ries, L., Freedman, L. S. *et al.* Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index. *Cancer* 68, 2142-2149 (1991).

90 Rakha, E. A., El-Sayed, M. E., Lee, A. H. *et al.* Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol* 26, 3153-3158 (2008).

91 Dowsett, M., Houghton, J., Iden, C. *et al.* Benefit from adjuvant tamoxifen therapy in primary breast cancer patients according oestrogen receptor, progesterone receptor, EGF receptor and HER2 status. *Ann Oncol* 17, 818-826 (2006).

92 Liu, S., Chia, S. K., Mehl, E. *et al.* Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. *Breast Cancer Res Treat* 119, 53-61 (2010).

93 Purdie, C. A., Quinlan, P., Jordan, L. B. *et al.* Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *Br J Cancer* 110, 565-572 (2014).

94 Bardou, V. J., Arpino, G., Elledge, R. M. *et al.* Progesterone receptor status

significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J Clin Oncol* 21, 1973-1979 (2003).

95 Stendahl, M., Ryden, L., Nordenskjold, B. *et al.* High progesterone receptor expression correlates to the effect of adjuvant tamoxifen in premenopausal breast cancer patients. *Clin Cancer Res* 12, 4614-4618 (2006).

96 Slamon, D. J., Clark, G. M., Wong, S. G. *et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235, 177-182 (1987).

97 Borg, A., Tandon, A. K., Sigurdsson, H. *et al.* HER-2/neu amplification predicts poor survival in node-positive breast cancer. *Cancer Res* 50, 4332-4337 (1990).

98 Piccart-Gebhart, M. J., Procter, M., Leyland-Jones, B. *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353, 1659-1672 (2005).

99 Dawood, S., Broglio, K., Buzdar, A. U. *et al.* Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. *J Clin Oncol* 28, 92-98 (2010).

100 Dowsett, M., Allred, C., Knox, J. *et al.* Relationship between quantitative estrogen and progesterone receptor expression and human epidermal growth factor receptor 2 (HER-2) status with recurrence in the Arimidex, Tamoxifen, Alone or in Combination trial. *J Clin Oncol* 26, 1059-1065 (2008).

101 Balduzzi, S., Mantarro, S., Guarneri, V. *et al.* Trastuzumab-containing regimens for metastatic breast cancer. *Cochrane Database Syst Rev*, CD006242 (2014).

102 Swain, S. M., Baselga, J., Kim, S. B. *et al.* Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med* 372, 724-734 (2015).

103 Urruticoechea, A., Smith, I. E. & Dowsett, M. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol* 23, 7212-7220 (2005).

104 de Azambuja, E., Cardoso, F., de Castro, G., Jr. *et al.* Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br*

J Cancer 96, 1504-1513 (2007).

105 Dowsett, M., Smith, I. E., Ebbs, S. R. *et al.* Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J Natl Cancer Inst* 99, 167-170 (2007).

106 Faneyte, I. F., Schrama, J. G., Peterse, J. L. *et al.* Breast cancer response to neoadjuvant chemotherapy: predictive markers and relation with outcome. *Br J Cancer* 88, 406-412 (2003).

107 Luo, J., Zhou, Z., Yang, Z. *et al.* The Value of 18F-FDG PET/CT Imaging Combined With Pretherapeutic Ki67 for Early Prediction of Pathologic Response After Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer. *Medicine (Baltimore)* 95, e2914 (2016).

108 Patani, N., Martin, L. A. & Dowsett, M. Biomarkers for the clinical management of breast cancer: international perspective. *Int J Cancer* 133, 1-13 (2013).

109 van 't Veer, L. J., Dai, H., van de Vijver, M. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536 (2002).

110 Wang, Y., Klijn, J. G., Zhang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679 (2005).

111 Foekens, J. A., Atkins, D., Zhang, Y. *et al.* Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. *J Clin Oncol* 24, 1665-1671 (2006).

112 Zhang, Y., Sieuwerts, A. M., McGreevy, M. *et al.* The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* 116, 303-309 (2009).

113 Paik, S., Shak, S., Tang, G. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351, 2817-2826 (2004).

114 Sotiriou, C., Wirapati, P., Loi, S. *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98, 262-272 (2006).

115 Ma, X. J., Wang, Z., Ryan, P. D. *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5, 607-616 (2004).

116 Jansen, M. P., Sieuwerts, A. M., Look, M. P. *et al.* HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol* 25, 662-668 (2007).

117 Jansen, M. P., Foekens, J. A., van Staveren, I. L. *et al.* Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 23, 732-740 (2005).

118 Gene expression profiling for guiding adjuvant chemotherapy decisions in women with early breast cancer: an evidence-based and economic analysis. *Ont Health Technol Assess Ser* 10, 1-57 (2010).

119 MacDonald, K. V., Bombard, Y., Deal, K. *et al.* The influence of gene expression profiling on decisional conflict in decision making for early-stage breast cancer chemotherapy. *Eur J Cancer* 61, 85-93 (2016).

120 Duffy, M. J., Harbeck, N., Nap, M. *et al.* Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer* 75, 284-298 (2017).

121 Jelisavac-Cosic, S., Sirotkovic-Skerlev, M., Kulic, A. *et al.* Prognostic significance of urokinase-type plasminogen activator (uPA) and plasminogen activator inhibitor (PAI-1) in patients with primary invasive ductal breast carcinoma - a 7.5-year follow-up study. *Tumori* 97, 532-539 (2011).

122 Duffy, M. J., McGowan, P. M., Harbeck, N. *et al.* uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Res* 16, 428 (2014).

123 Borstnar, S., Sadikov, A., Mozina, B. *et al.* High levels of uPA and PAI-1 predict a good response to anthracyclines. *Breast Cancer Res Treat* 121, 615-624 (2010).

124 Manders, P., Tjan-Heijnen, V. C., Span, P. N. *et al.* Predictive impact of urokinase-type plasminogen activator: plasminogen activator inhibitor type-1

complex on the efficacy of adjuvant systemic therapy in primary breast cancer. *Cancer Res* 64, 659-664 (2004).

125 Shah, N., Jin, K., Cruz, L. A. *et al.* HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ERalpha and inducing IL-6 expression. *Cancer Res* 73, 5449-5458 (2013).

Chapter 2

Aims and outline

Aims and outline of this thesis

Breast cancer is one of the most common malignancies and has a serious impact on female health. Fortunately, great progress has been achieved in the survival of breast cancer patients in recent years mostly due to earlier detection and more effective treatment regimens. To facilitate early detection of breast cancer, we need to be able to identify those individuals at significant risk for developing breast cancer at an early stage. Until now, several familial breast cancer susceptibility genes/alleles have been identified, which account for 51% of familial breast cancer risk. Novel breast cancer susceptibility genes/variants need to be identified to explain the remaining 49% of familial breast cancer risk. In addition, to further improve current anti-cancer treatment that patients undergo, we need knowledge to give the right drug to the right patient to achieve individualized or personalized treatment. To further improve the survival of breast cancer patients, it is thus important to identify novel genetic variants/factors associated with breast cancer risk, therapy response and aggressiveness that can be used to guide treatment decision making.

The aims of this thesis are focused on two main topics:

1. The discovery of novel genetic variants that predict an increased risk to develop breast cancer (**Chapter 3** and **4**).
2. The discovery of novel genetic markers that predict the prognosis of breast cancer patients and their response to therapy (**Chapter 5** and **6**).

In **Chapter 3**, we evaluated whether the p.G84E mutation or other mutations in the *HOXB13* gene are associated with an increased breast cancer risk. The *HOXB13* p.G84E mutation was reported to be a prostate cancer susceptibility allele[1]. HOXB13 plays an important role in breast tumor progression. Moreover, a high ratio of *HOXB13*:*IL17BR* expression is of prognostic and predictive value for ER-positive breast cancer patients[2,3]. Based on these observations, we hypothesized that *HOXB13* might also be a breast cancer susceptibility gene. So far, three studies had investigated the association between p.G84E and breast cancer risk, but obtained contradictory results[4-6]. We analyzed the entire coding region of the *HOXB13*

gene in 1,250 Dutch familial breast cancer cases and 800 geographically matched controls. Then two recurrent mutations were genotyped in 4,520 non-*BRCA1/2* cases and 3,127 controls to evaluate the association between the mutations and breast cancer risk.

Mutations in the *NBS1* gene have been reported to be associated with risks for several cancer types including breast cancer[7]. The functional variant rs2735383 locates to the 3'UTR of *NBS1*. The rs2735383CC variant could influence the binding ability of several microRNAs, thereby decreasing the expression level of the *NBS1* gene. Rs2735383 had been associated with an increased risk of lung cancer and colorectal cancer[8,9]. In this respect, the aim of the **Chapter 4** was to investigate whether *NBS1* rs2735383 was also associated with breast cancer risk. We analyzed *NBS1* rs2735383 in the Rotterdam Breast Cancer Study (RBCS) and in 45 studies of the Breast Cancer Association Consortium (BCAC) through imputation. Then we evaluated the association of *NBS1* rs2735383 with breast cancer risk in the overall and subgroup analyses.

APOBEC3B mRNA overexpression was associated with a hypermutator phenotype and poor outcomes in ER-positive breast cancer patients[10]. The 29.5 kb *APOBEC3B* deletion polymorphism results in a hypermutator phenotype[11] and has been associated with breast cancer risk. In **Chapter 5**, we investigated whether the 29.5 kb *APOBEC3B* deletion polymorphism associates with clinical outcome of breast cancer. Copy number analysis was performed by quantitative PCR in primary tumors of 1,756 Dutch breast cancer patients. The prognostic and predictive value of the *APOBEC3B* 29.5 kb deletion polymorphism was analyzed in four different clinical cohorts. As increased *APOBEC3B* mRNA levels had prognostic value and were associated with a poor prognosis in ER-positive breast cancer patients[10], we further evaluated the relation between *APOBEC3B* copy number and *APOBEC3B* mRNA expression.

Mutations in *GATA3* are frequently identified (*i.e.* larger than 10% of cases) among breast cancers[12]. *GATA3* mutations have been reported to be associated with ER-positive breast cancer[13] and the response to neoadjuvant aromatase inhibitor treatment[14]. Therefore, we aimed in **Chapter 6** to investigate whether *GATA3* mutations predicted the outcome of tamoxifen treatment in the advanced setting. We analyzed mutations in exons 5 and 6 of the

GATA3 gene in 235 ER-positive primary breast cancers and evaluated the association of the mutations with clinical outcome of patients who received tamoxifen as first-line therapy for recurrent disease. We also performed *GATA3* mRNA expression analysis in 148 out of the 235 tumors to evaluate the association of *GATA3* mRNA expression with clinical outcome of first-line tamoxifen therapy.

Finally, the results of **Chapter 3-6** are discussed in **Chapter 7** and are summarized in **Chapter 8**.

Reference

1 Ewing, C. M., Ray, A. M., Lange, E. M. *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 366, 141-149 (2012).

2 Ma, X. J., Wang, Z., Ryan, P. D. *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5, 607-616 (2004).

3 Jansen, M. P., Sieuwerts, A. M., Look, M. P. *et al.* HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol* 25, 662-668 (2007).

4 Alanee, S., Couch, F. & Offit, K. Association of a HOXB13 variant with breast cancer. *N Engl J Med* 367, 480-481 (2012).

5 Akbari, M. R., Kluzniak, W., Rodin, R. *et al.* The HOXB13 p.Gly84Glu mutation is not associated with the risk of breast cancer. *Breast Cancer Res Treat* 136, 907-909 (2012).

6 Laitinen, V. H., Wahlfors, T., Saaristo, L. *et al.* HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 22, 452-460 (2013).

7 di Masi, A. & Antoccia, A. NBS1 Heterozygosity and Cancer Risk. *Curr Genomics* 9, 275-281 (2008).

8 Gao, P., Ma, N., Li, M. *et al.* Functional variants in NBS1 and cancer risk: evidence from a meta-analysis of 60 publications with 111 individual studies. *Mutagenesis* 28,

683-697 (2013).

9 Yang, L., Li, Y., Cheng, M. *et al.* A functional polymorphism at microRNA-629-binding site in the 3'-untranslated region of NBS1 gene confers an increased risk of lung cancer in Southern and Eastern Chinese population. *Carcinogenesis* 33, 338-347 (2012).

10 Sieuwerts, A. M., Willis, S., Burns, M. B. *et al.* Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm Cancer* 5, 405-413 (2014).

11 Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B. *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 46, 487-491 (2014).

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

13 Jiang, Y. Z., Yu, K. D., Zuo, W. J. *et al.* GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. *Cancer* 120, 1329-1337 (2014).

14 Ellis, M. J., Ding, L., Shen, D. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486, 353-360 (2012).

Chapter 3

Recurrent *HOXB13* mutations in the Dutch population do not associate with increased breast cancer risk

Jingjing Liu, Wendy J.C. Prager-van der Smissen, Marjanka K. Schmidt, J. Margriet Collée, Sten Cornelissen, Roy Lamping, Anja Nieuwlaat, John A. Foekens, Maartje J. Hooning, Senno Verhoef, Ans M.W. van den Ouweland, Frans B.L. Hogervorst, John W.M. Martens, Antoinette Hollestelle.

Sci Rep 6, 30026 (2016).

Abstract

The *HOXB13* p.G84E mutation has been firmly established as a prostate cancer susceptibility allele. Although *HOXB13* also plays a role in breast tumor progression, the association of *HOXB13* p.G84E with breast cancer risk is less evident. Therefore, we comprehensively interrogated the entire *HOXB13* coding sequence for mutations in 1,250 non-*BRCA1/2* familial breast cancer cases and 800 controls. We identified two predicted deleterious missense mutations, p.G84E and p.R217C, that were recurrent among breast cancer cases and further evaluated their association with breast cancer risk in a larger study. Taken together, 4,520 familial non-*BRCA1/2* breast cancer cases and 3,127 controls were genotyped including the cases and controls of the whole gene screen. The concordance rate for the genotyping assays compared with Sanger sequencing was 100%. The prostate cancer risk allele p.G84E was identified in 18 (0.56%) of 3,187 cases and 16 (0.70%) of 2,300 controls (OR=0.81, 95% CI=0.41-1.59, *P*=0.54). Additionally, p.R217C was identified in 10 (0.31%) of 3,208 cases and 2 (0.087%) of 2,288 controls (OR=3.57, 95% CI=0.76-33.57, *P*=0.14). These results imply that none of the recurrent *HOXB13* mutations in the Dutch population are associated with breast cancer risk, although it may be worthwhile to evaluate p.R217C in a larger study.

Introduction

Breast cancer is the second leading cause of cancer death in Western countries and the most frequently diagnosed cancer in Western women. A family history of breast cancer is a major risk factor for developing breast cancer. Approximately 10-15% of breast cancer patients have at least one first-degree relative with breast cancer. Depending on the number of affected first-degree relatives, this implies risk ratios for breast cancer of 1.80 for one affected relative to 3.90 for three or more affected relatives[1].

Familial breast cancer has been associated with mutations in several high- and moderate-risk breast cancer susceptibility genes, as well as an increasing number of low-risk breast cancer susceptibility alleles. The two high-risk genes *BRCA1* and *BRCA2* were identified in the 1990s and germline mutations in these genes confer average cumulative lifetime breast cancer risks by age 70 of 65% and 45%, respectively[2-4]. Mutations in moderate-risk genes *ATM*, *CHEK2*, *PALB2,* and *RAD50* confer 2- to 4-fold increased breast cancer risks[5], although recent evidence suggests that the breast cancer risk conferred by *PALB2* mutations may be higher than initially thought[6,7]. The more than 90 identified common low-risk alleles, on the other hand, display small effect sizes (*i.e.* per allele odds ratios) of up to 1.3[8]. However, taken together in a polygenic risk score (PRS; calculated from 77 SNPs) the lifetime risk of breast cancer for women in the highest quantile of this PRS was 17%[9]. In total, these breast cancer susceptibility genes and alleles account for approximately 35% of the familial breast cancer risk, which means that the underlying cause of the majority of the familial breast cancer risk thus still remains unexplained.

In this respect, the rare variant c.251G>A (p.G84E; rs138213197) in the *HOXB13* gene was reported to be associated with prostate cancer[10]. Meta-analyses have estimated the increased prostate cancer risk from this mutation to be 4- to 5-fold and even higher among early onset prostate cancer patients and prostate cancer patients with a family history of prostate cancer[11-14]. Moreover, fine-scale mapping at the *HOXB* gene cluster at 17q21-22 had identified a number of highly correlated common SNPs that were associated with prostate cancer risk and tagging the rare *HOXB13* p.G84E variant. This not only further established the association between *HOXB13* p.G84E and prostate cancer risk, but also provided

evidence that GWAS associations could actually be driven by rare variants[15]. Interestingly, the *HOXB13* gene encodes a transcription factor that plays an important regulatory role during embryonic development, but also in tumorigenesis. For example, *HOXB13* was reported to regulate the transcription of androgen receptor (AR) target genes[16] and together with *HOXA9*, *HOXB13* is the most commonly deregulated gene in solid cancers[17]. Moreover, *HOXB13* was shown to preferentially bind a low-risk prostate cancer susceptibility allele located in an AR and FOXA1 binding site (*i.e.* rs339331), thereby enhancing *RFX6* expression and promoting metastasis[18].

In breast cancer, *HOXB13* gene expression is regulated by estrogen in an ER dependent manner[19]. Furthermore, a high *HOXB13*:*IL17BR* expression ratio was found to be a prognostic and predictive biomarker for ER-positive breast cancer patients[20,21]. The poor response to tamoxifen therapy that is predicted from high HOXB13 expression has been shown to be mediated by HOXB13 through the direct suppression of ER, the induction of IL6 expression and mTOR pathway activation[22]. Considering these observations, *HOXB13* might also be a likely candidate for being a breast cancer susceptibility gene. So far, three studies have investigated this hypothesis but obtained contradictory results. In the study by Alanee *et al.*[23], the *HOXB13* p.G84E mutation was shown to confer an increased breast cancer risk, however, Akbari *et al.* could not replicate this association in a larger study[24]. Laitinen *et al.* also found no association with breast cancer risk, but did observe a suggestive association in a particular high-risk subgroup[25]. Importantly, all three studies only investigated the prostate cancer risk-associated variant p.G84E.

In this study, we therefore analyzed the entire coding region of the *HOXB13* gene in 1,250 Dutch familial breast cancer cases and 800 geographically matched controls to establish whether the p.G84E mutation or other mutations in the *HOXB13* gene are associated with an increased breast cancer risk.

Methods

Study population

The samples included in this study were from two Dutch breast cancer case-control studies: RBCS and ABCS-F. RBCS cases (N=2,751) were selected from the database of the Clinical Genetics Centre at Erasmus University Medical Centre in Rotterdam, representing the Southwestern part of the Netherlands. Selected families included all families counselled between 1994 and 2014 that presented with at least two cases of female breast cancer or at least one case of female breast cancer and one case of ovarian cancer in first- or second-degree relatives. At least one of these two cases needed to be diagnosed before the age of 60. For each family, the youngest breast cancer patient who had been tested for *BRCA1* and *BRCA2* was assigned to be the index case and included in RBCS. Additionally, breast cancer cases were included that were diagnosed either before 40 years with unilateral breast cancer or before 50 years with bilateral breast cancer without having a first or second degree relative diagnosed with either breast or ovarian cancer. All cases and their tested relatives were negative for both *BRCA1* and *BRCA2* mutations. Median age of the RBCS cases was 44 years (range 18-92 years). The RBCS control population (N=1,159) was geographically matched and included women from cystic fibrosis families who were either spouses of individuals at risk of being carrier of a *CFTR* mutation or individuals who were tested negative for a *CFTR* mutation and were counselled between 1996 and 2010. Median age of the RBCS controls was 41 years (range 10-97 years).

ABCS-F cases (N=1,769) were selected from the linked databases of the Division of Diagnostic Oncology and the Tumor Registry of the Antoni van Leeuwenhoek hospital in Amsterdam[26]. We included female breast cancer patients of all ages (median age was 42 years (range 14-79 years)), without a pathogenic *BRCA1/2* mutation or unclassified variant, who were counselled in the Family Cancer Clinic and diagnosed and/or treated with cancer in the Antoni van Leeuwenhoek hospital in the period 1995-2012. For each family, only the youngest breast cancer patient who had been tested for *BRCA1* and *BRCA2* was included. ABCS-F controls (N=1,968) are healthy women of all ages (median age was 49 years (range

18-69 years)) from the general population and were recruited through the blood bank.

All experiments were performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.fmwv.nl). The RBCS and ABCS-F studies were approved by the Medical Ethical Committes of the Erasmus Medical Center Rotterdam and the Netherlands Cancer Institute, respectively. All individuals gave written informed consent.

PCR and Sanger sequencing

The entire *HOXB13* (RefSeq NM_006361.5) coding region was analyzed for sequence variations in 1,250 non-*BRCA1/2* familial breast cancer cases and 800 controls from RBCS (*i.e.* indexes from families counselled between 1995 and 2009 for cases and between 1996 and 2006 for controls) using PCR and Sanger sequencing. Twenty nanograms of DNA, extracted from peripheral blood, was PCR amplified in a final volume of 15µl containing 1X GoTaq buffer (Promega, Madison, WI), 1.5mM MgCl2, 200µM dNTPs (GE Healthcare, Waukesha, WI), 1µM of each primer and 0.75U of GoTaq polymerase (Promega) using an ABI2720 thermal cycler (Thermo Scientific, Waltham, MA). First, the PCR reaction was incubated for 5 minutes at 94ºC, followed by 35 cycles of 94ºC for 30 seconds, 58ºC for 1 minute and 72ºC for 1 minute. The PCR reaction ended with a final extension at 72ºC for 5 minutes. Removal of dNTPs and primers before sequencing was done by ExoSAP-IT PCR Product Cleanup (Affymetrix, Santa Clara, CA). Briefly, 2.5µl of PCR product was incubated with 0.5µl of ExoSAP-IT and 1x GoTaq buffer in a final volume of 12.5µl. Incubation took place in an ABI2720 thermal cycler for 15 minutes at 37ºC. Then enzymes were inactivated at 80ºC for 15 minutes before proceeding with Sanger sequencing. The sequencing reaction contained 2µl of ExoSAP-it treated PCR product, 1µl BigDye Terminator v3.1 reaction mix (Thermo Scientific), 1X BigDye Terminator sequencing buffer (Thermo Scientific) and 0.16µM of sequencing primer in a final volume of 10µl and was carried out in an ABI2720 thermal cycler according to the following protocol: 1 cycle of 96ºC for 2 minutes and 25 cycles of 96ºC for 30 seconds, 58ºC for 30 seconds and 72ºC for 2 minutes. Subsequently, the sequencing product was precipitated with absolute ethanol and 3M of NaAc, resuspended in

20µl of Hi-Di formamide (Thermo Scientific), and ran on an ABI3130XL Genetic Analyzer (Thermo Scientific). Sequencing electropherograms were analyzed using Mutation Surveyor v3.20 software (Softgenetics, State College, PA). Sanger sequencing was successful for 96.2% of the samples and PCR and sequencing primer sequences for the two exons of the *HOXB13* gene are available in Supplementary Table 3.1.

Taqman genotyping

Genotyping of the c.251G>A (p.G84E; rs138213197) and c.649C>T (p.R217C; rs139475791) mutations in the *HOXB13* gene was performed for all 7,647 DNA samples from RBCS and ABCS-F using custom-made Taqman genotyping assays (Thermo Scientific) on a Mx3000/ 3005P qPCR machine (Agilent Technologies, Santa Clara, CA). For genotyping p.G84E, 0.5X of Taqman genotyping assay and 0.5X of Taqman Genotyping Master Mix (Thermo Scientific) was added to 20ng of genomic DNA in a final volume of 10µl, whereas for p.R217C, 1X Taqman genotyping assay and 1X ABsolute qPCR Mix, low ROX (Thermo Scientific) was added to 20ng of genomic DNA in a final volume of 10µl. Cycling conditions were: 1 cycle of 10 minutes (for Taqman Genotyping Master Mix) or 15 minutes (for ABsolute qPCR Mix, low ROX) at 95ºC and 45 cycles of 15 seconds at 92ºC and 1 minute at 60ºC. The MxPro qPCR software v4.10 (Agilent) was used to visualize the genotyping results. The call rate of the genotyping assays was 97.9% for p.G84E and 98.1% for p.R217C, respectively, and Taqman assay design is specified in Supplementary Table 3.2. The accuracy of both genotyping assays was evaluated by comparing genotypes obtained from the 1,250 RBCS cases and 800 RBCS controls through Taqman genotyping with genotypes obtained from Sanger sequencing. For quality control, each 96-well plate included a wild-type and a heterozygous sample. Samples that were identified to be positive by either Sanger sequencing or in Taqman assays were independently confirmed by Sanger sequencing.

Statistical analyses

The association of both *HOXB13* mutations (*i.e.* p.G84E and p.R217C) with breast cancer

risk was evaluated by comparing the carrier allele frequency between cases and controls using either a χ^2 test or a Fisher's exact test (*i.e.* when the expected frequency $\leqslant$5 in any of the groups). Odds ratios and their 95% confidence intervals were calculated based on 2×2 table analysis of the cases and controls. All statistical tests were two-sided and *P*-values were considered statistically significant when smaller than 0.05.

Results

HOXB13 whole gene screen

We evaluated the entire coding sequence of the *HOXB13* gene for germline mutations in 1,250 non-*BRCA1/2* breast cancer patients and 800 controls from the Rotterdam Breast Cancer Study (RBCS) study. Using PCR and Sanger sequencing, we identified a total of eleven different rare variants (Table 3.1) and two more common variants (c.366C>T; p.S122S; rs8556; minor allele frequency (MAF) cases=0.126; MAF controls=0.138 and c.513T>C; p.S171S; rs9900627; MAF cases=0.079; MAF controls=0.091). Seven of the eleven rare variants were missense variants and five of these were present either in only one case or one control. The other two missense variants (*i.e.* c.251G>A and c.649C>T) were detected in multiple cases and controls (Figure 3.1). The c.251G>A (p.G84E) mutation was detected in 4 of 1,215 (0.33%) cases and 6 of 759 (0.79%) controls, whereas the c.649C>T mutation was detected in 6 of 1,206 (0.50%) cases and 1 of 765 (0.13%) controls (Table 3.1). For all identified missense variants the carrier frequency was low, resulting in insufficient power to draw meaningful statistical inferences from this sample size. However, it did appear that the prostate cancer risk variant p.G84E was less prevalent in breast cancer cases than controls, whereas the prevalence of the p.R217C variant appeared to be higher in breast cancer cases compared with controls. Interestingly, both missense mutations were predicted to be deleterious based on three different prediction classification tools: PredictSNP[27] (*i.e.* 87% for p.G84E and p.R217C), Meta-SNP[28] (*i.e.* 0.730 for p.G84E and 0.895 for p.R217C) and PON-P2[29] (*i.e.* 0.967 for p.G84E and 0.974 for p.R217C). Moreover, p.G84E is localized in the MEIS binding domain, whereas p.R217C is localized to the homeodomain of HOXB13,

further increasing the likelihood that these mutations are pathogenic[10]. For these reasons, we decided to further pursue these two variants in a second sample set by expanding RBCS (*i.e.* to all indexes from families counselled between 1994 and 2014 for cases and between 1996 and 2010 for controls) and by including the Amsterdam Breast Cancer Study (ABCS-F).

Table 3.1 Rare variants identified by PCR and Sanger sequencing of the *HOXB13* gene.

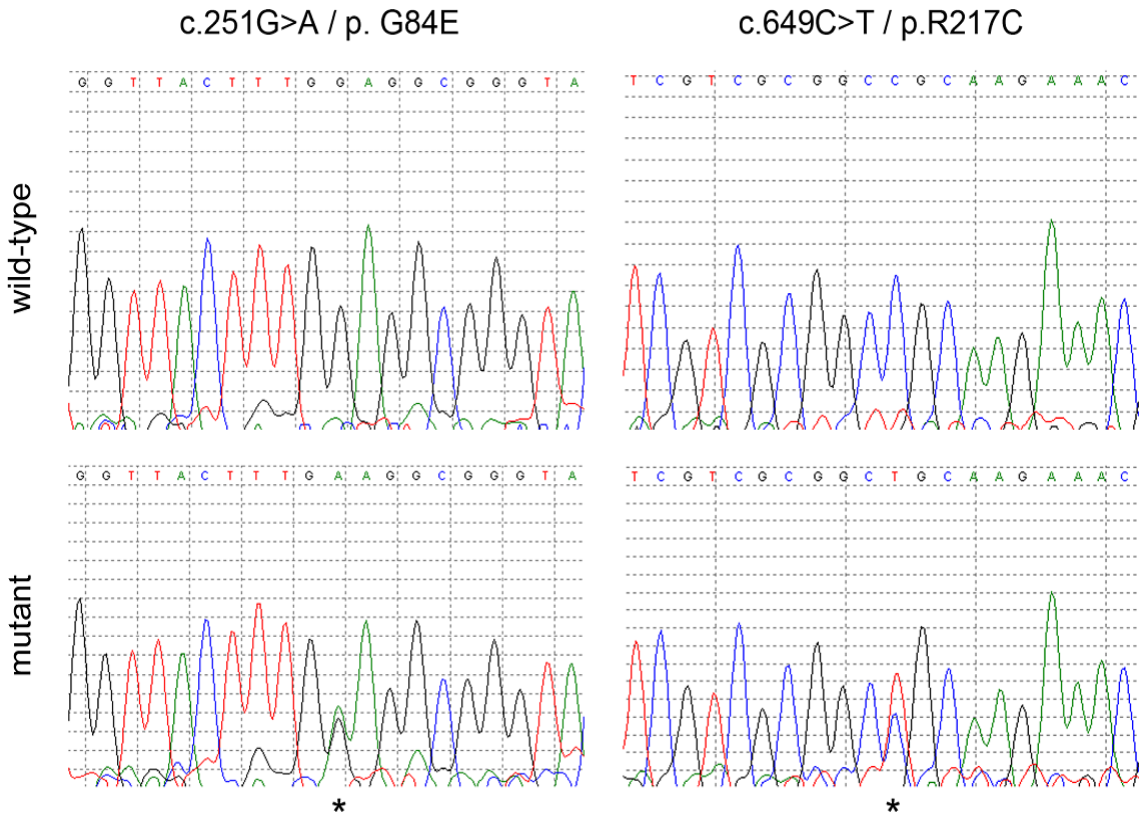
				Carrier allele frequency	
Position	**Nucleotide change**	**Amino acid change**	**Rs number**	**Controls**	**Cases**
5'UTR	c.1-6G>A			0/759 (0%)	1/1215 (0.08%)
Exon 1	c.251G>A	p.G84E	rs138213197	6/759 (0.79%)	4/1215 (0.33%)
Exon 1	c.328C>G	p.P110A		0/759 (0%)	1/1215 (0.08%)
Exon 1	c.330C>A	p.P110P	rs33993185	1/759 (0.13%)	0/1215 (0%)
Exon 1	c.332C>T	p.A111V		0/759 (0%)	1/1215 (0.08%)
Exon 1	c.569C>T	p.P190L		0/759 (0%)	1/1215 (0.08%)
Intron 1	c.601+49G>A		rs200606700	0/759 (0%)	1/1215 (0.08%)
Exon 2	c.649C>T	p.R217C	rs139475791	1/765 (0.13%)	6/1206 (0.50%)
Exon 2	c.803G>A	p.R268Q		0/765 (0%)	1/1206 (0.08%)
Exon 2	c.832G>T	p.V278L	rs200997384	1/765 (0.13%)	0/1206 (0%)
3'UTR	c.855+28C>A			3/765 (0.39%)	5/1206 (0.41%)

Genotyping *HOXB13* p.G84E and p.R217C

In order to facilitate fast and accurate screening of the p.G84E and p.R217C mutations, two custom-designed Taqman genotyping assays were developed for analyzing all samples from the RBCS and ABCS-F case-control studies. In total, all 4,520 non-*BRCA1/2* breast cancer patients and 3,127 controls were genotyped. These also included the 1,250 non-*BRCA1/2* breast cancer patients and 800 controls from the RBCS study that were used in the whole gene screen to evaluate the quality of the genotyping assay. The concordance between the results from the custom-designed Taqman genotyping assay and Sanger sequencing of these

patients was 100%. Interestingly, the p.G84E mutation was identified in 18 (0.56%) of 3,187 cases and 16 (0.70%) of 2,300 controls (Table 3.2). Consistent with the results from the whole gene screen, the p.G84E mutation was more prevalent in controls than cases, however, this was not statistically significant (OR=0.81, 95% CI=0.41-1.59, *P*=0.54). The p.R217C mutation was identified in 10 (0.31%) of 3,208 cases and 2 (0.087%) of 2,288 controls (Table 3.2). Consistent with the results of the whole gene screen, the p.R217C mutation was more prevalent in cases than in controls, but this difference was not significant (OR=3.57, 95% CI=0.76-33.57, *P*=0.14). These results imply that none of the recurrent *HOXB13* mutations in the Dutch population are associated with breast cancer risk.

Figure 3.1 Identification of the c.251G>A (p.G84E) and c.649C>T (p.R217C) mutations.



The lower electropherograms show the c.251C>A (left) and the c.649C>T (right) mutations which are indicated with an asterisk as compared with the wild-type sequences in the top panels.

Table 3.2 Association of *HOXB13* p.G84E and p.R217C with breast cancer risk.

		Carrier allele frequency			
Variant	**Study**	**Controls**	**Cases**	**OR (95% CI)**	***P*-value**
c.251G>A / p.G84E	RBCS	3/356 (0.84%)	9/1,465 (0.61%)		
	ABCS-F	13/1,944 (0.67%)	9/1,722 (0.52%)		
	Combined	16/2,300(0.70%)	18/3,187 (0.56%)	0.81 (0.41-1.59)	0.54
c.649C>T / p.R217C	RBCS	0/355 (0%)	5/1,473 (0.34%)		
	ABCS-F	2/1,933 (0.10%)	5/1,735 (0.29%)		
	Combined	2/2,288 (0.087%)	10/3,208 (0.31%)	3.57 (0.76-33.57)	0.14

OR: odds ratio, CI: confidence interval.

Discussion

The *HOXB13* c.251G>A (p.G84E) mutation has been shown to confer a 4- to 5-fold increased prostate cancer risk[11-13]. In this study, we have explored whether *HOXB13* gene mutations are also associated with breast cancer risk. Our results show that the prostate cancer risk variant p.G84E is not associated with breast cancer risk. Furthermore, another recurrent mutation in the *HOXB13* gene (*i.e.* c.649C>T; p.R217C) was also not associated with increased breast cancer risk, although it was more prevalent in cases than controls.

Interestingly, Alanee *et al.* had previously shown that the *HOXB13* p.G84E mutation conferred a moderate to high breast cancer risk[23]. The mutation was found in 6 (0.7%) of 877 familial, mostly Caucasian, non-*BRCA1/2* breast cancer cases, while the frequency in controls was 0.1% (OR=5.7, 95% CI=1.0-40.7, *P*=0.02). However, in a larger study (*i.e.* 4,037 cases of which 1,082 familial and 2,762 controls) conducted by Akbari *et al*., no association of the p.G84E mutation with breast cancer risk was observed among Canadian and Polish women of European origin[24]. The mutation was identified in 7 (0.17%) of 4,037 cases and 4 (0.14%) of 2,762 controls (OR=1.2, 95% CI=0.3-4.1, *P*=1.0). Also a third study by Laitinen *et al*. consisting of 986 cases (*i.e.* of which 323 familial non-*BRCA1/2* and 663 unselected) and 1,449 controls found no overall association between the p.G84E mutation and (familial)

breast cancer risk among Finnish women[25]. However, the authors did observe a suggestive association in a specific high-risk familial subgroup (*i.e.* 86 cases from the Pirkanmaa area of Finland; OR=3.2, 95% CI=0.9-11.9). Here in this study, we also did not observe an increased breast cancer risk associated with the p.G84E mutation in a relatively large study of 3,270 familial non-*BRCA1/2* breast cancer cases and 2,327 controls (OR=0.81, 95% CI=0.41-1.59, *P*=0.54). It thus appears that the *HOXB13* p.G84E mutation is not associated with increased breast cancer risk, although it cannot be excluded that it is associated with a specific high-risk subgroup. Considering the low population frequency, much larger studies are needed to determine whether the latter is indeed the case.

The whole gene screen for *HOXB13* also identified c.649C>T (p.R217C) as a recurrent mutation in the Dutch population, which was predicted to be pathogenic. In both the discovery as well as the validation phase of the study, the mutation was more prevalent in familial breast cases than controls, but the association with breast cancer risk was not significant. Considering the wide CIs and the very low population frequency, there is a chance that the study was underpowered and failed to detect the association. Evaluation of *HOXB13* p.R217C in a larger study or a population with a higher carrier allele frequency might therefore still be worthwhile to pursue. Since the p.G84E variant varies widely among different geographic populations (*i.e.* highest in North and West-Europeans and lowest in non-Europeans)[30,31], this may also be the case for p.R217C. Interestingly, the p.R217C mutation had been described before among a few prostate cancer cases[30,32], however, Xu *et al.* reported that p.R217C did not co-segregate with prostate cancer in the two families they identified[30]. Unfortunately, we were not able to perform informative segregation analysis in the present study as for only two families we had DNA available for two additional family members. In addition, we identified too little carriers of the mutation to say anything relevant regarding an excess of prostate cancer in their families as compared with non-carrier cases.

To conclude, none of the recurrent *HOXB13* mutations that we identified in the Dutch population were associated with breast cancer risk, although it may be worthwhile to evaluate p.R217C in a larger study or a population with a higher allele frequency.

Acknowledgements

We thank all patients and families included in this study. Furthermore, we want to thank Ellen van Schoot, Sanquin Bloodbank, Amsterdam for donating samples to this study and Renske Keeman and Carla van Tiggelen at the Antoni van Leeuwenhoek for database support. This study was funded by the Cancer Genomics Netherlands (CGC.nl) and a grant for the Netherlands Organization of Scientific Research (NWO). JL received a scholarship from the China Scholarship Council (Beijing, China). MKS is supported by grant NKI2009-4363 of the Dutch Cancer Society.

Author contributions

J.L. designed and performed the experiments, analyzed the data, wrote the paper and provided funding. W.J.C.P. designed and performed the experiments and analyzed the data. M.K.S. designed the study, contributed samples, provided clinical data, revised the manuscript and provided funding. J.M.C. designed the study, contributed samples and provided clinical data. S.C. performed experiments and contributed samples. R.L. performed experiments and contributed samples. A.N. provided clinical data. J.A.F. revised the manuscript and provided funding. M.J.H. designed the study, provided clinical data and revised the manuscript. S.V. designed the study, contributed samples and provided clinical data. A.M.W.O. designed the study, contributed samples, provided clinical data and revised the manuscript. F.B.L.H. designed the study, contributed samples, provided clinical data and revised the manuscript. J.W.M.M. designed the study, revised the manuscript and provided funding. A.H. conceived of the study, designed the study and the experiments, performed the experiments, analyzed the data and wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

References

1 Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* 358, 1389-1399 (2001).

2 Miki, Y., Swensen, J., Shattuck-Eidens, D. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66-71 (1994).

3 Wooster, R., Bignell, G., Lancaster, J. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789-792 (1995).

4 Antoniou, A., Pharoah, P. D., Narod, S. *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 72, 1117-1130 (2003).

5 Hollestelle, A., Wasielewski, M., Martens, J. W. *et al.* Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* 20, 268-276 (2010).

6 Antoniou, A. C., Casadei, S., Heikkinen, T. *et al.* Breast-cancer risk in families with mutations in PALB2. *N Engl J Med* 371, 497-506 (2014).

7 Cybulski, C., Kluzniak, W., Huzarski, T. *et al.* Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. *Lancet Oncol* 16, 638-644 (2015).

8 Michailidou, K., Beesley, J., Lindstrom, S. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47, 373-380 (2015).

9 Mavaddat, N., Pharoah, P. D., Michailidou, K. *et al.* Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107 (2015).

10 Ewing, C. M., Ray, A. M., Lange, E. M. *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 366, 141-149 (2012).

11 Shang, Z., Zhu, S., Zhang, H. *et al.* Germline homeobox B13 (HOXB13) G84E mutation and prostate cancer risk in Eurpean descendants: a meta-analysis of 24,213

cases and 73, 631 controls. *Eur Urol* 64, 173-176 (2013).

12 Witte, J. S., Mefford, J., Plummer, S. J. *et al.* HOXB13 mutation and prostate cancer: studies of siblings and aggressive disease. *Cancer Epidemiol Biomarkers Prev* 22, 675-680 (2013).

13 Huang, H. & Cai, B. G84E mutation in HOXB13 is firmly associated with prostate cancer risk: a meta-analysis. *Tumour Biol* 35, 1177-1182 (2014).

14 Cai, Q., Wang, X., Li, X. *et al.* Germline HOXB13 p.Gly84Glu mutation and cancer susceptibility: a pooled analysis of 25 epidemiological studies with 145,257 participates. *Oncotarget* 6, 42312-42321 (2015).

15 Saunders, E. J., Dadaev, T., Leongamornlert, D. A. *et al.* Fine-mapping the HOXB region detects common variants tagging a rare coding allele: evidence for synthetic association in prostate cancer. *PLoS Genet* 10, e1004129 (2014).

16 Norris, J. D., Chang, C. Y., Wittmann, B. M. *et al.* The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol Cell* 36, 405-416 (2009).

17 Bhatlekar, S., Fields, J. Z. & Boman, B. M. HOX genes and their role in the development of human cancers. *J Mol Med (Berl)* 92, 811-823 (2014).

18 Huang, Q., Whitington, T., Gao, P. *et al.* A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* 46, 126-135 (2014).

19 Wang, Z., Dahiya, S., Provencher, H. *et al.* The prognostic biomarkers HOXB13, IL17BR, and CHDH are regulated by estrogen in breast cancer. *Clin Cancer Res* 13, 6327-6334 (2007).

20 Ma, X. J., Hilsenbeck, S. G., Wang, W. *et al.* The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol* 24, 4611-4619 (2006).

21 Jansen, M. P., Sieuwerts, A. M., Look, M. P. *et al.* HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol* 25, 662-668 (2007).

22 Shah, N., Jin, K., Cruz, L. A. *et al.* HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ERalpha and inducing IL-6

expression. *Cancer Res* 73, 5449-5458 (2013).

23 Alanee, S., Couch, F. & Offit, K. Association of a HOXB13 variant with breast cancer. *N Engl J Med* 367, 480-481 (2012).

24 Akbari, M. R., Kluzniak, W., Rodin, R. *et al.* The HOXB13 p.Gly84Glu mutation is not associated with the risk of breast cancer. *Breast Cancer Res Treat* 136, 907-909 (2012).

25 Laitinen, V. H., Wahlfors, T., Saaristo, L. *et al.* HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 22, 452-460 (2013).

26 van den Broek, A. J., de Ruiter, K., van 't Veer, L. J. *et al.* Evaluation of the Dutch BRCA1/2 clinical genetic center referral criteria in an unselected early breast cancer population. *Eur J Hum Genet* 23, 588-595 (2015).

27 Bendl, J., Stourac, J., Salanda, O. *et al.* PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* 10, e1003440 (2014).

28 Capriotti, E., Altman, R. B. & Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* 14 Suppl 3, S2 (2013).

29 Niroula, A., Urolagin, S. & Vihinen, M. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS One* 10, e0117380 (2015).

30 Xu, J., Lange, E. M., Lu, L. *et al.* HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 132, 5-14 (2013).

31 Chen, Z., Greenwood, C., Isaacs, W. B. *et al.* The G84E mutation of HOXB13 is associated with increased risk for prostate cancer: results from the REDUCE trial. *Carcinogenesis* 34, 1260-1264 (2013).

32 Karlsson, R., Aly, M., Clements, M. *et al.* A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur Urol* 65, 169-176 (2014).

Chapter 4

rs2735383, located at a microRNA binding site in the 3'UTR of *NBS1*, is not associated with breast cancer risk

Jingjing Liu, Ivona Lončar, J. Margriet Collée, Manjeet K. Bolla, Joe Dennis, Kyriaki Michailidou, Qin Wang, Irene L. Andrulis, Monica Barile, Matthias W. Beckmann, Sabine Behrens, Javier Benitez, Carl Blomqvist, Bram Boeckx, Natalia V. Bogdanova, Stig E. Bojesen, Hiltrud Brauch, Paul Brennan, Hermann Brenner, Annegien Broeks, Barbara Burwinkel, Jenny Chang-Claude, Shou-Tung Chen, Georgia Chenevix- Trench, Ching Y. Cheng, Ji-Yeob Choi, Fergus J. Couch, Angela Cox, Simon S. Cross, Katarina Cuk, Kamila Czene, Thilo Dörk, Isabel dos-Santos-Silva, Peter A. Fasching, Jonine Figueroa, Henrik Flyger, Montserrat García-Closas, Graham G. Giles, Gord Glendon, Mark S. Goldberg, Anna González-Neira, Pascal Guénel, Christopher A. Haiman, Ute Hamann, Steven N. Hart, Mikael Hartman, Sigrid Hatse, John L. Hopper, Hidemi Ito, Anna Jakubowska, Maria Kabisch, Daehee Kang, Veli-Matti Kosma, Vessela N. Kristensen, Loic Le Marchand, Eunjung Lee, Jingmei Li, Artitaya Lophatananon, Jan Lubinski, Arto Mannermaa, Keitaro Matsuo, Roger L. Milne, NBCS Collaborators, Susan L. Neuhausen, Heli Nevanlinna, Nick Orr, Jose I.A. Perez, Julian Peto, Thomas C. Putti, Katri Pylkäs, Paolo Radice, Suleeporn Sangrajrang, Elinor J. Sawyer, Marjanka K. Schmidt, Andreas Schneeweiss, Chen-Yang Shen, Martha J. Shrubsole, Xiao-Ou Shu, Jacques Simard, Melissa C. Southey, Anthony Swerdlow, Soo H. Teo, Daniel C. Tessier, Somchai Thanasitthichai, Ian Tomlinson, Diana Torres, Thérèse Truong, Chiu-Chen Tseng, Celine Vachon, Robert Winqvist, Anna H. Wu, Drakoulis Yannoukakos, Wei Zheng, Per Hall, Alison M. Dunning, Douglas F. Easton, Maartje J. Hooning, Ans M.W. van den Ouweland, John W.M. Martens, Antoinette Hollestelle.

Sci Rep 6, 36874 (2016).

Abstract

NBS1, also known as NBN, plays an important role in maintaining genomic stability. Interestingly, rs2735383 G>C, located in a microRNA binding site in the 3'-untranslated region (UTR) of *NBS1*, was shown to be associated with increased susceptibility to lung and colorectal cancer. However, the relation between rs2735383 and susceptibility to breast cancer is not yet clear. Therefore, we genotyped rs2735383 in 1,170 familial non-*BRCA1/2* breast cancer cases and 1,077 controls using PCR-based restriction fragment length polymorphism (RFLP-PCR) analysis, but found no association between rs2735383CC and breast cancer risk (OR=1.214, 95% CI=0.936-1.574, *P*=0.144). Because we could not exclude a small effect size due to a limited sample size, we further analyzed imputed rs2735383 genotypes (r^2>0.999) of 47,640 breast cancer cases and 46,656 controls from the Breast Cancer Association Consortium (BCAC). However, rs2735383CC was not associated with overall breast cancer risk in European (OR=1.014, 95% CI=0.969-1.060, *P*=0.556) nor in Asian women (OR=0.998, 95% CI=0.905-1.100, *P*=0.961). Subgroup analyses by age, age at menarche, age at menopause, menopausal status, number of pregnancies, breast feeding, family history and receptor status also did not reveal a significant association. This study therefore does not support the involvement of the genotype at *NBS1* rs2735383 in breast cancer susceptibility.

Introduction

The DNA damage response (DDR) pathway maintains the stability of the human genome via a complex network of pathways integrating signal transduction, regulation of the cell cycle and repair of DNA. Double-strand breaks (DSBs), a particularly severe form of DNA damage, arise as a consequence of cell replication, programmed DNA rearrangements (*i.e.* meiosis and VDJ recombination) and exposure to carcinogens. When left unrepaired, DSBs may cause genomic instability, cell death and cancer[1,2]. In fact, mutations in genes involved in DSB repair, but also in the DDR pathway in general, are involved in the etiology of many human cancers. The two major repair pathways that mediate the repair of DSBs are the template-mediated homologous recombination repair pathway and the more error-prone non-homologous end-joining pathway[3,4]. The MRE11/RAD50/NBS1 complex is an important regulator of DSB repair through these pathways as this complex not only acts as a sensor of DSBs, but also recruits and activates the ATM protein to the break and activates it[5]. Activation of ATM, the central mediator of response to DSBs, initiates a cascade of signaling pathways involved in cell cycle checkpoint control, DNA repair and, when necessary, apoptosis by phosphorylation of p53, CHEK2, BRCA1, FANCD2 and NBS1 amongst others[6].

The DDR plays an important role in susceptibility to breast cancer. In fact, all of the currently identified high- and moderate-risk breast cancer genes (*i.e. BRCA1*, *BRCA2*, *CHEK2, ATM, NBS1* and *PALB2)* are involved in DNA repair[7,8]. As the majority of familial breast cancer risk is not yet attributable to known risk genes, this makes other genes encoding proteins involved in the DDR pathway attractive candidates for breast cancer susceptibility genes. The recent identification of the early DNA damage response gene *MCPH1* as a novel breast cancer susceptibility gene illustrates that this hypothesis still holds[9].

In this respect, the *NBS1* gene is located at chromosome 8q21 and bi-allelic germline mutations in *NBS1* cause the chromosomal instability syndrome Nijmegen breakage syndrome[10]. In addition, heterozygous carriers of *NBS1* mutations are at an increased risk to develop several types of cancer[11]. The *NBS1* c.657del5 founder mutation is the most prevalent mutation implicated in Nijmegen breakage syndrome (*i.e.* 90%) and has its origin in the Slavic population[12]. The mutation confers an overall 2.5- to 3-fold increased cancer risk and is

associated with increased risk for breast cancer, prostate cancer and lymphoma specifically[13]. Two other *NBS1* mutations implicated in Nijmegen breakage syndrome are p.I171V and p.R215W. Although both mutations associate with an overall cancer risk of 4-fold and 2-fold, respectively, there does not seem to be an increased risk to develop breast cancer specifically[13].

Besides the rare Nijmegen breakage syndrome-associated mutations, two common variants in *NBS1* (*i.e.* p.E185Q; rs1805794 and c.2265+541G>C; rs2735383) have also been reported to be associated with risks for several cancer types. Recent meta-analyses for *NBS1* rs1805794 have, however, shown that this variant does not associate with breast cancer risk[13-16], while associations with lung cancer and urinary system cancer are still inconclusive[13,16-18]. The functional variant rs2735383, localized in the 3'UTR of *NBS1*, has been shown to modulate the binding ability of microRNA-629 in lung cancer cells and microRNA-509-5p in colorectal cancer cells, affect *NBS1* transcriptional activity and decrease *NBS1* mRNA and NBS1 protein levels[19,20]. Although rs2735383 has been associated with an increased risk of lung cancer and colorectal cancer[13,20], its association with breast cancer risk is yet unclear. For this reason, we assessed whether *NBS1* rs2735383 is associated with breast cancer risk in the Rotterdam Breast Cancer Study (RBCS) by RFLP-PCR and in 45 studies of BCAC through imputation of the iCOGS array[21].

Materials and Methods

Study population

RBCS cases (N=1,269) came from the database of the Clinical Genetics Department at Erasmus University Medical Centre in Rotterdam, representing the Southwestern part of the Netherlands. First, we selected families that presented with at least two cases of female breast cancer or at least one case of female breast cancer and one case of ovarian cancer in first- or second-degree relatives. In addition, at least one of these two cases needed to be diagnosed before the age of 60. For each selected family, the youngest breast cancer patient who had been tested for *BRCA1* and *BRCA2* was then assigned to be the index case and included in RBCS. Furthermore, breast cancer cases were also included if they were diagnosed either

before the age of 40 years with unilateral breast cancer or before 50 years of age with bilateral breast cancer and did not report a family history of either breast or ovarian cancer in a first or second degree relative. All index cases and their tested relatives did not carry a *BRCA1* or *BRCA2* mutation. The median age of the RBCS cases was 44 years (range 18-92 years). RBCS controls (N=1,159) came from the same database and geographic location as the RBCS cases and included women from cystic fibrosis families who were either spouses of individuals at risk of being carrier of a *CFTR* mutation or individuals who were tested negative for a *CFTR* mutation. The median age of the RBCS controls was 41 years (range 10-97 years).

BCAC consists of case-control studies of unrelated women[21]. For the purpose of the current analyses, only studies with participants of European and Asian ancestry were included, resulting in a total of 45 case-control studies (Supplementary Table S4.3). Studies with participants of African ancestry (*i.e.* two studies) were not included because power in the analyses would be low due to a relatively low MAF (*i.e.* 0.123) and small amount of cases (*i.e.* 1,046). Each study was approved by its relevant governing research ethics committee and all study participants provided written informed consent. The experimental protocol was approved by the Medical Ethical Committee of the Erasmus Medical Center Rotterdam and the study was carried out in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.fmwv.nl). Following genotyping on the iCOGS array[21], quality control exclusions (described below), and analysis-specific exclusions, data from the following women were available for analysis: 47,640 patients with invasive breast cancer and 46,656 controls, totaling 94,296 BCAC participants.

PCR-based RFLP analysis

A 324bp fragment of the 3'UTR of *NBS1* including rs2735383 was amplified in a duplex PCR reaction together with a 713bp fragment of the *LRRC4* gene. Primers for *NBS1* and *LRRC4* were present in the PCR reaction at a final concentration of 0.25 and 1µM, respectively, and sequences are available in Supplementary Table S4.4. The amplified *LRRC4* fragment served as an internal digestion control and generated two fragments of 549 and

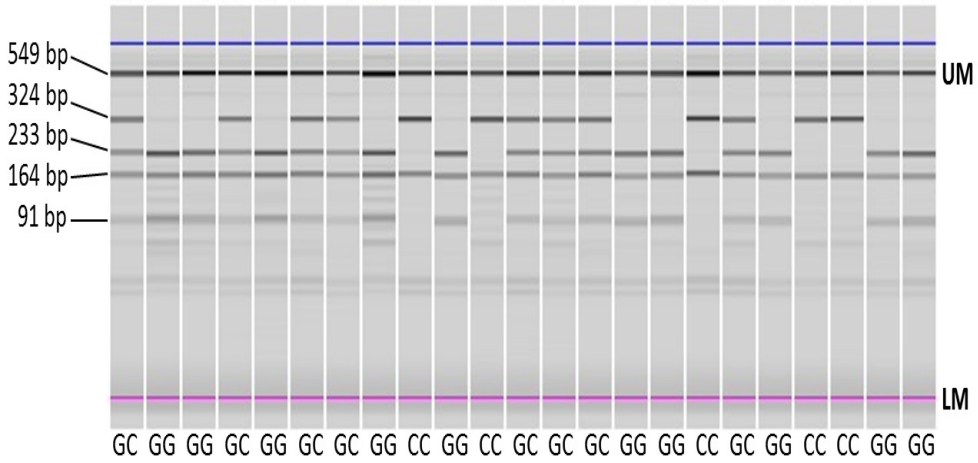
164bp upon complete digestion with ScfI (New England Biolabs, Frankfurt am Main, Germany). The 324bp amplified *NBS1* fragment was only digested when the major G allele was present at rs2735383, thereby generating two fragments of 233 and 91bp. Thus upon successful digestion with ScfI, samples with rs2735383GG generated four fragments, samples with rs2735383GC generated five fragments and samples with rs2735383CC generated three fragments (Figure 4.1).

iCOGS genotyping and imputation

Genotyping of BCAC studies was performed previously using the custom iCOGS Illumina Infinium iSelect BeadChip[21]. Briefly, DNA samples from 114,255 BCAC participants were genotyped, along with HapMap2 DNAs for European, African, and Asian populations. Raw intensity data files underwent centralized genotype calling and quality control[21]. The HapMap2 samples were used to identify women with predicted European and Asian ancestry by performing principal component (PC) analysis using a set of over 37,000 unlinked markers[22]. Nine European PCs and two Asian PCs were found to control adequately for residual population stratification in BCAC data. Samples with a low conversion rate, extreme heterozygosity, non-female sex, or one of a first-degree relative pair were excluded. Variants were excluded if they were monomorphic or had a call rate <95% (*i.e.* when MAF>0.05) or <99% (*i.e.* when MAF<0.05), deviation from HWE (*i.e.* P<10-7), or >2% duplicate discordance.

Imputation of genotypes was performed using 1000 Genomes Project data (v3 April 2012 release) as the reference panel[23,24]. To improve computation efficiency we used a two-step procedure which involved pre-phasing by chromosome and by chunk using SHAPEIT software in the first step[25] and imputation of the phased data using IMPUTE version 2 software in the second[26]. *NBS1* rs2735383 was imputed with an imputation $r^2 > 0.999$ in both Europeans and Asians.

Figure 4.1. Microchip electrophoresis of the RFLP-PCR products of 23 RBCS cases.



After PCR amplification of *NBS1* and *LRRC4* fragments (*i.e.* 324 and 713bp), digestion with SfcI generated four fragments (*i.e.* 549, 233, 164 and 91bp) for samples with rs2735383 GG genotypes, five fragments (*i.e.* 549, 324, 233, 164 and 91bp) for samples with rs2735383 GC genotypes and three fragments (*i.e.* 549, 324 and 164bp) for samples with rs2735383 CC genotypes. UM, upper marker; LM, lower marker.

Statistical analyses

The association between *NBS1* rs2735383 and invasive breast cancer risk was evaluated by logistic regression analysis providing ORs and 95% CIs. In the analyses of BCAC studies, ORs were adjusted for study, age, and PCs. In the analyses of RBCS, ethnicity was not a confounding factor thus reported ORs were unadjusted for PCs. For the European and Asian BCAC studies, we additionally performed the study-specific logistic regression analysis adjusting for age and PCs, and pooled the log ORs in a fixed-effects meta-analysis. Subgroup analyses within the European BCAC studies were based on age (*i.e.* ≤ 50 years and >50 years), age at menarche (*i.e.* ≤13 years and >13 years), age at menopause (*i.e.* ≤50 years and >50 years), menopausal status (*i.e.* premenopausal and postmenopausal), number of full-term pregnancies (*i.e.* ≤ 2 and >2), breast feeding (*i.e.* no and yes), first-degree family history of breast cancer and receptor status (*i.e.* ER positive, ER negative and triple negative). Clinical and demographic characteristics of the BCAC cases are presented in Supplementary Table

S4.5. Association between *NBS1* rs2735383 and the clinical and demographic characteristics were evaluated using a χ^2 test. All *P*-values were two-sided and $P<0.05$ was considered to be statistically significant after correction for multiple testing by the Bonferroni procedure. Logistic regression analyses were performed using SPSS statistics version 23 (IBM Corporation, Armonk, NY) and fixed-effects meta-analyses using Stata version 13 (StataCorp, College Station, TX).

Results

To evaluate the association between *NBS1* rs2735383 and breast cancer risk, we analyzed *NBS1* rs2735383 by RFLP-PCR in 1,269 non-*BRCA1/2* familial breast cancer patients and 1,159 controls from RBCS. Since genetic risk factors are usually enriched in familial/ early-onset breast cancer cases, specifically selecting these breast cancer patients improves statistical power. Among the cases, 516 had the GG genotype, 507 had the GC genotype and 147 had the CC genotype at rs2735383 (minor allele frequency (MAF)=0.342). Among the controls, 462 had the GG genotype, 501 had the GC genotype and 114 had the CC genotype (MAF=0.338). For both cases and controls, the genotypes of rs2735383 were in Hardy-Weinberg equilibrium (HWE). Because rs2735383 CC was associated with an increased risk of lung cancer and colorectal cancer under a recessive genetic model[13,20], we analyzed the association of rs2735383 with breast cancer in a similar way. However, rs2735383 was not significantly associated with the risk of breast cancer (OR=1.214, 95% CI=0.936-1.574, *P*=0.144; Table 4.1). In this respect, the lung cancer risk conferred by the rs2735383 CC genotype had been associated with an OR of 1.28 (95% CI=1.21-1.46, $P<0.001$), whereas the colorectal cancer risk had been associated with an OR of 1.55 (95% CI=1.27-1.94, $P<10^{-4}$)[13,20]. Here, we do not observe a similar effect size for breast cancer as for lung and colorectal cancer. However, RBCS is underpowered to detect effect sizes smaller than those observed for lung cancer *(i.e.* OR<1.28). Therefore, we cannot exclude rs2735383 CC is associated with breast cancer, but confers a smaller risk.

Table 4.1 Association of *NBS1* rs2735383 with breast cancer risk in the RBCS study.

Genetic model	N Controls	N Cases	OR (95% CI)	*P*-value
Recessive				
GG + GC	963	1023	1	
CC	114	147	1.214 (0.936-1.574)	0.144
	1077	1170		

N, number of; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

Table 4.2 Association of *NBS1* rs2735383 with overall breast cancer risk in the European and Asian BCAC studies.

Ethnicity	Genetic model	N Controls	N Cases	MAF Controls	MAF Cases	OR (95% CI)*	*P*-value*
European		40,0042	41,915	33.44%	33.39%		
	Recessive					1.014 (0.969-1.060)	0.556
	Dominant					1.006 (0.978-1.035)	0.684
	Additive					1.000 (0.979-1.021)	0.984
Asian		6,614	5,725	40.71%	40.58%		
	Recessive					0.998 (0.905-1.100)	0.961
	Dominant					0.995 (0.922-1.074)	0.900
	Additive					0.997 (0.946-1.050)	0.911

N, number of; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

* adjusted for age, study and principal components (PCs). In the European analyses nine PCs were added to the regression model and in the Asian analyses two PCs.

For this reason we analyzed *NBS1* rs2735383 in BCAC studies through imputation. Since we had data available for RBCS on rs2735383 from both the PCR-based RFLP and from imputation, we first evaluated the concordance between the two methods. In total, from 1,313 samples (*i.e.* 646 cases and 667 controls) we had genotypes for rs2738353 available from both RFLP-PCR and imputation. Importantly, the agreement between the two methods was 97.1% (*i.e.* concordance in 1,275 of 1,313 samples, r^2=0.933) and was similar among cases

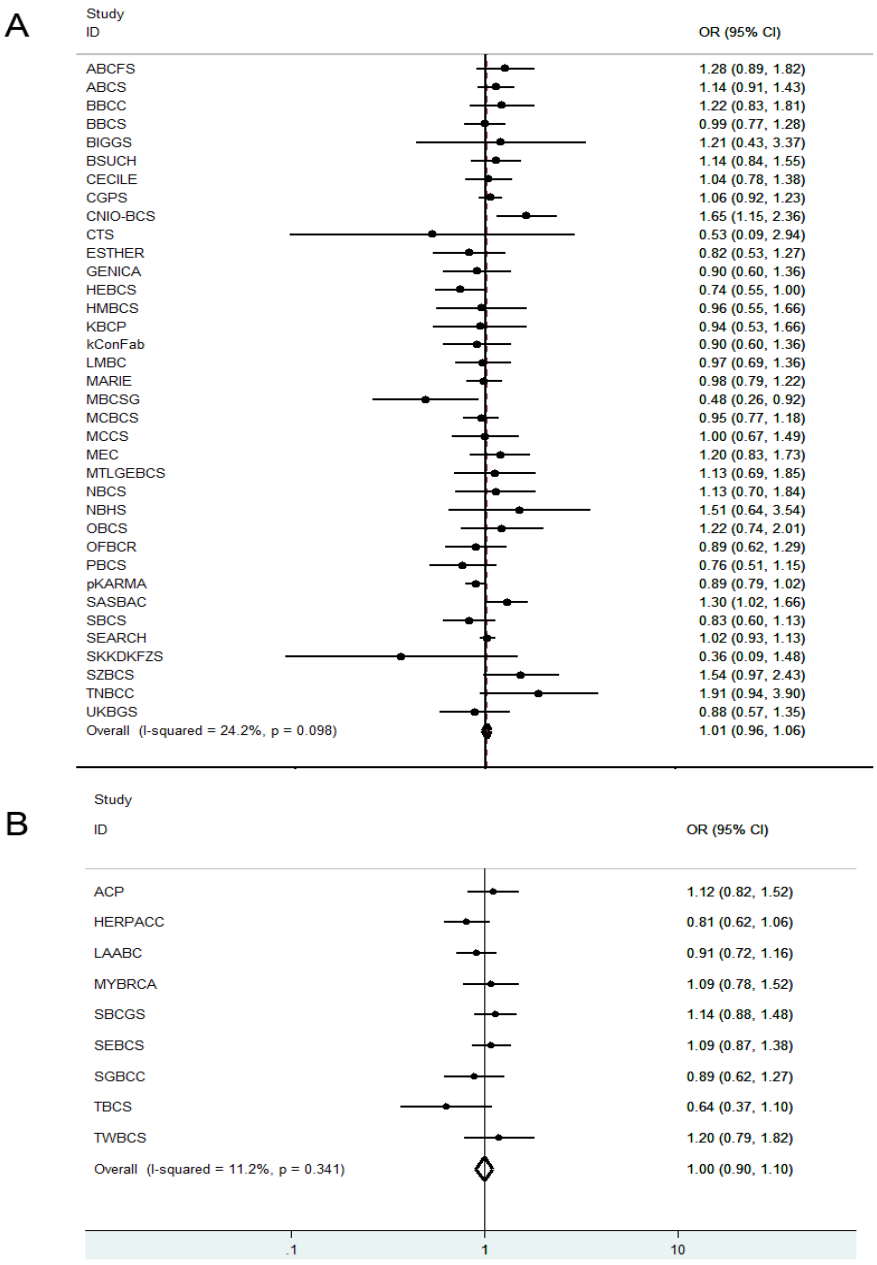
and controls (*i.e.* 98.1% versus 96.1%). Moreover, case-control ORs for imputed data were comparable to ORs obtained by RFLP-PCR (OR=1.14, 95% CI=0.80-1.62 versus OR=1.17, 95% CI=0.83-1.66). Therefore, we used the imputed data on rs2735383 to evaluate further its association with breast cancer risk.

For the overall analysis in Europeans we had 41,915 cases and 40,042 controls available from 36 case-control studies. However, rs2735383 was not associated with breast cancer risk in Europeans, neither under a recessive genetic model (OR=1.014, 95% CI=0.969-1.060, *P*=0.556; Table 4.2 and Figure 4.2), nor under a dominant (OR=1.006, 95% CI=0.978-1.035, *P*=0.684; Table 4.2) or additive model (per allele OR=1.000, 95% CI=0.979-1.021, *P*=0.984; Table 4.2). Because the association with increased lung and colorectal cancer risk was observed in the Asian population[13,20], we also performed the same analysis in the nine Asian BCAC studies. In total, we had 5,725 cases and 6,614 controls available for this analysis from nine case-control studies. Also in Asians we did not find any association between rs2735383 and breast cancer risk for either the recessive (OR=0.998, 95% CI=0.905-1.100, *P*=0.961; Table 4.2 and Figure 4.2), dominant (OR=0.995, 95% CI=0.922-1.074, *P*=0.900; Table 4.2) or additive genetic model (per allele OR=0.997, 95% CI=0.946-1.050, *P*=0.911; Table 4.2). These results imply that *NBS1* rs2735383 is not associated with an increased risk to develop invasive breast cancer.

A previous study had shown that rs2735383 may be associated with breast cancer risk in women >50 years, women with age at menarche >13 years, women with premenopausal status, women with number of abortions $\leqslant$2 and women who have breast fed, but not by age at menopause, number of pregnancies and family history[27]. Therefore, to exclude that an association of rs2735383 with breast cancer risk exists in a particular subgroup of individuals or breast cancer patients, we performed subgroup analysis according to age, age at menarche, age at menopause, menopausal status, number of full-term pregnancies, breast feeding, family history and receptor status. We did, however, not find any association between the genotype at rs2735383 and the risk of breast cancer in any of these subgroups for the recessive genetic model (Table 4.3). Also for the dominant and additive genetic models we found no association between the genotype at rs2735383 and breast cancer risk that would withstand

multiple testing correction (Supplementary Tables S4.1 and S4.2). *NBS1* rs2735383 is thus not associated with the risk for breast cancer, either in the overall analyses or in specific subgroups.

Figure 4.2 Forest plots for the association between rs2735383 and breast cancer risk.



A) for the 36 European BCAC studies and B) for the nine Asian BCAC studies. Study-specific (squares) were from a recessive genetic model and adjusted by age and PCs. Overall or pooled ORs (diamonds) were from a fixed-effects meta-analysis.

Table 4.3 Subgroup analysis of *NBS1* rs2735383 and breast cancer risk in the European BCAC studies.

Subgroup	N Controls	N Cases	MAF Controls	MAF Cases	OR (95% CI)*	*P*-value*
Age						
≤ 50 years	13,055	13,362	33.76%	33.41%	0.977 (0.899-1.062)	0.581
> 50 years	26,987	28,553	33.28%	33.38%	1.026 (0.971-1.084)	0.356
Age at menarche						
≤ 13 years	14,312	13,843	33.72%	33.13%	0.984 (0.914-1.060)	0.677
> 13 years	8,964	8,095	32.65%	33.63%	1.077 (0.978-1.187)	0.131
Age at menopause						
≤ 50 years	5,571	7,288	32,79%	33.43%	1.019 (0.906-1.146)	0.755
> 50 years	3,366	4,262	33.50%	33.24%	0.993 (0.855-1.154)	0.926
Menopausal status						
Premenopausal	8,974	7,412	33.66%	33.07%	0.981 (0.887-1.085)	0.715
Postmenopausal	19,648	17,353	33.39%	33.56%	1.007 (0.943-1.075)	0.844
Number of full-term pregnancies						
≤ 2	21,008	19,722	33.53%	33.20%	1.004 (0.942-1.071)	0.893
> 2	8,258	7,327	33.26%	33.44%	1.032 (0.931-1.144)	0.549
Breast feeding						
No	6,849	6,805	33.36%	33.25%	0.988 (0.884-1.104)	0.828
Yes	11,947	12,709	33.68%	33.28%	0.978 (0.903-1.060)	0.594
Family history						
1st degree relative with BC	23,648	4,119	33.21%	32.68%	0.990 (0.884-1.108)	0.859
Receptor status						
ER positive	39,699	25,959	33.47%	33.40%	1.021 (0.970-1.075)	0.427
ER negative	39,618	6,774	33.42%	32.90%	0.991 (0.908-1.082)	0.846
Triple negative	30,696	2,712	33.10%	32.04%	0.980 (0.847-1.134)	0.788

N, number; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; ER, estrogen receptor; BC, breast cancer

*recessive genetic model adjusted for age, study and nine principal components.

Discussion

The CC genotype of the common variant rs2735383 in the 3'UTR of *NBS1* has been shown to be associated with an increased cancer risk, specifically for lung and colorectal cancer (lung cancer: OR=1.28, 95% CI=1.21-1.46, $P<0.001$ and colorectal cancer: OR=1.55, 95% CI= 1.27-1.94, $P<10^{-4}$)[13,20]. In the current study, we evaluated the association of *NBS1* rs2735383 with breast cancer risk. We found that the CC genotype of rs2735383 did not confer an increased breast cancer risk, neither in the overall analyses nor in the subgroup analyses.

In agreement with these results, a small study by Han *et al.* consisting of 239 premenopausal breast cancer patients and 477 matched controls from the Nurses' Health Study II showed that rs2735383 did not associate with breast cancer risk under an additive genetic model (OR=0.92, 95% CI=0.72-1.16, $P=0.469$)[28]. Moreover, the study of Wu *et al*. consisting of 450 breast cancer patients and 450 cancer-free controls from the Henan Province in China also found no association with overall breast cancer risk[27]. However, after stratification according to reproductive factors, rs2735383 CC was found to be associated with an increased breast cancer risk for women >50 years, women with age at menarche >13 years, women with premenopausal status, women with number of abortions $\leqslant$2 and women who have breast fed, but not by age at menopause, number of pregnancies and family history[27]. In the current study we therefore also performed subgroup analysis by age, age at menarche, age at menopause, menopausal status, number of full-term pregnancies, breast feeding, family history and receptor status, but did not find any association between rs2735383 and risk of breast cancer in any of these subgroups that would withstand multiple testing correction. We thus could not replicate the earlier positive findings in women >50 years, women with age at menarche >13 years, premenopausal women and women who have breast fed. A possible, but not very likely, explanation for the difference in outcome between the studies may be the European versus Asian ethnicity. In the current study we chose to perform the subgroup analysis only in the European studies and not the Asian studies as this made sure that we had sufficient power in the subgroup analysis to identify smaller effects of rs2735383 on breast cancer risk. In this respect, a more plausible explanation would be that subgroup analyses, especially in a small study population (*i.e.* 450 cases and 450 controls in

the study of Wu *et al.*), could have easily given rise to false positive findings. Therefore, one should be careful when reporting positive findings from multiple small subgroup comparisons and always use appropriate levels of statistical significance[29]. Unfortunately, in the study from Wu *et al.*, there is no mention of multiple testing correction.

It was found that the rs2735383CC genotype significantly decreased the expression of the *NBS1* gene through either binding of microRNA-629 to the 3'UTR of *NBS1* gene in lung cancer cells or the binding of microRNA-509-5p to the 3'UTR of *NBS1* gene in colorectal cancer cells[19,20]. Since low expression of *NBS1* may reduce the efficiency of DSB repair, this way the rs2735383CC genotype likely confers an increased lung and colorectal cancer risk. According to our study, however, the rs2735383CC genotype does not confer an increased breast cancer risk. Considering that in lung and colorectal cancer cells different microRNAs appear to be downregulating *NBS1* expression, tissue specific expression of these microRNAs may likely play a role. Besides microRNA-509 and microRNA-629, the C allele at rs2735383 has also been predicted to enhance the binding of microRNA-499 and microRNA-508 to the 3'UTR of *NBS1*[27]. However, if these microRNAs are not expressed in normal breast tissue, the CC genotype of rs2735383 will not associate with breast cancer risk as *NBS1* cannot be downregulated by any of these microRNAs. At least in breast cancer cells, none of these microRNAs, except microRNA-629, are expressed at substantial levels (source: TCGA Research Network; http://cancergenome.nih.gov/). Since microRNAs are often deregulated between normal tissue and cancer tissue, this does not necessarily represent the situation in normal breast cells. Thus, further evaluation of the miRNA expression levels in normal (breast) tissue, but also their correlation with the genotype at rs2735383 should provide more insight for the tissue specificity of rs2735383 and cancer risk.

Importantly, in contrast to lung and colorectal cancer susceptibility, the results of this study do not support the presence of an association (*i.e.* OR>1.04 for Europeans and OR> 1.11 for Asians) between the genotype at rs2735383 in the 3'UTR of *NBS1* and breast cancer susceptibility.

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. This study would not have been possible without their contributions.

Author Contributions

J. Liu designed and performed the experiments, analyzed the data, wrote the paper and provided funding. I.L. designed and performed the experiments and analyzed the data. J.M.C. designed the study, contributed samples and provided clinical data. M.K.B. and Q.W. provided database management, contributed samples and provided
clinical data. J.D. and K.M. analyzed the data, contributed samples and provided clinical data. I.A., M.B., M.W.B., S.B., C.B., B.B., N.V.B., S.E.B., H. Brauch, P.B., A.B., B.B., S.-T.C., C.Y.C., J.-Y.C., F.J.C., A.C., S.S.C., K.C., K.C., I.d.-S.-S., P.A.F., J.F., H.F., M.G.-C., G.G.G., G.G., M.S.G., P.G., C.A.H., S.N.H., M.H., S.H., H.I., A.J., M.K., D.K., V.-M.K., V.N.K., L.L.M., E.L., J. Li, A.L., J. Lubinski, A.M., K.M., R.L.M., NBCS Collaborators, S.L.N., H.N., N.O., J.I.A.P., J.P., T.C.P., K.P., P.R., S.S., E.J.S., A.S., C.-Y.S., M.J.S., X.-O.S., M.C.S., S.H.T., S.T., I.T., D.T., T.T., C-C.T., C.V., R.W., A.H.W., D.Y. and W.Z. contributed samples and provided clinical data. H.Brenner, J.C.-C., J.L.H., G.C.-T., T.D., U.H., M.K.S. and A.S. contributed samples, provided clinical data and revised the manuscript. A.G.-N., J.B. and D.C.T. coordinated the study, contributed samples and provided clinical data. A.M.D. designed and coordinated the study, contributed samples and provided clinical data. J.S. designed and coordinated the study, contributed samples, provided clinical data and funding. P.H. led and designed the study, contributed samples, provided clinical data and funding. D.F.E. conceived of the study, designed the study, contributed samples, provided clinical data and funding. M.J.H. designed the study, provided clinical data and revised the manuscript. A.M.W.O. designed the study, contributed samples and provided clinical data. J.W.M.M. designed the study, revised the manuscript and provided funding. A.H. conceived of the study, designed the study and the experiments, analyzed the data and wrote the paper. All authors

read and approved the final version of the manuscript.

Additional Information

Supplementary information: accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

References

1 Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 27, 247-254 (2001).

2 Stracker, T. H. & Petrini, J. H. The MRE11 complex: starting from the ends. *Nat Rev Mol Cell Biol* 12, 90-103 (2011).

3 Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* 7, 739-750 (2006).

4 van Gent, D. C. & van der Burg, M. Non-homologous end-joining, a sticky affair. *Oncogene* 26, 7731-7740 (2007).

5 Lavin, M. F. ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene* 26, 7749-7758 (2007).

6 Lee, J. H. & Paull, T. T. Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. *Oncogene* 26, 7741-7748 (2007).

7 Hollestelle, A., Wasielewski, M., Martens, J. W. *et al.* Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* 20, 268-276 (2010).

8 Easton, D. F., Pharoah, P. D., Antoniou, A. C. *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* 372, 2243-2257 (2015).

9 Mantere, T., Winqvist, R., Kauppila, S. *et al.* Targeted Next-Generation Sequencing Identifies a Recurrent Mutation in MCPH1 Associating with Hereditary Breast Cancer Susceptibility. *PLoS Genet* 12, e1005816 (2016).

10 Varon, R., Vissinga, C., Platzer, M. *et al.* Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 93, 467-476 (1998).

11 di Masi, A. & Antoccia, A. NBS1 Heterozygosity and Cancer Risk. *Curr Genomics* 9, 275-281 (2008).

12 Varon, R., Seemanova, E., Chrzanowska, K. *et al.* Clinical ascertainment of Nijmegen breakage syndrome (NBS) and prevalence of the major mutation, 657del5, in three Slav populations. *Eur J Hum Genet* 8, 900-902 (2000).

13 Gao, P., Ma, N., Li, M. *et al.* Functional variants in NBS1 and cancer risk: evidence from a meta-analysis of 60 publications with 111 individual studies. *Mutagenesis* 28, 683-697 (2013).

14 Wang, F., Zhao, S. & Qin, H. NBS1 8360G > C polymorphism is associated with breast cancer risk was not credible: appraisal of a recent meta-analysis. *Breast Cancer Res Treat* 128, 291-292 (2011).

15 Yao, F., Fang, Y., Chen, B. *et al.* Association between the NBS1 Glu185Gln polymorphism and breast cancer risk: a meta-analysis. *Tumour Biol* 34, 1255-1262 (2013).

16 He, Y. Z., Chi, X. S., Zhang, Y. C. *et al.* NBS1 Glu185Gln polymorphism and cancer risk: update on current evidence. *Tumour Biol* 35, 675-687 (2014).

17 Fang, W., Qiu, F., Zhang, L. *et al.* The functional polymorphism of NBS1 p.Glu185Gln is associated with an increased risk of lung cancer in Chinese populations: case-control and a meta-analysis. *Mutat Res* 770, 61-68 (2014).

18 Zhang, Y., Huang, Y. S., Lin, W. Q. *et al.* NBS1 Glu185Gln polymorphism and susceptibility to urinary system cancer: a meta-analysis. *Tumour Biol* 35, 10723-10729 (2014).

19 Yang, L., Li, Y., Cheng, M. *et al.* A functional polymorphism at microRNA-629-binding site in the 3'-untranslated region of NBS1 gene confers an increased risk of lung cancer in Southern and Eastern Chinese population. *Carcinogenesis* 33, 338-347 (2012).

20 Li, J. T., Zhong, B. Y., Xu, H. H. *et al.* Associations between NBS1 Polymorphisms and Colorectal Cancer in Chinese Population. *PLoS One* 10, e0132332 (2015).

21 Michailidou, K., Hall, P., Gonzalez-Neira, A. *et al.* Large-scale genotyping identifies

41 new loci associated with breast cancer risk. *Nat Genet* 45, 353-361 (2013).

22 Price, A. L., Patterson, N. J., Plenge, R. M. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38, 904-909 (2006).

23 Michailidou, K., Beesley, J., Lindstrom, S. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47, 373-380 (2015).

24 Abecasis, G. R., Auton, A., Brooks, L. D. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65 (2012).

25 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 9, 179-181 (2012).

26 Howie, B., Fuchsberger, C., Stephens, M. *et al.* Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44, 955-959 (2012).

27 Wu, Z., Wang, P., Song, C. *et al.* Evaluation of miRNA-binding-site SNPs of MRE11A, NBS1, RAD51 and RAD52 involved in HRR pathway genes and risk of breast cancer in China. *Mol Genet Genomics* 290, 1141-1153 (2015).

28 Han, J., Haiman, C., Niu, T. *et al.* Genetic variation in DNA repair pathway genes and premenopausal breast cancer risk. *Breast Cancer Res Treat* 115, 613-622 (2009).

29 Kivimaki, M., Batty, G. D., Kawachi, I. *et al.* Don't let the truth get in the way of a good story: an illustration of citation bias in epidemiologic research. *Am J Epidemiol* 180, 446-448 (2014).

Chapter 5

The 29.5 kb *APOBEC3B* deletion polymorphism is not associated with clinical outcome of breast cancer

Jingjing Liu, Anieta M. Sieuwerts, Maxime P. Look, Michelle van der Vlugt-Daane, Marion E. Meijer-van Gelder, John A. Foekens, Antoinette Hollestelle, John W.M. Martens.

PLoS One 11, e0161731 (2016).

Abstract

Increased *APOBEC3B* mRNA levels are associated with a hypermutator phenotype and poor prognosis in ER-positive breast cancer patients. In addition, a 29.5 kb deletion polymorphism of *APOBEC3B*, resulting in an *APOBEC3A-B* hybrid transcript, has been associated with an increased breast cancer risk and the hypermutator phenotype. Here we evaluated whether the *APOBEC3B* deletion polymorphism also associates with clinical outcome of breast cancer. Copy number analysis was performed by quantitative PCR (qPCR) in primary tumors of 1,756 Dutch breast cancer patients. The *APOBEC3B* deletion was found in 187 patients of whom 16 carried a two-copy deletion and 171 carried a one-copy deletion. The prognostic value of the *APOBEC3B* deletion for the natural course of the disease was evaluated among 1,076 lymph-node negative (LNN) patients who did not receive adjuvant systemic treatment. No association was found between *APOBEC3B* copy number values and the length of metastasis-free survival (MFS; hazard ratio (HR)=1.00, 95% confidence interval (CI)=0.90-1.11, *P*=0.96). Subgroup analysis by ER status also did not reveal an association between *APOBEC3B* copy number values and the length of MFS. The predictive value of the *APOBEC3B* deletion was assessed among 329 ER-positive breast cancer patients who received tamoxifen as the first-line therapy for recurrent disease and 226 breast cancer patients who received first-line chemotherapy for recurrent disease. No association between *APOBEC3B* copy number values and the overall response rate (ORR) to either tamoxifen (odds ratio (OR)=0.88, 95% CI=0.69-1.13, *P*=0.31) or chemotherapy (OR=0.97, 95% CI= 0.71-1.33, *P*=0.87) was found. Thus, in contrast to *APOBEC3B* mRNA levels, the *APOBEC3B* deletion polymorphism has neither a prognostic nor a predictive value for breast cancer patients. Although a correlation exists between *APOBEC3B* copy number and mRNA expression, it is relatively weak. This suggests that other mechanisms exist that may affect and therefore determine the prognostic value of *APOBEC3B* mRNA levels.

Introduction

Breast cancer, like most cancer types, is a heterogeneous disease. The heterogeneous nature of breast cancer, however, provides challenges for identifying appropriate markers for disease susceptibility and progression, as well as treatment selection. Accordingly, transcriptional profiling has identified five molecular subtypes of breast cancer, which differ in prognosis, efficacy of treatment and preferred site of metastasis[1-5]. More recently, the catalogues of mutations across human cancers have provided us insight into the mutational processes that drive tumorigenesis[6,7]. For breast cancer, five distinct mutational signatures have been defined that contribute in varying degree to the final mutational catalogue of a breast tumor[7]. One of the most pronounced mutational processes impacting breast tumorigenesis is driven by the AID/APOBEC family of cytidine deaminases and gives rise to C>T and C>G substitutions at TpCpN nucleotides. Moreover, this mutational process associates with regional somatic hypermutation or kataegis[6-8].

The *APOBEC3* gene cluster is located on chromosome 22q13.1-q13.2 and harbors seven *APOBEC3* genes that have evolved in primates (*i.e. APOBEC3A*, *APOBEC3B*, *APOBEC3C*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G* and *APOBEC3H*)[9]. APOBEC3s play a role in intracellular defense through restriction of retroviral infections, but also of infections from the cancer-associated hepatitis B virus, the human papilloma virus and human T-lymphotropic virus[10]. Moreover, APOBEC3A, APOBEC3B, APOBEC3C and APOBEC3F are also able to inhibit LINE1 retrotransposition[11,12]. Besides its role in innate immunity, APOBEC3B has recently been identified as an endogenous source of mutation in breast cancer[13]. *APOBEC3B* mRNA expression was found to be upregulated in most breast cancers and tumors expressing high levels of *APOBEC3B* had a 2-fold increase in mutations compared with tumors expressing low *APOBEC3B* levels. This suggests that APOBEC3B, at least in part, underlies the APOBEC-driven mutational process in breast cancer, but also in other cancers[13,14]. In line with these findings, high levels of *APOBEC3B* mRNA were associated with a shorter disease-free survival in ER-positive, LNN, systemically untreated patients, as well as with earlier recurrence in luminal subtype patients and with a more aggressive phenotype in Japanese breast cancers[15-17]. Moreover, *APOBEC3B* expression has been reported to be

associated with a strong enrichment of mitotic and cell cycle-related genes[16].

A 29.5 kb germline deletion between the fifth exon of *APOBEC3A* and the eighth exon of *APOBEC3B* has been identified that essentially removes the complete *APOBEC3B* coding region from the genome and generates a fusion transcript of *APOBEC3A* with the 3' untranslated region (UTR) of *APOBEC3B*[18]. With a worldwide frequency of 22.5%, the frequency of the germline *APOBEC3B* deletion variant varies widely among the different ethnic groups, ranging from being rare in African and European populations (*i.e.* 0.9% and 6%, respectively) to being common in Asian and American populations (*i.e.* 36.9% and 57.7%, respectively)[18]. Through a genome-wide association study of copy number variation, Long *et al*. found that the *APOBEC3B* deletion variant was associated with an increased risk to develop breast cancer in Chinese women[19]. This finding was replicated among European[20] and Southeast Iranian women[21], but not among Swedish women[22]. Interestingly, carriers of the *APOBEC3B* deletion were shown to have a greater *APOBEC3A* mRNA stability resulting in higher APOBEC3A levels, increased activity of APOBEC-driven mutational processes and more severe DNA damage[23,24]. However, the *APOBEC3B* deletion polymorphism was not associated with the survival of breast cancer patients[16,22]. Thus, despite that *APOBEC3B* overexpression and the *APOBEC3B* deletion variant both result in a hypermutation phenotype, there seems to be a difference between the two mechanisms and their association with clinical outcome.

To investigate the relation between the *APOBEC3B* deletion polymorphism and clinical outcome of 1,756 Dutch breast cancer patients, we explored four different clinical cohorts: LNN patients who did not receive any adjuvant treatment, lymph node positive (LNP) patients who did receive adjuvant systemic treatment, hormone-naive ER-positive patients who received tamoxifen as first-line therapy for recurrent disease and patients that received first-line chemotherapy for recurrent disease. Furthermore, to investigate the relation between the clinical outcomes based on the *APOBEC3B* deletion polymorphism versus *APOBEC3B* overexpression, we investigated the correlation between *APOBEC3B* copy number and *APOBEC3B* mRNA expression.

Materials and Methods

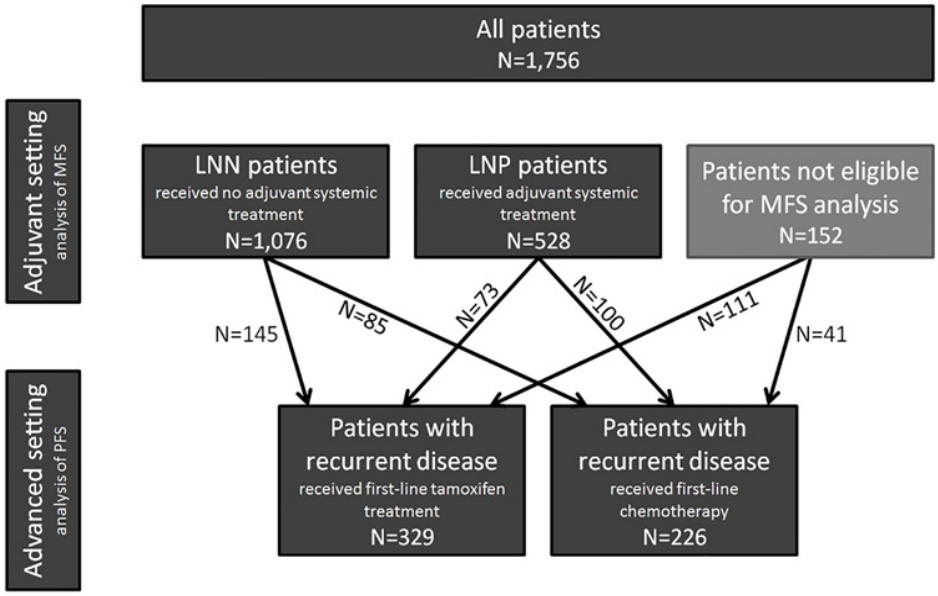
Study population

The retrospective study cohort consisted of 1,756 breast cancer patients who underwent surgery for an invasive primary breast cancer between 1978 and 2001. Inclusion criteria were no neoadjuvant treatment, no experience of a previous other cancer (except for basal cell carcinoma or stage Ia/Ib cervical cancer), a minimum of 100mg of freshly frozen primary tumor tissue available for downstream DNA isolation and DNA available from tissue with a tumor cell nuclei percentage $\geqslant$30%. Cytosolic ER and PR levels were determined by ligand binding assay or enzyme immunoassay[25,26]. ER and/or PR positivity was defined by $\geqslant$10 fmol/mg cytosolic protein and *ERBB2* overexpression was defined by a reverse transcriptase qPCR expression level $\geqslant$ 18[27]. In total, 796 patients underwent breast-conserving lumpectomy and 960 patients underwent modified radical mastectomy. In addition, 215 patients received adjuvant hormonal therapy, 308 patients received adjuvant chemotherapy and 6 patients received both adjuvant hormonal therapy and chemotherapy. There were 1,713 M0 patients and 43 M1 patients. The median age at the time of the primary surgery was 54 years, while the median age at the start of first-line treatment was 55 years. The clinicopathological variables of the patients are shown in Table 5.1.

The total study cohort consists of four specific studies that were grouped together: 1) 1,076 LNN patients who did not receive any adjuvant systemic treatment, 2) 528 LNP patients who received adjuvant systemic treatment, 3) 329 hormone-naive ER-positive patients who received first-line tamoxifen therapy for recurrent disease and 4) 226 patients who received first-line chemotherapy for recurrent disease as detailed in Figure 5.1. We included all eligible patients fulfilling the study criteria and the general inclusion criteria specified above. As a consequence 403 patients were included in two studies (*i.e.* one study in the adjuvant setting and one study in the advanced setting). The total study population, however, cannot be considered a consecutive series, since systemically treated LNN patients are missing. The total study population, however, cannot be considered a pure consecutive series, since systemically treated LNN patients were not included. The reason for this is that

we especially wished to study the association of *APOBEC3B* copy number with the natural course of the disease in LNN patients, not potentially confounded by adjuvant systemic therapy.

Figure 5.1 Schematic overview of study cohort.



In total, this retrospective study consists of 1,756 primary breast cancers from patients who underwent surgery between 1978 and 2001. Inclusion criteria are specified in the Materials and Methods section. In the adjuvant setting, there were 1,076 lymph node negative (LNN) patients who did not receive adjuvant systemic treatment and 528 lymph node positive (LNP) patients who received adjuvant systemic treatment for the analysis of MFS. In the advanced setting, a group of 329 hormone-naive patients with ER-positive breast cancer received first-line tamoxifen for recurrent disease. Of these, 145 patients came from the LNN patients group and 73 came from the LNP patients group. The remaining 111 patients in this group did not qualify for MFS analysis (*i.e.* 82 patients did not fulfill LNN or LNP study eligibility criteria and 29 patients already presented with metastasis at the time of diagnosis). Furthermore, a group of 226 patients received first-line chemotherapy for recurrent disease. Of these, 85 patients came from the LNN patients group and 100 came from the LNP patients group. The remaining 41 patients in this group did not qualify for MFS analysis (*i.e.* 27 patients did not fulfill LNN or LNP study eligibility criteria and 14 patients already presented with metastasis at the time of diagnosis).

Table 5.1 Association of *APOBEC3B* copy number status with clinicopathological variables in 1,756 primary breast cancers.

Variables	**Deleted**		**Balanced**		**Amplified**		***P*-value**
Total number	187		1260		309		
Age (in years)							0.54
⩽40	23	(12.3%)	160	(12.7%)	48	(15.5%)	
41-55	82	(43.9%)	498	(39.5%)	116	(37.5%)	
56-70	62	(33.2%)	412	(32.7%)	99	(32.0%)	
>70	20	(10.7%)	190	(15.1%)	46	(14.9%)	
Menopausal status							0.31
Premenopausal	89	(47.6%)	554	(44.0%)	149	(48.2%)	
Postmenopausal	98	(52.4%)	706	(56.0%)	160	(51.8%)	
Tumor size							1.00
pT1	71	(38.0%)	469	(37.2%)	115	(37.2%)	
pT2 + Unknown	96	(51.3%)	661	(52.5%)	163	(52.8%)	
pT3 + pT4	20	(10.7%)	130	(10.3%)	31	(10.0%)	
Nodal status							0.041
N0	117	(63.6%)	793	(63.5%)	173	(56.5%)	
N1-3	23	(12.5%)	212	(17.0%)	67	(21.9%)	
N>3	44	(23.9%)	244	(19.5%)	66	(21.6%)	
Tumor grade							0.052
Good/Moderate	34	(23.8%)	193	(21.9%)	66	(29.6%)	
Poor	109	(76.2%)	689	(78.1%)	157	(70.4%)	
Tumor histology							0.29
IDC	129	(83.2%)	833	(80.6%)	196	(78.4%)	
ILC	10	(6.5%)	119	(11.5%)	30	(12.0%)	
Other	16	(10.3%)	82	(7.9%)	24	(9.6%)	
ER status							0.24
Positive	130	(70.3%)	951	(75.8%)	226	(73.9%)	
Negative	55	(29.7%)	303	(24.2%)	80	(26.1%)	
PR status							0.99
Positive	114	(65.5%)	779	(65.8%)	187	(65.6%)	
Negative	60	(34.5%)	404	(34.2%)	98	(34.4%)	
HER2 status							0.59
Positive	18	(12.6%)	150	(15.7%)	39	(16.2%)	
Negative	125	(87.4%)	807	(84.3%)	201	(83.8%)	
Adjuvant systemic therapy							0.17
None	133	(73.5%)	856	(69.9%)	191	(63.0%)	
Chemotherapy	26	(14.4%)	219	(17.9%)	63	(20.8%)	
Hormonal therapy	22	(12.2%)	146	(11.9%)	47	(15.5%)	
Both	0	(0%)	4	(0.3%)	2	(0.7%)	

Note: For nodal status, tumor grade, tumor histology, ER, PR and HER2 status and adjuvant systemic therapy the number of patients do not add up to 1,756 because there were missing values for these variables. In addition, for the adjuvant systemic treatment variable: some patients were not eligible for adjuvant treatment because they were had distant metastasis at the time of primary tumor diagnosis.

Routine postsurgical follow-up and the definition of time to metastasis for LNN and LNP patients were as described previously[28]. The median follow-up for the 1,604 LNN and LNP patients included in the prognostic studies for the analysis of metastasis-free survival (MFS) was 114 months (range 10-354 months). Of these 1,604 patients, 686 (42.8%) had developed a distant metastasis and 691 (43.1%) of the 1,604 patients died during follow up. More specifically, 602 (37.5%) patients died after disease recurrence, whereas 89 (5.5%) patients died without evidence of disease recurrence at last follow-up. These 89 patients were censored at the date of last follow-up in the analysis of MFS.

Criteria for follow up and response to tamoxifen therapy were defined by standard International Union Against Cancer (Geneva, Switzerland) criteria of tumor response[29]. Complete and partial remission (together objective response) was observed in 11 and 48 patients, respectively, whereas 79 patients had progressive disease. From the patients with stable disease, 171 had no change for >6 months, whereas 20 patients had no change for $\leqslant$6 months. According to the advice of the European Organization for Research and Treatment of Cancer[30], we defined overall response as complete and partial remission including stable disease >6 months. As a result, 230 patients were classified as responders to tamoxifen and 99 patients showed no response to tamoxifen. The median follow up of patients after start of tamoxifen therapy was 49 months (range: 4-176 months). At the end of the follow up, 304 (92.4%) patients had developed tumor progression and were counted as events in the analysis of progression-free survival (PFS) and 264 patients had died.

Criteria for follow up and response to chemotherapy were similar to those for tamoxifen therapy with the exception that not all patients were chemotherapy-naive. In fact, 45 out of 226 patients had received adjuvant chemotherapy. Of those, 33 patients received cyclophosphamide/methotrexate/5-fluorouracil (CMF), 11 patients received anthracycline-based chemotherapy and 1 patient received both. Complete and partial remission was observed in 12 and 69 patients, respectively, whereas 66 patients had progressive disease. From the 77 patients with stable disease, 54 had no change for >6 courses, whereas 23 patients had no change for $\leqslant$6 courses. We defined overall response as complete and partial remission including stable disease >6 courses. As a result, 135 patients were classified as

responders to chemotherapy and 89 patients showed no response to chemotherapy. For two patients the type of response was ambiguous. The median follow up of patients after start of chemotherapy was 30 months (range: 4-153 months). Fifty-two patients received consolidation therapy after chemotherapy and PFS was right censored at two months after the start of consolidation therapy. At this time point, 138 patients had developed tumor progression and were counted as events in the analysis of PFS. At the end of follow up, 210 patients had died.

This study was approved by the Medical Ethical Committee of the Erasmus Medical Center Rotterdam, the Netherlands (MEC 02.953). As this is a retrospective study using remaining material from surgical resection of the patient's primary tumor, obtaining informed consent from the patient was not required provided patient records were anonymized and de-identified prior to analysis. Herewith, we adhered to the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.fmwv.nl). Results are reported in accordance with the REMARK criteria on clinical reporting[31].

Copy number analysis

Copy number analysis for *APOBEC3B* was performed on genomic DNA isolated from fresh-frozen primary tumor sections by qPCR on a Mx3000/3005P machine (Agilent Technologies, Santa Clara, CA). Briefly, genomic DNA was isolated from two to ten 30µm cryostat sections (5-20mg) with the NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany) according the protocol provided by the manufacturer. DNA quantity and quality was assessed by Nanodrop and the Quant-iT PicoGreen dsDNA HS Assay Kit (Thermo Scientific, Waltham, MA). Next, 0.5X Taqman Copy Number Assay for the *APOBEC3B* gene (Hs04504055_cn; Thermo Scientific), 0.5X Taqman Copy Number Reference Assay (*i.e.* for the *RNase P* gene; Thermo Scientific) and 0.5X ABsolute qPCR Mix, low ROX (Thermo Scientific) were added to 20ng of genomic DNA in a final volume of 17µl. Cycling conditions were: 1 cycle of 15 minutes at 95ºC and 45 cycles of 15 seconds at 92ºC and 1 minute at 60ºC. The MxPro qPCR software v4.10 (Agilent) was used to calculate the cycle threshold (Ct) values. Relative quantification analysis was performed within the CopyCaller

v2.0 software (Thermo Scientific). For this, the ΔCt value was calculated for each sample by subtracting the Ct value for the target gene (*i.e. APOBEC3B*) from the Ct value of the reference gene (*i.e. RNase P*). In the case of no Ct value for *APOBEC3B* after 45 cycles, while the *RNase P* gene was successfully amplified within 32 cycles for that sample, the ΔCt value could not be quantified and the sample was designated to have a two-copy deletion of the *APOBEC3B* gene. For samples where the ΔCt was quantified, ΔCt values were converted to calculated copy number values by the CopyCaller software as detailed in S5.1Fig. Then, samples with calculated copy numbers $\leqslant$0.2 were called as two-copy deletion of the *APOBEC3B* gene, samples with calculated copy numbers >0.20 and $\leqslant$1.41 as one-copy deletion of the *APOBEC3B* gene, samples with calculated copy numbers >1.41 and $\leqslant$3.44 as no copy number change or balanced and samples with calculated copy numbers >3.44 as amplified for the *APOBEC3B* gene. Furthermore, 325 samples were measured in duplicate distributed over the 96-well sample plates and each 96-well plate included genomic DNA from breast cancer cell line OCUB-F as a control since this cell line has a two-copy deletion of *APOBEC3B*.

Expression analysis

APOBEC3B mRNA expression analysis has been performed previously for 1,491 breast cancer patients[15]. Extraction of RNA, synthesis of cDNA, reverse transcriptase quantitative PCR and quantification of transcripts was also described before[32]. Out of the 1,756 patients for whom we performed *APOBEC3B* copy number analysis in the current study, *APOBEC3B* mRNA expression data was available for 1,132 patients.

Statistical analyses

Because the number of patients carrying a two-copy deletion of *APOBEC3B* was small (N=16, 0.91%), we grouped patients with one-copy and two-copy deletions together. A χ^2 or Fisher's exact test (when the expected frequency was $\leqslant$5 in any of the groups) was used to evaluate the association between the *APOBEC3B* copy number status (*i.e.* deleted, balanced

or amplified) and the clinicopathological variables. To assess the association between the *APOBEC3B* copy number status and the ORR to either first-line tamoxifen treatment or first-line chemotherapy, we used a logistic regression model to calculate ORs and their 95% CIs. For visualization purposes, we performed survival analysis by the Kaplan-Meier method. The difference between survival curves for patients with either a deletion, no copy number change, or an amplification was calculated using the 3-sample logrank test. In addition, univariate Cox proportional hazards regression models including continuous calculated copy number values as covariate were performed to assess the association between *APOBEC3B* copy number and survival times (*i.e.* MFS in the adjuvant setting and PFS in the advanced setting). Finally, the correlation between *APOBEC3B* copy number and *APOBEC3B* expression was evaluated using Spearman's rank correlation. All *P*-values were two-sided and *P*-values <0.05 were considered to be statistically significant. Analyses were performed using R, version 3.2.3, except for the power and sample size calculations shown in the Discussion. For these, we used the stpower cox tool in Stata version 13.1 (StataCorp, College Station, TX) assuming an alpha of 0.05, a beta of 0.2 (*i.e.* only in sample size calculations) and similar covariate standard deviations, allele frequencies, event probabilities and sample sizes (*i.e.* only in power calculations) as observed in the current study.

Results

We have performed *APOBEC3B* copy number analyses among 1,756 primary breast cancers and found no copy number change among 1,260 breast cancers. A two-copy deletion was identified in 16 (0.91%) breast cancers, whereas 171 (9.74%) breast cancers had a one-copy deletion. In addition, we detected an amplified *APOBEC3B* gene locus in 309 (17.26%) breast cancers. The minor allele frequency of the *APOBEC3B* deletion was 5.8% (203/3512) which is similar to the expected frequency in European population (*i.e*. 6.5%)[18].

Next, we evaluated the association between the *APOBEC3B* copy number status and the clinicopathological variables of the 1,756 breast cancer patients. We found that *APOBEC3B* copy number status was significantly associated with nodal status (*P*=0.041), but not with age, menopausal status, tumor size, tumor grade, tumor histology, ER, PR and *HER2* status, and

adjuvant systemic therapy (Table 5.1). Despite the significant association between *APOBEC3B* copy number status and nodal status, however, no meaningful trend was observed (Table 5.1).

To assess whether *APOBEC3B* copy numbers associate with clinical outcome in the adjuvant setting, we performed Kaplan-Meier survival analysis and Cox regression analysis in 1,604 LNN and LNP patients. No association between *APOBEC3B* copy number status and the length of MFS was found (*P*=0.14; Figure 5.2A). Moreover, calculated copy number values were also not associated with the length of MFS (HR=0.95, 95% CI=0.88-1.02, *P*=0.17; Table 5.2). Because all LNP patients were treated with adjuvant systemic therapy and we specifically wanted to evaluate the prognostic value of *APOBEC3B* copy number, we repeated these analyses in the cohort of 1,076 LNN patients that had not received any adjuvant systemic treatment. Again, we found no association between *APOBEC3B* copy number status and the length of MFS (*P*=0.84; Figure 5.2B), nor did we find an association between *APOBEC3B* calculated copy number values and the length of MFS (HR=1.00, 95% CI=0.90-1.11, *P*=0.96; Table 5.2). Also when we performed subgroup analysis in 769 ER-positive or 300 ER-negative untreated LNN patients (*i.e.* ER status was not available for 7 patients), *APOBEC3B* copy numbers did not appear to have any prognostic value (Figure 5.2C and 5.2D, Table 5.2).

To evaluate the predictive value of *APOBEC3B* copy numbers, we had two cohorts of breast cancer patients available that were treated with first-line therapy for recurrent disease. First, we evaluated whether the calculated *APOBEC3B* copy number values could predict the response to first-line tamoxifen therapy in a cohort of 329 hormone-naive breast cancer patients with ER-positive primary breast cancer. No significant association was observed between *APOBEC3B* copy number and the ORR for tamoxifen therapy (OR=0.88, 95% CI= 0.69-1.13, *P*=0.31; Table 5.3). Moreover, *APOBEC3B* copy number status was not associated with the length of PFS (*P*=0.25; Figure 5.2E), nor were calculated *APOBEC3B* copy number values associated with the length of PFS (HR=1.00, 95% CI=0.88-1.14, *P*=0.96; Table 5.4) in this cohort. Thus, *APOBEC3B* copy number is not a suitable biomarker to predict the type of response to tamoxifen therapy.

Table 5.2 Univariate Cox regression analysis to evaluate the association of calculated *APOBEC3B* copy number values with the length of MFS.

			Univariate analysis	
Study cohort	**N Patients**	**N Events**	**HR (95% CI)**	***P*-value**
LNN+LNP				
All patients	1,604	686	0.95 (0.88-1.02)	0.17
LNN				
All patients	1,076	366	1.00 (0.90-1.11)	0.96
ER+ patients	769	263	0.93 (0.82-1.06)	0.29
ER- patients	300	101	1.13 (0.94-1.35)	0.20

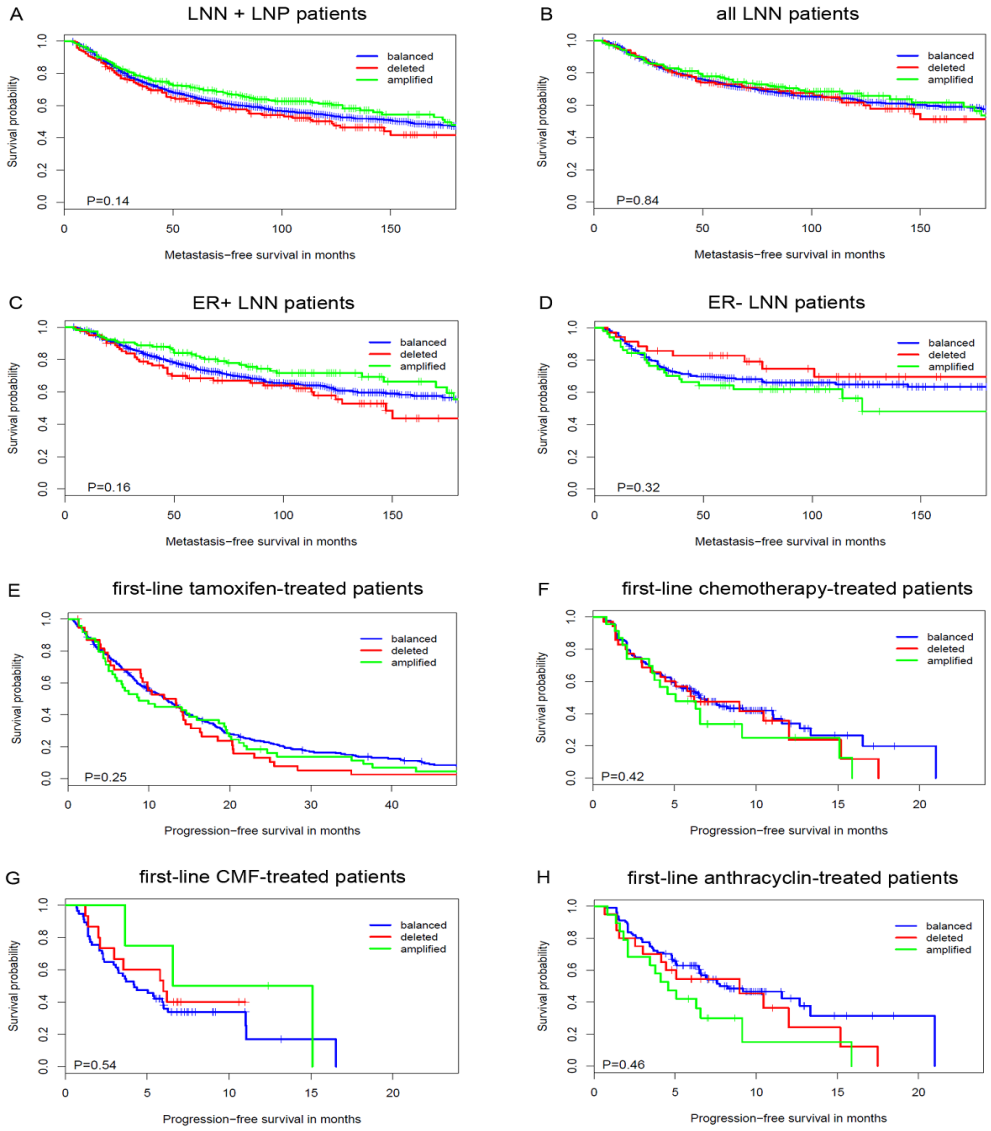
N, number of; HR, hazard ratio; CI, confidence interval; LNN, lymph node negative; LNP, lymph node positive.
Note: ER status was not available for 7 out of 1,076 patients.

Table 5.3 Univariate logistic regression analysis of the overall response rate in patients treated with first-line tamoxifen and in patients treated with first-line chemotherapy for recurrent disease.

		Univariate analysis	
Study cohort	**N Patients**	**OR (95% CI)**	***P*-value**
First-line tamoxifen			
All patients	329	0.88 (0.69-1.13)	0.31
First-line chemotherapy			
All patients	224	0.96 (0.71-1.31)	0.80
CMF-treated patients	75	1.25 (0.70-2.21)	0.45
Anthracyclin-treated patients	149	0.79 (0.53-1.16)	0.22

N, number of; OR, odds ratio; CI, confidence interval; CMF, cyclophosphamide/methotrexate/5- fluorouracil.
Note: the type of response was ambiguous for 2 patients.

Figure 5.2 Kaplan-Meier survival analysis as a function of *APOBEC3B* copy number status.



(A) In 1,604 patients of the lymph node negative (LNN) and lymph node positive (LNP) cohort combined. (B) In 1,076 LNN patients who did not receive any adjuvant systemic treatment. (C) In 769 ER-positive LNN patients who did not receive any adjuvant systemic treatment. (D) In 300 ER-negative LNN patients who did not receive any adjuvant systemic treatment. (E) In 329 ER-positive breast cancer patients who received first-line tamoxifen for recurrent disease. (F) In 226 breast cancer patients who received first-line chemotherapy for recurrent disease. (G) In 76 breast cancer patients who received first-line CMF-based chemotherapy for recurrent disease. (H) In 150 breast cancer patients who received first-line anthracycline based chemotherapy for recurrent disease. Differences between the survival curves were calculated with the 3-sample logrank test.

Table 5.4 Univariate Cox regression analysis to evaluate the association of calculated *APOBEC3B* copy number values with the length of PFS.

			Univariate analysis	
Study cohort	**N Patients**	**N Events**	**HR (95% CI)**	***P*-value**
First-line tamoxifen				
All patients	329	304	1.00 (0.88-1.14)	0.96
First-line chemotherapy				
All patients	226	138	1.06 (0.88-1.29)	0.53
CMF-treated patients	76	52	0.91 (0.65-1.28)	0.58
Anthracyclin-treated patients	150	86	1.19 (0.93-1.52)	0.17

N, number of; HR, hazard ratio; CI, confidence interval; CMF, cyclophosphamide/methotrexate/5-fluorouracil.

Note: The length of PFS was censored at two months after the start of consolidation therapy.

Next, we evaluated whether calculated *APOBEC3B* copy number values could predict the response to first-line chemotherapy in a cohort of 226 breast cancer patients. The calculated *APOBEC3B* copy number values were, however, not found to be associated with the ORR for chemotherapy (OR=0.96, 95% CI=0.71-1.31, *P*=0.80; Table 5.3). In addition, neither *APOBEC3B* copy number status, nor calculated *APOBEC3B* copy number values were associated with the length of PFS in these patients (*P*=0.42; Figure 5.2F and HR=1.06, 95% CI=0.88-1.29, *P*=0.53; Table 5.4, respectively). The lack of a significant association with the ORR to chemotherapy and the length of PFS was also observed when performing subgroup analysis by type of chemotherapy (*i.e.* CMF versus anthracyclines; Table 5.3, Figure 5.2G and 5.2H, Table 5.4). *APOBEC3B* copy number is thus also not a predictive biomarker for the type of response to chemotherapy.

In a previous study, we had analyzed *APOBEC3B* mRNA expression in 1,491 breast cancer patients and found that high *APOBEC3B* expression had prognostic value and was associated with poor outcome in untreated LNN patients with ER-positive breast cancer[15]. However, we did not find any association between *APOBEC3B* copy numbers and clinical outcome in the current study. As both *APOBEC3B* expression and *APOBEC3B* copy number

have been associated with a hypermutator phenotype[13,14,24], we attempted to clarify this discrepancy by evaluating the correlation between *APOBEC3B* copy number and *APOBEC3B* mRNA expression. Out of the 1,756 patients for whom we performed *APOBEC3B* copy number analysis in the current study, we had *APOBEC3B* mRNA expression data available for 1,132 patients. Interestingly, although a correlation among *APOBEC3B* copy number and mRNA expression was observed, the correlation coefficient was low (spearman's rho=0.26, P=2.2*10^{-16}). This suggests that other mechanisms exist that affect *APOBEC3B* mRNA expression besides *APOBEC3B* copy number.

Discussion

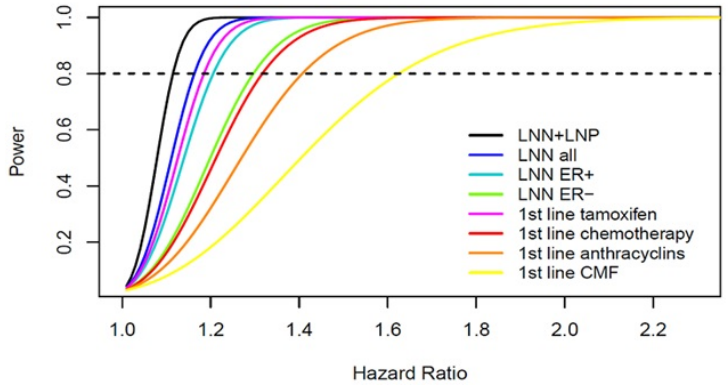
APOBEC3B mRNA expression is upregulated in multiple tumor types and this has been shown to correlate with an increased mutational load, particularly an increase in C>T transversions[13,14]. In line with these findings, increased expression of *APOBEC3B* mRNA was associated with a poor prognosis in ER-positive breast cancer[15]. At the same time, the 29.5 kb deletion polymorphism of *APOBEC3B* was found to be associated with the increased breast cancer risk in different populations[19-21], although these findings were not confirmed in a Swedish study[22]. Similar to *APOBEC3B* overexpression, the *APOBEC3B* deletion has been shown to correlate with an increased mutational load[24]. These findings may seem paradoxical, as loss of *APOBEC3B* should decrease the mutational load. However, the *APOBEC3B* deletion polymorphism not just deletes *APOBEC3B*. It also generates a novel fusion transcript (*i.e. APOBEC3A* under the control of the 3'UTR of *APOBEC3B*)[18]. Consequently, *APOBEC3A* mRNA was shown to become more stable, resulting in higher levels of APOBEC3A and, since APOBEC3A is a more efficient hypermutator than APOBEC3B, more severe DNA damage[23]. Thus, although the molecular mechanisms behind overexpression of *APOBEC3B* and the *APOBEC3B* deletion polymorphism are very different, they both result in a hypermutator phenotype. The hypothesis that the *APOBEC3B* deletion polymorphism thus may also be associated with clinical outcome is therefore plausible.

In a study by Gohler *et al*., the *APOBEC3B* deletion polymorphism was however, not

associated with breast cancer specific survival in 782 breast cancer cases[22]. Moreover, Cescon *et al*. showed that the deletion was not associated with recurrence after treatment for early breast cancer in METABRIC[16]. Because both studies also included patients that were treated, no distinction could be made between pure prognosis and therapy response. In the current study, we examined separately the prognostic and predictive value of *APOBEC3B* copy number in a total of 1,756 breast cancer patients. No association between *APOBEC3B* copy numbers and the length of MFS was found among 1,076 LNN patients who had not received adjuvant systemic treatment (Figure 5.2B, Table 5.2). In addition, an association with the length of MFS was also not observed among ER-positive or ER-negative breast cancer patients. These results imply that *APOBEC3B* copy number is not a prognostic biomarker for breast cancer. The association between *APOBEC3B* copy number and the response to treatment in breast cancer patients that received either first-line tamoxifen or chemotherapy for recurrent disease was also evaluated. However, we found no association between the type of response for either tamoxifen or chemotherapy and *APOBEC3B* copy number. Thus, besides not having any prognostic value, *APOBEC3B* copy numbers also do not appear to have a predictive value for breast cancer patients.

The copy number analyses for the adjuvant setting, except for the ER-negative LNN analysis, had sufficient power to conclude that there is no or only a marginal prognostic effect of *APOBEC3B* copy numbers in breast cancer patients (Figure 5.3). For the copy number analysis in the advanced setting involving patients treated with first-line tamoxifen, we could draw that same conclusion (Figure 5.3). However, for the copy number analysis involving patients treated with first-line chemotherapy, especially in the subgroup analyses by type of chemotherapy, we cannot exclude a modest effect of *APOBEC3B* copy numbers on the length of PFS (Figure 5.3). Therefore, replication of our results observed in the advanced chemotherapy setting in a larger sample size or population with a higher prevalence of the *APOBEC3B* deletion polymorphism could be needed.

Figure 5.3 Power as a function of the hazard ratio for *APOBEC3B* copy number in each of the analyzed subgroups.



The dashed horizontal line crosses the curve of each subgroup at the minimal hazard ratio for which we had 80% power in our *APOBEC3B* copy number analyses.

Another note is that the assay we used to determine *APOBEC3B* copy numbers does not discriminate between germline and tumor-specific (*i.e.* somatic) *APOBEC3B* deletion. This has no consequence for the accuracy of the germline copy number determination except for patients who carry a one-copy deletion and have strong amplification of the remaining *APOBEC3B* locus. Breast cancer patients who do not carry the germline deletion and have either a tumor-specific homozygous deletion or a heterozygous deletion of *APOBEC3B* in combination with a high tumor cell percentage may also be misclassified. In this respect, the observed amplification of the *APOBEC3B* allele is rather a somatic event than the result of an alteration in the germline. To date, there has been no report of the *APOBEC3B* locus being amplified in the germline. Interestingly, in a publically available SNP array data set of 344 breast cancers (accession number EGAS00001001178[33]), *APOBEC3A* copy number status was equal to *APOBEC3B* copy number status in all tumors, suggesting that *APOBEC3A* is always amplified or deleted simultaneously with *APOBEC3B* during breast tumorigenesis.

In our Kaplan-Meier survival analysis we grouped patients with one-copy deletion and two copy deletions together to provide a visualization of the estimated survival curves of patients with a deleted, balanced or amplified *APOBEC3B* gene. Unfortunately, the

population frequency of the 29.5 kb deletion polymorphism of *APOBEC3B* was too low in this Dutch cohort to analyze patients who carry a two-copy deletion separately. To illustrate this with a power calculation: we would need a sample size of 4,024 or 11,759 LNN patients to ensure a minimally detectable hazard ratio of 2 or 1.5, respectively, using a power of 80% and an alpha of 0.05. For this reason, evaluation of the clinical value of the two-copy deletion should preferably be done in breast cancer patients from the Asian or American population (*i.e.* population frequency of 36.9% and 57.7%, respectively)[18].

The observation that the *APOBEC3B* deletion polymorphism does not seem to have any impact on the clinical outcome for breast cancer patients, whereas increased expression of *APOBEC3B* mRNA does, strengthens the evidence that there are two different molecular mechanisms in place. Moreover, elevated levels of *APOBEC3B* mRNA have been shown to associate with cellular proliferation, whereas the *APOBEC3B* deletion polymorphism associated with activation of immune-related genes[16]. Interestingly, increased lymphocytic infiltration has been associated with a favorable outcome in some, but not all subtypes of breast cancer[34,35]. This is in contrast to what has been observed for increased *APOBEC3B* mRNA expression, which associated with increased proliferation and poor outcome[15].

Although we observed a correlation between *APOBEC3B* copy number and mRNA expression, this correlation was rather weak. As a consequence, the prognostic effect of increased levels of *APOBEC3B* mRNA is not necessarily caused by increased *APOBEC3B* copy numbers. Other mechanisms may exist that elevate *APOBEC3B* mRNA levels. Recently, it was shown that APOBEC3B interacts with ER to bind to ER binding sites, where it generates C to U transitions. Furthermore, the presence of APOBEC3B was necessary for histone modification, but also in order to recruit chromatin remodeling factors to ER binding sites[36]. This may very well explain why elevated levels of *APOBEC3B* mRNA are only prognostic in ER-positive breast cancer patients. Although the *APOBEC3B* deletion polymorphism does not have any prognostic or predictive value, it does appear to contribute to breast tumorigenesis by conferring an increased risk to develop breast cancer. However, still little is known regarding the role of APOBEC3A in breast cancer. Therefore, more studies should be done in order to investigate the precise effect of the *APOBEC3B* deletion

polymorphism and *APOBEC3A-B* hybrid transcript resulting from this deletion. This will provide more insight into APOBEC-mediated hypermutation.

Supporting Information

S5.1 Fig. Histogram of the ΔCt values for all 1,756 measured DNA samples and cell line OCUB-F.

S5.1 Table. Clinical data of all 1,756 breast cancer patients included in the study

References

1 Perou, C. M., Sorlie, T., Eisen, M. B. *et al.* Molecular portraits of human breast tumours. *Nature* 406, 747-752 (2000).

2 Sorlie, T., Perou, C. M., Tibshirani, R. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 98, 10869-10874 (2001).

3 Sorlie, T., Tibshirani, R., Parker, J. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100, 8418-8423 (2003).

4 Prat, A., Pineda, E., Adamo, B. *et al.* Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* (2015).

5 Smid, M., Wang, Y., Zhang, Y. *et al.* Subtypes of breast cancer show preferential site of relapse. *Cancer Res* 68, 3108-3114 (2008).

6 Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979-993 (2012).

7 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C. *et al.* Signatures of mutational processes in human cancer. *Nature* 500, 415-421 (2013).

8 Taylor, B. J., Nik-Zainal, S., Wu, Y. L. *et al.* DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *Elife* 2, e00534 (2013).

9 Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. *et al.* Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22, 367-377 (2005).

10 Harris, R. S. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res* 17, 8 (2015).

11 Bogerd, H. P., Wiegand, H. L., Hulme, A. E. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* 103, 8780-8785 (2006).

12 Muckenfuss, H., Hamdorf, M., Held, U. *et al.* APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem* 281, 22161-22172 (2006).

13 Burns, M. B., Lackey, L., Carpenter, M. A. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366-370 (2013).

14 Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet* 45, 977-983 (2013).

15 Sieuwerts, A. M., Willis, S., Burns, M. B. *et al.* Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm Cancer* 5, 405-413 (2014).

16 Cescon, D. W., Haibe-Kains, B. & Mak, T. W. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc Natl Acad Sci U S A* 112, 2841-2846 (2015).

17 Tsuboi, M., Yamane, A., Horiguchi, J. *et al.* APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. *Breast Cancer* (2015).

18 Kidd, J. M., Newman, T. L., Tuzun, E. *et al.* Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet* 3, e63 (2007).

19 Long, J., Delahanty, R. J., Li, G. *et al.* A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst* 105, 573-579 (2013).

20 Xuan, D., Li, G., Cai, Q. *et al.* APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* 34, 2240-2243

(2013).

21 Rezaei, M., Hashemi, M., Hashemi, S. M. *et al.* APOBEC3 Deletion is Associated with Breast Cancer Risk in a Sample of Southeast Iranian Population. *Int J Mol Cell Med* 4, 103-108 (2015).

22 Gohler, S., Da Silva Filho, M. I., Johansson, R. *et al.* Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population. *J Cancer Res Clin Oncol* (2015).

23 Caval, V., Suspene, R., Shapira, M. *et al.* A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. *Nat Commun* 5, 5129 (2014).

24 Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B. *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 46, 487-491 (2014).

25 Foekens, J. A., Portengen, H., Look, M. P. *et al.* Relationship of PS2 with response to tamoxifen therapy in patients with recurrent breast cancer. *Br J Cancer* 70, 1217-1223 (1994).

26 Foekens, J. A., Portengen, H., van Putten, W. L. *et al.* Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols. *Cancer Res* 49, 5823-5828 (1989).

27 van Agthoven, T., Sieuwerts, A. M., Meijer-van Gelder, M. E. *et al.* Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J Clin Oncol* 27, 542-549 (2009).

28 Wang, Y., Klijn, J. G., Zhang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679 (2005).

29 Hayward, J. L., Carbone, P. P., Heuson, J. C. *et al.* 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. *Cancer* 39, 1289-1294

(1977).

30 European Organization for Research and Treatment of Cancer & Breast Cancer Cooperative Group. Manual for clinical research and treatment in breast cancer. *Excerpta Medica. Almere, the Netherlands*, 116-117 (2000).

31 McShane, L. M., Altman, D. G., W., S. *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 97, 1180-1184 (2005).

32 Sieuwerts, A. M., Meijer-van Gelder, M. E., Timmermans, M. *et al.* 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 (2005).

33 Nik-Zainal, S., Davies, H., Staaf, J. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47-54 (2016).

34 Loi, S., Sirtaine, N., Piette, F. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol* 31, 860-867 (2013).

35 Loi, S., Michiels, S., Salgado, R. *et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol* 25, 1544-1550 (2014).

36 Periyasamy, M., Patel, H., Lai, C. F. *et al.* APOBEC3B-Mediated Cytidine Deamination Is Required for Estrogen Receptor Action in Breast Cancer. *Cell Rep* 13, 108-121 (2015).

Chapter 6

GATA3 mRNA expression, but not mutation, associates with longer progression-free survival in ER-positive breast cancer patients treated with first-line tamoxifen for recurrent disease

Jingjing Liu, Wendy J.C. Prager-van der Smissen, Maxime P. Look, Anieta M. Sieuwerts, Marcel Smid, Marion E. Meijer-van Gelder, John A. Foekens, Antoinette Hollestelle, John W.M. Martens.

Cancer Lett 376, 104-109 (2016).

Abstract

In breast cancer, *GATA3* mutations have been associated with a favorable prognosis and the response to neoadjuvant aromatase inhibitor treatment. Therefore, we investigated whether *GATA3* mutations predict the outcome of tamoxifen treatment in the advanced setting. In a retrospective study consisting of 235 hormone-naive patients with ER-positive breast cancer who received tamoxifen as first-line treatment for recurrent disease, *GATA3* mutations (in 14.0% of patients) did not significantly associate with either the overall response rate (ORR) or with the length of progression-free survival (PFS) after start of tamoxifen therapy. Interestingly, among 148 patients for whom both mutation and mRNA expression data was available, *GATA3* mutations associated with an increased expression of *GATA3*. However, only 23.7% of *GATA3* high tumors had a mutation. Evaluation of the clinical significance of *GATA3* mRNA revealed that it was associated with prolonged PFS, but not with the ORR, also in multivariate analysis. Thus, *GATA3* mRNA expression, but not *GATA3* mutation, is an independent predictor of prolonged PFS in ER-positive breast cancer patients who received first-line tamoxifen for recurrent disease. Besides *GATA3* mutation other mechanisms must exist that underlie increased *GATA3* levels.

Introduction

Breast cancer is one of the most frequently diagnosed cancers in Western women. About 70% of all diagnosed breast cancers are estrogen receptor α (ER) positive. ER-positive breast cancers are well-differentiated and have a better outcome compared to other subtypes[1,2]. In this respect, tamoxifen is a frequently used and effective drug for patients diagnosed with ER-positive disease. However, half of ER-positive patients who receive tamoxifen as first-line therapy for recurrent disease do not respond to the treatment, due to intrinsic resistance, while the other half initially responding patients become resistant during treatment[3]. To better understand the mechanism involved in this intrinsic and acquired resistance and to be able to predict which patients are likely to respond to tamoxifen, the identification of novel markers predicting the efficacy of tamoxifen treatment is highly needed.

GATA3 belongs to a family of zinc-finger transcription factors and is involved in embryogenesis and the differentiation of a variety of human tissues, including kidney, skin, breast and the central nervous system[4-8]. Both in the normal mammary gland and breast cancer tissue, GATA3 and ER expression are highly correlated[6,9]. In fact, GATA3 is expressed in the normal luminal epithelial cells where it maintains luminal cell differentiation[7], whereas in breast cancer *GATA3* is highly expressed in the luminal subtype, regulating differentiation and suppressing dissemination[7,10,11]. Furthermore, GATA3 is an integral component of the ER pathway as it regulates the pioneer factor FOXA1 and mediates ER binding by shaping enhancer accessibility[7,12]. Consequently, a large overlap exists between co-expressed genes for ER and GATA3, including many well-known ER pathway genes[13]. Since the expression of ER has important implications for both prognosis and treatment of breast cancers, several studies have assessed the association of GATA3 with clinical outcome. High GATA3 protein expression was shown to be associated with a lower grade, smaller tumor size and increased ER and PR expression[14-18]. In line with these findings, some, but not all, studies have shown that both *GATA3* mRNA and GATA3 protein expression are independent prognostic markers, where high levels of GATA3 associate with a longer disease-free and overall survival in breast cancer patients[14,15,18-21]. Furthermore, in a small study of Parikh *et al*. high levels of GATA3 protein were predictive of hormone

responsiveness in ER-positive breast cancer patients[22]. In the neoadjuvant setting both *GATA3* mRNA and GATA3 protein expression were shown to be predictive of a favorable response to chemotherapy[17,23].

The human *GATA3* gene is a highly conserved gene located at 10p14-15 and consists of six exons which encode a protein of 444 residues[24]. Germline mutations of *GATA3* cause a rare and complex disease of hypoparathyroidism, sensorineural deafness and renal insufficiency (HDR syndrome)[25]. In breast cancer, *GATA3* is one of the most frequently mutated genes[26-29] and sporadic heterozygous *GATA3* mutations have been identified in approximately 5-20% of ER-positive breast cancers[30]. These mutations mostly cluster in the vicinity of the second zinc finger of GATA3[31] and are virtually absent among ER-negative breast cancers. Interestingly, *GATA3* mutations were correlated with improved disease-free and overall survival in breast cancer patients overall, but also in ER-positive breast cancer patients who received adjuvant endocrine therapy[32]. Furthermore, mutations in *GATA3* were also shown to be correlated with response to neoadjuvant aromatase inhibition treatment[33]. This suggests that *GATA3* mutation may be a determinant of the response to hormonal treatment.

To investigate this hypothesis, we analyzed exons 5 and 6 of the *GATA3* gene for mutations in 235 ER-positive primary breast cancers and evaluated the association of the identified mutations with the ORR and PFS of first-line tamoxifen therapy given for recurrent disease, as well as with *GATA3* mRNA expression levels.

Materials and methods

Study population

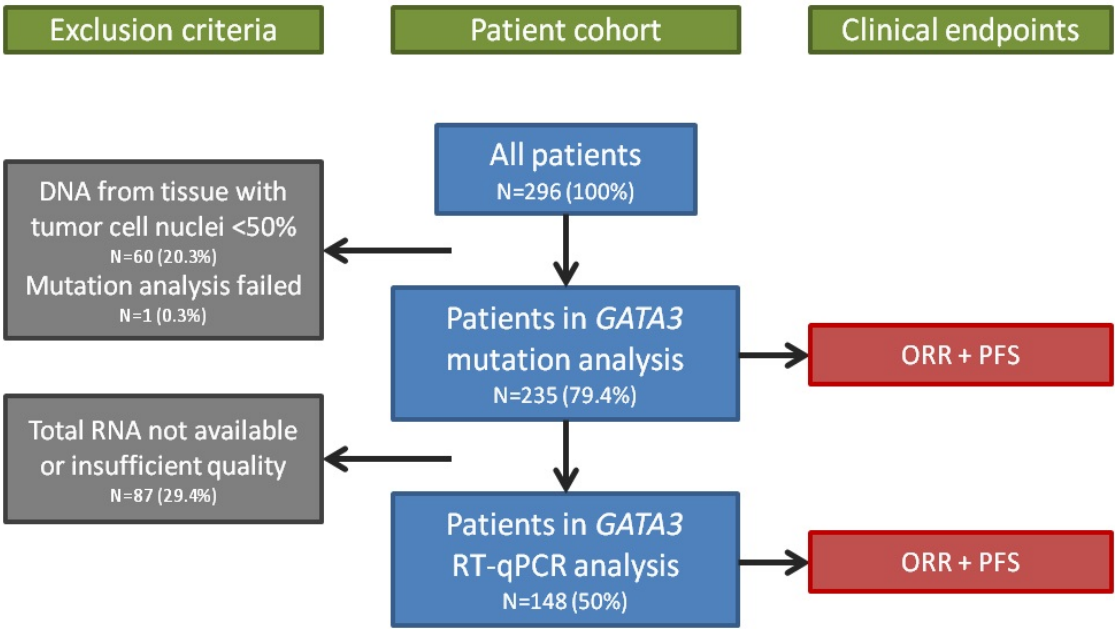
This retrospective study included 235 female breast cancer patients (Figure 6.1) and was approved by the Medical Ethical Committee of the Erasmus Medical Center Rotterdam, the Netherlands (MEC 02.953). In this study we adhered to the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.fmwv.nl) and results are reported in accordance with the REMARK criteria on clinical reporting[34]. All

patients were diagnosed between 1979 and 1996 with measurable breast cancer disease, underwent primary surgery and were treated with tamoxifen as first-line treatment that was given for recurrent disease. Primary tumors were ER-positive and a minimum of 100mg of fresh frozen tissue was required for downstream DNA and RNA extraction[35]. ER and/or PR positivity was defined by ⩾10fmol/mg cytosolic protein[36,37] and *ERBB2* overexpression was defined by a reverse transcriptase quantitative PCR (RT-qPCR) expression level ⩾18[38]. The patients did not receive neo-adjuvant therapy or adjuvant hormonal treatment, did not experience previous other cancers and did not show subjective or objective toxicity[35]. There were 296 patients that fulfilled these criteria and from whom detailed clinical follow up and primary tumor DNA was available. However, 60 patients were excluded as the percentage of tumor cell nuclei was below 50%, precluding reliable mutation analysis. Furthermore, mutation analysis failed in 1 patient, totaling to n=235 included in the present study. From these, 84 patients underwent breast- conserving lumpectomy and 151 underwent modified mastectomy. In addition, 17 patients received adjuvant anthracycline-containing chemo-therapy and 14 received adjuvant chemotherapy without anthracyclines. There were 209 M0 patients and 26 M1 patients. The median age at the time of the primary surgery was 57 years, while the median age at the start of first-line treatment was 61 years. Criteria for follow up and response to tamoxifen therapy were defined by standard International Union Against Cancer criteria of tumor response[39]. Complete and partial remission (together objective response) was observed in 4 and 34 patients, respectively, whereas 52 patients had progressive disease. From the patients with stable disease, 132 had no change for longer than 6 months, whereas 13 patients had no change ⩽6 months. According to the advice of the European Organization for Research and Treatment of Cancer[40], we defined overall response as complete and partial remission including stable disease >6 months. As a result, 170 patients were classified as responders to tamoxifen and 65 patients showed no response to tamoxifen. The median follow up of patients after start of tamoxifen therapy was 49 months (range: 4-208 months). At the end of the follow up, 224 patients had developed tumor progression and 196 patients had died.

From 148 of the 235 patients we had total RNA of sufficient quality from the primary

tumor available (*i.e.* at an input of 10ng total RNA amplifiable for 3 reference genes within 25 cycles) in order to perform *GATA3* mRNA expression analysis by RT-qPCR. The clinicopathological variables of the patients are shown in Table 6.1.

Figure 6.1 Study design and patient subsets analyzed for *GATA3* mutation status and *GATA3* mRNA expression.



The "All patients (N=296)" box represents all hormone naive patients diagnosed with recurrent breast cancer between 1979 and 1996 and were treated with first-line tamoxifen (details provided in the Materials and methods section). For the *GATA3* mutation analysis 60 patients were excluded whose tumor cell nuclei percentage was below 50% and mutation analysis failed in one patient. From these 235 patients, total RNA of sufficient quality was available for 148 patients for *GATA3* RT-qPCR analysis. Clinical endpoints were the overall response rate (ORR) and progression-free survival (PFS).

Mutation analysis

Genomic DNA previously extracted from the fresh frozen primary breast tumor of 235 patients[41] and quantified by Picogreen was used at an input of 20ng to amplify *GATA3* exon 5 and 6 sequences. Subsequently, PCR amplicons were subjected to Sanger sequencing analysis on an ABI3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA). All mutations were confirmed by Sanger sequencing using an independently amplified template. For the splice acceptor site mutations we performed an exonic reverse transcriptase PCR (RT-PCR) on a RNA template instead of PCR on a DNA template. We reported *GATA3* mutations and predicted protein changes according the HGVS recommendations for the description of sequence variants[42]. PCR and sequencing primer sequences are available in Table S6.1A and B.

Expression analysis

Total RNA was extracted and cDNA was synthesized previously from the fresh frozen primary breast tumor of 148 patients as described before[35]. qPCR for *GATA3* was performed in a Mx3000P^{TM} Real-Time PCR System (Agilent, Amsterdam, the Netherlands) using SensiFast Probe Lo-Rox master mix (GC biotech, Alphen aan den Rijn, the Netherlands) and a Taqman Gene expression Assay kit from Applied Biosystems (Hs00231122_m1; spanning exon 2 to 3; Nieuwerkerk aan den IJssel, the Netherlands) with 40 rounds of amplification as recommended by the manufacturer. In addition to a negative control (*i.e.* genomic DNA), we also included a standard curve of a serially diluted cDNA sample consisting of pooled breast cancer cDNA samples in each PCR plate. The latter was done to ensure that the PCR efficiency between plates was comparable and to normalize the data obtained from different plates and experiments. *GATA3* mRNA expression levels for the samples were determined relative to the average Cq value of our reference gene set consisting of hydroxymethylbilane synthase (*HMBS*), hypoxanthine phosphoribosyltransferase 1 (*HPRT1*) and TATA-box binding protein (*TBP*) and quantified as follows: *GATA3* expression = $2^{\text{Cq reference gene set} - \text{Cq } GATA3 \text{ gene}}$[35]. PCR primer sequences for the reference genes are available in Table S6.1C.

Table 6.1 Association of *GATA3* mutation status and *GATA3* gene expression levels with clinicopathological variables in 235 ER-positive primary breast cancers.

	***GATA3* mutation status**			***GATA3* expression level**		
Variable	**Number of wild-type patients**	**Number of mutant patients**	***P*-value**	**Number of patients**	**Median *GATA3* expression[a]**	***P*-value**
Total number	202	33		148	1.224	
Menopausal status[b]			0.22			0.25
premenopausal	44	11		33	0.891	
postmenopausal	157	22		115	1.238	
Tumor grade			0.68			0.24
good/moderate	26	5		17	1.723	
poor	116	15		86	1.010	
unknown	60	13		45	1.242	
Tumor size			0.057			0.41
pT1	54	4		39	1.106	
pT2 + unknown	120	27		93	1.276	
pT3 + pT4	28	2		16	2.053	
Nodal status			0.36			0.37
N0	83	18		71	1.287	
N1-3	43	7		30	1.172	
N>3	63	7		37	0.767	
unknown	13	1		10	0.778	
Dominant site of relapse			0.25			0.019
soft	24	1		13	0.276	
bone	107	17		83	1.463	
visceral	71	15		52	0.872	
Disease-free interval (m)			0.53			0.42
⩽12	53	8		39	1.225	
13-36	78	16		64	1.144	
>36	71	9		45	1.340	
PR protein status			0.28			0.0015
Positive	155	22		116	1.082	
Negative	46	11		32	1.646	
Unknown	1	0				
ERBB2 mRNA status			0.54			0.084
Positive	20	2		18	0.807	
Negative	149	29		130	1.259	
Unknown	33	2				

[a] Log2-transformed *GATA3* gene expression levels.

[b] At the start time of first-line tamoxifen treatment; m, months.

Statistical analyses

A $\chi 2$ or a Fisher's exact test (when the expected frequency $\leqslant 5$ in any of the groups) was used to evaluate the relation between *GATA3* mutation status and the clinicopathological variables. The relation between *GATA3* mRNA expression levels and the clinicopathological variables was evaluated using either the two-sample Wilcoxon rank-sum test (for 2 categories) or the Kruskal-Wallis equality-of- populations rank test (for 3 categories). The association with tamoxifen response was analyzed with a logistic regression model to calculate odds ratios (ORs) and their 95% confidence intervals (95% CIs). PFS analysis was performed by the Kaplan-Meier method for visualization purposes and differences between survival curves were calculated by the Peto & Peto modification of the Gehan-Wilcoxon test (which puts more weight on the earlier events) for *GATA3* mutation status and the log-rank test for *GATA3* gene expression. Univariate and multivariate Cox proportional hazards regression models were applied to calculate the hazards ratios (HRs) and 95% CIs in the analysis for the PFS. Log2-transformed expression values for *GATA3* mRNA, ER and PR protein and *ERBB2* mRNA were used in logistic and Cox regression analysis. All *P*-values were two-sided and *P*-values smaller than 0.05 were considered to be statistically significant. Analyses were performed using R, version 3.2.3.

Results

Since at the start of this study all of the *GATA3* mutations reported so far clustered in exons 5 and 6 of the *GATA3* gene, which encode the highly conserved second zinc finger required for DNA binding, we limited sequence analysis to these two exons. In total, we identified at least one *GATA3* sequence variant in 54 out of the 235 primary tumors of patients with ER-positive recurrent breast cancer. A silent mutation in exon 5 (rs11567941; c.1257G>A; p.T122T; minor allele frequency (MAF)=0.02) was identified in 24 tumors, however, we did not consider this mutation to be pathogenic. Furthermore, we identified a frameshift insertion in 22 tumors, a frameshift deletion in five tumors and a splice site deletion in six tumors (Table 6.2). These mutations predicted prematurely truncated proteins in 14 tumors and proteins with

a longer C-terminal tail in 19 tumors. In total, we identified 33 *GATA3* mutations that we considered to be pathogenic in 33 (14.0%) out of 235 ER-positive primary breast tumors.

Next, we evaluated the association between *GATA3* mutation status and the clinicopathological variables (Table 6.1), the ORR (Table S6.2) and the length of PFS after start of tamoxifen treatment (Table S6.2 and Figure 6.2A). We found no relation between *GATA3* mutation and any of the clinicopathological variables (Table 6.1). Furthermore, *GATA3* mutations did not significantly associate with the ORR for tamoxifen therapy in univariate logistic regression analysis (70.8% vs. 81.8%; OR=1.86, 95% CI=0.73-4.73, *P*= 0.19; Table S6.2) or with the length of PFS in Kaplan-Meier (*P*=0.80; Figure 6.2A) and Cox regression analysis (HR=0.95, 95% CI=0.65-1.40, *P*=0.81; Table S6.2). Also subsetting by the type of mutation (*i.e.* mutations predicted to truncate versus elongate the protein) did not yield any significant differences. Although the number of patients were small, survival curves appeared very similar. However, logistic regression analysis did show that the traditional predictive factor disease-free interval was associated with the efficacy of tamoxifen therapy (Table S6.2). Similarly, the traditional predictive factors dominant site of relapse, disease-free interval and the level of PR protein expression were found to be associated with PFS (Table S6.2). These results implied that *GATA3* mutation status is not a significant predictor for the outcome of tamoxifen therapy in patients with recurrent disease.

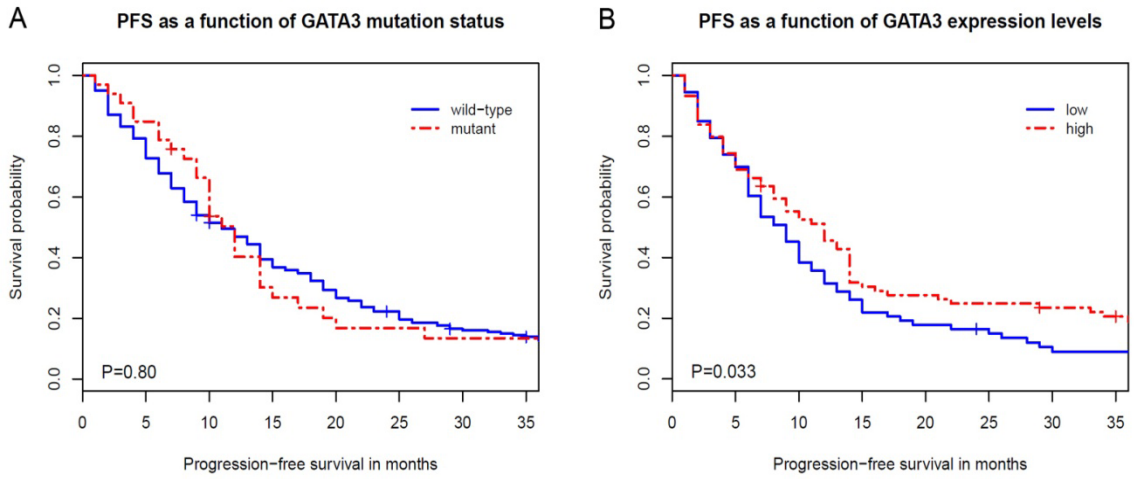
Out of the 235 tumors for which we performed *GATA3* mutation analysis, we were able to perform *GATA3* mRNA expression analysis by RT-qPCR for 148 tumors. In 25 out of these 148 tumors we identified a *GATA3* mutation and 123 tumors were wild-type. Interestingly, *GATA3* expression levels were higher among mutant *GATA3* tumors (*i.e.* irrespective of the predicted effect of the mutation) than wild-type *GATA3* tumors (*P*=0.0019). Eighteen tumors (72.0%) with *GATA3* mutations had high *GATA3* expression levels (*i.e.* above the median) while only seven tumors (28.0%) with *GATA3* mutations had low *GATA3* expression levels (*i.e.* below the median). However, out of the 76 *GATA3* high expressing tumors, only 18 (23.7%) had a mutation in the *GATA3* gene. Because the high levels of *GATA3* mRNA were only partially explained by a mutation in *GATA3* itself, we hypothesized that *GATA3* expression instead of mutation might be associated with the outcome of tamoxifen treatment.

Table 6.2 Identified *GATA3* mutations among 235 ER-positive primary breast cancers.

Location	Nucleotide change	Predicted protein change	Number of patients
Exon 5	c.925-3_925-2delCA	p.S309Pfs*45	6
Exon 5	c.961_962delTG	p.C321Sfs*31	1
Exon 5	c.983_984insC	p.W329Lfs*25	1
Exon 5	c.1002_1003insG	p.G335Gfs*18	1
Exon 5	c.1003delG	p.G335Gfs*20	1
Exon 5	c.1007_1008insC	p.V341Vfs*15	1
Exon 5	c.1021_1022insC	p.A341Afs*11	1
Exon 5	c.1033_1034insAC	p.Y345Yfs*11	1
Exon 5	c.1035_1036insT	p.Y346Lfs*7	1
Exon 6	c.1195_1196delAG	p.R399Tfs*108	1
Exon 6	c.1202_1203insG	p.S402Vfs*106	1
Exon 6	c.1202_1203insGTCC	p.S403Vfs*106	1
Exon 6	c.1206_1207insT	p.S403Ffs*105	2
Exon 6	c.1207_1208 insC	p.L404Pfs*103	1
Exon 6	c.1222_1223insC	p.P409Afs*99	2
Exon 6	c.1223_1224insT	p.P409Sfs*100	1
Exon 6	c.1223_1224insG	p.P409Afs*99	1
Exon 6	c.1257_1258insC	p.T421Hfs*87	1
Exon 6	c.1263_1282del20	p.M423Vfs*78	1
Exon 6	c.1271_1272insC	p.P425Afs*82	2
Exon 6	c.1277_1278insA	p.S427Ifs*81	2
Exon 6	c.1304_1305insC	p.S437Lfs*71	2
Exon 6	c.1305delC	p. S437Pfs*39	1

Nomenclature for the identified nucleotide changes and predicted protein changes is according the HGVS recommendations for the description of sequence variants[42]. * Stop codon.

Figure 6.2 Kaplan-Meier analysis of progression-free survival.



(A) according to *GATA3* mutation status for 235 ER-positive breast cancer patients who received first-line tamoxifen therapy for recurrent disease. The difference between the survival curves was calculated using the Peto & Peto modification of the Gehan–Wilcoxon test. (B) Dichotomized at median *GATA3* expression level for 148 ER-positive breast cancer patients who received first-line tamoxifen therapy for recurrent disease. The difference between the survival curves was calculated using the log-rank test.

To evaluate this, we made use of the *GATA3* mRNA expression data for all 148 primary breast tumors from recurrent breast cancer patients that we had generated by RT-qPCR. We found that *GATA3* expression was associated with dominant site of relapse and PR protein status, but not with menopausal status, tumor grade, tumor size, nodal status, disease-free interval or *ERBB2* mRNA status (Table 6.1). In univariate logistic regression analysis, we found no association of *GATA3* expression level with the ORR for tamoxifen (64.9% versus 66.2%; OR=1.12, 95% CI=0.90-1.41, *P*=0.31; Table 6.3). Additionally, menopausal status, dominant site of relapse and ER protein, PR protein and *ERBB2* mRNA expression levels were also not associated with the ORR for tamoxifen, in contrast to disease-free interval (Table 6.3). *GATA3* expression was, however, associated with the length of PFS, as it was prolonged for patients with tumors with high *GATA3* mRNA levels compared to those with low levels (*P*=0.033; Figure 6.2B). Concordantly, in univariate Cox regression analysis, high *GATA3* expression levels were significantly associated with a prolonged PFS (HR=0.87, 95%

CI=0.78-0.98, *P*=0.017; Table 6.4). Besides *GATA3* expression levels, also disease-free interval, but not menopausal status, dominant site of relapse and ER protein, PR protein and *ERBB2* mRNA expression levels, were associated with the length of PFS (Table 6.4). In multivariate analysis, by including *GATA3* expression in a model with all the traditional predictive factors, *GATA3* expression levels were significantly associated with a prolonged PFS (HR=0.85, 95% CI=0.75-0.96), *P*=0.0079; Table 6.4). These results imply that *GATA3* mRNA expression, rather than genetic aberration of the gene alone, is an independent predictor for the length of PFS in hormone-naive ER-positive breast cancer patients treated with first-line tamoxifen for recurrent disease.

Table 6.3 Univariate logistic regression analysis of the overall response rate in 148 ER positive breast cancer patients treated with first-line tamoxifen for recurrent disease.

	Univariate analysis	
Variable	**OR (95% CI)**	***P*-value**
Base model:		
Menopausal status[a]		
premenopausal	1	
postmenopausal	1.83 (0.83-4.03)	0.13
Dominant site of relapse		
soft	1	
bone	0.43 (0.11-1.69)	0.23
visceral	0.81 (0.20-3.40)	0.78
Disease-free interval (m)		
≤12	1	
13-36	4.70 (1.98-11.11)	0.00043
>36	3.54 (1.43-8.76)	0.0063
ER protein expression	1.12 (0.95-1.33)	0.17
PR protein expression	1.06 (0.95-1.18)	0.30
ERBB2 mRNA expression	1.07 (0.95-1.22)	0.26
Additions to the base model:		
GATA3 mRNA expression	1.12 (0.90-1.41)	0.31

[a] At the start time of first-line tamoxifen treatment; m, months.

Table 6.4 Univariate and multivariate Cox regression analyses of progression-free survival in 148 ER-positive breast cancer patients treated with first-line tamoxifen for recurrent disease.

	Univariate analysis		Multivariate analysis	
Variable	**HR (95% CI)**	***P*-value**	**HR (95% CI)**	***P*-value**
Base model:				
Menopausal status[a]				
premenopausal	1		1	
postmenopausal	0.73 (0.49-1.09)	0.12	0.84 (0.54-1.29)	0.42
Dominant site of relapse				
soft	1		1	
bone	1.88 (0.99-3.57)	0.054	2.36 (1.21-4.63)	0.012
visceral	1.47 (0.76-2.86)	0.26	1.74 (0.86-3.53)	0.13
Disease-free interval (m)				
⩽12	1		1	
13-36	0.66 (0.43-0.99)	0.046	0.61 (0.40-0.93)	0.021
>36	0.56 (0.36-0.88)	0.012	0.59 (0.37-0.94)	0.026
ER protein expression	0.95 (0.87-1.03)	0.22	0.97 (0.88-1.06)	0.49
PR protein expression	0.95 (0.90-1.00)	0.051	0.95 (0.89-1.01)	0.077
ERBB2 mRNA expression	0.98 (0.91-1.05)	0.62	0.98 (0.91-1.06)	0.62
Additions to the base model:				
GATA3 mRNA expression	0.87 (0.78-0.98)	0.017	0.85 (0.75-0.96)	0.0079

[a] At the start time of first-line tamoxifen treatment; m, months.

Discussion

GATA3 is one of the most frequently mutated genes in breast cancer[26-29] and mutations in *GATA3* are associated with improved survival[32]. Because *GATA3* mutations are also associated with both a favorable outcome among ER-positive patients who received adjuvant endocrine treatment as well as response to neoadjuvant aromatase inhibitors[32,33], we here evaluated whether *GATA3* mutations measured in the primary tumor (*i.e.* all ER positive) can determine the outcome of patients treated with first-line tamoxifen for recurrent disease. However, *GATA3* mutations were not significantly associated with either the ORR or with PFS in 235 ER-positive breast cancer patients who received tamoxifen as a first-line therapy for recurrent disease (Table S6.2 and Figure 6.2A). Even though *GATA3* mutations were

associated with increased levels of *GATA3* expression, only *GATA3* expression was found to be an independent predictor for prolonged PFS (Table 6.4). Our results suggest that not *GATA3* mutation, but rather *GATA3* expression predicts the length of PFS. This result will need to be validated in an independent patient population.

Jiang *et al.* have previously shown that *GATA3* mutations were associated with improved survival in both the TCGA cohort as well as the Chinese FUSCC cohort[32]. In the TCGA cohort, however, this prognostic effect was limited to ER-positive breast cancer cases (overall survival P=0.041) in contrast to all cases in the FUSCC cohort (overall survival P=0.033). Furthermore, in the FUSCC cohort, *GATA3* mutations were also associated with longer disease-free survival in ER-positive patients who received adjuvant endocrine treatment (P=0.046), which may suggest a role for *GATA3* mutation in the efficacy of endocrine therapy. However, our results do not show that *GATA3* mutations are associated with the outcome of tamoxifen treatment in 235 ER-positive patients who were treated with first-line tamoxifen for recurrent disease (Table S6.2 and Figure 6.2A). This suggests that the improved disease-free survival of patients with *GATA3* mutated tumors in the FUSCC cohort can be attributed to a pure prognostic association of *GATA3* mutation rather than its role as a predictive factor for tamoxifen efficacy. Noteworthy, however, is that *GATA3* mutations in the neoadjuvant setting were a predictive marker of favorable outcome of aromatase inhibitor treatment[33]. In that study, 77 ER-positive breast cancer samples were sequenced and *GATA3* mutations were more frequently present in aromatase inhibitor sensitive tumors (P=0.01). The apparent discrepancy between our study and the study of Ellis *et al.* might be attributable to a difference in the mechanism of action between aromatase inhibitors and tamoxifen or, probably more likely, due to the difference in primary versus recurrent disease receiving endocrine treatment and the used endpoints.

At the gene expression level, high *GATA3* has consistently and independent of other clinicopathological predictors been linked to a better outcome[14,19], but at the protein level the prognostic effect of GATA3 remains controversial[15,18,20,21]. Higher sensitivity and/or accuracy of gene expression compared with protein expression measurement methods could very well explain poor consistency at the protein level. Interestingly, in a small study including only 28

patients and examining the expression of GATA3 by immunohistochemistry, Parikh *et al.* found that GATA3 expression predicted hormone responsiveness in breast cancer as six of 14 (43%) cancers were GATA3 negative in the hormone-unresponsive group and 0 of 14 (0%) cancers were GATA3 negative in the hormone-responsive group (P=0.031)[22]. These results are in line with the current study, where we analyzed *GATA3* gene expression levels in 148 ER-positive recurrent breast cancer patients who were treated with first-line tamoxifen and found that high levels of *GATA3* were associated with a prolonged PFS (Figure 6.2B). Moreover, in multivariate analysis, *GATA3* expression was an independent predictor of progression-free survival (Table 6.4). The predictive effect of GATA3 at the protein level, however, requires independent examination.

In the current study, we also observed that breast cancers with a *GATA3* mutation had significantly higher *GATA3* expression levels, although only *GATA3* expression appeared to be associated with prolonged PFS. Importantly, from the 76 breast cancers with high *GATA3* expression levels, 18 (23.7%) had a *GATA3* mutation. Thus, a significant fraction (n=58, 76.3%) of the *GATA3* high breast cancers does not have a *GATA3* mutation. In order to be certain that we did not miss any mutations located outside exon 5 and 6, we additionally sequenced the other coding exons (*i.e.* exon 2-4) of the *GATA3* gene in these 58 breast cancers, but did not find any additional mutations. This not only confirms that the vast majority of *GATA3* mutations are actually located in exon 5 and 6, but this also suggests that there are other mechanisms besides *GATA3* mutation that may be responsible for the high *GATA3* expression in ER-positive breast cancers with a wild-type *GATA3* gene. As the GATA3 transcription factor reshapes gene loci by recruiting chromatin remodeling complexes, mutations of these proteins present in these complexes or other upstream pathway members could very well be involved in these mechanisms. For example, GATA3 expression was recently also reported to be increased by Wnt/β-catenin pathway activation in adipocytes[43]. Identification of these players may lead to a better understanding of the mechanisms of resistance to tamoxifen treatment by transcriptional regulation through GATA3, the crucial transcription factor regulating luminal differentiation in the mammary gland.

In conclusion, not *GATA3* mutation, but *GATA3* gene expression is associated with prolonged PFS in ER-positive breast cancer patients who received first-line tamoxifen treatment for recurrent disease. In addition, *GATA3* mutation leads to an increased *GATA3* mRNA expression, but besides genetic aberration of *GATA3*, other mechanisms are in place to explain the increased *GATA3* levels in *GATA3* wild-type tumors with high *GATA3* mRNA.

Funding

This study was funded by Cancer Genomics Netherlands (CGC.nl) and a grant for the Netherlands Organization of Scientific Research (NWO). JL received a scholarship from the China Scholarship Council (Beijing, China).

Authors' contributions

Jingjing Liu contributed to analysis and interpretation of data and wrote the manuscript. Wendy J.C. Prager-van der Smissen performed experiments and acquired data. Maxime P. Look and Marcel Smid contributed to analysis and interpretation of the data. Anieta M. Sieuwerts contributed to study concept design, performed experiments and acquired data. Marion E. Meijer-van Gelder contributed to data acquisition. John A. Foekens and John W.M. Martens contributed to study concept design and wrote the manuscript. Antoinette Hollestelle contributed to study concept design, performed experiments and acquired data, contributed to analysis and interpretation of the data and wrote the manuscript.

Acknowledgements

We want to thank Michelle van der Vlugt-Daane and Vanja de Weerd for technical assistance.

References

1 Sims, A. H., Howell, A., Howell, S. J. *et al.* Origins of breast cancer subtypes and therapeutic implications. *Nat Clin Pract Oncol* 4, 516-525 (2007).

2 Frkovic-Grazio, S. & Bracko, M. Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma. *J Clin Pathol* 55, 88-92 (2002).

3 Clarke, R., Tyson, J. J. & Dixon, J. M. Endocrine resistance in breast cancer - An overview and update. *Mol Cell Endocrinol* 418, 220-234 (2015).

4 Grote, D., Souabni, A., Busslinger, M. *et al.* Pax 2/8-regulated Gata 3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney. *Development* 133, 53-61 (2006).

5 Kaufman, C. K., Zhou, P., Pasolli, H. A. *et al.* GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev* 17, 2108-2122 (2003).

6 Asselin-Labat, M. L., Sutherland, K. D., Barker, H. *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 9, 201-209 (2007).

7 Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. *et al.* GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 127, 1041-1055 (2006).

8 Lim, K. C., Lakshmanan, G., Crawford, S. E. *et al.* Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet* 25, 209-212 (2000).

9 Hoch, R. V., Thompson, D. A., Baker, R. J. *et al.* GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer* 84, 122-128 (1999).

10 van 't Veer, L. J., Dai, H., van de Vijver, M. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536 (2002).

11 Kouros-Mehr, H., Bechis, S. K., Slorach, E. M. *et al.* GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* 13, 141-152 (2008).

12 Theodorou, V., Stark, R., Menon, S. *et al.* GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* 23, 12-22 (2013).

13 Wilson, B. J. & Giguere, V. Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* 7, 49 (2008).

14 Mehra, R., Varambally, S., Ding, L. *et al.* Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 65, 11259-11264 (2005).

15 Yoon, N. K., Maresh, E. L., Shen, D. *et al.* Higher levels of GATA3 predict better survival in women with breast cancer. *Hum Pathol* 41, 1794-1801 (2010).

16 Albergaria, A., Paredes, J., Sousa, B. *et al.* Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res* 11, R40 (2009).

17 Tominaga, N., Naoi, Y., Shimazu, K. *et al.* Clinicopathological analysis of GATA3-positive breast cancers with special reference to response to neoadjuvant chemotherapy. *Ann Oncol* 23, 3051-3057 (2012).

18 Gulbahce, H. E., Sweeney, C., Surowiecka, M. *et al.* Significance of GATA-3 expression in outcomes of patients with breast cancer who received systemic chemotherapy and/or hormonal therapy and clinicopathologic features of GATA-3-positive tumors. *Hum Pathol* 44, 2427-2431 (2013).

19 Jenssen, T. K., Kuo, W. P., Stokke, T. *et al.* Associations between gene expressions in breast cancer and patient survival. *Hum Genet* 111, 411-420 (2002).

20 Ciocca, V., Daskalakis, C., Ciocca, R. M. *et al.* The significance of GATA3 expression in breast cancer: a 10-year follow-up study. *Hum Pathol* 40, 489-495 (2009).

21 Bollet, M. A., Savignoni, A., De Koning, L. *et al.* Tumor aromatase expression as a prognostic factor for local control in young breast cancer patients after breast-conserving treatment. *Breast Cancer Res* 11, R54 (2009).

22 Parikh, P., Palazzo, J. P., Rose, L. J. *et al.* GATA-3 expression as a predictor of hormone response in breast cancer. *J Am Coll Surg* 200, 705-710 (2005).

23 Yu, K. D., Zhu, R., Zhan, M. *et al.* Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin Cancer Res* 19, 2723-2733 (2013).

24 Labastie, M. C., Bories, D., Chabret, C. *et al.* Structure and expression of the human GATA3 gene. *Genomics* 21, 1-6 (1994).

25 Van Esch, H., Groenen, P., Nesbit, M. A. *et al.* GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* 406, 419-422 (2000).

26 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70 (2012).

27 Banerji, S., Cibulskis, K., Rangel-Escareno, C. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486, 405-409 (2012).

28 Kandoth, C., McLellan, M. D., Vandin, F. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333-339 (2013).

29 Lawrence, M. S., Stojanov, P., Mermel, C. H. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495-501 (2014).

30 Gaynor, K. U., Grigorieva, I. V., Allen, M. D. *et al.* GATA3 mutations found in breast cancers may be associated with aberrant nuclear localization, reduced transactivation and cell invasiveness. *Horm Cancer* 4, 123-139 (2013).

31 Usary, J., Llaca, V., Karaca, G. *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* 23, 7669-7678 (2004).

32 Jiang, Y. Z., Yu, K. D., Zuo, W. J. *et al.* GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. *Cancer* 120, 1329-1337 (2014).

33 Ellis, M. J., Ding, L., Shen, D. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486, 353-360 (2012).

34 McShane, L. M., Altman, D. G., W., S. *et al.* Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 97, 1180-1184 (2005).

35 Sieuwerts, A. M., Meijer-van Gelder, M. E., Timmermans, M. *et al.* 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 15, 7311-7321 (2005).

36 Foekens, J. A., Portengen, H., van Putten, W. L. *et al.* Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols. *Cancer Res* 49, 5823-5828 (1989).

37 Foekens, J. A., Portengen, H., Look, M. P. *et al.* Relationship of PS2 with response to tamoxifen therapy in patients with recurrent breast cancer. *Br J Cancer* 70, 1217-1223 (1994).

38 van Agthoven, T., Sieuwerts, A. M., Meijer-van Gelder, M. E. *et al.* Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. *J Clin Oncol* 27, 542-549 (2009).

39 Hayward, J. L., Carbone, P. P., Heuson, J. C. *et al.* 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. *Cancer* 39, 1289-1294 (1977).

40 European Organization for Research and Treatment of Cancer & Breast Cancer Cooperative Group. Manual for clinical research and treatment in breast cancer. *Excerpta Medica. Almere, the Netherlands*, 116-117 (2000).

41 Riaz, M., Berns, E. M., Sieuwerts, A. M. *et al.* Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. *Breast Cancer Res Treat* 133, 843-851 (2012).

42 den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15, 7-12 (2000).

43 L., W. & Di, L. J. Wnt/β-Catenin Mediates AICAR Effect to Increase GATA3 Expression and Inhibit Adipogenesis. *J Biol Chem* 290, 19458-19468 (2015).

Chapter 7

General discussion

Discussion

Breast cancer is the most common cancer in women worldwide while incidence rates vary widely. The survival rates are heading toward a positive trend[1]. Intensive research efforts have been performed to develop novel technologies and therapeutics for a clinical application in patients suffering from breast cancer. The purpose of this thesis is to contribute to the development of breast cancer survival. In this thesis, we aimed to identify novel genetic variants associated with breast cancer risk to facilitate early detection of breast cancer. Furthermore, we aimed to discover new prognostic and predictive biomarkers for breast cancer to improve personalized treatment.

7.1 Discovering genetic variants for breast cancer susceptibility

Breast cancer is a genetic disease and arises as a consequence of genetic mutations including somatic and germline mutations. The development of breast cancer is related with many factors and a family history of breast cancer is a major one. Familial breast cancer has been associated with germline mutations in susceptibility genes. Three well-defined classes of breast cancer susceptibility genes have been identified: high-, moderate-, and low-risk genes/alleles. High-risk susceptibility genes include *BRCA1*[2], *BRCA2*[3], *CDH1*[4], *PALB2*[5], *PTEN*[6], *STK11*[7] and *TP53*[8]. Mutations in high-risk genes are very rare in frequency in the population and account for around 30% of the familial risk of breast cancer[9]. Moderate-risk genes include *ATM*, *CHEK2*, *NBS1* and *RAD50*[10]. The moderate-risk genes confer 5% increased familial breast cancer risk and mutations in those genes are rare in frequency[11]. More than 90 low-risk genes/alleles have been identified through genome wide association studies (GWAS). The low-risk genes/alleles are quite common in frequency and confer 16% of the familial breast cancer risk[12]. Altogether, these three classes of susceptibility genes/alleles explain about 51% of familial breast cancer risk and the remaining 49% of familial breast cancer risk still remains to be explained since twin studies have suggested that a large part of the remaining familial breast cancer risk is due to genetic factors[13].

Early detection of breast cancer is important for a favorable survival of breast cancer

patients. For early detection of breast cancer, we need to estimate the significant risk for developing breast cancer of individuals at an early stage. Breast cancer susceptibility genes/alleles can be applied clinically in risk estimation and risk prediction models combining breast cancer susceptibility genes/alleles are useful to make decisions on screening and prevention[14,15]. The combined effects of the low-risk SNPs are able to increase breast cancer risk significantly and thus polygenic risk scores can be used to stratify breast cancer risk[16]. In addition, the inclusion of non-genetic risk factors, such as mammographic density, weight gain and vitamin D deficiency into existing models may also improve breast cancer risk prediction[17]. For example, the Gail model has been commonly used to estimate breast cancer risk in women[18] and it is valuable to add genetic information from low-risk SNPs to the Gail model[19-21]. Moreover, the BOADICEA model which is a genetic model for familial breast cancer susceptibility predict the risk of developing breast cancer by combining the effects of *BRCA1*, *BRCA2* and all the known low-risk genes[14,22]. However, these risk prediction models are not good enough yet, as the current susceptibility genes/alleles only explain about 51% of familial breast cancer risk. Therefore, it is necessary to identify more susceptibility genes/alleles based on further genome-based research[23]. The improvement of clinical risk prediction is necessary to detect individuals with breast cancer risk earlier so that they might benefit from preventive measures.

7.1.1 Methods to identify novel breast cancer susceptibility genes

In the past, three main methods have been used to identify genetic breast cancer susceptibility genes: linkage analysis, mutational screening of candidate genes, and association studies[24]. Linkage analysis is suitable for mapping high-risk susceptibility genes only. *BRCA1* and *BRCA2* were identified by linkage analysis in breast cancer families and positional cloning. Other high-risk susceptibility genes such as *CDH1*, *PTEN* and *STK11* were also identified by positional cloning. Mutations in these genes were found to be associated with hereditary diffuse gastric cancer syndrome (*CDH1*), Cowden syndrome (*PTEN*) and Peutz-Jeghers syndrome (*STK11*), syndromes which all are also associated with an increased risk of breast cancer. Since additional high-risk genes, if these exist at all, are probably very rare, linkage

analysis is not a suitable method to identify them. If a susceptibility gene/mutation is infrequently present within the breast cancer families under investigation, linkage analysis becomes less powerful. Genome-wide linkage analyses for lower frequency genes is only valuable in a more homogenous collection of families. Alternatively, large scale sequencing projects will probably be the most optimal method to identify further high-risk genes.

BRCA1 and BRCA2 play important roles in the maintenance of genomic stability and are involved in the DNA repair pathway. Mutations in genes whose protein products operate in DNA repair pathways, like BRCA1 and BRCA2, may also be associated with breast cancer risk. Actually, the known intermediate-risk genes are all involved in DNA repair pathways[10]. All of the intermediate-risk genes were identified by candidate-gene or population-based approaches. NBS1 together with MRE11 and RAD50 plays an important role in the early response to double-strand breaks. *NBS1* and *RAD50* are current intermediate-risk genes[10] and it seems that *MRE11* is also involved in breast cancer susceptibility[25,26]. Additional genes involved in DNA repair and related pathways are thus candidate breast cancer susceptibility genes. Over the past years mutational screening of candidate genes has been a successful approach to identify novel intermediate-risk genes. However, in the current era, whole genome sequencing in familial case-control studies will be more efficient to reveal further intermediate-risk genes.

Genome wide association studies (GWAS) have been performed to identify low-risk alleles. Until today, more than 90 common low-risk alleles have been identified. It is estimated that another 1,000 low-risk loci may still remain unidentified and could confer another 14% of familial breast cancer risk[27]. Further low-risk alleles may be identified through even larger GWAS combined with imputation or genome- wide sequencing projects. As the majority of identified low-risk susceptibility alleles are not located in or near genes, post-GWAS analysis is needed to identify the causal variants and determine their function[28]. Importantly, further identification of breast cancer susceptibility genes/alleles is needed to improve upon the current breast cancer risk prediction models to facilitate early detection of breast cancer.

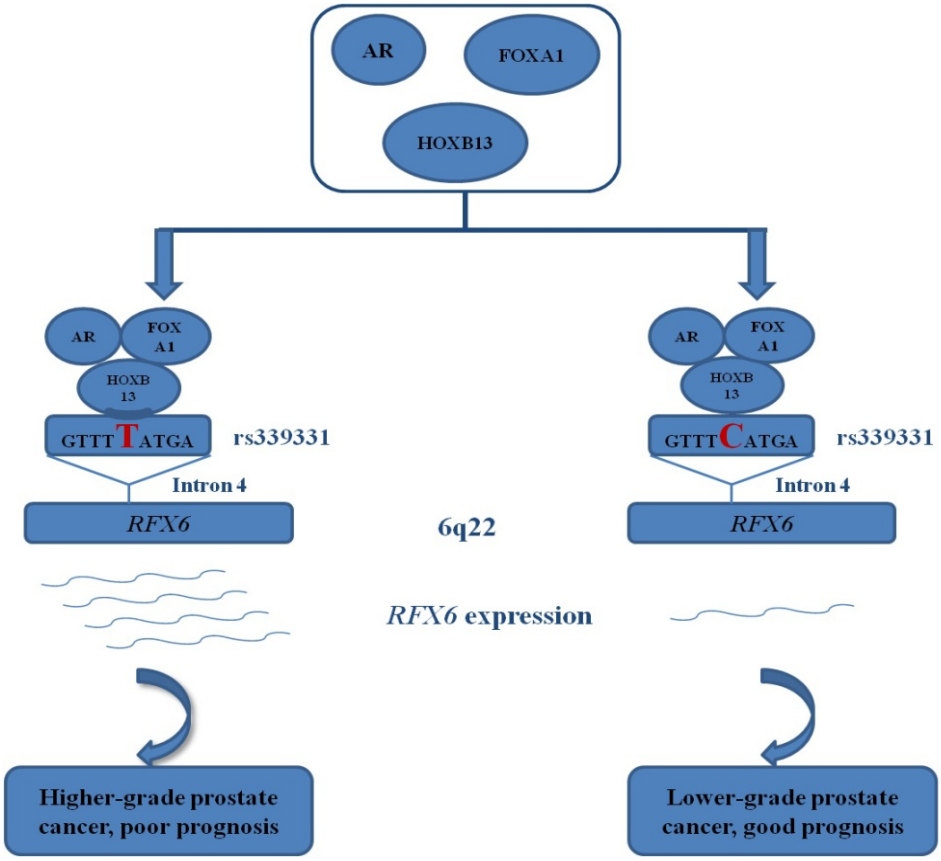
In **Chapter 3** and **4**, we aimed to discover new genes or alleles associated with breast

cancer risk by using a candidate-gene approach. Based on previous studies, we focused on *HOXB13* gene mutations and the functional variant rs2735383 in *NBS1*.

7.1.2 Mutations in the *HOXB13* gene and breast cancer susceptibility

In **Chapter 3**, we investigated the association of *HOXB13* mutations with breast cancer risk. *HOXB13* p.G84E had been reported to be associated with prostate cancer and conferred a 4- to 5-fold increased prostate cancer risk[29]. HOXB13 plays an important role in prostate cancer development[30]. The rare *HOXB13* p.G84E variant was observed to be associated with prostate cancer risk by linkage analysis and candidate gene sequencing of 200 genes at the 17q21-22 linkage region[29]. Then several groups validated the association of *HOXB13* p.G84E with prostate cancer risk[31-33]. In breast cancer tissue, high expression of *HOXB13* has been shown to mediate poor response to tamoxifen therapy[34]. Moreover, a high ratio of *HOXB13*: *IL17BR* expression was found to be a prognostic and predictive biomarker for ER-positive breast cancer patients[35,36]. Based on these observations, we hypothesized that the *HOXB13* p.G84E mutation might also be associated with breast cancer risk. At the start of the project, three studies had analyzed this association and obtained contradictory results[37-39]. In the study by Alanee *et al.* the *HOXB13* p.G84E mutation was shown to confer an increased breast cancer risk[37], however, this association was not found in the other two studies[38,39]. In our study, *HOXB13* p.G84E was found not to be associated with breast cancer risk. We also detected another recurrent mutation, p.R217C, which had not been associated with prostate cancer risk[40,41]. Although we found that this mutation was 3.5 fold more prevalent in cases than controls, the association between *HOXB13* p.R217C and breast cancer risk was not statistically significant. The limitation of our study was that the sample size was small. Due to the low population frequency, producing wide confidence intervals, the association may have failed to be detected. Therefore, it is worthwhile to evaluate the association of *HOXB13* p.R217C with breast cancer risk in a larger study of 6,237 cases and 6,237 controls. Therefore, the *HOXB13* p.R217C mutation, G84E mutation, as well as two other less frequent mutations that were identified from our study (*i.e.* p.190L, p.R268Q) have been genotyped within the Breast Cancer Association Consortium (BCAC) on the OncoArray.

Figure 7.1 The mechanism of *HOXB13* as a prostate cancer susceptibility gene.



HOXB13 is recruited to a prostate cancer-associated SNP rs339331 and enhances RFX6 expression to promote prostate cancer metastasis[44].

In prostate cancer, HOXB13 regulates the transcription of androgen receptor (AR) target genes[42]. HOXB13 was shown to bind to a prostate cancer-associated SNP located in a FOXA1 and AR binding site that enhanced RFX6 expression to promote prostate cancer metastasis (Figure 7.1)[43,44]. In breast cancer, ER and HOXB13 have been shown to regulate each other's expression[34,45]. Breast cancer risk-associated SNPs are enriched in the cistromes of FOXA1 and ER[46]. If the association of *HOXB13* mutations is found in a larger study, it should be established for instance by ChIP-seq if the HOXB13 cistrome overlaps with ER-FOXA1 cistromes providing support for the involvement of HOXB13 in ER-FOXA1 transcriptional regulation. Furthermore, by the determining the overlap between risk-associated SNPs and HOXB13 binding sites as identified by ChIP-seq, risk-associated SNPs

directly affect HOXB13 binding can be identified. Moreover, the genomic regions to which HOXB13 when bound to these SNPs interact and thus potentially regulate their gene expression can be revealed by 3C (chromosome conformation capture) which is a technique to study the spatial organization of chromatin in 3D. Finally, to investigate whether HOXB13 p.R217C affects the HOXB13 binding function, ChIP-seq profiles can be generated for p.R217C mutant transfectants and compared with the ChIP-seq profile for wild-type *HOXB13*.

7.1.3 The *NBS1* rs2735383 variant and breast cancer susceptibility

In **Chapter 4**, we investigated the association of *NBS1* rs2735383 with breast cancer risk. The DNA damage response plays an important role in susceptibility to breast cancer. NBS1 is one of the mediators involved in the DNA damage response. *NBS1* is a moderate-risk breast cancer gene, and the 657del5 mutation confers a three-fold increased breast cancer risk[47]. The functional variant rs2735383 in the 3'UTR of *NBS1,* which could regulate the expression of *NBS1,* has been reported to be associated with the risk of lung cancer[48] and colorectal cancer[49]. However, the association of *NBS1* rs2735383 with breast cancer risk was still unclear. Based on these findings in lung and colorectal cancer, we hypothesized that *NBS1* rs2735383 might be associated with breast cancer risk. In our study, *NBS1* rs2735383 was firstly genotyped in 1,159 controls and 1,269 cases from RBCS study but we did not observe an association of *NBS1* rs2735383 with increased breast cancer risk. As the RBCS cohort is relatively small and thus our analysis likely underpowered to find a small effect size, we analyzed *NBS1* rs2735383 through imputation in 47,640 breast cancer cases and 46,656 controls from BCAC studies. But also in this sufficiently sized cohort to do a properly powered analysis, we did, however, not find any evidence for an association of rs2735383 and increased breast cancer risk in the overall analysis and also not in appropriate subgroup analysis.

In a small study by Wu *et al.*, rs2735383 was shown not to be associated with overall breast cancer risk, while an association was identified with reproductive factors including age, age at menarche and menopause, menopausal status, number of pregnancies and abortions,

breast-feeding status and family history of breast cancer in first-degree relatives[50]. We did not find an association in these same subgroups. The sample size in the study by Wu *et al*. was not large and thus statistical power was only around 60%. All studies presenting novel risk alleles need to be validated in larger studies and different populations to provide accurate risk estimates. We did not replicate the findings by Wu *et al*. and we concluded *NBS1* rs2735383 is not associated with increased breast cancer risk.

In lung cancer, rs2735383 influences the binding ability of microRNA-629 and affects *NBS1* transcriptional activity and as result expression level. Low expression of *NBS1* may decrease DNA repair efficiency and thus increase lung cancer risk[48]. In colorectal cancer, microRNA-509-5p only binds to the 3'UTR of *NBS1* containing the rs2735383 C allele and as a results decreases the expression level of *NBS1*[49]. MicroRNA-499 and microRNA-508 were also shown to bind to the 3'UTR of *NBS1*[50]. In breast cancer cells, only microRNA-629 is expressed at substantial levels, however we do not know the expression of these microRNAs in normal mammary epithelial tissue. In our study, rs2735383 was not associated with breast cancer risk which might indicate that these microRNAs are not expressed in normal breast tissue and thus cannot downregulate *NBS1* or other factor compromise their effects. In further work, the microRNA expression levels in normal mammary tissue and their correlation with rs2735383 should be evaluated to understand the lack of an association of this SNP with breast cancer risk.

7.2 Discovering genetic markers for clinical outcome of breast cancer

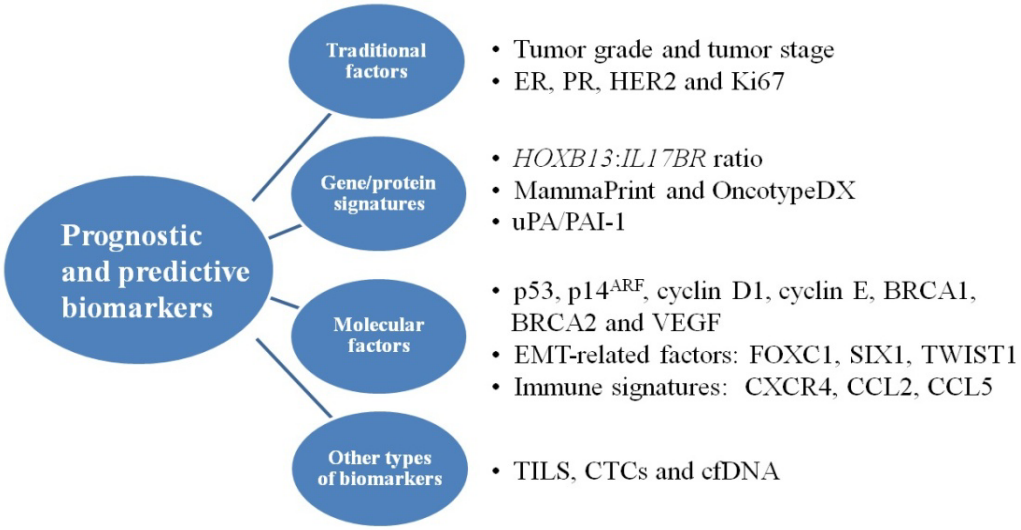
Breast cancer is a heterogeneous disease and its classification has evolved over the years. Intrinsic subtypes are essential in decision on the type of therapies as these subtypes show different biological and clinical behavior[51]. In order to provide patients with individualized treatments, prognostic and predictive biomarkers are needed (Figure 7.2). Tumor grade[52] and tumor stage[53] are traditional clinical factors while ER, PR, HER2 and Ki67 are traditional molecular prognostic and/or predictive biomarkers[54]. The use of these classical biomarkers for the prediction of treatment response and clinical outcome of breast cancer patients and

part of the current guideline for the treatment of breast cancer has been well established. Recently various prognostic and predictive gene expression signatures have been identified. For example, a high *HOXB13*:*IL17BR* ratio has been found to be associated with a high risk of recurrence and predicted poor response to tamoxifen[35,36]. The multiprotein signature uPA/PAI-1 and several multigene signatures such as MammaPrint and OncotypeDX have also been identified as prognostic and predictive biomarkers[55]. However, only a few of them have been brought to clinical utility so far. To further improve individualized treatment for breast cancer patients as well as to understand the disease, we need to identify novel prognostic and predictive biomarkers. Further molecular markers for breast cancer might be involved in tumor driver and suppressor pathways, the DNA damage response pathway and pathways promoting metastasis[56]. For example, several molecular factors such as p53, $p14^{ARF}$, cyclin D1, cyclin E, BRCA1, BRCA2 and VEGF amongst others have been identified. In addition, epithelial-mesenchymal transition (EMT) is one of the mechanisms of cancer metastasis and several transcription factors have been shown to promote EMT. Recently, the prognostic value of transcription factors promoting EMT such as FOXC1, SIX1 and TWIST1 in breast cancer patients has been studied and were associated with adverse outcome in breast cancer[57,58]. The prognostic and predictive value of the immune signatures and numbers of tumor-infiltrating lymphocytes (TILs) in relationship with breast cancer outcome has also been revealed[59]. High expression of CXCR4 in triple negative breast cancer might indicate a poorer survival[60] while high *CXCR4* mRNA expression was associated with a good survival[61]. Two chemokines CCL2[62] and CCL5[63] have been shown to be unfavorable prognostic markers for breast cancer patients. Total TILs have been shown to be associated with a better prognosis[64,65] and $CD8^+$ T cells, which are the majority of TILs, are also favorable prognostic factors for breast cancer patients[66,67]. In addition to the molecular tumor derived biomarkers, blood-derived biomarkers including circulating tumor cells (CTCs) and cell-free DNA (cfDNA) have been demonstrated to have the prognostic value in breast cancer. Elevated numbers of CTCs are associated with unfavorable clinical outcomes[68]. CfDNA are short fragments of nucleic acids in the circulation and, like CTCs, can be detected in blood. In breast cancer, higher total serum cfDNA was associated with worse survival[69,70]. In additional,

ER mutations in cfDNA are associated with poor outcome in metastatic breast cancer[71]. However, the prognostic and predictive value of these novel biomarkers need to be validated in clinical trials before clinical utility is warranted.

In **Chapter 5** and **6**, we aimed to discover new genetic markers that predict the prognosis of breast cancer patients and response to therapy. We focused on two genes: *APOBEC3B* and *GATA3*.

Figure 7.2 Prognostic and predictive biomarkers in breast cancer.



7.2.1 Discovering the prognostic and predictive value of the *APOBEC3B* deletion polymorphism

Somatic mutations do not arise at random as they occur in a genetic context suggesting specific processes are involved. These characteristic patterns of somatic mutations were designated mutational signatures and 12 distinct mutational signatures in total have been defined in breast cancer[72,73]. Signature 2 and 13 are characterized by C>T and C>G substitutions that are attributed to the AID/APOBEC family of cytidine deaminases. APOBEC3B is a member of APOBEC3 gene family and able to convert cytosine to uracil[74]. APOBEC3B was shown to be one of the underlying enzymatic sources of these types of mutations in breast cancer[75,76]. The overexpression of *APOBEC3B* was shown to be

associated with increased mutational load[75], poor outcome in ER-positive breast cancer[77] and an aggressive phenotype in Japanese breast cancers[78,79]. A 29.5 kb *APOBEC3B* deletion polymorphism results in the complete removal of the *APOBEC3B* coding region and generates a novel fusion transcript with a protein sequence of *APOBEC3A*, but with a 3'UTR of *APOBEC3B*. The 29.5 kb *APOBEC3B* deletion polymorphism has been associated with increased breast cancer risk in different populations[80-82] and potentially also result in a hypermutator phenotype and loss this polymorphism counterintuitively also correlated with an increased mutational load[83]. In this respect, we hypothesized that the *APOBEC3B* deletion polymorphism might also be associated with clinical outcome of breast cancer patients. In our study (**Chapter 5**), we investigated the association between the *APOBEC3B* deletion polymorphism and clinical outcome of 1,756 breast cancer patients. Our study showed that the *APOBEC3B* deletion polymorphism was neither a prognostic nor a predictive biomarker for tamoxifen therapy and chemotherapy in breast cancer. We did find a correlation between *APOBEC3B* copy number and *APOBEC3B* mRNA expression but this correlation was lower than expected. In contrast to high levels of *APOBEC3B* expression, the 29.5 kb *APOBEC3B* deletion polymorphism did not have clinical value.

The *APOBEC3B* deletion was also found not to be associated with the survival of breast cancer patients by Gohler *et al.*[84], however, the patients in that study received various kind of treatments. In our study we specifically analyzed pure prognosis and response to tamoxifen treatment and chemotherapy, separately. Our results showed that the *APOBEC3B* deletion polymorphism was not a prognostic biomarker for breast cancer. Furthermore, we did not find any predictive value of the *APOBEC3B* deletion polymorphism for the type of response to chemotherapy and tamoxifen therapy. The sample size of the patients who received first-line chemotherapy was small, so the predictive value to first-line chemotherapy should be revisited in larger studies.

Furthermore, we observed that the correlation between *APOBEC3B* copy number and *APOBEC3B* mRNA expression was rather weak. Thus, *APOBEC3B* copy number is not the only mechanism to regulate the mRNA level of *APOBEC3B*. Multiple groups have reported higher levels of *APOBEC3B* mRNA to be prognostic in ER-positive breast cancer patients

only and *APOBEC3B* expression was also predictive for tamoxifen resistance[85]. These findings suggested that APOBEC3B particularly plays an important role in ER-positive breast cancer but it does not explain how *APOBEC3B* expression is controlled. The signaling pathway responsible for *APOBEC3B* upregulation might be the PKC-NFκB signaling pathway[86]. PKC and NF-κB inhibitors could suppress breast cancer mutagenesis and tumor evolution, therefore the inhibition of the PKC/NF-κB signaling pathway may be a target for primary tumor treatment. Recently, DNA replication stress caused by cytotoxic agents or oncogenic signaling was also shown to modulate APOBEC3 activity which implicated the ability of therapeutics[87]. In a future study, it is important to unravel the mechanism of *APOBEC3B* upregulation as the members involved in the pathway could be a target for breast cancer therapy.

Mutational signature 2 and 13 in breast cancer are attributed to the AID/APOBEC family. Next to APOBEC3B, other APOBEC members have also been connected to these mutational signatures. APOBEC3A could also mediate somatic mutations and APOBEC3A-mediated mutagenesis is much more frequent the mutagenesis due to APOBEC3B[88]. Recently, APOBEC1[89] and APOBEC3H[90] have been proven to contribute to APOBEC mutagenesis in breast cancer. These members may be involved in the increased mutational load in various breast cancer subtypes and the roles of these APOBEC members in breast cancer need to be investigated in the future.

Tumors with the 29.5 kb *APOBEC3B* deletion polymorphism were detected to show predominantly *APOBEC3A*-like mutational signatures[88]. The role of elevated levels of *APOBEC3B* mRNA expression and the *APOBEC3B* deletion polymorphism are likely different. *APOBEC3B* mRNA expression was strongly associated with cellular proliferation, while the 29.5 kb *APOBEC3B* deletion polymorphism was related to immune activation[91]. We have found that APOBEC mutagenesis in general lead increased immune-response gene expression and tumor infiltrate[92]. APOBEC3B does play a role in innate cellular immunity against retroviral infections and the retrotransposition of endogenous elements[93,94]. Absence of *APOBEC3B* was linked to immune response-related gene expression[91,95] and *APOBEC3B* deletion carriers also had increased tumor-infiltrating immune cells[95]. Increasing lymphocytic

infiltration was reported to be associated with excellent prognosis in node-positive, ER-negative/HER2-negative breast cancer[96]. In line with this observation, the *APOBEC3B* deletion polymorphism may be a predictor for anticancer immunotherapy[91]. The novel fusion transcript presumably acts like APOBEC3A in breast cancer. However, the role of the APOBEC3A in breast cancer is still unclear. Therefore, the role of the *APOBEC3B* deletion polymorphism and the novel fusion transcript should be investigated in more studies.

7.2.2 Discovering the predictive value of *GATA3* mutations and expression

Breast cancer arises due to genetic mutations including somatic and germline mutations, while sporadic breast cancer develops only from somatic mutations. A large number of somatic mutations have been identified particularly through recent large genome sequencing studies and two different types of somatic mutations are recognized: 'driver' and 'passenger' mutations. Until today, around 140 genes that contain driver mutations have been identified[97,98]. All of the identified driver genes can be classified into 12 signaling pathways. A mutation in *GATA3* is a frequent breast cancer-specific driver event and GATA3 is involved in transcriptional regulation. GATA3 belongs to a family of zinc-finger transcription factors and plays an important role in the embryogenesis of a variety of tissues[99-103]. *GATA3* is one of the most frequently mutated genes (*i.e.* around 15% in ER^+ subtype) among breast cancers[104]. Many studies have shown that both *GATA3* mRNA and GATA3 protein expression were independent favorable prognostic markers in breast cancer[105-107]. Moreover, GATA3 was suggested to be predictive of a favorable response to hormone therapy[108] and chemo-therapy[109]. *GATA3* is a highly conserved gene. The vast majority of the *GATA3* mutations are found in exons 5 and 6, which encode a zinc finger and the C-terminal domain[110-112]. *GATA3* mutations were observed mostly in ER-positive tumors and correlated with improved overall survival in ER-positive breast cancer patients who received adjuvant endocrine therapy[113]. Recently, mutant *GATA3* was shown to be correlated with suppression of proliferation upon aromatase inhibitor treatment[114]. These findings suggested that mutant *GATA3* might predict response to tamoxifen treatment. To investigate this hypothesis, we analyzed *GATA3* mutations in 235 ER-positive breast cancer cases who received tamoxifen as a first line

therapy for recurrent disease (**Chapter 6**). We did not identify an association of *GATA3* mutations with clinical outcome, which indicated that mutant *GATA3* was not a significant predictor for tamoxifen therapy in patients with recurrent breast cancer. However, we found that *GATA3* mRNA expression was associated with an increased progression-free survival which implied that *GATA3* mRNA expression was predictive for the type of response to tamoxifen therapy.

In the very small study of Parikh *et al*., 14 patients with ER-positive hormone-unresponsive patients were compared with 14 ER-positive hormone-responsive patients and found high levels of GATA3 was also predictive of hormone response in ER-positive breast cancer patients[108]. Thus, this and our study in **Chapter 6**, underline a predictive role for *GATA3* expression. Still considering the size of both studies the current finding requires validation in an appropriate independent dataset.

Apart from the clinical association, we observed that *GATA3* mutations were associated with increased levels of *GATA3* mRNA expression, an observation that was confirmed by others in an independent cohort[92]. Importantly, the major fraction of the *GATA3* high breast cancers did not harbor a *GATA3* mutation and *GATA3* mutations thus only explained a small part of the high *GATA3* mRNA expression levels. Therefore, there must be other mechanisms besides *GATA3* mutations to explain *GATA3* overexpression. As a transcription factor, GATA3 can interact with various other factors to control its downstream pathway(s). Genetic of epigenetic regulation of proteins in these cascades may regulate the expression of *GATA3*. *GATA3* expression was recently reported to be increased by Wnt/β-catenin pathway activation in preadipocytes[115]. Moreover, GATA3 activity was shown to be modulated by ZPO2. More specifically ZPO2 when associated with ZBTB32 could downregulate the expression of *GATA3* and as a result promote aggressive breast cancer development. ZPO2 is therefore a possible candidate gene for future diagnostic and, if found clinically relevant, therapeutic strategies[116]. In the future study, expression of *GATA3* could related with ZPO2 expression in our breast cancer patient cohort and if a positive relation was found to determine their predictive value for response to tamoxifen. Identification of these factors may lead to a better understanding of *GATA3* expression regulation and its mechanistic role in resistance to

tamoxifen treatment. To understand the predictive value of *GATA3* mRNA expression, additional studies in clinical cohorts are needed for validation of the current findings and more mechanistic studies to reveal its mechanistic role.

7.3 General conclusion

In conclusion, to facilitate early detection of breast cancer, it is necessary to improve risk prediction models that are useful for early detection of breast cancer by identification of novel breast cancer risk genes/alleles. In our studies, two recurrent *HOXB13* mutations in the Dutch population and *NBS1* rs2735383 in the European and Asian populations were not associated with increased breast cancer risk. Furthermore, to improve personalized treatment it is necessary to discover new prognostic and predictive biomarkers for breast cancer. We found that the 29.5 kb *APOBEC3B* deletion polymorphism was neither a prognostic nor a predictive biomarker for breast cancer. Furthermore, *GATA3* mRNA expression, but not *GATA3* mutation, is an independent predictor for first-line tamoxifen therapy. We concluded sufficiently that *NBS1* rs2735383 was not associated with increased breast cancer risk, and the 29.5 kb *APOBEC3B* deletion polymorphism showed no prognostic value nor predictive value to first-line tamoxifen therapy. However, the status of one of the two recurrent *HOXB13* mutations, p.R217C, the predictive value to first-line chemotherapy of the 29.5 kb *APOBEC3B* deletion polymorphism and the predictive role for *GATA3* mRNA expression should be validated in larger studies. Lager studies for validation and more studies to understand the mechanisms are necessary in the future.

Reference

1 Eifel, P., Axelson, J. A., Costa, J. *et al.* National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 93, 979-989 (2001).

2 Miki, Y., Swensen, J., Shattuck-Eidens, D. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66-71 (1994).

3 Wooster, R., Bignell, G., Lancaster, J. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789-792 (1995).

4 Kaurah, P., MacMillan, A., Boyd, N. *et al.* Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA* 297, 2360-2372 (2007).

5 Cybulski, C., Kluzniak, W., Huzarski, T. *et al.* Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. *Lancet Oncol* 16, 638-644 (2015).

6 Lynch, E. D., Ostermeyer, E. A., Lee, M. K. *et al.* Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* 61, 1254-1260 (1997).

7 Leggett, B. A., Young, J. P. & Barker, M. Peutz-Jeghers syndrome: genetic screening. *Expert Rev Anticancer Ther* 3, 518-524 (2003).

8 Birch, J. M., Alston, R. D., McNally, R. J. *et al.* Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* 20, 4621-4628 (2001).

9 Fachal, L. & Dunning, A. M. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr Opin Genet Dev* 30, 32-41 (2015).

10 Hollestelle, A., Wasielewski, M., Martens, J. W. *et al.* Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* 20, 268-276 (2010).

11 Melchor, L. & Benitez, J. The complex genetic landscape of familial breast cancer. *Hum Genet* 132, 845-863 (2013).

12 Michailidou, K., Beesley, J., Lindstrom, S. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47, 373-380 (2015).

13 Mack, T. M., Hamilton, A. S., Press, M. F. *et al.* Heritable breast cancer in twins. *Br J Cancer* 87, 294-300 (2002).

14 Antoniou, A. C., Cunningham, A. P., Peto, J. *et al.* The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 98, 1457-1466 (2008).

15 Dite, G. S., MacInnis, R. J., Bickerstaffe, A. *et al.* Breast Cancer Risk Prediction

Using Clinical Models and 77 Independent Risk-Associated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev* 25, 359-365 (2016).

16 Pashayan, N., Duffy, S. W., Chowdhury, S. *et al.* Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer* 104, 1656-1663 (2011).

17 Amir, E., Freedman, O. C., Seruga, B. *et al.* Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst* 102, 680-691 (2010).

18 Gail, M. H., Brinton, L. A., Byar, D. P. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 81, 1879-1886 (1989).

19 Gail, M. H. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst* 101, 959-963 (2009).

20 Mealiffe, M. E., Stokowski, R. P., Rhees, B. K. *et al.* Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* 102, 1618-1627 (2010).

21 Wacholder, S., Hartge, P., Prentice, R. *et al.* Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 362, 986-993 (2010).

22 Antoniou, A. C., Pharoah, P. P., Smith, P. *et al.* The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer* 91, 1580-1590 (2004).

23 van Zitteren, M., van der Net, J. B., Kundu, S. *et al.* Genome-based prediction of breast cancer risk in the general population: a modeling study based on meta-analyses of genetic associations. *Cancer Epidemiol Biomarkers Prev* 20, 9-22 (2011).

24 Turnbull, C. & Rahman, N. Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* 9, 321-345 (2008).

25 Yuan, S. S., Hou, M. F., Hsieh, Y. C. *et al.* Role of MRE11 in cell proliferation, tumor invasion, and DNA repair in breast cancer. *J Natl Cancer Inst* 104, 1485-1502 (2012).

26 Bartkova, J., Tommiska, J., Oplustilova, L. *et al.* Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer:

MRE11 as a candidate familial cancer-predisposing gene. *Mol Oncol* 2, 296-316 (2008).

27 Michailidou, K., Hall, P., Gonzalez-Neira, A. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45, 353-361, 361e351-352 (2013).

28 Edwards, S. L., Beesley, J., French, J. D. *et al.* Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* 93, 779-797 (2013).

29 Ewing, C. M., Ray, A. M., Lange, E. M. *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 366, 141-149 (2012).

30 Kim, Y. R., Kim, I. J., Kang, T. W. *et al.* HOXB13 downregulates intracellular zinc and increases NF-kappaB signaling to promote prostate cancer metastasis. *Oncogene* 33, 4558-4567 (2014).

31 Witte, J. S., Mefford, J., Plummer, S. J. *et al.* HOXB13 mutation and prostate cancer: studies of siblings and aggressive disease. *Cancer Epidemiol Biomarkers Prev* 22, 675-680 (2013).

32 Huang, H. & Cai, B. G84E mutation in HOXB13 is firmly associated with prostate cancer risk: a meta-analysis. *Tumour Biol* 35, 1177-1182 (2014).

33 Cai, Q., Zhang, B., Sung, H. *et al.* Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* 46, 886-890 (2014).

34 Shah, N., Jin, K., Cruz, L. A. *et al.* HOXB13 mediates tamoxifen resistance and invasiveness in human breast cancer by suppressing ERalpha and inducing IL-6 expression. *Cancer Res* 73, 5449-5458 (2013).

35 Ma, X. J., Wang, Z., Ryan, P. D. *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 5, 607-616 (2004).

36 Jansen, M. P., Sieuwerts, A. M., Look, M. P. *et al.* HOXB13-to-IL17BR expression ratio is related with tumor aggressiveness and response to tamoxifen of recurrent breast cancer: a retrospective study. *J Clin Oncol* 25, 662-668 (2007).

37 Alanee, S., Couch, F. & Offit, K. Association of a HOXB13 variant with breast cancer. *N Engl J Med* 367, 480-481 (2012).

38 Akbari, M. R., Kluzniak, W., Rodin, R. *et al.* The HOXB13 p.Gly84Glu mutation is not associated with the risk of breast cancer. *Breast Cancer Res Treat* 136, 907-909 (2012).

39 Laitinen, V. H., Wahlfors, T., Saaristo, L. *et al.* HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 22, 452-460 (2013).

40 Xu, J., Lange, E. M., Lu, L. *et al.* HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 132, 5-14 (2013).

41 Karlsson, R., Aly, M., Clements, M. *et al.* A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur Urol* 65, 169-176 (2014).

42 Norris, J. D., Chang, C. Y., Wittmann, B. M. *et al.* The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol Cell* 36, 405-416 (2009).

43 Huang, Q., Whitington, T., Gao, P. *et al.* A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding. *Nat Genet* 46, 126-135 (2014).

44 Mills, I. G. HOXB13, RFX6 and prostate cancer risk. *Nat Genet* 46, 94-95 (2014).

45 Wang, Z., Dahiya, S., Provencher, H. *et al.* The prognostic biomarkers HOXB13, IL17BR, and CHDH are regulated by estrogen in breast cancer. *Clin Cancer Res* 13, 6327-6334 (2007).

46 Cowper-Sal lari, R., Zhang, X., Wright, J. B. *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* 44, 1191-1198 (2012).

47 Steffen, J., Nowakowska, D., Niwinska, A. *et al.* Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. *Int J Cancer* 119, 472-475 (2006).

48 Yang, L., Li, Y., Cheng, M. *et al.* A functional polymorphism at

microRNA-629-binding site in the 3'-untranslated region of NBS1 gene confers an increased risk of lung cancer in Southern and Eastern Chinese population. *Carcinogenesis* 33, 338-347 (2012).

49 Li, J. T., Zhong, B. Y., Xu, H. H. *et al.* Associations between NBS1 Polymorphisms and Colorectal Cancer in Chinese Population. *PLoS One* 10, e0132332 (2015).

50 Wu, Z., Wang, P., Song, C. *et al.* Evaluation of miRNA-binding-site SNPs of MRE11A, NBS1, RAD51 and RAD52 involved in HRR pathway genes and risk of breast cancer in China. *Mol Genet Genomics* 290, 1141-1153 (2015).

51 Gnant, M., Thomssen, C. & Harbeck, N. St. Gallen/Vienna 2015: A Brief Summary of the Consensus Discussion. *Breast Care (Basel)* 10, 124-130 (2015).

52 Elston, C. W. & Ellis, I. O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19, 403-410 (1991).

53 Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 17, 1471-1474 (2010).

54 Weigel, M. T. & Dowsett, M. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocr Relat Cancer* 17, R245-262 (2010).

55 Duffy, M. J., Harbeck, N., Nap, M. *et al.* Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer* 75, 284-298 (2017).

56 Taneja, P., Maglic, D., Kai, F. *et al.* Classical and Novel Prognostic Markers for Breast Cancer and their Clinical Significance. *Clin Med Insights Oncol* 4, 15-34 (2010).

57 Matysiak, M., Kapka-Skrzypczak, L., Jodlowska-Jedrych, B. *et al.* EMT promoting transcription factors as prognostic markers in human breast cancer. *Arch Gynecol Obstet* 295, 817-825 (2017).

58 Riaz, M., Sieuwerts, A. M., Look, M. P. *et al.* 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* 14, R123 (2012).

59 Bedognetti, D., Hendrickx, W., Marincola, F. M. *et al.* Prognostic and predictive immune gene signatures in breast cancer. *Curr Opin Oncol* 27, 433-444 (2015).

60 Chen, H. W., Du, C. W., Wei, X. L. *et al.* Cytoplasmic CXCR4 high-expression exhibits distinct poor clinicopathological characteristics and predicts poor prognosis in triple-negative breast cancer. *Curr Mol Med* 13, 410-416 (2013).

61 Dalm, S. U., Sieuwerts, A. M., Look, M. P. *et al.* Clinical Relevance of Targeting the Gastrin-Releasing Peptide Receptor, Somatostatin Receptor 2, or Chemokine C-X-C Motif Receptor 4 in Breast Cancer for Imaging and Therapy. *J Nucl Med* 56, 1487-1493 (2015).

62 Fang, W. B., Jokar, I., Zou, A. *et al.* CCL2/CCR2 chemokine signaling coordinates survival and motility of breast cancer cells through Smad3 protein- and p42/44 mitogen-activated protein kinase (MAPK)-dependent mechanisms. *J Biol Chem* 287, 36593-36608 (2012).

63 Velasco-Velazquez, M. & Pestell, R. G. The CCL5/CCR5 axis promotes metastasis in basal breast cancer. *Oncoimmunology* 2, e23660 (2013).

64 Hida, A. I., Sagara, Y., Yotsumoto, D. *et al.* Prognostic and predictive impacts of tumor-infiltrating lymphocytes differ between Triple-negative and HER2-positive breast cancers treated with standard systemic therapies. *Breast Cancer Res Treat* 158, 1-9 (2016).

65 Salgado, R., Denkert, C., Demaria, S. *et al.* The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol* 26, 259-271 (2015).

66 Liu, S., Lachapelle, J., Leung, S. *et al.* CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer. *Breast Cancer Res* 14, R48 (2012).

67 Ali, H. R., Provenzano, E., Dawson, S. J. *et al.* Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. *Ann Oncol* 25, 1536-1543

(2014).

68 Lee, J. S., Magbanua, M. J. & Park, J. W. Circulating tumor cells in breast cancer: applications in personalized medicine. *Breast Cancer Res Treat* 160, 411-424 (2016).

69 Fujita, N., Nakayama, T., Yamamoto, N. *et al.* Methylated DNA and total DNA in serum detected by one-step methylation-specific PCR is predictive of poor prognosis for breast cancer patients. *Oncology* 83, 273-282 (2012).

70 De Mattos-Arruda, L. & Caldas, C. Cell-free circulating tumour DNA as a liquid biopsy in breast cancer. *Mol Oncol* 10, 464-474 (2016).

71 Angus, L., Beije, N., Jager, A. *et al.* ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients. *Cancer Treat Rev* 52, 33-40 (2017).

72 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C. *et al.* Signatures of mutational processes in human cancer. *Nature* 500, 415-421 (2013).

73 Nik-Zainal, S., Davies, H., Staaf, J. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47-54 (2016).

74 Conticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. *et al.* Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22, 367-377 (2005).

75 Burns, M. B., Lackey, L., Carpenter, M. A. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366-370 (2013).

76 Harris, R. S. Molecular mechanism and clinical impact of APOBEC3B-catalyzed mutagenesis in breast cancer. *Breast Cancer Res* 17, 8 (2015).

77 Sieuwerts, A. M., Willis, S., Burns, M. B. *et al.* Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm Cancer* 5, 405-413 (2014).

78 Tsuboi, M., Yamane, A., Horiguchi, J. *et al.* APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. *Breast Cancer* (2015).

79 Tokunaga, E., Yamashita, N., Tanaka, K. *et al.* Expression of APOBEC3B mRNA in

Primary Breast Cancer of Japanese Women. *PLoS One* 11, e0168090 (2016).

80 Long, J., Delahanty, R. J., Li, G. *et al.* A common deletion in the APOBEC3 genes and breast cancer risk. *J Natl Cancer Inst* 105, 573-579 (2013).

81 Xuan, D., Li, G., Cai, Q. *et al.* APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* 34, 2240-2243 (2013).

82 Rezaei, M., Hashemi, M., Hashemi, S. M. *et al.* APOBEC3 Deletion is Associated with Breast Cancer Risk in a Sample of Southeast Iranian Population. *Int J Mol Cell Med* 4, 103-108 (2015).

83 Nik-Zainal, S., Wedge, D. C., Alexandrov, L. B. *et al.* Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. *Nat Genet* 46, 487-491 (2014).

84 Gohler, S., Da Silva Filho, M. I., Johansson, R. *et al.* Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population. *J Cancer Res Clin Oncol* (2015).

85 Law, E. K., Sieuwerts, A. M., LaPara, K. *et al.* The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci Adv* 2, e1601737 (2016).

86 Leonard, B., McCann, J. L., Starrett, G. J. *et al.* The PKC/NF-kappaB signaling pathway induces APOBEC3B expression in multiple human cancers. *Cancer Res* 75, 4538-4547 (2015).

87 Kanu, N., Cerone, M. A., Goh, G. *et al.* DNA replication stress mediates APOBEC3 family mutagenesis in breast cancer. *Genome Biol* 17, 185 (2016).

88 Chan, K., Roberts, S. A., Klimczak, L. J. *et al.* An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet* 47, 1067-1072 (2015).

89 Saraconi, G., Severi, F., Sala, C. *et al.* The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal

adenocarcinomas. *Genome Biol* 15, 417 (2014).

90 Starrett, G. J., Luengas, E. M., McCann, J. L. *et al.* The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat Commun* 7, 12918 (2016).

91 Cescon, D. W., Haibe-Kains, B. & Mak, T. W. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc Natl Acad Sci U S A* 112, 2841-2846 (2015).

92 Smid, M., Rodriguez-Gonzalez, F. G., Sieuwerts, A. M. *et al.* Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nat Commun* 7, 12910 (2016).

93 Doehle, B. P., Schafer, A. & Cullen, B. R. Human APOBEC3B is a potent inhibitor of HIV-1 infectivity and is resistant to HIV-1 Vif. *Virology* 339, 281-288 (2005).

94 Bogerd, H. P., Wiegand, H. L., Doehle, B. P. *et al.* APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res* 34, 89-95 (2006).

95 Wen, W. X., Soo, J. S., Kwan, P. Y. *et al.* Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. *Breast Cancer Res* 18, 56 (2016).

96 Loi, S., Sirtaine, N., Piette, F. *et al.* Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol* 31, 860-867 (2013).

97 Vogelstein, B., Papadopoulos, N., Velculescu, V. E. *et al.* Cancer genome landscapes. *Science* 339, 1546-1558 (2013).

98 Chen, J. C., Alvarez, M. J., Talos, F. *et al.* Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. *Cell* 166, 1055 (2016).

99 Grote, D., Souabni, A., Busslinger, M. *et al.* Pax 2/8-regulated Gata 3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing

kidney. *Development* 133, 53-61 (2006).

100 Kaufman, C. K., Zhou, P., Pasolli, H. A. *et al.* GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev* 17, 2108-2122 (2003).

101 Asselin-Labat, M. L., Sutherland, K. D., Barker, H. *et al.* Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* 9, 201-209 (2007).

102 Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. *et al.* GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 127, 1041-1055 (2006).

103 Lim, K. C., Lakshmanan, G., Crawford, S. E. *et al.* Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet* 25, 209-212 (2000).

104 Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70 (2012).

105 Mehra, R., Varambally, S., Ding, L. *et al.* Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* 65, 11259-11264 (2005).

106 Yoon, N. K., Maresh, E. L., Shen, D. *et al.* Higher levels of GATA3 predict better survival in women with breast cancer. *Hum Pathol* 41, 1794-1801 (2010).

107 Jenssen, T. K., Kuo, W. P., Stokke, T. *et al.* Associations between gene expressions in breast cancer and patient survival. *Hum Genet* 111, 411-420 (2002).

108 Parikh, P., Palazzo, J. P., Rose, L. J. *et al.* GATA-3 expression as a predictor of hormone response in breast cancer. *J Am Coll Surg* 200, 705-710 (2005).

109 Tominaga, N., Naoi, Y., Shimazu, K. *et al.* Clinicopathological analysis of GATA3-positive breast cancers with special reference to response to neoadjuvant chemotherapy. *Ann Oncol* 23, 3051-3057 (2012).

110 Arnold, J. M., Choong, D. Y., Thompson, E. R. *et al.* Frequent somatic mutations of GATA3 in non-BRCA1/BRCA2 familial breast tumors, but not in BRCA1-, BRCA2- or sporadic breast tumors. *Breast Cancer Res Treat* 119, 491-496 (2010).

111 Usary, J., Llaca, V., Karaca, G. *et al.* Mutation of GATA3 in human breast tumors. *Oncogene* 23, 7669-7678 (2004).

112 Nesbit, M. A., Bowl, M. R., Harding, B. *et al.* Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem* 279, 22624-22634 (2004).

113 Jiang, Y. Z., Yu, K. D., Zuo, W. J. *et al.* GATA3 mutations define a unique subtype of luminal-like breast cancer with improved survival. *Cancer* 120, 1329-1337 (2014).

114 Ellis, M. J., Ding, L., Shen, D. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486, 353-360 (2012).

115 Wang, L. & Di, L. J. Wnt/beta-Catenin Mediates AICAR Effect to Increase GATA3 Expression and Inhibit Adipogenesis. *J Biol Chem* 290, 19458-19468 (2015).

116 Shahi, P., Wang, C. Y., Lawson, D. A. *et al.* ZNF503/Zpo2 drives aggressive breast cancer progression by down-regulation of GATA3 expression. *Proc Natl Acad Sci U S A* (2017).

Chapter 8

Summary/Samenvatting

Summary

Breast cancer is the second leading cause of deaths by cancers worldwide. The incidence of breast cancer is rising in most countries, while survival of patients with breast cancer has improved. Breast cancer is a genetic disease and arises as a consequence of genetic mutations including both sporadic driver mutations and germline breast cancer susceptibility alleles. Breast cancer is also a heterogeneous disease and prognostic and predictive biomarkers are important to provide patients with personalized treatment.

In the first part of the thesis (**Chapter 3** and **4**), we aimed to discover new breast cancer susceptibility alleles. In **Chapter 3**, we evaluated whether mutations in the *HOXB13* gene were associated with breast cancer risk. The *HOXB13* gene encodes a transcription factor and is important in tumorigenesis. The rare variant p.G84E in the *HOXB13* gene was reported to be associated with an increased risk to develop prostate cancer, while in breast cancer, a high *HOXB13*:*IL17BR* ratio was found to be a prognostic and predictive biomarker for ER-positive breast cancer patients. Based on these findings, we hypothesized that *HOXB13* might also be a breast cancer susceptibility gene. We firstly analyzed the entire *HOXB13* coding sequence in 1,250 non-*BRCA1/2* familial breast cancer cases and 800 controls and identified two predicted deleterious missense mutations, p.G84E and p.R217C. Then, in total 4,520 familial non-*BRCA1/2* breast cancer cases and 3,127 controls were further genotyped for these two recurrent mutations. The prostate cancer risk allele p.G84E, but also mutation p.R217C were found not to be associated with breast cancer risk. Considering the low population frequency and the wide confidence intervals for the p.R217C mutation, the association may have failed to be detected. Therefore, it is recommended that p.R217C is evaluated in a larger study.

In **Chapter 4,** we investigated whether the single nucleotide variant *NBS1* rs2735383 was associated with an increased breast cancer risk. The DNA damage response (DDR) pathway maintains human genome stability and plays an important role in susceptibility to breast cancer. NBS1 is involved in double-strand break repair and germline *NBS1* mutations are associated with the risk to develop several types of cancer. The germline *NBS1* mutation c.657del5 was shown to be associated with breast cancer risk. The genotype of functional

variant rs2735383, localized in the 3'UTR of *NBS1,* regulates the expression level of the *NBS1* gene by influencing the binding ability of several microRNAs. Furthermore, rs2735383 has been reported to be associated with lung cancer and colorectal cancer risk. The association of *NBS1* rs2735383 with breast cancer risk was yet unclear. Therefore we aimed to investigate this association. We analyzed *NBS1* rs2735383 by RFLP-PCR in 1,159 controls and 1,269 cases of the Rotterdam Breast Cancer Study (RBCS) study and found that rs2735383 was not significantly associated with breast cancer risk. As the RBCS study was relatively small and underpowered to find a small effect size, we additionally analyzed *NBS1* rs2735383 in 47,640 breast cancer cases and 46,656 controls from Breast Cancer Association Consortium (BCAC) studies through imputation. For the overall analysis, *NBS1* rs2735383 was not associated with an increased risk to develop breast cancer in Europeans or in Asians. Also for the subtype-specific analysis according to age, age at menarche, age at menopause, menopausal status, number of full-term pregnancies, breast feeding, family history and receptor status, we did not find any association between *NBS1* rs2735383 and breast cancer risk.

The results from **Chapter 3** and **4** showed that two recurrent *HOXB13* mutations in the Dutch population and *NBS1* rs2735383 in the European and Asian population were not associated with increased breast cancer risk.

In the second part of the thesis (**Chapter 5** and **6**), we aimed to discover new prognostic and predictive biomarkers for breast cancer. In **Chapter 5**, the prognostic and predictive value of the *APOBEC3B* deletion polymorphism was evaluated for breast cancer patients. APOBEC3B is an endogenous source of mutation in breast cancer and high levels of *APOBEC3B* mRNA has been shown to be associated with a poor prognosis in ER-positive breast cancer. A 29.5 kb *APOBEC3B* deletion polymorphism has also been found to increase the mutational load and associate with an increased breast cancer risk. We hypothesized that the 29.5 kb *APOBEC3B* deletion polymorphism was associated with clinical outcome of breast cancer patients. To evaluate the prognostic and predictive value, we analyzed the *APOBEC3B* copy number in 1,756 patients who were divided in four cohorts: 528 LNP patients who received adjuvant systemic treatment, 1,076 LNN patients who did not receive

any adjuvant systemic treatment, 329 ER-positive patients who received tamoxifen as first-line treatment for recurrent breast cancer and 226 patients who received first-line chemotherapy for recurrent disease. No association between *APOBEC3B* copy numbers and clinical outcome was observed in any of these four cohorts, which implied that *APOBEC3B* copy number was not a prognostic nor a predictive biomarker for breast cancer outcome. We further identified that the association between *APOBEC3B* copy number and *APOBEC3B* mRNA levels was low.

In **Chapter 6**, we evaluated whether the *GATA3* mutation was an independent predictor for the clinical outcome of breast cancer patients with recurrent disease who had received tamoxifen as first-line therapy. GATA3 is a component of the ER pathway and is highly expressed in the luminal breast cancer subtype. *GATA3* expression has been shown to be an independent prognostic marker for breast cancer. Mutated *GATA3* was frequently present in ER-positive breast cancers and has been associated with the favorable response to neoadjuvant aromatase inhibitor treatment. These findings suggested that *GATA3* mutations may be a predictive biomarker for the type of response to hormonal treatment in recurrent breast cancer. We performed sequence analysis for exons 5 and 6 of the *GATA3* gene in 235 ER-positive primary breast cancers and found mutations in 14% of these tumors. However, *GATA3* mutations were not significantly associated with either the overall response rate or with the length of progression-free survival. The *GATA3* mutation status was thus not a predictor for the type of response to tamoxifen therapy. Out of the 235 tumors, we also performed *GATA3* mRNA expression analysis for 148 tumors. In contrast to *GATA3* mutation status, *GATA3* expression was shown to be an independent predictor for good prognosis for breast cancer patients who received tamoxifen as first-line therapy for recurrent disease.

In the second part of the thesis, we found that the 29.5 kb *APOBEC3B* deletion polymorphism was neither a prognostic nor a predictive biomarker for breast cancer outcome. Furthermore, *GATA3* mRNA expression, but not *GATA3* mutation, was an independent predictor for the efficacy of first-line tamoxifen therapy.

Samenvatting

Borstkanker is wereldwijd de op een na belangrijkste doodsoorzaak door kanker. De incidentie van borstkanker stijgt in de meeste landen nog steeds, terwijl de overleving van patiënten met borstkanker verbetert. Borstkanker is een genetische ziekte en ontstaat als gevolg van genetische mutaties, zowel sporadische tumor-drijvende mutaties als kiembaan borstkanker predispositie allelen. Borstkanker is ook een heterogene ziekte en prognostische en predictieve biomarkers zijn belangrijk om patiënten te kunnen voorzien van gepersonaliseerde therapie.

In het eerste deel van dit proefschrift (**Hoofdstuk 3 en 4**) was ons doel om nieuwe predispositie allelen voor borstkanker te identificeren. In **Hoofdstuk 3** hebben we geëvalueerd of mutaties in het *HOXB13* gen associeerden met een risico op het ontstaan van borstkanker. Het *HOXB13* gen codeert voor een transcriptiefactor welke belangrijk is tijdens de tumorgenese. Er is gerapporteerd dat de zeldzame variant p.G84E in het *HOXB13* gen geassocieerd was met een verhoogd risico op het ontwikkelen van prostaatkanker, terwijl een hoge *HOXB13*:*IL17BR* expressieratio een prognostische en predictieve factor is voor ER positieve borstkanker patiënten. Gebaseerd op deze bevindingen was onze hypothese dat *HOXB13* ook een borstkanker predispositiegen kan zijn. Daarom hebben we allereerst de complete coderende sequentie van *HOXB13* geanalyseerd in het bloed van 1250 non-*BRCA1/2* familiaire borstkankerpatiënten en identificeerden twee terugkerende en waarschijnlijk pathogene missense mutaties, p.G84E en p.R217C. Vervolgens werden deze terugkerende mutaties gegenotypeerd in een totaal aantal van 4520 familiaire non-*BRCA1/2* borstkankerpatiënten en 3127 controle individuen. Het prostaatkanker risico allel p.G84E, maar ook de p.R217C mutatie, bleken niet te associëren met een risico op het ontstaan van borstkanker. Wanneer er rekening wordt gehouden met de lage populatiefrequentie en de brede betrouwbaarheidsintervallen voor de p.R217C mutatie is het mogelijk dat we niet in staat waren om de associatie te detecteren in de huidige situatie. Om deze reden wordt het aanbevolen om de p.R217C mutatie nogmaals te evalueren in een grotere studie.

In **Hoofdstuk 4** hebben we onderzocht of de *NBS1* rs2735383 variant associeerde met een verhoogd risico op het ontstaan van borstkanker. Het mechanisme dat DNA schade

repareert in de cel beschermt de stabiliteit van het genoom en speelt een belangrijke rol bij de gevoeligheid voor borstkanker. NBS1 is betrokken bij het herstel van dubbelstrengs DNA breuken en kiembaan mutaties in *NBS1* associëren met een verhoogd risico op het ontwikkelen van verschillende typen kanker. De kiembaan *NBS1* mutatie c.657del5 is verbonden met een verhoogd risico op het ontwikkelen van borstkanker. Het genotype van de functionele variant rs2735383, welke gelokaliseerd is in het 3' niet getransleerde gebied van het *NBS1* gen, reguleert de expressie van het *NBS1* gen door de bindingsaffiniteit van verschillende microRNA's te beïnvloeden. Daarnaast werd gerapporteerd dat rs2735383 geassocieerd is met een verhoogd long-en darmkanker risico. De associatie tussen *NBS1* rs2735383 en het risico op het ontwikkelen van borstkanker was nog onbekend. In dit hoofdstuk was ons doel deze associatie te onderzoeken. Hiervoor analyseerden we *NBS1* rs2735383 door middel van een RFLP-PCR in bloed van 1159 controle en 1269 aangedane individuen uit de Rotterdamse Borstkanker Studie (RBCS) en ontdekten dat rs2735383 niet significant associeerde met een borstkanker risico. Omdat RBCS een relatief kleine studie was en dus niet genoeg statistische power bevatte om kleine effecten te detecteren hebben we *NBS1* rs2735383 door middel van imputatie aanvullend geanalyseerd in 47640 borstkankerpatiënten en 46656 controle individuen van het Borstkanker Associatie Consortium (BCAC). In de globale analyse was *NBS1* niet geassocieerd met een verhoogd risico op het ontwikkelen van borstkanker in Europeanen noch in Aziaten. Ook in de subtype specifieke analyses gebaseerd op leeftijd, leeftijd bij menarche, leeftijd bij menopause, menopausale status, aantal voldragen zwangerschappen, borstvoeding, familiegeschiedenis en hormoonreceptor status vonden we geen enkele associatie tussen *NBS1* 2735383 en het risico op borstkanker.

The resultaten van **Hoofdstuk 3 en 4** lieten zien dat twee terugkerende *HOXB13* mutaties in de Nederlandse populatie en *NBS1* rs2738353 in de Europese en Aziatische populatie niet geassocieerd zijn met een verhoogd risico op het krijgen van borstkanker.

In het tweede deel van dit proefschrift (**Hoofdstuk 5 en 6**) was het ons doel om genetische markers met prognostische en/of predictieve waarde voor borstkanker te ontdekken. In **Hoofdstuk 5** werd de prognostische en predictieve waarde van het 29.5 kb

APOBEC3B deletie polymorfisme geëvalueerd in borstkankerpatiënten. APOBEC3B is een endogene bron van mutaties in borstkanker en hoge *APOBEC3B* mRNA niveaus associëren met een slechte prognose voor patiënten met ER positieve borstkanker. Daarnaast is gevonden dat een 29.5 kb APOBEC3B deletie polymorfisme associeert met een verhoogde hoeveelheid aan mutaties en een verhoogd borstkanker risico. Onze hypothese was daarom dat het 29.5 kb APOBEC3B deletie polymorfisme geassocieerd was met de klinische uitkomst van borstkankerpatiënten. Om de prognostische en predictieve waarde te evalueren hebben we het aantal kopieën van het *APOBEC3B* gen bepaald in tumorweefsel van 1756 patiënten welke voor de analyses opgedeeld werden in 4 cohorten, n.l. 528 lymfeklier-positieve patiënten die adjuvant systemisch behandeld werden, 1076 lymfeklier-negatieve patiënten die geen adjuvante systemische therapie ondergingen, 329 ER-positieve patiënten die behandeld werden met tamoxifen in de eerste lijn voor terugkerende ziekte en 226 patiënten die behandeld werden met chemotherapie in de eerste lijn voor terugkerende ziekte. We vonden geen associatie tussen het aantal kopieën van *APOBEC3B* en de klinische uitkomst van de patiënten in elk van de vier cohorten. Dit impliceerde dat het aantal kopieën van het *APOBEC3B* gen dus geen prognostische noch predictieve waarde heeft voor de uitkomst na de diagnose borstkanker. Verder hebben we vastgesteld dat de associatie tussen het aantal kopieën van APOBEC3B en de APOBEC3B mRNA expressie laag was.

In **Hoofdstuk 6** hebben we geëvalueerd of de *GATA3* mutatie status een onafhankelijke voorspeller is voor de klinische uitkomst van borstkankerpatiënten met terugkerende ziekte en welke tamoxifen als eerste lijns therapie ontvingen. GATA3 is onderdeel van het ER-signaaltransductiepad en komt hoog tot expressie bij het luminale borstkanker subtype. Eerder onderzoek heeft laten zien dat *GATA3* expressie een onafhankelijke prognostische factor is voor borstkanker. *GATA3* is frequent gemuteerd in ER-positief borstkanker en deze mutaties bleken geassocieerd met een gunstige respons op de behandeling met neoadjuvante aromataseremmers. Deze bevindingen suggereren dat *GATA3* mutatie mogelijk een predictieve biomarker is voor het type respons op hormonale behandeling bij borstkanker. We voerden een sequentie analyse uit voor de exonen 5 en 6 van het *GATA3* gen in 235

ER-positieve primaire borstkankers and ontdekten mutaties in 14% van de tumoren. Echter, *GATA3* mutaties associeerden niet significant met de respons op eerste lijns tamoxifen therapie of met de lengte van de progressie-vrije overleving na start van de eerste lijns behandeling. De *GATA3* mutatie status is dus geen voorspeller voor het type respons op tamoxifen behandeling. Voor 148 van de 235 tumoren hebben we additioneel expressieanalyse voor *GATA3* mRNA uitgevoerd. In tegenstelling tot *GATA3* mutatiestatus vonden we dat *GATA3* expressie wel een onafhankelijke voorspeller voor een goede prognose is in borstkankerpatiënten met terugkerende ziekte die tamoxifen als eerste lijns therapie ontvingen.

In het tweede deel van dit proefschrift vonden we dat het 29.5 kb APOBEC3B deletie polymorfisme niet een prognostische en ook geen predictieve biomarker was voor borstkanker. Daarnaast bleek GATA3 mRNA expressie, maar niet GATA3 mutatiestatus een onafhankelijke voorspeller te zijn voor de werkzaamheid van tamoxifen in de eerste lijn.

Appendices

Acknowledgements

List of Publications

PhD Portfolio

Curriculum Vitae

Acknowledgements

Finally I reached this section and finished this PhD thesis. This journey was indeed challenging. I would not be able to finish this without the help around me. I want to thank everyone who facilitated me in this journey.

First of all, I would like to express my sincere gratitude to my promotor Prof. John Foekens and co-promotors John Martens and Antoinette Hollestelle. I am very grateful for the great opportunity you provided me to start my PhD journey in your lab. Dear John and John, thank you for your guidance of my PhD study and the thesis. Dear Antoinette, thank you for giving me clear comments and insights down to details of developing studies and manuscripts. Without your help, I will not finish my scientific papers and this thesis. You are so nice and friendly to me. I am very lucky to have you as my supervisors and I had a very nice time in your lab. I also express my gratitude to my committee. Thanks for your comments and approving my thesis.

I would also like to thank all my colleagues from the department of Medical Oncology in Erasmus MC. Muhammad, NingQing, Tommaso, Diana, Katherina, René, Marjolein, Nick, Joan, Wendy, Anita, Anieta, Maxime, Marion, Marcel, Vanja, Michelle, Jean, Corine, Mieke, Reneé, Maurice, Mai, Jaco, Kirsten, Rosita, thank you all for creating such great atmosphere. Furthermore, Anieta, Anita, Corine, Marcel, Maxime, Marion, Mieke, Wendy, Michelle, Reneé and the internship students Priyanca and Ivona, thanks to your work on my projects.

My lovely Chinese friends: Qiushi, Diya, Wenting, Yanwei, Changbin, Bin, Tao, Kuikui, Ruoyu, Yingying, Zhanmin, Ruxi, Wei, Qiye, shanshan, hui, thanks for the great time we had together making dishes, playing games and travelling other cities. I have a nice spare time in the Netherlands.

At last, I would like to thank my parents and two elder sisters, for their unconditional support and trust. I owe my sincere to my grandpa and grandma who passed away during my PhD. I wish you can rest in peace in heaven.

List of Publications

1. rs2735383, located at a microRNA binding site in the 3'UTR of the *NBS1* gene, is not associated with breast cancer risk.

   Jingjing Liu, Ivona Lončar, J. Margriet Collée, Manjeet K. Bolla, Joe Dennis, Kyriaki Michailidou, Qin Wang, Irene L. Andrulis, Monica Barile, Matthias W. Beckmann, Sabine Behrens, Javier Benitez, Carl Blomqvist, Bram Boeckx, Natalia V. Bogdanova, Stig E. Bojesen, Hiltrud Brauch, Paul Brennan, Hermann Brenner, Annegien Broeks, Barbara Burwinkel, Jenny Chang-Claude, Shou-Tung Chen, Georgia Chenevix- Trench, Ching Y. Cheng, Ji-Yeob Choi, Fergus J. Couch, Angela Cox, Simon S. Cross, Katarina Cuk, Kamila Czene, Thilo Dörk, Isabel dos-Santos-Silva, Peter A. Fasching, Jonine Figueroa, Henrik Flyger, Montserrat García-Closas, Graham G. Giles, Gord Glendon, Mark S. Goldberg, Anna González-Neira, Pascal Guénel, Christopher A. Haiman, Ute Hamann, Steven N. Hart, Mikael Hartman, Sigrid Hatse, John L. Hopper, Hidemi Ito, Anna Jakubowska, Maria Kabisch, Daehee Kang, Veli-Matti Kosma, Vessela N. Kristensen, Loic Le Marchand, Eunjung Lee, Jingmei Li, Artitaya Lophatananon, Jan Lubinski, Arto Mannermaa, Keitaro Matsuo, Roger L. Milne, NBCS Collaborators, Susan L. Neuhausen, Heli Nevanlinna, Nick Orr, Jose I.A. Perez, Julian Peto, Thomas C. Putti, Katri Pylkäs, Paolo Radice, Suleeporn Sangrajrang, Elinor J. Sawyer, Marjanka K. Schmidt, Andreas Schneeweiss, Chen-Yang Shen, Martha J. Shrubsole, Xiao-Ou Shu, Jacques Simard, Melissa C. Southey, Anthony Swerdlow, Soo H. Teo, Daniel C. Tessier, Somchai Thanasitthichai, Ian Tomlinson, Diana Torres, Thérèse Truong, Chiu-Chen Tseng, Celine Vachon, Robert Winqvist, Anna H. Wu, Drakoulis Yannoukakos, Wei Zheng, Per Hall, Alison M. Dunning, Douglas F. Easton, Maartje J. Hooning, Ans M.W. van den Ouweland, John W.M. Martens, Antoinette Hollestelle.
   Sci Rep 6, 36874 (2016).

2. The 29.5 kb *APOBEC3B* deletion polymorphism is not associated with clinical outcome of breast cancer.

   Jingjing Liu, Anieta M. Sieuwerts, Maxime P. Look, Michelle van der Vlugt-Daane, Marion E. Meijer-van Gelder, John A. Foekens, Antoinette Hollestelle, John W.M. Martens.
   PLoS One 11, e0161731 (2016).

3. Recurrent *HOXB13* mutations in the Dutch population do not associate with increased breast cancer risk.

   Jingjing Liu, Wendy J.C. Prager-van der Smissen, Marjanka K. Schmidt, J. Margriet Collée, Sten Cornelissen, Roy Lamping, Anja Nieuwlaat, John A. Foekens, Maartje J. Hooning, Senno Verhoef, Ans M.W. van den Ouweland, Frans B.L. Hogervorst, John W.M. Martens, Antoinette Hollestelle.
   Sci Rep 6, 30026 (2016).

4. *GATA3* mRNA expression, but not mutation, associates with longer progression-free survival in ER-positive breast cancer patients treated with first-line tamoxifen for recurrent disease.

   Jingjing Liu, Wendy J.C. Prager-van der Smissen, Maxime P. Look, Anieta M. Sieuwerts, Marcel Smid, Marion E. Meijer-van Gelder, John A. Foekens, Antoinette Hollestelle, John W.M. Martens.
   Cancer Lett 376, 104-109 (2016).

5. A new artemisinin derivative SM1044 induces apoptosis of Kasumi-1 cells and its mechanism.

Jingjing Liu, Aimei Fei, Ruimin Nie, Jin Wang, Ying Li, Zhenyi Wang, Jianqing Mi
Journal of Experimental Hematology 19, 607-611 (2011).

6. Effect of arsenic trioxide on induction of apoptosis in MCL cell line and its possible mechanisms

   Aimei Fei, Chaoming Mao, **Jingjing Liu**, Jiang Zhu, Jianqing Mi
   Journal of Experimental Hematology 18, 909-913 (2010).

PhD portfolio

Name PhD student	Jingjing Liu
Erasmus MC Department	Medical Oncology
Research School	Postgraduate school Molecular Medicine
PhD period	September 2011 to February 2016
Promotor	Prof. dr. John A. Foekens
Copromotor	Dr. A. Hollestelle

PhD training

General academic skills	**Year**	**ECTS**
-Biomedical English writing and communication	2014	3.0
-Laboratory animal science	2012	5.0

In-depth courses

-Molecular Medicine	2012	0.7
-Basic and Translational Oncology	2012	1.8
-Molecular Diagnostics	2012	1.0
-SNP course	2012	2.0
-Biomedical research	2012	1.5
-Basic introduction course on SPSS	2013	1.0
-Biostatistical Methods I: Basic Principles	2013	5.7
-Genome Maintenance and Cancer	2014	0.8
-Epigenetic regulation	2014	0.8

Presentations

-Journal club	2011-2015	1.5
-Oral presentation at TCGP meeting	2013	0.3

-Oral presentation at JNI Scientific Lab meeting	2014	0.3
-Oral presentation at TCGP meeting	2014	0.3
-Poster presentation at MolMed day	2015	0.9
-Poster presentation at the annual meeting of the American Association for Cancer Research	2015	0.3

International conferences

-Annual meeting of the American Association for Cancer Research	2015	1.5

Seminars and workshops

-Monthly TCGP meeting	2011-2015	1.0
-MolMed day	2012-2015	2.0
-JNI Scientific Lab meeting	2011-2015	2.5
-Journal club	2011-2015	1.0

Supervising practicals and bachelor's thesis

-Internship of HLO student Priyanca Asra August 2013-March 2014	2013-2014	4.0

Total ECTS	38.9

Curriculum Vitae

Jingjing Liu was born in Puyang City in People's Republic of China on May 21, 1987. She studied in the college of life science in Henan University in China from 2004 and obtained a bachelor degree in 2008. That same year she started her master education in school of medicine, Shanghai Jiao Tong University in China. In 2011 she obtained her master degree and was awarded a scholarship to conduct a PhD at the department of Medical Oncology supervised by Professor John Foekens at Erasmus MC in the Netherlands. During her PhD study, she worked on human breast cancer, which included to identify novel breast cancer susceptibility variants and new prognostic and predictive biomarkers. Her PhD research has led to four published scientific articles and has been presented at several scientific meetings.