The genetic etiology of familial breast cancer: Assessing the role of rare genetic variation using next generation sequencing

Hilbers, F.S.M.

The genetic etiology of breast cancer

Assessing the role of rare genetic variation using next generation sequencing

Florentine Hilbers

The genetic etiology of familial breast cancer

Assessing the role of rare genetic variation using next generation sequencing

Florentine Hilbers

Cover & layout: Florentine Hilbers
Printing: Print Service Ede - www.proefschriftenprinten.nl

The work presented in this thesis was financially supported by the Dutch Cancer Society.

ISBN: 978-90-830704-2-1

The genetic etiology of familial breast cancer

Assessing the role of rare genetic variation using next generation sequencing

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 7 juli 2020
klokke 13:45 uur

door
Florentine Susanna Maria Hilbers
Geboren te Wilnis in 1986

Promotoren
Prof. dr. P. Devilee
Prof. dr. C.J. van Asperen

Leden promotiecommissie
Prof. dr. H. van Attikum
Prof. dr. J. Morreau
Dr. A.R. Mensenkamp, Radboud Universtitair Medisch Centrum
Dr. M.A. Rookus, Nederlands Kanker Instituut

Table of Contents

Chapter 1

General introduction

1 Clinical aspects of familial breast cancer

1.1 Breast cancer tumorigenesis

Healthy breast tissue consists of a broad range of cell types meticulously arranged into highly organized structures of branched ducts and lobules supported by fatty and fibrous connective tissue. Homeostasis is maintained by tight regulation of the morphology and proliferation of each cell via both intra- and extracellular signals. These signals comprise multiple barriers which a cell has to overcome in order to become malignant i.e. to be able to survive and proliferate, and potentially invade and metastasize, regardless of external signals. To achieve this, a cell has to acquire, among others, the capabilities to ensure continued growth signaling, ignore growth inhibition and apoptotic signals, adjust its energy metabolism and stimulate angiogenesis.[1] Traditionally this malignant behavior was thought to be the result of the accumulation of somatic genetic changes within a single cell: the so-called somatic mutation theory. Nowadays, it is appreciated that the process of tumorigenesis is much more complex and involves changes at various organization levels and a dynamic reciprocity between them, including genetic mutations, epigenetic changes and alteration of the microenvironment.[2–4]

Although no longer thought to be the only facet in tumorigenesis, somatic genetic alterations that critically contribute to malignancy, the so-called driver mutations, remain one of the most well studied aspects of breast tumors. Genes frequently affected by mutations or structural variants include *TP53*, *PIK3CA* and *GATA3*.[5] However, the mutational landscape of breast tumors is far from homogeneous; a large diversity in the combination of alterations in over 50 significantly mutated genes has been found.[6–11] Besides genetic alterations, it has become clear that epigenetic changes play an important role in breast tumor development as well. For many cancers, including those of the breast, a global increase in CpG-island hypermethylation has been observed. Methylation of the CpG-islands in a promoter region generally leads to reduced expression of the corresponding gene. Interestingly, hypermethylation of promoter regions is especially frequently seen in tumor suppressor genes, including *CDH1*, *CDKN2A*, *PTEN* and *BRCA1* (reviewed in [12,13]). Nonetheless, with regard to epigenetic alterations, substantial heterogeneity has been observed between tumors: a large study analyzing genome-wide methylation data found that breast tumors can be classified into at least five different methylation groups by unsupervised clustering.[7]

Another important factor in breast cancer tumorigenesis is the microenvironment. The tumor microenvironment comprises the extracellular matrix and multiple cell types such as fibroblasts and immune cells. Moreover, it acts as the medium for many important soluble factors such as cytokines, growth factors and enzymes. The microenvironment can both inhibit and facilitate the malignant behavior of the tumor cells.[4] For example, while normal myoepithelial cells in the tumor microenvironment may inhibit the growth of breast tumor cells, cancer-associated fibroblasts can promote growth and invasion.[14]

1.2 Breast cancer subtypes

Although all tumors in the end acquire approximately the same set of capabilities, the examples described above illustrate the tremendous variation in the way they achieve this. Because of this heterogeneity, breast tumors are often classified in subtypes. The large majority of breast tumors, roughly 95%, are adenocarcinomas arising in the breast epithelium, while a small percentage consists of sarcomas originating from the stromal cells of the connective tissue of the breast. Adenocarcinomas are further classified according to their morphological and cytological patterns. By far the largest group are invasive ductal carcinomas of "no special

type", which is a diagnosis of exclusion used for tumors that do not possess characteristics that would classify them as one of the special subtypes. These special subtypes comprise ~25% of all breast tumors and consist, among others, of lobular, tubular, cribiform and metaplastic carcinomas.[15] Histological subtypes can be associated with specific molecular characteristics, for example, while lobular carcinomas have fewer somatic genetic aberrations overall, they almost always lose E-cadherin function, through inactivating mutations or promotor hypermethylation of *CDH1* (the gene encoding E-Cadherin) or impaired integrity of the E-cadherin-catenin complex.[16]

Another important factor used in defining breast cancer subtypes is whether or not they express the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and proliferation marker Ki67. The presence of these proteins is routinely assessed in clinical practice by immunohistochemistry and tumors are subsequently classified into four subtypes: Luminal A (ER+/PR+/HER2-/Ki67 low), Luminal B (ER+/PR+/HER2-/Ki67 high), HER2-overexpressing (HER2+, any ER/PR/Ki67) and basal-like (ER-/PR-/HER2-, any Ki67).[17] The presence of the three receptors indicate the pathways through which the cells receive growth signals and, more importantly, offer opportunities for therapeutic targeting. While HER2 overexpression in breast tumors was once associated with a poor prognosis, this drastically changed when the first HER2-targeting agent, trastuzumab, became available.[18] Also the dependence of ER-positive/PR-positive breast cancer on estrogen signalling can be targeted by several drugs, such as selective estrogen receptor modulators (e.g. tamoxifen) and aromatase inhibitors (e.g. letrozole). These therapies are not indicated for breast cancers lacking all three receptors, known as "triple negative" breast cancer, which are associated with poor prognosis. For this group of tumors, chemotherapy has long been the only available treatment option. However, recently clinical trials have started exploring immunotherapy as a treatment option for triple negative breast cancer as these tumors have been shown to be more immunogenic, with higher PD-L1 expression and T-cell infiltration than other subtypes.[19]

Besides these subtypes commonly used in clinical practice, further classification of breast tumors has been proposed based on gene expression analysis and mutational signatures. Gene expression analysis has resulted in a similar but more detailed range of subtypes, including luminal A, luminal B, HER2-enriched, basal-like, normal-like, Claudin-low and apocrine (reviewed in [20]). These subtypes are often called the "intrinsic subtypes". The luminal A, luminal B, HER2-enriched and basal-like subtypes largely overlap with their immunohistochemistry counterparts. The claudin-low subtype is thought to represent a group of tumors characterized by epithelial-to-mesenchymal transition and stem cell like features and which are usually negative for ER, PR and HER2.[21] The apocrine expression subtype represents another subset of triple negative breast cancers which express the androgen receptor.[22] Mutational signatures describe the kind of nucleotide substitutions and structural variations that are overrepresented in a tumor. These signatures are thought to provide insight into the mutagenic processes and possible DNA repair deficiencies that have shaped the tumor's genome. For breast cancer twelve different nucleotide substitution and six rearrangement signatures have been described, some of which have been linked to specific tumor attributes, such as deficiency in homologous recombination or mismatch repair.[6,23]

1.3 Breast cancer risk and familial clustering of breast cancer

Breast cancer is the most common cancer diagnosis and represents the most common cancer-associated cause of mortality in women worldwide.[24] In the Netherlands, approximately one in eight women will develop breast cancer at some point during her life.[25] The identification of risk factors is important for both prevention and the development of efficient screening programs aimed at early detection. However, the etiology of breast cancer is complex; among others, genetics, physical characteristics, lifestyle and reproductive factors are known to affect disease risk.[26] Gender is an obvious factor in breast cancer; the life time risk to develop breast cancer is approximately 13% for females, while for males it is only 0.11%.[25] Geographical location is another important factor associated with the risk of developing breast cancer. The incidence of breast cancer is remarkably higher in "western" countries (Europe, North America, Australia and New Zealand) compared to the rest of the world.[24] The incidence of breast cancer, alike that of many other cancers, also increases with age.[25] This association between cancer and age is generally attributed to the time needed for the stochastic process of accumulating the tumorigenic capabilities described above, through somatic mutations, epigenetic changes, telomere shortening, declined DNA repair efficiency and changes in the microenvironment.[27] In addition, early menarche and late menopause are associated with an increased risk of breast cancer.[28] The estrogen and progesterone fluctuation during menstrual cycles induces repetitive phases of mammary epithelium proliferation and regression, which cause an increased chance of genetic errors.[29] At the same time, reproductive factors, such as a first-full term pregnancy at relatively young age, a higher number of full-term pregnancies and total duration of breast feeding are associated with an decreased risk of breast cancer.[30] Pregnancy and lactation are thought to induce several, long-term, systemic and local changes that could explain their associated with a decreased breast cancer risk, which include changes in circulating hormone levels, estrogen responsiveness, number of mammary stem cells and differentiation status.[31] Furthermore, several lifestyle factors such as alcohol consumption, lack of physical activity, post-menopausal obesity and exposure to exogenous estrogen via oral contraception or hormone replacement therapy increase the risk of breast cancer (reviewed in [26]).

Family history is another well-established risk factor for breast cancer. Women who have one first-degree relative diagnosed with breast cancer, have a relative risk (RR) of approximately 1.8 to develop breast cancer themselves. Having three or more affected first degree relatives is associated with a RR of 3.9.[32,33] Moreover, a lower age of onset is associated with a higher risk in first-degree relatives compared to late onset breast cancer.[32,33] This relative risk associated with a family history is known as the familial relative risk (FRR). Several algorithms have been developed to calculate an individual's risk of breast cancer based on a specific family history, including BOADICEA,[34] BRCAPRO,[35] the Tyrer-Cuzick model,[36] the Claus model[37] and Gail model.[38] Studies in monozygotic and dizygotic twins have determined that genetic factors account for approximately 27% of the variance in breast cancer susceptibility. Although genetic factors likely play at least some role the etiology of every breast cancer case, only ~15% of cases are considered "familial". Of these familial breast cancer cases, approximately 5-10% carry a mutation in a known high-risk breast cancer susceptibility gene. Breast cancer susceptibility loci associated with a moderate or small risk increase have also been identified (see sections 2.1, 2.2 and 2.3 below). However, most familial breast cancer cases are not yet tested for these moderate and low-risk loci in diagnostic settings in the Netherlands and most EU countries. Moreover, all known genetic risk factors jointly still explain less than half of the familial relative risk.

1.4 Clinical management of hereditary and familial breast cancer

When a woman has multiple first or second-degree family members affected with breast cancer, genetic testing for mutations in breast cancer susceptibility genes is indicated. Moreover, if a woman herself has been diagnosed with breast cancer at a particularly young age, or with both breast and ovarian cancer or with bilateral disease, referral for genetic counseling might be appropriate. (See [39–41] for the exact indications for genetic counseling in the Netherlands.) When genetic testing is warranted, at a minimum *BRCA1* and *BRCA2* will be tested. Nowadays, however, gene panels containing additional moderate and high-risk susceptibility genes (see section 2.1 and 2.2) are commonly assessed using next generation sequencing (NGS, also known as massively parallel sequencing (MPS)). The results of these genetic tests can be classified in three categories. In the far majority, no potentially causal mutation will be detected. On the other hand, there are cases in which a mutation is found that is clearly associated with an increased risk of breast cancer. Besides these straightforward results, there is a third category of cases in which a genetic variant is identified for which the association with breast cancer risk is unclear. These variants are referred to as variants of uncertain significance (VUS). For *BRCA1* and *BRCA2* testing alone, in 5-10% of cases a clearly pathogenic mutation is found, while in approximately 15% a VUS detected.

With regard to clinical surveillance, a distinction is made between "hereditary" breast cancer, when a causal mutation has been found, and "familial" breast cancer, when either no mutation or a VUS has been detected. Guidelines for the clinical management of familial and hereditary breast cancer describe screening strategies including age to start screening, the frequency and the methods to be employed.[39–42] In the Netherlands, there are separate guidelines for families with *BRCA1*, *BRCA2*, *PALB2*, and *CHEK2* mutations in addition to guidelines for rare syndromes associated with an increased breast cancer risk such as Li-Fraumeni syndrome. Women from breast cancer families in which no clearly pathogenic mutation has been detected are classified based on family history in either moderate high (RR 2-3) or high risk (RR 3-4). Women at higher risk start screening at younger age and at higher frequency compared with women in lower risk categories. In addition, for women at a very high risk, risk reduction via prophylactic mastectomy or salpingo-oophorectomy is an option. Woman who carry a high-risk mutation might also opt for pre-implantation genetic diagnosis (PGD) also called embryo selection, to ensure that the predisposing mutation is not passed on to their children.

Given these options for prevention and early detection, it can make a dramatic difference for women from breast cancer families to know their mutational status, as women from families in which no causal mutation has been detected face much larger uncertainties. While in a hereditary breast cancer family mutation testing can identify the family members carrying the high risk variant, sisters within a familial breast cancer family are all assumed to have the same risk and all receive the same screening advice. In addition, the magnitude of the increased risk might be more uncertain. Risks for *BRCA1* and *BRCA2* mutation carriers have been determined based on studies with large numbers of carriers.[43–45] Although these risk estimates are likely to have been biased by the enrichement of carriers with a familial in these studies. In familial cancer clinics, risk in women from familial breast cancer families is presently estimated on the basis of family history alone; these estimates derive from large case-control analyses[32,33] and have been incorporated into models mentioned in paragraph 1.3. As the number of affected female relatives is an important variable in all commonly used risk prediction models, small families or families with few women are less informative and risk estimates might be less precise, and possibly underestimated in these cases. Consequently,

decisions on preventive measures are much more complicated in the presence of this uncertainty about risk.

2 The genetic landscape of breast cancer

2.1 High risk breast cancer susceptibility genes

Genetic variants associated with breast cancer risk are often subdivided into categories according to the magnitude of the risk increase and their population allele frequency. Although there is no generally accepted cut-off, genetic variants that are associated with a risk that is more than three times that of the general population are often considered high risk in clinical guidelines in the EU and the USA. This generally translates to a lifetime risk of higher than 30%. The two most well-known high-risk breast cancer susceptibility genes, *BRCA1* and *BRCA2*, were discovered after linkage analysis and subsequent sequencing in multi-case breast cancer families. Mutations in *BRCA1* and *BRCA2* are associated with a risk to develop breast cancer before the age of 70 of approximately 60% and 55%, respectively.[43–45] Both genes also give an increased risk of ovarian cancer with a 59% and 16% risk by age 70 for *BRCA1* and *BRCA2* respectively. In addition mutations in *BRCA1* and *BRCA2* have been linked to pancreatic[46–48] and prostate cancer.[49–51] The chance for a male mutation carrier to develop breast cancer before age 70 is approximately 1.2% for *BRCA1* and 6.8% for *BRCA2*.[52] Interestingly, while tumors arising in *BRCA1* mutation carriers are strongly enriched for the "triple negative" (lacking the receptors ER, PR and HER2) phenotype as well as basal-like molecular subtype expression profiles, *BRCA2* associated tumors are much more heterogeneous[53] and akin to "sporadic" breast cancer. Yet, both *BRCA1* and *BRCA2* play a crucial role in repair of DNA double-strand breaks via homologous recombination. In addition, *BRCA1* is involved in several cell cycle checkpoints that prevent a cell with DNA damage from entering mitosis.[54] Consequently, mutations in *BRCA1* and *BRCA2* are thought to contribute to tumorigenesis via accelerated accumulation of somatic mutations.

More recently, *PALB2* has been established as a breast cancer susceptibility gene.[55,56] The risk of developing breast cancer before the age of 70 for a carrier of a protein-truncating PALB2 mutation is approximately 35%. Moreover, in the context of a family history with two affected first-degree relatives the life-time risk for breast cancer is increased to 58%. In addition to increasing the risk of breast cancer, protein-truncating variants (PTV) in *PALB2* are associated with an increased pancreatic cancer risk.[57] Like mutations in *BRCA2*[58] and specific, more moderate-risk, mutations in *BRCA1*,[59] homozygous or compound-heterozygous mutations in *PALB2* cause Fanconi Anaemia,[60,61] a disease characterized by bone marrow failure, congenital anomalies and predisposition to several malignancies. Interestingly, PALB2 binds directly to both BRCA1 and BRCA2[62] and, like its binding partners, is involved in homologous recombination.

In addition to variants in these three genes, variants associated with a high risk for developing breast cancer risk can be found in *CDH1*,[63–65] *PTEN*,[66] *STK11*[67,68] and *TP53*.[69] In contrast to the variants in *BRCA1*, *BRCA2* and *PALB2*, variants in these genes cause cancer syndromes, defined by an increased risk for multiple types of cancer and sometimes other phenotypical features. It is therefore rare to find mutations in these genes in families with only an increased occurrence of breast cancer.[70–74] Due to the relative low numbers of patients with these syndromes, risk estimates vary widely, but all of these genes are generally regarded as high-risk breast cancer susceptibility genes. Biologically, these genes play very diverse roles in the cell. *TP53* is the most frequently somatically mutated gene in human cancer.

Via the regulation of expression of several genes it can, among others, activate DNA repair, arrest the cell cycle or induce apoptosis upon DNA damage.[75] PTEN is a negative regulator survival and proliferation via inhibition of the PI3K/AKT pathway.[76] While STK11 and CDH1 are important for cell polarity, cell-cell adhesion and energy metabolism, thereby suppressing cell proliferation and migration.[77–80]

2.2 Moderate risk breast cancer susceptibility genes

A second category of risk alleles consists of variants in moderate risk genes associated with a two- to four-fold increased risk. Well established moderate risk variants have been found in *ATM*,[81] *CHEK2*,[82] *BARD1*,[83–85] *FANCM*,[86,87] *NBS1*[88] and *RECQL*.[89,90] Variants in these genes can be relatively common in the general population. This is for example the case for the c.1100delC variant in *CHEK2*, which has an allele frequency of ~1% in north-western Europe.[82] However, for most of these genes pathogenic variants are still very rare, which makes estimating risks challenging. Moreover, it has been shown that breast cancer risks can vary between variants within the same gene. For example, while most variants in *ATM* are thought to be associated with an approximately 2- to 3-fold increased risk, a specific missense variant, c.7271T>G (p.V2424G), has been associated with risks 8-10 times as high as in the general population.[91–93] This strongly increased risk is probably caused by a dominant-negative effect of this specific variant on ATM protein function. [94]

The relatively moderate increase in risk associated with genetic variants in these genes, also makes the link with a family history of breast cancer less strong. Contrarily to the high-risk genes discussed above, most of these moderate risk genes have therefore not been identified via family-based linkage analysis. *ATM* and *NBS1*, which cause the rare recessive disorders ataxia-telangiectasia and Nijmegen breakage syndrome respectively, have been identified as breast cancer susceptibility genes because first-degree relatives of patients diagnosed with these syndromes had a markedly increased incidence of breast cancer. For the other moderate risk genes an increased breast cancer risk was observed after "candidate-gene" re-sequencing, meaning that genes with a function similar to that of *BRCA1* and *BRCA2* have selectively been tested for an association in either family-based or case-control studies. It is therefore not surprising that all of these genes have a function in DNA damage response because this strongly determined their candidacy. In addition to the genes mentioned above, there are a number of genes for which a link with breast cancer has been suggested, but not yet convincingly established. These include: *FAM175A*,[95] *MEN1*,[96] *MRE11A*,[97,98] *MSH6*,[85,99,100] *NF1*,[85,101] *RAD50*,[98,102–104] *RAD51C*,[85,99,105–110] *RAD51D*,[85,99,111,112] *RINT1*[113] and *XRCC2*.[114–116]

Until recently, moderate-risk genes were not routinely assessed in clinical practice. Therefore, the exact contribution of variants in these genes to familial breast cancer remained uncertain. Lately, two large studies reporing the results of gene panel testing of (familial) breast cancer cases found that mutations in these moderate risk genes are present in approximately 4% of the tested breast cancer patients.[85,117] Due to the relative rarity of these variants no specific guidelines for clinical management of carriers exist for the moderate risk genes other than *CHEK2*. Therefore, the breast cancer risk for an individual carrier has to be estimated from the family history and the risk associated with the mutation (for *ATM*, *CHEK2* and *PALB2* this can now also be done using BOADICEA[118]). After this, the guidelines for familial breast cancer can be followed.

2.3 Low risk breast cancer susceptibility alleles

A last category of breast cancer susceptibility alleles is formed by those associated with low RRs, usually between 0.7 and 1.3. Currently there are more than 300 single nucleotide polymorphisms (SNPs) found to be associated with breast cancer overall (i.e., irrespective of subtype), while a handful of variants are associated with specific subtypes of breast cancer.[119] This type of susceptibility alleles can be very common, with most risk alleles being present in more than five percent of the general population. Virtually all low risk genetic variants have been identified with so-called genome wide association studies (GWAS) studies. These are large case-control studies in which hundreds of thousands of single nucleotide polymorphisms are genotyped in each individual. Associations with breast cancer are calculated not only for the directly genotyped SNPs, but also for SNPs in close proximity that are "tagged" due to linkage disequilibrium and can therefore be imputed on the basis of a haplotype reference panel. This linkage disequilibrium only exists between SNPs with similar allele frequencies. Most SNP arrays are therefore not suitable to detect the risk associated with rare variants with allele frequencies <5%.

The linkage disequilibrium between SNPs also makes it challenging to pinpoint the causal variant in a region with multiple associated SNPs. So-called "fine-mapping" studies, often combined with functional in vitro assays of the variant, try to discover the causal variant by genotyping a larger number of variants within a susceptibility locus in more ethnically diverse populations.[120] Interestingly, most low risk genetic variants known today are not located in the protein-encoding regions of the genome. Low-risk variants for which a mechanism has been unraveled are often present in regulatory regions and affect gene expression. Such effects have for example been shown for a locus on 11q13. In this region two functional SNPs were uncovered: one reducing the binding of transcription factor *ELK4* to an enhancer region, the other increasing binding of GATA3 to a silencer. The risk alleles of both SNPs reduced transcriptional activity of *CCND1*.[121] Although *CCND1* has traditionally been considered to be a oncogene, it also promotes the recruitment of RAD51 to double strand breaks and reduction of CCND1 levels inhibits homologous recombination.[122] More complex mechanisms are likely to be revealed, as single susceptibility loci can affect multiple genes in both cis and trans, and not only in the tissue from which the tumor arises.[123] This is in line with the complex interactions between cell types within a tumor mass during tumorigenesis (see 1.1).

Due to the large number of low-risk variants and the small risks associated with them, the genotype of an individual for a single SNP has very little positive prediction power for the occurrence of breast cancer. Currently, these low-risk variants do not have any clinical implications. However, several studies have shown that these SNPs can be combined multiplicatively into a single risk score which predicts risk in a much more discriminatory way.[124,125] A recent modeling study has suggested that tailoring population-based screening programs for breast cancer based on polygenic risk scores could improve the cost-effectiveness and benefit-to-harm ratio of such programs.[126] Several ongoing clinical trials are assessing if screening strategies can be improved by considering an individual's genetic risk (e.g. the WISDOM study (NCT02620852) and the MyPeBS study (NCT03672331)). In general, it would be valuable to integrate the low-risk variants with the rare moderate and high risk variants and non-genetic factors into risk-prediction models so that their accuracy to predict breast cancer increases. It has already been shown that low-risk variants can be combined in to a model with moderate risk gene *CHEK2*.[127] The effect of these low risk variants in *BRCA1* or *BRCA2* mutation carriers has also been studied extensively. Many of the genetic variants that

are associated with small increases in breast cancer in the general population, have similar effects in carriers.[128] Moreover, several polymorphism have been identified that specifically modify risk for *BRCA1* and *BRCA2* mutation carriers.[129]

3. The missing heritability of breast cancer and the potential role of next generation sequencing

3.1 What type of genetic variants could explain the missing heritability

Despite countless studies in very large numbers of familial breast cancer cases and controls, more than half of the FRR remains unexplained. Of note, the techniques employed for the discovery of breast cancer susceptibility alleles have had a substantial impact on the types of genetic variants that studies have been able to discover. Genetic linkage studies, which statistically weigh co-segregation of genomic regions with disease within families, are only able to detect regions harboring high-risk breast cancer alleles in extended "informative" pedigrees. GWAS on the other hand, by nature of their design are limited to susceptibility alleles that are relatively common but can detect weak associations of small increased risk. They perform very well in situations of genetic heterogeneity. Most SNP arrays genotype and tag SNPs with a minor allele frequency of 5% or higher in the general population. The number of cases and controls in a study, together with the MAF of a SNP dictate the effect sizes and allele frequencies that can be detected. The largest GWAS study to date has included 122,977 cases and 105,974 controls and was able to detect significant ORs as low as 1.03 and associations with SNPs with a MAF of 1%.[130] However, the MAF of variants associated with a larger increase in risk (OR >2) are typically much lower than 1% and can therefore not be detected with the current GWAS and imputation strategies. Candidate gene re-sequencing studies are able to detect moderate and high-risk breast cancer alleles. The statistical power of this approach is strongly determined by the number of genes tested and the sample-size. Selecting familial cases, in which breast cancer risk alleles are thought to be enriched, and comparing the allele frequency in this population with that in population controls, allows for the dection of moderate and high risk alleles with population frequencies of less than 0.5%. Moreover, the focus on a single gene reduces issues with multiple testing and lowered p-value cut-offs to correct for that. In this way, moderate risks genes can be detected, without the need for very large studies. A clear downside, however, is that selecting candidates depends heavily on our assumptions on the pathways and genes involved in breast cancer susceptibility.

After the eras of linkage studies, candidate gene re-sequencing and GWAS, two main unexplored areas of genetic variation remained in which the missing heritability of breast cancer might reside. First of all, additional risk alleles might be found among the very rare genetic variants. The allele frequency cut-off for variants not detectable by the studies published to date, ranges from smaller than ~1% for low-risk polymorphisms to smaller than ~0.01% combined frequency in areas with high-risk alleles. In order to explore this area, larger case-control studies are needed. In addition, selecting phenotypically more homogenous groups of (familial) cases, might increase the efficiency of the study by also selecting for a more homogeneous genetic etiology. The second main unexplored area consists of moderate-risk variants in regions currently not linked with pathways involved in DNA damage repair. To explore this area, we need to be able to detect relatively rare (>0.1%) and potentially novel variants in all gene-coding regions of the genome. Moreover, this needs to be done fast and cost-effectively enough to be able to assess multiple cases from multiple unexplained breast

cancer families. Also for these variants, it will be important to select phenotypically more homogenous groups of cases to increase the efficiency of the study.

3.2 Challenges in the use of NGS to discover novel breast cancer risk alleles

By sequencing millions of short DNA fragments in parallel, NGS makes it possible to analyze large parts of the genome in a timely manner. However, although costs of NGS continue to decline, they still warrant carefully designed studies to increase the cost-effectiveness. One common choice is to focus on the protein coding regions of the genome, also known as the exome. This still enables searching for new breast cancer susceptibility genes without first narrowing down the list of genes of interest, allowing for a more agnostic view on which genes might be involved in breast cancer susceptibility. However, by focusing on the protein encoding regions it becomes more difficult to accurately detect copy number and structural variation, because of this many studies limit themselves to single nucleotide variants and small insertions and deletions. Moreover, it is usually not possible to assess the association of genetic variants with breast cancer within the limited number of individuals that have been exome sequenced. Therefore, many studies opt for a two stage design, where potentially interesting variants are discovered in a relatively small number of families using NGS, after which an association with breast cancer is assessed in a much larger group of familial cases and controls employing different genotyping techniques. However, the relatively small number of familial cases available and the need to control for multiple testing in the statistical analysis, make that only a hand full of variants can be tested with sufficient power in the second phase of the study. Given that, on average, an exome of a person from European descent contains ~12,000 non-synonymous variants,[131] selection of potentially interesting variants is not a trivial task and will often depend on assumptions about the features of a causal variant.

When narrowing down the list of potentially interesting genetic variants, there are several characteristics that can be taken into account. A common selection factor is to assess the severity of the variant, assuming that variants leading to a completely inactivated protein are more likely to cause disease than variants that only partially impair protein function, and that these variants are in turn more likely to cause disease than synonymously coding variants. Thus PTVs are generally considered pathogenic whereas the severity of missense variants is commonly assessed by in silico prediction tools, which take into account information on factors such as evolutionary conservation, known functional protein-domains, three-dimensional structure and the characteristic of the changed amino acid. These tools, however, have various limitations. For example, a variant that truly affects an exonic splice enhancer is likely to be wrongly classified as "benign" because of our limited abilities to predict this effect. Another commonly assessed variant characteristic is the allele frequency of a variant in reference datasets containing data from the general population. For example, if a variant is relatively frequent in the general population it cannot be associated with a high risk of breast cancer. Contrarily, if a variant is detected in multiple families within a study, while being rare or absent in the general population, this can be an indication to select a variant for further analysis. After the selection of a list of potentially causal variants, the next step depends on the availability of additional DNA samples. If the family of the index case has been extensively sampled, co-segregation analysis is a powerful way to assess the association with breast cancer. However, often only a limited number of family members can be assessed causing this analysis to be inconclusive. In this case, additional (familial) cases and controls will need to be genotyped.

4 Scope and outline of this thesis

This thesis aims to contribute to our understanding of the genetic etiology of breast cancer with the help of next generation sequencing. It will focus on families with a clear clustering of breast cancer, but in which no mutations in *BRCA1* or *BRCA2* have been detected. This thesis intends to give new insights into the genetic factors that are responsible for the clustering of breast cancer and provide clues for a better risk prediction in these families.

Chapter 2 of this thesis describes an exome sequencing effort in six breast cancer families. In order to select a potentially more homogeneous set of breast cancer cases, families were selected in which multiple women had tumor that showed a specific array CGH profile.

Chapter 3 reports on a large international case-control study that aimed to validate *XRCC2* as a new breast cancer susceptibility gene. This gene had recently been discovered using exome sequencing.

Chapter 4 reports on the functional analysis of missense variants detected in *XRCC2*. By selecting those variants that affect XRCC2 function, a more accurate burden analysis could be performed on the data of the case-control study described in chapter 3.

Chapter 5 describes an exome sequencing effort in families with an potential recessive mode of inheritance. All families selected for this study had at least three siblings affected with breast cancer and no breast cancer cases in first degree relatives from the previous of following generation. This study combined the exome sequencing results with haplotype-sharing data to more efficiently filter the genetic variants.

Chapter 6 is a review of the literature on the methods to determine the role of extremely rare genetic variants in familial breast cancer.

Chapter 7 discusses the main findings of this thesis, their potential consequences, clinical implications and future directions.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
2. Vineis P, Schatzkin A, Potter JD. Models of carcinogenesis: an overview. Carcinogenesis. 2010;31(10):1703-1709. doi:10.1093/carcin/bgq087
3. Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell. 2013;153(1):38-55. doi:10.1016/j.cell.2013.03.008
4. Place AE, Jin Huh S, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. Breast Cancer Res BCR. 2011;13(6):227. doi:10.1186/bcr2912
5. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(Database issue):D805-811. doi:10.1093/nar/gku1075
6. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486(7403):400-404. doi:10.1038/nature11017
7. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70. doi:10.1038/nature11412
8. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012;486(7403):405-409. doi:10.1038/nature11154
9. Dorman SN, Viner C, Rogan PK. Splicing mutation analysis reveals previously unrecognized pathways in lymph node-invasive breast cancer. Sci Rep. 2014;4:7063. doi:10.1038/srep07063
10. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 2012;486(7403):353-360. doi:10.1038/nature11143
11. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395-399. doi:10.1038/nature10933
12. Stefansson OA, Esteller M. Epigenetic modifications in breast cancer and their role in personalized medicine. Am J Pathol. 2013;183(4):1052-1063. doi:10.1016/j.ajpath.2013.04.033
13. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128(4):683-692. doi:10.1016/j.cell.2007.01.029
14. Hu M, Yao J, Carroll DK, et al. Regulation of in situ to invasive breast carcinoma transition. Cancer Cell. 2008;13(5):394-406. doi:10.1016/j.ccr.2008.03.007
15. SR L, IO E, SJ S, PH T, MJ van de V. WHO Classification of Tumours of the Breast. http://publications.iarc.fr/Book-And-Report-Series/Who-Iarc-Classification-Of-Tumours/WHO-Classification-Of-Tumours-Of-The-Breast-2012. Accessed September 24, 2018.
16. Rakha EA, Patel A, Powe DG, et al. Clinical and biological significance of E-cadherin protein expression in invasive lobular carcinoma of the breast. Am J Surg Pathol. 2010;34(10):1472-1479. doi:10.1097/PAS.0b013e3181f01916
17. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn H-J. 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. 2011;22(8):1736-1747. doi:10.1093/annonc/mdr304
18. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med. 2005;353(16):1659-1672. doi:10.1056/NEJMoa052306
19. Katz H, Alsharedi M. Immunotherapy in triple-negative breast cancer. Med Oncol Northwood Lond Engl. 2017;35(1):13. doi:10.1007/s12032-017-1071-6
20. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol. 2011;5(1):5-23. doi:10.1016/j.molonc.2010.11.003
21. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res BCR. 2010;12(5):R68.

doi:10.1186/bcr2635
22. Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular apocrine breast tumours by microarray analysis. Oncogene. 2005;24(29):4660-4671. doi:10.1038/sj.onc.1208561
23. Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54. doi:10.1038/nature17676
24. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424. doi:10.3322/caac.21492
25. IKNL. Cijfers over kanker, Nederlandse Kankerregistratie. http://www.cijfersoverkanker.nl/. Accessed July 23, 2017.
26. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? J Cell Mol Med. 2005;9(1):208-221.
27. DePinho RA. The age of cancer. Nature. 2000;408(6809):248-254. doi:10.1038/35041694
28. Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. Lancet Oncol. 2012;13(11):1141-1151. doi:10.1016/S1470-2045(12)70425-4
29. Atashgaran V, Wrin J, Barry SC, Dasari P, Ingman WV. Dissecting the Biology of Menstrual Cycle-Associated Breast Cancer Risk. Front Oncol. 2016;6:267. doi:10.3389/fonc.2016.00267
30. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet Lond Engl. 2002;360(9328):187-195. doi:10.1016/S0140-6736(02)09454-0
31. Dall G, Risbridger G, Britt K. Mammary stem cells and parity-induced breast cancer protection- new insights. J Steroid Biochem Mol Biol. 2017;170:54-60. doi:10.1016/j.jsbmb.2016.02.018
32. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer. 1997;71(5):800-809.
33. 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 Lond Engl. 2001;358(9291):1389-1399. doi:10.1016/S0140-6736(01)06524-2
34. Antoniou AC, Pharoah PDP, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer. 2002;86(1):76-83. doi:10.1038/sj.bjc.6600008
35. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. Am J Hum Genet. 1998;62(1):145-158.
36. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004;23(7):1111-1130. doi:10.1002/sim.1668
37. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet. 1991;48(2):232-242.
38. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81(24):1879-1886.
39. STOET, VKGN. Landelijke Richtlijn Familiair Mamma/Ovariumcarcinoom.; 2010. http://www.oncoline.nl/familiair-mamma-ovariumcarcinoom.
40. STOET. Erfelijke En Familiaire Tumoren: Richtlijnen Voor Diagnostiek En Preventie.; 2017. https://www.stoet.nl/wp-content/uploads/2017/04/STOET-Richtlijnenboekje-april2017_DEF.pdf.
41. NABON. Landelijke Richtlijn Mammacarcinoom.; 2017. http://www.oncoline.nl/mammacarcinoom.
42. STOET, VKGN. Landelijke Richtlijn Hereditair Mamma/Ovariumcarcinoom.; 2010. http://www.

oncoline.nl/hereditair-mamma-ovariumcarcinoom.
43. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol Off J Am Soc Clin Oncol. 2007;25(11):1329-1333. doi:10.1200/JCO.2006.09.1066
44. Mavaddat N, Peock S, Frost D, et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst. 2013;105(11):812-822. doi:10.1093/jnci/djt095
45. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317(23):2402-2416. doi:10.1001/jama.2017.7112
46. Murphy KM, Brune KA, Griffin C, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res. 2002;62(13):3789-3793.
47. Hahn SA, Greenhalf B, Ellis I, et al. BRCA2 germline mutations in familial pancreatic carcinoma. J Natl Cancer Inst. 2003;95(3):214-221.
48. Couch FJ, Johnson MR, Rabe KG, et al. The prevalence of BRCA2 mutations in familial pancreatic cancer. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2007;16(2):342-346. doi:10.1158/1055-9965.EPI-06-0783
49. Lecarpentier J, Silvestri V, Kuchenbaecker KB, et al. Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. J Clin Oncol Off J Am Soc Clin Oncol. 2017;35(20):2240-2250. doi:10.1200/JCO.2016.69.4935
50. Thompson D, Easton DF, Breast Cancer Linkage Consortium. Cancer Incidence in BRCA1 mutation carriers. J Natl Cancer Inst. 2002;94(18):1358-1365.
51. Breast Cancer Linkage Consortium. Cancer risks in BRCA2 mutation carriers. J Natl Cancer Inst. 1999;91(15):1310-1316.
52. Tai YC, Domchek S, Parmigiani G, Chen S. Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst. 2007;99(23):1811-1814. doi:10.1093/jnci/djm203
53. Mavaddat N, Barrowdale D, Andrulis IL, et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2012;21(1):134-147. doi:10.1158/1055-9965.EPI-11-0775
54. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer. 2011;12(1):68-78. doi:10.1038/nrc3181
55. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 2007;39(2):165-167. doi:10.1038/ng1959
56. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497-506. doi:10.1056/NEJMoa1400382
57. Jones S, Hruban RH, Kamiyama M, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science. 2009;324(5924):217. doi:10.1126/science.1171202
58. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. Science. 2002;297(5581):606-609. doi:10.1126/science.1073834
59. Sawyer SL, Tian L, Kähkönen M, et al. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov. 2015;5(2):135-142. doi:10.1158/2159-8290.CD-14-1156
60. Xia B, Dorsman JC, Ameziane N, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet. 2007;39(2):159-161. doi:10.1038/ng1942
61. Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet. 2007;39(2):162-164. doi:10.1038/ng1947
62. Sy SMH, Huen MSY, Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci U S A. 2009;106(17):7155-7160. doi:10.1073/pnas.0811159106
63. Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402-405. doi:10.1038/32918

64. Keller G, Vogelsang H, Becker I, et al. Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. Am J Pathol. 1999;155(2):337-342. doi:10.1016/S0002-9440(10)65129-2
65. Pharoah PD, Guilford P, Caldas C, International Gastric Cancer Linkage Consortium. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 2001;121(6):1348-1353.
66. Nelen MR, van Staveren WC, Peeters EA, et al. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Hum Mol Genet. 1997;6(8):1383-1387.
67. Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature. 1998;391(6663):184-187. doi:10.1038/34432
68. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 1998;18(1):38-43. doi:10.1038/ng0198-38
69. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233-1238.
70. FitzGerald MG, Marsh DJ, Wahrer D, et al. Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. Oncogene. 1998;17(6):727-731. doi:10.1038/sj.onc.1201984
71. Ginsburg OM, Akbari MR, Aziz Z, et al. The prevalence of germ-line TP53 mutations in women diagnosed with breast cancer before age 30. Fam Cancer. 2009;8(4):563-567. doi:10.1007/s10689-009-9287-z
72. Guénard F, Pedneault CS-L, Ouellette G, et al. Evaluation of the contribution of the three breast cancer susceptibility genes CHEK2, STK11, and PALB2 in non-BRCA1/2 French Canadian families with high risk of breast cancer. Genet Test Mol Biomark. 2010;14(4):515-526. doi:10.1089/gtmb.2010.0027
73. Rapakko K, Allinen M, Syrjäkoski K, et al. Germline TP53 alterations in Finnish breast cancer families are rare and occur at conserved mutation-prone sites. Br J Cancer. 2001;84(1):116-119. doi:10.1054/bjoc.2000.1530
74. Schrader KA, Masciari S, Boyd N, et al. Germline mutations in CDH1 are infrequent in women with early-onset or familial lobular breast cancers. J Med Genet. 2011;48(1):64-68. doi:10.1136/jmg.2010.079814
75. Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. Nat Rev Cancer. 2014;14(5):359-370. doi:10.1038/nrc3711
76. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. Nat Rev Mol Cell Biol. 2012;13(5):283-296. doi:10.1038/nrm3330
77. Shackelford DB, Shaw RJ. The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. Nat Rev Cancer. 2009;9(8):563-575. doi:10.1038/nrc2676
78. Li J, Liu J, Li P, et al. Loss of LKB1 disrupts breast epithelial cell polarity and promotes breast cancer metastasis and invasion. J Exp Clin Cancer Res CR. 2014;33:70. doi:10.1186/s13046-014-0070-0
79. Bays JL, Campbell HK, Heidema C, Sebbagh M, DeMali KA. Linking E-cadherin mechanotransduction to cell metabolism through force-mediated activation of AMPK. Nat Cell Biol. 2017;19(6):724-731. doi:10.1038/ncb3537
80. Chen A, Beetham H, Black MA, et al. E-cadherin loss alters cytoskeletal organization and adhesion in non-malignant breast cells but is insufficient to induce an epithelial-mesenchymal transition. BMC Cancer. 2014;14:552. doi:10.1186/1471-2407-14-552
81. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38(8):873-875. doi:10.1038/ng1837
82. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-59. doi:10.1038/ng879
83. De Brakeleer S, De Grève J, Loris R, et al. Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. Hum Mutat.

2010;31(3):E1175-1185. doi:10.1002/humu.21200
84. Ratajska M, Antoszewska E, Piskorz A, et al. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. Breast Cancer Res Treat. 2012;131(1):89-97. doi:10.1007/s10549-011-1403-8
85. Couch FJ, Shimelis H, Hu C, et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol. 2017;3(9):1190-1196. doi:10.1001/jamaoncol.2017.0424
86. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci U S A. 2014;111(42):15172-15177. doi:10.1073/pnas.1407909111
87. Peterlongo P, Catucci I, Colombo M, et al. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. Hum Mol Genet. 2015;24(18):5345-5355. doi:10.1093/hmg/ddv251
88. Steffen J, Nowakowska D, Niwińska A, et al. Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. Int J Cancer. 2006;119(2):472-475. doi:10.1002/ijc.21853
89. Cybulski C, Carrot-Zhang J, Kluźniak W, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet. 2015;47(6):643-646. doi:10.1038/ng.3284
90. Sun J, Wang Y, Xia Y, et al. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet. 2015;11(5):e1005228. doi:10.1371/journal.pgen.1005228
91. Southey MC, Goldgar DE, Winqvist R, et al. PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet. 2016;53(12):800-811. doi:10.1136/jmedgenet-2016-103839
92. Bernstein JL, Teraoka S, Southey MC, et al. Population-based estimates of breast cancer risks associated with ATM gene variants c.7271T>G and c.1066-6T>G (IVS10-6T>G) from the Breast Cancer Family Registry. Hum Mutat. 2006;27(11):1122-1128. doi:10.1002/humu.20415
93. Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res BCR. 2011;13(4):R73. doi:10.1186/bcr2919
94. Waddell N, Jonnalagadda J, Marsh A, et al. Characterization of the breast cancer associated ATM 7271T>G (V2424G) mutation by gene expression profiling. Genes Chromosomes Cancer. 2006;45(12):1169-1181. doi:10.1002/gcc.20381
95. Solyom S, Aressy B, Pylkäs K, et al. Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions. Sci Transl Med. 2012;4(122):122ra23. doi:10.1126/scitranslmed.3003223
96. Dreijerink KMA, Goudet P, Burgess JR, Valk GD. Breast-Cancer Predisposition in Multiple Endocrine Neoplasia Type 1. N Engl J Med. 2014;371(6):583-584. doi:10.1056/NEJMc1406028
97. 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. 2008;2(4):296-316. doi:10.1016/j.molonc.2008.09.007
98. Damiola F, Pertesi M, Oliver J, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. Breast Cancer Res BCR. 2014;16(3):R58. doi:10.1186/bcr3669
99. Lu H-M, Li S, Black MH, et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. JAMA Oncol. August 2018. doi:10.1001/jamaoncol.2018.2956
100. Win AK, Lindor NM, Jenkins MA. Risk of breast cancer in Lynch syndrome: a systematic review. Breast Cancer Res BCR. 2013;15(2):R27. doi:10.1186/bcr3405
101. Seminog OO, Goldacre MJ. Age-specific risk of breast cancer in women with neurofibromatosis type 1. Br J Cancer. 2015;112(9):1546-1548. doi:10.1038/bjc.2015.78
102. Heikkinen K, Rapakko K, Karppinen S-M, et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis. 2006;27(8):1593-

1599. doi:10.1093/carcin/bgi360
103. Tommiska J, Seal S, Renwick A, et al. Evaluation of RAD50 in familial breast cancer predisposition. Int J Cancer. 2006;118(11):2911-2916. doi:10.1002/ijc.21738
104. Uhrhammer N, Delort L, Bignon Y-J. Rad50 c.687delT does not contribute significantly to familial breast cancer in a French population. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2009;18(2):684-685. doi:10.1158/1055-9965.EPI-08-0971
105. Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010;42(5):410-414. doi:10.1038/ng.569
106. Clague J, Wilhoite G, Adamson A, Bailis A, Weitzel JN, Neuhausen SL. RAD51C germline mutations in breast and ovarian cancer cases from high-risk families. PloS One. 2011;6(9):e25632. doi:10.1371/journal.pone.0025632
107. De Leeneer K, Van Bockstal M, De Brouwer S, et al. Evaluation of RAD51C as cancer susceptibility gene in a large breast-ovarian cancer patient population referred for genetic testing. Breast Cancer Res Treat. 2012;133(1):393-398. doi:10.1007/s10549-012-1998-4
108. Loveday C, Turnbull C, Ruark E, et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. Nat Genet. 2012;44(5):475-476; author reply 476. doi:10.1038/ng.2224
109. Thompson ER, Boyle SE, Johnson J, et al. Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. Hum Mutat. 2012;33(1):95-99. doi:10.1002/humu.21625
110. Blanco A, Gutiérrez-Enríquez S, Santamariña M, et al. RAD51C germline mutations found in Spanish site-specific breast cancer and breast-ovarian cancer families. Breast Cancer Res Treat. 2014;147(1):133-143. doi:10.1007/s10549-014-3078-4
111. Chen X, Li Y, Ouyang T, et al. Associations between RAD51D germline mutations and breast cancer risk and survival in BRCA1/2-negative breast cancers. Ann Oncol Off J Eur Soc Med Oncol. August 2018. doi:10.1093/annonc/mdy338
112. Loveday C, Turnbull C, Ramsay E, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet. 2011;43(9):879-882. doi:10.1038/ng.893
113. Park DJ, Tao K, Le Calvez-Kelm F, et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. Cancer Discov. 2014;4(7):804-815. doi:10.1158/2159-8290.CD-14-0212
114. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
115. Hilbers FS, Wijnen JT, Hoogerbrugge N, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. J Med Genet. 2012;49(10):618-620. doi:10.1136/jmedgenet-2012-101191
116. Hilbers FS, Luijsterburg MS, Wiegant WW, et al. Functional Analysis of Missense Variants in the Putative Breast Cancer Susceptibility Gene XRCC2. Hum Mutat. 2016;37(9):914-925. doi:10.1002/humu.23019
117. Kurian AW, Ward KC, Howlader N, et al. Genetic Testing and Results in a Population-Based Cohort of Breast Cancer Patients and Ovarian Cancer Patients. J Clin Oncol Off J Am Soc Clin Oncol. 2019;37(15):1305-1315. doi:10.1200/JCO.18.01854
118. Lee AJ, Cunningham AP, Tischkowitz M, et al. Incorporating truncating variants in PALB2, CHEK2, and ATM into the BOADICEA breast cancer risk model. Genet Med Off J Am Coll Med Genet. 2016;18(12):1190-1198. doi:10.1038/gim.2016.31
119. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
120. Rivandi M, Martens JWM, Hollestelle A. Elucidating the Underlying Functional Mechanisms of Breast Cancer Susceptibility Through Post-GWAS Analyses. Front Genet. 2018;9:280. doi:10.3389/fgene.2018.00280

121. French JD, Ghoussaini M, Edwards SL, et al. Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet. 2013;92(4):489-503. doi:10.1016/j.ajhg.2013.01.002
122. Jirawatnotai S, Hu Y, Michowski W, et al. A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers. Nature. 2011;474(7350):230-234. doi:10.1038/nature10155
123. Ferreira MA, Gamazon ER, Al-Ejeh F, et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. Nat Commun. 2019;10. doi:10.1038/s41467-018-08053-5
124. Mavaddat N, Pharoah PDP, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015;107(5). doi:10.1093/jnci/djv036
125. Muranen TA, Mavaddat N, Khan S, et al. Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. Breast Cancer Res Treat. 2016;158(3):463-469. doi:10.1007/s10549-016-3897-6
126. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. JAMA Oncol. July 2018. doi:10.1001/jamaoncol.2018.1901
127. Muranen TA, Greco D, Blomqvist C, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. Genet Med Off J Am Coll Med Genet. October 2016. doi:10.1038/gim.2016.147
128. Kuchenbaecker KB, Neuhausen SL, Robson M, et al. Associations of common breast cancer susceptibility alleles with risk of breast cancer subtypes in BRCA1 and BRCA2 mutation carriers. Breast Cancer Res BCR. 2014;16(6):3416. doi:10.1186/s13058-014-0492-9
129. Couch FJ, Wang X, McGuffog L, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. 2013;9(3):e1003212. doi:10.1371/journal.pgen.1003212
130. Michailidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. October 2017. doi:10.1038/nature24284
131. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291. doi:10.1038/nature19057

Chapter 2

Exome sequencing of non-BRCA1/2 hereditary breast cancer: no genetic evidence for a sub-group defined by aCGH profiling

Florentine S. Hilbers, Caro M. Meijers, Jeroen F.J. Laros, Michiel van Galen, Nicoline Hoogerbrugge, Hans F.A. Vasen, Petra M. Nederlof, Juul T. Wijnen, Christi J. van Asperen, Peter Devilee

PLoS ONE 2013; 8, e55734

Abstract

Most familial breast cancer risk (~70%) is not explained by mutations in the known predisposition genes, primarily *BRCA1* and *BRCA2*. Exome sequencing of cases from non-*BRCA1/2* breast cancer families might be an effective way to detect new high-risk predisposing genes, but only when the underlying genetic heterogeneity would be limited. We selected six families in which the tumours of multiple cases showed a specific genomic profile as detected by array comparative genomic hybridization (aCGH). Linkage analysis of these families revealed a region on chromosome 4 with a LOD score of 2.49 under homogeneity. We then applied massive parallel sequencing after exome capture on germline DNA of two patients from each family. Whereas many variants were revealed within the linked region, no genes were found with a potentially pathogenic variant in more than one family. When selecting variants outside the linked chromosome 4 region, we detected multiple possibly pathogenic variants in genes that encode for DNA integrity maintenance proteins. However, all of these were rejected as high-risk variants because of poor co-segregation or relatively high allele frequency in a control population. These results suggest that the aCGH-profile shared by the tumours in these six families does not define a sub-group of non-*BRCA1/2* familial breast cancer with a shared genetic aetiology. Our data also support the emerging view that the remainder of hereditary breast cancer has a very heterogeneous genetic basis.

Introduction

The genetic landscape of breast cancer susceptibility known to date is constituted by more than 30 gene loci. Mutations in some of these, like *BRCA1* and *BRCA2*, are extremely rare, but confer high risks to breast cancer, others are common but only confer a minor increase in risk. However, jointly these alleles explain less than 30% of the familial breast cancer risk.[1–3] When considering families with multiple cases of early-onset breast cancer in which mutations in the known high-risk genes have been excluded (hereafter: "BRCAx" families), an unknown, rare, highly penetrant allele would appear to be the most parsimonious genetic explanation. However, linkage studies have not discovered any major breast cancer susceptibility gene since the identification of *BRCA1* and *BRCA2*. This suggests that these high-risk alleles are too rare to be detected by linkage studies in unselected BRCAx families.

Therefore, an important factor determining the success of a genome-wide search for linkage in a set of BRCAx families is the extent of underlying genetic heterogeneity. Simulation studies have shown that study power drops sharply if mutations in the sought-after new gene explain <30% of the investigated families. Selecting families based on a shared phenotype might lead to a genetically more homogeneous group of families, which are more likely to share variants in the same gene. A shared phenotype might be defined by the presence of certain cancer types in the family, for example, linkage analysis of non-*BRCA1* breast cancer families with a case of male breast cancer, led to the discovery of the *BRCA2* locus.[4] Also, certain histopathological features of tumours might be used to identify subgroups. It has been shown that breast tumours from *BRCA1* and *BRCA2* mutation carriers show specific genomic profiles as determined by comparative genomic hybridization (CGH).[5–10]

We recently described a specific array comparative genomic hybridization (aCGH) profile in a subgroup of BRCAx breast tumours.[11] This aCGH-profile is characterized by a gain of almost whole chromosome 22, in combination with some other specific changes, and was observed to be present in multiple breast cancer cases within six of the 27 analyzed BRCAx families. We hypothesized that these six families might have mutations in the same high-risk breast cancer gene. Here we present linkage analysis of these six families as well as exome sequencing of two family-members from each.

Methods

Patients

Previously, we determined the aCGH profiles of 58 breast tumours from 27 BRCAx families. A detailed description of the original selection criteria of the BRCAx families is given in Didraga et al.[11] We selected six of these families in which the tumours of multiple cases showed the 22-gain-like profile. The pedigrees of these families are depicted in Figure 1a-f. The occurrence of cancer was assessed through the index case and whenever possible verified with pathology reports. The number of breast cancer cases per family ranged from five to eleven, with a mean age of onset of 54 years. No male breast cancer cases and no ovarian cancer cases were reported. In total 46 breast tumours were diagnosed in these families, of which four were second primary tumours. One breast cancer case developed a kidney tumour and another breast cancer case was diagnosed with colon cancer. Other cancers that occurred in these families were liver cancer (n=3), stomach/oesophagus cancer (n=3), colon cancer (n=2), melanoma (n=1), cervical cancer (n=1), prostate cancer (n=1) and two cancers of unknown type. All participants provided written informed consent and approval of the medical ethical committee at the Leiden University Medical Centre was obtained.

Linkage Analysis

The six selected families are part of a larger cohort (n=55 families), which was genotyped before by Oldenburg et al.[12] for a genome-wide linkage analysis study. In brief, all individuals from who DNA was available were genotyped using the Linkage Mapping Set MD10 from Applied Biosystems. Genotypes were called automatically using Genemapper software (Applied Biosystems) and checked manually. Allele frequencies were calculated based on one randomly chosen individual from each family. We performed a multipoint linkage analysis using Genehunter software (version 2.1 B).[13] We assumed a model with a dominant susceptibility allele with an allele frequency of 0.003. Breast cancer risk at age 80 for carriers of the risk allele was assumed to be 0.85. For non-carriers we assumed a risk of 0.096. Risks were modelled in seven age categories as described by Easton et al.[14] Under the assumption of homogeneity, the LOD scores of the six families linked to the 22-gain profile were added up. To define the limits of a linkage region we took the maximum LOD score minus one as a cut-off.

Massive Parallel Sequencing

Genomic DNA was extracted from peripheral blood using standard protocols. Samples were prepared according to the manufacturers protocol (SureSelect All Exon (v1), Agilent Technologies) with some minor adjustments. In brief, for each individual 5 µg DNA was fragmented using adaptive focused acoustics (Covaris S-series single tube) in order to get fragments of 200-300 bp. Primer oligonucleotides for paired-end sequencing (Illumina) were ligated to both ends of the fragment. Of each sample 500 ng was then hybridized with 2.5 µl SureSelect Oligo Capture Library for 20 hours. After multiple washing steps, the captured DNA was amplified in order to get sufficient DNA for the sequencing experiment. Paired-end flow cells were then prepared on a cluster station according to the manufactures protocol (Illumina), using one lane per sample. Sequencing was the performed on a Genome Analyzer IIx (Illumina) with a paired-end module, generating 75 base pair reads.

Data Analysis

Alignment of the reads was done using the GAPPSv3 pipeline. Before alignment raw reads were filtered for adapter sequences and low quality bases using the FastxToolkit.[15] Alignment to the human reference genome (hg19, GRCh37) was done using Stampy[16] which integrates BWA[17] for bulk alignment and its own algorithm for complex regions. For detailed settings see supplementary Table 1. Variants were called with VarScan.[18] Filter settings applied a minimum coverage of 10 times at the variant position, and a variant allele frequency of at least 30% of the reads. In the region of the linkage peak we increased the sensitivity by calling variants if the variant allele was supported by at least 15% of the reads. Annotation of the variants was done using SeattleSeq (version 7.01).[19] Assuming that causal variants are rare, we removed all variants with an allele frequency >1% in either HapMap[20], 1000 genomes (phase 1),[21] exome variant server (v.0.0.11, ESP5400)[22] or our in-house variant database. In addition, variants that were found in a homozygous state in at least one of the twelve individuals were removed.

Sanger sequencing and melting curve analysis (MCA)

Validation of variants was done using PCR following standard protocols, followed by Sanger sequencing on an ABI3730XL sequencer. To assess variant frequencies in familial breast cancer cases and controls, high resolution melting curve analysis was performed. Non-

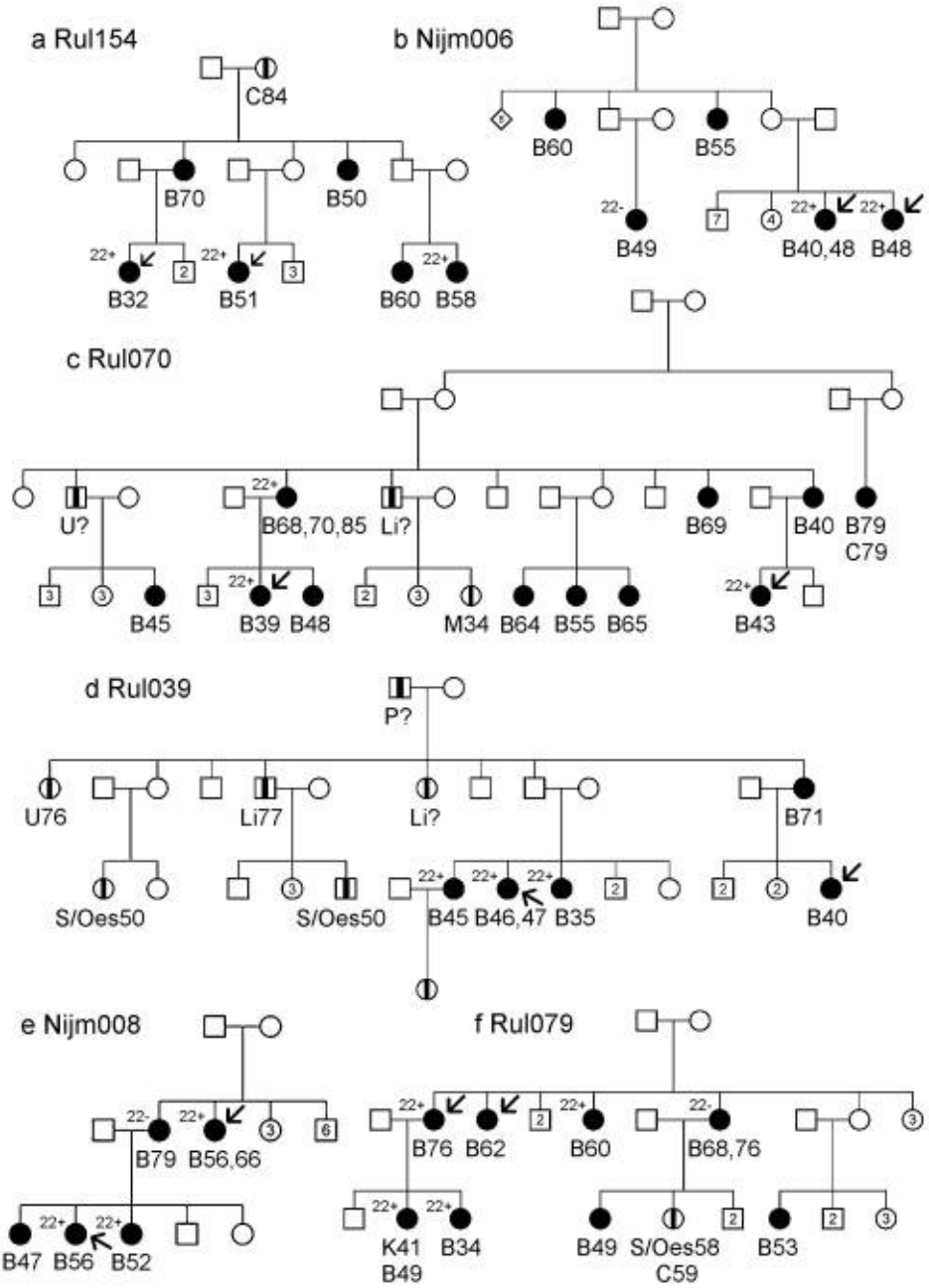


Figure 1 a-f

Pedigrees of the families in which multiple tumours showed the "22-gain-like" aCGH profile. Individuals affected with breast cancer are represented by a filled square or circle. Individuals affected by another type of cancer are represented by a square or circle with a vertical black stripe. Below the age at diagnosis and type of cancer can be found: B stands for breast cancer, Li or liver cancer, S for stomach cancer, Oes for oesophagus cancer, C for colon cancer, M for melanoma, Cvx for Cervix cancer, K for kidney cancer, P for prostate cancer and U for type of cancer unknown. Arrows point at the individuals at whose DNA was used for exome sequence. Individuals with tumours with and without the "22-gain-like profile" are represented by "22+" and "22-".

BRCA1/2 familial breast cancer cases (n=531) were obtained from the clinical genetics centre Leiden and healthy controls (n=458) were obtained from the Dutch blood bank, Sanquin. PCR was performed in a 1:10:10 forward primer: reverse primer: probe ratio in the presence of LC green (Idaho Technology Inc.). Melting curves were assessed on a LightScanner (Idaho Technology Inc.) for temperatures between 50°C and 90°C and analyzed with Call-IT software (Idaho Technology Inc.). All primer and probe sequences are available upon request.

Results

We previously analysed the breast tumours of 58 patients from 27 BRCAx families using aCGH.[11] Hierarchical clustering identified several subgroups of BRCAx tumours, one of which was characterized by a gain of chromosome 22. Remarkably, in 6 families (Figure 1), tumours from multiple patients displayed this chromosome 22 gain profile. Linkage analysis under homogeneity revealed a linkage peak with a LOD score of 2.49 on chromosome 4 in these six families (Figure 2). The next highest linkage peak was 1.04 at 10q and no other linkage peaks with a LOD score greater than 1.0 were detected.

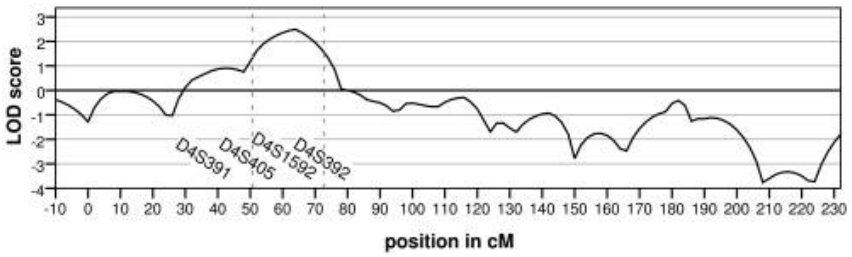


Figure. 2
Linkage on chromosome 4 for the families in which multiple tumours showed the "22-gain-like" aCGH profile. The LOD-score was calculated under the assumption of homogeneity. The dashed lines indicate the maximum LOD-score -1interval. The X-axis shows the position on chromosome 4 in centimorgan and the markers closest to the linkage peak are indicated.

A 25-Mb candidate region (chr4:40.000.000-65.000.000) was defined as the region showing a LOD score greater than the peak LOD score minus one. Two individuals per family were selected for exome sequencing, usually at least second-degree relatives (figure 1). (Details on coverage of the individual exomes can be found in supplementary figure 1 and 2.) This revealed on average 499 variants in the candidate region that were shared by both individuals of a family. After removing intergenic and non-conserved variants in non-coding regions, five variants remained (Table 1). However, none of the genes carrying these variants were found to do so in two or more families. Hence mutations in a single gene are less likely to explain the linkage result. We then considered the possibility that two or more genes in the chromosome 4 region each fortuitously carries a high-risk mutation in one of the six families. Of the detected variants, three synonymous variants in three genes (*FRYL*, *AASDH*, *PPAT*) were not further examined, because these variants are unlikely to affect protein function. A missense variant in *REST* and a well-conserved 3'UTR variant in *LNX1* were validated by Sanger sequencing. The *LNX1* variant was present in five of eight cases in family RUL070. The missense variant in *REST* was detected in six out of seven cases in family RUL079, however Grantham and conservation scores for this variant were low (Grantham=45, Phastcons=0.00, GERP= -3.56) and Polyphen[23] predicts it to be benign.

Table 1 Well conserved or coding variants in the linkage region on chromosome 4.

Variant	Gene	Family	Effect	rs-number	Phast Cons[1]	GERP[1]
Chr4:g.48545947T>C	*FRYL*	RUL070	Synonymous	-	1.00	1.33
Chr4:g.54327036_54327037insATT	*LNX1*	RUL070	3' UTR	57366823	0.97	4.56
Chr4:g.57248742A>C	*AASDH*	RUL070	Synonymous	146114987	1.00	-0.43
Chr4:g.57261623G>A	*PPAT*	RUL070	Synonymous	-	0.22	-5.98
Chr4:g.57797037G>T	*REST*	RUL079	Missense [2]	138787075	0.00	-3.56

[1]Phastcons and GERP are both regional conservation algorithms ranging from 0 to 1 and -12.3 to 6.17 respectively (1 and 6.17 being most conserved). [2] Grantham = 45, PolyPhen prediction = Benign.

Finally, we examined the possibility that the six families shared variants in a gene outside the linkage peak region. We first focused on variants that were likely to result in a truncated protein (gained stop-codon, frameshift and splice-site variants). In the six families we found in total 49 different, rare protein-truncating variants in 48 genes. A number of genes showed a protein-truncating variant shared by several families. However, all these variants were present in regions whose sequences showed large similarities with regions elsewhere in the genome. When examining the unprocessed sequence-reads of the families in which the variants were not called, in most instances the variant could be detected, but in fewer reads than the required threshold of 30%. Thus, we considered all these variants to be false-positive findings resulting from sequence read-mapping errors. Indeed, the only one of these variants that we followed up by Sanger sequencing was a splice-site mutation in *FANCD2*. *FANCD2* is a Fanconi Anaemia gene and therefore a candidate breast cancer gene. However, upon re-sequencing, this variant was not present in *FANCD2*, but in a region with a similar sequence elsewhere on chromosome 3 near *EMC3* (data not shown).

Of the truncating variants that were present in a single family, a frameshift mutation, c.811delT, in *HAUS3* was potentially interesting, because *HAUS3* has been reported to be somatically mutated in a lobular breast tumour.[18] Sanger sequencing showed that five out of seven breast cancer patients in RUL079 had this deletion. High resolution melting curve analysis of this specific variant did not reveal any additional carriers among 531 familial breast cancer cases. However, three individuals in a group of 458 healthy controls were found to carry the c.811delT, dismissing it as a high-risk breast cancer allele.

We also took into account possibly damaging missense variants. This was defined as missense variants with either a Grantham score >100, a GERP conservation score > 3, a PhastCons conservation score > 0.7 or a "probably damaging" PolyPhen2 prediction. Due to the large number of variants remaining (n=657), following up all variants with Sanger sequencing was deemed impractible. We therefore selected variants with a function in DNA integrity maintenance, because the majority of breast cancer susceptibility genes identified to date have a function in this pathway (table 2). Again, no genes were found to have a variant in more than one family. However, some individual families showed possibly damaging variants in genes (n=8) involved DNA damage repair or chromosome segregation, shared by both assayed individuals. Sanger sequencing showed that the variants in *CASC5*, *CUL9*, *MUTYH*, *SMC6*, *TTK* and *XRCC2* had a poor or moderate co-segregation with disease (Supplementary figure 3). The variant in *RBMX* could not be validated. However, the variant in *HLTF*, p.S378T, was present in five out of five cases of family NIJM008. This variant was selected because of a high GERP conservation score (3.15). The PhastCons conservations score, however, was only 0.21 and this variant was predicted to be benign by Polyphen2.

Table 2 Possibly damaging or well conserved variants in genes encoding proteins involved in DNA integrity maintaince.

Gene	Variant	Grantham	GERP[1]	PhastCons[1]	PolyPhen2	Function
CASC5	p.I26L	5	4.53	0.999	Probably damaging	Spindle-assembly checkpoint signaling and chromosome alignment
CUL9	p.S2328F	155	5.03	0.989	Possibly damaging	Regulates the subcellular localization of p53 and subsequent function
HLTF	p.S378T	58	3.15	0.208	Benign	Error-free postreplication repair of damaged DNA
MUTYH	p.I223V	29	5.43	1	Benign	Oxidative DNA damage repair
RBMX	p.Y357H	83	5.66	1	Probably damaging	Regulation of programmed cell death in breast cancer and homologous recombination
SMC6	p.R403W	101	2.65	0.998	Probably damaging	DNA damage repair via homologous recombination
TTK	p.R185W	101	4.04	0.004	Probably damaging	Chromosome alignment, centrosome duplication and critical mitotic checkpoint
XRCC2	p.R91W	101	4.48	0.742	Probably damaging	DNA damage repair via homologous recombination

Variants were selected if either of these criteria was met: Grantham score>100, GERP conservation score> 3, PhastCons conservation score>0.7, or a "Probably damaging" Polyphen2 prediction. 1Phastcons and GERP are both regional conservation algorithms ranging from 0 to 1 and -12.3 to 6.17 respectively (1 and 6.17 being most conserved).

Discussion

The landscape of genetic risk factors for breast cancer is known to be diverse, ranging from rare high-risk alleles, like *BRCA1* and *BRCA2*, to common polymorphisms that only confer a minor breast cancer risk increase. The large proportion of familial breast cancer cases that is not explained by the genetic risk factors known to date is thought to be very heterogeneous. Both segregation analysis[24–26] and the fact that no major high-risk breast cancer genes have been identified since *BRCA2* suggest that additional high-risk alleles are most likely much rarer than mutations in *BRCA1* and *BRCA2*. Massive parallel sequencing might be a very useful tool to identify these very rare high-risk alleles. However, finding novel disease alleles among thousands of not-pathogenic variants might be more complex in a common and genetically heterogeneous disease like breast cancer, than in the rare Mendelian phenotypes in which massive parallel sequencing has been very successful to date.[27] Therefore, selecting a genetically more homogeneous patient subgroup might be crucial.

We hypothesized that by selecting BRCAx families with a similar phenotype, we would enrich our study population for families with germline mutations in the same gene. In this study six BRCAx families in which the majority of tumours show a previously identified aCGH profile[11] were selected. Linkage analysis in these families showed a peak on chromosome 4, which suggested that these families might share a genetic aetiology. Massive parallel sequencing after whole-exome capture was performed on two individuals per family. However, no genes could be identified in which more than one family showed a likely pathogenic variant after assessing the predicted effect on the protein and co-segregation. In addition, we did detect multiple possibly pathogenic variants in genes that encode for DNA integrity maintenance proteins outside the linkage peak region. All of these were however discarded as likely causes of familial clustering of breast cancer because of poor co-segregation or relative high frequency of the specific variant in a control population.

Multiple studies have shown that aCGH classifiers can be build to distinguish *BRCA1* and *BRCA2* tumours from sporadic tumours and each other.[5–10] These studies suggest that tumours of patients with mutations in the same gene also share a somatic genetic aetiology. Alvarez[28] and colleagues found that part of the BRCAx tumours showed aCGH profiles similar to those of *BRCA1* tumours. A large proportion of these tumours turned out to have hypermethylation of *BRCA1*. Some studies that performed aCGH profiling on BRCAx tumours find similarities with profiles of *BRCA2* tumours,[29,30] suggesting that either a cause of *BRCA2* inactivation in these tumours has yet to be detected or that inactivation of a number of genes can lead to a similar aCGH profile. It might be that patients with the 22-gain profile do not share mutations in the same gene, but in the same pathway. In order to detect an enrichment of deleterious variants in a specific pathway, a larger number of familial patients with 22-gain tumours may need to be sequenced. In this case also expression profiling of tumours might be helpful; however, it will be challenging to collect sufficient numbers samples for such an effort.

Another possibility is that patients with a 22-gain tumour have mutations in a moderate risk gene. Muranen et al.[31] have shown that specific aCGH features occur significantly more often in tumours of patients with a *CHEK2**1100delC mutation. This suggests that also moderate risk germline mutations can lead to a homogenous phenotype. By only assessing variants that are shared by both family members and discarding variants that show poor co-segregation, we would most likely have missed variants in a moderate risk gene. In addition, moderate risk variants might have an allele frequency of more than 1% as has been shown to be true for the *CHEK2**1100delC mutation in some populations.[32] However, without using

these selection criteria, it would not have been possible to limit possibly interesting variants to a number that is manageable to follow-up. Therefore, a study design which includes exome sequencing in a very limited number of familial cases has little power to detect moderate risk variants.

In general, the balance between stringent selection criteria to limit the number of variants for follow-up and not excluding too many possibly interesting variants is not easy to find. Keinan and Clark recently described the excess of rare genetic variants due to recent explosive growth of the human population.[33] This excess of rare variants causes large numbers of variants to remain after selection on low population frequency. In addition, this makes it hard to validate the effect of a very rare variant on breast cancer risk outside the family it was originally detected in. An example of, in hindsight, too stringent selection is the missense variant in *XRCC2* on which we did originally not further follow-up because it did not show convincing co-segregation with disease. However, via personal communication we became aware that Park and colleagues had found a truncation variant in the same gene. After mutations scanning a large population of familial breast cancer cases and controls a significant association with breast cancer risk was found.[34] These points stress that international collaboration and sharing of data, both in the variant selection and validation phase might be very important.

In conclusion, we did not find evidence for mutations in a rare high-risk gene in a subgroup of BRCAx cases defined by an aCGH profile. Although, we can not rule out that these families have mutations in genes belonging to the same pathway or in a non-coding region. Massive parallel sequencing efforts in large cohorts of BRCAx cases are needed to definitively unravel the genetic basis underlying the aetiology of unexplained familial clustering of breast cancer and its link with tumour characteristics.

Acknowledgements

We thank Sophie Greve-Onderwater and Yavuz Ariyurek for their practical advice on massive parallel sequencing. This work was financially supported by the Dutch Cancer Society (grant UL 2009-4388)

References

1. Ahmed M, Lalloo F, Evans DG. Update on genetic predisposition to breast cancer. Expert Rev Anticancer Ther. 2009;9(8):1103-1113. doi:10.1586/era.09.38
2. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. Nat Genet. 2008;40(1):17-22. doi:10.1038/ng.2007.53
3. Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. Annu Rev Genomics Hum Genet. 2008;9:321-345. doi:10.1146/annurev.genom.9.081307.164339
4. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science. 1994;265(5181):2088-2090.
5. Jonsson G, Naylor TL, Vallon-Christersson J, et al. Distinct genomic profiles in hereditary breast tumors identified by array-based comparative genomic hybridization. Cancer Res. 2005;65(17):7612-7621. doi:10.1158/0008-5472.CAN-05-0570
6. Joosse SA, van Beers EH, Tielen IHG, et al. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. Breast Cancer Res Treat. 2009;116(3):479-489. doi:10.1007/s10549-008-0117-z
7. Joosse SA, Brandwijk KIM, Devilee P, et al. Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. Breast Cancer Res Treat. 2012;132(2):379-389. doi:10.1007/s10549-010-1016-7
8. Tirkkonen M, Johannsson O, Agnarsson BA, et al. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. Cancer Res. 1997;57(7):1222-1227.
9. van Beers EH, van Welsem T, Wessels LFA, et al. Comparative genomic hybridization profiles in human BRCA1 and BRCA2 breast tumors highlight differential sets of genomic aberrations. Cancer Res. 2005;65(3):822-827.
10. Wessels LFA, van Welsem T, Hart AAM, van't Veer LJ, Reinders MJT, Nederlof PM. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. Cancer Res. 2002;62(23):7110-7117.
11. Didraga MA, van Beers EH, Joosse SA, et al. A non-BRCA1/2 hereditary breast cancer subgroup defined by aCGH profiling of genetically related patients. Breast Cancer Res Treat. 2011;130(2):425-436. doi:10.1007/s10549-011-1357-x
12. Oldenburg RA, Kroeze-Jansema KHG, Houwing-Duistermaat JJ, et al. Genome-wide linkage scan in Dutch hereditary non-BRCA1/2 breast cancer families identifies 9q21-22 as a putative breast cancer susceptibility locus. Genes Chromosomes Cancer. 2008;47(11):947-956. doi:10.1002/gcc.20597
13. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58(6):1347-1363.
14. Easton DF, Bishop DT, Ford D, Crockford GP. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. Am J Hum Genet. 1993;52(4):678-701.
15. FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit/. Accessed September 13, 2019.
16. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res*. 2011;21(6):936-939. doi:10.1101/gr.111120.110
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
18. Koboldt DC, Chen K, Wylie T, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283-2285. doi:10.1093/bioinformatics/btp373
19. SeattleSeq Variation Annotation. https://snp.gs.washington.edu/SeattleSeqAnnotation134/. Accessed September 13, 2019.
20. The International HapMap Project. *Nature*. 2003;426(6968):789-796. doi:10.1038/nature02168

21. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073. doi:10.1038/nature09534
22. Exome Variant Server. http://evs.gs.washington.edu/EVS/. Accessed June 1, 2015.
23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248
24. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol*. 2001;21(1):1-18. doi:10.1002/gepi.1014
25. Antoniou AC, Pharoah PDP, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer*. 2002;86(1):76-83. doi:10.1038/sj.bjc.6600008
26. Cui J, Antoniou AC, Dite GS, et al. After BRCA1 and BRCA2-what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. *Am J Hum Genet*. 2001;68(2):420-431. doi:10.1086/318187
27. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790-793. doi:10.1038/ng.646
28. Alvarez S, Diaz-Uriarte R, Osorio A, et al. A predictor based on the somatic genomic changes of the BRCA1/BRCA2 breast cancer tumors identifies the non-BRCA1/BRCA2 tumors with BRCA1 promoter hypermethylation. *Clin Cancer Res*. 2005;11(3):1146-1153.
29. Gronwald J, Jauch A, Cybulski C, et al. Comparison of genomic abnormalities between BRCAX and sporadic breast cancers studied by comparative genomic hybridization. *Int J Cancer*. 2005;114(2):230-236. doi:10.1002/ijc.20723
30. Mangia A, Chiarappa P, Tommasi S, et al. Genetic heterogeneity by comparative genomic hybridization in BRCAx breast cancers. *Cancer Genet Cytogenet*. 2008;182(2):75-83. doi:10.1016/j.cancergencyto.2008.01.002
31. Muranen TA, Greco D, Fagerholm R, et al. Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. *Breast Cancer Res*. 2011;13(5):R90. doi:10.1186/bcr3015
32. CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet*. 2004;74(6):1175-1182. doi:10.1086/421251
33. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*. 2012;336(6082):740-743. doi:10.1126/science.1217283
34. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet*. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027

Supplementary Data

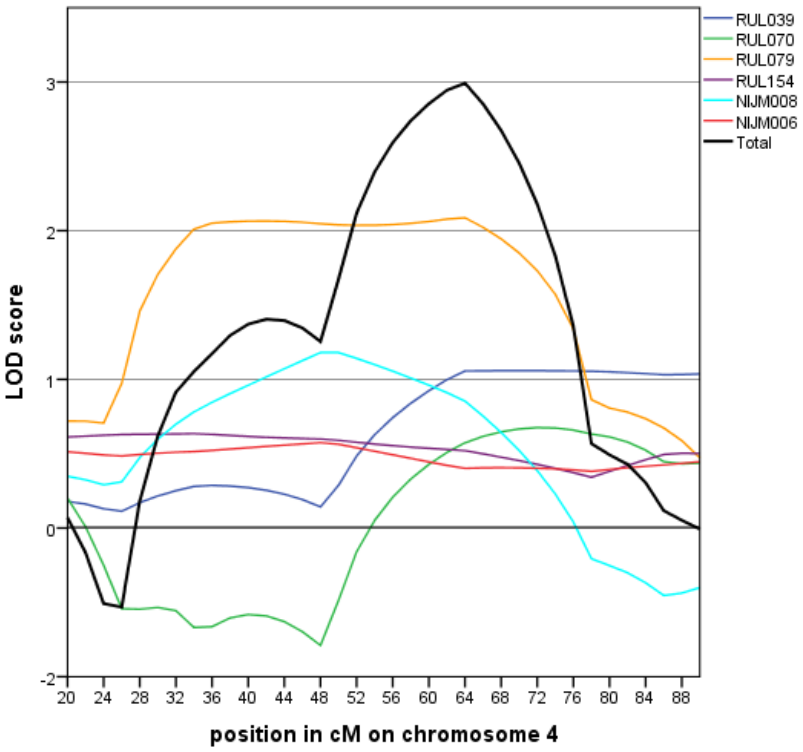


Figure S1
Parametric LOD scores of the individual families in the linkage region on chromosome 4. The X-axis shows the position on chromosome 4 in centimorgan.

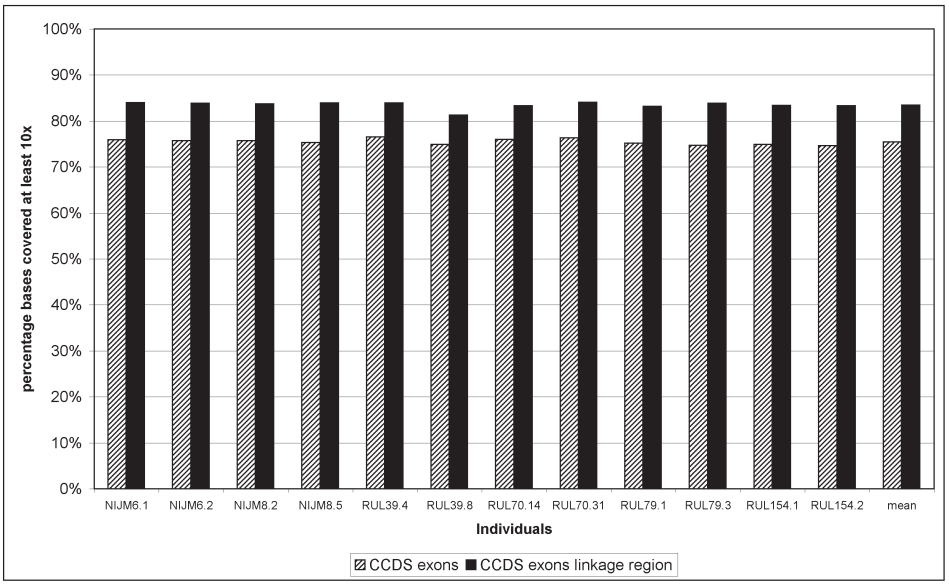


Figure S2
Percentage of CCDS exon bases covered at least 10× per individual. CCDS=consensus coding sequence.

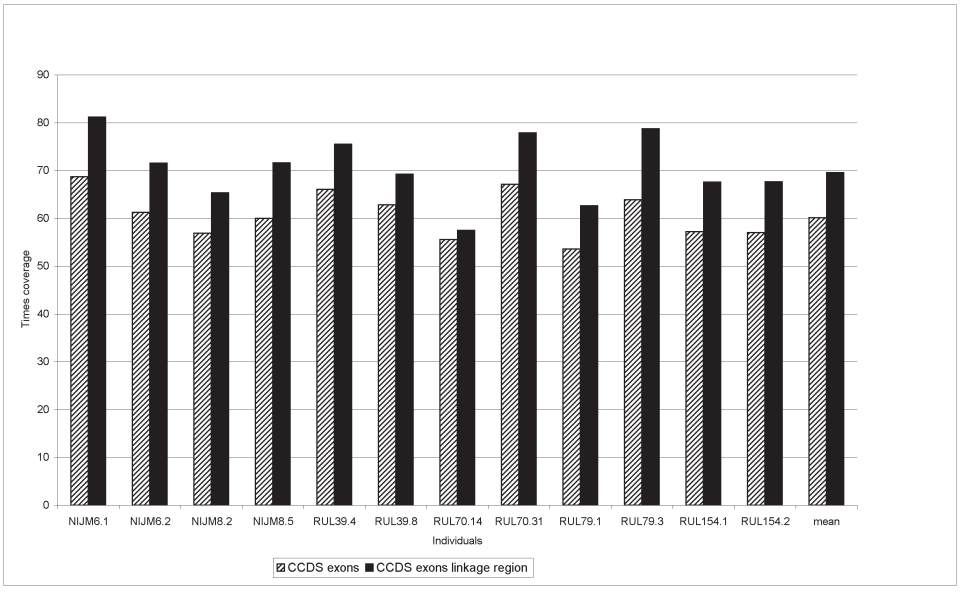


Figure S3
Mean coverage of CCDS exons per individual. CCDS=consensus coding sequence.

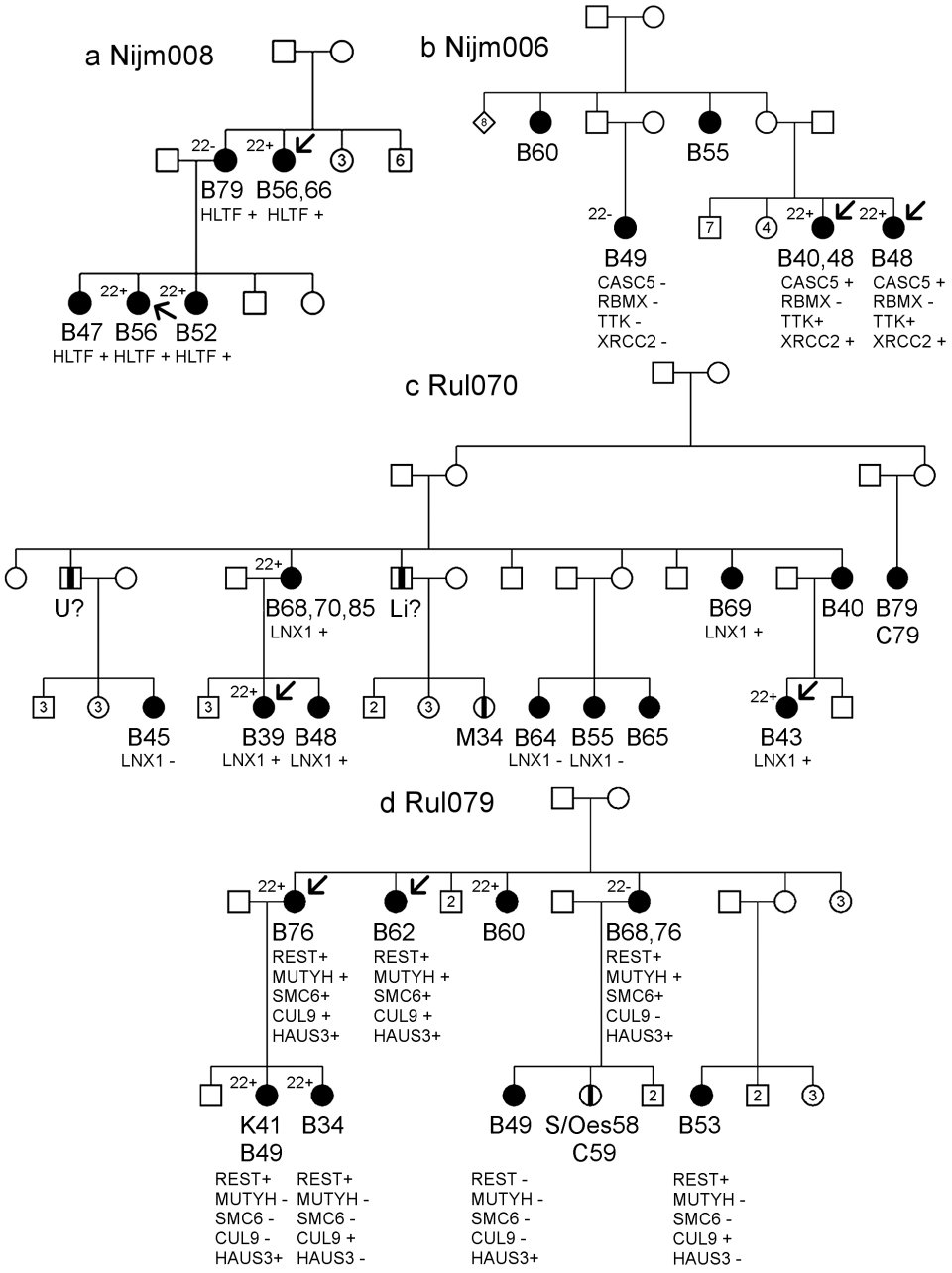


Figure S4
Segregation of selected variants within the families (a–d). Individuals carrying or not carrying a specific variant are indicated with a "+" or with a "–"respectively. The p.Y357H variant in RBMX, which was detected by massive parallel sequencing in family Nijm006, could not be validated by Sanger sequencing.

Table S1 Description of the data analysis settings

Software	Version	Settings
FastxToolkit	0.0.13	minimum base quality=20 minimum percentage base with minimum quality=80 minimum read length after clipper/trimmer=25
FastQC	v0.8.0	Default
Stampy	v1.0.12	Default
BWA	0.5.6	Default
Samtools	0.1.14	Default
VarScan	v2.2	See material and methods section

Table S2 Truncating variants all detected in only one of the six families

Gene	Variant	Function
A4GNT	p.R226X	Glycosyltransferase
ADAMTS7	p.N1353fs	Metallopeptidase
ANO8	p.? *	Ca2+-activated Cl– channel
CRNKL1	p.Q50X	Pre-mRNA splicing
CPA3	p. ? **	Secretory granule metalloexopeptidase
HAUS3	p.C271fs	Microtubule generation within the mitotic spindle
HIST1H2BE	p.S7fs	Member of the histone H2B family
IFNK	p.W13X	Glycoprotein, important in host defenses against viral infections
KCNQ5	p.Q824X	Potassium channel
KIAA1751	p.A732fs	Unknown
LOC100132900	p.R90X	Unknown
MSGN1	p.Q106X	Mesoderm maturation
NIPSNAP3A	p.Q124fs	Putative role in vesicular transport
OTOP3	p.Q82X	Unknown
PKD1L2	p.G235fs	May function as a component of cation channel pores
RETNLB	p.I12fs	Important for epithelial barrier function and integrity
TAS2R8	p.F245fs	Taste receptor
TBC1D17	p.F546fs	Unknown

Chapter 3

Rare variants in *XRCC2* as breast cancer susceptibility alleles

Florentine S. Hilbers, Juul T. Wijnen, Nicoline Hoogerbrugge, Jan C. Oosterwijk, Margriet J. Collee, Paolo Peterlongo, Paolo Radice, Siranoush Manoukian, Irene Feroce, Fabio Capra, Fergus J. Couch, Xianshu Wang, Lucia Guidugli, Kenneth Offit, Sohela Shah, Ian G. Campbell, Ella R. Thompson, Paul A. James, Alison H. Trainer, Javier Gracia, Javier Benitez, Christi J. van Asperen, Peter Devilee

J Med Genet. 2012 Oct; 49(10): 618-20

Abstract

Background Recently, rare germline variants in *XRCC2* were detected in non-*BRCA1/2* familial breast cancer cases, and a significant association with breast cancer was reported. However, the breast cancer risk associated with these variants needs further evaluation.

Methods The coding regions and exon–intron boundaries of *XRCC2* were scanned for mutations in an international cohort of 3548 non-*BRCA1/2* familial breast cancer cases and 1435 healthy controls using various mutation scanning methods. Predictions on functional relevance of detected missense variants were obtained from three different prediction algorithms.

Results The only protein-truncating variant detected was found in a control. Rare non-protein-truncating variants were detected in 20 familial cases (0.6%) and nine healthy controls (0.6%). Although the number of variants predicted to be damaging or neutral differed between prediction algorithms, in all instances these categories were evenly represented among cases and controls.

Conclusions Our data do not confirm an association between *XRCC2* variants and breast cancer risk, although a relative risk smaller than two could not be excluded. Variants in *XRCC2* are unlikely to explain a substantial proportion of familial breast cancer.

Recently, Park and colleagues identified germline variants in x-ray repair cross complementing gene-2 (*XRCC2* (MIM 600375; NM_005431.1)) in a small number of breast cancer cases with a positive family history for the disease.[1] The overall difference in the prevalence of protein-truncating and potentially deleterious rare missense variants between cases and controls was reported to be statistically significant. XRCC2 is involved in the repair of double-strand breaks via homologous recombination. In addition, a homozygous protein-truncating variant in *XRCC2* has been detected in a Fanconi anaemia patient with consanguineous parents.[2] Since most known high and moderate-risk breast cancer genes have a function in DNA damage repair, and some are also Fanconi anaemia genes, it seems plausible that mutations in *XRCC2* represent breast cancer susceptibility alleles. In order to evaluate the association between *XRCC2* variants and breast cancer risk, we analysed the coding regions of *XRCC2* in a cohort of 3548 non-*BRCA1/2* familial breast cancer cases and 1435 healthy controls derived from various geographical locations. A more detailed description of the study population and mutations-scanning methods can be found in the online supplementary table S1.

We detected only one protein-truncating variant, a one-base-pair deletion, c.343T[8]>[7], present in a 41-year-old Italian control. Rare non-protein truncating variants in *XRCC2* were detected in 20 familial cases (0.6%) and nine healthy controls (0.6%). Polyphen2,[3] SIFT[4] and AlignGVGD[5] were used to predict the effect of detected missense variants on XRCC2 protein function. Although the number of variants predicted to be damaging or neutral differed between the prediction algorithms (table 1), in all instances these categories were evenly represented among cases and controls. The only common variant in the coding region of *XRCC2*, c.563G>A (rs3218536), was found to have equal minor allele frequencies in familial cases (0.085) and healthy controls (0.086).

Thus, our data do not confirm an association between *XRCC2* variants and breast cancer risk. It is possible that the study by Park et al represents a false-positive finding, or, alternatively, our data are a false-negative finding. The association reported by Park et al was based on six likely pathogenic variants in 1308 cases and zero in 1120 controls, which was significant in Fisher's exact test. However, it should be noted that of these six variants, four were considered possibly or probably damaging, based on in silico prediction. Moreover, the number of positives in cases and controls is extremely small, and the statistic is, therefore, likely to be very unstable. In the NHLBI ESP Exome Variant Server,[6] a publicly available database describing variants found in exomes of patients with heart, lung and blood disease and healthy controls, rare variants in the coding region of *XRCC2* are reported in 0.5% of all exomes, interestingly including two variants resulting in a gained stop codon. Whereas an allele frequency of 0.5% corresponds quite well with our findings, and with the frequency found by Park et al in their cases, it seems at odds with what they detect in controls. Even when including the innocuous missense change, Park et al found one rare *XRCC2* variant in 1120 controls (0.09%). This frequency might have been an underestimation given that the study employed High Resolution Melting Curve analyses for mutation detection rather than direct sequencing. Indeed, when specifically testing two truncating variants in another 1436 controls, they found one of them to be positive, indicating that the frequency of pathogenic variants among controls is greater than what is suggested by their finding of zero out 1120.

Our study had 80% power to detect a relative risk of at least 2.1 ($p<0.05$) for a variant with an allele frequency of 0.5% (see online supplementary figure S1). Our results could, therefore, represent a false-negative finding if the relative risk associated with *XRCC2* variants would be lower than two. We note that variants in a number of other DNA damage repair genes, such as *CHEK2* and *BRIP1*, have been associated with breast cancer with such low

risks.[7,8] Park et al did not provide a quantitative estimate of the risk, but in the two families with an *XRCC2* variant for which other family members were also available for DNA analysis, cosegregation of the variant with breast cancer was incomplete. This suggests that if an association between *XRCC2* and breast cancer exists at all, it may not be very strong. A much larger sample than that studied by us here would need to be analysed to address this.

Another potential source of controversy is the selection criteria used for constituting the case- and control-series. Park and colleagues analysed 1308 breast cancer cases diagnosed before age 45 years and 1120 healthy controls recruited through population-based sampling by the Australian Breast Cancer Family Registry. In addition, they scanned 689 index cases from multicase breast cancer families and 150 male breast cancer cases. The current study included mostly clinic-based cohorts of cases that were forwarded for *BRCA*-mutation analysis because the prior probability of detecting a *BRCA1* or *BRCA2* mutation exceeded 10%.[9] Hence, both studies attempted to enrich for 'genetic' familial breast cancer cases, but in slightly different ways, perhaps leading to different representations of certain case subgroups. Ethnic backgrounds of the cohorts may also differ proportionally between the studies, although European ancestry was represented in both, and variants were detected in comparable frequencies among individuals of Italian, Spanish, Dutch and US origin.

Neither group studied the effect of missense variants on protein function other than by in silico prediction algorithms. Whereas truncating variants are likely to cause reduction in activity of XRCC2 in homologous recombination, the effect of missense variants (if any) may be subtler. If and how loss of XRCC2 function translates into breast cancer risk will be difficult to assess, but the existence of an effect cannot be excluded at this moment. In mice, complete loss of Xrcc2 is embryonic lethal, and leads to increased genetic instability at the cellular level,[10,11] a hallmark of many breast cancers.

In summary, our data do not confirm an association between *XRCC2* variants and breast cancer risk, although a relative risk smaller than two could not be excluded. Our inability to reproduce the previously reported association might point at a more general obstacle in applying exome sequencing in order to find new genes involved in common complex diseases. Exome sequencing, typically, yields many rare candidate variants. Accordingly, the prior odds that any of these variants are truly associated with disease are small, even when such a variant has been detected in two independent exomes. In any case, our data suggest that variants in *XRCC2* are unlikely to explain a substantial proportion of familial breast cancer.

Acknowledgements

The LUMC was supported by the Dutch Cancer Society (grant UL 2009-4388). MBCSG thanks Bernard Peissel and Daniela Zaffaroni of Fondazione IRCCS Istituto Nazionale Tumori; Bernardo Bonanni and Monica Barile of Istituto Europeo di Oncologia, and the personnel of the CGT laboratory at IFOM-IEO Campus. MBCSG was funded by grants from Fondazione Italiana per la Ricerca sul Cancro (Special Project 'Hereditary tumors'), Italian Ministry of Health ('Progetto Tumori Femminili'), and by Italian citizens who allocated the 5×1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects '5×1000'). The CNIO was partially supported by the Spanish Association against Cancer and FIS08-1120 from the Health Ministry. At MSKK support was from the Breast Cancer Research Fund and Miele Fund.

Table 1 Rare variants detected in the coding region of *XRCC2*

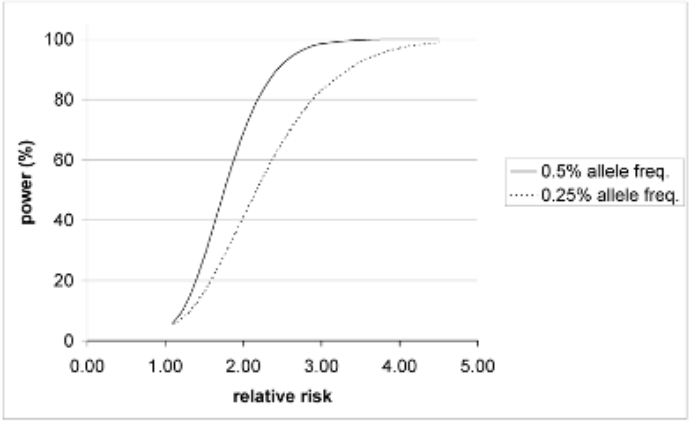
Variant*	Protein alteration†	rs-number	Polyphen-2	Align GVGD‡	SIFT§	Cases (%) n=3548	Controls (%) n=1435
c.140A>G	p.H47R	-	Poss. damaging	C0	0.01	1 (0.03)	0 (0)
c.189A>G	p.A63A	-	Silent	-	0.37	2 (0.06)	0 (0)
c.223G>C	p.E75Q	-	Benign	C0	0.14	1 (0.03)	0 (0)
c.283A>G	p.I95V	rs140214637	Benign	C0	0.38	6 (0.17)	2 (0.14)
c.343T[8]>[7]	p.L117fs	-	Truncating	-	-	0 (0)	1 (0.07)
c.353T>C	p.V118A	rs185815454	Benign	C0	1.00	1 (0.03)	0 (0)
c.359G>A	p.C120Y	-	Prob. Damaging	C65	0.00	0 (0)	1 (0.07)
c.398T>C	p.L133P	-	Prob. Damaging	C65	0.07	0 (0)	1 (0.07)
c.490G>C	p.E164Q	-	Prob. Damaging	C0	0.04	1 (0.03)	0 (0)
c.509A>C	p.E170A	-	Poss. damaging	C65	0.00	1 (0.03)	2 (0.14)
c.562C>T	p.R188C	rs139219364	Prob. Damaging	C15	0.01	1 (0.03)	0 (0)
c.581C>T	p.T194M	-	Prob. Damaging	C15	0.00	1 (0.03)	0 (0)
c.595A>C	p.M199L	-	Benign	C0	0.16	0 (0)	1 (0.07)
c.620A>G	p.E207G	rs61762969	Benign	C0	0.39	1 (0.03)	1 (0.07)
c.659A>T	p.D220V	-	Benign	C0	0.45	1 (0.03)	0 (0)
c.714G>C	p.R238S	-	Prob. Damaging	C65	0.05	1 (0.03)	0 (0)
c.742C>G	p.Q248E	rs190900560	Benign	C0	0.46	1 (0.03)	0 (0)
c.772C>T	p.R258C	-	Benign	C0	0.27	0 (0)	1 (0.07)
c.808T>G	p.F270V	rs145085742	Prob. Damaging	C45	0.00	1 (0.03)	0 (0)

*Based on XRCC2 transcript NM_005431.1
†Based on peptide sequence NP_005422.1
‡Depth: until Branchiostoma floridae (lancelet). C65 is the most likely deleterious category, while C0 is most likely neutral.
§Normalised probabilities smaller than 0.05 are predicted to be deleterious.

References

1. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
2. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. J Med Genet. 2012;49(3):184-186. doi:10.1136/jmedgenet-2011-100585
3. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249. doi:10.1038/nmeth0410-248
4. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863-874. doi:10.1101/gr.176601
5. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet. 2006;43(4):295-305. doi:10.1136/jmg.2005.033878
6. Exome Variant Server. https://evs.gs.washington.edu/EVS/. Accessed September 13, 2019.
7. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-59. doi:10.1038/ng879
8. 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. 2006;38(11):1239-1241. doi:10.1038/ng1902
9. Frank TS, Deffenbaugh AM, Reid JE, et al. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. J Clin Oncol. 2002;20(6):1480-1490. doi:10.1200/JCO.2002.20.6.1480
10. Deans B, Griffin CS, O'Regan P, Jasin M, Thacker J. Homologous recombination deficiency leads to profound genetic instability in cells derived from Xrcc2-knockout mice. Cancer Res. 2003;63(23):8181-8187.
11. Deans B, Griffin CS, Maconochie M, Thacker J. Xrcc2 is required for genetic stability, embryonic neurogenesis and viability in mice. EMBO J. 2000;19(24):6675-6685. doi:10.1093/emboj/19.24.6675

Supplementary data



Supplementary fig. 1 Power to detect a specific relative risk.
For a study with 3548 cases and 1435 controls and α is 0.05.

Supplementary table 1. Description of the study populations and mutation scanning methods

	Centre	Country	n	Mutation scanning method	Description
Cases	LUMC	The Netherlands	847	Sanger sequencing	Probands from the clinical genetics centre of Leiden, Rotterdam, Nijmegen and Groningen, who tested negative for BRCA1/2 mutations, either unselected but with an a priori change of a BRCA1 or BRCA2 mutation of ≥ 10% (n=472) or at least 2 cases with age of onset <50 (n=375)
	MSKCC	USA	739	Sanger sequencing	Probands from families with 3 or more breast cancer cases, no age restrictions, negative for mutations in BRCA1/2
	Peter MacCallum Cancer Centre	Australia	590	HRMCA[1,3]	Affected probands from the familial cancer clinic, who tested negative for BRCA1/2 mutations from multigenerational affected families with an a priori chance of a BRCA1 or BRCA2 mutation of ≥ 10%.
	CNIO	Spain	468	DHPLC[2,3]	Patients from families with at least 2 breast cancer cases (at least one of the <50 years) and tested negative for BRCA1/2 mutations
	MBCSG	Italy	463	Sanger sequencing	Probands ascertaind at the Istituto Nazionale Tumori (INT) and Istituto Europeo di Oncologia (IEO), Milan, who tested negative for BRCA1/2 mutations and with early onset of disease (<36 years) or from families with at least 2 breast cancer cases diagnosed before age 50 or 1 breast cancer case and 1 ovrian cancer case, .
	Mayo Clinic	USA	441	Sanger sequencing	Cases from BRCA1/2 negative multi-case breast cancer families, from some families more than 1 case was screened for XRCC2 variants
Controls	MBCSG	Italy	444	Sanger sequencing	Age-matched female blood donors recruited through the Immunohematology and Transfusion Medicine Service of INT and Associazione Volontari Italiani Sangue (AVIS) of Milan
	CNIO	Spain	392	DHPLC[2,3]	Healthy controls from the Spanish College of Lawyers
	LUMC	The Netherlands	367	Sanger sequencing	Healthy controls from the Dutch blood bank
	Peter MacCallum Cancer Centre	Australia	232	HRMCA[1,3]	Healthy Caucasian female volunteers from southern England

All primer sequences are available upon request.
[1] HRMCA = High resolution melting curve analysis
[2] DHPLC = Denaturating high performance liquid chromatography
[3] Detected variants were validated by Sanger sequencing

Chapter 4

Functional analysis of missense variants in the putative breast cancer susceptibility gene *XRCC2*

Florentine S. Hilbers, Martijn S. Luijsterburg, Wouter W. Wiegant, Caro M. Meijers, Moritz Völker-Albert, Rick A. Boonen, Christi J. van Asperen, Peter Devilee, Haico van Attikum

Hum Mutat. 2016 Sep; 37(9): 914-25

Abstract

XRCC2 genetic variants have been associated with breast cancer susceptibility. However, association studies have been complicated because *XRCC2* variants are extremely rare and consist mainly of amino acid substitutions whose grouping is sensitive to misclassification by the predictive algorithms. We therefore functionally characterized variants in *XRCC2* by testing their ability to restore XRCC2-DNA repair deficient phenotypes using a cDNA-based complementation approach. While the protein-truncating variants p.Leu117fs, p.Arg215* and p.Cys217* were unable to restore XRCC2 deficiency, 19 out of 23 missense variants showed no or just a minor (<25%) reduction in XRCC2 function. The remaining 4 (p.Cys120Tyr, p.Arg91Trp, p.Leu133Pro and p.Ile95Lx) had a moderate effect. Overall, measured functional effects correlated poorly with those predicted by in silico analysis. After regrouping variants from published case-control studies based on the functional effect found in this study and reanalysis of the prevalence data, there was no longer evidence for an association with breast cancer. This suggests that if breast cancer susceptibility alleles of *XRCC2* exist, they are likely restricted to protein-truncating variants and a minority of missense changes. Our study emphasizes the use of functional analyses of missense variants to support variant classification in association studies.

Introduction

One of the most important risk factors for breast cancer is having family members affected with the disease.[1] Mutations in two of the most well-characterized susceptibility genes, *BRCA1* (OMIM 113705) and *BRCA2* (OMIM 600185), explain about 17% of this familial relative risk.[2–4] Although a number of additional moderate and high-risk genes have been identified,[5] these explain at most another 5%. [4] Any yet unidentified moderate or high-risk breast cancer gene is likely to have extremely low mutation frequencies, hampering its discovery. Next-generation sequencing has provided the possibility to detect these rare variants on a genome-wide scale. However, due to the very low allele frequencies of these variants, obtaining conclusive evidence for an association with breast cancer has proven a considerable challenge.

XRCC2 (OMIM 600375) was among the first genes in which variants were reported to be associated with breast cancer risk by a study employing next-generation sequencing.[6] However, variants in *XRCC2* are very rare, with about 0.5% of healthy individuals carrying one (with the exception of the common single nucleotide polymorphism (SNP) rs3218536), and most have been detected only once. Nonetheless, by genotyping familial or early onset cases and healthy controls a significant association with breast cancer was detected.[6] *XRCC2* was a very interesting candidate breast cancer gene for several reasons. Firstly, it has been identified as a potential Fanconi anemia gene in a Saudi Arabian patient from consanguineous parents.[7] Currently, for at least four of the sixteen known Fanconi anemia genes, it has been shown that, while two mutated alleles cause Fanconi anemia, a single affected allele increases the risk of breast cancer.[8–12] Secondly, *XRCC2*, like many other confirmed breast cancer genes, encodes a protein involved in the repair of DNA double-strand breaks (DSB) via homologous recombination (HR). XRCC2 forms a complex with several members of the RAD51 protein family, including RAD51B, RAD51C and RAD51D. This complex is required for the localization of RAD51, a key mediator of HR, to DSB sites.[13] During HR, the ends of a DSB are resected to form single-strand DNAs, which are coated with RPA. RAD51 replaces RPA and subsequently initiates the search for a homologous sequence to enable repair of the DSB.[14]

Although *XRCC2* is a promising candidate breast cancer gene, a second, larger case-control study initiated by our lab, did not confirm its association with breast cancer since an equal percentage of variants in familial cases and healthy controls was found.[15] However, both case-control studies lacked the power to examine associations with individual variants or even with variant subtypes, and therefore protein-truncating variants and missense variants predicted to be damaging by in silico analysis were pooled in order to calculate the association with breast cancer. Given the very low number of variants detected, misclassification of only a few variants can already strongly influence the results of an association analysis, which might explain the discrepancy between the two studies.

Here we aimed to functionally characterize the *XRCC2* non-synonymous coding variants by testing their ability to complement the DNA repair phenotype of XRCC2-deficient cells. Remarkably, in our functional assays the majority of missense variants found in the population did not affect XRCC2's function in DNA repair, although in silico analyses predicted several of them to be damaging. When we re-classified the *XRCC2* variants based on these functional effects and re-analyzed the data from the two previously published case-controls studies, evidence for an association with breast cancer disappeared. These findings illustrate how functional analyses may aid the interpretation of rare genetic variation in the context of disease susceptibility.

Materials and Methods

Variant selection and in silico analysis

We used four different prediction algorithms to estimate how likely variants from breast cancer cases and healthy controls[6,15] as well as a Fanconi Anemia case[7] would affect XRCC2 protein function. PolyPhen2 (v2.2.2r398),[16] SIFT (JCVI-SIFT v.1.03, cut-off 0.05)[17] and AlignGVGD (depth until Branchiostoma floridae (Lancelet), cut-off C45)[18] predict the effect of a single amino acid change, while CADD (v1.2, cut-off 20)[19] can also provide predictions for small insertions or deletions and variants that result in a premature stop codon. All variants mentioned in this paper are described based on DNA sequence NM_005431.1 and protein sequence NP_005422.1. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1.

We used five different splice site prediction algorithms to predict the effect of the selected variants on splicing: SpliceSiteFinder (cut-off 70), MaxEntScan (cut-off 0),[20] Splice Site Prediction by Neural Network (cut-off 0.4),[21] GeneSplicer (cut-off 0),[22] and Human Splicing Finder (cut-off 65).[23] Prediction algorithms were run by the integrated software package Alamut (version 4.7.1) using default settings. An effect on splicing was considered when more than 10% difference in at least two algorithms was observed[24].

Cell lines and culture conditions

All cell lines were grown in DMEM-F12 (Gibco) containing 10% fetal bovine serum (Bodinco), Glutamax (Gibco, 1x), HEPES (Gibco, 10mM), Na-pyruvate (Lonza, 1mM) and penicillin/streptomycin (Gibco, 100 units/mL /100ug/mL). *irs1* cells were kindly provided by John Thacker.[25,26] *irs1* cells carrying the sister chromatid reporter (SCR) (irs-SCR) were a kind gift from Ralph Scully.[27,28] HEK293 cells containing a DR-GFP reporter were obtained from Maria Jasin.[29]

siRNAs, plasmids and site-directed mutagenesis

An mCherry-XRCC2 expression vector was created by inserting the human *XRCC2* cDNA sequence (OriGene, RG208330) into mCherry-C1. All selected variants were introduced into this vector using two-step site directed mutagenesis or, in case of variants in regions with high A/T content, by mutagenic overlap PCR followed by cloning. For both methods we made use of complementary primers containing the desired variant and an approximately 20bp sequence on both sites. The presence of the variants and the lack of additional mutations in the cDNA region were confirmed with Sanger sequencing (data not shown). An siRNA-resistant version of mCherry-XRCC2 was generated by mutating all wobble bases in the siRNA-targeting sequence using mutagenic overlap PCR such that the amino acid sequence of the XRCC2 protein would not change. Variants of interest were introduced in this vector using the aforementioned methods. In addition, wildtype and p.Arg17*-containing *XRCC2* cDNAs were amplified by PCR to remove the stop codon and introduce a C-terminal FLAG tag. The PCR products were subsequently inserted into mCherry-C1. All primer and siRNA sequences can be found in Supplementary Material, Tables S1 and S2.

Western blot analysis

Cell pellets we lysed in RIPA buffer and subsequently incubated at 95°C for 10 minutes in Laemmli buffer. Equal amounts of each sample were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis and transferred to PVDF membranes (Millipore).

Membranes were probed with primary antibodies to XRCC2 (N-20, 73278, Santa Cruz), mCherry (ab125096, Abcam) or FLAG (#F1804, Sigma-Aldrich), and appropriate secondary antibodies (LI-COR Biosciences) followed by protein detection using the Odyssey imaging scanning system (LI-COR Biosciences).

RAD51 foci formation after MMC treatment in *irs1*

The localization of RAD51 to DSBs was examined by assessing RAD51 foci formation after mitomycin C (MMC) treatment. On day one *irs1* cells were seeded in 12-wells plates containing 12mm coverslips at 30.000 cells /well. On day two the cells were transfected with 500ng mCherry or mCherry-XRCC2 expression vector using Effectene Transfection Reagent (Qiagen) according to the manufactor's protocol. On day three the medium containing the transfection reagents was removed and the cells were incubated for one hour with medium containing 6μM MMC. Cells were then incubated with fresh medium lacking MMC and six hours later fixed using 4% formaldehyde and permeabilized with 0.5% Triton X-100. Subsequently, the coverslips were incubated with 100mM glycine for 10 minutes and phosphate buffered saline containing 0.5% bovine serum albumin and 0.05% Tween 20 for 10 minutes. Finally, coverslips were incubated with anti-RAD51 antibody (H-92, Santa Cruz) at a 1:100 dilution for one hour, followed by incubation with goat anti-rabbit-488 (A-11034, Life Technologies) at 1:1000 and DAPI at 0.1 μg/ml for one hour. The fraction of mCherry-positive cells containing ≥5 RAD51 foci was scored manually. For each variant the experiment was repeated at least three times. Between 100 and 300 cells were scored in each experiment.

Sister chromatid recombination reporter assay in *irs1*

This sister chromatid recombination (SCR) reporter assay was essentially performed as described previously.[27] Briefly, *irs1* cells containing the SCR reporter were seeded at day one at 30.000 cells/well in a 12-well plate format. On day two the cells were transfected with 200ng of mCherry or mCherry-XRCC2 expression vector in combination with 600ng of either a I-SceI expression vector (pCBASce) or a control vector (pCAGGS).[29] Three days later the fraction of mCherry- and GFP-positive cells was determined on a LSRII flow cytometer (BD Biosciences). Data analysis was performed using Cytobank.[30] HR efficiencies were calculated as GFP-to-mCherry ratios. For each variant the experiment was repeated at least three times.

Microscopic analysis of fixed cells

Images of fixed samples were acquired on a Zeiss AxioImager M2 widefield fluorescence microscope equipped with 40x, 63x and 100x Planar apochromatic (1.4 numerical aperture) oil-immersion objectives (Zeiss) and an HXP 120 metal-halide lamp used for excitation. Fluorescent probes were detected using the following filters: DAPI (excitation filter: 350/50 nm, dichroic mirror: 400 nm, emission filter: 460/50 nm), GFP/Alexa 488 (excitation filter: 470/40 nm, dichroic mirror: 495 nm, emission filter: 525/50 nm) and mCherry (excitation filter: 560/40 nm, dichroic mirror: 585 nm, emission filter: 630/75 nm). Images were recorded using ZEN 2012 software (Zeiss).

DR-GFP reporter assay in HEK293

HEK293 cells carrying the DR-GFP reporter[29] were used to measure HR efficiencies essentially as described.[31] Briefly, on day one cells were seeded in 12-well plates at 225.000 cells/well. Cells were then transfected the same day and the second day with siRNA against *XRCC2* or a

control siRNA (Sigma, Universal Negative Control #1) at 20nM using Lipofectamine RNAiMax (Life Technologies) according to the manufacturer's protocol. At day three the cells were co-transfected with 500ng mCherry or siRNA-resistant mCherry-XRCC2 expression vector and 4µg of either an I-SceI expression vector (pCBASce) or a control vector (pCAGGS) using Lipofectamine 2000 (Life Technologies) (Pierce et al., 1999). Two days later cells were analyzed on a LSRII flow cytometer. Data analysis was performed using Cytobank.[30] HR efficiencies were calculated as GFP-to-mCherry ratios. For each variant the experiment was repeated at least three times. For cell cycle analysis, on day five cells were pulse labeled with 1µM EdU for two hours and then fixed and stained using the Click-iT EdU Flow Cytometry Kit (Invitrogen) according to protocol using an Alexa Fluor 488 dye. Knockdown of XRCC2 was validated using reverse-transcriptase quantitative PCR (RT-qPCR) as described in [32] using primer sets that either detect mRNA from the endogenous *XRCC2* or siRNA-resistant *XRCC2* cDNA (Table S1).

Table 1. In silico predictions and measured functional effect for all *XRCC2* genetic variants

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Poly-Phen-2	SIFT[a]	Align-GVGD[b]	CADD[c]	RAD51 foci - Hamster	SCR - Hamster	DR-GFP - Human
c.46 G>T [d]	p.Ala16Ser	(5)	Possibly damaging	Tolerated	C0	24.300	81%	100%	NA
c.49C>T	p.Arg17*	(5)	-	-	-	26.600	66%	111%	77%
c.140A>G	p.His47Arg	(15)	Possibly damaging	Tolerated	C0	24.000	90%	89%	NA
c.181C>A	p.Leu61Ile	(5)	Possibly damaging	Damaging	C0	22.600	83%	96%	NA
c.223G>C	p.Glu75Gln	(15)	Possibly damaging	Tolerated	C0	22.100	88%	89%	NA
c.247dup[e]	p.Thr83fs	-	-	-	-	29.000	11%	7%	48%
c.271C>T	p.Arg91Trp	(5)	Probably damaging	Damaging	C65	27.400	71%	66%	62%
c.283A>C [f]	p.Ile95Leu	(5)	Benign	Damaging	C0	0.001	77%	91%	70%
c.283A>G	p.Ile95Val	(5,15)	Benign	Tolerated	C0	0.002	90%	95%	NA
c.350delT	p.Leu117fs	(15)	-	-	-	29.400	51%	81%	56%
c.353T>C [g]	p.Val118Ala	(15)	Benign	Tolerated	C0	4.263	81%	101%	NA
c.359G>A	p.Cys120Tyr	(15)	Probably damaging	Damaging	C65	24.900	77%	70%	60%
c.398T>C	p.Leu133Pro	(15)	Probably damaging	Damaging	C65	24.000	62%	82%	67%
c.490G>C	p.Glu164Gln	(15)	Probably damaging	Damaging	C0	25.000	96%	75%	NA
c.509A>C	p.Glu170Ala	(15)	Probably damaging	Damaging	C65	27.400	101%	98%	NA
c.562C>T	p.Arg188Cys	(15)	Probably damaging	Damaging	C15	29.300	73%	102%	81%
c.563G>A [h]	p.Arg188His	(5,15)	Benign	Tolerated	C0	18.980	88%	86%	97%
c.581C>T	p.Thr194Met	(15)	Probably damaging	Damaging	C15	27.000	82%	92%	NA
c.595A>C	p.Met199Ile	(15)	Benign	Tolerated	C0	9.678	69%	97%	99%
c.620A>G [i]	p.Glu207Gly	(15)	Benign	Tolerated	C0	14.470	91%	84%	82%
c.643C>T	p.Arg215*	(6)	-	-	-	34.000	41%	79%	42%
c.651_652delTG	p.Cys217*	(5)	-	-	-	24.400	27%	61%	31%
c.659A>T	p. Asp220Val	(15)	Benign	Damaging	C0	13.880	86%	96%	NA
c.693G>T	p.Trp231Cys	(5)	Probably damaging	Damaging	C65	32.000	88%	88%	NA
c.714G>C	p.Arg238Ser	(15)	Probably damaging	Tolerated	C65	18.570	77%	94%	NA
c.742C>G [j]	p.Gln248Glu	(15)	Benign	Tolerated	C0	0.001	94%	92%	NA
c.772C>T	p.Arg258Cys	(15)	Benign	Tolerated	C0	22.500	71%	88%	108%
c.808T>G [k]	p.Phe270Val	(5,15)	Probably damaging	Damaging	C45	24.900	76%	85%	NA

All non-synonymous coding *XRCC2* variants from Park et al. (2012) and Hilbers et al. (2012) were selected for functional testing and in silico analysis by Polyphen-2, SIFT, AlignGVGD and CADD. In addition, a variant described by Shamseldin et al. (2012), detected in a Fanconi Anemia case and a truncating variant, p.Thr83fs reported by Douglas Easton (personal communication) were analyzed. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. [a] based on a 0.05 cut-off, [b] AlignGVGD depth until Branchiostoma floridae (Lancelet) Class 0 is the least likely to be damaging, Class C65 is the most likely to be damaging, [c] PHRED scaled C-score, a high score indicates a high change for the variant to be deleterious, a score of 20 is commonly used as a cut-off, [d] rs4987090, [e] reported by Douglas Easton (personal communication), [f] rs140214637, [g] rs185815454, [h] rs3218536, [i] rs61762969, [j] rs190900560 [k] rs145085742 .

Results

In silico prediction tools yield conflicting results

We selected all non-synonymous coding variants from the two major studies assessing the association of rare genetic variants in *XRCC2* and breast cancer risk for functional analysis in this study (Table 1).[6,15] Among these variants was p.Arg188His (rs3218536), the only variant in *XRCC2* common enough to directly establish its association with breast cancer. However, two meta-analyses, both including over 30,000 cases and controls, did not find association with breast cancer risk for this SNP.33,34 In addition we included a variant found in a Fanconi Anemia case (c.643C>T, p.Arg215*).[7] These variants comprised twenty-three amino acid changes, two deletions and two premature stop codons (Fig. 1A). In silico prediction algorithms PolyPhen-2,[16] SIFT,[17] AlignGVGD[18] and CADD[19] were used to predict if these variants would affect XRCC2 protein function. Polyphen and SIFT both use multiple sequence alignment to assess the effect of amino acid changes on protein function. Polyphen also takes into account information on protein domains and three-dimensional structures. AlignGVGD combines Grantham difference of amino acids with ortholog multiple sequence alignment, whereas CADD combines 63 different annotations, including PolyPhen, SIFT and Grantham differences of amino acids, to predict the effect of amino acid changes on protein function. The number of amino acid changes predicted to be "damaging" based on these prediction tools ranged from seven (AlignGVGD) to fourteen (CADD). Six variants, p.Arg91Trp, p.Cys120Tyr, p.Leu133Pro, p.Glu170Ala, p.Trp231Cys and p.Phe270Val (26%), were classified as damaging by all prediction algorithms. For eleven variants (48%) the algorithms yielded conflicting results and for six variants (26%) all agreed on them being benign. The level of disparity between the in silico tools demonstrates the need for more conclusive data on the functional effects of these variants.

Few XRCC2 variants affect RAD51 foci formation in hamster cells

To assess the effect of the missense variants on *XRCC2* function, we introduced all variants in a human XRCC2 cDNA expression vector and assessed to which extent these variants could complement the repair defect of XRCC2-deficient *irs1* hamster cells (Fig. 1B). One of the main phenotypic characteristics of *irs1* cells is that the accumulation of RAD51 into distinct foci containing DSBs is strongly reduced.[26,35] Indeed, we could confirm that in response to treatment with the DSB-inducing agent Mytomycin C (MMC) RAD51 foci formation was almost completely absent in *irs1* cells that were transfected with an mCherry expression vector (Fig. 1C). In contrast, the expression of mCherry-tagged human *XRCC2* restored RAD51 recruitment to sites of DNA damage in *irs1* cells. This is consistent with the results of previous studies showing that human *XRCC2* expression can complement the RAD51 foci formation defect in *irs1*.[26,35]

To assess the functionality of the selected *XRCC2* missense variants, we generated a collection of mCherry-tagged XRCC2 fusions containing these variants and measured their ability to restore RAD51 foci formation following MMC exposure in *irs1* cells (Fig. 1B). We found most missense variants, including the common SNP p.R188H for which no association with breast cancer risk was found, to rescue RAD51 foci formation to more than 75% of the levels seen for wild-type XRCC2 (Fig. 1D). Only five out of twenty-three variants were able to rescue RAD51 foci formation to levels below 75% (p.Arg91Trp, p.Leu133Pro, p.Arg188Cys, p.Met199Ile

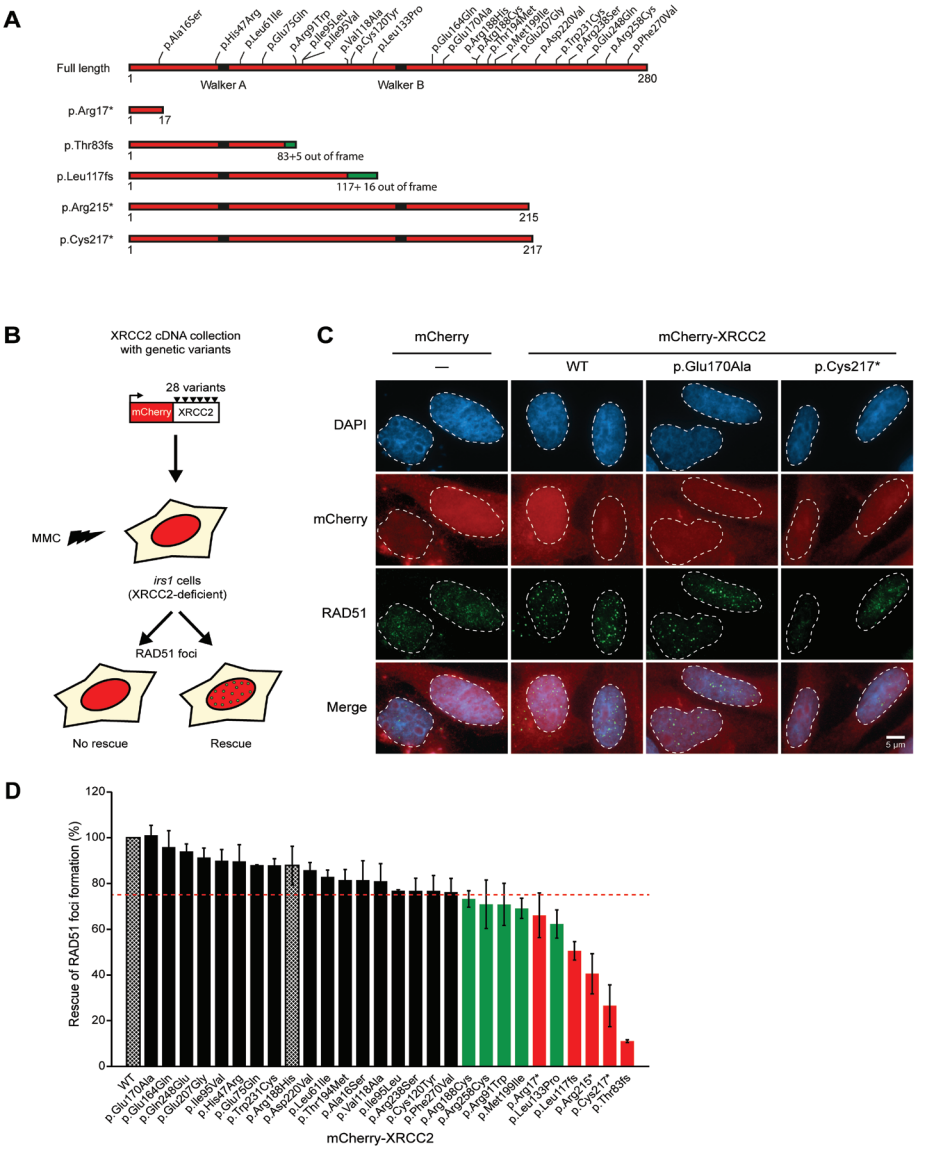


Figure 1. Functional analysis of the effect of *XRCC2* variants on RAD51 foci formation in hamster cells.
(A) Schematic overview of the *XRCC2* variants. Red indicates the in-frame *XRCC2* sequence, while green represents out-of-frame amino acids. (B) Workflow for the functional analysis of the effect of *XRCC2* variants on RAD51 foci formation in XRCC2-deficient *irs1* hamster cells. A collection of mCherry-tagged *XRCC2* cDNAs containing selected genetic variants was introduced in *irs1* cells. Following treatment of the cells with MMC, RAD51 foci formation was monitored to determine whether the *XRCC2* variants would rescue the defect in RAD51 foci formation of the *irs1* cells. (C) Examples of RAD51 foci formation in *irs1* cells expressing mCherry, mCherry-XRCC2 (wildtype; WT), mCherry-XRCC2 carrying a functional variant (p.E170A), and mCherry-XRCC2 carrying a protein-truncating variant (p.C217*). (D) Effect of *XRCC2* variants on the rescue of RAD51 foci formation in *irs1* cells. The percentage of cells with >5 MMC-induced RAD51 foci was scored. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

and p.Arg258Cys; from 62%-73%). Importantly, three protein-truncating variants, p.Leu177fs, p.Arg215* and p.Arg217*, rescued RAD51 foci formation to levels between 26% and 50%, suggesting impaired protein function. Unexpectedly, a fourth protein-truncating variant, p.Arg17*, showed a rescue of 66%. To verify that our assay was sensitive enough to detect differences in complementation for variants with a drastic effect on the protein produced, we also tested another variant, p.Thr83fs (Douglas Easton, personal communication), that would result in a truncated protein much shorter than the three giving a moderate effect. This variant rescued RAD51 foci formation to only 11%. Western blot analysis confirmed that expression of all protein-truncating variants produced a protein with a molecular weight consistent with the position of the premature stop-codon in the *XRCC2* coding sequence (Supplementary Material, Fig. S1), ruling out that the observed defects were due to impaired protein production. Thus, these results reveal that the truncating variants in *XRCC2* affect protein function, while most missense variants do not.

Few XRCC2 variants affect DSB repair via HR in hamster cells

RAD51 foci formation represents an intermediate step in the process of HR. Therefore, it is important to also establish the effect of genetic variants in *XRCC2* on the completion of DSB repair via this pathway. To this end, we employed *irs1* cells carrying a sister chromatid-recombination (SCR) reporter (*irs1*-SCR)[27] (Fig. 2A). This construct consists of two differentially mutated GFP sequences, one of which harbors an I-SceI restriction site (Fig. 2B). Expression of the I-SceI nuclease in these cells generates a DSB that is primarily repaired via HR, which results in the restoration of a functional GFP gene. Hence, the percentage of GFP-positive cells after I-SceI expression is a measure of the efficiency of HR. When we co-transfected *irs1*-SCR cells with an I-SceI expression and mCherry control vector hardly any GFP positive cells were detected, indicating that the cells are indeed incapable of repairing DSBs via HR. However, when mCherry-tagged wild-type *XRCC2* was introduced, a 16-fold increase in GFP-positive cells compared to mCherry alone was observed (Fig. 2C), comparable with previous results.[28] This also indicates that under our experimental conditions, HR can be rescued with an efficiency that allows us to assess the functionality of *XRCC2* variants.

We subsequently introduced our collection of mCherry-tagged *XRCC2* expression vectors containing the selected variants into the *irs1*-SCR cells and measured HR by flow cytometry. Similar to the results of our RAD51 foci analysis, we found that most missense variants, including the common SNP p.R188H, had no or only a very small effect on the ability to complement the HR repair defect in these cells (Fig. 2D). On average, the effects of these genetic variants in the SCR reporter assay were, however, smaller than those in the RAD51 foci formation assay. In fact, only three missense variants (p.Glu164Gln, p.Cys120Tyr and p.Arg91Trp) rescued to 75% or slightly less when compared to wild-type XRCC2. Surprisingly, for three of the four protein-truncating variants the effects were also smaller than those seen in the RAD51 foci analysis, ranging from 60% to 81% rescue. However, for the shortest truncated protein (p.Thr83fs), we observed a very strong defect in HR with only 7% rescue compared to wild-type, indicating that the assay is capable of identifying alleles that have impaired protein function. Interestingly, p.Arg17*, the protein-truncating variant that showed unexpectedly good rescue in the RAD51 foci assay, also fully restored HR in this assay. Overall the results from the sister chromatid reporter assay concur with those from the RAD51 foci assay and indicate that only a small minority of the missense variants affects XRCC2 function.

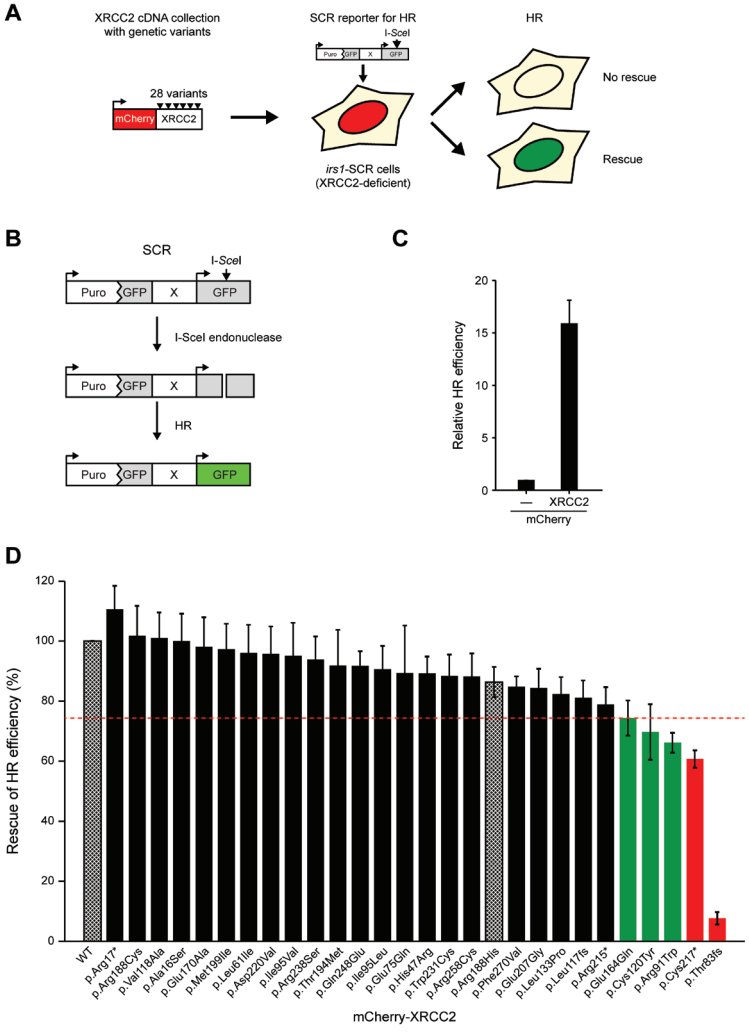


Figure 2. Functional analysis of the effect of *XRCC2* variants on HR in hamster cells.
(A) Workflow for the functional analysis of the effect of *XRCC2* variants on HR in XRCC2-deficient *irs1* hamster cells. A collection of mCherry-tagged XRCC2 cDNAs containing selected genetic variants was introduced in *irs1* cells containing the SCR reporter for HR (*irs1*-SCR cells). Following expression of the I-SceI nuclease, the fraction of GFP-expressing cells among the mCherry-positive cells was determined to assess whether the *XRCC2* variants would rescue the HR defect of *irs1*-SCR cells. (B) Schematic of the sister chromatid reporter (SCR) assay. (C) Relative HR efficiency of *irs1*-SCR cells expressing mCherry-XRCC2 compared to mCherry alone. The repair efficiency in cells expressing mCherry was set to 1. (D) Effect of *XRCC2* variants on the rescue of HR in *irs1*-SCR cells. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

Validation of the effect of XRCC2 variants on HR in human cells

Given our observation that only few *XRCC2* variants fail to complement the repair phenotype of *irs1* hamster cells, we next asked if these variants would show a similar effect in human cells. To this end, we used human HEK293 cells carrying the HR reporter DR-GFP to validate the results of a subset of *XRCC2* variants[29] (Fig. 3A). Like the SCR reporter in *irs1* cells, the DR-GFP reporter allows us to measure the efficiency of HR by flow cytometry as the percentage of GFP-positive cells after I-SceI expression (Fig. 3B). We used siRNA-mediated knockdown to reduce the expression of endogenous XRCC2 in cells that either expressed mCherry alone or an siRNA-resistant version of mCherry-tagged XRCC2. Western blot analysis and quantitative reverse transcription PCR (RT-qPCR) with primers that specifically recognize transcripts from the endogenous *XRCC2* gene confirmed the efficient knockdown at the protein and mRNA level, respectively (Fig. 3C and 3D). Importantly, knockdown of endogenous XRCC2 led to an ~80% reduction in HR efficiency in mCherry-expressing cells, indicating a clear defect in executing HR (Fig. 3E). Expression of mCherry-XRCC2 restored HR levels to almost 70% of that observed in control cells expressing mCherry (Fig. 3E), showing that our experimental setup allows the functional characterization of *XRCC2* variants. We selected all variants that showed <75% rescue compared to wild-type in at least one of the two *irs1* assays in addition to a subset of variants that showed full rescue in either assay and examined their ability to rescue the HR-deficient phenotype of human XRCC2-depleted cells. The protein-truncating variants p.Thr83fs, p.Leu117fs, p.Arg215* and p.Cys217* restored the HR-defect to 48%, 56%, 42% and 31%, respectively (Fig. 3F). As observed before, the truncating variant p.Arg17* showed a more moderate effect and rescued HR up to 67%. We considered the possibility that this truncating mutant showed a considerable rescue in all functional assays due to re-initiation of translation downstream of the stop codon (Supplementary Material, Fig. S2A). To study this, we expressed mCherry-XRCC2-FLAG and mCherry-XRCC2-p.Arg17*-FLAG constructs in HEK293T cells and examined the expression of mCherry- and FLAG-tagged XRCC2 products by Western blot analysis using antibodies against mCherry or FLAG (Supplementary Material, Fig. S2B). While we detected full-length mCherry-XRCC2-FLAG using either the mCherry or FLAG antibody, we detected the presence of a truncated mCherry-XRCC2 protein in mCherry-XRCC2-p.Arg17*-FLAG- expressing cells when using the mCherry antibody. We were unable to detect expression of a truncated XRCC2-FLAG protein in these cells when using the FLAG antibody (Supplementary Material, Fig. S2B). This suggests there is no re-initiation of translation downstream of the stop codon. Surprisingly, when we examined higher exposure Western blots, we detected low levels of mCherry-XRCC2-FLAG protein in mCherry-XRCC2-p.Arg17*-FLAG expressing cells. This suggests a certain degree of read-through of the stop codon introduced by the p.Arg17* mutation (Supplementary Material, Fig. S2B and S2C). Indeed, the p.Arg17* mutation introduces a TGA stop codon, which has been reported to result in translational read-through in mammalian genes.36 This may explain why we observed a considerable rescue for p.Arg17* in the functional assays. Finally, also most missense variants showed considerable rescue of the HR defect with the common polymorphism p.Arg188His even up to 97%, while some others, such as p.Cys120Tyr were less efficient in restoring HR (~60% rescue; Fig. 3F).

Given that HR is most prominent in S-phase,[37] we examined if the expression of *XRCC2* variants could impacted HR indirectly by affecting cell cycle progression. Flow cytometry analysis, however, did not show any major differences in cell cycle progression between HEK293 DR-GFP cells expressing wildtype or variant *XRCC2* (Supplementary Material, Fig. S3), indicating that the outcome of the assay was not influenced by alterations in cell cycle

distribution. Thus, overall our results in human cells confirm that truncating variants more strongly impair XRCC2's DNA repair function compared to missense variants.

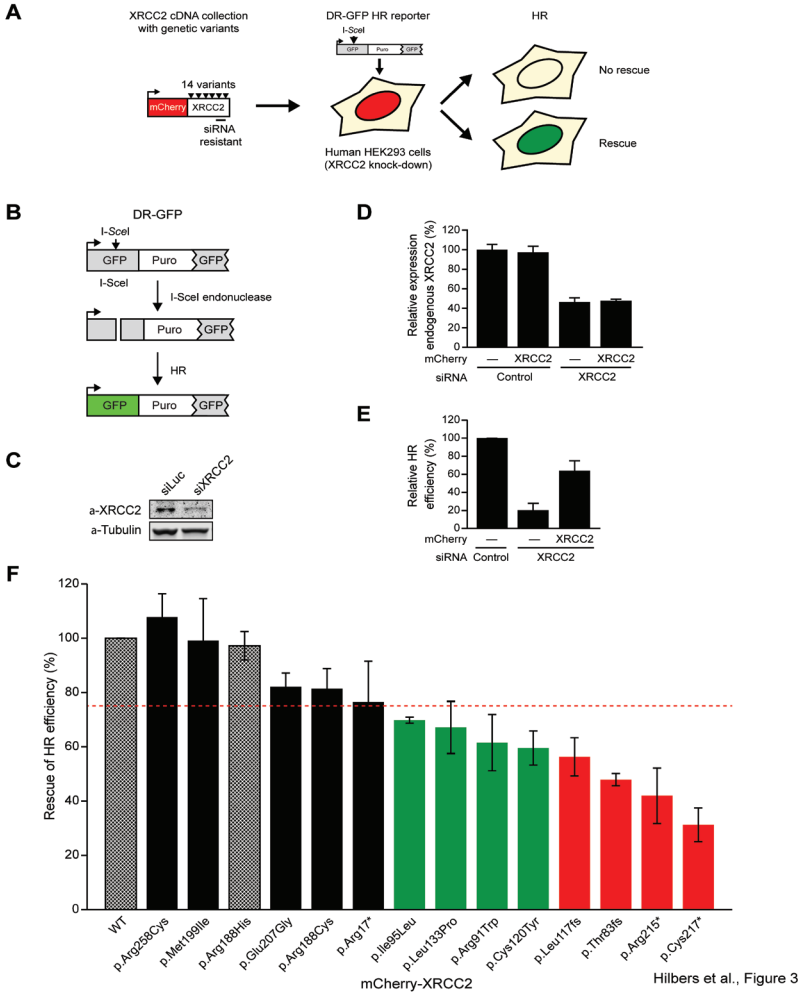


Figure 3. Functional analysis of the effect of *XRCC2* variants on HR in human cells.
(A) Workflow for the functional analysis of the effect of *XRCC2* variants on HR in human cells. A collection of mCherry-tagged XRCC2 cDNAs containing selected genetic variants was introduced in HEK293 cells containing the DR-GFP reporter for HR after knockdown of *XRCC2* by siRNA interference. Following expression of the I-SceI nuclease, the fraction of GFP-expressing cells among mCherry-positive cells was determined to assess whether the *XRCC2* variants would rescue the HR defect of the *XRCC2* knockdown cells. (B) Schematic of the DR-GFP reporter assay. (C) Validation of *XRCC2* knock-down after siRNA transfection (80 nM) of HEK293T cells using a specific XRCC2 antibody. (D) Relative expression of endogenous *XRCC2* in HEK293 cells expressing siRNA-resistant wild-type mCherry-XRCC2 cDNA or mCherry alone, following treatment with control or *XRCC2* siRNAs. Expression was determined by RT-qPCR using primers specific for the endogenous *XRCC2* transcript. The XRCC2 expression level in control cells expressing mCherry alone was set to 100%. (E) Effect of siRNA-mediated XRCC2 knockdown and subsequent re-expression of siRNA-resistant mCherry-XRCC2 cDNA on HR in HEK293 DR-GFP cells. The bars indicate the relative percentage of GFP-positive cells after DSB induction by I-SceI. The HR efficiency of mCherry-expressing cells transfected with control siRNAs were set to 100%. (F) Effect of XRCC2 variants on the rescue of HR in XRCC2-depleted HEK293 DR-GFP cells. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

The outcome of functional assays correlates poorly with in silico predictions

In order to assess the robustness of the functional assays, we examined the correlation between the three HR assays (RAD51 foci and GFP-based reporter assays in hamster cells and human cells) in more detail. We found moderately high Pearson correlation coefficients for these assays ranging from 0.57 to 0.72 (Fig. 4A-C). This degree of correlation likely reflects the variation within the assays and the fact that each measures different but related features. We also compared the results of the assays with the predictions of the in silico algorithms. For the prediction tools that give categorical results for amino acid changes (PolyPhen2, SIFT and AlignGVGD), there seems to be no correlation with protein function in either of the assays (Supplementary Material, Fig. S4). Likewise, we found that CADD, which assigns a continuous prediction score instead of an effect category to each variant, shows a rather poor correlation with all three assays ranging from -0.31 to -0.45 (Fig. 4D-F). Based on this we conclude that the outcome of these predictive algorithms poorly correlates with the effect of missense variants on XRCC2's function in DNA repair.

Lack of association with breast cancer upon functional reclassification of variants

Based on the results of our functional assays, we reclassified the *XRCC2* variants and re-assessed their association with breast cancer. We used the results of the common SNP p.R188H as a benchmark for neutrality. The rescue of HR found for this variant ranged between 86% to 97% when compared to wild-type in the different assays. Any variant showing a similar or better rescue of HR in these assays is unlikely to have an association with breast cancer. In contrast, no individual *XRCC2* variant has been convincingly shown to increase breast cancer risk. The only variant with a demonstrable clinical association is the p.R215* variant found in the Saudi Arabian Fanconi Anemia patient. Given the overlap between Fanconi Anemia genes and breast cancer susceptibility genes, this variant currently gives the best estimation of the range of functional effects that could be relevant for breast cancer susceptibility. The variant shows an HR rescue of 41%, 79% and 42% in the RAD51 foci formation assay, SCR assay and DR-GFP assay, respectively.

Using these extremes we defined three functional categories. One category with variants strongly affecting XRCC2 function, with <50% HR rescue in at least two out of three assays. This category contains the protein-truncating variants p.Thr83fs, p.Arg215*and p.Cys217*. A second set of variants with a moderate effect on XRCC2 function defined as 50-75% HR rescue in two out of three assays. This category contains missense variants, p.Arg91Trp, p.Ile95Leu, p.Cys120Tyr and p.Leu133Pro, and the frame-shifting variant p.Leu117fs. The last category includes variants that showed 75-100% HR rescue in at least two assays. This category includes all other missense variants and protein-truncating variant p.Arg17* (for an overview of the results in all three assays Table 1). When we reexamine the association data provided by Park et al.[6] and Hilbers et al.[15] we do not find any evidence for an association with breast cancer (Table 2). For the >50% and 25-50% reduction in complementation categories combined we find an odds ratio of 0.75 (95% CI: 0.18-3.13). It should, however, be noted that the number of individuals with a variant in the <50% rescue category is too low for a meaningful association analysis. An association with breast cancer can therefore not be excluded for these variants. In conclusion, our functional classification data suggest that it is highly unlikely that the missense variants analyzed in this study are associated with breast cancer risk.

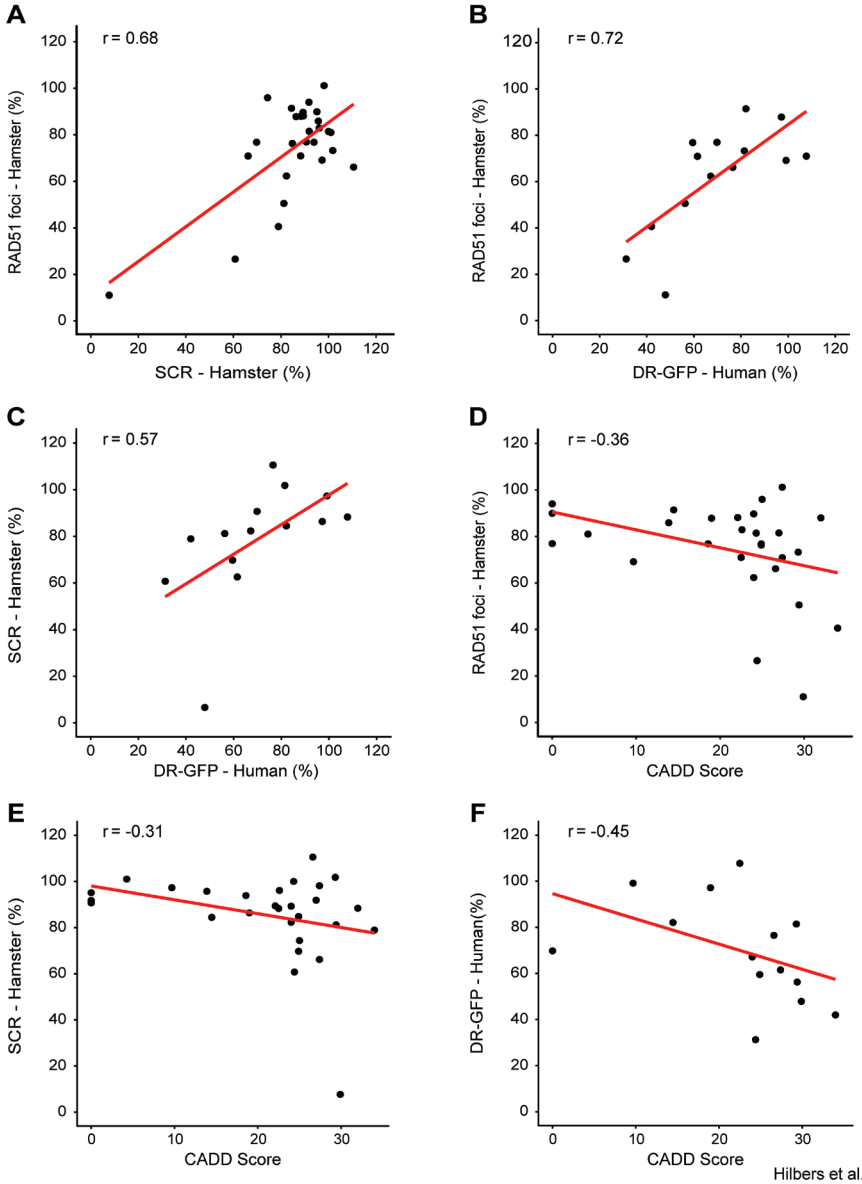


Figure 4. Correlation between the outcome of functional assays and in silico analysis.
(A) Correlation between the outcome of RAD51 foci formation and SCR assays in hamster cells, (B) RAD51 foci formation assay in hamster cells and DR-GFP assays in human cells, (C) SCR assays in hamster cells and DR-GFP assays in human cells, (D) RAD51 foci formation assays in hamster cells and CADD prediction analysis, (E) SCR assays in hamster cells and CADD prediction analysis, and (F) RAD51 foci formation assays in hamster cells and CADD prediction analysis. r indicates the Pearson correlation coefficient.

Table 2. Association analysis of functionally characterized XRCC2 variants

	Degree of variant complementation	Cases (%) (n=2147)	Controls (%) (n=1120)	
Park et al. (5)	**All**	12 (0.56)	1 (0.09)	
	75-100%	7 (0.33)	1 (0.09)	
	50-75%	3 (0.14)	0 (0)	
	<50%	2 (0.09)	0 (0)	
	<75%	5 (0.23)	0 (0)	
	Degree of variant complementation	**Cases (%) (n=3548)**	**Controls (%) (n=1435)**	
Hilbers et al. (15)	**All**	18 (0.51)	10 (0.70)	
	75-100%	18 (0.51)	7 (0.49)	
	50-75%	0 (0)	3 (0.21)	
	<50%	0 (0)	0 (0)	
	<75%	0 (0)	3 (0.21)	
	Degree of variant complementation	**Cases (%) (n=5695)**	**Controls (%) (n=2555)**	**OR (95% CI)**
Total	**All**	30 (0.53)	11 (0.43)	1.22 (0.61-2.45)
	75-100%	25 (0.44)	8 (0.31)	1.40 (0.63-3.12)
	50-75%	3 (0.05)	3 (0.12)	0.45 (0.09-2.22)
	<50%	2 (0.04)	0 (0)	NA
	<75%	5 (0.09)	3 (0.12)	0.75 (0.18-3.13)

The *XRCC2* genetic variants reported by Park et al. (2012) and Hilbers et al. (2012) were classified according to the level of complementation compared to wild-type.

Discussion

Here we report the functional analysis of all *XRCC2* genetic variants detected previously in two large case-control studies,[6,15] which assessed their association with breast cancer risk. We show that most non-synonymous variants have no, or only a small effect on XRCC2 function, with the exception of p.Arg91Trp, p.Ile95Leu, p.Cys120Tyr and p.Leu133Pro, which displayed moderate effects in our functional assays. While three protein-truncating variants had a strong effect on XRCC2 function in our assays, two other such variants, p.Arg17* and p.Leu117fs show unexpectedly good complementation of the XRCC2-deficient phenotypes. Upon reclassification of the *XRCC2* genetic variants according to the results of our functional analyses, the originally reported association with breast cancer was no longer evident.

Most prediction algorithms classify protein-truncating variants as pathogenic and this is also predicted for p.Arg17*. However, a closer inspection of this variant, which showed unexpected functionality in our assays, suggested that the introduced TGA stop codon frequently leads to translational read-through as has been reported for a number of mammalian genes.[36] Indeed, we can detect small amounts of full-length XRCC2 protein produced from this allele, which is a likely explanation for the considerable rescue of p.Arg17*

in the functional assays. However, we should note that it is unclear whether the endogenous *XRCC2* sequence harboring this mutation would produce sufficient amounts of full-length protein to support HR at normal levels. Nonetheless, these results suggest a mechanism through which protein-truncating mutations might still retain function and illustrate that our functional assays are sensitive enough to detect this phenomenon.

Two *XRCC2* variants assayed in our study have previously been studied by others. The common SNP p.Arg188His was shown by Rafii and colleagues to have a negligible effect on XRCC2 function in conferring resistance to Mitomycin C in clonogenic survival assay,[38] consistent with the results of our three functional assays. The frame-shifting variant p.Leu117fs has previously been identified in the mismatch-repair deficient uterine sarcoma cell line SKUT-1.[39] Functional analysis of this variant has shown that, although it has a reduced function compared to wild-type XRCC2, the effect is not as strong as other truncating or frame-shifting mutations in the same region.[40] In our analysis, p.Leu117fs retained approximately 55% function compared to wild-type, whereas more C-terminal variants resulting in protein truncation seem to have stronger inactivating effects. Although our findings are consistent with the previously reported results,[40] we can at this moment not explain why p.Leu117fs has this relatively mild effect.

The two main conserved domains in XRCC2 are the ATP-binding Walker A (amino acids 48-55) and Walker B (amino acids 145-149) motifs. It has been shown that mutations in the most conserved part of the Walker A motif do not affect XRCC2 function.[26,28] Moreover, one study suggested that the Walker A motif is important for the interaction between XRCC2 and RAD51D, whereas another study suggested that the RAD51D-binding region of XRCC2 is located in the N-terminus of XRCC2 (amino acids 1-42).[40] On the other hand, the C-terminal half of the XRCC2 protein (amino acids 144-280) is required for the interaction with RAD51.40 Given that the variants p.Leu117fs, p.Arg215* and p.Cys217*, which result in the expression of C-terminally truncated XRCC2 proteins, which are still partly functional in our assays, it is unlikely that the localization of RAD51 and the repair of DSBs via HR only depends on this interaction. The notion that the Walker A motif does not seem to be important for XRCC2 function might explain why all in silico prediction tools overestimated the number of missense variants affecting XRCC2 function, as these tools weigh protein conservation so heavily. XRCC2 is one of five RAD51-paralogs that have no functional redundancy.[35] Especially in the case of a paralogue, which originates from a segmental duplication and subsequently evolved into a gene with a different function, in silico predictions tools based on sequence conservation might have trouble correctly predicting functional effects. Interestingly, the four missense variants with the strongest effect in our assays are all located in between the Walker A and Walker B motif. This region has no known function at present, but might affect the three-dimensional structure of the protein.

It remains possible that because of limitations of the assays we have used, we missed certain variant effects. For example, the use of cDNA-constructs instead of a construct with an intron-exon structure, does not allow the detection of effects of variants on RNA splicing. In fact, our in silico analysis of possible effects on splicing suggested that 4 of the 28 variants examined in this study may affect splicing. Whether this prediction has relevance for the splicing of *XRCC2* mRNAs in vivo remains to be established. In addition, we are not able to detect the effect of nonsense mediated decay (NMD) on variants resulting in a premature stop-codon. *XRCC2* consists of three exons of which the third exon is by far the largest. All truncating variants except p.R17* are located in this last exon. Given that NMD is thought to only affect mRNA molecules with a premature stop codon upstream of the last exon-exon

junction, it seems highly unlikely that this process will affect any of the truncating variants other than p.R17*.[41]

Crucial to the functional analysis of variants in any gene, is a full understanding of the processes the encoded protein is involved in. In the case of XRCC2 it is widely accepted that its main function is in RAD51-dependent HR. There are however some indications that XRCC2 is also involved in other pathways. A study in Arabidopsis thaliana has implicated a role of XRCC2 in the repair of DSBs via RAD51-independent single-strand annealing, a pathway that by definition is mutagenic.[42] In addition, it has been shown that XRCC2, like many other HR proteins, is also involved in centrosome integrity and correct segregation of chromosomes.[43,44] However, it is not known what its exact role is in this process and whether it depends on interactions other than those important for HR. We can therefore not exclude that the variants that showed full complementation of the XRCC2-deficient phenotype in our assays, affect an XRCC2-regulated process that was not assessed in our study. A last limitation of our study is the use of transient transfections. For example, the use of siRNAs in the DR-GFP assay resulted in a partial knockdown of endogenous XRCC2. A stable knockdown or knockout of XRCC2, would most likely reduce inter-experiment variability and might provide a larger window for functional analysis of *XRCC2* variants.

In general, this study demonstrates the pitfalls in establishing an association between rare genetic variation and disease risk. Now that next-generation sequencing makes it possible to explore variation in large regions of the genome at reasonable costs at epidemiological scale, the number of extremely rare variants with a possible relation to disease risk will increase tremendously. Likewise, diagnostic testing of breast cancer genes other than *BRCA1* and *BRCA2* is likely to result in a steep increase in variants of uncertain significance. When no extensively sampled families are available for co-segregation analysis of individual variants, a gene-centered burden analysis is the only way to establish a disease association, but this crucially depends on their correct classification. Functional analyses as shown here for *XRCC2* variants can prove a crucial step in this process. To this end, future efforts should establish a more sensitive and robust cellular system that allows for the high-throughput testing of large numbers of variants in (putative) breast cancer susceptibility genes.

In conclusion, this study has shown that most rare *XRCC2* missense variants have little effect on XRCC2 protein function and are unlikely to be associated with familial and early onset breast cancer. In the group of *XRCC2* variants that resulted in a >25% reduction in complementation no evidence of an association with breast cancer has been found. Although we cannot exclude that the few *XRCC2* variants with a very strong effect on protein function increase breast cancer risk, these results suggests that the previously reported association with breast cancer is a false-positive finding. Unfortunately, none of the studies published to date had sufficient families in which one of the variants showing reduced complementation was found and for which DNA samples from additional family members were available for a co-segregation analysis. Much larger case-control studies are necessary to provide conclusive results on the association between the variants with >50% reduction in complementation and disease risk. Nevertheless, based on the current data, we conclude that *XRCC2* variants are unlikely to explain a significant fraction of the familial risk to breast cancer.

Funding

This work was supported by the Netherlands Organisation for Scientific Research [NWO-VENI to M.S.L.]; and the Dutch Cancer Society [KWF UL 2009-4388 to P.D., KWF UL 2014-6853 to P.D.,

C.J.v.A. and H.v.A.].

Acknowledgements

The authors would like to thank John Thacker, Ralph Scully and Maria Jasin for generously providing *irs1*, *irs1*-SCR and HEK293 DR-GFP cells, respectively, and Maaike Vreeswijk for help with the in silico splice prediction analysis.

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. 2001;358(9291):1389-1399. doi:10.1016/S0140-6736(01)06524-2
2. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer. 2000;83(10):1301-1308. doi:10.1054/bjoc.2000.1407
3. Peto J, Collins N, Barfoot R, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst. 1999;91(11):943-949. doi:10.1093/jnci/91.11.943
4. Lalloo F, Evans DG. Familial breast cancer. Clin Genet. 2012;82(2):105-114. doi:10.1111/j.1399-0004.2012.01859.x
5. Easton DF, Pharoah PDP, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-2257. doi:10.1056/NEJMsr1501341
6. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
7. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. J Med Genet. 2012;49(3):184-186. doi:10.1136/jmedgenet-2011-100585
8. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497-506. doi:10.1056/NEJMoa1400382
9. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med. 2010;362(20):1909-1919. doi:10.1056/NEJMra0809889
10. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. Science. 2002;297(5581):606-609. doi:10.1126/science.1073834
11. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci USA. 2014;111(42):15172-15177. doi:10.1073/pnas.1407909111
12. Sawyer SL, Tian L, Kähkönen M, et al. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov. 2015;5(2):135-142. doi:10.1158/2159-8290.CD-14-1156
13. Chun J, Buechelmaier ES, Powell SN. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the. Mol Cell Biol. 2013;33(2):387-395. doi:10.1128/MCB.00465-12
14. Krejci L, Altmannova V, Spirek M, Zhao X. Homologous recombination and its regulation. Nucleic Acids Res. 2012;40(13):5795-5818. doi:10.1093/nar/gks270
15. Hilbers FS, Wijnen JT, Hoogerbrugge N, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. J Med Genet. 2012;49(10):618-620. doi:10.1136/jmedgenet-2012-101191
16. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20. doi:10.1002/0471142905.hg0720s76
17. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-3814. doi:10.1093/nar/gkg509
18. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet. 2006;43(4):295-305. doi:10.1136/jmg.2005.033878
19. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-315. doi:10.1038/ng.2892
20. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2-3):377-394. doi:10.1089/1066527041410418
21. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J

Comput Biol. 1997;4(3):311-323. doi:10.1089/cmb.1997.4.311
22. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 2001;29(5):1185-1190. doi:10.1093/nar/29.5.1185
23. Desmet F-O, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67. doi:10.1093/nar/gkp215
24. Thery JC, Krieger S, Gaildrat P, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. Eur J Hum Genet. 2011;19(10):1052-1058. doi:10.1038/ejhg.2011.100
25. Jones NJ, Cox R, Thacker J. Isolation and cross-sensitivity of X-ray-sensitive mutants of V79-4 hamster cells. Mutat Res. 1987;183(3):279-286. doi:10.1016/0167-8817(87)90011-3
26. O'Regan P, Wilson C, Townsend S, Thacker J. XRCC2 is a nuclear RAD51-like protein required for damage-dependent RAD51 focus formation without the need for ATP binding. J Biol Chem. 2001;276(25):22148-22153. doi:10.1074/jbc.M102396200
27. Puget N, Knowlton M, Scully R. Molecular analysis of sister chromatid recombination in mammalian cells. DNA Repair (Amst). 2005;4(2):149-161. doi:10.1016/j.dnarep.2004.08.010
28. Nagaraju G, Hartlerode A, Kwok A, Chandramouly G, Scully R. XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversions between sister chromatids. Mol Cell Biol. 2009;29(15):4283-4294. doi:10.1128/MCB.01406-08
29. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. Genes Dev. 1999;13(20):2633-2638. doi:10.1101/gad.13.20.2633
30. Kotecha N, Krutzik PO, Irish JM. Web-based analysis and publication of flow cytometry experiments. Curr Protoc Cytom. 2010;Chapter 10:Unit10.17. doi:10.1002/0471142956.cy1017s53
31. Smeenk G, Wiegant WW, Marteijn JA, et al. Poly(ADP-ribosyl)ation links the chromatin remodeler SMARCA5/SNF2H to. J Cell Sci. 2013;126(Pt 4):889-903. doi:10.1242/jcs.109413
32. Typas D, Luijsterburg MS, Wiegant WW, et al. The de-ubiquitylating enzymes USP26 and USP37 regulate homologous recombination by counteracting RAP80. Nucleic Acids Res. 2015;43(14):6919-6933. doi:10.1093/nar/gkv613
33. He Y, Zhang Y, Jin C, et al. Impact of XRCC2 Arg188His polymorphism on cancer susceptibility: a meta-analysis. PLoS One. 2014;9(3):e91202. doi:10.1371/journal.pone.0091202
34. Yu K-D, Chen A-X, Qiu L-X, Fan L, Yang C, Shao Z-M. XRCC2 Arg188His polymorphism is not directly associated with breast cancer risk: evidence from 37,369 subjects. Breast Cancer Res Treat. 2010;123(1):219-225. doi:10.1007/s10549-010-0753-y
35. Takata M, Sasaki MS, Tachiiri S, et al. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. Mol Cell Biol. 2001;21(8):2858-2866. doi:10.1128/MCB.21.8.2858-2866.2001
36. Loughran G, Chou M-Y, Ivanov IP, et al. Evidence of efficient stop codon readthrough in four mammalian genes. Nucleic Acids Res. 2014;42(14):8928-8938. doi:10.1093/nar/gku608
37. Karanam K, Kafri R, Loewer A, Lahav G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. Mol Cell. 2012;47(2):320-329. doi:10.1016/j.molcel.2012.05.052
38. Rafii S, O'Regan P, Xinarianos G, et al. A potential role for the XRCC2 R188H polymorphic site in DNA-damage repair and breast cancer. Hum Mol Genet. 2002;11(12):1433-1438. doi:10.1093/hmg/11.12.1433
39. Mohindra A, Hays LE, Phillips EN, Preston BD, Helleday T, Meuth M. Defects in homologous recombination repair in mismatch-repair-deficient tumour cell lines. Hum Mol Genet. 2002;11(18):2189-2200. doi:10.1093/hmg/11.18.2189
40. Tambini CE, Spink KG, Ross CJ, Hill MA, Thacker J. The importance of XRCC2 in RAD51-related DNA damage repair. DNA Repair (Amst). 2010;9(5):517-525. doi:10.1016/j.

dnarep.2010.01.016
41. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. Nat Rev Genet. 2012;13(4):246-259. doi:10.1038/nrg3160
42. Serra H, Da Ines O, Degroote F, Gallego ME, White CI. Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. PLoS Genet. 2013;9(11):e1003971. doi:10.1371/journal.pgen.1003971
43. Cappelli E, Townsend S, Griffin C, Thacker J. Homologous recombination proteins are associated with centrosomes and are required for mitotic stability. Exp Cell Res. 2011;317(8):1203-1213. doi:10.1016/j.yexcr.2011.01.021
44. Griffin CS, Simpson PJ, Wilson CR, Thacker J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. Nat Cell Biol. 2000;2(10):757-761. doi:10.1038/35036399

Supplementary data

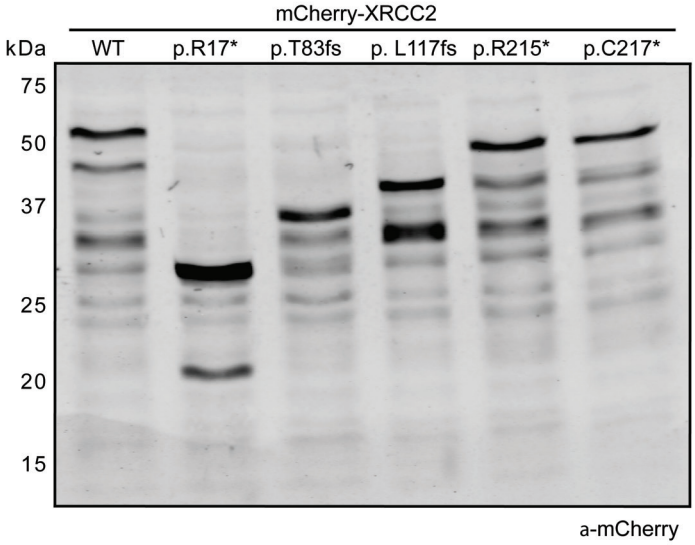


Figure S1. Western blot analysis of selected *XRCC2* variants.
Wild-type (WT) mCherry-XRCC2 and mCherry-XRCC2 carrying different truncating variants were expressed in HEK293 cells. Cell lysates were prepared and subjected to Western blot (WB) analysis using anti-mCherry antibody.

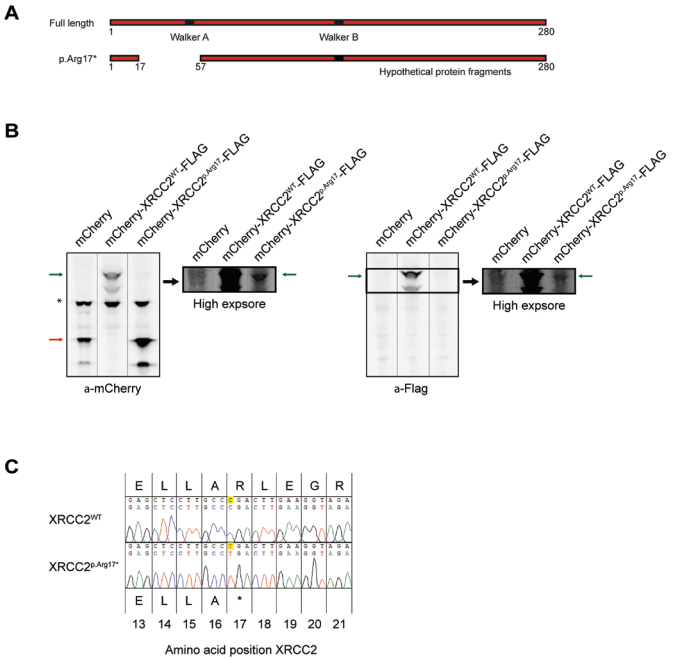


Figure S2.
(A) Schematic overview of *XRCC2* wildtype and XRCC2-p.Arg17* mutant proteins. The hypothetical protein generated by re-initiation of translation downstream of the stop codon introduced by the p.Arg17* mutation is shown. (B) Western blot analysis of HEK293T cells expressing mCherry, mCherry-XRCC2-FLAG or mCherry-XRCC2p.Arg17-FLAG. The star (*) indicates a non-specific band detected with the mCherry antibody. The red arrow indicates the position of mCherry protein. The green arrow indicates the position of the mCherry-XRCC2-FLAG fusion protein. (C) Mutation detection analysis by sequencing of mCherry-XRCC2-FLAG and mCherry-XRCC2p.Arg17-FLAG constructs. Amino acids and corresponding DNA sequences in *XRCC2* are shown.

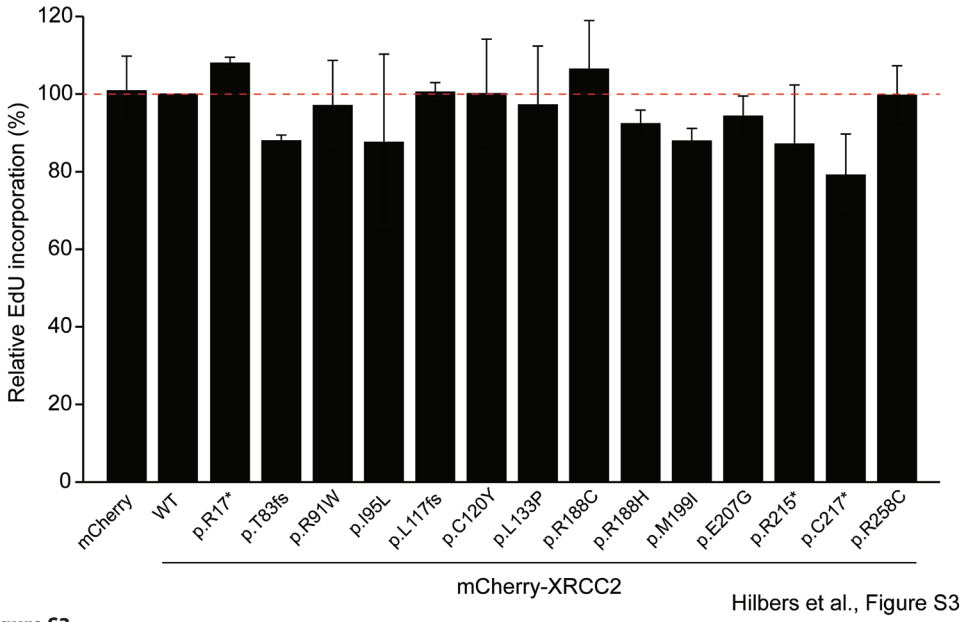


Figure S3.
Cell cycle analysis of HEK293 DR-GFP cells expressing selected *XRCC2* variants. *XRCC2* knockdown cells expressing I-SceI and either mCherry alone, wild-type (WT) mCherry-XRCC2 or different mCherry-XRCC2 variants were labeled with EdU and examined by flow cytometry to determine the fraction of cells in early, mid and late S phase. The fraction of EdU-positive cells for each sample is presented relative to that for the mCherry-XRCC2 sample, which was set to 100%.

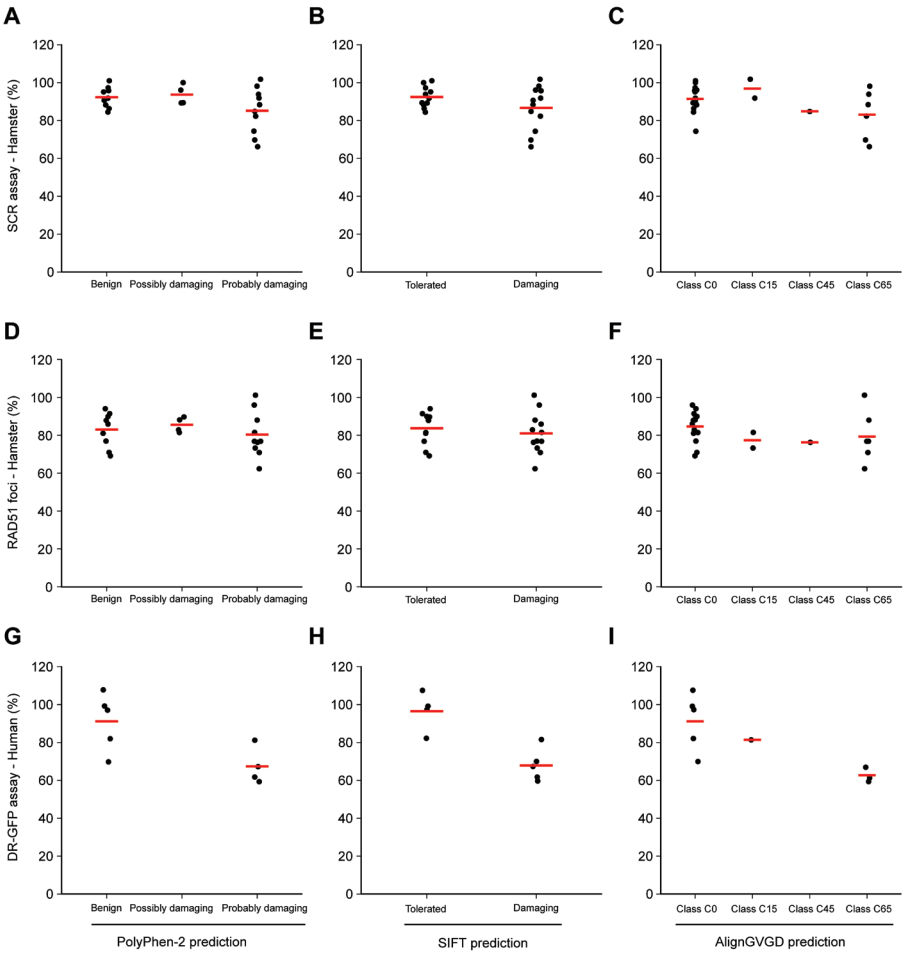


Figure S4.
Correlation between the outcome of functional assays and in silico predictions. The correlation between the functional effects of all *XRCC2* missense variants as found in the SCR assay in hamster cells and the predictions from in silico analysis by (A) PolyPhen-2, (B) SIFT and (C) AlignGVGD are shown. (D-F) As in A-C, except that the functional effects of *XRCC2* variants on RAD51 foci formation in hamster cells were compared. (G-I) As in A-C, except that the functional effects of *XRCC2* variants on HR in the DR-GFP assay in HEK293 cells were compared.

Table S1. Primers used for PCR, site-directed mutagenesis and RT-qPCR

Description		Forward primer	Reverse primer
SDM[a]	c.46 G>T	GGACCGAGCTCCTTTCCCGACTTGAAGG	CCTTCAAGTCGGGAAAGGAGCTCGGTCC
	c.49C>T	CCGAGCTCCTTGCCTGACTTGAAGGTAGAAG	CTTCTACCTTCAAGTCAGGCAAGGAGCTCGG
	c.140A>G	GGTGATATTCTTGAATTTCGTGGCCCAGAAGGAACAGG	CCTGTTCCTTCTGGGCCACGAAATTCAAGAATATCACC
	c.181C>A	CAGAAATGCTTTATCACATAACAGCACGATGTATACTTCCC	GGGAAGTATACATCGTGCTGTTATGTGATAAAGCATTTCTG
	c.223G>C	CCAAATCAGAAGGTGGCCTGCAAGTAGAAGTCTTATTTATT	AATAAATAAGACTTCTACTTGCAGGCCACCTTCTGATTTGG
	c.247dup	CCTGGAAGTAGAAGTCTTATTTATTGATAACAGATTACCACTTTGATATGCTCC	GGAGCATATCAAAGTGGTAATCTGTTATCAATAAATAAGACTTCTACTTCCAGG
	c.271C>T	CCACTTTGATATGCTCTGGCTAGTTACAATTCTTGAGC	GCTCAAGAATTGTAACTAGCCAGAGCATATCAAAGTGG
	c.283A>C	GCTCCGGCTAGTTACACTTCTTGAGCACAGACTATCCC	GGGATAGTCTGTGCTCAAGAAGTGTAACTAGCCGGAGC
	c.283A>G	GCTCCGGCTAGTTACAGTTCTTGAGCACAGACTATCCC	GGGATAGTCTGTGCTCAAGAACTGTAACTAGCCGGAGC
	c.350delT	TCAAATACTGCCTGGGAAGATTTTTTTGGTGTACTGCAGTAGTAGCA	TGCTACTACTGCAGTACACCAAAAAAATCTTCCCAGGCAGTATTTGA
	c.353T>C	GCCTGGGAAGATTTTTTTTGGCGTACTGCAGTAGTAGCACCC	GGGTGCTACTACTGCAGTACGCCAAAAAAAATCTTCCCAGGC
	c.359G>A	GGGAAGATTTTTTTTGGTGTACTACAGTAGTAGCACCCACTTAC	GTAAGTGGGTGCTACTACTGTAGTACACCAAAAAAAATCTTCCC
	c.398T>C	CCCACTTACTTCTTACACTTTACTCACCAGAAAGTATGTTTTGTAGTCACCC	GGGTGACTACAAAACATACTTTCTGGTGAGTAAAGTGTAAGAAGTAAGTGGG
	c.490G>C	CCGCGTCAATGGAGGACAAAGTGTGAACTTACAGG	CCTGTAAGTTCACACTTTGTCCTCCATTGACGCGG
	c.509A>C	GGAGGAGAAAGTGTGAACTTACAGGCGTCTACTCTGAGG	CCTCAGAGTAGACGCCTGTAAGTTCACACTTTCTCCTCC
	c.562C>T	GCTTGTAAATGACTATTGCCTGGTTCTTTTTGCAACGACAC	GTGTCGTTGCAAAAAGAACCAGGCAATAGTCATTTACAAGC
	c.563G>A	GCTTGTAAATGACTATCACCTGGTTCTTTTTGCAACGACAC	GTGTCGTTGCAAAAAGAACCAGGTGATAGTCATTTACAAGC
	c.581C>T	CGCCTGGTTCTTTTTGCAATGACACAAACTATAATGCAG	CTGCATTATAGTTTGTGTCATTGCAAAAAGAACCAGGCG
	c.595A>C	GCAACGACACAAACTATACTGCAGAAAGCCTCGAGCTCA	TGAGCTCGAGGCTTTCTGCAGTATAGTTTGTGTCGTTGC
	c.620A>G	GAAAGCCTCGAGCTCATCAGGAGAACCTTCTCATGCCTCTCG	CGAGAGGCATGAGAAGGTTCTCCTGATGAGCTCGAGGCTTTC
	c.643C>T	AACCTTCTCATGCCTCTCGATGACTGTGTGATGTGGACATAG	CTATGTCCACATCACACAGTCATCGAGAGGCATGAGAAGGTT
	c.651_652delTG	GCCTCTCGACGACTGTGATGTGGACATAG	CTATGTCCACATCACAGTCGTCGAGAGGC
	c.659A>T	GCCTCTCGACGACTGTGTGATGTGGTCATAGACTACAGACC	GGTCTGTAGTCTATGACCACATCACACAGTCGTCGAGAGGC
	c.693G>T	CAGTTGCTGACATGCCTTACAG	CTGTAAGGCATGTCAGCAACTG
	c.714G>C	GGCAGCAACTGGTGAAGCACAGCATGTTTTTCTCC	GGAGAAAAACATGCTGTGCTTCACCAGTTGCTGCC
	c.742C>G	GGATGTTTTTCTCCAAACAAGATGATTCTGAAAGCAGCAACC	GGTTGCTGCTTTCAGAATCATCTTGTTTGGAGAAAAACATCC
	c.772C>T	ATTTTCATTAGTTTCATGTTGTTTAAAAAGTA	TACTTTTTAAACAACATGAAACTAATGAAAAT
	c.808T>G	CTCCAATAATAACAAAATGTTTTTTTAAACTG	CAGTTTAAAAAAACATTTTGTTATTATTGGAG
SDM siRNA resistance		GTTCTTTTCGCTACTACTCAGACAATTATGCAG	CTGCATAATTGTCTGAGTAGTAGCGAAAAGAAC
SDM siRNA res.[b]	c.562C>T	GCTTGTAAATGACTATTGCCTGGTTCTTTTCGCTACTACTC	GAGTAGTAGCGAAAAGAACCAGGCAATAGTCATTTACAAGC
	c.563G>A	GCTTGTAAATGACTATCACCTGGTTCTTTTCGCTACTACTC	GAGTAGTAGCGAAAAGAACCAGGTGATAGTCATTTACAAGC
	c.581C>T	CGCCTGGTTCTTTTCGCTATGACTCAGACAATTATGCAG	CTGCATAATTGTCTGAGTCATAGCGAAAAGAACCAGGCG
	c.595A>C.	GCTACTACTCAGACAATTCTGCAGAAAGCCTCGAGCTCA	TGAGCTCGAGGCTTTCTGCAGAATTGTCTGAGTAGTAGC
mChrerry-XRCC2 [c]		CACCATCGTGGAACAGTACG	ACCTCTACAAATGTGGTATGGCTG
mChrerry-XRCC2 siRNA res.[c]		CACCATCGTGGAACAGTACG	GCCAAAAGACGGCAATATGGTG
mCherry-XRCC2-FLAG		CTCAGATCTGCCGCCGCGATCGCCATG	ATAGAATTCTTACTTGTCATCGTCGTCCTTGTAGTCACAAAATTCAACCCCAC
RT-qPCR	endo XRCC2[d]	CGTCAATGGAGGAGAAAGTG	TGCATTATAGTTTGTGTCGTTGC
	cDNA XRCC2[d]	CGCTACTACTCAGACAATTATGC	ATCCTGTGCTTCACCAGTTG

[a] Site-directed mutagenesis (SDM). [b] Site-directed mutagenesis specific for the siRNA resistant mCherry-XRCC2 construct. [c] Primers for the amplification of the mCherry-*XRCC2* sequence in the mCherry-*XRCC2* construct, used during site-directed mutagenesis and for verification of the presence of the desired variants and the lack of additional mutations with the help of Sanger sequencing. [d] RT-qPCR primers that either detect mRNA from the endogenous XRCC2 (endo) or siRNA-resistant XRCC2 cDNA (cDNA).

Table S2. siRNAs

Target	Sequence
XRCC2	5'-UUAUAGUUUGUGUCGUUGCAA-3'
Control	Universal Negative Control #1 (Sigma, SIC001)

Table S3. In silico predictions on splicing for all *XRCC2* genetic variants

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.46 G>T [a]	p.Ala16Ser	(5)	Exon 2 –c.40N	AS: 78.99 = 78.99	AS: 7.24 = 7.24	AS: 0.41 => NI (-100%)	AS: 6.42 => 6.01 (-6.4%)	AS: 79.84 = 79.84	-
			Exon 2 – c.59	AS: 71.70 => 77.3 (+7.8%)	AS: 5.33 => 7.11 (+33.4)	AS: NI => 0.47 (+100%)		AS: 79.82 = 79.82	+
c.49C>T	p.Arg17*	(5)	Exon 2 –c.40N	AS: 87.99 = 78.99	AS: 7.24 = 7.24	AS: 0.41 => 0.56 (+36.2%)	AS: 6.42 => 6.04 (-5.9%)	AS: 79.84 = 79.84	-
			Exon 2 – c.59	AS: 71.70 +> 74.98 (+4.6%)	AS: 5.33 => 5.82 (+9.4%)			AS: 79.82 => 80.53 (+0.9%)	-
c.140A>G	p.His47Arg	(15)	Exon 3 – c.122N	AS: 82.42 = 82.42	AS: 7.33 = 7.33		4.1 = 4.1	86.84 = 86.84	-
			Exon 3 – c.149	AS: 77.54 => 77.80 (+0.3%)	AS: 2.5 => 2.31 (-7.7%)			84.56 => 84.68 (+0.1%)	-
			Exon 3 – c.152		AS: NI => 0.17 (+100%)			76.13 => 76.32 (+0.2%)	-
c.181C>A	p.Leu61Ile	(5)	Exon 3 – c.122N	AS: 82.42 = 82.42	AS: 7.33 = 7.33		AS: 4.1 => 4.21 (+2.7%)	AS: 86.84 = 86.84	-
			Exon 3 – c.188	AS: 79.67 => 76.18 (-4.4%)	AS: 5.95 => 3.54 (-40.5%)			AS: 83.03 => 80.45 (-3.1%)	-
c.223G>C	p.Glu75Gln	(15)	Exon 3 – c.221	DS: NI => 75.43 (+100%)					-
			Exon 3 – c.225					DS: NI => 66.58 (+100%)	-
			Exon 3 – c.227					AS: 71.3 => 71.93 (+0.9%)	-
			Exon 3 – c.230					AS: 69.94 => 67.83 (+3.7%)	-
			Exon 3 – c.233					65.46 => 67.83 (+3.6%)	-
c.247dup [b]	p.Thr83fs	-	Exon 3 – c.251	AS: 83.89 => 83.23 (-0.8%)	AS: 4.83 => 5.0 (+3.5%)	AS: 0.54 => 0.51 (-5.3%)		AS: 86.47 => 83.08 (-3.9%)	-
c.271C>T	p.Arg91Trp	(5)	Exon 3- c.272	DS: 76.35 => 76.81 (+0.6%)					-
			Exon 3- c.278		AS: 3.43 => 4.13 (+20.2%)			AS: 76.55 => 76.13 (-0.5%)	-
c.283A>C [c]	p.Ile95Leu	(5)	Exon 3 – c.292	AS: NI => 72.72 (+100%)				AS: 71.47 => 73.5 (+2.8%)	-
c.283A>G	p.Ile95Val	(5,15)	Exon 3 – c.284					AS: NI => 81.75 (+100%)	-
			Exon 3 – c.292					AS: 71.47 => 71.59 (+0.2%)	-

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.350delT	p.Leu117fs	(15)	Exon 3-c.363	AS: 78.97 = 78.97	AS: 6.52 => 5.74 (-11.9%)			AS: 81.2 = 81.2	-
			Exon 3 – c.401	AS: 75.64 = 75.64	AS: 6.07 = 6.07		AS: 3.25 => 3.44 (+5.6%)	AS: 81.22 = 81.22	-
c.353T>C [d]	p.Val118Ala	(15)	Exon 3-c.351					DS: 68.97 => NI (-100%)	-
			Exon 3-c.363	AS: 78.97 => 75.68 (-4.2%)	AS: 6.52 => 4.72 (-27.5%)			AS: 81.2 => 80.5 (-0.9%)	-
			Exon 3-c.401	AS: 75.64 = 75.64	AS: 6.07 =6.07		AS: 3.25 => 2.85 (-12.5%)		-
c.359G>A	p.Cys120Tyr	(15)	Exon 3 – c.363	AS: 78.97 = 78.97	AS: 6.52 => 5.8 (-11.0%)	AS: NI => 0.49 (+100%)		AS: 81.2 => 81.27 (+0.1%)	-
			Exon3 – c.366					AS: 72.3 => 72.34 (+0.1%)	-
			Exon 3-c.369					AS: 71.81 => 71.69 (-0.2%)	-
			Exon 3-c.401	AS: 75.64 = 75.64	AS: 6.07 = 6.07		AS: 3.25 => 3.42 (+5.0%)	AS: 81.22 = 81.22	-
c.398T>C	p.Leu133Pro	(15)	Exon 3 – c.403	DS: 77.38 = 77.38	DS: 6.93 = 6.93	DS: 0.53 => 0.8 (+50.4%)		DS: 79.66 = 79.66	-
			Exon 3 – c.401	AS: 75.64 => 81.82 (+8.2%)	AS: 6.07 => 6.53 (+7.7%)		AS: 3.25 =>2.79 (-14.4%)	AS: 81.22 => 88.91 (+9.5%)	-
			Exon 3-c.405					AS: 72.35 => 72.77 (+0.6%)	-
c.490G>C	p.Glu164Gln	(15)	Exon 3-c.473	AS: 80.38 = 80.38	AS: 3.79 = 3.79	AS: 0.62 => 0.57 (-7.5%)		AS: 77.04 = 77.04	-
			Exon 3 – c.491					AS: 66.64 => NI (-100%)	-
			Exon 3 – c.495					AS: NI => 65.73 (+100%)	-
			Exon 3 – c.508	AS: 80.85 = 80.85	AS: 1.61 => 3.34 (+107.7%)			AS: 85.96 = 85.96	-
c.509A>C	p.Glu170Ala	(15)	Exon 3 – c.508	AS: 80.85 = 80.85	AS: 1.61 => 1.58 (-1.6%)			AS: 85.96 => 85.89 (-0.1%)	-
			Exon 3 – c.511					AS: 69.22 => NI (-100%)	-
			Exon 3 – c.522	AS: NI => 71.6 (+100%)				AS: 80.25 = 80.25	-
c.562C>T	p.Arg188Cys	(15)	-						-
c.563G>A [e]	p.Arg188His	(5,15)	-						-
c.581C>T	p.Thr194Met	(15)	Exon 3 – c.601		AS: 1.03 => 2.34 (+127.3%)			AS: 80.46 = 80.46	-
c.595A>C	p.Met199Ile	(15)	Exon 3 – c.601	AS: NI => 72.3 (+100%)	AS: 1.03 => 2.45 (+137.8%)			AS: 80.46 => 83.87 (+4.2%)	+
			Exon 3 – c.605					AS: 68.36 => 70.84 (+3.6%)	-

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.620A>G [f]	p.Glu207Gly	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.35 (+0.7%)	DS: 80.47 = 80.47	-
			Exon 3 – c.620					AS: 80.26 => 83.39 (+3.9%)	-
			Exon 3 – c.623					AS: 73.31 => 72.4 (-1.2%)	-
c.643C>T	p.Arg215*	(6)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 3.79 (-12.3%)	DS: 80.47 = 80.47	-
c.651_652delTG	p.Cys217*	(5)	Exon 3 – c.647					DS: 71.13 => 68.02 (-4.4%)	-
			Exon 3 – c.649					DS: 66.88 => NI (-100%)	-
			Exon 3 – c.682		DS: 4.7 = 4.7 AS: 3.89 => 2.75 (-29.1%)		DS: 4.33 => 4.23 (-2.2%)	DS: 80.47 = 80.47	-
			Exon 3 – c.665		AS: 3.89 => 2.75 (-29.1%)			AS: 71.89 = 71.89	-
c.659A>T	p. Asp220Val	(15)	Exon 3 – c.657					DS: NI => 73.81 (+100%)	-
			Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 3.68 (-14.9%)	DS: 80.47 = 80.47	-
			Exon 3 – c.665		AS: 3.89 => 4.93 (+26.7%)			AS: 71.89 => 74.65 (+3.8%)	-
c.693G>T	p.Trp231Cys	(5)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 2.94	DS: 80.47 = 80.47	-
			Exon 3 – c.687	DS: 81.88 => 89.67 (+9.5%)					-
			Exon 3 – c.691	DS: NI => 71.72 (+100%)				DS: NI => 81.93 (+100%)	+
			Exon 3 – c.697					AS: 74.15 => 74.11 (-0.0%)	-
c.714G>C	p.Arg238Ser	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.21 (-2.6%)	DS: 80.47 = 80.47	-
			Exon 3 – c.714					AS: 79.2 => 75.04 (-5.3%)	-
			Exon 3 – c.734	AS: 70.48 = 70.48	AS: 3.1 => 3.88 (+25.3%)			AS: 74.9 = 74.9	-
c.742C>G [g]	p.Gln248Glu	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.54 (+4.9%)	DS: 80.47 = 80.47	-
			Exon 3 –c.747					AS: 76.11 => 73.38 (-3.6%)	-
			Exon 3 – c.750					AS: 84.19 => 81.65 (-3%)	-
c.772C>T	p.Arg258Cys	(15)	Exon 3 – c.786	AS: NI => 70.24 (+100%)	AS: 2.36 => 2.79 (+18.4%)			AS: 73.9 = 73.9	+

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.808T>G [h]	p.Phe270Val	(5,15)	Exon 3 – c.807					DS: NI => 65.06 (+100%)	-
			Exon 3 – c.821	AS: 84.42 => 78.83 (-6.6%)				AS: 78.15 = 78.15	-

All non-synonymous coding *XRCC2* variants from Park et al. (5) and Hilbers et al. (15) were selected for in silico analysis by SpliceSiteFinder, MaxEntScan, NNSplice, GeneSplicer and Human Splice Finder using the integrated Alamut software package. In addition, a variant described by Shamseldin et al. (2012), detected in a Fanconi Anemia case and a truncating variant, p.Thr83fs reported by Douglas Easton (personal communication) were analyzed. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. [a] rs4987090, [b] reported by Douglas Easton (personal communication), [c] rs140214637, [d] rs185815454, [e] rs3218536, [f] rs61762969, [g] rs190900560 [h] rs145085742. [i] a variation of more than 10 % in at least two algorithms was considered as having an effect on splicing (Théry et al., 2011). The scores indicate the values for splice donor (SD) or splice acceptor (SA) sites, respectively. Changes relative to wild-type sequences are indicated in %. Th. = threshold, NI = not identified.

Chapter 5

Clustering of known low and moderate risk alleles rather than a novel recessive high-risk gene in non-*BRCA1/2* sib trios affected with breast cancer

Florentine S. Hilbers, Peter J. van 't Hof, Caro M. Meijers, Hailiang Mei, Kyriaki Michailidou, Joe Dennis, Frans B. L. Hogervorst, Petra M. Nederlof, Christi J. van Asperen, Peter Devilee

Int J Cancer. 2020 May 7

Abstract

Breast cancer risk is approximately twice as high in first-degree relatives of female breast cancer cases than in women in the general population. Less than half of this risk can be attributed to the currently known genetic risk factors. Recessive risk alleles represent a relatively underexplored explanation for the remainder of familial risk. To address this, we selected 19 non-*BRCA1/2* breast cancer families in which at least three siblings were affected, while no first-degree relatives of the previous or following generation had breast cancer. Germline DNA from one of the siblings was subjected to exome sequencing, while all affected siblings were genotyped using SNP arrays in order to assess haplotype sharing and to calculate a polygenic risk score (PRS) based on 160 low risk variants. We found no convincing candidate recessive alleles among exome sequencing variants in genomic regions for which all three siblings shared two haplotypes. However, we found two families in which all affected siblings carried the *CHEK2**1100delC. In addition, the average normalized PRS of the "recessive" family probands (0.81) was significantly higher than that in both general population cases (0.35, p=0.026) and controls (p=0.0004). These findings suggest that the familial aggregation is, at least in part, explained by a polygenic effect of common low risk variants and rarer intermediate risk variants, while we did not find evidence of a role for novel recessive risk alleles.

Introduction

Breast cancer is the most common cancer in females in the Western world and has a complex etiology in which both genetic and environmental factors affect disease risk. Having a family member affected with the disease is one of the most important risk factors.[1] Pathogenic variants in the two most well-known high risk breast cancer genes, *BRCA1* and *BRCA2*, explain approximately 17% of the familial relative risk.[2] In addition, a number of less frequently mutated high risk genes (e.g. *TP53*) and a number of genes in which pathogenic variants are associated with a more moderately increased risk (e.g. *CHEK2*) together explain another 5%. Moreover, approximately 160 common polymorphisms have been associated with small increases in risk, which jointly explain about 18% of the excess familial risk.[3]

Since the discovery of *BRCA1* and *BRCA2*, several segregation studies have concluded that a polygenic model, or a model with a recessive allele would best explain the remaining familial risk.[4–7] Genetic searches for new loci, while successful, have focused on detecting rare dominant high-risk alleles (by candidate gene re-sequencing) or common low risk variants. Systematic searches for recessive alleles have not been conducted, despite evidence suggesting that such alleles could play a role in the genetic etiology of breast cancer. For example, a large meta-analysis on familial breast cancer risk has shown that having a sister affected with breast cancer is associated with a stronger increase in risk than having a mother with breast cancer.[8] In addition, an increased breast cancer risk has been reported in the offspring of consanguineous parents.[9] Studies assessing regions of homozygosity in outbred populations have not shown more or larger regions of homozygosity in breast cancer cases, but some have suggested an increased frequency of homozygosity in specific genomic regions.[10,11]

We performed a small-scale search for recessive breast cancer risk alleles in families with at least three affected siblings and no other first or second-degree relatives with early onset breast cancer. The regions in which all affected siblings shared two haplotypes, as determined by low-density SNP arrays, were identified and used to filter the exome sequence

data that was generated for one of the siblings. This approach significantly reduces the number of potentially interesting variants, allowing for less stringent filters on allele frequency and hence fewer assumptions about the characteristics of a novel breast cancer risk-associated variant. In addition, we calculated a polygenic risk score based on 160 known breast cancer risk-associated polymorphisms and assessed the contribution of exonic variants in known breast cancer susceptibility genes that were predicted to be damaging by in silico prediction algorithms.

Methods

Selection of Families

Families were ascertained through the clinical genetics centers of two Dutch hospitals, the Leiden University Medical Center (LUMC) and the Netherlands Cancer institute Antoni van Leeuwenhoek hospital (NKI-AvL) and from a previously described set of breast cancer families collected throughout the Netherlands.[12] We enriched for families with a presumed recessive mode of inheritance by selecting families in which at least three siblings were affected with breast cancer at any age. Sib-ships that had first-degree relatives with breast cancer in the previous or following generation were excluded, as were families with second degree relatives with breast cancer diagnosed before age 50. DNA from blood lymphocytes had to be available for at least two affected siblings. Availability of DNA samples from parents or other family members was not a selection criterion. In every family at least one affected individual had been extensively tested according to local testing standards for pathogenic variants in *BRCA1* and *BRCA2*, and all families with a pathogenic variant or variant of uncertain significance in *BRCA1* or *BRCA2* were excluded.

Haplotype analysis

We genotyped all available DNA samples from the affected siblings using the HumanLinkage V Panel from Illumina. Sample preparation was done according to the manufacturer's protocol (Rev. B October 2010). Samples were hybridized to GoldenGate Universal-32 BeadChip (Illumina) and chips were scanned using a Bead Array Reader (Illumina). The GenomeStudio software (version 2011.1, Illumina) was used to call genotypes. We used Merlin (v1.12) to calculate, for each sib pair and marker position, the probability that at this position the sib pair shared zero, one or two alleles identical by decent (IBD).[13] On average a sib pair is expected to share two haplotypes in 25% of their genomes. To decrease the chance of false-negative regions, we set a probability cut-off such that for all sib pairs at least 25% of the markers were selected as sharing two alleles IBD (cut-off: $p > 0.05$). We then selected all positions in which all siblings shared two alleles IBD or, for the analysis allowing for one phenocopy, all positions in which all but one sib shared two alleles IBD. These positions were converted into a BED file describing the regions IBD for both haplotypes. Each of these regions started one base pair after the last upstream position for which the affected siblings did not share two alleles IBD and ran until one base pair before the first downstream position for which they did not share two alleles IBD.

Exome sequencing and analysis

From each family one affected individual was selected for exome sequencing of germline DNA. In most instances this was the individual with the youngest age of diagnosis; however

in two families another individual was selected due to limited availability of DNA. Samples were prepared using Illumina's Paired-End Library Preparation Kit, after which the coding regions of the genome were captured using SeqCap EZ Exome v3.0 (Nimblegen). Sequencing was done on a HiSeq 2000 (Illumina), generating 2x100 base pair reads. We used GATK for indel realignment, base recalibration and finally variant calling using Haplotypecaller.[14] These analyses were done according to the GATK best practices guidelines for DNA sequencing analysis. A detailed description of the settings and version numbers of the used software is given in the supplementary methods.

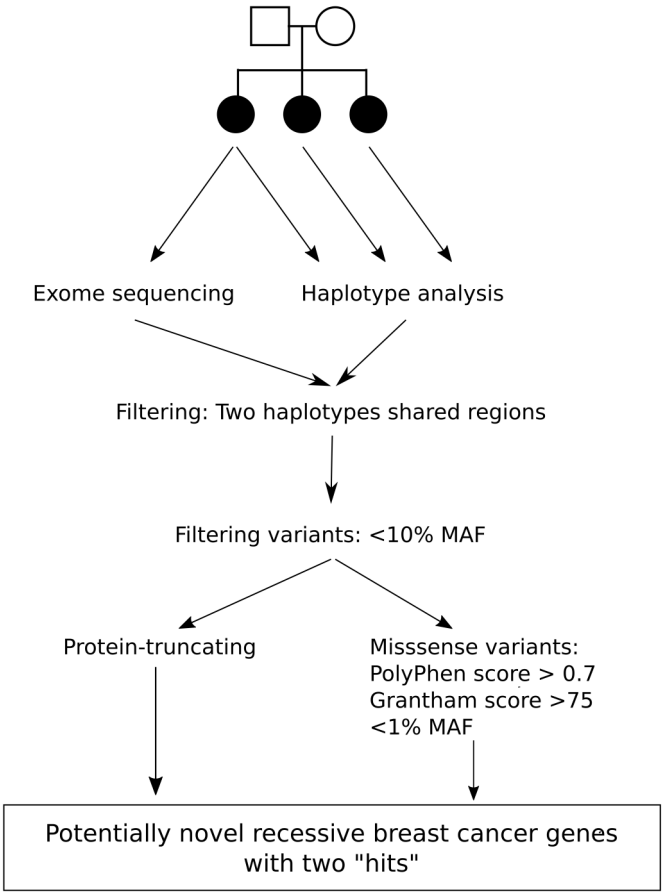


Figure 1. Strategy for the identification of recessively predisposing genetic variants
This overview presents our strategy for exome variant filtering in order to detect potential new breast cancer risk alleles with a recessive mode of inheritance.

Variant filtering and validation

Figure 1 outlines our strategy for identifying recessively predisposing genetic variants in the affected sib ships. We first selected, for each individual, variants in regions in which they shared two haplotypes IBD with their siblings, using the family-specific BED files. We then annotated

the variants using Seattleseq (138, v9.03).[15] Next, we selected all stop-gained, frameshift and canonical splice site variants. These predicted protein-truncating variants (PTV) could be either heterozygous or homozygous. We removed variants with an allele frequency >10% in either the exome variant server, Hapmap, 1000 genomes, ExAC or Genome of the Netherlands (GoNL) data.[16–20] In addition, we removed all variants with an allele frequency of >30% in our own dataset, since these are likely to be experiment-specific artefacts. All remaining variants were manually inspected in the Integrative Genomics Viewer (IGV; v2.3.34) in order to remove any clear misalignments or other calling errors.[21] In the genes in which a heterozygous potential PTV was found, we searched for a "second hit", defined as either another potential PTV or a missense variant, satisfying the same frequency cut-off. When two (or more) "hits" in a gene were identified, these variants were validated using Sanger sequencing. Primer sequences are available upon request.

We also considered a scenario in which two missense changes in a gene on two haplotypes could cause a recessive inheritance (i.e., either homozygous or compound heterozygous). For this, we selected all missense changes in the regions specified by the BED files, with allele frequencies <1% and in silico annotations suggestive of deleteriousness (PolyPhen score > 0.7; Grantham score >75).

Variants in known and suspected breast cancer genes

We examined a set of 35 known and suspected breast cancer susceptibility genes (derived from commercially available multigene panels, Supplementary Table 1) for genetic variants regardless of haplotype sharing. The genes were assigned into 4 categories, based on the level of evidence for being associated with breast cancer risk (strong to unlikely); a separate category consisted of "syndromic" genes, in which variants have been associated with a range of cancers typical of certain familial cancer syndromes (*TP53*, *CDH1*, *PTEN*). PTVs in level 1/2 genes were filtered on allele frequency in the general population (exome variant server, Hapmap, 1000 genomes or GoNL) with a cut-off of 0.1% for the high-risk genes (*BRCA1*, *BRCA2*, *PALB2*, *TP53*, *PTEN* and *CDH1*) and 2% for the moderate risk genes (*ATM*, *CHEK2*), allowing for the observation that some PTVs in moderate risk genes (such as the c.1100delC in *CHEK2*) occur at >0.5% allele frequencies in some populations. All missense variants in the 35 genes were selected if their allele frequency in the general population was <2% and they had either a CADD[22] score > 20 or were found in one of the level 1/2 genes. All selected variants were inspected manually in the IGV to remove misalignments. Variants that were both rare and not likely to result from a misalignment were then validated using Sanger sequencing.

Validation of potential recessive risk alleles

In order to further assess the association of selected variants with recessive breast cancer, we selected a set of 111 women diagnosed with breast cancer 35 or younger, through the clinical genetics center of the LUMC.

Polygenic risk score analysis

All affected sibs for whom DNA was available were genotyped using one of two SNP arrays partly designed to study SNPs associated with breast cancer risk: the iCOGs array and the OncoArray. In order to calculate polygenic risk scores, we selected all independent SNPs shown to be significantly ($p<5x10^{-8}$) associated with overall breast cancer by the Breast Cancer Association Consortium (BCAC), the largest case-control study to date.[23] The selected

SNPs and respective ORs are shown in Supplementary Table 2. A small number of known low risk variants were not included on the arrays. These variants were imputed with the help of IMPUTE2 based on the genome of the Netherlands (GoNL release 5.3) and 1000 genomes (Phase 3) data (see supplementary methods).[18,19,24] Polygenic risk scores were calculated using:

$$PRS_j = \sum_{i=1}^{160} n_{ij} \ln (OR_i)$$

Where n_{ij} is the number of risk alleles (0,1 or 2) SNP *i* carried by individual *j* and OR_i is the per-allele odds ratio associated with SNP *i* (derived from [23], see Supplementary Table 2). We compared the PRS of the family probands (the same individuals subjected to exome sequencing) with 357 sporadic cases and 327 age-matched controls from the ORIGO study.[25] These individuals were genotyped using the iCOGS array and imputed in the same way as the familial cases. The PRS was normalized based on the mean and standard deviation of the ORIGO controls so that one unit in PRS corresponded to one standard deviation. The odds ratio per unit standard deviation of the PRS was obtained via univariate logistic regression within the ORIGO population. The null hypothesis of there not being a true difference in mean PRS between the "recessive" family probands, population cases and population controls was tested using a Welch two sample t-test. All analyses were performed using R version 3.4.1.

All individuals provided informed consent and approval of the medical ethical committee at the LUMC was obtained.

Results

Selected families and haplotype analysis

Nineteen families were selected for analysis (Supplementary Figure 1). Samples were available from two affected siblings for 3 families, three affected siblings for 14 families and four affected siblings for 2 families. The average age at diagnosis of first primary breast cancer was 49.9. One family included a male breast cancer patient diagnosed at age 65. The "two haplotypes shared IBD" regions for each family covered on average 31.6%, 10.1% and 2.9% of the genome for families with two, three or four DNA samples available respectively. This is slightly higher than predicted proportions (25%, 6.25% and 1.6% respectively), but this was expected given our conservative IBD probability cut-off (see methods).

Exome sequencing

Exome sequencing of one affected individual per family achieved 51x average on target coverage and detected on average 28,724 variants per individual. After filtering these variants based on the family-specific haplotype sharing regions, an average of 10,775 (37.5%), 3,222 (11.2%) and 734 (2.6%) variants remained in families with two, three or four individuals genotyped respectively. We first focused on variants that were predicted to result in a truncated protein. When a heterozygous protein-truncating variant (PTV) was found, we assessed the gene for a second hit which could also be a missense variant (Table 1).

We originally set the PTV allele frequency cut-off relatively high (<10%) to allow for the possibility of a single variant that was homozygous in multiple families. No such variants were detected in our dataset, but we did find six genes with two or more heterozygous positions in six different families. For compound heterozygotes, we assumed that the allele

frequency of a potentially causal variant was lower (<2%), rendering the variants in *TLR5*, *TRPM1*, *UNC93A*, *PLXNB3* and *CCHCR1* unlikely candidates. In the remaining gene, *PDIA2*, we identified a PTV p.R148* and a missense variant p.R473Q, shared IBD in one family. *PDIA2* encodes an oxidoreductase involved in protein folding and specifically expressed in the pancreas.[26–28] In addition, it binds estrogen (specifically 17β-estradiol) and might buffer the local estrogen levels in the pancreas.[29] In order to further examine the possibility that variants in *PDIA2* are associated with breast cancer, we genotyped a set of 111 patients diagnosed with breast cancer before the age of 35 for the two variants detected in family RF1. The PTV p.R148* was not observed, while the missense variant p.R473Q was detected twice (0.9%). The allele frequency of 1.3% in the Genome of the Netherlands, also suggests that this variant is not associated with breast cancer.[19]

A similar filter for missense variants revealed two rare homozygous missense variants, *SERINC2* p.R126W in family RF4 and *ZNF717* p.H63L in family RF7 (Supplementary table 3). SERINC2 regulates lipid biosynthesis and incorporates serine into membrane lipids, while the function of ZNF71 is unknown. The CADD scores for both variants were <20. Based on this, neither variant was considered as a serious candidate for follow-up studies.

Table 1: rare protein-truncating and missense variants found in the regions where the sibships share two haplotypes

Family	Gene	Variant (coding DNA)	Variant (protein)	Rs-number	Co-segregation[a]	Frequency in GoNL[b]
RF1	PDIA2	c.442C>T	p.R148*	rs370453080	2/3	0%
		c.1418G>A	p.R473Q	rs116969376	3/3	1.3%
RF4	TLR5	c.1174C>T	p.R392*	rs5744168	3/3	6.5%
		c.541C>A	p.Q181K	rs45528236	3/3	6.5%
RF6	TRPM1	c.4240G>T	p.E1414*	rs3784589	2/3	4.9%
		c.1930G>A	p.V644M	rs17815774	3/3	4.7%
RF13	UNC93A	c.625+1G>C	p.?	rs113906647	1/3	3.3%
		c.1159T>C	p.Y387H	rs663227	1/3	0.7%
RF14	PLXNB3	c.1629+2C>T	p.?	-	1/3	0%
		c.4787T>A	p.V1596E	rs146832392	3/3	6.0%
RF17	CCHCR1	c.121G>T	p.E41*	rs72856718	3/3	9.6%
		c.2147G>A	p.R716Q	rs130072	3/3	9.6%
		c.803T>A	p.L268Q	rs11540822	3/3	9.6%

[a]Indicates the number of siblings carrying the allele out of the total number do siblings from this family tested.
[b] Frequency in Genome of the Netherlands: genome sequences of 998 independent Dutch individuals.[22] Accession numbers for the transcripts and protein sequences used to describe the variants: PDIA2: NM_006849.2, NP_006840.2; TLR5: NM_003268.5, NP_003259.2; TRPM: NM_001252020.1, NP_001238949.1; UNC93A: NM_018974.3, NP_061847.2; PLXNB3: NM_005393.2, NP_005384.2; CCHCR1: NM_001105564.1, NP_001099034.1.

Analyses allowing for one phenocopy

Since breast cancer is a common disease, there is a high probability that a case in a family is not genetic (i.e., a phenocopy). Therefore, we assessed the regions of the genome where only

two out of three (or three out of four) affected sisters share two haplotypes. PTVs obtained in this way were then filtered as in the previous analysis (see Table 2). Again, most variants were relatively common, but did not occur in multiple families. The only gene in which variants are rare enough to be a possible candidate was *SLC26A10*, with variants c.1206G>A (p.W402*) and c.1247T>G (p.L416R) found in family RF2. Both variants were shared by two of the three affected sisters. However, in GoNL, both variants were present in the same 7 individuals and predicted to be on the same haplotype, excluding the possibility of compound heterozygosity. SCL26A10 has no known function and has been suggested to be an imprinted, maternally expressed, pseudogene.[30,31]

Table 2: Rare protein-truncating and missense variants found in the regions where the sibships share two haplotypes, allowing for one phenocopy

Family	Gene	Variant (coding DNA)	Variant (protein)	Rs-number	Co-segregation[a]	Frequency in GoNL[b]
RF2	*ZAN*	c.1249+1G>A	p.?	rs117406702	3/3	3.8%
		c.8132C>T	p.P2711L	rs201771583	3/3	0
	SLC26A10	c.1206G>A	p.W402*	rs113207856	2/3	0.7%
		c.1247T>G	p.L416R	rs111924104	2/3	0.7%
RF6	*CCHCR1*	c.121G>T	p.E41*	rs72856718	1/3	9.6%
		c.803T>C	p.L232Q	rs11540822	1/3	9.6%
RF8	*PLA2G4C*	c.893delC	p.P298fs	rs11564598	3/3	2.9%
		c.452C>T	p.P151L	rs11564538	1/3	5.0%
RF14	*PKHD1L1*	c.7246+1G>C	p.?	rs17368310	3/3	4.5%
		c.10310A>G	p.D3437G	rs118053060	2/3	2.5%

[a]Indicates the number of siblings carrying the allele out of the total number do siblings from this family tested.
[b] Frequency in Genome of the Netherlands: genome sequences of 998 independent Dutch individuals.[22]
Accession numbers for the transcripts and protein sequences used to describe the variants: *ZAN*: NM_003386.2, NP_003377.2; *SLC26A10*: NM_133489.2, NP_597996.2; *CCHCR1*: NM_001105564.1, NP_001099034.1; *PLA2G4C*, NM_003706.2, NP_003697.2; *PKHD1L1*: NM_177531.4, NP_803875.2.

Known and suspected moderate and high-risk genes

We next examined 35 genes in which PTVs have been demonstrated or suspected to be associated with breast cancer risk (Table 3 and Supplementary Table 1). We found two rare missense variants in known high-risk genes, one in *PALB2* and one in *BRCA2*. ClinVar lists the variant in *PALB2* as benign, the one in *BRCA2* as variant of uncertain significance (VUS). Family RF17 was included in our study as being non-*BRCA1/2* because the sister not carrying the missense variant was the one tested in the clinical setting. No studies on the functional effects of this variant have been published to date, but the CADD score of 35 indicates that it might affect protein function. Therefore, it is possible that this family harbors a pathogenic *BRCA2* variant.

The c.1100delC pathogenic variant in *CHEK2*, associated with an odds ratio (OR) of approximately 2.3,[32] was found in all affected individuals of families RF4 and RF8, with all individuals being heterozygous. We found several missense variants in the (suspected) moderate-risk genes *ATM*, *CHEK2* and *RAD51C*. The effect of missense changes in *ATM* and *CHEK2* on breast cancer risk is, besides a few specific examples, largely uncertain.[33–35] None of the variants listed in Table 3 belong to any of these exceptions, but some do have CADD scores

>20 suggestive of pathogenicity. Two other variants have previously been associated with breast cancer risk, although association data have been conflicting. *ATM* c.146C>G (p.S49C) was detected in families RF6 and RF20; its associated breast cancer risk is unlikely to be larger than 1.5.[34,36,37] Likewise, conflicting results were obtained for the breast and/or ovarian cancer risk of *RAD51C* c.790G>A (p.G264S) in families RF8 and RF19.[38] The contribution of these variants to breast cancer susceptibility, if any, is therefore uncertain.

Table 3: Rare genetic variant in known and suspected breast cancer genes

Gene	Family	Variant (coding DNA)	Variant (protein)	Rs-number	Co-segregation[a]	Frequency[b]
ATM	RF6	c.146C>G	p.S49C	rs1800054	2/3	1.7%
ATM	RF7	c.2531G>A	p.G844E	rs587781808	2/3	0.002%
ATM	RF10	c.2991A>G	p.(=)	rs1203368496	3/3	0
ATM	RF18	c.584C>T	p.T195I	rs1196611507	2/3	-
ATM	RF20	c.146C>G	p.S49C	rs1800054	3/3	1.7%
BRCA2	RF17	c.8290G>A	p.A2764T	rs786202189	2/3	-
CDH1	RF21	c.1689C>T	p.(=)	rs587780786	2/2	0.007%
CHEK2	RF4	c.1100delC	p.T367fs	rs555607708	3/3	1%
CHEK2	RF8	c.1100delC	p.T367fs	rs555607708	3/3	1%
CHEK2	RF14	c.556A>C	p.N186H	rs146198085	1/3	0.01%
PALB2	RF20	c.150A>T	p.K50N	-	1/2	-
RAD51C	RF8	c.790G>A	p.G264S	rs147241704	3/3	0.3%
RAD51C	RF19	c.790G>A	p.G264S	rs147241704	1/2	0.3%

[a] Indicates the number of siblings carrying the allele out of the total number do siblings from this family tested. [b] Highest frequency in either ESP, ExAc, gnomAD, or GoNL; - if no entry listed; Accession numbers for the transcripts and protein sequences used to describe the variants: *ATM*: NM_000051.3, NP_000042.3; *BRCA1*: NM_007294.3, NP_009225.1; *BRCA2*: NM_000059.3, NP_000050.2; *CDH1*: NM_004360.3, NP_004351.1; *CHEK2*: NM_007194.3, NP_009125.1; *PALB2*: NM_024675.3, NP_078951.2; *RAD51C*: NM_058216.2, NP_478123.1.

Polygenic risk score analysis

Over 160 independent common SNPs have been found to be convincingly associated with breast cancer and can be combined into a PRS ([23]; Supplementary Table 2). To examine the effect of the PRS on the breast cancer cases in our families, we genotyped or imputed these SNPs for all individuals from whom DNA was available. The PRS was normalized such that the mean and standard deviation of the population controls were 0 and 1 respectively. Figure 2 shows the difference in distribution between our familial cases and a set of population cases and controls, clearly showing a strong skewing towards PRS >0 for the familial cases. The odds ratio per unit standard deviation of the PRS was 1.46. The average PRS of all the affected siblings in the families was 0.63, corresponding to an odds ratio (OR) of 1.27. The average score of the family probands (0.81, OR 1.36) was significantly higher than that in both population cases (0.35, OR 1.14, $p = 0.026$) and controls ($p = 0.0004$).

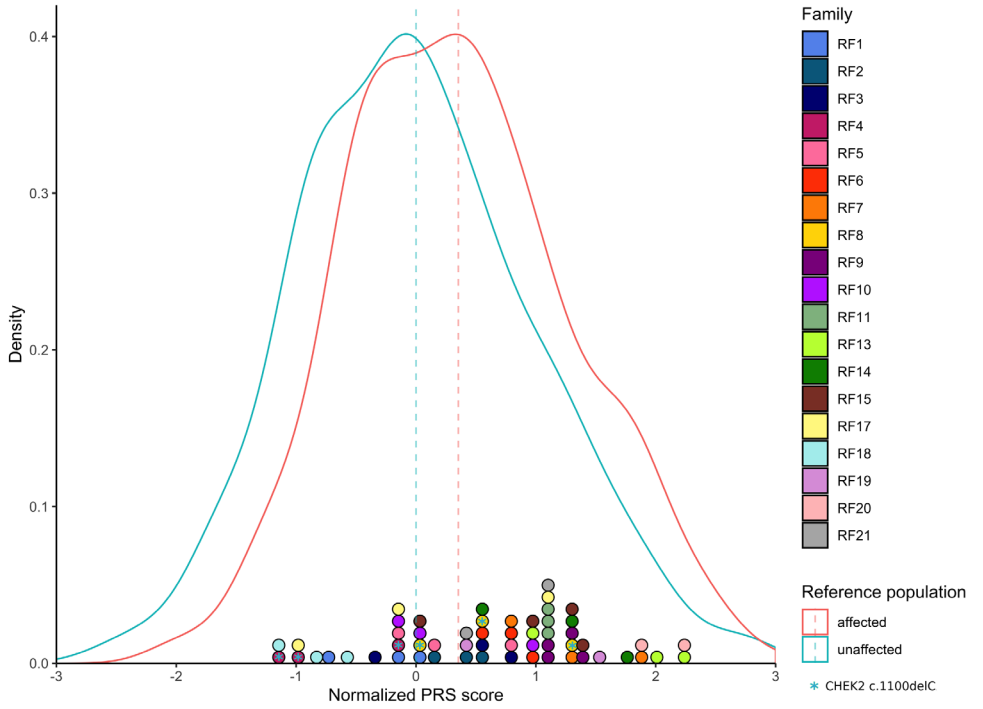


Figure 2. PRS scores for recessive families compared to population cases and controls
The blue and red line represents the density plot of PRS for population controls and cases, respectively. Colored circles at the ordinate each represent one individual from the 19 investigated families, circles with the same color belong to the same family. Circles with a blue star represent carriers of the *CHEK2* c.1100delC variant.The dotted lines represent the mean PRS for the population controls and familial cases.

Discussion

In this study we assessed whether breast cancer in families with at least three affected siblings, can be explained by a susceptibility gene with a recessive mode of inheritance. After a haplotype-guided exome analysis, we identified no homozygous or compound-heterozygous variants that were likely to explain the clustering of breast cancers in the selected families. We did identify two families in which all affected individuals carry the known moderate risk variant *CHEK2**1100delC. Furthermore, we showed that on average, the affected women in these families had significantly higher PRS than both sporadic cases and population controls. Together, these results indicate that our selection criteria enrich for these factors and suggest that, rather than being caused by a single highly penetrant variant, increased breast cancer risk in some of these families may be due to the combined effect of multiple rare and common genetic variants with varying effect-sizes, and perhaps other non-genetic risk factors as well.

Due to a few limitations of our study we cannot completely rule out that some of our families are nonetheless explained by recessive risk alleles. First, some of the variants we

identified (e.g. *PDIA2* p.R148*) are so rare in the general population that they would require very large case-control populations to assess their association with breast cancer. As they grow in size, publicly available reference datasets and databases in which variants in potential disease-associated genes can be reported are becoming very valuable for this purpose. Secondly, a recessive risk allele might be located outside the protein coding regions of the genome and thus not be captured by an exome sequencing approach. Moreover, structural variation, affecting more than a few base pairs, is mostly undetectable with the methods used in this study. Whole genome sequencing would identify these, but their mostly poor genomic annotation will make their filtering for follow-up analyses very hard.

Thirdly, our family selection has led to many sibships that could also be explained by a dominant allele with incomplete penetrance. While our study design had advantages for the variant filtering, there are alternative ways to enrich for recessive alleles, such as population-based sib pairs, or early-onset cases with unaffected parents. Such studies haven't yet been published for breast cancer but would probably also suffer from severe genetic model heterogeneity. Thus, the existence of recessive breast cancer alleles remains possible, although it is remarkable in this regard, that only a handful of the >160 common breast cancer loci derived from population-based genome-wide association studies affect risk in a recessive mode, rather than in a co-dominant way.[25]

Nonetheless, our results are in agreement with previous exome sequencing studies in non-*BRCA1/2* familial breast cancer cases. Although more than twenty such studies have been published, only two new breast cancer genes suggested by these studies were replicated independently: *FANCM* and *RECQL*.[39–42] Most of these studies, however, reported pathogenic variants in known moderate risk genes. Studies employing gene panel sequencing in a large numbers of familial breast cancer cases suggest that approximately 4% carry a pathogenic or likely pathogenic variant in a breast cancer gene other than *BRCA1* or *BRCA2*.[43–45] We found two index cases carrying the *CHEK2**1100delC pathogenic variant (consistent with high frequency of this variant in the Dutch population), and four possibly pathogenic variants in other susceptibility genes. At least for *CHEK2**1100delC it has been shown that the risk associated with this pathogenic variant and the risk associated with a PRS combine multiplicatively.[46] With regard to the common low risk variants, our results are consistent with studies which have found that non-*BRCA1/2* familial breast cancer cases have a higher PRS than both cases from the general population and cases who carry a *BRCA1* or *BRCA2* pathogenic variant.[47–50] Whether the prevalence of rare missense variants in the known breast cancer genes we observed in our families is causally linked to breast cancer, will need very large case-control studies to substantiate further.

The enrichment of moderate and low risk alleles among the cases of at least part of the families in our study adds to a growing body of evidence on the importance of this type of risk alleles in causing familial breast cancer. Multigene panel sequencing has rendered the detection of rare variation in known risk genes standard clinical genetic practice, but the genotyping of the many common low risk alleles is not yet routinely performed in this setting. Nonetheless, the risks associated with the PRS and the likely multiplicative way in which it combines with those of pathogenic variants in moderate risk genes argue for a more comprehensive approach to genetic testing and counseling. This calls for the development of integrative risk prediction models, including the effect of mammographic density, lifestyle and environmental risk factors.

Competing Interests

The authors declare that they have no conflicts of interest

Funding

This work was supported by the Dutch Cancer Society [KWF UL 2009-4388], CR-UK, Genome Canada and NIH/NCI. This study makes use of data generated by the Genome of the Netherlands Project. A full list of the investigators is available from www.nlgenome.nl. Funding for the project was provided by the Netherlands organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobank and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI).

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 Lond Engl 2001;358:1389–99.
2. Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DGR, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med 2015;372:2243–57.
3. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen T-H, Wang Q, Bolla MK, Yang X, Adank MA, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet 2019;104:21–34.
4. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol 2001;21:1–18.
5. Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002;86:76–83.
6. Cui J, Antoniou AC, Dite GS, Southey MC, Venter DJ, Easton DF, Giles GG, McCredie MR, Hopper JL. After BRCA1 and BRCA2-what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. Am J Hum Genet 2001;68:420–31.
7. Kaufman DJ, Beaty TH, Struewing JP. Segregation analysis of 231 Ashkenazi Jewish families for evidence of additional breast cancer susceptibility genes. Cancer Epidemiol Biomarkers Prev 2003;12:1045–52.
8. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer 1997;71:800–9.
9. Liede A, Malik IA, Aziz Z, Rios Pd P de los, Kwan E, Narod SA. Contribution of BRCA1 and BRCA2 mutations to breast and ovarian cancer in Pakistan. Am J Hum Genet 2002;71:595–606.
10. Assié G, LaFramboise T, Platzer P, Eng C. Frequency of germline genomic homozygosity associated with cancer cases. JAMA 2008;299:1437–45.
11. Enciso-Mora V, Hosking FJ, Houlston RS. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. Eur J Hum Genet EJHG 2010;18:909–14.
12. Oldenburg RA, Kroeze-Jansema KHG, Houwing-Duistermaat JJ, Bayley J-P, Dambrot C, van Asperen CJ, van den Ouweland AMW, Bakker B, van Beers EH, Nederlof PM, Vasen H, Hoogerbrugge N, et al. Genome-wide linkage scan in Dutch hereditary non-BRCA1/2 breast cancer families identifies 9q21-22 as a putative breast cancer susceptibility locus. Genes

Chromosomes Cancer 2008;47:947–56.
13. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30:97–101.
14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
15. SeattleSeq Variation Annotation [Internet]. [cited 2015 May 1];Available from: http://snp.gs.washington.edu/SeattleSeqAnnotation138/index.jsp
16. Exome Variant Server [Internet]. [cited 2015 Jun 1];Available from: http://evs.gs.washington.edu/EVS/
17. International HapMap Consortium. A haplotype map of the human genome. Nature 2005;437:1299–320.
18. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
19. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 2014;46:818–25.
20. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91.
21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6.
22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5.
23. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
25. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353–61, 361e1-2.
26. Desilva MG, Lu J, Donadel G, Modi WS, Xie H, Notkins AL, Lan MS. Characterization and chromosomal localization of a new protein disulfide isomerase, PDIp, highly expressed in human pancreas. DNA Cell Biol 1996;15:9–16.
27. Fu X-M, Zhu BT. Human pancreas-specific protein disulfide-isomerase (PDIp) can function as a chaperone independently of its enzymatic activity by forming stable complexes with denatured substrate proteins. Biochem J 2010;429:157–69.
28. Klappa P, Stromer T, Zimmermann R, Ruddock LW, Freedman RB. A pancreas-specific glycosylated protein disulphide-isomerase binds to misfolded proteins and peptides with an interaction inhibited by oestrogens. Eur J Biochem 1998;254:63–9.
29. Fu X-M, Zhu BT. Human pancreas-specific protein disulfide isomerase homolog (PDIp) is an intracellular estrogen-binding protein that modulates estrogen levels and actions in target cells. J Steroid Biochem Mol Biol 2009;115:20–9.
30. Alper SL, Sharma AK. The SLC26 gene family of anion transporters and channels. Mol Aspects Med 2013;34:494–515.
31. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome Res 2007;17:1723–30.
32. CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065

controls from 10 studies. Am J Hum Genet 2004;74:1175–82.
33. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, Zemankova P, Havranek O, Vocka M, Cerna M, Lhotova K, Borecka M, Janatova M, et al. Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. Int J Cancer 2019;145:1782–97.
34. Fletcher O, Johnson N, dos Santos Silva I, Orr N, Ashworth A, Nevanlinna H, Heikkinen T, Aittomäki K, Blomqvist C, Burwinkel B, Bartram CR, Meindl A, et al. Missense variants in ATM in 26,101 breast cancer cases and 29,842 controls. Cancer Epidemiol Biomarkers Prev 2010;19:2143–51.
35. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA, Berman SH, Blecua P, Riaz N, Huo Y, Ding YC, Neuhausen SL, et al. Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene 2017;36:4161–70.
36. Buchholz TA, Weil MM, Ashorn CL, Strom EA, Sigurdson A, Bondy M, Chakraborty R, Cox JD, McNeese MD, Story MD. A Ser49Cys variant in the ataxia telangiectasia, mutated, gene that is more common in patients with breast carcinoma compared with population controls. Cancer 2004;100:1345–51.
37. Stredrick DL, Garcia-Closas M, Pineda MA, Bhatti P, Alexander BH, Doody MM, Lissowska J, Peplonska B, Brinton LA, Chanock SJ, Struewing JP, Sigurdson AJ. The ATM missense mutation p.Ser49Cys (c.146C>G) and the risk of breast cancer. Hum Mutat 2006;27:538–44.
38. Sopik V, Akbari MR, Narod SA. Genetic testing for RAD51C mutations: in the clinic and community. Clin Genet 2015;88:303–12.
39. Kiiski JI, Pelttari LM, Khan S, Freysteinsdottir ES, Reynisdottir I, Hart SN, Shimelis H, Vilske S, Kallioniemi A, Schleutker J, Leminen A, Bützow R, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci U S A 2014;111:15172–7.
40. Peterlongo P, Catucci I, Colombo M, Caleca L, Mucaki E, Bogliolo M, Marin M, Damiola F, Bernard L, Pensotti V, Volorio S, Dall'Olio V, et al. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. Hum Mol Genet 2015;24:5345–55.
41. Sun J, Wang Y, Xia Y, Xu Y, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Lou H, Xie Y. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet 2015;11:e1005228.
42. Cybulski C, Carrot-Zhang J, Kluźniak W, Rivera B, Kashyap A, Wokołorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, Huzarski T, Gronwald J, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet 2015;47:643–6.
43. Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, et al. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer 2015;121:25–33.
44. Thompson ER, Rowley SM, Li N, McInerny S, Devereux L, Wong-Brown MW, Trainer AH, Mitchell G, Scott RJ, James PA, Campbell IG. Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. J Clin Oncol 2016;34:1455–9.
45. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, Vaccari EM, Bissonnette J, Booker JK, Cremona ML, Gibellini F, Murphy PD, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med 2016;18:823–32.
46. Muranen TA, Greco D, Blomqvist C, Aittomäki K, Khan S, Hogervorst F, Verhoef S, Pharoah PDP, Dunning AM, Shah M, Luben R, Bojesen SE, et al. Genetic modifiers of CHEK2*1100delC-associated breast cancer risk. Genet Med 2016;
47. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015;107.

48. Muranen TA, Mavaddat N, Khan S, Fagerholm R, Pelttari L, Lee A, Aittomäki K, Blomqvist C, Easton DF, Nevanlinna H. Polygenic risk score is associated with increased disease risk in 52 Finnish breast cancer families. Breast Cancer Res Treat 2016;158:463–9.
49. Sawyer S, Mitchell G, McKinley J, Chenevix-Trench G, Beesley J, Chen XQ, Bowtell D, Trainer AH, Harris M, Lindeman GJ, James PA. A role for common genomic variants in the assessment of familial breast cancer. J Clin Oncol Off J Am Soc Clin Oncol 2012;30:4330–6.
50. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, Lee A, Vreeswijk MPG, Hollestelle A, Seynaeve C, Meijers-Heijboer H, Oosterwijk JC, Hoogerbrugge N, Olah E, Vasen HFA, et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. J Med Genet 2019;56:581–9.

Supplementary Methods

Exome sequencing read alignment and variant calling

Analysis of the sequence data was performed within the in-house pipeline framework Biopet (version 0.1.3)[1]. First, raw reads were trimmed based on quality using Sickle (version 1.200)[2] and adapter sequences were removed with the help of Cutadapt (version 1.1)[3]. Afterwards, the quality of the reads was assessed using Fastqc (version 0.10.1)[4]. The reads were then aligned to the human reference genome (hg19) with the help of BWA (version 0.7.8-r455)[5]. After alignment Picard (version 1.109.1722)[6] was used to sort the bam files and to mark duplicate reads. Using GATK (version 3.1-1-gcfc45fd)[7] we applied base quality score recalibration, indel realignment, called variants using HaplotypeCaller and recalibrated variant quality scores using Hapmap, Omni, 1000G and dbSNP for single nucleotide variants and Mills and dbSNP for indels (datasets as provided in gatk_bundle_2.5). Next, variants in the regions described in the family-specific BED files were selected using vcftools (v1.12b)[8].

Imputation of SNP array data for PRS calculation

Imputation was performed (without pre-phasing) with IMPUTE2 (version 2.3.2)[9] using both 1000G (phase 3 b37 haplotypes)[10] and GoNL (release5.3, imputation ready haplotypes)[11] as a reference. We imputed a region of 1Mb around every SNP of interest with a buffer of 500 kb. We set "k" to 200 and "k_hap" to 2000 and 998, for 1000G and GoNL respectively. Before imputation the reference panels were merged" using the "merge_ref_panels" option. The "effective size" of the population (Ne) was set to 20000. To replicate this analysis use "seed" 8256245.

Supplementary Table 1 Classification of genes according to level of evidence that protein-truncating variants in these genes are associated with breast cancer

Evidence level[a]		Genes
1.	Strong	*ATM, BRCA1, BRCA2, CHEK2, PALB2*
2.	Syndromic	*CDH1, PTEN, TP53*
3.	Likely	*BARD1, BRIP1, FANCC, FANCM, NF1, RAD51C, RAD51D*
4.	Suggestive	*AKT1, MEN1, MSH6, NBN, PIK3CA, RECQL, STK11*
5.	Unlikely	*ATR, EPCAM, FAM175A, GEN1, MLH1, MRE11A, MSH2, MUTYH, PMS2, PPM1D, RAD50, RINT1, XRCC2*

[a] 1: association has been demonstrated in multiple publications; 2: familial cancer syndromes in which breast cancer is a linked feature; 3: some controversy among studies, but meta-analyses positive; 4: a few positive studies, but no firm replication yet; 5: anecdotal evidence, or studies controversial

Supplementary Table 2 Overview of the SNPs included in the polygenic risk score

SNP	Ref	Alt	OR[a]	Beta[b]
rs616488	A	G	0.94	-0.06
rs2992756	T	C	0.94	-0.06
rs4233486	C	T	1.03	0.03
rs79724016	T	G	0.93	-0.07
rs1707302	A	G	1.04	0.04
rs140850326	Del	Ins	0.97	-0.03

SNP	Ref	Alt	OR[a]	Beta[b]
rs17426269	G	A	1.05	0.05
rs12022378	C	T	1.04	0.04
rs7529522	T	C	1.06	0.06
rs11249433	A	G	1.11	0.10
rs12405132	C	T	0.97	-0.03
rs12048493	A	C	1.04	0.04
rs4971059	G	A	1.05	0.05
rs35383942	C	T	1.12	0.11
rs11117758	G	A	0.95	-0.05
rs72755295	A	G	1.15	0.14
rs113577745	C	G	1.08	0.08
rs12710696	T	C	0.97	-0.03
rs6725517	A	G	0.96	-0.04
rs3833441	Del	Ins	1.09	0.09
rs4849887	T	C	1.10	0.09
rs2016394	G	A	0.95	-0.05
rs1550623	G	A	1.05	0.05
rs1830298	C	T	0.94	-0.06
rs34005590	C	A	0.82	-0.20
rs4442975	G	T	0.89	-0.12
rs16857609	C	T	1.06	0.06
rs12479355	A	G	0.96	-0.04
rs6762644	A	G	1.05	0.05
rs4973768	C	T	1.11	0.10
rs12493607	G	C	1.05	0.05
rs6796502	G	A	0.92	-0.08
rs1053338	A	G	1.05	0.05
rs6805189	T	C	0.97	-0.03
rs13066793	A	G	0.94	-0.06
rs9833888	G	T	1.06	0.06
rs34207738	Del	Ins	1.06	0.06
rs58058861	G	A	1.06	0.06
rs6815814	A	C	1.06	0.06
rs10718573	Ins	Del	0.96	-0.04
rs10022462	C	T	1.04	0.04
rs9790517	C	T	1.04	0.04
rs77528541	G	T	0.95	-0.05
rs6828523	C	A	0.91	-0.09
rs116095464	T	C	1.06	0.06

SNP	Ref	Alt	OR[a]	Beta[b]
rs3215401	Del	Ins	0.93	-0.07
rs10069690	C	T	1.06	0.06
rs2012709	C	T	1.02	0.02
rs10941679	A	G	1.15	0.14
rs62355902	A	T	1.18	0.17
rs10472076	T	C	1.03	0.03
rs1353747	T	G	0.96	-0.04
rs72749841	T	C	0.93	-0.07
rs35951924	Del	Ins	0.95	-0.05
rs7707921	T	A	1.06	0.06
rs10474352	C	T	0.94	-0.06
rs6882649	G	T	1.03	0.03
rs6596100	C	T	0.94	-0.06
rs1432679	C	T	0.93	-0.08
rs4562056	G	T	1.05	0.05
rs204247	G	A	0.96	-0.04
rs3819405	C	T	0.96	-0.04
rs2223621	T	C	0.96	-0.04
rs71557345	G	A	0.92	-0.08
rs17529111	T	C	1.02	0.02
rs12207986	G	A	1.03	0.03
rs3757322	T	G	1.08	0.08
rs9397437	G	A	1.17	0.16
rs2747652	T	C	1.06	0.06
rs6569648	C	T	1.06	0.06
rs7971	A	G	0.96	-0.04
rs17156577	T	C	1.05	0.05
rs6964587	G	T	1.03	0.03
rs17268829	T	C	1.05	0.05
rs71559437	G	A	0.93	-0.07
rs4593472	C	T	0.97	-0.03
rs11977670	G	A	1.06	0.06
rs720475	G	A	0.96	-0.04
rs9693444	A	C	0.94	-0.06
rs13365225	A	G	0.91	-0.09
rs6472903	G	T	1.06	0.06
rs2943559	A	G	1.10	0.10
rs514192	A	T	0.95	-0.05
rs12546444	A	T	0.93	-0.07

SNP	Ref	Alt	OR[a]	Beta[b]
rs13267382	A	G	0.97	-0.03
rs58847541	G	A	1.08	0.08
rs13281615	A	G	1.11	0.10
rs11780156	C	T	1.05	0.05
rs1011970	G	T	1.07	0.07
rs10759243	C	A	1.06	0.06
rs676256	C	T	1.10	0.09
rs10816625	A	G	1.11	0.10
rs13294895	C	T	1.06	0.06
rs1895062	A	G	0.94	-0.06
rs10760444	G	A	0.97	-0.03
chr9:136151579	Ins	Del	1.03	0.03
rs67958007	Ins	Del	1.09	0.09
rs7072776	A	G	0.95	-0.05
rs11814448	A	C	1.12	0.11
rs10995201	A	G	0.90	-0.11
rs704010	T	C	0.93	-0.07
rs140936696	Ins	Del	0.96	-0.04
rs7904519	A	G	1.03	0.03
rs11199914	C	T	0.96	-0.04
rs35054928	Ins	Del	0.79	-0.24
rs45631563	A	T	1.23	0.21
rs2981578	C	T	0.81	-0.21
rs3817198	T	C	1.05	0.05
rs6597981	A	G	1.04	0.04
rs3903072	G	T	0.97	-0.03
rs554219	C	G	1.21	0.19
chr11:69088342	C	A	1.28	0.25
rs11820646	T	C	1.04	0.04
rs12422552	G	C	1.06	0.06
rs7297051	C	T	0.89	-0.12
rs202049448	T	C	0.95	-0.05
rs17356907	A	G	0.91	-0.09
rs1292011	A	G	0.92	-0.08
rs206966	C	T	1.05	0.05
rs11571833	A	T	1.35	0.30
rs6562760	A	G	1.05	0.05
rs2236007	G	A	0.93	-0.07
rs2588809	T	C	0.94	-0.06

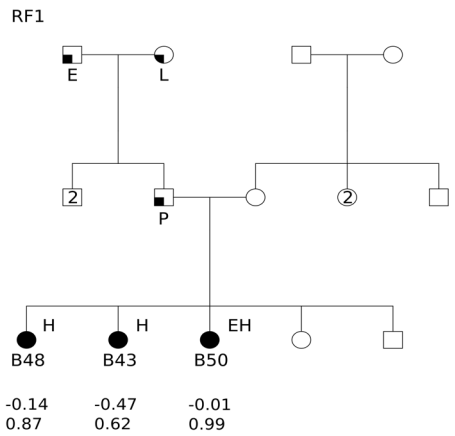
SNP	Ref	Alt	OR[a]	Beta[b]
rs999737	C	T	0.91	-0.09
rs941764	A	G	1.03	0.03
rs11627032	T	C	0.96	-0.04
rs10623258	Del	Ins	1.04	0.04
rs2290203	G	A	0.94	-0.06
rs4784227	C	T	1.23	0.21
rs17817449	T	G	0.95	-0.05
rs11075995	A	T	0.97	-0.03
rs28539243	G	A	1.05	0.05
rs2432539	A	G	0.97	-0.03
rs13329835	A	G	1.07	0.07
rs4496150	C	A	0.96	-0.04
rs146699004	Ins	Del	0.97	-0.03
rs72826962	C	T	1.20	0.18
chr17:44252468	G	A	0.95	-0.05
rs2787486	A	C	0.93	-0.07
rs745570	A	G	1.05	0.05
rs527616	C	G	1.03	0.03
rs1436904	T	G	0.95	-0.05
rs117618124	T	C	0.89	-0.12
rs6507583	A	G	0.92	-0.08
rs78269692	T	C	1.09	0.09
rs2594714	G	A	0.97	-0.03
rs2965183	G	A	1.04	0.04
chr19:17262404	C	G	1.03	0.03
rs4808801	A	G	0.93	-0.07
rs71338792	Del	Ins	1.05	0.05
rs3760982	A	G	0.95	-0.05
rs16991615	G	A	1.10	0.10
rs6122906	A	G	1.05	0.05
rs2823093	G	A	0.94	-0.06
rs17879961	A	G	1.26	0.23
rs132390	C	T	0.96	-0.04
rs6001930	T	C	1.12	0.11
rs738321	C	G	0.95	-0.05
rs73161324	C	T	1.06	0.06
rs28512361	G	A	1.05	0.05

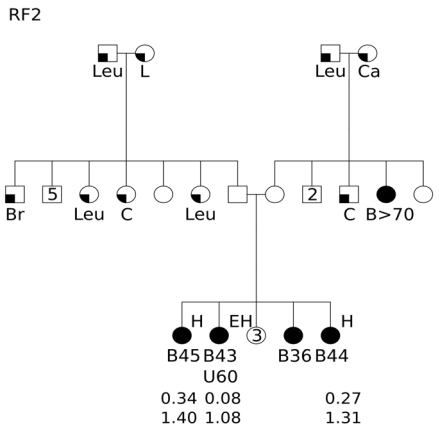
[a] odds ratios (OR) for the alternative (Alt) allele derived from Michailidou et al. 2017[12], [b] beta for the Alt allele calculated based on the ORs from Michailidou et al. 2017[12]

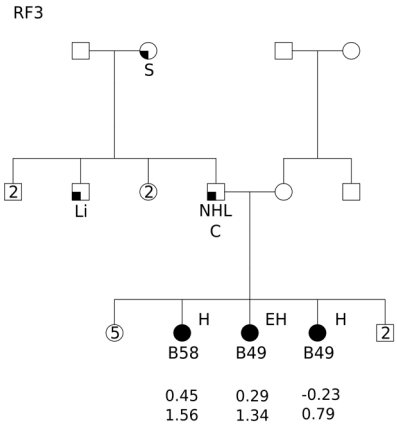
Supplementary Table 3 Rare missense variants found in the regions where the sibships share two haplotypes

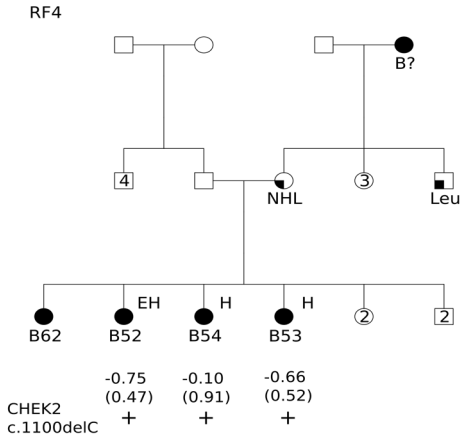
Gene	Family	Variant (coding DNA)	Variant (protein)	Rs-number	Frequency[a]	CADD
SERINC2	RF4	c. 364C>T	p.R126W	rs183001614	0.0053	19.010
ZNF717	RF7	c.188T>A	p.H63L	rs201105907	0.0001	0.015

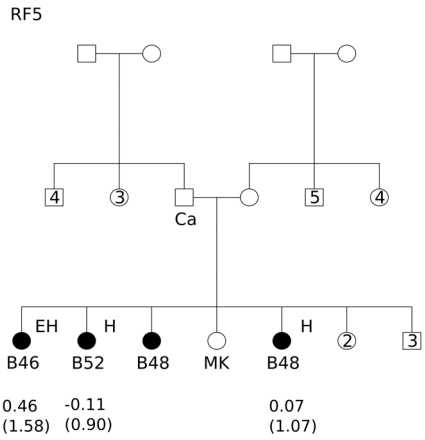
[a] Frequency in GnomAD[13] Accession numbers for the transcripts and protein sequences used to describe the variants can be found in Supplementary table 1

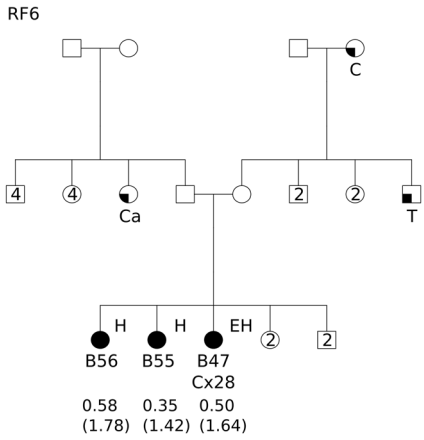

RF1
E
L
2
P
②
H
H
EH
B48
B43
B50
-0.14
0.87
-0.47
0.62
-0.01
0.99

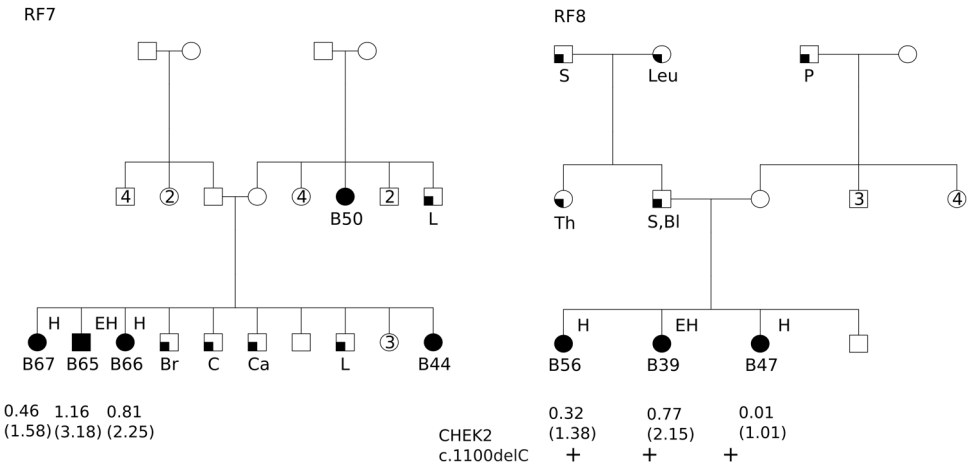

RF2
Leu
L
Leu
Ca
Br
5
Leu
C
Leu
2
C
B>70
H
EH
③
H
B45
B43
U60
B36
B44
0.34
0.08
1.40
1.08
0.27
1.31

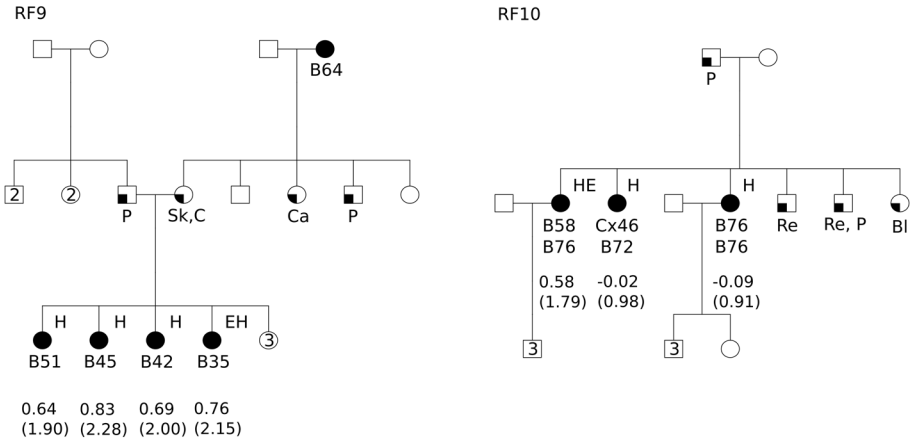

RF3
S
2
Li
②
NHL
C
⑤
H
EH
H
B58
B49
B49
2
0.45
1.56
0.29
1.34
-0.23
0.79

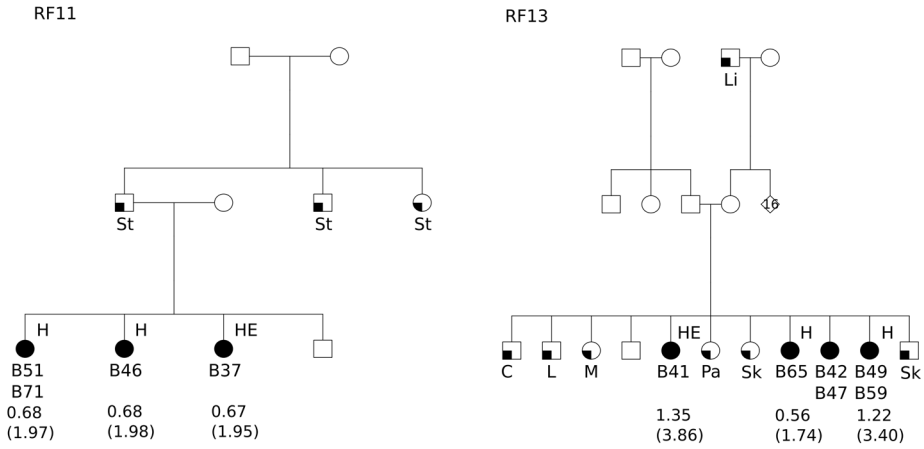

RF4
B?
4
NHL
③
Leu
EH
H
H
B62
B52
B54
B53
②
2
-0.75
(0.47)
-0.10
(0.91)
-0.66
(0.52)
CHEK2
c.1100delC
+
+
+


RF5
4
③
Ca
5
④
EH
H
H
B46
B52
B48
MK
B48
②
3
0.46
(1.58)
-0.11
(0.90)
0.07
(1.07)

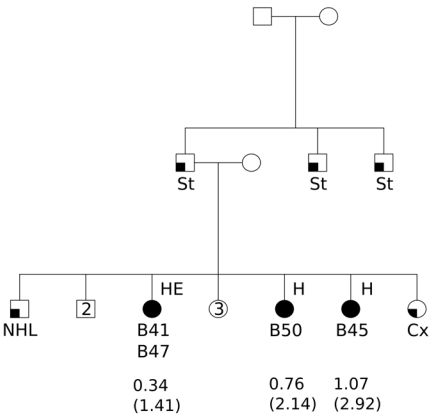

RF6
C
4
④
Ca
2
②
T
H
H
EH
②
B56
B55
B47
Cx28
2
0.58
(1.78)
0.35
(1.42)
0.50
(1.64)


RF7
4
②
④
B50
2
L
H
EH
H
B67
B65
B66
Br
C
Ca
L
③
B44
0.46 1.16 0.81
(1.58) (3.18) (2.25)
RF8
S
Leu
P
Th
S,Bl
3
④
H
EH
H
B56
B39
B47
0.32 0.77 0.01
(1.38) (2.15) (1.01)
CHEK2
c.1100delC
+ + +

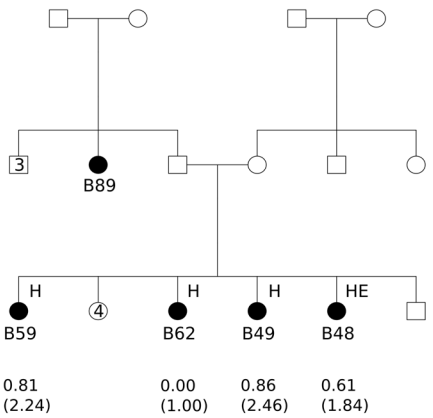

RF9
B64
2
②
P
Sk,C
Ca
P
H
H
H
EH
③
B51
B45
B42
B35
0.64 0.83 0.69 0.76
(1.90) (2.28) (2.00) (2.15)
RF10
P
HE
H
H
B58
B76
Cx46
B72
B76
B76
Re
Re, P
Bl
0.58 -0.02 -0.09
(1.79) (0.98) (0.91)
3
3


RF11
St
St
St
H
H
HE
B51
B71
B46
B37
0.68 0.68 0.67
(1.97) (1.98) (1.95)
RF13
Li
16
HE
H
H
C
L
M
B41
Pa
Sk
B65
B42
B47
B49
B59
Sk
1.35 0.56 1.22
(3.86) (1.74) (3.40)

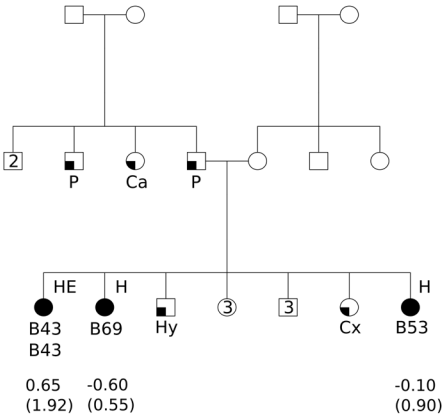
RF14



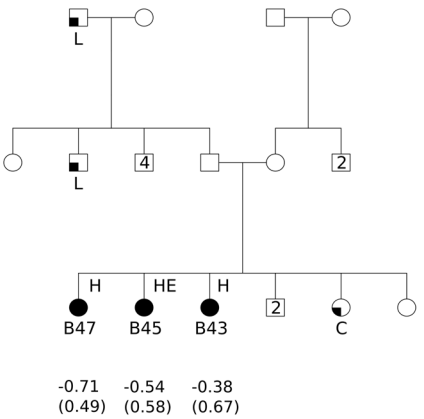
RF15



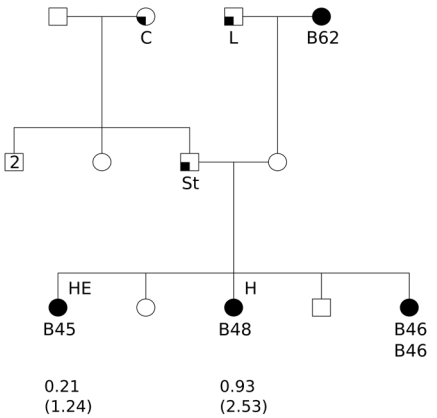
RF17



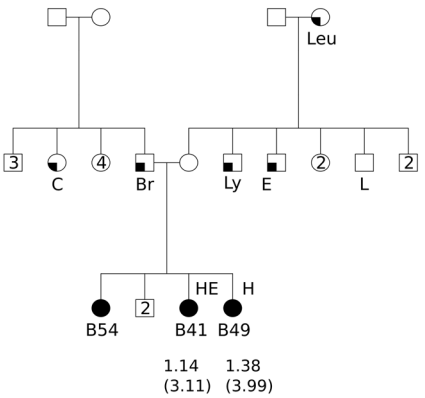
RF18

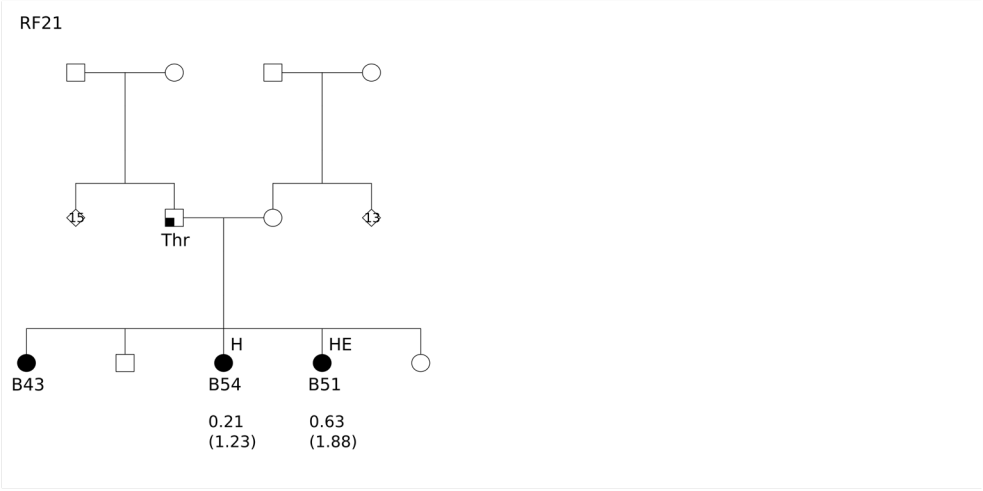


RF19



RF20





Supplementary Figure 1. Pedigrees of the families included in this study
H indicates that an individual's germline DNA was haplotyped, HE indicates that an individual's germline DNA was both haplotyped and exome sequenced. Under each affected individual for whom germline DNA was available the normalized PRS and OR (in italics) are indicated. The * -symbol indicates individuals carrying the CHEK2*1100delC variant. B= breast cancer, Bl= bladder cancer, Br= brain cancer, C= colon cancer, Ca= cancer not otherwise specified, Cx= cervical cancer, E= esophagus cancer, Hy= hypophysis cancer, L= lung cancer, Leu= leukemia, Li= liver cancer, Ly= lymphoma (not specified), M= melanoma, NHL= non-Hodgkin lymphoma, P= Prostate cancer, Pa= pancreas cancer, Re= renal cancer, S= stomach cancer, Sk = skin cancer (not specified), T= testicular cancer, Th= thyroid cancer.

References

1 The Leiden University Medical Centre. Biopet. (0.1.3). 2014.
2 Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Published Online First: 2011.http://github.com/najoshi/sickle (accessed 14 Apr2017).
3 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011;17:10–2.
4 Babraham Institute. Fastqc. (0.10.1). 2012.
5 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 2009;25:1754–60.
6 Broad Institute. Picard. (1.109.1722). 2014.
7 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
8 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinforma Oxf Engl 2011;27:2156–8.
9 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
10 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
11 Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet 2014;46:818–25.
12 Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, Bolla MK, Wang Q, Tyrer J, Dicks E, Lee A, Wang Z, Allen J, Keeman R, Eilber U, French JD, Qing Chen X, Fachal L, McCue K, McCart Reed AE, Ghoussaini M, Carroll JS, Jiang X, Finucane H, Adams M, Adank MA, Ahsan H, Aittomäki K, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Arun B, Auer PL, Bacot F, Barrdahl M, Baynes C, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Børresen-Dale A-L, Brand JS, Brauch H, Brennan P, Brenner H, Brinton L, Broberg P, Brock IW, Broeks A, Brooks-Wilson A, Brucker SY, Brüning T, Burwinkel B, Butterbach K, Cai Q, Cai H, Caldés T, Canzian F, Carracedo A, Carter BD, Castelao JE, Chan TL, David Cheng T-Y, Seng Chia K, Choi J-Y, Christiansen H, Clarke CL, NBCS Collaborators, Collée M, Conroy DM, Cordina-Duverger E, Cornelissen S, Cox DG, Cox A, Cross SS, Cunningham JM, Czene K, Daly MB, Devilee P, Doheny KF, Dörk T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Elvira M, Engel C, Eriksson M, Fasching PA, Figueroa J, Flesch-Janys D, Fletcher O, Flyger H, Fritschi L, Gaborieau V, Gabrielson M, Gago-Dominguez M, Gao Y-T, Gapstur SM, García-Sáenz JA, Gaudet MM, Georgoulias V, Giles GG, Glendon G, Goldberg MS, Goldgar DE, González-Neira A, Grenaker Alnæs GI, Grip M, Gronwald J, Grundy A, Guénel P, Haeberle L, Hahnen E, Haiman CA, Håkansson N, Hamann U, Hamel N, Hankinson S, Harrington P, Hart SN, Hartikainen JM, Hartman M, Hein A, Heyworth J, Hicks B, Hillemanns P, Ho DN, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Hou M-F, Hsiung C-N, Huang G, Humphreys K, Ishiguro J, Ito H, Iwasaki M, Iwata H, Jakubowska A, Janni W, John EM, Johnson N, Jones K, Jones M, Jukkola-Vuorinen A, Kaaks R, Kabisch M, Kaczmarek K, Kang D, Kasuga Y, Kerin MJ, Khan S, Khusnutdinova E, Kiiski JI, Kim S-W, Knight JA, Kosma V-M, Kristensen VN, Krüger U, Kwong A, Lambrechts D, Le Marchand L, Lee E, Lee MH, Lee JW, Neng Lee C, Lejbkowicz F, Li J, Lilyquist J, Lindblom A, Lissowska J, Lo W-Y, Loibl S, Long J, Lophatananon A, Lubinski J, Luccarini C, Lux MP, Ma ESK, MacInnis RJ, Maishman T, Makalic E, Malone KE, Kostovska IM, Mannermaa A, Manoukian S, Manson JE, Margolin S, Mariapun S, Martinez ME, Matsuo K, Mavroudis D, McKay J, McLean C, Meijers-Heijboer H, Meindl A, Menéndez P, Menon U, Meyer J, Miao H, Miller N, Taib NAM, Muir K, Mulligan AM, Mulot C, Neuhausen SL, Nevanlinna H, Neven P, Nielsen SF, Noh D-Y, Nordestgaard BG, Norman

A, Olopade OI, Olson JE, Olsson H, Olswold C, Orr N, Pankratz VS, Park SK, Park-Simon T-W, Lloyd R, Perez JIA, Peterlongo P, Peto J, Phillips K-A, Pinchev M, Plaseska-Karanfilska D, Prentice R, Presneau N, Prokofyeva D, Pugh E, Pylkäs K, Rack B, Radice P, Rahman N, Rennert G, Rennert HS, Rhenius V, Romero A, Romm J, Ruddy KJ, Rüdiger T, Rudolph A, Ruebner M, Rutgers EJT, Saloustros E, Sandler DP, Sangrajrang S, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schürmann P, Scott RJ, Scott C, Seal S, Seynaeve C, Shah M, Sharma P, Shen C-Y, Sheng G, Sherman ME, Shrubsole MJ, Shu X-O, Smeets A, Sohn C, Southey MC, Spinelli JJ, Stegmaier C, Stewart-Brown S, Stone J, Stram DO, Surowy H, Swerdlow A, Tamimi R, Taylor JA, Tengström M, Teo SH, Beth Terry M, Tessier DC, Thanasitthichai S, Thöne K, Tollenaar RAEM, Tomlinson I, Tong L, Torres D, Truong T, Tseng C-C, Tsugane S, Ulmer H-U, Ursin G, Untch M, Vachon C, van Asperen CJ, Van Den Berg D, van den Ouweland AMW, van der Kolk L, van der Luijt RB, Vincent D, Vollenweider J, Waisfisz Q, Wang-Gohrke S, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Wu AH, Xia L, Yamaji T, Yang XR, Har Yip C, Yoo K-Y, Yu J-C, Zheng W, Zheng Y, Zhu B, Ziogas A, Ziv E, ABCTB Investigators, ConFab/AOCS Investigators, Lakhani SR, Antoniou AC, Droit A, Andrulis IL, Amos CI, Couch FJ, Pharoah PDP, Chang-Claude J, Hall P, Hunter DJ, Milne RL, García-Closas M, Schmidt MK, Chanock SJ, Dunning AM, Edwards SL, Bader GD, Chenevix-Trench G, Simard J, Kraft P, Easton DF. Association analysis identifies 65 new breast cancer risk loci. Nature Published Online First: 23 October 2017. doi:10.1038/nature24284

13 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Consortium TGAD, Neale BM, Daly MJ, MacArthur DG. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 2019;:531210.

Chapter 6

The impact of next generation sequencing on the analysis of breast cancer susceptibility: a role for extremely rare genetic variation?

Florentine S. Hilbers, Maaike P.G. Vreeswijk, Christi J. van Asperen, Peter Devilee

Clin Genet. 2013 Nov; 84(5): 407-14

Abstract

Women with a family history of breast cancer have an approximately twofold elevated risk of the disease. Even though an array of genes has been associated with breast cancer risk the past two decades, variants within these genes jointly explain at most 40% of this familial risk. Many explanations for this 'missing heritability' have been proposed, including the existence of many very rare variants, interactions between genetic and environmental factors and structural genetic variation. In this review, we discuss how next generation sequencing will teach us more about the genetic architecture of breast cancer, with a specific focus on very rare genetic variants. While such variants potentially explain a substantial proportion of familial breast cancer, assessing the breast cancer risks conferred by them remains challenging, even if this risk is relatively high. To assess more moderate risks, epidemiological approaches will require very large patient cohorts to be genotyped for the variant, only achievable through international collaboration. How well we will be able to eventually resolve the missing heritability for breast cancer in a clinically meaningful way crucially depends on the underlying complexity of the genetic architecture.

The genetic landscape of breast cancer

Genetic variation in over 75 loci has been significantly associated with breast cancer risk the past 20 years, either by linkage studies in multiple-case families, genome-wide association scans, or candidate gene mutation scanning. Despite this impressive progress, currently known risk alleles explain only about 40% of familial breast cancer risk.[1] These known risk alleles can be roughly subdivided in three groups based on their relative breast cancer risk and population frequencies. Broadly speaking, these are very rare high risk alleles, common low risk alleles, and an intermediate group of 'uncommon' alleles associated with a two- to threefold elevated risk (see Fig. 1 and Box 1 for definitions). The existence and potential impact of a fourth group, rare to very rare low risk alleles, has thusfar remained in the realm of speculation, as these are extremely difficult to detect.

Box 1 Rare alleles

The terms rare and very rare are used somewhat arbitrary in literature.
In this review we use the following alternative allele frequencies as cut-offs:
Common (50–5%)
Low frequency (5–0.5%)
Rare (0.5–0.05%)
Very rare (<0.05%)

BRCA1 and BRCA2 remain the two most significant genes to date. Mutations in these genes confer high breast cancer risks and explain approximately 20% of familial breast cancer.[2–4] The spectrum of mutations found in either gene is extremely complex, with almost 2000 unique mutations (or 'alleles') having been documented for each.[5] Interestingly, over half of these have been detected only once so far, and thus represent extremely rare population frequencies, whereas other have been found recurrently, often within specific ethnic minorities as a result of a founder effect. Well known examples include the *BRCA1* c.66_67delAG mutation and the *BRCA2* c.5946delT and *BRCA2* c.771_775delTCAAA mutations, which have been detected in 0.5–1% of the general populations of Ashkenazi Jews, and Iceland, respectively. The first estimates of breast and ovarian cancer risks conferred by *BRCA1* and *BRCA2* were linkage-based, i.e. regardless of mutation-type. These analyses indicated that *BRCA1* mutations confer a cumulative breast cancer risk of 87% by age 70, and an ovarian cancer risk of 45% by age 70. A wealth of data on mutation carriers have since documented that patterns of risk change significantly for both *BRCA1* and *BRCA2* mutations conditional on their relative position within the gene.[6,7] These data show that the relation between mutation position and cancer risk is complex, and as such provide an example as to what to expect at other susceptibility loci for which it will be more difficult to obtain such a large dataset. The classical view is that mutant alleles that inactivate *BRCA1* or *BRCA2* cellular functions, e.g. through premature protein-truncation or aberrant mRNA-splicing, confer high breast and ovarian cancer risks. Two lessons can be gleaned from the *BRCA1/2* case, to be taken into account when trying to interpret very rare variants in other breast cancer susceptibility genes. First, while the majority of mutations in *BRCA1* and *BRCA2* indeed seem to functionally inactivate the resulting protein, very few analyses have been done, besides for the few well known founder mutations, to actually establish that each unique mutation causes a high risk of breast and ovarian cancer. Instead, the risk associated with very rare mutations has been inferred from the joint analyses

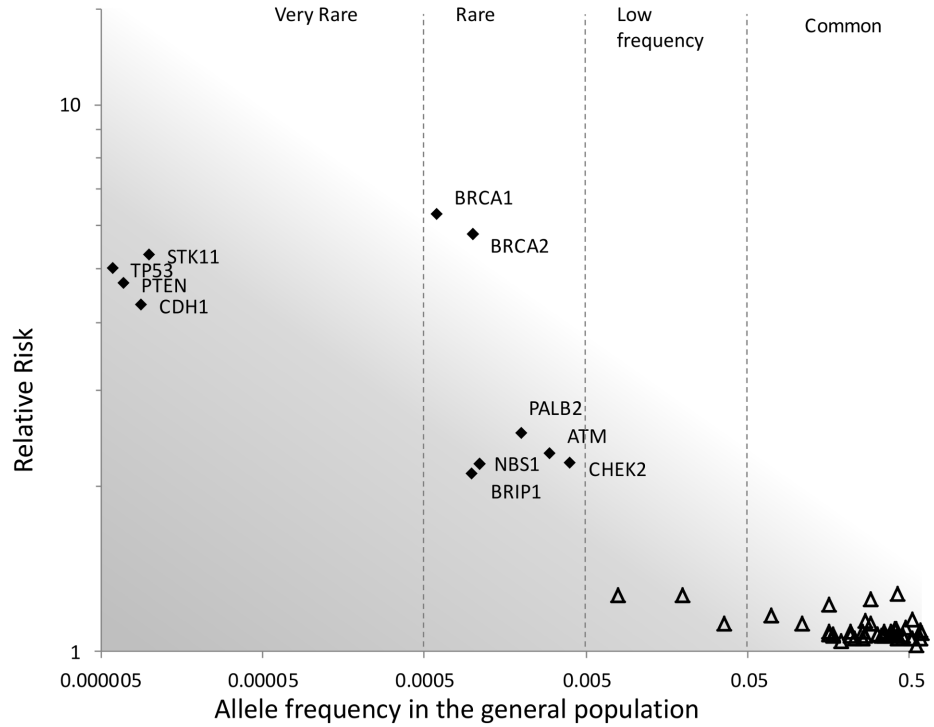


Figure 1 The genetic landscape of breast cancer
This figure shows the allele frequencies and risk distributions for the currently known breast cancer risk alleles. Genes represented by a filled diamond indicate a joint allele frequency, and an average risk associated with all observed alleles/mutations in this gene. Hence there might be variants within this gene that are associated with a much higher or lower risk than this given average, and individual risk alleles will have much lower allele frequencies than suggested by the value on the x-axis. In the case of genes in which the breast cancer associated variants are very rare, an approximation of the risk is given since there is still much uncertainty about these risks. Genes indicated by an open triangle represent a single variant for which the associated breast cancer risk was derived from case-control analyses.

of all these alleles in families (also called 'burden analysis', see below). Yet, in the daily clinical practice of counseling familial breast cancer, every protein-truncating variant in *BRCA1/2* is interpreted as 'high risk'. Second, several notable deviations from the expected risk-patterns have been detected, in that some mutations in *BRCA1* or *BRCA2*, cause either a low (e.g. *BRCA2* p.Lys3326*)[8] or moderately increased risks to breast cancer (e.g. *BRCA1* c.Arg1966Gln).[9] This implies that for many very rare variants for which the effect on protein function is less clear a causal relation with breast cancer is hard to establish.

Other genes with high risk breast cancer alleles were discovered because of their strong association with familial cancer syndromes in which breast cancer was one of the defining components, and include the genes *TP53* (Li-Fraumeni syndrome),[10] *PTEN* (Cowden syndrome),[11] *STK11* (Peutz-Jeghers syndrome)[12,13] and *CHD1* (hereditary diffuse gastric cancer syndrome).[14–16] Mutations in these genes are very rare in breast cancer families that do not fit the clinical criteria for these syndromes,[17–21] but are associated with a 2- to 10-fold increased risk of breast cancer. Likewise, *ATM* was an immediate candidate to contain breast cancer susceptibility alleles. Ataxia telangiectasia, caused by *ATM* mutations, is a recessive disorder,

and mothers of patients had already been documented to have an elevated breast cancer risk.[22]

Once it became clear that *BRCA1* and *BRCA2* encode proteins with a role in the DNA damage response to double-strand DNA breaks, and could be connected to a gene network underlying the recessive disorder Fanconi anemia (FA),[23] re-sequencing of other genes constituting these pathways in large patient cohorts resulted in the discovery of a second group of genes in which variants are associated with a more moderately increased risk of breast cancer. This group includes the genes *ATM*,[24] *CHEK2*,[25] *BRIP1*,[26] *PALB2*[27] and *NBS1*.[28] Most variants in these genes are thought to be associated with an approximately twofold increased risk. Some of these variants are relatively common, with allele frequencies in the general population up to 1%. However, most variants are very rare and their relation with breast cancer could only be established by a burden analysis pooling very rare variants and comparing their combined frequency between cases and controls. Accordingly, they explain a relatively small proportion of familial breast cancer.[27] Other genes for which associations have been found through candidate gene approaches are *FAM175A* (Abraxas),[29] *BARD1*,[30,31] *RAD51C*,[32] *MRE11*[33] and *RAD50*.[34] However, evidence for these genes is still limited and sometimes contradicting.[35–42] The last group of genetic variants associated with breast cancer consists of low risk variants of which the minor allele frequency is usually higher than 5%. Currently over 60 of these common low risk variants have been identified through large genome-wide association studies with the per allele odds ratios for the risk allele ranging from 1.02 to 1.27[8]. Interestingly, these low risk variants are usually found outside protein coding regions and most cannot directly be linked to a gene encoding a DNA damage response-related protein.

Missing heritability

Multiple models have been proposed to explain the 'missing heritability' in breast cancer. After the discovery of *BRCA1* and *BRCA2*, segregation studies have suggested that a polygenic model most probably explains the majority of remaining familial risk.[43–45] The inability to identify a third '*BRCA1/2* -like' gene by several genome-wide linkage analyses suggests that additional high risk alleles are probably very rare and scattered across several loci. As a result of recent extreme human population growth, the distribution of variant allele frequencies is strongly skewed towards very rare variants.[46] Conceivably, these variants could confer high, low or moderate breast cancer risks, but genetic drift and selection have not yet had the time to mould their allele frequencies. Jointly therefore, they potentially explain a major part of the polygenic risk of breast cancer. Candidate gene association studies have been underpowered to detect these variants and the single-nucleotide polymorphism (SNP) arrays typically used in genome-wide association studies have been designed to only tag common variation. A specific class of rare variants that might contribute to familial breast cancer risk are structural variants like large deletions and copy number variations. These variants are not detected by most genotyping methods used in association studies. Instead, a sufficiently sensitive detection would require a technique (or combination of techniques) that assesses both break point sequences and copy number variation in a genome wide manner.

Finally, interactions between genetic variants and/or environmental risk factors remain a largely unexplored area as most case–control studies that could address this are typically underpowered to detect such associations. Even very large cohort-studies such as those undertaken by the Breast Cancer Association Consortium have been able to investigate only two-way interactions. This consortium has now reported a few significant interactions between common low risk variants and environmental risk factors.[47] In addition, many of the

common SNPs associated with breast or ovarian cancer risk have been found to be able to modify the risks conferred by *BRCA1* or *BRCA2*, indicating that genetic interactions in fact do exist.[48] To detect additional interactions, (environmental) risk factor data has to be collected for even larger cohorts and rare variants might have to be pooled based on their effect on a specific gene or pathway.

Identifying new breast cancer risk alleles

Although a number of additional breast cancer genes have been identified by candidate gene approaches, an important disadvantage is that this approach is limited by our current understanding of the pathways involved in breast cancer pathogenesis. Next generation sequencing (NGS, Box 2) has brought the opportunity to discover rare alleles associated with breast cancer risk in a more agnostic way by allowing researchers to sequence whole exomes or even whole genomes. However, also this approach comes with challenges. Most studies applying NGS in familial breast cancer cases have taken a whole exome sequencing approach.[49–53] A typical exome sequencing experiment results in tens of thousands of heterozygous variants, which somehow will have to be reduced to a number manageable for validation. Many bioinformatics tools exist to assist these complex analyses; however, many different combinations of tools and settings for data analysis and variant selection are reported in literature. Common strategies include focusing on variants predicted to result in a truncated protein or missense changes probably to affect protein function by in silico prediction algorithms. Also the removal of variants with an allele frequency of more than 1% (in publicly available databases) is a common filtering step. However, all these filtering steps come with the risk of discarding a causal variant.

Some exome sequencing studies have suggested *FANCM*,[49] *BLM*,[53] *FANCC*[53] and *XRCC2*[52] as potential new breast cancer genes, while others have not reported likely new risk alleles.[50,51] Of note, none of the exome sequencing studies highlighting new genes provide conclusive evidence for their involvement in breast cancer, but rely on previous data and functional connotation to support their candidacy. *FANCC* and *FANCM* are both FA genes and were obvious candidates ever since *BRCA2* was found to be a FA gene.[54] An association between *FANCC* mutations and breast cancer had been reported before,[55] but for *FANCM* the available data are conflicting.[56] Likewise, *XRCC2* has been suggested to be a FA gene,[57] but the association between *XRCC2* variants and familial breast cancer was not detected by two other case–control studies.[58,59] Mutations in *BLM* are known to cause Bloom syndrome, a very rare recessive disorder characterized by short stature and high incidence of multiple cancers.[60] An association between heterozygous *BLM* mutations and breast cancer has been reported before, although not all truncating mutations seem to be associated with a similar risks.[61–64]

An interesting alternative to the exome sequencing approach is a study by Ruark et al.,[65] who used NGS to analyse a gene-panel of 507 genes implicated in DNA damage response to search for new familial breast cancer associated genes. This allowed them to sequence more cases and generate a deeper coverage at each DNA-base sequenced. Focusing on protein truncating variants, they reported such variants in the *PPM1D* gene to be significantly associated with breast and ovarian cancer, but at very rare allele frequencies. Interestingly, all protein truncating variants were mosaic in blood lymphocyte DNA, while functional analyses revealed that the tumorigenic mechanism probably does not comply with that of simple tumor suppressor gene inactivation.

Owing to high sequencing costs only a limited number of samples have been sequenced to date. This hampers selection of variants or genes on basis of their variant

frequencies in cases and controls. When sequencing costs drop further and the number of sequenced samples increases, the full potential of NGS in finding new breast cancer genes would be exploited if, through international collaboration, large numbers of cases would become accessible for sequence comparisons. The discovery of extremely rare breast cancer risk alleles would be particularly enhanced, as no single research centre is predicted to amass sufficient numbers of cases to detect these. However, in order to pool NGS data in an efficient and meaningful manner, consensus needs to be reached on data analysis strategies. At the moment consortia are being formed to pool NGS data in order to increase power.[66] As we can also expect a shift from exome sequencing to whole-genome sequencing, this will give more insight into the role of structural variants and variants in nonprotein- coding regions.

Box 2 Next generation sequencing

Next generation sequencing is a common used term for a set of sequencing techniques that are also known as Massively Parallel Sequencing (MPS). With this technique many different sequences can be determined in one reaction. For most experiments genomic DNA is randomly fragmented. Subsequently, these fragments can be enriched for all coding regions (exome) or a specific set of disease associated genes (gene panels). These fragments are then loaded on a chip and for each location on this chip, corresponding with a specific fragment, the sequence is determined. For the actual sequencing many different techniques exist, each with their own error rates and artifacts which have to be taken into account in the analysis. In the end, every position in the genome or region of interest will have been sequenced multiple times. These reads will have to be combined to a consensus sequence. The accuracy of this consensus sequence depends on how many times a position is sequenced, also known as the sequencing depth.

Detecting the effect of very rare variants

While NGS will identify many potential variants conferring breast cancer risk, assessing this risk purely on sequencing data usually has insufficient statistical power. Therefore, the next step is usually to perform a case–control study, examining the allele frequency of a specific variant in cases and controls. This strategy can be very successful if it is relatively common in the population under study, for example in the case of *CHEK2* c.1100delC.[25] However, most potentially causal variants are very rare even among familial cases, making it necessary to genotype a very large number of cases and controls. For example, in order to detect the effect of a single variant with an allele frequency of 0.05% and a relative risk of two with 80% power and an alpha of 5%, at least 22,000 cases and an equal number of controls are needed.[67] In some cases, the efficiency of a case–control study might be improved by selecting subjects from a certain geographical region or with a specific phenotype. For example, mutations in *BRCA1*, *BRCA2*,[68,69] *BRIP1*[70] and *RAD51C*[32] are also associated with an increased risk of ovarian cancer. Selecting cases from families with both breast and ovarian cancer could increase the power of a case–control study assessing variants in these genes.

A typical 'burden analysis' potentially also increases the power of a case–control study. By pooling variants that are probable to affect the function of the gene of interest and comparing the total number of these variants in cases and controls, associations can be found for a gene in which individual variants are too rare for risk assessment. This type of analysis has been used successfully in the case of most genes with moderate risk alleles

found through candidate gene studies.[24,26,27] However, predicting the functional effect of a variant is not straightforward. Although a number of in silico tools exist that can predict the effect of a variant, these prediction tools might misclassify some missense variants.[71] Erroneous inclusion of neutral variants will dilute the effect of truly pathogenic variants and thus decrease statistical power. In addition, allelic risk heterogeneity, as already set forth above for *BRCA1* and *BRCA2*, complicates the interpretation of the risk estimated from burden analysis, as this risk might not apply to all pooled variants. An extreme example has also been documented for *ATM*, in which pooled truncating variants are associated with a moderate risk ratio of around 2.3, but in which a single missense variant thought to have a dominant-negative effect on ATM function, p.V2424G, has been found associated with a more than 10-fold increased risk.[72,73]

Functional analysis of genetic variants can be very helpful to select variants with functional effect that can be grouped in a burden analysis. For BRCA1 and BRCA2 many such functional assays exist.[74] Also for variants in other breast cancer genes results from functional assays have been reported.[75–79] However, functional data alone are never sufficient to determine if a variant increases breast cancer risk. It is important to keep in mind that most breast cancer associated genes have many cellular functions and not all of these might be equally important for breast cancer risk. In addition, functional assays might result in variants that do not clearly cluster with either neutral or pathogenic variants. The clinical relevance of such variants is hard to determine.

On the basis of the classical two-hit model,[80,81] many studies assess loss of the wild type allele in the tumor when examining a potential breast cancer risk allele. Although loss of the wild type allele seems to occur in most (but not all) tumors in *BRCA1/2* mutation carriers,[82–85] for other genes loss of the wild type allele is less frequently seen. For example, Goldgar et al.[73] showed that only 1 of 18 breast tumors of carriers of an *ATM* mutation, showed loss of the wild type allele, while 6 showed loss of the mutant allele. Other studies confirm that loss of the wild type allele seems not necessary for carcinogenesis in *ATM* mutation carriers.[72,86] Although there is one *ATM* variant suggested to have a dominant-negative effect,[72] other variants might be haploinsufficient. In general, if the wild type allele is lost more often than the mutant allele, this can be regarded as evidence for pathogenicity, but the lack of it provides little evidence against pathogenicity.

Another way of assessing risk is to analyze cosegregation of the variant with breast cancer in the family where it was found,[87] which has been applied to very rare variants detected in *BRIP1* and *PALB2*.[26,27] However, unless multiple families with the same variant can be analyzed in this way, or the variant shows (near-)perfect co-segregation in a very large family with many cases of breast cancer, this approach is unlikely to provide accurate quantitative estimates of the risk conferred by the variant. Nonetheless, family-based analyses, despite the practical difficulties surrounding the sampling of family-members of the proband, remain attractive because of their better statistical power over case–control analysis in the general population, and because of their immediate relevance for clinical genetic counseling purposes. Here too, international collaboration will be pivotal to enable convincing evidence to be compiled. Specific gene variant databases exist for most genes in which variants have been found to be associated with breast cancer risk. By collating variant data from all over the world, the number of variants for which enough data is available to perform co-segregation analysis will increase. For *BRCA1* and *BRCA2* the ENIGMA consortium is dedicated to classifying variants of uncertain clinical significance.[88] Similar collaborations will be useful in the classification of variants in other breast cancer associated genes.

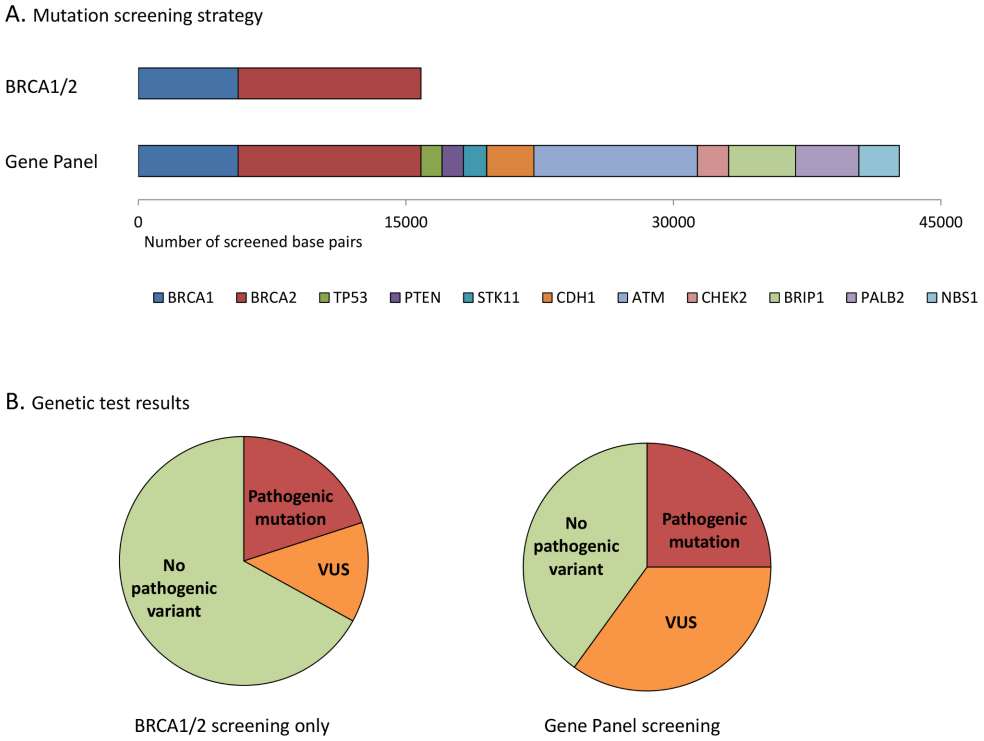


Figure 2 Test results from different mutation screening strategies in the clinic
This figure shows the correlation between mutation screening strategy and the distribution of test results. A shows the screened genes for the different mutation screening strategies and the corresponding number of screened coding base pairs. B shows the distribution of test results for BRCA1/2 mutation screening based on Frank et al. (93). The number of VUSs for the gene panel mutation screening was calculated under the assumption that the rate of VUS/ base pair for the additional genes would be similar to that of BRCA1 and BRCA2. Note that with the gene panel approach the expected increase of pathogenic mutations will be relatively small compared to the strong expected increase in the number of individuals with a VUS.

How will NGS impact clinical genetic counselling for breast cancer?

For clinical genetic services, NGS offers the possibility to screen additional risk loci at minimal additional costs. Therefore, many centers now consider the use of gene panels containing all known breast cancer associated loci for mutation screening. However, detection of variants in these additional risk loci introduces new challenges.[89] For many very rare variants, or combinations of variants, much uncertainty about the associated risk exists. Therefore, mutational screening of these genes should initially take place in a research setting until these can be established with at least some accuracy. A sobering fact is also that even for established high risk genes such as *BRCA1* and *BRCA2*, a variant of uncertain clinical significance (VUS) is uncovered in about 13% of new families tested,[90] which comprise approximately 40% of families testing 'positive'. When additional genes are included in the test panel for familial breast cancer, the percentage of families with a VUS in at least one of these genes will increase strongly (Fig. 2). Even though most methods discussed above can be used to assess the risk associated with these VUS, the lack of an epidemiological support, such as for *BRCA1/2*, will strongly limit the clinical utility of these test-results.

At some point in the not too distant future, DNA diagnostic laboratories might also start to use whole exome or genome analyses. This is expected to result in large numbers of 'hard to interpret' variants. As stipulated above, many potential explanations for the missing heritability of breast cancer still exist, and until we know more about these, it will remain extremely difficult to make clinical inferences on the basis of very rare variant data. On the bright side, however, NGS offers an unprecedented opportunity to get more insight into this genetic architecture. Much of this will initially remain within the realm of research, rather than clinical application, and will strongly depend on international collaboration and data collation in the public domain. Functional laboratory assays should complement genetic epidemiological data, as it did for *BRCA1* and *BRCA2*. Thus clinical genetic testing using NGS could greatly assist the advancement of knowledge on the genetic complexity underlying familial breast cancer. However, genetic counsellors should prepare for a situation in which very few of the detected very rare variants in genes other than the high risk genes we know today are 'actionable' in terms of disease management.

Acknowledgements
The authors were supported by the Dutch Cancer Society (grant UL 2009-4388)

References

1. Lalloo F, Evans DG. Familial breast cancer. Clin Genet. 2012;82(2):105-114. doi:10.1111/j.1399- 0004.2012.01859.x
2. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer. 2000;83(10):1301-1308. doi:10.1054/bjoc.2000.1407
3. Easton DF. How many more breast cancer predisposition genes are there? Breast Cancer Res. 1999;1(1):14-17.
4. Peto J, Collins N, Barfoot R, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst. 1999;91(11):943-949. doi:10.1093/jnci/91.11.943
5. NHGRI: Breast Cancer Information Core. https://research.nhgri.nih.gov/projects/bic/index.shtml. Accessed September 14, 2019.
6. Thompson D, Easton D. Variation in BRCA1 cancer risks by mutation position. Cancer Epidemiol Biomarkers Prev. 2002;11(4):329-336.
7. Thompson D, Easton D. Variation in cancer risks, by mutation position, in BRCA2 mutation carriers. Am J Hum Genet. 2001;68(2):410-419. doi:10.1086/318181
8. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353-361, 361e1-2. doi:10.1038/ng.2563
9. Spurdle AB, Whiley PJ, Thompson B, et al. BRCA1 R1699Q variant displaying ambiguous functional abrogation confers intermediate breast and ovarian cancer risk. J Med Genet. 2012;49(8):525-532. doi:10.1136/jmedgenet-2012-101037
10. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233-1238.
11. Nelen MR, van Staveren WC, Peeters EA, et al. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Hum Mol Genet. 1997;6(8):1383-1387.
12. Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature. 1998;391(6663):184-187. doi:10.1038/34432
13. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 1998;18(1):38-43. doi:10.1038/ng0198-38
14. Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402-405. doi:10.1038/32918
15. Keller G, Vogelsang H, Becker I, et al. Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. Am J Pathol. 1999;155(2):337-342. doi:10.1016/S0002-9440(10)65129-2
16. Pharoah PD, Guilford P, Caldas C, International Gastric Cancer Linkage Consortium. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 2001;121(6):1348-1353.
17. Ginsburg OM, Akbari MR, Aziz Z, et al. The prevalence of germ-line TP53 mutations in women diagnosed with breast cancer before age 30. Fam Cancer. 2009;8(4):563-567. doi:10.1007/s10689-009-9287-z
18. Rapakko K, Allinen M, Syrjäkoski K, et al. Germline TP53 alterations in Finnish breast cancer families are rare and occur at conserved mutation-prone sites. Br J Cancer. 2001;84(1):116-119. doi:10.1054/bjoc.2000.1530
19. FitzGerald MG, Marsh DJ, Wahrer D, et al. Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. Oncogene. 1998;17(6):727-731. doi:10.1038/sj.onc.1201984
20. Guenard F, Pedneault CS-L, Ouellette G, Labrie Y, Simard J, Durocher F. Evaluation of the contribution of the three breast cancer susceptibility genes CHEK2, STK11, and PALB2 in non-BRCA1/2 French Canadian families with high risk of breast cancer. Genet Test Mol Biomarkers. 2010;14(4):515-526. doi:10.1089/gtmb.2010.0027

21. Schrader KA, Masciari S, Boyd N, et al. Germline mutations in CDH1 are infrequent in women with early-onset or familial lobular breast cancers. J Med Genet. 2011;48(1):64-68. doi:10.1136/jmg.2010.079814
22. Swift M, Morrell D, Massey RB, Chase CL. Incidence of cancer in 161 families affected by ataxia-telangiectasia. N Engl J Med. 1991;325(26):1831-1836. doi:10.1056/NEJM199112263252602
23. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med. 2010;362(20):1909-1919. doi:10.1056/NEJMra0809889
24. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38(8):873-875. doi:10.1038/ng1837
25. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-59. doi:10.1038/ng879
26. 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. 2006;38(11):1239-1241. doi:10.1038/ng1902
27. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 2007;39(2):165-167. doi:10.1038/ng1959
28. Steffen J, Nowakowska D, Niwińska A, et al. Germline mutations 657del5 of the NBS1 gene contribute significantly to the incidence of breast cancer in Central Poland. Int J Cancer. 2006;119(2):472-475. doi:10.1002/ijc.21853
29. Solyom S, Aressy B, Pylkäs K, et al. Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions. Sci Transl Med. 2012;4(122):122ra23. doi:10.1126/scitranslmed.3003223
30. De Brakeleer S, De Grève J, Loris R, et al. Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. Hum Mutat. 2010;31(3):E1175-1185. doi:10.1002/humu.21200
31. Ratajska M, Antoszewska E, Piskorz A, et al. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. Breast Cancer Res Treat. 2012;131(1):89-97. doi:10.1007/s10549-011-1403-8
32. Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010;42(5):410-414. doi:10.1038/ng.569
33. 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. 2008;2(4):296-316. doi:10.1016/j.molonc.2008.09.007
34. Heikkinen K, Rapakko K, Karppinen S-M, et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis. 2006;27(8):1593-1599. doi:10.1093/carcin/bgi360
35. Clague J, Wilhoite G, Adamson A, Bailis A, Weitzel JN, Neuhausen SL. RAD51C germline mutations in breast and ovarian cancer cases from high-risk families. PLoS One. 2011;6(9):e25632. doi:10.1371/journal.pone.0025632
36. De Leeneer K, Van Bockstal M, De Brouwer S, et al. Evaluation of RAD51C as cancer susceptibility gene in a large breast-ovarian cancer patient population referred for genetic testing. Breast Cancer Res Treat. 2012;133(1):393-398. doi:10.1007/s10549-012-1998-4
37. Gorringe KL, Choong DYH, Visvader JE, Lindeman GJ, Campbell IG. BARD1 variants are not associated with breast cancer risk in Australian familial breast cancer. Breast Cancer Res Treat. 2008;111(3):505-509. doi:10.1007/s10549-007-9799-x
38. Thompson ER, Boyle SE, Johnson J, et al. Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients. Hum Mutat. 2012;33(1):95-99. doi:10.1002/humu.21625
39. Zheng Y, Zhang J, Hope K, Niu Q, Huo D, Olopade OI. Screening RAD51C nucleotide

alterations in patients with a family history of breast and ovarian cancer. Breast Cancer Res Treat. 2010;124(3):857-861. doi:10.1007/s10549-010-1095-5
40. Uhrhammer N, Delort L, Bignon Y-J. Rad50 c.687delT does not contribute significantly to familial breast cancer in a French population. Cancer Epidemiol Biomarkers Prev. 2009;18(2):684-685. doi:10.1158/1055-9965.EPI-08-0971
41. Tommiska J, Seal S, Renwick A, et al. Evaluation of RAD50 in familial breast cancer predisposition. Int J Cancer. 2006;118(11):2911-2916. doi:10.1002/ijc.21738
42. Loveday C, Turnbull C, Ruark E, et al. Germline RAD51C mutations confer susceptibility to ovarian cancer. Nat Genet. 2012;44(5):475-476; author reply 476. doi:10.1038/ng.2224
43. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol. 2001;21(1):1-18. doi:10.1002/gepi.1014
44. Antoniou AC, Pharoah PDP, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer. 2002;86(1):76-83. doi:10.1038/sj.bjc.6600008
45. Cui J, Antoniou AC, Dite GS, et al. After BRCA1 and BRCA2-what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. Am J Hum Genet. 2001;68(2):420-431. doi:10.1086/318187
46. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science. 2012;336(6082):740-743. doi:10.1126/science.1217283
47. Nickels S, Truong T, Hein R, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet. 2013;9(3):e1003284. doi:10.1371/journal.pgen.1003284
48. Milne RL, Antoniou AC. Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers. Ann Oncol. 2011;22 Suppl 1:i11-17. doi:10.1093/annonc/mdq660
49. Gracia-Aznarez FJ, Fernandez V, Pita G, et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. PLoS One. 2013;8(2):e55681. doi:10.1371/journal.pone.0055681
50. Hilbers FS, Meijers CM, Laros JFJ, et al. Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected on the basis of aCGH tumor profiling. PLoS ONE. 2013;8(1):e55734. doi:10.1371/journal.pone.0055734
51. Park DJ, Odefrey FA, Hammet F, et al. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. Breast Cancer Res Treat. 2011;130(3):1043-1049. doi:10.1007/s10549-011-1704-y
52. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
53. Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. PLoS Genet. 2012;8(9):e1002894. doi:10.1371/journal.pgen.1002894
54. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. Science. 2002;297(5581):606-609. doi:10.1126/science.1073834
55. Berwick M, Satagopan JM, Ben-Porat L, et al. Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. Cancer Res. 2007;67(19):9591-9596. doi:10.1158/0008-5472.CAN-07-1501
56. Garcia MJ, Fernandez V, Osorio A, et al. Mutational analysis of FANCL, FANCM and the recently identified FANCI suggests that among the 13 known Fanconi Anemia genes, only FANCD1/BRCA2 plays a major role in high-risk breast cancer predisposition. Carcinogenesis. 2009;30(11):1898-1902. doi:10.1093/carcin/bgp218
57. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. J Med Genet. 2012;49(3):184-186. doi:10.1136/jmedgenet-2011-100585

58. Hilbers FS, Wijnen JT, Hoogerbrugge N, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. J Med Genet. 2012;49(10):618-620. doi:10.1136/jmedgenet-2012-101191
59. Rodriguez-Lopez R, Osorio A, Sanchez-Pulido L, et al. No mutations in the XRCC2 gene in BRCA1/2-negative high-risk breast cancer families. Int J Cancer. 2003;103(1):136-137. doi:10.1002/ijc.10766
60. Ellis NA, Groden J, Ye TZ, et al. The Bloom's syndrome gene product is homologous to RecQ helicases. Cell. 1995;83(4):655-666. doi:10.1016/0092-8674(95)90105-1
61. Prokofyeva D, Bogdanova N, Dubrowinskaja N, et al. Nonsense mutation p.Q548X in BLM, the gene mutated in Bloom's syndrome, is associated with breast cancer in Slavic populations. Breast Cancer Res Treat. 2013;137(2):533-539. doi:10.1007/s10549-012-2357-1
62. Sokolenko AP, Iyevleva AG, Preobrazhenskaya EV, et al. High prevalence and breast cancer predisposing role of the BLM c.1642 C>T (Q548X) mutation in Russia. Int J Cancer. 2012;130(12):2867-2873. doi:10.1002/ijc.26342
63. Cleary SP, Zhang W, Di Nicola N, et al. Heterozygosity for the BLM(Ash) mutation and cancer risk. Cancer Res. 2003;63(8):1769-1771.
64. Baris HN, Kedar I, Halpern GJ, et al. Prevalence of breast and colorectal cancer in Ashkenazi Jewish carriers of Fanconi anemia and Bloom syndrome. Isr Med Assoc J. 2007;9(12):847-850.
65. Ruark E, Snape K, Humburg P, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature. 2013;493(7432):406-410. doi:10.1038/nature11725
66. Southey MC, Park DJ, Nguyen-Dumont T, et al. COMPLEXO: identifying the missing heritability of breast cancer via next generation collaboration. Breast Cancer Res. 2013;15(3):402. doi:10.1186/bcr3434
67. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003;19(1):149-150. doi:10.1093/bioinformatics/19.1.149
68. Zhang S, Royer R, Li S, et al. Frequencies of BRCA1 and BRCA2 mutations among 1,342 unselected patients with invasive ovarian cancer. Gynecol Oncol. 2011;121(2):353-357. doi:10.1016/j.ygyno.2011.01.020
69. Pal T, Permuth-Wey J, Betts JA, et al. BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. Cancer. 2005;104(12):2807-2816. doi:10.1002/cncr.21536
70. Rafnar T, Gudbjartsson DF, Sulem P, et al. Mutations in BRIP1 confer high risk of ovarian cancer. Nat Genet. 2011;43(11):1104-1107. doi:10.1038/ng.955
71. Tchernitchko D, Goossens M, Wajcman H. In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics. Clin Chem. 2004;50(11):1974-1978. doi:10.1373/clinchem.2004.036053
72. Chenevix-Trench G, Spurdle AB, Gatei M, et al. Dominant negative ATM mutations in breast cancer families. J Natl Cancer Inst. 2002;94(3):205-215. doi:10.1093/jnci/94.3.205
73. Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res. 2011;13(4):R73. doi:10.1186/bcr2919
74. Carvalho MA, Couch FJ, Monteiro ANA. Functional assays for BRCA1 and BRCA2. Int J Biochem Cell Biol. 2007;39(2):298-310. doi:10.1016/j.biocel.2006.08.002
75. Osorio A, Endt D, Fernandez F, et al. Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families. Hum Mol Genet. 2012;21(13):2889-2898. doi:10.1093/hmg/dds115
76. Roeb W, Higgins J, King M-C. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet. 2012;21(12):2738-2744. doi:10.1093/hmg/dds101
77. Bell DW, Kim SH, Godwin AK, et al. Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. Int J Cancer. 2007;121(12):2661-2667. doi:10.1002/ijc.23026

78. Shaag A, Walsh T, Renbaum P, et al. Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. Hum Mol Genet. 2005;14(4):555-563. doi:10.1093/hmg/ddi052
79. Scott SP, Bendix R, Chen P, Clark R, Dork T, Lavin MF. Missense mutations but not allelic variants alter the function of ATM by dominant interference in patients with breast cancer. Proc Natl Acad Sci U S A. 2002;99(2):925-930. doi:10.1073/pnas.012329699
80. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA. 1971;68(4):820-823.
81. Ponder B. Cancer. Gene losses in human tumours. Nature. 1988;335(6189):400-402. doi:10.1038/335400a0
82. Collins N, McManus R, Wooster R, et al. Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12-13. Oncogene. 1995;10(8):1673-1675.
83. Gudmundsson J, Johannesdottir G, Bergthorsson JT, et al. Different tumor types from BRCA2 carriers show wild-type chromosome deletions on 13q12-q13. Cancer Res. 1995;55(21):4830-4832.
84. Osorio A, de la Hoya M, Rodriguez-Lopez R, et al. Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. Int J Cancer. 2002;99(2):305-309. doi:10.1002/ijc.10337
85. Tung N, Miron A, Schnitt SJ, et al. Prevalence and predictors of loss of wild type BRCA1 in estrogen receptor positive and negative BRCA1-associated breast cancers. Breast Cancer Res. 2010;12(6):R95. doi:10.1186/bcr2776
86. Broeks A, Urbanus JH, Floore AN, et al. ATM-heterozygous germline mutations contribute to breast cancer-susceptibility. Am J Hum Genet. 2000;66(2):494-500. doi:10.1086/302746
87. Mohammadi L, Vreeswijk MP, Oldenburg R, et al. A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example. BMC Cancer. 2009;9:211. doi:10.1186/1471-2407-9-211
88. Spurdle AB, Healey S, Devereau A, et al. ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat. 2012;33(1):2-7. doi:10.1002/humu.21628
89. Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME. Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? J Clin Oncol. 2013;31(10):1267-1270. doi:10.1200/JCO.2012.46.9403
90. Frank TS, Deffenbaugh AM, Reid JE, et al. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. J Clin Oncol. 2002;20(6):1480-1490. doi:10.1200/JCO.2002.20.6.1480

Chapter 7

Summary and Discussion

Introduction

A family history of breast cancer is one of the most important risk factors for the disease.[1] Over the last decades many genetic loci associated with breast cancer risk have been discovered. In spite of this, only approximately half of the familial relative risk (FRR) for breast cancer can be explained by the currently known genetic risk factors.[2–4] In families where no genetic explanation has been found for the clustering of breast cancer, uncertainty remains about who is at increased risk and to which extend this risk is increased. This hampers decisions on screening strategies and preventive measures. Next generation sequencing (NGS) offers new possibilities to explore the genetic etiology of unexplained familial clustering of breast cancer. It allows for the detection of genetic variants regardless of their frequency in the general population and without the need of a prior hypothesis about which genes or genomic regions might be involved in breast cancer susceptibility. The aim of this thesis was to get a better understanding of the genetic etiology of non-*BRCA1/2* familial breast cancer with the help of NGS.

The selection of a homogeneous phenotype: a failed strategy?

The average exome of an individual from European descent has approximately 12,000 non-synonymous genetic variants.[5] However, at the time the studies described in this thesis were conducted, the cost of NGS only allowed us to sequence the exomes of a relatively small number of familial cases. Therefore, further association analysis using different techniques was necessary to follow up on potentially interesting genetic variants. However, also the number of familial cases available in downstream case-control analyses was limited. Thus, it was crucial to carefully select and strongly reduce the candidate variants for follow-up analysis. Selection based on the predicted effect of a genetic variant on protein function is often not able to sufficiently reduce the number of potentially interesting variants, while selection based on the pathways in which a gene is involved, strongly depend on our assumptions on which pathways play a role in breast cancer. Similar to classical association studies, an exome sequencing effort would ideally find multiple families with a mutation in the same gene while these mutations are absent or extremely rare in the general population. However, as the genetic etiology of breast cancer is already known to be very heterogeneous, a large number of familial cases would need to be sequenced in order to find two of them with a mutation in the same gene. Interestingly, mutations in *BRCA1* are strongly associated with a number of tumor characteristics. Notably, tumors of *BRCA1* mutation carriers are strongly enriched for the "triple negative" (lacking the receptors ER, PR and HER2) immunohistochemistry phenotype and a basal-like expression profile.[6] Based on this, we hypothesized that by selecting non-*BRCA1/2* breast cancer patients or families that share a certain phenotype, we would also select for a more homogenous genetic etiology and increase our chances of finding multiple cases with mutations in the same gene.

In **Chapter 2** of this thesis we selected six non-*BRCA1/2* families in which the majority of tumors show a specific, previously identified array comparative genome hybridization (CGH) profile.[7] Subsequent linkage analysis in these families showed a peak with a LOD score of 2.49 on chromosome 4, which suggested that the clustering of breast cancer in these families might be caused by mutations in a gene in this linkage region. Therefore, whole-exome sequencing was performed on two individuals per family. However, no genes with a likely pathogenic variant in more than one family were found. Not on chromosome 4, nor

elsewhere in the genome. Similarly, **Chapter 5** describes an exome sequencing study in which we focused on families with a possible recessive mode of inheritance. For this study we selected 19 non-*BRCA1/2* breast cancer families in which at least three siblings were affected, while no first-degree relatives in the previous or following generation had breast cancer. The germline DNA from one of the siblings was subjected to exome sequencing, while all affected siblings were genotyped using a SNP arrays in order to assess haplotype sharing. This allowed us to focus on the exome sequencing variants in the regions where all affected siblings shared two haplotypes. However, also this exome sequencing study did not yield any potential novel susceptibility genes. It is possible that in these two studies we have missed high-risk susceptibility alleles that were in fact present in the sequenced individuals. We might have discarded a variant as it seemed unlikely to affect protein function or we might not have detected it at all as it resides outside the protein-coding regions. Moreover, the results of these studies do not exclude that there are additional high-risk breast cancer susceptibility genes that are strongly associated with a specific phenotype. We might simply have selected the wrong tumor and family characteristics. As conventionally a LOD score greater than 3.0 is considered evidence for linkage, our LOD score of 2.49 in **Chapter 2** might have been spurious. And, while our selection of families with at least three affected siblings in **Chapter 5** had important advantages for the variant filtering, other selection criteria, such as very early onset cases, might have been more likely to enrich for recessive susceptibility alleles. Moreover, as the families we selected had a large number of breast cancer cases in one generation, it is not impossible that these families are in fact explained by a dominant allele.

Biologically, there are several ways in which an inherited genetic variant can be associated with a tumor phenotype. Cancer susceptibility genes are typically thought to be tumor suppressors, which require the loss of both copies for malignant transformation. However, on a cellular level the loss of one copy of a gene can already have effects on downstream signalling, gene expression and cellular functions. This is called haploinsufficiency. For example, lymphoblastic cell lines derived from carrier of a heterozygous deleterious *PALB2* mutation show aberrant DNA replication and a shift to error-prone DNA repair mechanisms.[8,9] This might result in characteristics that most tumors associated with a specific susceptibility gene have in common, for example, altered expression of genes controlled by the pathway in which the gene with the inherited mutation functions or altered phosphorylation of proteins in this same pathway. However, due to the large number of genetic and epigenetic changes a tumor cell acquires during tumorigenesis, characteristics associated with an inherited mutation might be partly masked and become difficult to detect. Association between inherited mutations and tumor phenotype can also occur indirectly due to synergy with other genetic, epigenetic or microenvironmental changes, which are subsequently selected for because of increased fitness. One of the best-studied examples is so-called loss of heterozygosity (LOH). As mentioned above, while a cell is thought to be relatively unaffected by the loss of one copy of a tumor suppressor gene, loss of the second copy will have a much more dramatic effect and contribute significantly to tumorigenesis. Somatic loss of one copy of a susceptibility gene therefore will be selected for in the context of an inherited pathogenic mutation, but not in the absence of such a mutation. This phenomenon is frequently observed in high-risk breast cancer susceptibility genes *BRCA1*, *BRCA2*[10] and *PALB2*.[11] LOH at these loci is very high in tumors of gene carriers, but much lower in sporadic cases. In genes associated with a more moderate increase in risk of breast cancer this association is less clear: while loss of the wild-type allele is frequently observed in the breast tumors of *ATM* mutation carriers,[12,13] tumors of *CHEK2* mutation carriers show no strong enrichment for the loss of the wild-type *CHEK2*

allele.[14–16] More complex relationships between germline mutations and tumor characteristics have also been described. For example, tumors of *BRCA1* and *BRCA2* mutation carriers have been found to be associated with two specific mutation signatures and two rearrangement signatures, which are thought to be the "genomic scar" shaped by the absence of *BRCA1* or *BRCA2* function and the resulting DNA repair deficiencies.[17] Interestingly, this also shows that mutations in two different genes can result in a similar phenotype. Lastly, it has been hypothesized that part of the phenotypical heterogeneity of breast cancer stems from the existence of two different of cell types of origin, myoepithelial and luminal cells.[18] In a full-grown tumor, epigenetic features and gene expression patterns, might still be traced back to this cell of origin. Therefore, if a genetic risk factor more strongly predisposes to cancer in one of these two cell types, these epigenetic and gene expression features would also be associated with this genetic risk factor.

Although the two studies described in **Chapters 2** and **5** of this thesis have been unsuccessful in discovering novel breast cancer risk alleles, it might be too early to completely dismiss the strategy of selecting a more homogeneous group of familial breast cancer patients when aiming to find new breast cancer risk alleles. Several of the known breast cancer susceptibility genes are associated with a specific phenotype (see **Chapter 1** of this thesis), although, apart from a few of the cancer syndromes, these phenotypes were typically discovered after the association of breast cancer with a specific gene or genomic region had been detected. The CGH profile as used in **Chapter 2** of this thesis provides a relatively low-resolution picture of the tumor genome. Over the past years, technical advances and decreasing sequencing costs have provided new opportunities to assess tumor characteristic and therefore to potentially select tumors with a more homogeneous etiology. Most importantly, it has become possible to apply massive parallel sequencing on DNA and RNA from formalin fixed paraffin embedded (FFPE) material. In addition, copy number aberrations can now be characterized more precisely using SNPs array-based techniques. The molecular tumor characteristics described in the introduction of this thesis, e.g. based on mutations, copy number variations, "intrinsic" gene expression-based subtypes and mutational signatures, could potentially be used to select for a more homogeneous population of familial breast cancer cases. Currently, little or no data exists on whether these molecular features cluster within families. Exploring this would be a first step to decide if it is worthwhile further pursuing the "homogeneous phenotype strategy".

The challenge of establishing the risk associated with extremely rare variants

As discussed above, arguably the most difficult and laborious step in the analysis of exome sequencing data is the variant filtering. If no genes are identified in which multiple families carry a variant that is likely to affect protein function, other strategies are needed to select those variants most likely to be associated with breast cancer risk for downstream validation. **Chapter 6** of this thesis reviews several approaches. As a first step, genetic variants are often filtered based on minor allele frequency in one of the many available reference data sets of healthy individuals. This is rationalized by the prevalence of breast cancer in the general population, which must be consistent with the presumed risk associated with the variant, i.e. a high-risk variant cannot be too common, otherwise breast cancer would be more prevalent than it is. Although somewhat arbitrary, a cut-off of 0.1% allele frequency is often used.

However, it is important to keep in mind that there are several examples of founder mutations, such as *CHEK2**1100delC, that are associated with a moderately or strongly increased breast cancer risk and have an allele frequency larger than 0.1%. Moreover, due to recent explosive human population growth, most variants are rare regardless of association with disease.[19] For example, within the ExAC dataset containing the exome data of more than 60,000 individuals, approximately 99% of detected high quality variants had an allele frequency of less than 1%.[5] Hence, selecting for variants occurring <0.1% in reference data sets will not dismiss a large proportion of candidates. A next obvious step is to focus on protein truncating variants as they are almost certain to affect protein function. However, these variants make up only a small minority of the detected variants. Further filtering can be done using in silico prediction algorithms. These tools use information such as evolutionary conservation, known functional domains and three-dimensional structure, to estimate the likelihood of a missense variant to affect protein function. Unfortunately, the sensitivity and specificity of these tools is known to be far from optimal.[20,21] However, the limited number of alternative filtering options besides in silico prediction algorithms make that these algorithms are frequently used. Lastly, variants are often filtered based on the function of the gene in which the variant is found, where genes with roles cancer-related processes such as cell proliferation and DNA repair are prioritized. This filter depends heavily on our knowledge of the pathways involved in carcinogenesis. In addition, one could argue that if we are only considering variants in genes related to carcinogenesis, a sequencing of a cancer specific gene panel might be a better approach than exome sequencing.

Follow-up association analysis can be done on the variant level, i.e. by genotyping a set of cases and controls for that specific variant only. However, as explained above many variants will be very rare and even large case control sets will often lack the power to detect a significant association. An alternative approach to a variant-level analysis is a so-called burden analysis, where in every case and control the whole coding region of the respective gene is sequenced. In this case, the association analysis is not done based on individual variants, but rather compares the total number of likely damaging variants between cases and controls. This requires a decision on which variants to consider as (likely) damaging. Some of the same filters as discussed above, such as allele frequency and in silico prediction can be used. However, as the number of variants to be assessed is now considerably smaller, additional options are available. These include co-segregation analysis in families with a variant, assessment of loss of heterozygosity in the tumors of carriers, and functional assays. It is worth investing time and effort in the selection of variants, as both the inclusion of benign variants and the exclusion of variants that truly affect the protein of interest reduce the power of the association analysis.

Chapter 3 and **4** of this thesis give a good example of the difficulties associated with establishing the risk associated with very rare variants, in this case in the gene *XRCC2*. The possible association between variants in *XRCC2* and familial breast cancer was first reported by a research group at the university of Melbourne, Australia. They had found one protein-truncating variant in the exome of a familial breast cancer patient and, based on the function of XRCC2 in DNA repair, had decided to further explore the possibility of this being a breast cancer susceptibility allele. Subsequently, they requested access to the exome sequencing data from several other groups, among which the data from our study reported in **Chapter 2** of this thesis. This pooled analysis of exomes and a subsequent case-control study provided a suggestion that variants in *XRCC2* might indeed be associated with breast cancer risk and the results were subsequently published.[22] **Chapter 3** of this thesis reports the data of a large

international case-control study aiming to validate the results of this initial publication. This study applied a burden analysis strategy, classifying variants based on in silico prediction algorithms (Polyphen-2[23], SIFT[24] and AlignGVGD[25]). Regardless of the prediction algorithm used, this study did not find an association with breast cancer. However, as prediction algorithms are known to be imperfect, we decided to further explore this result using functional assays to assess the effect of the genetic variants on the XRCC2 protein and the DNA repair pathways in which it functions. Chapter 4 of this thesis reports the results of this effort. It showed that, based on a RAD51 foci formation assay and two reporter constructs, the SCR reporter and the DR-GFP reporter, only a limited number of variants actually affected protein function. When only taking into account these variants, again no association with breast cancer risk was found, although an association could not be ruled out for those variants which strongly affect XRCC2 function, due to the low number of variants in this group.

This example underlines the difficulties with variant selection and the issues related to very rare genetic variants. A big challenge for the future will be to conduct exome sequencing studies with sufficient sample size to at least allow for gene-level association analyses. Recently, an exome sequencing study in over 20,000 cases with type 2 diabetes showed that it was able to identify four susceptibility genes at exome-wide significance based on a rare variant (<0.5% minor allele frequency) burden analysis.[26] Of note, the effective power of this study was increased by selecting an ethnically diverse population, thereby sampling a broader range of haplotypes. In the near future, similar efforts for breast cancer susceptibility will likely be pursued in the context of existing international collaborations, such as the Breast Cancer Association Consortium (BCAC). However, in order to solve variant-level associations for very rare variants, case-control studies with the currently available sample sizes will likely not suffice. A possible alternative method of classifying rare variants of uncertain significance (VUS) is by way of co-segregation analysis. Co-segregation analysis assesses the association between a genetic variant and a disease by quantifying, based on a pedigree and the breast cancer cases occurring within it, the extent to which a genetic variant co-occurs with disease more often than expected. However, co-segregation analysis requires extensive DNA sampling within families carrying a VUS. It will therefore be crucial to invest in the collection of such DNA samples.

The updated landscape of breast cancer susceptibility

Before embarking on further efforts to discover novel breast cancer risk alleles, it is important to reflect on what we have learned about the landscape of genetic susceptibility over the last few years. To tailor our efforts, we need to understand how likely it is that further high-risk alleles explain a considerable proportion of the currently unexplained familial clustering of breast cancer.

Historically, research into the genetic etiology of breast cancer mainly focused on the discovery of high-risk genes. Naturally, families with very strong clustering of breast cancer are a logical starting point for the discovery of genetic risk factors. Moreover, for a long time, research on the genetic etiology of breast cancer was limited by the fact that the sequence of most the human genome was unknown, and no technologies existed that allowed for the analysis of large genomic regions. Therefore, the region of interest first needed to be narrowed using linkage analysis with low-density microsatellite genotyping, after which the regions significantly more often shared by the affected individuals in a set of families could be further explored. As linkage analysis is only able to find regions in which alleles with a

relatively high penetrance are located, this limited the scope of early research into the genetic etiology of breast cancer. When in 2001 the first draft of the human genome was released, it was accompanied by a manuscript describing the construction of a map of 1.4 million SNPs in the human genome, providing for the first time sufficient density to study human haplotype structure and allowing for subsequent genome-wide studies assessing the association between common genetic variation and disease in the general population. From 2007 onward, several genome-wide association studies (GWAS) together have reported over 300

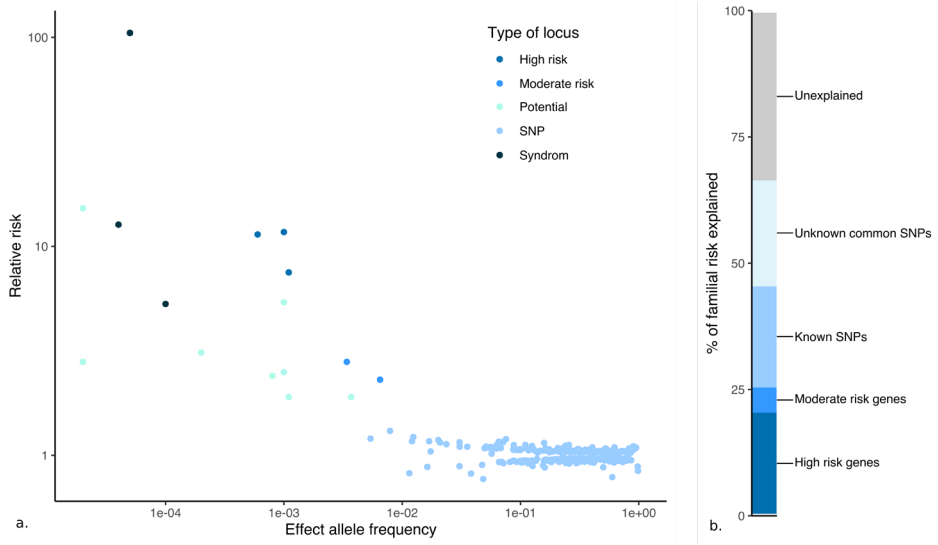


Figure 1. The current landscape of breast cancer susceptibility alleles
a. Relative risk and effect allele frequency for the currently known breast cancer susceptibility alleles. For underlying data please see supplementary table 1. b. percentage of familial risk explained by the currently known breast cancer susceptibility alleles. For references and underlying data see supplementary table 1.

The invention of NGS brought new hope for the discovery of novel high-risk susceptibility genes as it allows for cost-effective genome-wide detection of genetic variants in individual familial breast cancer cases. However, as outlined above, we have not been able to find any novel high-risk breast cancer alleles in the two exome sequencing studies described in this thesis, **Chapter 2** and **Chapter 5**. In **Chapter 5** of this thesis, besides exploring the potential role of recessive high-risk alleles, we genotyped the families in this study for all the known and suspected moderate and high-risk genes in addition to genotyping the 160 SNPs currently known to be associated with a small increase in breast cancer risk. This study found that the average normalized PRS of the familial cases was significantly higher than that in both general population cases and controls. Indicating that the low risk variants do contribute to familial clustering of breast cancer, although it is difficult to estimate to which extend due to the atypical breast cancer families represented in this study. Moreover, in several families we detected a moderate risk variant in *ATM* or *CHEK2*. In another study (not included in this thesis), we have reported that in a set of 101 unselected non-*BRCA1/2* breast cancer families, familial breast cancer cases have on average a higher PRS. Moreover, taking into account the PRS can change risk management recommendations in 10-20% of the women in these families depending on the guideline used.[29] Interestingly, also work from others has shown limited value of NGS for the discovery of novel high-risk genes. Although

over 30 exome sequencing studies have been published to date, the number of potential high-risk susceptibility genes identified is limited. Several genes, such as *KAT6B*,[30] *RINT1*,[31] *APOBEC3B*,[32] *XRCC2*[22] and *RCC1*,[33] have been suggested as novel susceptibility genes but external validation has either not been performed or resulted in conflicting results. Only two promising novel susceptibility genes coming out of exome sequencing studies have now been validated independently by several other studies: *FANCM*[34] and *RECQL*.[35,36] Remarkably, many exome sequencing studies report pathogenic variants in known moderate risk genes such as *ATM,*[37–42] *CHEK2*,[39,41,43–45] and *PALB2*,[37,39,41,45,46] suggesting that indeed a substantial proportion of familial cases might be explained by a combination of low and moderate risk susceptibility alleles. In fact, already shortly after the discovery of *BRCA1* and *BRCA2* as breast cancer susceptibility genes, several segregation analyses have suggested a that a polygenic model would best explain the remaining familial clustering of breast cancer.[47,48] This now seems to be confirmed by the results of the latest GWAS analysis in which it is estimated that that all, currently known and yet to be discovered, common low risk alleles explain together approximately 41% of the familial risk of breast cancer.[2] Figure 1 provides an overview of our current understanding of the genetic landscape of breast cancer. While moderate and high-risk alleles are thought to explain 5% and 20% respectively,[49] 41% is thought to be explained by low risk, common variants of which ~20% by the currently known low risk susceptibility loci.[2] The remaining 35% would then be explained by currently unknown factors such as, rare variants, interactions between risk factors, inherited epigenetic factors and environmental risk factors that are shared between family members.

Now that it has become clear that many genetic factors contribute to familial clustering of breast cancer and that individual factors are often associated with just a small increase in risk, there is a clear need to combine these factors into risk prediction models that are able to provide insights in an individual's risk of breast cancer. Several studies have aimed to combine the effects of low risk loci into a polygenic risk score (PRS). The latest of these, combines the effect of 313 SNPs.[28] Predating the GWAS era, there are also many models aiming to predict the risk of breast cancer based on non-genetic factors and high-risk mutations. Arguably the most extensive model to date is the BOADICEA model, which has very recently been updated to include the effect of the 313 currently known low risk loci. This model now uses truncating mutations in *BRCA1*, *BRCA2*, *ATM*, *CHEK2* and *PALB2*; 313 low risk loci; age at menarche; age at menopause; parity; age at first live birth; oral contraceptive (OC) use; hormone replacement therapy (HRT) use; height; BMI; alcohol intake; family history and a residual polygenic component to predict lifetime breast cancer risk.[50] In the UK population, this model would predict approximately 15% of women to have moderate lifetime risk of breast cancer (≥17% and <30% according to the NICE guidelines) and approximately 1% to have a high risk (>30%). This model has not yet been prospectively tested. Moreover, it does not take into account interactions between risk factors beyond the log-additive model, nor variant-specific risks in moderate and high risk genes, genetic variants in genes associated with cancer syndromes i.e. *TP53*, *CDH1* and *PTEN*, genetic variants in likely novel breast cancer susceptibility genes such as *FANCM* and *RECQL*, subtype specific effects, time varying variables for BMI, alcohol, OC and HRT use and the exact timing of pregnancies. A large prospective validation effort could be combined with an attempt to include these factors and improve the model, either through classic association analysis or deep learning. An advantage of the latter method is that it allows for continuous learning, making use of all available data, although the lack of formal statistics to express uncertainty is a disadvantage. The most optimal approach would probably be to prospectively calculate, for example for a large cohort of women in the context of population

screening, risk of breast cancer based on the current BOADICEA model in order to validate it and to simultaneously try to optimize the model using deep learning on prospectively collected data from these same women. The use of a model like BOADICEA would also make it possible to focus future research on (familial) cases that are unexplained by the currently known risk factors, e.g. familial or early onset cases with a very low predicted risk

Conclusions

To summarize, the work reported in this thesis has not been able to identify any novel high-risk breast cancer susceptibility alleles. Although there are likely still several extremely rare risk alleles to be discovered and the presence of high-risk alleles outside of protein-coding regions cannot be excluded, it seems presently unlikely that these will explain a substantial proportion of familial breast cancer. Both our work and that of others has suggested that most non-*BRCA1/2* familial breast cancer cases are likely explained by a combination of low-, and moderate-risk susceptibility alleles.

As expected, the largest challenge associated with the use of exome sequencing in the context of familial breast cancer has been the large number of genetic variants detected in a relatively small set of familial cases, which, with the sample sizes used to date, prohibits any formal association testing in the variant selection process. Therefore, until we are able to conduct exome sequencing studies with at least sufficient power to allow for exome-wide gene-level association analyses, the discovery of novel risk alleles in an assumption-free manner is still not a reality. The selection of a more homogeneous phenotype in hopes of selecting for a more homogeneous genetic etiology, has not resulted in any potential risk alleles being detected in more than one family. Although, with more advance techniques becoming available for the phenotyping of tumors, there might still be value in this approach for future attempts to discover novel high-risk alleles. Our experience attempting to validate rare genetic variants in *XRCC2* as breast cancer susceptibility alleles has served as a reminder of the limited value of in silico prediction algorithms, which can lead to misleading results of burden analyses and incorrect conclusions about a gene's role in breast cancer susceptibility. Although functional assays can give important insights, in many cases the time needed for the set-up and conduct of these assays, makes that they will likely only be used for strong candidate genes.

Taken together, these findings lead to the conclusion that if we want to be able to provide better risk prediction to the familial breast cancer cases who remain unexplained by susceptibility alleles currently tested in clinical practice, and to any woman for that matter, future efforts should focus on developing models that combine all currently known susceptibility alleles and take into account other risk factors. After initial validation of such a model, deep-learning techniques could be employed to continuously improve them based on real-world data. This prospective might mean that the dichotomy of sporadic and familial breast cancer with regard to genetic susceptibility disappears, which would require a change in perspective for both breast cancer susceptibility research and genetic counseling.

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 Lond. Engl. 358, 1389–1399 (2001).
2. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. Nature (2017) doi:10.1038/nature24284.
3. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br. J. Cancer 83, 1301–1308 (2000).
4. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353–361, 361e1-2 (2013).
5. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291 (2016).
6. Mavaddat, N. et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 21, 134–147 (2012).
7. Didraga, M. A. et al. A non-BRCA1/2 hereditary breast cancer sub-group defined by aCGH profiling of genetically related patients. Breast Cancer Res. Treat. 130, 425–436 (2011).
8. Nikkilä, J. et al. Heterozygous mutations in PALB2 cause DNA replication and damage response defects. Nat. Commun. 4, 2578 (2013).
9. Obermeier, K. et al. Heterozygous PALB2 c.1592delT mutation channels DNA double-strand break repair into error-prone pathways in breast cancer patients. Oncogene 35, 3796–3806 (2016).
10. Maxwell, K. N. et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat. Commun. 8, 319 (2017).
11. Lee, J. E. A. et al. Molecular analysis of PALB2-associated breast cancers. J. Pathol. 245, 53–60 (2018).
12. Weigelt, B. et al. The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J. Natl. Cancer Inst. 110, 1030–1034 (2018).
13. Renault, A.-L. et al. Morphology and genomic hallmarks of breast tumours developed by ATM deleterious variant carriers. Breast Cancer Res. BCR 20, 28 (2018).
14. Muranen, T. A. et al. Breast tumors from CHEK2 1100delC-mutation carriers: genomic landscape and clinical implications. Breast Cancer Res. BCR 13, R90 (2011).
15. Oldenburg, R. A. et al. The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. Cancer Res. 63, 8153–8157 (2003).
16. Massink, M. P. G., Kooi, I. E., Martens, J. W. M., Waisfisz, Q. & Meijers-Heijboer, H. Genomic profiling of CHEK2*1100delC-mutated breast carcinomas. BMC Cancer 15, 877 (2015).
17. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 534, 47–54 (2016).
18. Anderson, W. F., Rosenberg, P. S., Prat, A., Perou, C. M. & Sherman, M. E. How Many Etiological Subtypes of Breast Cancer: Two, Three, Four, Or More? JNCI J. Natl. Cancer Inst. 106, (2014).
19. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 336, 740–743 (2012).
20. Miosge, L. A. et al. Comparison of predicted and actual consequences of missense mutations. Proc. Natl. Acad. Sci. U. S. A. 112, E5189-5198 (2015).
21. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. 18, 225 (2017).
22. Park, D. J. et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am. J. Hum. Genet. 90, 734–739 (2012).
23. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. Chapter 7, Unit7.20 (2013).

24. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. Genome Res. 11, 863–874 (2001).
25. Tavtigian, S. V. et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J. Med. Genet. 43, 295–305 (2006).
26. Flannick, J. et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature 570, 71–76 (2019).
27. Easton, D. F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087–1093 (2007).
28. Mavaddat, N. et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am. J. Hum. Genet. 104, 21–34 (2019).
29. Lakeman, I. M. M. et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. J. Med. Genet. 56, 581–589 (2019).
30. Lynch, H. et al. Can unknown predisposition in familial breast cancer be family-specific? Breast J. 19, 520–528 (2013).
31. Park, D. J. et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. Cancer Discov. 4, 804–815 (2014).
32. Radmanesh, H. et al. Assessment of an APOBEC3B truncating mutation, c.783delG, in patients with breast cancer. Breast Cancer Res. Treat. 162, 31–37 (2017).
33. Riahi, A. et al. Exome sequencing and case-control analyses identify RCC1 as a candidate breast cancer susceptibility gene. Int. J. Cancer 142, 2512–2517 (2018).
34. Kiiski, J. I. et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc. Natl. Acad. Sci. U. S. A. 111, 15172–15177 (2014).
35. Cybulski, C. et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat. Genet. 47, 643–646 (2015).
36. Sun, J. et al. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet. 11, e1005228 (2015).
37. Cybulski, C. et al. Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland. Clin. Genet. 88, 366–370 (2015).
38. Määttä, K. et al. Whole-exome sequencing of Finnish hereditary breast cancer families. Eur. J. Hum. Genet. EJHG 25, 85–93 (2016).
39. Maxwell, K. N. et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am. J. Hum. Genet. 98, 801–817 (2016).
40. Tavera-Tapia, A. et al. Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the ATM gene. Breast Cancer Res. Treat. 161, 597–604 (2017).
41. Lu, H.-M. et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. JAMA Oncol. 5, 51–57 (2019).
42. Guo, X. et al. Discovery of a Pathogenic Variant rs139379666 (p. P2974L) in ATM for Breast Cancer Risk in Chinese Populations. Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 28, 1308–1315 (2019).
43. Snape, K. et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res. Treat. 134, 429–433 (2012).
44. Gracia-Aznarez, F. J. et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. PloS One 8, e55681 (2013).
45. Shahi, R. B. et al. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. BMC Cancer 19, 313 (2019).
46. Silvestri, V. et al. Whole-exome sequencing and targeted gene sequencing provide insights into the role of PALB2 as a male breast cancer susceptibility gene. Cancer 123, 210–218

(2017).
47. Antoniou, A. C. et al. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet. Epidemiol. 21, 1–18 (2001).
48. Antoniou, A. C. et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br. J. Cancer 86, 76–83 (2002).
49. Antoniou, A. C. & Easton, D. F. Models of genetic susceptibility to breast cancer. Oncogene 25, 5898–5905 (2006).
50. Lee, A. et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet. Med. Off. J. Am. Coll. Med. Genet. 21, 1708–1718 (2019).

Supplementary data

locus	EAF	RR	group
BRCA1	6.00E-04[1]	11.4[2]	High risk
BRCA2	0.001[1]	11.7[2]	High risk
PALB2	0.0011[3]	7.5[3]	High risk
ATM	0.0034[3]	2.8[3]	Moderate risk
CHEK2	0.0065[3]	2.3[3]	Moderate risk
TP53	5.00E-05[4]	4.5[5]	Syndrome
CDH1	1.00E-04[3]	5.3[3]	Syndrome
PTEN	4.00E-05[3]	12.7[3]	Syndrome
BARD1	0.001[6]	5.4[6]	Potential
FANCC	8.00E-04[7]	2.4[8]	Potential
FANCM	0.0037[9]	1.9[10]	Potential
MEN1	2.00E-05[11]	2.8[12]	Potential
MSH6	0.001[13]	1.9[3]	Potential
RECQL	0.001[13]	2.5[14]	Potential
STK11	2.00E-05[15]	15.2[16]	Potential
N=313	various[17]	various[17]	SNP

Supplementary table 1.
References for the effect allele frequencies (EAF) and relative risks (RR) for the breast cancer susceptibility alleles from Figure 1.

Supplementary references

1. Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer. 2008;98(8):1457-1466. doi:10.1038/sj.bjc.6604305
2. Easton DF, Pharoah PDP, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-2257. doi:10.1056/NEJMsr1501341
3. Couch FJ, Shimelis H, Hu C, et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol. 2017;3(9):1190-1196. doi:10.1001/jamaoncol.2017.0424
4. Gonzalez KD, Noltner KA, Buzin CH, et al. Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations. J Clin Oncol. 2009;27(8):1250-1256. doi:10.1200/JCO.2008.16.6959
5. LaDuca H, Polley EC, Yussuf A, et al. A clinical guide to hereditary cancer panel testing: evaluation of gene-specific cancer associations and sensitivity of genetic testing criteria in a cohort of 165,000 high-risk patients. Genet Med. August 2019. doi:10.1038/s41436-019-0633-8
6. Weber-Lassalle N, Borde J, Weber-Lassalle K, et al. Germline loss-of-function variants in the BARD1 gene are associated with early-onset familial breast cancer but not ovarian cancer. Breast Cancer Res. 2019;21(1):55. doi:10.1186/s13058-019-1137-9
7. FANCC | gnomAD. https://gnomad.broadinstitute.org/gene/ENSG00000158169. Accessed December 8, 2019.
8. Berwick M, Satagopan JM, Ben-Porat L, et al. Genetic heterogeneity among Fanconi anemia

SNPs associated with breast cancer.[27,28] Together these low risk alleles explain approximately 20% of the familial relative risk of breast cancer.[28]

heterozygotes and risk of cancer. Cancer Res. 2007;67(19):9591-9596. doi:10.1158/0008-5472.CAN-07-1501

9. FANCM | gnomAD. https://gnomad.broadinstitute.org/gene/ENSG00000187790. Accessed December 8, 2019.
10. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci USA. 2014;111(42):15172-15177. doi:10.1073/pnas.1407909111
11. Kamilaris CDC, Stratakis CA. Multiple Endocrine Neoplasia Type 1 (MEN1): An Update and the Significance of Early Genetic and Clinical Diagnosis. Front Endocrinol (Lausanne). 2019;10:339. doi:10.3389/fendo.2019.00339
12. Dreijerink KMA, Goudet P, Burgess JR, Valk GD. Breast-Cancer Predisposition in Multiple Endocrine Neoplasia Type 1. N Engl J Med. 2014;371(6):583-584. doi:10.1056/NEJMc1406028
13. RECQL | gnomAD. https://gnomad.broadinstitute.org/gene/ENSG00000004700. Accessed December 8, 2019.
14. Bogdanova N, Pfeifer K, Schurmann P, et al. Analysis of a RECQL splicing mutation, c.1667_1667+3delAGTA, in breast cancer patients and controls from Central Europe. Fam Cancer. 2017;16(2):181-186. doi:10.1007/s10689-016-9944-y
15. Beggs AD, Latchford AR, Vasen HFA, et al. Peutz-Jeghers syndrome: a systematic review and recommendations for management. Gut. 2010;59(7):975-986. doi:10.1136/gut.2009.198499
16. Giardiello FM, Brensinger JD, Tersmette AC, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology. 2000;119(6):1447-1453. doi:10.1053/gast.2000.20228
17. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002

Nederlandse samenvatting

Achtergrond

Wereldwijd is borstkanker de meest voorkomende vorm van kanker en meest voorkomende kanker-gerelateerde doodsoorzaak in vrouwen.[1] In Nederland ontwikkelt ongeveer één op acht vrouwen borstkanker gedurende haar leven.[2] Ondanks dat alle borsttumoren uiteindelijk ongeveer dezelfde maligne eigenschappen ontwikkelen, is er veel variatie in de onderliggende genetische, epigenetische veranderingen en de tumormicro-omgeving die hieraan ten grondslag ligt. Vanwege deze heterogeniteit worden tumoren vaak onderverdeeld in subtypes op basis van morfologie, histologie, DNA-expressie en mutatieprofielen.

De etiologie van borstkanker is complex, onder andere genetische factoren, fysieke eigenschappen, leefstijlfactoren en reproductiefactoren hebben invloed op het borstkankerrisico.[3] Ook familiegeschiedenis is een belangrijke risicofactor voor deze aandoening. Vrouwen met een eerstegraads familielid gediagnosticeerd met borstkanker hebben een relatief risico (RR) van ongeveer 1.8 om zelf borstkanker te ontwikkelen.4,5 Meerdere familieleden met borstkanker en een lagere leeftijd ten tijde van de diagnose zijn geassocieerd met een hoger risico.[4,5]

De twee bekendste borstkankergenen, *BRCA1* en *BRCA2*, zijn geassocieerd met een risico om borstkanker te ontwikkelen voor de leeftijd van 70 jaar van respectievelijk 60% en 55%.[6–8] Naast deze twee genen zijn er nog een aantal andere hoog-risico genen bekend. *PALB2* is een relatief recent ontdekt hoog-risico gen. Vrouwen met een pathogene genetisch variant in dit gen hebben een risico van ongeveer 35% om voor de leeftijd van 70 jaar borstkanker te ontwikkelen.[9,10] Daarnaast zijn er nog een aantal hoog-risico genen die geassocieerd zijn met een syndroom dat gekenmerkt wordt door een verhoogd risico op diverse vormen van kanker. Dit betreft, de genen *CDH1*,[11–13] *PTEN*,[14] *STK11*[15,16] en *TP53*.[17]

Naast hoog-risico genen zijn er ook een aantal genen geassocieerd met een matig verhoogd (RR 2-3) risico op borstkanker. Voor de Nederlandse populatie is het gen *CHEK2* het belangrijkst, omdat er een founder mutatie bestaat, c.1100delC, die in de Noordwest-Europese populatie een allel frequentie in de algemene populatie heeft van ongeveer 1%.[18] Daarnaast zijn ook genetische varianten in *ATM*[19] geassocieerd met een matig verhoogd risico op borstkanker. Tenslotte zijn er ook nog meer dan 300 laag-risico (RR <1.5) varianten geassocieerd gevonden met borstkanker.[20] Echter verklaren al deze hoog-, mattig-, en laag-risico varianten samen minder dan de helft van het familiare risico op borstkanker.

Voor families waarin een pathogene variant is gevonden in één van de bekende hoog- of matig-risico genen bestaan richtlijnen die beschrijven vanaf welke leeftijd, met welke frequentie en met welke techniek de draagsters van pathogenen variant gescreend zouden moeten worden voor borstkanker.[21–23] Bovendien kan onderscheid worden gemaakt tussen vrouwen die wel en niet draagster zijn van de genetisch variant geassocieerd met borstkankerrisico. Echter, in families waarin geen verklaring is gevonden voor het clusteren van borstkanker, blijft er onzekerheid bestaan over wie er een verhoogd risico heeft en in hoeverre dit risico verhoogd is. Dit maakt beslissingen met betrekking tot screening ingewikkelder.

Een relatief nieuwe techniek, "next generation sequencing" (NGS), maakt het mogelijk om in parallel miljoenen DNA-fragmenten te onderzoeken en daarmee in korte tijd grote delen van het genoom te analyseren. Deze techniek zou belangrijke mogelijkheden kunnen bieden voor het ontdekken van nieuwe borstkankergenen, omdat het de ontdekking van genetische varianten mogelijk maakt onafhankelijk van de allelfrequentie in de algemene populatie en zonder een aanname te hoeven maken over welke genetische regio's er mogelijk geassocieerd zijn met een verhoogd borstkankerrisico. Doordat deze techniek nog relatief

kostbaar is, was het voor dit onderzoek echter niet mogelijk om grote aantallen familiare borstkankerpatiënten te onderzoeken. Dit maakt dat het grote aantal genetische varianten dat gevonden wordt in deze patiënten teruggebracht moet worden tot een behapbaar aantal voor validatiestudies zonder daarbij gebruik te kunnen maken van statistische associatietesten.

Er bestaan diverse manieren om varianten te selecteren op basis van voorspelde effecten op het functioneren van het eiwit. Deze aanpak heeft echter als nadeel dat de beschikbare predictiealgoritme niet erg precies zijn. Ook het selecteren van varianten op basis van allel frequentie in de algemene populatie is niet erg effectief, aanzien het grootste deel van de gedetecteerde varianten zeldzaam is.[24,25] Een potentieel interessante strategie is om te focussen op genen waarin meerdere onderzochte families een potentieel pathogene variant hebben. De grote etiologische heterogeniteit in borstkanker maakt echter dat de a priori kans hierop erg klein is. Voor sommige bekende borstkankergenen, waaronder *BRCA1*, is er een associatie gevonden met specifieke borstkanker subtypes. Daarom zou het de moeite waard kunnen zijn om familiare borstkankerpatiënten te selecteren met vergelijkbare tumorkarakteristieken in de hoop zo ook een etiologisch homogenere populatie te selecteren.

Doel van het onderzoek

Het doel van het onderzoek beschreven in dit proefschrift is om bij te dragen aan de kennis van de genetische etiologie van borstkanker met behulp van NGS. Het focust op families met een duidelijk clustering van borstkanker, maar waarin geen pathogene varianten in *BRCA1* of *BRCA2* zijn gevonden. Hiermee hoopt dit onderzoek nieuwe inzichten te geven in de genetische risicofactoren die verantwoordelijk zijn voor het clusteren van borstkanker in deze families.

Resultaten

Voor het onderzoek beschreven in **hoofdstuk 2** van dit proefschrift hebben we zes niet-*BRCA1/2* gemuteerde families geselecteerd op basis van een "array comparative genome hybridization" profiel van de tumoren.[26] Linkage analyse in deze families had eerder een piek laten zien op chromosoom 4, wat suggereerde dat de clustering van borstkanker in deze families verklaard zou kunnen worden door een genetische variant in deze linkage regio. Daarom werden de exomen van twee individuen per familie onderzocht met behulp van NGS. Echter, geen enkel gen had een mogelijk pathogene variant in meer dan één familie en in het algemeen werden er geen waarschijnlijke kandidaat-risicoallelen gevonden.

Ook in het onderzoek beschreven in **hoofdstuk 5** hebben we geprobeerd een homogenere groep familiare borstkankerpatiënten te selecteren, ditmaal op basis van een mogelijk recessieve overerving binnen de families. In deze families met elk minstens drie zussen met borstkanker hebben we bij één zus het exoom geanalyseerd, terwijl we bij alle aangedane zussen de haplotypes hebben bepaald met behulp van "single nucleotide polymorphisme" (SNP) arrays. Dit maakte het mogelijk om bij de analyse van de exoom resultaten te focussen op varianten in regio's waarin alle aangedane zussen twee haplotypes delen. Ook deze studie vond echter geen waarschijnlijke kandidaat-hoog-risicoallelen. Wel werden er varianten in bekende matig-risico genen gevonden in verschillende families en werd daarnaast een verhoogd aantal laagrisico-allelen gevonden in de familiare borstkankerpatiënten in vergelijking met sporadisch borstkankerpatiënten.

Het is mogelijk dat we in deze twee studies een hoogrisico-allel gemist hebben. Het zou kunnen dat we zo'n variant niet geselecteerd hebben omdat het onwaarschijnlijk leek dat het de eiwitfunctie zou beïnvloeden of dit überhaupt niet gedetecteerd hebben omdat het zich buiten de eiwitcoderende regio's bevond. Bovendien sluiten deze twee studies niet uit dat er nog additionele hoog-risico genen bestaan met een sterke associatie met een bepaald fenotype, we zouden simpelweg het verkeerde fenotype gekozen kunnen hebben. De linkage piek in **hoofdstuk 2** had slechts een LOD-score van 2.49, terwijl een score hoger dan 3.0 typisch als significant gezien wordt. De linkage van deze families met de regio op chromosoom 4 zou daarom toeval kunnen zijn. De selectie van families in **hoofdstuk 5** had weliswaar belangrijke voordelen voor de filtering van varianten. Het is echter mogelijk dat een andere selectie van families, bijvoorbeeld op basis van jonge borstkankerpatiënten zonder familiegeschiedenis, beter verrijkt zou zijn geweest voor mogelijke recessieve allelen. In de laatste jaren zijn er diverse technieken ontwikkeld die het mogelijk maken een tumor in veel meer detail te karakteriseren. Het zou een goede eerste stap voor verder onderzoek zijn om op basis van deze nieuwe technieken uit te zoeken of er fenotypes gevonden kunnen worden die clusteren in families.

In **hoofdstuk 6** van dit proefschrift worden verschillende strategieën besproken voor het filteren van met behulp van NGS gevonden genetische varianten en het vaststellen van het borstkankerrisico geassocieerd met zeer zeldzame genetische varianten. Als er geen genen zijn waarin meerdere families een zeldzame en mogelijk pathogene variant hebben, is het filteren op basis van allel frequentie in de algemene populatie en op de functionele effecten zoals voorspeld door in silico predictiealgoritme een veel gebruikte strategie. Vaak is hiermee het aantal varianten nog niet voldoende teruggebracht om validatiestudies mogelijk te maken. Daarom wordt er vaak voor gekozen te filteren op basis van de functie van de betreffende genen en daarmee, op basis van de huidige beschikbare kennis, de waarschijnlijkheid dat een gen betrokken in bij de ontwikkeling van borstkanker.

Omdat veel van de overgebleven varianten erg zeldzaam zijn, is het vaak niet zinvol om te proberen een specifieke variant te valideren in case-controle studies. In plaats daarvan is een "burden" analyse een betere strategie. Hierbij wordt de hele eiwitcoderende regio van een gen onderzocht en het aantal mogelijk pathogene varianten vergeleken tussen (familiare) borstkankerpatiënten en gezonde controles. Een cruciaal aspect bij deze analyse is de definitie van een "mogelijk pathogene variant". Dit kan zowel met behulp van in silico predictiealgoritmen als met behulp van functionele testen gedaan worden. Verkeerde aanname met betrekking tot de functionele effecten van genetische varianten kunnen gemakkelijk leiden tot incorrecte conclusies met betrekking tot het borstkankerrisico geassocieerd met varianten in een specifiek gen.

Hoofdstuk 3 en **4** geven een goed voorbeeld van hoe ingewikkeld het kan zijn om het risico geassocieerd met zeer zeldzame varianten in een gen, in dit geval *XRCC2*, vast te stellen. Een mogelijke associatie tussen *XRCC2* en borstkanker werd als eerste gepubliceerd door Park et al.[27] **Hoofdstuk 3** van dit proefschrift beschrijft een studie waarin wij geprobeerd hebben, in een grote internationale case-controle studie, deze associatie te valideren. Voor de burden analyse in deze studie is gebruik gemaakt van verschillende in silico predictiealgoritmen, op basis daarvan kon echter geen associatie met borstkanker vastgesteld worden. Het onderzoek beschreven in **hoofdstuk 4** van dit proefschrift heeft vervolgens met behulp van drie verschillende functionele testen vastgesteld dat de meeste gevonden varianten in *XRCC2* nauwelijks effect hebben op eiwitfunctie. Voor de varianten die wel een effect op eiwitfunctie hadden, werd geen associatie met borstkanker gevonden,

al kon voor de varianten met het sterkste functionele effect een associatie niet uitgesloten worden door de extreme zeldzaamheid van deze varianten.

Ook andere studies die gebruik maakten van NGS om nieuwe borstkankergenen te vinden hebben relatief weinig succes gehad. De meer dan 30 gepubliceerde exoom sequencing studies hebben wel een aantal mogelijke nieuwe borstkankergenen gesuggereerd, waaronder *KAT6B*,[28] *RINT1*,[29] *APOBEC3B*[30], *XRCC2*[27] en *RCC1*.[31] Externe validatie van deze genen heeft echter nog niet plaatsgevonden of heeft geresulteerd in tegenstrijdige resultaten. Er zijn maar twee nieuwe genen gevonden waarbij de associatie met borstkanker ook in diverse andere studiepopulaties is gevalideerd: *FANCM*[32,33] en *RECQ*.[34,35]

In **hoofdstuk 5** hebben we naast de zoektocht naar een mogelijk recessief gen, ook de mogelijkheid verkend dat deze families verklaard worden door bekende risicogenen of polymorfismen. We hebben laten zien dat de familiare borstkankerpatiënten gemiddeld meer laagrisico-allelen hebben dan borstkankerpatiënten uit de algemene populatie. Daarnaast waren borstkankerpatiënten uit twee van deze families draagster van de bekende matig-risico variant *CHEK2**11000delC. In een andere studie, die geen deel uitmaakt van dit proefschrift, hebben we laten zien dat ook familiare borstkankerpatiënten die niet geselecteerd zijn voor een mogelijk recessieve overerving gemiddeld meer laagrisico-allelen bij zich dragen dan borstkankerpatiënten uit de algemene populatie.[36] Diverse andere exoom sequencing studies in non-*BRCA1/2* families hebben ook varianten gevonden in bekende risico genen zoals *ATM*,[37–42] *CHEK2*[39,42–45] en *PALB2*.[37,39,42,45,46] Dit bevestigt dat een substantieel deel van de onverklaarde familiare clustering van borstkanker waarschijnlijk verklaard wordt door een combinatie van matig en laag risico genen.

Een aantal studies heeft geprobeerd een zogenaamde "polygene risico score" (PRS) te berekenen op basis van een combinatie van bekende laag-risico varianten. De meest recente studie gebruikte hiervoor 313 verschillende genetische varianten.[20] Daarnaast zijn er modellen die zowel genetische als niet genetische risicofactoren meewegen. Het meest geavanceerde model is op dit moment de laatste versie van het BOADICEA model waarbij mutaties in *BRCA1*, *BRCA2*, *ATM*, *CHEK2* en *PALB2*, 313 laag risico SNPS en diverse niet-genetische risicofactoren worden gebruikt in de berekening van borstkankerrisico.[47]

Conclusie

Het onderzoek beschreven in dit proefschrift heeft geen nieuwe borstkankergenen kunnen identificeren. Ondanks dat er waarschijnlijk nog verscheidene extreem zeldzame hoogrisico-allelen te ontdekken zijn en er mogelijk genetische varianten buiten de eiwit-coderende regio's geassocieerd zijn met een sterk verhoogd risico op borstkanker, lijkt het onwaarschijnlijk dat deze varianten een substantieel deel van de familiare borstkanker kunnen verklaren. Zowel het onderzoek beschreven in dit proefschrift als werk van anderen suggereert dat een groot deel van de onverklaarde familiare clustering van borstkanker verklaard wordt door een combinatie van genetische varianten geassocieerd met een laag of middelmatig verhoogd risico op borstkanker.

Zoals verwacht was de grootste uitdaging bij het gebruik van NGS in familiare borstkanker de detectie van een groot aantal genetische varianten in relatief klein aantal familiare borstkankerpatiënten, waardoor het gebruik van statistische testen voor associatie niet zinvol is in de selectie van mogelijk interessante varianten. Hierdoor zal een aannamevrije strategie voor het ontdekken van nieuwe borstkankergenen alleen realiteit worden wanneer we in staat zijn studies te doen die voldoende groot zijn om tenminste op gen-niveau exoom-brede associatie analyses uit te kunnen voeren.

Het selecteren van een fenotypisch homogenere groep familiare borstkankerpatiënten in de hoop te selecteren voor homogenere etiologie, heeft niet geresulteerd in het ontdekken van mogelijke nieuwe risico-allelen in meer dan één familie. Echter, aangezien er nu meer geavanceerde technieken beschikbaar zijn voor het fenotyperen van tumoren, kan onderzocht worden of er mogelijk toch nog waarde in deze strategie zit voor toekomstige pogingen tot het ontdekken van nieuw hoogrisico-allelen.

Onze ervaring in het valideren van *XRRC2* als een borstkankergen heeft ons er wederom geleerd dat in silico predictiealgoritmen van beperkte waarde zijn voor het classificeren van genetische variante. De misclassificatie van varianten kan leiden tot misleidende resultaten in burden analyses en incorrecte conclusies over de associatie van een gen met borstkankerrisico. Hoewel functionele testen belangrijke inzichten kunnen geven in het effect van genetische varianten op het functioneren van een eiwit, is het opzetten en uitvoeren van deze testen vaak tijdrovend. Hierdoor is dit vaak alleen de moeite waard voor de meest waarschijnlijke kandidaatsgenen.

Deze bevindingen tezamen leiden tot de conclusie dat als we de risicovoorspelling voor vrouwen in onverklaarde borstkankerfamilies en voor vrouwen in de algemene populatie willen verbeteren, we verder moeten focussen op het ontwikkelen en valideren van risicomodellen die alle bekende risico-allelen combineren met niet-genetische riscofactoren. Na het valideren van deze modellen, zouden we gebruik kunnen maken van deep learning technieken om deze modellen verder te verbeteren en nieuwe risicofactoren te identificeren.

Dit vooruitzicht zou ertoe kunnen leiden dat de tweedeling tussen sporadische en familiare borstkanker met betrekking tot genetische risicofactoren verdwijnt. Dit zou een aanpassing vergen in zowel het perspectief van het onderzoek naar borstkankerrisico als voor de werkwijze en benadering voor moleculair pathologisch en klinisch genetisch onderzoek.

Referenties

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018;68(6):394-424. doi:10.3322/caac.21492
2. IKNL. Cijfers over kanker, Nederlandse Kankerregistratie. http://www.cijfersoverkanker.nl/. Accessed July 23, 2017.
3. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? J Cell Mol Med. 2005;9(1):208-221.
4. Pharoah PD, Day NE, Duffy S, Easton DF, Ponder BA. Family history and the risk of breast cancer: a systematic review and meta-analysis. Int J Cancer. 1997;71(5):800-809.
5. 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. 2001;358(9291):1389-1399. doi:10.1016/S0140-6736(01)06524-2
6. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 2007;25(11):1329-1333. doi:10.1200/JCO.2006.09.1066
7. Mavaddat N, Peock S, Frost D, et al. Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst. 2013;105(11):812-822. doi:10.1093/jnci/djt095
8. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017;317(23):2402-2416. doi:10.1001/jama.2017.7112
9. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is

a breast cancer susceptibility gene. Nat Genet. 2007;39(2):165-167. doi:10.1038/ng1959
10. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497-506. doi:10.1056/NEJMoa1400382
11. Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402-405. doi:10.1038/32918
12. Keller G, Vogelsang H, Becker I, et al. Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. Am J Pathol. 1999;155(2):337-342. doi:10.1016/S0002-9440(10)65129-2
13. Pharoah PD, Guilford P, Caldas C, International Gastric Cancer Linkage Consortium. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology. 2001;121(6):1348-1353.
14. Nelen MR, van Staveren WC, Peeters EA, et al. Germline mutations in the PTEN/MMAC1 gene in patients with Cowden disease. Hum Mol Genet. 1997;6(8):1383-1387.
15. Hemminki A, Markie D, Tomlinson I, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature. 1998;391(6663):184-187. doi:10.1038/34432
16. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 1998;18(1):38-43. doi:10.1038/ng0198-38
17. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990;250(4985):1233-1238.
18. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet. 2002;31(1):55-59. doi:10.1038/ng879
19. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 2006;38(8):873-875. doi:10.1038/ng1837
20. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
21. STOET. Erfelijke En Familiaire Tumoren: Richtlijnen Voor Diagnostiek En Preventie.; 2017. https://www.stoet.nl/wp-content/uploads/2017/04/STOET-Richtlijnenboekje-april2017_DEF.pdf.
22. STOET, VKGN. Landelijke Richtlijn Familiair Mamma/Ovariumcarcinoom.; 2010. http://www.oncoline.nl/familiair-mamma-ovariumcarcinoom.
23. NABON. Landelijke Richtlijn Mammacarcinoom.; 2017. http://www.oncoline.nl/mammacarcinoom.
24. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291. doi:10.1038/nature19057
25. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science. 2012;336(6082):740-743. doi:10.1126/science.1217283
26. Didraga MA, van Beers EH, Joosse SA, et al. A non-BRCA1/2 hereditary breast cancer subgroup defined by aCGH profiling of genetically related patients. Breast Cancer Res Treat. 2011;130(2):425-436. doi:10.1007/s10549-011-1357-x
27. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
28. Lynch H, Wen H, Kim YC, et al. Can unknown predisposition in familial breast cancer be family-specific? Breast J. 2013;19(5):520-528. doi:10.1111/tbj.12145
29. Park DJ, Tao K, Le Calvez-Kelm F, et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. Cancer Discov. 2014;4(7):804-815. doi:10.1158/2159-8290.CD-14-0212
30. Radmanesh H, Spethmann T, Enssen J, et al. Assessment of an APOBEC3B truncating mutation, c.783delG, in patients with breast cancer. Breast Cancer Res Treat. 2017;162(1):31-37. doi:10.1007/s10549-016-4100-9
31. Riahi A, Radmanesh H, Schurmann P, et al. Exome sequencing and case-control

analyses identify RCC1 as a candidate breast cancer susceptibility gene. Int J Cancer. 2018;142(12):2512-2517. doi:10.1002/ijc.31273
32. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci USA. 2014;111(42):15172-15177. doi:10.1073/pnas.1407909111
33. Peterlongo P, Catucci I, Colombo M, et al. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. Hum Mol Genet. 2015;24(18):5345-5355. doi:10.1093/hmg/ddv251
34. Cybulski C, Carrot-Zhang J, Kluźniak W, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet. 2015;47(6):643-646. doi:10.1038/ng.3284
35. Sun J, Wang Y, Xia Y, et al. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet. 2015;11(5):e1005228. doi:10.1371/journal.pgen.1005228
36. Lakeman IMM, Hilbers FS, Rodriguez-Girondo M, et al. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. J Med Genet. 2019;56(9):581-589. doi:10.1136/jmedgenet-2019-106072
37. Cybulski C, Lubiński J, Wokołorczyk D, et al. Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland. Clin Genet. 2015;88(4):366-370. doi:10.1111/cge.12524
38. Määttä K, Rantapero T, Lindström A, et al. Whole-exome sequencing of Finnish hereditary breast cancer families. Eur J Hum Genet. 2016;25(1):85-93. doi:10.1038/ejhg.2016.141
39. Maxwell KN, Hart SN, Vijai J, et al. Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet. 2016;98(5):801-817. doi:10.1016/j.ajhg.2016.02.024
40. Tavera-Tapia A, Pérez-Cabornero L, Macías JA, et al. Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the ATM gene. Breast Cancer Res Treat. 2017;161(3):597-604. doi:10.1007/s10549-016-4058-7
41. Guo X, Lin W, Bai M, et al. Discovery of a Pathogenic Variant rs139379666 (p. P2974L) in ATM for Breast Cancer Risk in Chinese Populations. Cancer Epidemiol Biomarkers Prev. 2019;28(8):1308-1315. doi:10.1158/1055-9965.EPI-18-1294
42. Lu H-M, Li S, Black MH, et al. Association of Breast and Ovarian Cancers With Predisposition Genes Identified by Large-Scale Sequencing. JAMA Oncol. August 2018. doi:10.1001/jamaoncol.2018.2956
43. Snape K, Ruark E, Tarpey P, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. Breast Cancer Res Treat. 2012;134(1):429-433. doi:10.1007/s10549-012-2057-x
44. Gracia-Aznarez FJ, Fernandez V, Pita G, et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. PLoS ONE. 2013;8(2):e55681. doi:10.1371/journal.pone.0055681
45. Shahi RB, De Brakeleer S, Caljon B, et al. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. BMC Cancer. 2019;19(1):313. doi:10.1186/s12885-019-5494-7
46. Silvestri V, Zelli V, Valentini V, et al. Whole-exome sequencing and targeted gene sequencing provide insights into the role of PALB2 as a male breast cancer susceptibility gene. Cancer. 2017;123(2):210-218. doi:10.1002/cncr.30337
47. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019;21(8):1708-1718. doi:10.1038/s41436-018-0406-9

Dankwoord

Promoveren is hard werken, maar promotieonderzoek doe je gelukkig niet alleen. Hieronder wil ik graag een aantal mensen in het bijzonder bedanken die direct of indirect hebben bijgedragen aan het tot stand komen van dit proefschrift.

Peter Devilee, bedankt dat je mij de mogelijkheid hebt geboden om dit promotieonderzoek in jouw groep uit te voeren. Jouw vertrouwen in een goede afloop en het geduld in de laatste fase zijn van groot belang geweest voor de succesvolle afloop van dit traject.

Christi van Asperen, ik ben je erg dankbaar voor je aanmoedigingen en positieve blik. Jij was voor de dit proefschrift de cruciale link met de klinische praktijk en de families waarvoor we dit onderzoek uiteindelijk voor gedaan hebben.

Caro Meijers, bedankt voor jouw ondersteuning in het lab. Jouw efficiënte en goed georganiseerde manier van werken hebben zeker bijgedragen aan de kwaliteit van dit onderzoek.

Maaike Vreeswijk, ik ben je erg dankbaar voor het feit ik altijd mijn verhaal bij je kwijt kon. De gesprekken met jou hebben me meer dan eens belangrijke inzichten opgeleverd, zowel voor mijn promotieonderzoek als voor mijn verdere wetenschappelijke carrière.

Ook mijn andere kamergenoten, Juul Wijnen en later Ellen Thomassen, en alle andere leden van de Tumgen groep wil ik bedanken voor vele gezellige gesprekken en levendige wetenschappelijke discussies.

Haico van Attikum, ik ben je erg dankbaar voor de mogelijkheid om de functionele analyse van *XRCC2* varianten in jouw groep uit te voeren samen met Martijn Luijsterburg en Wouter Wiegant. Jullie advies en scherpe analytische blik is van groot belang geweest voor dit hoofdstuk.

Mijn promotieonderzoek heeft gebruik gemaakt van een aantal complexe technieken en analysemethode. Daarom wil ik graag ook de personen bedanken die mij over de jaren van advies voorzien hebben op het gebied van next generation sequencing en bioinformatica, met name Sophie Greve, Yavuz Ariyurek, Jeroen Laros, Peter van 't Hof en Leon Mei.

Ook familie en vrienden wil ik bedanken voor hun steun, in het bijzonder mijn ouders die altijd voor mij klaar hebben gestaan. Mijn zusje, Francisca, ik ben ik erg dankbaar voor het luisterend oor dat je zo vaak hebt geboden, de aanmoedigingen en, in deze laatste fase, jouw praktische hulp als paranimf.

Tenslotte wil ik ook alle patiënten en hun familieleden bedanken die toestemming hebben gegeven voor het gebruik van hun DNA in dit onderzoek. Zonder hen had dit proefschrift niet tot stand kunnen komen.

Curriculum Vitae

Florentine Hilbers werd geboren op 14 juni 1986 in Wilnis. In 2004 behaalde ze haar Vwo-diploma aan het Veenlanden College in Mijdrecht en begon ze aan haar studie Medische Biologie aan de Universiteit van Amsterdam. Na het behalen van haar bachelor diploma, vervolgde ze haar studie met een master Epidemiologie aan het Netherlands Institute for Health Sciences (NIHES) van de Erasmus Universiteit in Rotterdam.

Tijdens haar master liep zij stage op de afdeling Psychosociaal onderzoek en Epidemiologie van het Nederlands Kanker Instituut onder begeleiding van Floor van Leeuwen en Marjanka Schmidt. Hier deed zij onderzoek naar de associatie van veel voorkomende genetische varianten in *TGFβ1* en *PAI-1* en het ontwikkelen van cardiovasculaire ziekte in patiënten behandeld met radiotherapie voor borstkanker. Dit project resulteerde in een publicatie en een presentatie op het nationaal congres van de Vereniging voor Epidemiologie (WEON).

In mei 2010 begon zij met haar promotieonderzoek in de Tumor Genetica groep van de afdeling Humane Genetica in het LUMC onder begeleiding van Professor Peter Devilee en Professor Christi van Asperen. Dit door het KWF gefinancierde onderzoek richtte zich op het verhelderen van de genetische basis van familiare borstkanker met behulp van next generation sequencing. Naast de publicaties in dit proefschrift, heeft zij de resultaten van dit onderzoek onder meer gepresenteerd op het jaarlijkse congres van de American Society of Human Genetics.

In Februari 2016 verruilde zij Leiden voor Brussel om daar als scientific advisor aan de slag te gaan bij Breast International Group, een netwerk van academische onderzoeksgroepen op het gebied van klinisch borstkankeronderzoek. Hier draagt zij bij aan diverse klinische trials op het gebied van borstkanker, waaronder MINDACT, (Neo)ALTTO, PYTHIA en LORELEI. Binnen deze studies focust zij vooral op het vinden van nieuwe biomarkers die de therapiegevoeligheid van een tumor voorspellen. Daarnaast is zij nauw betrokken bij de AURORA-studie, een moleculair screening programma voor patiënten met gemetastaseerde borstkanker.

Publicatielijst

1. **Hilbers FSM**, Boekel NB, van den Broek AJ, van Hien R, Cornelissen S, Aleman BMP, van 't Veer LJ, van Leeuwen FE, Schmidt MK. Genetic variants in TGFbeta-1 and PAI-1 as possible risk factors for cardiovascular disease after radiotherapy for breast cancer. *Radiother Oncol 2012;102:115–21*.
2. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, Neuhausen SL, John EM, Andrulis IL, Terry MB, Daly M, Buys S, Le Calvez-Kelm F, Lonie A, Pope BJ, Tsimiklis H, Voegele C, **Hilbers FM**, Hoogerbrugge N, Barroso A, Osorio A, Giles GG, Devilee P, Benitez J, Hopper JL, Tavtigian SV, Goldgar DE, Southey MC. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet 2012;90:734–9*.
3. **Hilbers FS**, Wijnen JT, Hoogerbrugge N, Oosterwijk JC, Collee MJ, Peterlongo P, Radice P, Manoukian S, Feroce I, Capra F, Couch FJ, Wang X, Guidugli L, Offit K, Shah S, Campbell IG, Thompson ER, James PA, Trainer AH, Gracia J, Benitez J, van Asperen CJ, Devilee P. Rare variants in XRCC2 as breast cancer susceptibility alleles. *J Med Genet 2012;49:618–20*.
4. Gracia-Aznarez FJ, Fernandez V, Pita G, Peterlongo P, Dominguez O, de la Hoya M, Duran M, Osorio A, Moreno L, Gonzalez-Neira A, Rosa-Rosa JM, Sinilnikova O, Mazoyer S, Hopper J, Lazaro C, Southey M, Odefrey F, Manoukian S, Catucci I, Caldes T, Lynch HT, **Hilbers FSM**, van Asperen CJ, Vasen HFA, Goldgar D, Radice P, Devilee P, Benitez J. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One 2013;8:e5568*1.
5. **Hilbers FS**, Meijers CM, Laros JFJ, van Galen M, Hoogerbrugge N, Vasen HFA, Nederlof PM, Wijnen JT, van Asperen CJ, Devilee P. Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected on the basis of aCGH tumor profiling. *PLoS One 2013;8:e55734*.
6. Southey MC, Park DJ, Nguyen-Dumont T, Campbell I, Thompson E, Trainer AH, Chenevix-Trench G, Simard J, Dumont M, Soucy P, Thomassen M, Jonson L, Pedersen IS, Hansen TV, Nevanlinna H, Khan S, Sinilnikova O, Mazoyer S, Lesueur F, Damiola F, Schmutzler R, Meindl A, Hahnen E, Dufault MR, Chris Chan T, Kwong A, Barkardottir R, Radice P, Peterlongo P, Devilee P, **Hilbers F**, Benitez J, Kvist A, Torngren T, Easton D, Hunter D, Lindstrom S, Kraft P, Zheng W, Gao Y-T, Long J, Ramus S, Feng B-J, Weitzel JN, Nathanson K, Offit K, Joseph V, Robson M, Schrader K, Wang S, Kim YC, Lynch H, Snyder C, Tavtigian S, Neuhausen S, Couch FJ, Goldgar DE. COMPLEXO: identifying the missing heritability of breast cancer via next generation collaboration. *Breast Cancer Res 2013;15:402*.
7. **Hilbers FSM**, Vreeswijk MPG, van Asperen CJ, Devilee P. The impact of next generation sequencing on the analysis of breast cancer susceptibility: a role for extremely rare genetic variation? *Clin Genet 2013;84:407–14*.
8. Peterlongo P, Catucci I, Colombo M, Caleca L, Mucaki E, Bogliolo M, Marin M, Damiola F, Bernard L, Pensotti V, Volorio S, Dall'Olio V, Meindl A, Bartram C, Sutter C, Surowy H, Sornin V, Dondon M-G, Eon-Marchais S, Stoppa-Lyonnet D, Andrieu N, Sinilnikova OM, Mitchell G, James PA, Thompson E, Marchetti M, Verzeroli C, Tartari C, Capone GL, Putignano AL, Genuardi M, Medici V, Marchi I, Federico M, Tognazzo S, Matricardi L, Agata S, Dolcetti R, Della Puppa L, Cini G, Gismondi V, Viassolo V, Perfumo C, Mencarelli MA, Baldassarri M, Peissel B, Roversi G, Silvestri V, Rizzolo P, Spina F, Vivanet C, Tibiletti MG, Caligo MA, Gambino G, Tommasi S, Pilato B, Tondini C, Corna C, Bonanni B, Barile M, Osorio A, Benitez J, Balestrino L, Ottini L, Manoukian S, Pierotti MA, Renieri A, Varesco L, Couch FJ, Wang X, Devilee P, **Hilbers FS**, van Asperen CJ, Viel A, Montagna M, Cortesi L, Diez O, Balmana J, Hauke J, Schmutzler RK, Papi L, Pujana MA, Lazaro C, Falanga A, Offit K, Vijai J, Campbell I, Burwinkel B, Kvist A, Ehrencrona H, Mazoyer S, Pizzamiglio S, Verderio P, Surralles J, Rogan PK, Radice P. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet 2015;24:5345–55*.
9. **Hilbers FS**, Luijsterburg MS, Wiegant WW, Meijers CM, Volker-Albert M, Boonen RA, van Asperen CJ, Devilee P, van Attikum H. Functional Analysis of Missense Variants in the

Putative Breast Cancer Susceptibility Gene XRCC2. *Hum Mutat 2016;37:914–25*.
10. Risi E, Biagioni C, Benelli M, Migliaccio I, McCartney A, Bonechi M, Guarducci C, **Hilbers F**, Di Cosimo S, Huober J, Romagnoli D, Boccalini G, Vitale S, Sotiriou C, Biganzoli L, Di Leo A, Malorni L. An RB-1 loss of function gene signature as a tool to predict response to neoadjuvant chemotherapy plus anti-HER2 agents: a substudy of the NeoALTTO trial (BIG 1-06). *Ther Adv Med Oncol 2019;11:1758835919891608*.
11. Lambertini M, Martel S, Campbell C, Guillaume S, **Hilbers FS**, Schuehly U, Korde L, Azim HAJ, Di Cosimo S, Tenglin RC, Huober J, Baselga J, Moreno-Aspitia A, Piccart-Gebhart M, Gelber RD, de Azambuja E, Ignatiadis M. Pregnancies during and after trastuzumab and/or lapatinib in patients with human epidermal growth factor receptor 2-positive early breast cancer: Analysis from the NeoALTTO (BIG 1-06) and ALTTO (BIG 2-06) trials. *Cancer 2019;125:307–16*.
12. Di Cosimo S, Appierto V, Pizzamiglio S, Tiberio P, Iorio MV, **Hilbers F**, de Azambuja E, de la Pena L, Izquierdo M, Huober J, Baselga J, Piccart M, de Braud FG, Apolone G, Verderio P, Daidone MG. Plasma miRNA Levels for Predicting Therapeutic Response to Neoadjuvant Treatment in HER2-positive Breast Cancer: Results from the NeoALTTO Trial. *Clin Cancer Res 2019;25:3887–95*.
13. Lambertini M, Campbell C, Gelber RD, Viale G, McCullough A, **Hilbers F**, Korde LA, Werner O, Chumsri S, Jackisch C, Wolff AC, Vaz-Luis I, Ferreira AR, Prat A, Moreno-Aspitia A, Piccart M, Loi S, de Azambuja E. Dissecting the effect of hormone receptor status in patients with HER2-positive early breast cancer: exploratory analysis from the ALTTO (BIG 2-06) randomized clinical trial. *Breast Cancer Res Treat 2019;177:103–14*.
14. Lakeman IMM, **Hilbers FS**, Rodriguez-Girondo M, Lee A, Vreeswijk MPG, Hollestelle A, Seynaeve C, Meijers-Heijboer H, Oosterwijk JC, Hoogerbrugge N, Olah E, Vasen HFA, van Asperen CJ, Devilee P. Addition of a 161-SNP polygenic risk score to family history-based risk prediction: impact on clinical management in non-BRCA1/2 breast cancer families. *J Med Genet 2019;56:581–9*.
15. Di Cosimo S, Appierto V, Pizzamiglio S, Silvestri M, Baselga J, Piccart M, Huober J, Izquierdo M, de la Pena L, **Hilbers FS**, de Azambuja E, Untch M, Pusztai L, Pritchard K, Nuciforo P, Vincent-Salomon A, Symmans F, Apolone G, de Braud FG, Iorio MV, Verderio P, Daidone MG. Early Modulation of Circulating MicroRNAs Levels in HER2-Positive Breast Cancer Patients Treated with Trastuzumab-Based Neoadjuvant Therapy. I*nt J Mol Sci 2020;21. doi:10.3390/ijms21041386*
16. Delaloge S, Piccart M, Rutgers E, Litiere S, van 't Veer LJ, van den Berkmortel F, Brain E, Dudek-Peric A, Gil-Gil M, Gomez P, **Hilbers FS**, Khalil Z, Knox S, Kuemmel S, Kunz G, Lesur A, Pierga J-Y, Ravdin P, Rubio IT, Saghatchian M, Smilde TJ, Thompson AM, Viale G, Zoppoli G, Vuylsteke P, Tryfonidis K, Poncet C, Bogaerts J, Cardoso F. Standard Anthracycline Based Versus Docetaxel-Capecitabine in Early High Clinical and/or Genomic Risk Breast Cancer in the EORTC 10041/BIG 3-04 MINDACT Phase III Trial. *J Clin Oncol 2020;:JCO1901371*.
17. Eiger D, Pondé NF, Agbor-Tarh D, Moreno-Aspitia A, Piccart M, **Hilbers FS**, Werner O, Chumsri S, Dueck A, Kroep JR, Gomez H, Láng I, Rodeheffer RJ, Ewer MS, Suter T, de Azambuja E. Long-term cardiac outcomes of patients with HER2-positive breast cancer treated in the adjuvant lapatinib and/or trastuzumab Treatment Optimization Trial. *Br J Cancer 2020;122:1453–60*.
18. **Hilbers FS**, van 't Hof PJ, Meijers CM, Mei H, Michailidou K, Dennis J, Hogervorst FBL, Nederlof PM, van Asperen CJ, Devilee P. Clustering of known low and moderate risk alleles rather than a novel recessive high-risk gene in non-BRCA1/2 sib trios affected with breast cancer. *Int J Cancer Published Online First: 7 May 2020. doi:10.1002/ijc.33039*
19. **Hilbers FS**, Aftimos P. Expanding the landscape of actionable genomic alterations in metastatic breast cancer: comprehensive genomic profiling for all? *Annals of Oncology Published Online First: 16 May 2020. doi:10.1016/j.annonc.2020.05.004*
20. Piccart MJ, **Hilbers FS**, Bliss JM, Caballero C, Frank ES, Renault P, Naït-Kaoudjt R, Schumacher E, Spears PA, Regan MM, Gelber RD, Davidson NE, Norton L, Winer EP, On Behalf of the

BIG-NABCG Collaboration. The road map to safe and well-designed de-escalation trials of systemic adjuvant therapy for solid tumors. *Submitted*

21. Aftimos P, Oliveira M, Irrthum A, Benelli M, Nili Gal-Yam E, Robson M, Fumagalli D, Ndozeng J, di Leo A, Sotiriou C, Ciruelos E, De Azambuja E, Viale G, Scheepers E, Curigliano G, Bliss JM, Reis-Filho J, Colleoni M, Balic M, Cardoso F, Albanell J, Duhem C, Marreaud S, Rojas B, Gombos A, Wildiers H, Guerrero Zotano AL, Hall P, Bonetti A, Larsson K, De Giorgis M, Khodaverdi S, Greil R, Sverrisdóttir Á, Caballero C, Loibl S, Linderholm B, Zoppoli G, Davidson NE, Johannsson O, Bedard P, Loi S, Knox S, Cameron D, Harbeck N, Yates L, Lasa-Montoya M, Brandao M, Vingiani A, Venet D, **Hilbers FS**, Piccart MJ. Genomics and transcriptomics analyses of breast cancer primaries and matched metastases in AURORA, the Breast International Group (BIG) molecular screening initiative. *Submitted*
22. Pondé N, Agbor-Tarh D, Dal Lago L, Korde LA, **Hilbers F**, Jackisch C, Werner O, Gelber RD, Jatoi A, Dueck AC, Moreno-Aspitia A, Sotiriou C, de Azambuja E, Piccart M. Tolerability and Toxicity of Trastuzumab or Trastuzumab + Lapatinib in Older Patients: A Sub-analysis of the ALTTO Trial (BIG 2-06; NCCTG (Alliance) N063D). *Submitted*.