Chapter

Whole exome sequencing of germline DNA from familial non-\textit{BRCA1/2} breast cancer cases with a homogeneous tumour profile

5

Maarten P.G. Massink\textsuperscript{1}, Saskia E. van Mil\textsuperscript{1}, Ans M.W. van den Ouweland\textsuperscript{2}, Quinten Waisfisz\textsuperscript{1}, Hanne Meijers-Heijboer\textsuperscript{1}.

\textsuperscript{1}Departments of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands.

\textsuperscript{2}Department of Clinical Genetics\textsuperscript{2}, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, The Netherlands.
Abstract

Introduction A large proportion of familial breast cancer susceptibility is still unexplained. Inherited germline mutations in the high-risk BRCA1, BRCA2, and PALB2 genes account for approximately 10 to 20 percent of familial breast cancer risk. The failure to discover other high risk genes by means of linkage studies suggests a great degree of genetic heterogeneity underlying breast cancer susceptibility and that novel high-risk genes either do not exist or are rare and difficult to detect. In a previous study, we identified a small number of familial non-BRCA1/2 mutated (BRCAX) basal-like breast carcinomas with shared genomic features. These features include recurrent characteristic copy number aberrations and most notably, similar to the BRCA1-related tumours in this study, copy number neutral LOH of large overlapping genomic regions on chromosome 17. We hypothesized these BRCAX samples to constitute a genetically more homogeneous group as compared to unselected BRCAX samples. This homogeneity could be caused by pathogenic germline mutations in the same high or moderate-risk breast cancer gene, or alternatively, multiple mutated genes acting in the same pathway.

Methods We performed whole exome sequencing analysis of germline DNA from the five basal-like BRCAX cases with shared characteristic copy number aberrations and copy neutral LOH on chromosome 17 in an effort to identify putative novel high or moderate-risk breast cancer genes.

Conclusion A novel pathogenic RAD51C germline mutation was identified in one of the BRCAX cases. We did not find conclusive evidence for causal mutations in unknown breast cancer risk genes in the other four patients. Nevertheless, the combination of tumour profiling (gene expression and genomic profiling) and whole exome_genome sequencing could prove useful in unraveling the underlying genetic basis of breast cancer.
Introduction

Both genetic and non-genetic factors are involved in the aetiology of breast cancer. Non-genetic factors include age, body mass index, alcohol intake and menstrual and reproductive history [1]. A family history of breast cancer is indicative for the presence of inherited genetic events that predispose to this disease. Approximately 5 to 10 percent of all breast cancer cases arise within such a familial clustering of multiple breast cancers [2]. The known genes/genetic variants associated with breast cancer risk can be divided according to the breast cancer risk they confer: high-risk genes typically confer lifetime risks of breast cancer of >50%, moderate-risk genes confer lifetime risks of >20% and common low-risk polymorphisms which confer smaller increases in risk (i.e., 10% to 20% lifetime risk).

Inherited germline mutations in the high-risk BRCA1, BRCA2, and PALB2 genes are identified in about 10 to 20 percent of breast cancer families. The BRCA1 and BRCA2 genes also predispose to ovarian cancer and a substantial fraction of families with breast and ovarian cancer harbour mutations in BRCA1 or BRCA2. Most of the deleterious mutations in these genes are small deletions or insertions that result in a truncated protein product [3]. The frequency of these mutations may vary in different populations [4]. Loss of heterozygosity of the normal allele is frequently observed in BRCA1/2 associated tumours [5]. Other high risk genes have been identified as part of inherited cancer syndromes such as TP53 mutations in Li-Fraumeni syndrome, PTEN germline mutations in Cowden syndrome, STK11 mutations in Peutz-Jegher syndrome, and mutations in the CDH1 gene involved in hereditary diffuse gastric cancer [6].

Genome-wide linkage studies have not discovered other high risk breast cancer genes, suggesting that no further high-risk genes of comparable importance to BRCA1 and BRCA2 exist [7]. If additional high-risk genes do exist these are too rare to detect by linkage analysis in unselected familial cohorts.

Moderate-risk variants have been found by re-sequencing candidate genes, such as those interacting with BRCA1 and BRCA2, or acting in the same pathway. Genes in this category include amongst others: CHEK2, ATM, PALB2, BRIP1, RAD51C, RAD51D, BARD1 [6]. The above mentioned susceptibility alleles are rare and do not reach a frequency of >1% of all breast cancer cases even when combined, suggesting a selective disadvantage. High and moderate risk variants in known genes together account for less than 25% of the familial risk of breast cancer. However,
since few genes have been studied by such methods, it is likely that additional susceptibility genes of this class exist [8].

Most of the unexplained genetic susceptibility to breast cancer is likely to be explained by a polygenic model involving a combination of many low-risk variants. The evaluation of common genetic variation in breast cancer has rapidly evolved by the implementation of genome-wide association studies (GWAS) in large cohort studies. GWAS in breast cancer so far have led to the discovery of more than 70 established breast cancer susceptibility loci, of which most are stronger associated to ER+ breast cancer than to ER- breast cancer [9-11].

BRCA1-related tumours are described to have distinct biological features, suggesting the involvement of specific oncogenic pathways in the development of these tumours. BRCA1-related tumours are often poorly differentiated ductal carcinomas of higher grade and higher mitotic count as compared to age matched sporadic tumours. Where most BRCA2-related tumours are estrogen receptor (ER) and progesterone receptor (PR) positive, the most striking feature of BRCA1-related tumours is their lack of ER, PR and HER2NEU expression (triple negative phenotype) [12]. Gene expression profiling has shown BRCA1-related tumours to be mostly of the basal-like subtype [12,13]. Basal tumours are frequently of the triple negative phenotype, and it has been suggested that sporadic triple negative/basal-like tumours in fact display defects in the BRCA1 DNA repair pathway [14]. Furthermore, characteristic genomic profiles have been shown for BRCA1/2-related tumours by array-comparative genomic hybridization (aCGH) and SNP-array analyses [15,16].

We recently described a characteristic genomic profile for BRCA1-mutated basal-like breast carcinomas [17]. In this study, BRCA1-mutated tumours were compared to a small group of familial breast carcinomas with no mutations in either BRCA1 or BRCA2 (BRCAX). These BRCAX samples were all basal-like breast carcinomas and a subset of these samples appeared to have shared genomic features. These features include recurrent characteristic copy number aberrations and most notably, similar to the BRCA1-related tumours, copy number neutral LOH of large overlapping genomic regions on chromosome 17. Methylation specific PCR indicated no evidence of hypermethylation of the BRCA1 promoter region in these samples, indicative that BRCA1 dysfunction in tumourigenesis can be excluded. Therefore, these BRCAX samples may constitute a genetically more homogeneous sample group as compared to unselected BRCAX samples.
We hypothesized that these basal-like BRCAX samples shared pathogenic germline mutations in the same high or moderate breast cancer risk gene, with similar biological functions as BRCA1 or a BRCA1-related DNA damage repair pathway. As loss of the wild-type BRCA1 allele is suggested to be the most common mechanism of inactivation in tumours from patients who carry a deleterious BRCA1 mutation [5], overlapping regions of LOH in BRCAX samples could hint towards the genomic region(s) harboring a novel breast cancer susceptibility gene(s). For this reason we performed whole exome sequencing analysis of germline DNA of the five basal-like BRCAX cases with shared characteristic copy number aberrations and copy neutral LOH on chromosome 17 in their tumours.

Materials and Methods

Patients
In previous work, we determined the genomic tumour profiles of 120 familial breast cancer cases. To find characteristic genomic aberrations for BRCA1-mutated basal-like breast carcinomas we compared copy number aberration (CNA) data of BRCA1-mutated and BRCAX basal-like breast carcinomas. During this analysis, a genomically homogeneous subgroup of 5 BRCAX samples was identified. These samples are characterised by CNA frequencies comparable to the BRCA1-mutated samples. Furthermore, these samples share a number of characteristic copy number gains and losses and an overlapping region of copy number neutral LOH on the long arm of chromosome 17.

Whole exome sequencing
Genomic DNA was extracted using standard methods. Exonic targets were enriched with SeqCap EZ Human Exome Library v3.0 kit (Nimblegen). Sequencing was performed with 100 bp paired-end reads on a Hiseq2000 (Illumina, San Diego, CA). Data analysis was performed using an in-house pipeline as described [18].

First pass filter settings applied a minimum coverage of 10 times. All variants with an allele frequency >1% in the 1000 Genomes Project (release February 2012), and the Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP, ESP5400 release), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) databases were excluded. Variants were restricted to nonsense, splice-site, frame shift and missense variants (restricted to genes found in two or more samples) located on chromosome
17 in regions of LOH in matching tumour material. These mutations were visualized in the integrated genomics viewer (IGV) [18], checked for their function and prediction of pathogenicity in de prediction tools SIFT [19], AlignGVGD [20,21], Mutation taster [22] and Polyphen [23]. Secondly, the curated BIOGRID (www.thebiogrid.org) database of genetic and protein interactions was used to identify 411 BRCA1 associated genes. These 411 genes were all scanned for possible pathogenic germline mutations if present in regions of LOH in matching tumour material.

**Follow up of variants**
Validation of the RECQL5 and BABAM1 variants was done using PCR and Sanger sequencing. Follow up of RECQL5 was performed by allele-specific PCR in Non-BRCA1/2 familial breast cancer cases (n = 356) obtained from the DNA diagnostic laboratory of VUmc. DNA samples from irreversibly anonymised healthy controls (n = 191) were used following the code for Proper Secondary Use of Human Tissue and were obtained from the DNA diagnostic laboratory of VUmc. All primer sequences are available upon request.

**RAD51C deletion screening**
The exome hidden Markov model software suite (XHMM, version 7653993a257a) [24] was used to search for any structural variation in the RAD51C gene. This statistical tool uses principal-component analysis (PCA) to normalize exome read depth and a hidden Markov model (HMM) to discover exon-resolution copy number variation (CNV) and genotype variation across samples.

**Results**
We previously analysed the genomic profiles of 120 familial breast tumours and identified five non-BRCA1/2 mutated (BRCAX) basal-like breast carcinomas with shared characteristic genomic features [ref Massink et al]. These features include recurrent characteristic copy number aberrations and copy number neutral LOH of large genomic regions on chromosome 17. We hypothesized these BRCAX samples to constitute a genetically homogeneous sample group caused by shared pathogenic germline mutations in the same high or moderate-risk breast cancer gene, or alternatively multiple mutated genes acting in the same pathway. As BRCA1-mutated basal-like breast tumours were all found to have copy neutral LOH at the BRCA1 locus, the regions of copy neutral LOH in the BRCAX samples
could point towards genomic regions harboring the unknown causative breast cancer susceptibility gene (s). We therefore performed whole-exome sequencing analysis of germline DNA for these 5 basal-like BRCA1X samples.

After filtering and restricting the reported variants to nonsense, splice-site, frameshift and missense variants located in regions of LOH in matching tumour material, 15 variants in 11 genes remained (Table 1). These variants include 2 nonsense variants, 1 frameshift deletion, 4 splice site altering variants and 8 missense variants. However, except for the missense variants, none of the genes carrying these variants were found in more than one sample.

Table 1. Variants detected by whole exome sequencing.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Gene</th>
<th>AAChange</th>
<th>Effect</th>
<th>dbSNP137</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>SPNS3</td>
<td>NM_182538:c.319C&gt;T:p.R107*</td>
<td>stopgain SNV</td>
<td>rs150923668</td>
</tr>
<tr>
<td></td>
<td>NAGLU</td>
<td>NM_000263:c.2209C&gt;A:p.R737S</td>
<td>SNV</td>
<td>rs86312</td>
</tr>
<tr>
<td></td>
<td>RECQL5</td>
<td>NM_004259:c.2828G&gt;A:p.R943H</td>
<td>SNV</td>
<td>rs200535477</td>
</tr>
<tr>
<td></td>
<td>FMNL1</td>
<td>NM_005892:c.486G&gt;A:p.T162T</td>
<td>SNV splicing</td>
<td>rs142930756</td>
</tr>
<tr>
<td>M2</td>
<td>CARD14</td>
<td>NM_052819:c.488G&gt;A:p.R163H</td>
<td>SNV</td>
<td>rs147466598</td>
</tr>
<tr>
<td>M3</td>
<td>KRT13</td>
<td>NM_002274:c.1253_1256del</td>
<td>frameshift deletion</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>RAD51C</td>
<td>NM_058216:c.955C&gt;T:p.R319*</td>
<td>stopgain SNV</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>MRC2</td>
<td>NM_006039:c.3333G&gt;A:p.T1111T</td>
<td>SNV splicing</td>
<td>rs144406119</td>
</tr>
<tr>
<td></td>
<td>FSCN2</td>
<td>NM_001077182:c.538C&gt;T:p.R180W</td>
<td>SNV</td>
<td>rs200332556</td>
</tr>
<tr>
<td></td>
<td>SDK2</td>
<td>NM_001144952:c.226T&gt;C:p.Y76H</td>
<td>SNV splicing</td>
<td>rs117687984</td>
</tr>
<tr>
<td>M4</td>
<td>NAGLU</td>
<td>NM_000263:c.944A&gt;T:p.N315I</td>
<td>SNV</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>CARD14</td>
<td>NM_024110:c.2539C&gt;T:p.L847F</td>
<td>SNV</td>
<td>_</td>
</tr>
<tr>
<td>M5</td>
<td>ABCA10</td>
<td>(NM_080282:exon21:c.2330+1G&gt;A)</td>
<td>splicing</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>RECQL5</td>
<td>NM_004259:c.2828G&gt;A:p.R943H</td>
<td>SNV</td>
<td>rs200535477</td>
</tr>
</tbody>
</table>

Table 1. List of genes in which nonsense, splice-site, frame shift and missense variants found with an occurrence <1% in the 1000genome and ESP5400 databases and located in regions of LOH in matching tumour material on chromosome 17. All SNVs are nonsynonymous SNVs.
For one patient we believe to have identified the causative variant in a known breast cancer gene. This variant is a novel nonsense mutation in the \textit{RAD51C} gene (c.C955T:p.R319X) located on the long arm of chromosome 17.

An interesting missense variant in the \textit{RECQL5} (ATP-dependent DNA helicase RecQ-like type 5) gene was found in two samples. This variant (c.G2828A:p.R943H, rs200535477) is reported in low frequencies in control databases (ESP5400: 0.0031, 1000genomes: 0.0014). Interestingly, this gene is located on chromosome 17q25, a region where copy neutral LOH is present in tumour material of both samples harbouring this variant. Analysis on tumour material showed that this variant allele was lost in one tumour and retained in the other. Despite this result the variant was followed up by allele specific PCR and was found with a frequency of 1.3% in breast cancer patients (9/356 cases in total) and a frequency of 1.8% in healthy controls (7/191). These data suggest that rs200535477 is not a high or moderate-risk allele for breast cancer.

Next, 411 \textit{BRCA1} associated genes, obtained from the BIOGRID database, were all scanned for possible pathogenic germline mutations if present in regions of LOH in matching tumour material. This resulted in two genes with possibly damaging missense variants, based on functional effect scores calculated by existing methods in PolyPhen2, SIFT, Align GVGD and Mutation Taster.

The first was the \textit{RECQL5} variant described above, the second variant was found in the \textit{BABAM1/NBA1} gene located on chromosome 19 in a region of LOH and copy number loss in the matching tumour material. The missense variant p.Y278C is predicted to be pathogenic by all prediction methods. However, the variant allele was found to be lost in matching tumour material, therefore no further analysis of this variant was performed.

\textbf{Discussion/Conclusion}

The majority of familial breast cancer risk is not explained by the genetic risk factors known to date, and are thought to have a heterogeneous genetic basis in which the occurrence of additional unknown high-risk alleles is rare [25,26]. As shown for rare Mendelian diseases, exome sequencing is useful to find such rare high-risk alleles [27]. However, this is more complex in a genetically heterogeneous disease like breast cancer. We hypothesized that by selecting BRCAX samples with a similar molecular breast cancer subtype and shared genomic characteristics, we would enrich for patients with germline mutations in the same gene, or pathway.
By whole exome sequencing, we found a novel nonsense mutation in the \textit{RAD51C} gene. This gene encodes a DNA double-strand break repair protein and biallelic mutations in this gene result in Fanconi anemia-like syndrome [28]. Several studies confirmed the occurrence of monoallelic \textit{RAD51C} germline mutations predominantly in breast and ovarian cancer families, although with varying frequencies, clearly establishing \textit{RAD51C} as a cancer-predisposing gene [29-31]. The patient carrying the \textit{RAD51C} nonsense mutation in the present study has a first degree relative diagnosed with breast cancer and a second degree relative diagnosed with ovarian cancer. Interestingly, by \textit{RAD51C} deletion screening, Schnurbein et al found a recurrent gross deletion in three breast tumours from two independent families. In line with our sample group characteristics, all three cases were classified as high-grade, invasive ductal, triple negative breast tumours [32]. No such deletions were seen in the remaining four BRCAX samples after copy number variation analysis by the XHMM methodology.

We identified a putative pathogenic missense variant in the \textit{RECQL5} gene in two patients (c.G2828A:p.R943H, rs200535477). This gene encodes an ATP-dependent DNA helicase from the RecQ family of genes. RecQ helicases are a conserved family of DNA helicases that display highly specialized roles in the maintenance of genome stability. Mutations in three of the five human RecQ helicases, \textit{BLM}, \textit{WRN} and \textit{RECQL4} are associated with the autosomal recessive diseases Bloom syndrome, Werner syndrome and Rothmund-Thomson syndrome respectively, which are characterized by chromosomal instability, premature aging and predisposition to cancer [33]. Although there is no human disease associated with lack of \textit{RECQL5}, \textit{RECQL5} deficiency in mice results in cancer predisposition indicating that it functions as a potential tumour suppressor [34]. Based on chromosomal location and gene function, the \textit{RECQL5} gene appeared to be a strong candidate gene. However the data from LOH analyses of the tumours and variant frequencies in breast cancer cases and controls suggests that a possible role for the \textit{RECQL5} variant in breast cancer susceptibility is unlikely.

We also identified a novel missense variant in the \textit{BABAM1} gene (19p13.11), (c.A833G:p.Y278C). This gene encodes a BRCA1 interacting protein that is essential for protein-protein interactions of a BRCA1 complex also containing FAM175A (Abraxas), BRCC3 and BRE. The complex acts early in DNA damage response and regulates damage-dependent BRCA1 localization [35]. Genome-wide association studies found that the 19p13.11 locus is associated with risk of serous epithelial ovarian cancer [36], and specifically with risk of triple negative breast cancer [37]. \textit{BABAM1} has been
regarded as a novel candidate gene for hereditary breast cancer. However, a comprehensive screening effort of this gene in affected cases of breast cancer families resulted in only a small number of sequence variants of which none appeared to be disease related. For this reason it was suggested that germline mutations in BABAM1 are rare or absent in familial breast cancer patients [38]. Further analysis in tumour material of the patient carrying the p.Y278C variant showed that the variant allele was lost. This observation and data from literature suggest that BABAM1 is not a likely cause in tumourigenesis and therefore no further follow-up analyses were performed.

In conclusion, we did not find conclusive evidence for the involvement of currently unknown breast cancer susceptibility genes in a subgroup of basal-like BRCAX cases selected by genomic profiling. This result could partially be explained by the fact that causative mutations might have been missed by the exome sequencing procedure. Despite our selection criteria, the sample group still appears to be genetically heterogeneous as the RAD51C variant was only found in one of the samples. Nevertheless, this finding also suggests the approach is valid and can be optimized for future endeavors. The combination of tumour profiling (gene expression and genomic profiling) and whole exome/genome sequencing in large cohorts could prove useful in unraveling the underlying genetic basis of breast cancer.
References


