Identifying New Genes for Inherited Breast Cancer by Exome Sequencing

Cailyn Helen Spurrell

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Mary-Claire King, Chair
Jay Shendure
Marshall Horwitz

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

Identifying New Genes for Inherited Breast Cancer by Exome Sequencing

Cailyn H Spurrell

Chair of the Supervisory Committee:
Professor Mary-Claire King
Department of Genome Sciences

Breast cancer is the most common cancer among American women and family history is an important risk factor for its occurrence. More than 20 genes have been identified with inherited mutations that lead to significantly increased risk of breast cancer. However, most familial breast cancer is not explained by mutations in these genes. The goal of this project was to identify additional breast cancer genes by exome sequencing.

In order to select families for gene discovery, we first screened families for mutations in all known breast cancer genes using targeted capture and massively parallel sequencing (BROCA). Families that remained unsolved after screening with BROCA were evaluated by exome sequencing. Germline DNA of 144 subjects with breast cancer from 54 high incidence families was sequenced. All truncating mutations shared by at least two affected persons in a family were genotyped in all participating members of that family in order to evaluate co-segregation with cancer. Rare truncating mutations co-segregating with breast and ovarian cancer were detected in *ATR*, *CHEK1*, and *GEN1*, each in one of the 54 families. ATR is recruited to sites of DNA damage; ATR phosphorylates CHEK1 in response to DNA damage, leading to a halt in cell cycle progression; GEN1 is an

endonuclease that resolves Holliday junctions. Like BRCA1 and BRCA2, all three candidate genes function in biological pathways related to homologous recombination repair.

In order to identify additional mutations in these three genes, unrelated women with breast cancer or ovarian cancer and controls were evaluated with high-throughput targeted sequencing approaches including BROCA custom capture and Molecular Inversion Probes (MIPs). Public databases were also reviewed. In *ATR*, truncating mutations were identified in 4 of 2544 cases and 3 of 7652 controls (P = 0.049). In *CHEK1* truncating mutations were identified in 5 of 2544 cases and 1 of 7652 controls (P = 0.004). In *GEN1*, truncating mutations were identified in 2 of 1717 cases and 3 of 7652 controls (P = 0.21). This study suggests new candidate genes for inherited predisposition to breast cancer while also demonstrating the challenges facing gene discovery for this complex disease.

Table of Contents

Acknowledgements

I would like to thank my advisor, Mary-Claire King. It has been a great honor and privilege to be a part of this lab, and Mary-Claire has been an exceptional mentor. Her teaching, guidance, and dedication to excellence have made me a better scientist. I would also like to thank my committee, Marshall Horwitz, Jay Shendure, Evan Eichler, and Rachel Klevit for their support throughout my dissertation.

I thank past and present members of the King lab, who have been another source of wisdom and support through grad school. I would like to particularly acknowledge Tom Walsh, Liz Swisher, Ming Lee, Silvia Casadei, Caitlin Rippey, Alex Nord, Diane Dickel, and Sarah Pierce.

On the road to graduate school I had several mentors who taught me to love science and practice it well. I especially thank Tom Gibson, John Budrick, and Len Pennacchio.

I want to thank all my amazing friends who have given me love and advice (and kept me sane) throughout graduate school. Thanks to my friends and classmates Renee George, Sarah Ng, and Ray Malfavon-Borja. I would separately like to thank Jessica Brown and Monika Tomsinski for their enduring friendship.

Finally, I would like to thank my family, who have loved and encouraged me at every step.

Chapter 1: Introduction: Genetics of inherited predisposition to breast cancer

Breast cancer is the second most common malignancy in American women, after skin cancer. Currently, an American woman has a 1 in 8 lifetime risk of being diagnosed with breast cancer(1). More than 240,000 new cases of invasive breast cancer are diagnosed in the United States each year. More than 40,000 deaths from breast cancer occur each year in the United States(1).

1.1 Breast cancer genes

Breast cancer was one of the first complex diseases to be evaluated using molecular genetics techniques. Initial studies in high-risk families were complicated by genetic heterogeneity of loci and alleles, incomplete penetrance, phenocopies, and expression dependent on age and sex(2). Despite this, linkage analysis led to the discovery and cloning of the breast cancer susceptibility genes *BRCA1* and *BRCA2* in severely affected families(3),(4). More than 1,000 different breast cancer-associated mutations in each gene have been documented (5). A study based on much earlier and less complete screening tools suggested that mutations in *BRCA1* and *BRCA2* account for 5% of all breast cancer diagnoses, and 10% of all breast cancer diagnosed before age 45(6). We are presently undertaking a new estimate based on complete sequencing of both genes. Mutations in these genes also confer an increased risk for ovarian cancer(7) and male breast cancer(8).

Screening for genes that interact with domains of BRCA1 and BRCA2 identified additional candidate genes for breast cancer susceptibility. These genes were sequenced in severely affected families that were negative for mutations in *BRCA1* and *BRCA2*. Loss of function mutations co-segregating with breast cancer in families were identified in *BRIP1*(9), *CHEK2*(10), *PALB2*(11), *ATM*(12), *BARD1*(13), *MRE11A*(14), and *NBN*(15); each mutation was individually rare. These studies supported the model that inherited breast cancer is extremely genetically heterogeneous, with respect to both loci and alleles. Like *BRCA1* and *BRCA2*, the other breast cancer genes function in pathways that maintain genomic integrity(16). Relative risks estimated for common alleles were 2- to 3-fold for *CHEK2 c.1100delC*(17) and *NBN c.657del5*(18) and as high as 4-fold for *PALB2 c.1592delT*(19).

Mutations in genes that cause some rare cancer syndromes also increase breast cancer risk. Increased breast cancer risk among female carriers is a component of Li-Fraumeni syndrome (*TP53*(20)), Cowden syndrome (*PTEN*(21)), hereditary diffuse gastric cancer (*CDH1*(22)), and Peutz-Jegher syndrome (*STK11*(23)). Mutations in Lynch syndrome genes (*MSH2*, *MSH6*, *MLH1*, *PMS2*(24)) increase risk of ovarian cancer, but not breast cancer.

Homozygosity or compound heterozygosity for loss-of-function mutations in *BRCA2*(25), *PALB2*(26) and *BRIP1*(27) cause Fanconi Anemia, a genomic instability disorder characterized by developmental abnormalities, bone marrow failure, and cancer predisposition(28). The convergence of the BRCA and Fanconi Anemia pathways suggests that some proportion of inherited breast cancer may be explained by mutations in as-yet-undiscovered genes within this or other DNA damage response pathways.

Recently, additional genes for breast and ovarian cancer susceptibility have been discovered both by candidate gene sequencing and by unbiased exome sequencing.

These genes also act in DNA damage response pathways. Candidate gene resequencing in *BRCA1/BRCA2*-negative probands identified *RAD51C*(29) and *RAD51D*(30) as susceptibility genes in breast and ovarian cancer families. Exome sequencing identified *XRCC2*(31) and *Abraxas/FAM175A*(32) as candidate genes for breast cancer susceptibility. In the case of *XRCC2*, a large validation study did not replicate the association(33). Additional studies are needed to confirm whether variants in *XRCC2* and *Abraxas* confer an elevated breast cancer risk, and if so, the relative risk of deleterious mutations in these genes. This highlights one difficulty of gene identification for inherited breast cancer. Case-control analysis provides excellent evidence for causality, but is often statistically inconclusive given such extreme genetic heterogeneity.

An alternative hypothesis for remaining genetic influences on inherited breast cancer is the "common disease-common allele" model(34). Under this hypothesis, a combination of alleles, each with a modest effect, would lead to increased susceptibility to breast cancer. Multiple genome-wide association studies (GWAS) of breast cancer have suggested common SNPs with statistically significant effects on risk associated with low odds ratios(35–38). None of the loci found by GWAS overlap with any known breast cancer gene. This is due to the heterogeneity of loss-of-function alleles in these genes. The results of GWAS have been difficult to interpret.

Functional relationships between breast cancer and candidate SNPs have generally not been assessed.

Clinical genetic testing of *BRCA1* and *BRCA2* in the United States is predominately carried out by Myriad Genetics, Inc. As of summer 2013, their main test uses PCR and Sanger sequencing to screen for deleterious mutations in these genes. In 2006, an additional test to detect copy number variants in *BRCA1* and *BRCA2* was added for high-risk patients who did not have damaging single nucleotide variants. Testing for mutations in other breast cancer genes is available in various research laboratories, but is not routinely performed. These tests have also been performed with Sanger sequencing with a laborious gene-by-gene approach.

1.2 Next generation sequencing

In 2001 the draft human genome was published after 10 years of effort by hundreds of collaborators(40). The project used Sanger sequencing technology(41) at a cost of approximately $300 million(42). Having a reference genome makes it possible to resequence additional individuals with shotgun sequencing techniques and align reads to unique portions of the genome for the study of any phenotype.

Recent advances in technology have decreased sequencing cost and time. In 2008 whole genome sequences were reported in several healthy individuals(43),(44). The main advantage of next generation technology is the massively parallel nature of the sequencing reaction. Millions of DNA templates are immobilized and clustered on a solid surface. Sequencing-by-synthesis

technologies use various methods to incorporate fluorescent nucleotides into each DNA molecule one base at a time. After each cycle, all DNA clusters are imaged to identify the incorporated nucleotide(45). In this project, I used the Illumina platform for all next-generation sequencing.

While next generation sequencing technologies make sequencing faster and more economical, there are several disadvantages. These technologies have shorter read lengths than Sanger sequencing making it difficult to assemble repeat-rich regions of the genome such as *Alu* elements, segmental duplications, microsatellites, and centromeres. Error rates are higher for next generation technologies, with a specific type of error profile. Several groups have developed strategies to overcome these limitations(46).

1.3 Targeted sequencing strategies

Massively parallel sequencing technologies developed in the last 5 years have greatly increased the sequencing capacity of individual labs. Targeted sequencing approaches interrogate a directed subset of the genome in a high-throughput manner(47). Genomic DNA is hybridized to complementary probes that are specific to the target regions and isolated and purified using magnetic beads(42). Often a unique barcode is added to all targeted fragments from one individual to facilitate downstream multiplexing.

The first targeted sequencing strategy assays the exome, the set of all protein coding exons in the human genome(47). The haploid human exome contains 30 Mb, about 1% of the genome. By

targeting the consensus coding DNA sequence (CCDS), microRNAs, and other ncRNAs, it is possible to ascertain protein coding variants and other single nucleotide variants and small insertions and deletions with interpretable functional consequences. Recent advances enable identification of copy number variants (>3 exons long) using exome data(48). By capturing the exome, we also avoid sequencing the majority of the repetitive regions of the genome, which are hard to map and assemble.

A second targeted sequencing strategy is to design custom probes to a region of interest, usually much smaller than the exome. One application of targeted sequencing is re-sequencing many genes or loci implicated in a genetically heterogeneous phenotype. Another application of targeted sequencing is re-sequencing a large linkage region. Linkage analysis in a large pedigree may lead to a candidate chromosomal location for gene discovery that is unwieldy for Sanger sequencing. Targeting the entire genomic locus tests all genes and regulatory sequence in one experiment, and is an unbiased analysis of the linkage region. These approaches query a smaller fraction of the genome than the exome and are a high-throughput way to screen cases or populations with a specific genetic question.

1.4 BROCA

An approach named BROCA, developed in our lab, exploits targeted capture followed by massively parallel sequencing in order to simultaneously test for mutations in all known breast individuals from 800 severely affected breast cancer families and ovarian cancer susceptibility genes(49,50). BROCA v.4 captures 1.1Mb of genomic sequence from 24 genes. To date we

have used BROCA to evaluate more than 2400 affected. Including only germline truncating mutations, complete gene deletions, and missenses experimentally shown to be damaging, we resolved 26% of families (unpublished data). In these families, we found damaging mutations in 19 genes. BROCA is now available clinically, with results returned for genes including *BRCA1* and *BRCA2*. Knowing genetic status will influence medical treatment of the patients harboring mutations.

Most high incidence breast cancer families in our series remained unsolved after sequencing all known breast cancer susceptibility genes. We hypothesized that damaging loss of function variants in as-yet-undiscovered genes would explain some fraction of the unsolved families. The discovery of such genes became the subject of my dissertation project. While missense substitutions or in-frame indels may also have a loss of function phenotype, the consequences of these classes of mutations are more difficult to predict with *in silico* tools. Furthermore, in complex diseases, phenocopies, incomplete penetrance and age- and sex-specific expressivity add to the difficulty in determining which segregating variants were causal. Hence, I focused on variants most likely to involve loss of function, defined here as nonsense variants, frameshifting indels, and copy number variants that caused early truncations or completely deleted a gene.

Chapter 2: Exome sequencing of severely affected families

2.1 Approach

Methods for targeted enrichment and sequencing of the human exome are now widely used to assay the protein coding and miRNA containing portions of the genome. Exome sequencing has been successful at revealing the underlying causal mutations in many Mendelian disorders. Recently, it has been applied to more complex phenotypes, including cancer.

For this project, I sequenced germline DNA from144 individuals with breast cancer (143 women and 1 man) from 54 families using exome capture and massively parallel sequencing. Of the 54 families, 17 families included cases of breast cancer and ovarian cancer, 4 families included cases of both female and male breast cancer, and 33 families included only cases of female breast cancer. Depending on family structure, 2-4 affected individuals per family were sequenced. When possible, I chose affected cousins or individuals whose cancer would most likely share a common genetic origin. In all individuals, I identified loss of function mutations that were not present in dbSNP137 or 1000 Genomes Project, then evaluated co-segregation of these mutations with breast cancer in the family. Overall, I evaluated co-segregation of 388 rare, likely loss-of-function variants. The most promising candidate genes were revealed to be *ATR, CHEK1* and *GEN1*, each of which harbored a damaging mutation in one of the 54 families. Re-sequencing of these genes in other affected individuals revealed additional rare, truncating mutations in the same genes.

With few exceptions, each individual was sequenced with Nimblegen v2.0 and the Illumina HiSeq 2000/2500. Table 1 indicates the unique identifier for each family, the date sequenced, and the sequencing platform used. Libraries were sequenced to an average 100x depth with 93% of bases covered by >10 reads. After alignment, I reviewed all rare insertion/deletion and single nucleotide variants in order to identify all possible truncating mutations including frameshifts, nonsense and potential splice altering variants that were present in at least 2 related, sequenced cases. I then Sanger sequenced these variants in all members of the same family (not only those exome-sequenced) in order to determine segregation of the variant in the pedigree. The number of variants in each family evaluated for co-segregation is presented in Table 1. In some cases no truncations were present in more than one sequenced individual. In these families, I evaluated truncations present in only one individual or potentially damaging missenses shared by multiple sequenced individuals.

Copy number variants (CNVs) were called in 80 exomes using Conifer(48), which can detect CNVs that span at least 3 exons. Families in which CNVs were called are noted in Table 1. Exomes that did not have CNV analysis were sequenced using an earlier platform or could not be analyzed with the Burrows-Wheeler Alignment tool. No potentially damaging CNVs were identified.

Table 1: Exome sequencing and evaluation of co-segregation with cancer in 54 families.
*Loss-of-function variants evaluated for co-segregation with cancer

Family	Proband	V2 exome	Instrument	CNV analysis	Variants*
CF43	CF4301	10/27/11	HiSeq	x	9
CF55	CF5504	6/10/11	HiSeq	x	1
CF57	CF5718	2/25/11	GAIIx	x	14
CF61	CF6101	6/11/11	HiSeq	x	18
CF62	CF6210	6/12/11	HiSeq	x	2
CF94	CF9401	2/25/11	GAIIx	x	15
CF265	CF26516	6/13/11	HiSeq	x	8
CF448	CF44801	6/14/11	HiSeq	x	8
CF450	CF45001	3/7/10	GAIIx		9
CF470	CF47003	6/15/11	HiSeq	x	6
CF515	CF51501	7/15/11	HiSeq	x	9
CF548	CF54801	6/16/11	HiSeq	x	6
CF590	CF59001	6/17/11	HiSeq	x	4
CF689	CF68901	10/27/11	HiSeq	x	1
CF737	CF73701	4/13/10	GAIIx		8
CF758	CF75801	9/14/11	HiSeq	x	14
CF779	CF77901	9/14/11	HiSeq	x	16
CF823	CF82301	10/4/11	HiSeq	x	9
CF910	CF91004	9/14/11	HiSeq	x	9
CF923	CF92305	6/18/11	HiSeq	x	5
CF944	CF94403	10/1/12	HiSeq		9
CF963	CF96301	10/1/12	HiSeq		11
CF985	CF98502	10/4/11	HiSeq	x	2
CF997	CF99701	10/4/11	HiSeq	x	3
CF998	CF99801	10/1/12	HiSeq		4
CF1026	CF102601	10/4/11	HiSeq	x	7
CF1036	CF103602	9/14/11	HiSeq	x	7
CF1041	CF104101	9/14/11	HiSeq	x	8
CF1085	CF108501	9/14/11	HiSeq	x	17
CF1105	CF110503	9/14/11	HiSeq	x	10
CF1152	CF115203	10/27/11	HiSeq	x	8
CF1249	CF124901	10/1/12	HiSeq	x	15
CF1257	CF125704	5/16/10	GAIIx		10
CF1292	CF129201	2/14/12	HiSeq	x	4
CF1329	CF132902	4/2/12	HiSeq	x	2
CF1439	CF143902	9/7/12	HiSeq		7
CF1445	CF144501	4/6/12	HiSeq		1
CF1604	CF160404	5/18/12	HiSeq	x	15
CF1760	CF176001	4/2/12	HiSeq	x	13
CF1778	CF177803	1/25/13	HiSeq		6
CF1790	CF179002	2/14/12	HiSeq	x	3
CF1793	CF179302	1/25/13	HiSeq		5
CF1794	CF179405	1/25/13	HiSeq		1
CF1830	CF183002	1/25/13	HiSeq		10
CF1831	CF183101	1/25/13	HiSeq		4
CF1859	CF185906	5/18/12	HiSeq	x	3
CF1863	CF186301	9/7/12	HiSeq		2
CF1888	CF188801	9/7/12	HiSeq		2
CF1903	CF190303	9/7/12	HiSeq		2
CF1919	CF191901	1/25/13	HiSeq		8
CF2047	CF204701	1/25/13	HiSeq		6
CF2185	CF218501	1/25/13	HiSeq		5
CF3503	CF350301	1/25/13	HiSeq		2
JBC267	JBC26702	10/23/11	HiSeq		5

In principle, the Exome Variant Server (EVS)(51) would be a valuable tool for determining whether a candidate variant could be involved in breast cancer susceptibility, because a heterozygous variant seen in multiple control exomes is not likely to be cancer associated. However, the EVS is of limited usefulness for breast cancer studies, because sex of contributors is not identified and women with breast cancer were not excluded. As is clear from the mutational profiles of *BRCA1* and *BRCA2*, variants in breast cancer genes that occur in male controls may nonetheless increase cancer risk among female carriers. In addition, most samples from females sequenced by the EVS were collected from the Women's Health Initiative (WHI), which ascertained women at high risk of developing breast cancer. WHI participants who subsequently developed cancer were not excluded or identified as affected on the EVS. Indeed, the EVS contains multiple rare truncations in known cancer genes. In 6500 exomes, there are 7 truncations in *BRCA1* and 21 truncations in *BRCA2*, whether in males, affected females, or unaffected females is not indicated. Finally, insertions and deletions are not validated by the EVS, and at this date there are still indel calls that are extremely likely to be errors: for example, *BRCA2 c.1806insA* is indicated with a frequency of 150 of 6174 exomes. Despite these problems, I used the EVS as a source of variant frequencies in controls, but recognize that comparisons involving EVS control data are very conservative.

Of the 388 variants evaluated in families, three variants in three different genes proved the most promising candidates. These genes were *ATR*, *CHEK1* and *GEN1*. These genes shared the features that (a) they harbored loss-of-function mutations that were either *de novo* in a breast cancer patient or segregated reasonably well with breast cancer in a family; (b) they did not

include multiple loss of function mutations on the EVS, and (c) they were biologically compelling. For these three genes, I re-sequenced additional unrelated familial breast cancer cases in order to identify other additional damaging variants. I also investigated the presence of rare mutations in these genes in two series of controls. The following sections present this information about each new candidate gene.

2.2 ATR

In family CF1026, exome sequencing revealed truncating mutation *ATR c.4468 G>A* (p.W1490X) in the proband diagnosed with breast cancer at age 34 (Figure 1). This variant was not present on the Exome Variant Server, dbSNP137 and 1000 Genomes. The variant was also not present in any other relative in CF1026, including both parents, suggesting that it occurred *de novo* in the proband. The variant sequence was verified by Sanger sequencing, both from the original sample from the proband and from a second blood sample. Read depth at this site from exome sequence of the proband was 39 variant reads and 95 wildtype reads. Although the read ratio is not 1:1, it is sufficiently balanced to suggest that the variant was not an artifact of circulating tumor DNA. It is possible that the mutation is *de novo* and somatic. In order to verify that relationships in the family were correctly identified, I carried out haplotype analysis of the family by genotyping microsatellite markers near *ATR* in the proband, her parents, and other relatives. I concluded that the *ATR* mutation is very likely to be *de novo* in this patient. Four other breast cancer cases in this family remain unsolved by exome sequencing.

ATR is a member of the phosphatidylinositol kinase-related kinase (PIKK) family and is an essential regulator of genomic integrity(52). The ATR protein and its binding partner ATRIP are

recruited to ssDNA at stalled replication forks. ATR phosphorylates the effector kinase CHEK1 and other downstream proteins to halt cell cycle progression and repair DNA damage. The classical interpretation of this relationship is that ATR/CHEK1 and ATM/CHEK2 perform analogous but independent roles in response to double stranded breaks in DNA(53). However, recent publications suggest that there are also points of intersection between the pathways(54). *ATM* and *CHEK2* are known breast cancer susceptibility genes(10,12), making *ATR* and *CHEK1* excellent candidates for new breast cancer genes.

An inherited missense mutation in *ATR* was recently implicated in a single large kindred as the causal variant for a syndrome involving oropharyngeal cancer, telangiectasias, and mild developmental defects(52). One female heterozygous for the missense mutation was diagnosed with breast cancer prior to age 35.

Given the damaging nature of the variant in the proband of family CF1026 and the role of *ATR* in DNA damage response, we decided to investigate this pathway further. *ATR* and *CHEK1* were added to the fourth version of our BROCA capture as candidate genes. We sequenced 1177 cases from high-incidence families with BROCA v.4 containing *ATR* and *CHEK1*. We then designed Molecular Inversion Probes to sequence an additional 540 breast cancer cases specifically for *ATR* and *CHEK1*.

With targeted sequencing we found two additional rare, loss-of-function *ATR* variants in breast cancer patients (Figure 1). In family CF1347, of Ashkenazi Jewish (AJ) ancestry, the proband diagnosed with breast cancer at age 40 and her mother diagnosed with breast cancer at age 63

shared a genomic deletion spanning 3 exons of *ATR* (chr3: 142,176,991-142,181,325; *ATR* p.C2348_K2501del). The deletion is in-frame and removes 153 amino acids of the 310 amino acids of the kinase domain. This deletion was not present in 665 other AJ breast cancer cases or in 760 European-American controls (or in 192 Palestinian controls). The second additional deleterious *ATR* mutation was identified in the proband of family CF2497, of Brazilian ancestry. This patient was diagnosed at age 46 with bilateral, triple negative (ER-/PR-/HER-) breast cancer. We do not yet know her family history. The mutation is *ATR c.3045 G>A* (p.R1015X). Neither mutation was present in the EVS, dbSNP137, or 1000 Genomes.

BROCA sequencing of 773 patients with high grade ovarian cancer, not selected for family history or age at diagnosis, yielded two additional ATR truncations: *ATR c.4957 C>T* (p.R165X), and *ATR c.7215insT* (p.E2409EfsX13). Neither mutation was present on the EVS, dbSNP137, or 1000 Genomes.

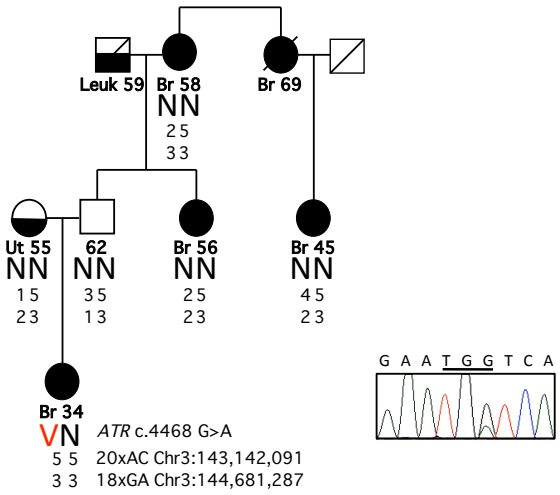
Next, using Molecular Inversion Probes (MIPs), I sequenced the entire *ATR* gene in genomic DNA of 1152 controls, all females older than age 50 of European-American ancestry. Average coverage was 40X with 96% of the targeted region was covered by >10 reads. No additional truncations of *ATR* were identified in these controls. The Exome Variant Server lists 3 rare truncations in *ATR,* each in one of 6,500 individuals. *ATR* has an average coverage of 87x on the EVS.

Comparing frequencies of truncating alleles in *ATR* in cases and controls, we count 4 events in 2544 cases and 3 events in 7652 controls. (The fifth ATR truncation is a CNV, and cannot be

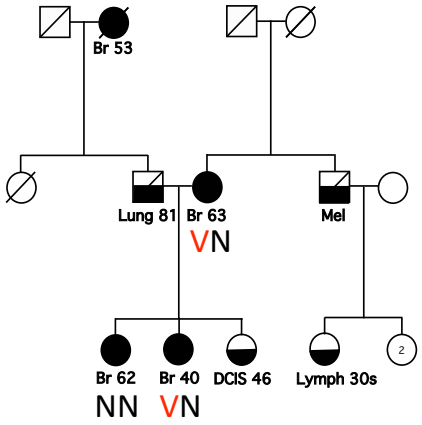
included for this comparison, as EVS and MIP controls were not screened for CNVs.) Based on a test of differences in proportions, $P = 0.049$.

(a)

CF 1026: *ATR* c.4468 G>A (p.W1490X)



CF1347: *ATR* c.7042_7503del (p.C2348_K2501del)



(b)

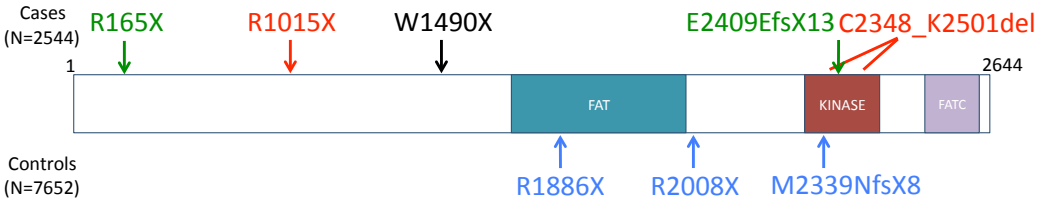


Figure 1. Families and controls with *ATR* mutations
(a) Pedigrees of cancer families with *ATR* variants
(b) Protein diagram of ATR with variants found by exome sequencing (black), BROCA of cancer families (red), BROCA of ovarian case series (green), and controls (blue).

2.3 *CHEK1*

CHEK1 and *CHEK2* are functionally similar but not structurally homologous kinases. Studies in *Saccharomyces pombe* suggest that the yeast *Chk1* C-terminal domain has a global regulatory effect on gene function. Truncating mutations in the C-terminal domain of *Chk1* cause loss of function *in vivo*(55). Missense mutations in this regulatory domain can have either activating or loss-of-function phenotypes(56). The C-terminal domain is well conserved from human to *S. pombe* and contains activating phosphorylation sites for *ATM* and *ATR*(57).

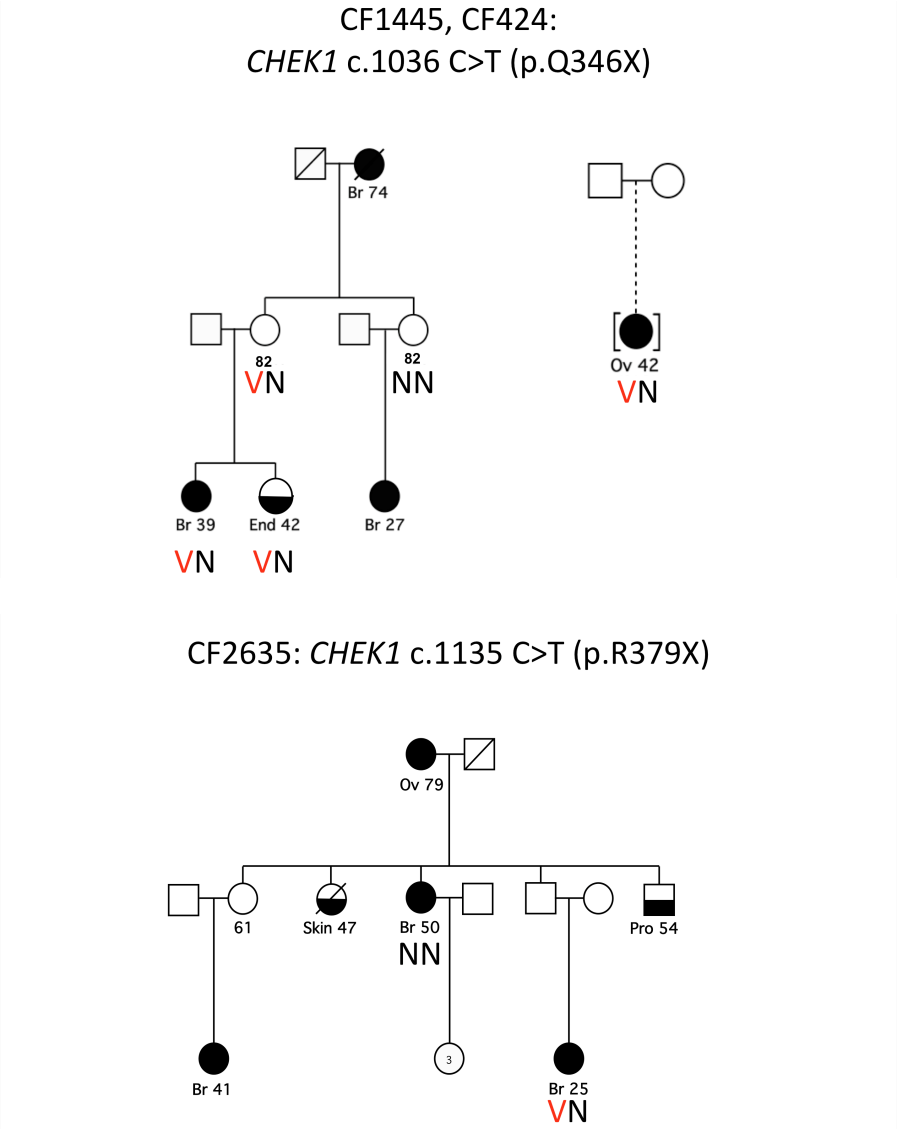
In family CF1445, exome sequencing revealed truncating mutation *CHEK1 c.1036 C>T* (p.Q346X), carried by the proband diagnosed with breast cancer at age 39 (Figure 2). Her sister, who was diagnosed with endometrial cancer at age 42, shares this variant. The truncation was inherited from their mother, who is unaffected at age 82. Re-sequencing other breast and ovarian cancer cases with BROCA and MIPs revealed the same variant in a patient diagnosed with ovarian cancer at age 42. Re-sequencing also revealed *CHEK1* c.1135 C>T (p.R379X) in the proband of family CF2635, who was diagnosed with breast cancer at age 25. Targeted sequencing of 773 ovarian cancer patients yielded two additional *CHEK1* truncation carriers, one with CHEK1 p.Q346X and one with CHEK1 p.R379X. Both *CHEK1* mutations truncate the protein downstream of the N-terminal kinase domain.

No truncating mutations in *CHEK1* were present in any of the 1152 controls evaluated by MIPs. MIP coverage of *CHEK1* was particularly poor. 30% of targeted bases were not captured despite several rounds of optimization. Therefore I re-sequenced *CHEK1* exons 9 and 10 with PCR and

Sanger sequencing in all cases and controls as MIPs perpetually missed this region. The EVS includes one truncating mutation in *CHEK1,* CHEK1 p.Q346X, found in 1 of 6500 individuals. On the EVS, *CHEK1* has an average coverage of 93x.

Overall, *CHEK1* truncations are seen in 5 of 2544 cases and 1 of 7652 controls. This difference is statistically significant, with P = 0.004. Interestingly, 4 of the 5 *CHEK1* truncations are seen in ovarian cancer patients or in families with both breast and ovarian cancer.

(a)



(b)

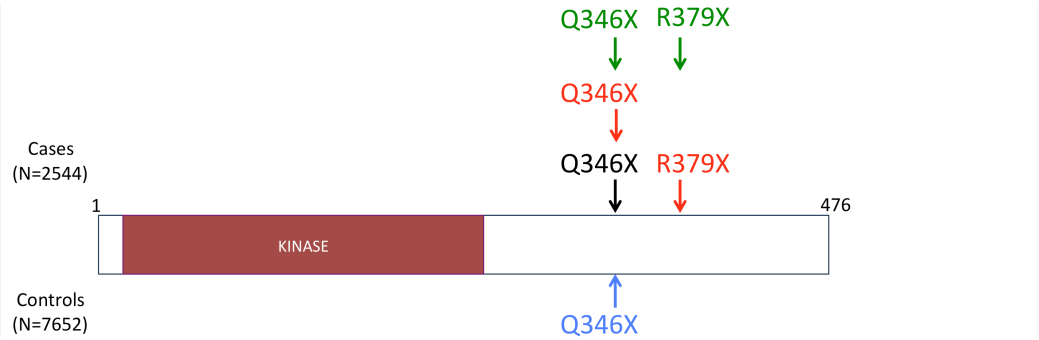


Figure 2. Families and controls with *CHEK1* mutations
(a) Pedigrees of cancer families with *CHEK1* variants
(b) Protein diagram of CHEK1 with variants found by exome sequencing (black), BROCA of cancer families (red), BROCA of ovarian case series (green), and controls (blue).

2.4 *GEN1*

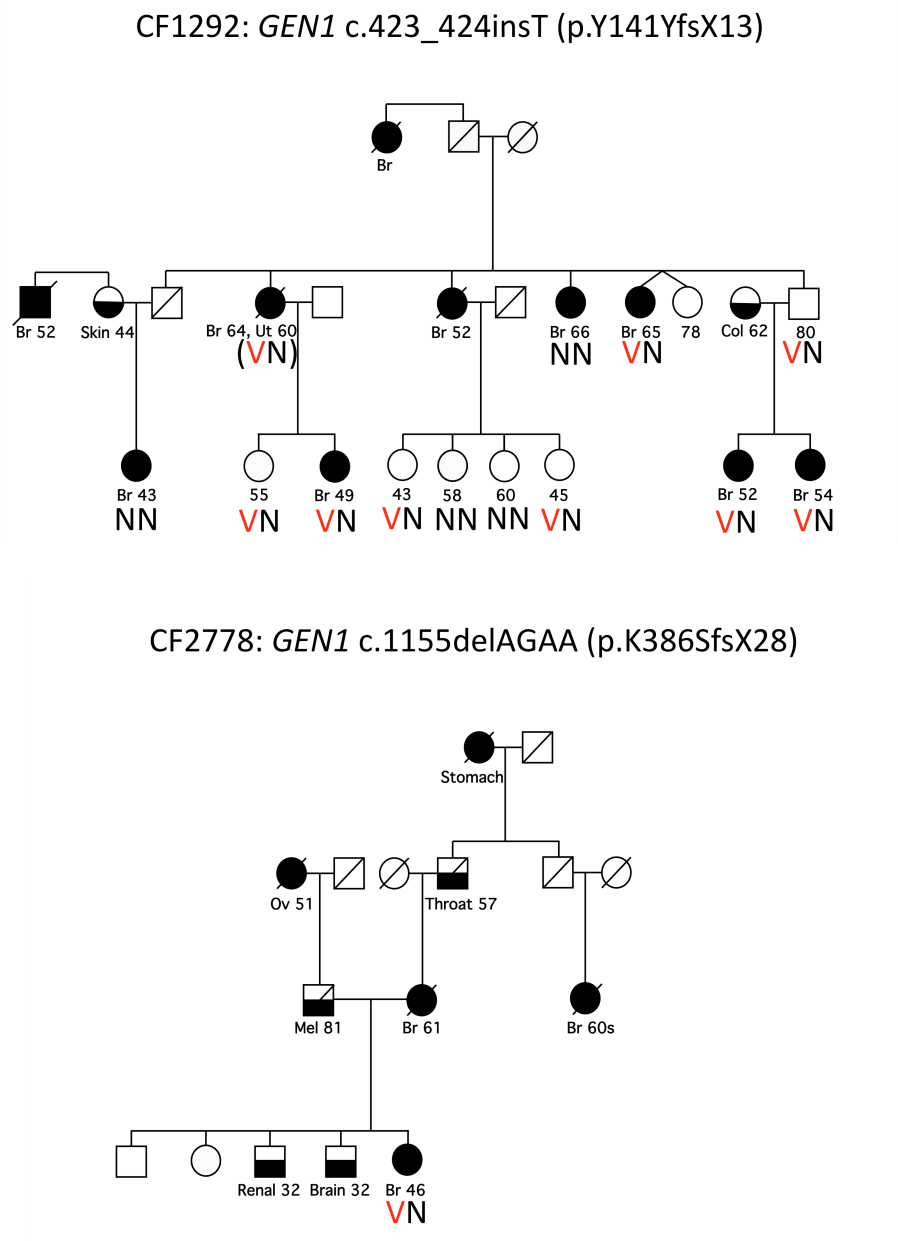
In family CF1292 exome sequencing revealed a truncating mutation, *GEN1 c.423_424insT* (p.Y141YfsX13). The mutation is present in 6 of the 8 breast cancer cases in the family for whom genotypes could be determined (Figure 3). *GEN1* is a resolvase that binds specifically to Holliday junctions during homologous recombination. *GEN1* functions as a dimer to cleave at symmetrical locations in the Holliday junction(58). There is some redundancy in the resolution of double stranded breaks, but each enzyme recognizes a specific DNA structure(59). Studies in *C. elegans* suggest that *GEN1* may have a secondary function in the initial response to DNA damage, acting in parallel to *ATR*/*ATM* and *CHEK1*(60).

Because of its involvement in the double stranded break repair pathway and the presence of somatic *GEN1* mutations in breast cancers, *GEN1* was previously sequenced for germline mutations in 176 breast cancer cases(61). One C-terminal frameshift was identified (c.2515_2519delAAGTT) but was present in 4% of cases and in 4% of controls.

Like *ATR* and *CHEK1*, *GEN1* was added to BROCA v.4 and sequenced in 1177 familial breast cancer cases. An additional 540 cases were sequenced with MIPs with 6240 bases tiled across *GEN1*. Average coverage of *GEN1* was 190x, with 12% bases not covered. Targeted sequencing of these 1717 cases revealed a second truncation, *GEN1 c.1155delAGAA* (p. K386SfsX28) in the proband of family CF2778 (Figure 3). Both frameshifts are in the N-terminal half of the protein: *GEN1 c.423_424insT* disrupts the endonuclease domain, and *GEN1 c.1155delAGAA* eliminates over half of the protein (Figure 3).

Re-sequencing *GEN1* in 1152 controls did not reveal any truncations. Three rare *GEN1* truncations are present on the EVS: *GEN1 c.1771G>T* (p.E591X), *GEN1 c.531_532insC* (p.H178PfsX3), and *GEN1 c.1929_1932del4* (p.K645CfsX29), each present in 1 of 6500 individuals. Overall, rare truncations were identified in 2 of 1717 cases and in 3 of 7652 controls. This difference was not statistically significant (P = 0.21).

(a)



(b)

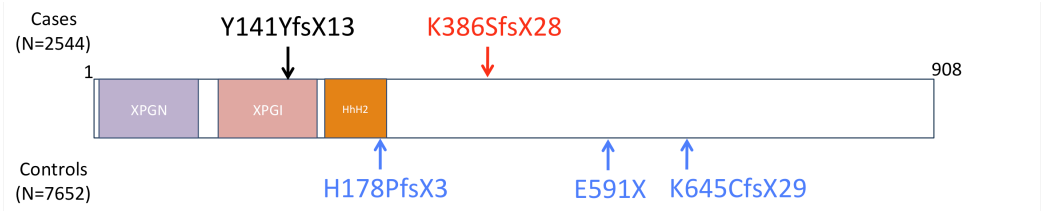


Figure 3. Families and controls with *GEN1* mutations
(a) Pedigrees with *GEN1* variants
(b) Protein diagram of GEN1 with variants found by exome sequencing (black), BROCA of cancer families (red), and controls (blue)

2.5 Additional candidate genes in DNA damage response pathways

ATRIP and *FANCL* are intriguing candidate genes because of their involvement in DNA damage response pathways. One truncating mutation was found in each gene, both mutations near the C-terminus. In family CF1888, *ATRIP* c. 2257delG (p. P753PfsX12) is present in a proband diagnosed with breast cancer at age 49 and her mother diagnosed with breast cancer at age 55. The proband's aunt had ovarian cancer at age 61, and subsequently passed away; we do not have the aunt's DNA. ATRIP is the binding partner of ATR and is essential for the recruitment of ATR to sites of DNA damage(62)(63). In family CF1329, *FANCL* c.1114 ins TAAT (p.T372IfsX12) alters the last 8 amino acids of the protein, 7 of which are highly conserved. The variant was present in 5 of the 8 breast cancer cases whose genotypes could be determined. *FANCL* is a ubiquitin ligase that acts in a complex of Fanconi Anemia proteins in the homologous recombination pathway(64). *ATRIP* and *FANCL* genes have not yet been sequenced in additional breast cancer cases.

2.6 Remaining families

Of the 54 families evaluated by exome sequencing, only 3 were solved. Many other families contained 10 or more damaging variants that appeared in at least 2 breast cancer cases. These variants either did not co-segregate with cancer in the family as a whole, or were truncations very near an unconserved C-terminus, or had multiple truncations listed on the EVS. However by re-sequencing additional unrelated cases, it is possible that additional mutations might be identified in some of these genes.

2.7 Methods

Subjects. Cases for this project are breast cancer patients and their families enrolled in our lab since 1974. In each family, at least four relatives have developed breast cancer. Contact and counseling of participating families is the responsibility of a certified genetic counselor, Jessica Mandell, MS, CGC. Participants are enrolled on University of Washington IRB protocol 95-1216-G/E, and have consented to their DNA being used for discovery of new breast cancer genes. Pedigree information, clinical information, DNA, RNA, and lymphoblast cell lines were already in place for this project. No families were recruited specifically for the present project.

Preparation of DNA. From each participant, blood was drawn into two 10ml ACD vacutainer tubes. DNA was extracted directly from one tube by a standard salting out procedure. The second tube was used for EBV immortalization of lymphoblasts in the King lab. All DNA samples were evaluated by agarose gel electrophoresis and Nanodrop ND100. All samples are high molecular weight DNA without degradation, are free of RNA and have $A_{260}/A_{280}>1.8$ and $A_{260}/A_{230}>1.5$

Library construction. Genomic DNA from each sample (1 μg) was sheared (Covaris AFA) in 55 μl resuspension buffer(RSB, Illumina) using the settings: duty cycle 10%, intensity 5, and cycle/burst 200 for 120 sec. Fragmented DNA ends were repaired for 30 min at 30°C with 40 μl End Repair Mix in a total volume of 100 μl (Illumina) and eluted in 17.5 μl RSB after clean-up. A-tails were then added to the end-repaired DNA at 37°C for 30 mins in a total volume of 30 μl

(A-tailing Mix, Illumina). Indexing adaptors were ligated to the A-tailed fragments at 30°C for 10 min in a total volume of 37.5 µl (Ligation Mix and Adapter Index). The reactions were stopped by adding Stop Ligation Buffer. Samples were eluted in 25 µl of water after clean-up. All clean-up steps were performed with 1.8X AmpureXP beads as directed by Agencourt. The adaptor-ligated fragments were PCR amplified according to the Nimblegen SeqCap EZ Human Exome Library v2.0 protocol, each in a total volume of 50 µl. Each library was purified following the Ampure XP protocol and run on a high-sensitivity chip on the Agilent 2100 Bioanalyzer to determine the library concentration.

Library capture and sequencing. Each library (1µg) was hybridized to SeqCap EZ Exome probes (v2.0, Nimblegen) according to manufacturer's protocols and blocked with 100 µl of 1 mg/ml human COT1 DNA (Invitrogen) and 10 µl of both SLXA_Pair_For_Amp (100 µM) and SLXA_Pair_Rev_Amp (100 µM). The hybridized library was captured and washed as directed by Nimblegen and eluted in 50 µl of water. The enriched library was PCR amplified in a total volume of 50 µl (4 µl library, 1X iProof High Fidelity Master Mix and 0.625 µM of both SLXA_Pair_For_Amp and SLXA_Pair_Rev_Amp). The PCR conditions were: 98°C for 30 sec, 20 cycles of 98°C for 10 sec, 60°C for 30 sec and 72°C for 30 sec, followed by a final 72°C for 5 min. Each library was purified with Agencourt beads, and concentration determined on the 2100 Bioanalyzer. One lane of 76 bp paired-end reads was generated for 3-4 samples on an Illumina HiSeq using the Illumina pipeline v1.6. Illumina sequencing was performed by Anne Thornton and Tom Walsh.

Exome target. The exome includes all unique, well-annotated protein coding regions (including

flanking regions for exons smaller than 100 bp) from the NCBI Reference Sequence (Jan 2010), CCDS (Sept 2009) as well as miRNAs from miRBase (v.14) and additional customer inputs. The total target size is 36.5 Mb. However, the total size of all regions covered by probes is 44.1 Mb and includes some flanking regions.

Sequence Alignment. We used MAQ v0.7.1 to align the sequence reads to the February 2009 human reference sequence of the UCSC Genome Browser (GRCh37/hg19). An average of 5.7 Gb of sequence aligned to the exome target. The proportion of the entire targeted exome covered by > 10 reads of Q30 or greater quality was 93.1%. DNA variants were filtered against dbSNP137 and Phase 3 of the 1000 Genomes project (March 2010 data release), then classified by predicted function to include all missense, nonsense, frameshift, or splice-site alleles. Given that dbSNP includes both disease-associated and benign alleles, we included in the analysis known SNPs identified by dbSNP as clinically associated.

Validation. Primers were designed with Primer3(65), ordered from IDT. PCR was performed with Invitrogen Taq and Buffers. Sequencing chemistry performed with BigDye (Applied Biosystems) and Better Buffer, sequenced on an ABI 3130. Sequencing traces were aligned and viewed with Sequencher.

Exome Variant Server. Variant results and gene coverage data were downloaded from the EVS(51) August 2013. Limitations of the EVS are described in section 2.1.

Molecular Inversion Probes (MIPs). Molecular inversion probes were designed to all exons of *GEN1*, *ATR*, and *CHEK1*, including 10 intronic base pairs at each splice junction. 70 bp oligos at 100uM were ordered from IDT at 25nmole scale. Oligos were pooled (5ul each) and a 50ul aliquot was kinase treated with 10ul T4 DNA ligase buffer and 5ul T4 PNK in a total volume of 100ul for 45 minutes at 37°C followed by 20 minutes at 65°C. MIP capture was performed with 100ng DNA in 22.4ul autoclaved water, 2.5ul 10x Ampligase Buffer (Epicentre), kinase-treated MIPs (diluted 1:100), 0.04ul 0.10mM dNTPs, 0.04ul 10U/ul Stoffel (Applied Biosystems), and 0.01ul 100U/ul Ampligase (Epicentre) in a total reaction volume of 25ul. Samples were heated to 95°C for 10 min with the lid tracking 10°C then lowered to 60°C at 0.1°C/sec ramp. Samples hybridized at 60°C for 72 hours. Non-circularized DNA was eliminated by incubation with ExoI and ExoIII (New England Biolabs) and 10x Ampligase buffer for 30 minutes at 37°C, followed by enzyme deactivation at 95°C for 2 minutes. Barcode PCR was performed on 2.5ul of the MIP reaction with 2.5ul of the 10uM Barcode primer (including the SLXA_PE_MIPBC_RV primer sequence), 12.5ul 2X iProof Master Mix (Bio-Rad), and 0.125ul 100uM SLXA_PE_MIPBC_FOR primer in a total volume of 25ul. The PCR conditions were: 98°C for 30 sec, 15 cycles of 98°C for 10 sec, 60°C for 30 sec and 72°C for 30 sec, followed by a final 72°C for 2min. Each library was pooled and purified with Agencourt beads, and concentration determined on the 2100 Bioanalyzer. Samples were run 384 libraries per lane on the HiSeq2500.

Chapter 3: Discussion and future directions

ATR, *CHEK1,* and *GEN1* are new candidate breast cancer susceptibility genes. These three genes expand the proportion of inherited breast cancer that may be associated with rare variants in DNA damage response pathways. *CHEK1* and *ATR* are compelling candidate genes, for both biological and genetic reasons, and appear to be involved in both breast and ovarian cancer pathogenesis. The overall contribution of mutations in these genes will become clearer as additional unrelated breast and ovarian cancer cases are sequenced by exome and targeted sequencing approaches. *GEN1* should be sequenced in more cases and controls to determine if it contributes to cancer susceptibility.

3.1 Heterogeneity in breast cancer genetics

Inherited breast cancer is extremely genetically heterogeneous with respect to both alleles and loci. Variants that segregate within a family and are not seen in controls are good candidates for disease association. However, as the results of this project indicate, every candidate gene requires detailed analysis.

This study expands the proportion of breast cancer that is associated with rare, severe mutations in DNA repair pathways (Figure 4). Recent exome sequencing studies identified rare mutations in *XRCC2* (31) and *Abraxas* (32), which function in homologous recombination DNA repair.

The heterogeneity of loci is contained within these related pathways; other genes involved in the maintenance of genetic integrity are good candidates for inherited breast cancer susceptibility.

The initial linkage of breast cancer to *BRCA1* was detected in families with young ages at breast cancer onset. Later it became clear that linkage was also significant in families with both breast and ovarian cancer. Linkage of breast cancer to *BRCA2* included families with multiple cases of male breast cancer. In contrast, many as-yet-unresolved families are female-breast-cancer specific and are characterized by diagnosis at older ages, although familial cases as a whole are still younger at diagnosis than cases in the general population. More recent linkage studies suggest that no single gene is likely to account for a large fraction of the remaining high incidence families(66). Genetic heterogeneity, phenocopies, and incomplete penetrance are even more of a challenge than in the past.

This exome sequencing study suggests that large numbers of very rare genetic mutations underlie common human diseases. In particular, the rare alleles model for susceptibility to cancer was supported by a recent study of ovarian carcinoma, which revealed individually rare deleterious germline mutations in 12 different genes (50). An analogous result is seen in the genetics of blood pressure variation, where individually rare loss-of-function mutations converge on renal salt handling genes(67),(68). These results support a rare variant-common disease model with variants converging in a biologically relevant pathway. In addition, as indicated above, common variants implicated in inherited breast cancer by genome-wide association studies do not overlap with *ATR*, *CHEK1* and *GEN1*, or any of the known breast cancer genes.

We solved only 3 of the 54 cancer families that were exome sequenced. A few more families may be solved as we find more genetic evidence for some candidates. Additional evidence could include damaging mutations in the same genes segregating in additional cancer families, tumor loss of heterozygosity, and biological assays implicating specific rare missenses as loss-of-function alleles.

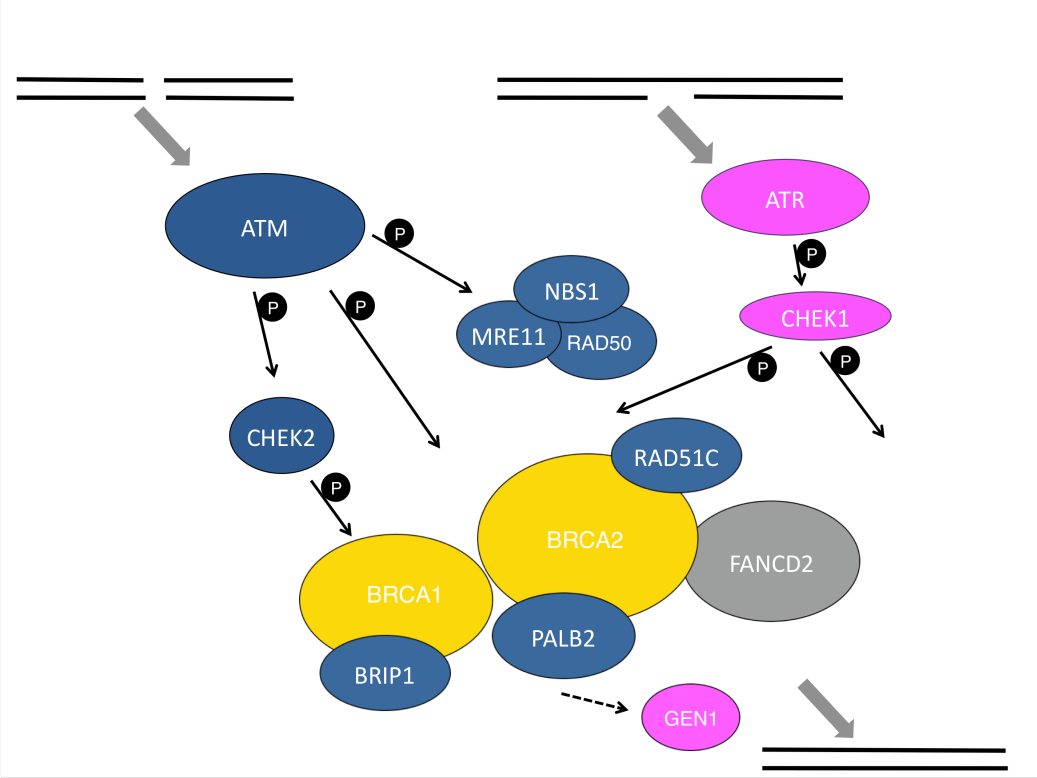


Figure 4. Schematic of DNA repair pathways involved in inherited predisposition to breast cancer. Candidate proteins ATR, CHEK1 and GEN1 are in pink.

3.2 Whole genome v. targeted sequencing

Sequencing costs continue to decrease, making whole genome sequencing a more attractive prospect. BROCA targeted sequencing and exome sequencing are still comparatively inexpensive, but target specific regions of the genome. For the immediate future, specific sequencing technologies will continue to be applied to different types of medical genetics efforts.

BROCA has a high read depth, captures *cis* regulatory regions, and by tiling introns can accurately call even small CNVs. Currently, 96 samples are multiplexed per lane. To achieve high coverage and quality, only a fraction of the genome is sequenced. This trade off of high sample throughput but fewer genomic regions covered is ideal for clinical applications. Results can only be reported for variants whose significance is understood; sequencing the entire genome or exome yields too many variants of unknown significance.

Exome sequencing captures all protein-coding regions of the genome, which have traditionally been easier to study and assign causality. Indeed, approximately 85% of known mutations that cause human disease are in coding sequence(69). By focusing on the coding sequence, exome sequencing captures variants that are more likely to affect protein function, but misses regulatory variation, unannotated genes/RNAs, and many structural variants.

Whole genome sequencing identifies variants in regulatory regions, but these variants may be difficult to interpret. Tools such as promoter/enhancer bashing(70), data on the VISTA enhancer

browser(71), and ENCODE(72) data on the UCSC genome browser will make determining the effect of non-coding mutations more tenable.

The high physical coverage of whole genome sequence facilitates the calling of all structural variants. Exome and BROCA sequencing can currently only reveal CNVs based on read depth, or if the breakpoints fall in targeted sequence. Whole genome sequencing can call insertions and deletions as well as inversions and translocations that do not change the copy number content.

Whole genome sequencing yields high physical coverage, but to identify single nucleotide variants it is essential to have high sequence coverage as well. Studies comparing exome sequencing and whole genome sequencing show that exome sequencing calls more variants in coding sequence(73). While all sequencing methods are less robust in regions of repeats, secondary structure, and high GC content, in targeted sequencing this can be ameliorated by designing more probes to difficult regions and omitting probes from intronic *Alu* and segmentally duplicated regions.

Our pipeline currently uses BROCA as a screen for deleterious variants in all known genes. Families that remain unsolved are next sequenced for all coding variation. In the future, families that remain unsolved after exome sequencing will be subjected to whole genome sequencing. Exome sequencing will likely remain well used for the foreseeable future because it is more cost effective, has more power to detect variants, and it is easier to interpret these coding variants. Targeted sequencing will also remain a cost-effective way to screen large numbers of patients for genetically well-understood disorders.

3.3 Future directions: Missense mutations in breast cancer genes

As medical genetics labs continue to sequence the exomes and genomes of more individuals, a major issue will be determining the functional relevance of rare loss-of-function alleles. The genomes of all individuals contain rare truncations. It is even more difficult to interpret the functional significance of missense mutations and in frame deletions. *In silico* tools are not sufficient to determine whether these changes lead to loss of function *in vivo*.

DNA damage repair genes are well conserved; many have homologues in model organisms. For some tumor suppressor genes, there are functional assays to measure the response to DNA damage. Recently, an *in vivo* assay examined rare missense mutations in *CHEK2* and determined whether each variant conferred a loss of function phenotype(74).

Another strategy uses deep mutational scanning of protein domains to assess the function of all possible single nucleotide changes. In a pool of all protein variants, selection followed by high-throughput sequencing is carried out to determine the functional consequences of each change(75). The challenge for genes with multiple functions will be to develop high-throughput assays that meaningfully evaluate each domain.

3.4 Future directions: Regulatory mutations in breast cancer genes

BROCA sequencing includes 10kb of intergenic sequencing on both the 5' and 3' end of each gene, as well as all intronic variation. This intronic and intergenic data can be used to look for rare variants that segregate with breast cancer in each unsolved family. Rare variants can be analyzed with splice site and splice enhancer software (e.g. Flybase, ESE Finder), then assayed by sequencing cDNA from patient RNA or by creating mini-genes if RNA is not available from a patient.

Rare noncoding variation could be functional if the variant lies in a transcription factor or enhancer-binding site. A second BROCA library made from patient RNA, rather than DNA, could be used to look for allelic imbalance in the transcripts of all captured genes. To find causal regulatory variants, the best present source is the subset of ENCODE regions that have been documented in normal breast tissue (BC_Breast_02-03015), mammary epithelial cells (HMEC), or breast derived tumor cell lines (MCF7, MCF10A-Er-Src, MDA-MB-231).

3.5 Clinical applications

Through exome sequencing, we identified additional candidate breast cancer genes involved in DNA damage response pathways. As relative risk data are collected through additional sequencing studies, prevention and treatment strategies can be extended to women with mutations in these genes.

The discovery of *BRCA1* and *BRCA2* led to major changes in clinical practice. Increased screening and risk-reducing mastectomies and salpingo-oophorectomies have significantly reduced breast cancer incidence in mutation carriers(76–78). Increased surveillance of patients with mutations in other DNA damage response genes will be beneficial as more is learned about the risks associated with mutations in these genes.

Poly(ADP-ribose)polymerase (PARP) inhibitors show promise in *BRCA1/2* mutant tumors. PARP inhibitors use a synthetic lethal strategy to target cells with defects in the homologous recombination DNA repair pathway(79). These therapies may also be effective in patients with mutations in other members of the HR pathway. Inhibitors of other cell-cycle proteins are being tested for efficacy in HR-deficient tumors(80). For example, therapies involving CHEK1 inhibitors show promise in human tumor xenographs and mouse models(81).

This study demonstrated both the power and limitations of next generation sequencing approaches for identifying new candidate breast cancer susceptibility genes. Breast and ovarian cancer patients will benefit as comprehensive genetic testing options become available and as more is learned about the functional consequences of individually rare mutations.

References

1. Howlader N, Noone A, Krapcho M, Garshell J, Neyman N, Altekruse S, et al. SEER Cancer Statistics Review, 1975-2010, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, April 2013.

2. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–9, 1990.

3. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66–71, 1994..

4. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene *BRCA2*. Nature 378:789–92, 1995.

5. Szabo C, Masiello A, Ryan JF, The BIC Consortium, Brody LC. The Breast Cancer Information Core: Database design, structure, and scope. Hum Mutat. 16:123–31, 2000.

6. Malone KE, Daling JR, Doody DR, Hsu L, Bernstein L, Coates RJ, et al. Prevalence and predictors of *BRCA1* and *BRCA2* mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. Cancer Res. 66:8297–308, 2006.

7. Pal T, Permuth-Wey J, Betts JA, Krischer JP, Fiorica J, Arango H, et al. *BRCA1* and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. Cancer 104:2807–16, 2005.

8. Weitzel JN, Lagos VI, Cullinane CA, Gambol PJ, Culver JO, Blazer KR, et al. Limited family structure and *BRCA* gene mutation status in single cases of breast cancer. JAMA 297:2587, 2007

9. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet. 2006 Nov;38(11):1239–41, 2006.

10. Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, et al. Spectrum of mutations in *BRCA1, BRCA2, CHEK2*, and *TP53* in families at high risk of breast cancer. JAMA 295:1379–88.

11. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. *PALB2*, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet. 39:165–7, 2007

12. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. *ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 38:873–5, 2006.

13. Karppinen S-M, Heikkinen K, Rapakko K, Winqvist R. Mutation screening of the *BARD1* gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. J Med Genet. 41:e114, 2004.

14. Zhong Q, Chen CF, Li S, Chen Y, Wang CC, Xiao J, et al. Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. Science 285:747–50, 1999.

15. Heikkinen K. *RAD50* and *NBS1* are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis 27:1593–9, 2005

16. Wang Y, Cortez D, Yazdi P, Neff N, Elledge SJ, Qin J. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev. 14:927–39, 2000.

17. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. J Clin Oncol. 26:542–8, 2008.

18. Seemanova E, Jarolim P, Seeman P, Varon R, Digweed M, Swift M, et al. Cancer risk of heterozygotes with the *NBN* founder mutation. JNCI 99:1875–80, 2007.

19. Erkko H, Dowty JG, Nikkila J, Syrjakoski K, Mannermaa A, Pylkas K, et al. Penetrance analysis of the *PALB2 c.1592delT* founder mutation. Clin Cancer Res. 14:4667–71, 2008.

20. Gonzalez KD, Noltner KA, Buzin CH, Gu D, Wen-Fong CY, Nguyen VQ, et al. Beyond Li-Fraumeni syndrome: clinical characteristics of families with p53 germline mutations. J Clin Oncol. 27:1250–6, 2009.

21. FitzGerald MG, Marsh DJ, Wahrer D, Bell D, Caron S, Shannon KE, et al. Germline mutations in *PTEN* are an infrequent cause of genetic predisposition to breast cancer. Oncogene 17:727–31, 1998.

22. Schrader KA, Masciari S, Boyd N, Wiyrick S, Kaurah P, Senz J, et al. Hereditary diffuse gastric cancer: association with lobular breast cancer. Fam Cancer 7:73–82, 2008.

23. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJP, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res. 12:3209–15, 2006.

24. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 81:214–8, 1999.

25. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in *FANCD1/BRCA2*. J Med Genet. 44:1–9, 2007.

26. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, et al. Biallelic mutations in *PALB2* cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet. 39:162–4, 2007.

27. Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, et al. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat Genet. 37:934–5, 2005.

28. Joenje H, Patel KJ. The emerging genetic and molecular basis of Fanconi anaemia. Nat Rev Genet. 2:446–57, 2001.

29. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish *RAD51C* as a human cancer susceptibility gene. Nat Genet. 42:410–4, 2010.

30. Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, et al. Germline mutations in *RAD51D* confer susceptibility to ovarian cancer. Nat Genet. 43:879–82, 2011.

31. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, et al. Rare mutations in *XRCC2* increase the risk of breast cancer. Am J Hum Genet. 90:734–9, 2012.

32. Solyom S, Aressy B, Pylkäs K, Patterson-Fortin J, Hartikainen JM, Kallioniemi A, et al. Breast cancer-associated *Abraxas* mutation disrupts nuclear localization and DNA damage response functions. Sci Transl Med. 4:122ra23, 2012.

33. Hilbers FS, Wijnen JT, Hoogerbrugge N, Oosterwijk JC, Collee MJ, Peterlongo P, et al. Rare variants in *XRCC2* as breast cancer susceptibility alleles. J Med Genet. 49:618–20, 2012.

34. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet. 33:177–82 2003

35. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 447:1087–93 2007

36. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 39:870–4 2007

37. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A. 105:4340–5, 2008

38. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet. 44:312–8 2012

39. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 461:747–53 2009

40. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 409:860–921 2001

41. Sanger F. Sequences, sequences, and sequences. Annu Rev Biochem. 57:1–28, 1988.

42. Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 11:31–46, 2009

43. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–6, 2008.

44. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, et al. The diploid genome sequence of an Asian individual. Nature 456:60–5, 2008.

45. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 26:1135–45, 2008.

46. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. Parallel, tag-directed assembly of locally derived short sequence reads. Nat Methods 7:119–22, 2010.

47. Turner EH, Ng SB, Nickerson DA, Shendure J. Methods for genomic partitioning. Annu Rev Genomics Hum Genet. 10:263–84, 2009.

48. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. Genome Res. 22:1525–32, 2012.

49. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci. USA 107:12629–33, 2010.

50. Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. Proc Natl Acad Sci USA. 108:18032–7, 2011

51. NHLBI Exome Sequencing Project. Exome Variant Server, cited 1 Aug 2013, at http://evs.gs.washington.edu/EVS/

52. Tanaka A, Weinel S, Nagy N, O'Driscoll M, Lai-Cheong JE, Kulp-Shorten CL, et al. Germline mutation in *ATR* in autosomal- dominant oropharyngeal cancer syndrome. Am J Hum Genet. 90:511–7, 2012.

53. Bakkenist CJ, Kastan MB. Initiating cellular stress responses. Cell. 118:9–17, 2004.

54. Smith J, Tho LM, Xu N, Gillespie DA. The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. Adv Cancer Res. 108:73–112, 2010.

55. Kosoy A, O'Connell MJ. Regulation of Chk1 by its C-terminal Domain. Mol Biol Cell. 19:4546–53, 2008.

56. Tapia-Alveal C, Calonge TM, O'Connell MJ. Regulation of Chk1. Cell Div. 4:8, 2009.

57. Liu Q, Guntuku S, Cui XS, Matsuoka S, Cortez D, Tamai K, et al. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. Genes Dev. 14:1448–59, 2000

58. Rass U, Compton SA, Matos J, Singleton MR, Ip SCY, Blanco MG, et al. Mechanism of Holliday junction resolution by the human GEN1 protein. Genes Dev. 24:1559–69, 2010.

59. Dehé P-M, Coulon S, Scaglione S, Shanahan P, Takedachi A, Wohlschlegel JA, et al. Regulation of Mus81–Eme1 Holliday junction resolvase in response to DNA damage. Nat Struct Mol Biol. 20:598–603, 2013.

60. Bailly AP, Freeman A, Hall J, Déclais A-C, Alpi A, Lilley DMJ, et al. The *Caenorhabditis elegans* homolog of Gen1/Yen1 resolvases links DNA damage signaling to DNA double-strand break repair. PLoS Genet. 6:e1001025, 2010.

61. Breast Cancer Susceptibility Collaboration (UK), Turnbull C, Hines S, Renwick A, Hughes D, Pernet D, et al. Mutation and association analysis of *GEN1* in breast cancer susceptibility. Breast Cancer Res Treat. 124:283–8, 2010.

62. Cortez D, Guntuku S, Qin J, Elledge SJ. ATR and ATRIP: partners in checkpoint signaling. Science 294:1713–6, 2001.

63. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300:1542–8, 2003.

64. Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, et al. A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet. 35:165–70, 2003.

65. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365-86, 2000.

66. Turnbull C, Rahman N. Genetic predisposition to breast cancer: Past, present, and future. Annu Rev Genomics Hum Genet. 29:321–45, 2008.

67. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet. 40:592–9, 2008

68. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. Cell. 104:545–56, 2001 .

69. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 106:19096–101, 2009.

70. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol. 27:1173–5, 2009..

71. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 35:D88–92, 2007. .

72. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74, 2012.

73. Clark MJ, Chen R, Lam HYK, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol. 29:908–14, 2011.

74. Roeb W, Higgins J, King M-C. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet. 21:2738–44, 2012.

75. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods 7:741–6, 2010.

76. Kauff ND, Satagopan JM, Robson ME, Scheuer L, Hensley M, Hudis CA, et al. Risk-reducing salpingo-oophorectomy in women with a *BRCA1* or *BRCA2* mutation. N Engl J Med. 346:1609–15, 2002.

77. Eisen A, Lubinski J, Klijn J, Moller P, Lynch HT, Offit K, et al. Breast cancer risk following bilateral oophorectomy in *BRCA1* and *BRCA2* mutation carriers: an international case-control study. J Clin Oncol 23:7491–6, 2005.

78. Rebbeck TR, Lynch HT, Neuhausen SL, Narod SA, Van't Veer L, Garber JE, et al. Prophylactic oophorectomy in carriers of *BRCA1* or *BRCA2* mutations. N Engl J Med. 346:1616–22, 2002.

79. Banerjee S, Kaye S. PARP inhibitors in *BRCA* gene-mutated ovarian cancer and beyond. Curr Oncol Rep. 13:442–9, 2011.

80. Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: new genetic developments, new therapeutic avenues. Hum Genet. 124:31–42, 2008.

81. Walton MI, Eve PD, Hayes A, Valenti MR, De Haven Brandon AK, Box G, et al. CCT244747 is a novel potent and selective CHK1 inhibitor with oral efficacy alone and in combination with genotoxic anticancer drugs. Clin Cancer Res 18:5650–61, 2012.

Appendix A
Table of all variants found in cancer families

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
43	2	111,430,237	C > A	BUB1	splice	1	6
43	2	202,025,221	A > G	CFLAR	H287R	4	4
43	3	50,103,769	C > G	RBM6	P926R	7	1
43	3	52,554,531	G > A	STAB1	R1872H	6	2
43	4	170,321,414	A > C	NEK1	I1242S	4	4
43	15	41,796,352	C > A	LTK	E752X	6	1
43	16	81164184	del 1	PKD1L2	del G	6	1
43	17	26,918,827	del 2	SPAG5	del AA	1	7
43	19	50,983,413	C > T	C19orf63	R115X	5	2
55	6	57,013,135	del 2	ZNF451	del TG	6	1
57	1	161,140,217	C > G	PPOX	P339A	2	4
57	3	50,326,292	C > T	IFRD2	R388H	2	4
57	5	149,391,845	C > T	HMGXB3	P280S	5	2
57	6	51,944,735	del 1	PKHD1	del C	1	5
57	6	142,723,781	A > G	GPR126	Q590R	6	0
57	6	148,867,277	A > G	SASH1	M1165V	8	0
57	7	8,124,623	G > T	GLCCI1	R425L	7	0
57	7	15,652,188	C > T	MEOX2	G248S	7	0
57	12	123,687,253	G > T	MPHOSPH9	Q417K	2	4
57	15	48,580,665	G > C	SLC12A1	R949T	2	4
57	19	38,385,154	del 2	WDR87	del CT	1	5
57	20	44,416,546	G > C	WFDC3	P96R	6	1
57	20	48,156,148	T > C	PTGIS	D213G	6	1
57	X	71,830,942	C > T	PHKA1	R828H	2	4
61	1	12,353,624	C > T	VPS13D	R1966C	3	2
61	1	12,726,169	G > A	AADACL4	R216Q	3	2
61	1	15,639,551	A > G	FHAD1	splice	3	2
61	1	27,684,750	G > A	MAP3K6	P946L	3	2
61	1	45,295,296	C > T	PTCH2	R358H	3	2
61	2	219,146,842	G > C	TMBIM1	P8R	2	3
61	2	219,290,475	C > A	VIL1	H96Q	2	3
61	5	66,462,494	C > A	MAST4	A2496D	4	1
61	5	180,477,254	del 1	BTNL9	del C	3	2
61	6	139,202,128	G > A	ECT2L	R567Q	4	1
61	9	6,255,967	G > C	IL33	splice	3	2

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
61	10	117,309,039	G > A	ATRNL1	R1263Q	4	1
61	12	6,691,312	A > C	CHD4	I1502M	4	1
61	12	7,520,724	G > A	CD163L1	T1381M	4	1
61	12	31,131,551	C > T	TSPAN11	A111V	2	3
61	14	61,997,170	A > G	PRKCH	M540V	3	2
61	17	45,822,209	A > G	TBX21	Q362R	3	2
61	19	38,996,525	G > A	RYR1	R2827K	2	3
62	1	40,319,722	del 1	TRIT1	del G	2	3
62	1	62,733,958	del 1	KANK4	del C	1	4
94	1	32,669,918	G > A	CCDC28B	G156R	4	0
94	1	147,095,669	C > T	BCL9	R1066X	3	1
94	2	85,922,515	del 1	GNLY	del C	4	0
94	3	81,698,005	A > G	GBE1	splice	3	1
94	3	113,729,787	del 1	KIAA1407	del G	2	2
94	3	126,218,903	C > T	UROC1	A478T	2	2
94	6	28,120,089	G > T	ZNF192	W236C	4	0
94	7	107,664,505	del 1	LAMB4	del T	3	1
94	8	37,696,528	G > A	GPR124	V777M	4	0
94	9	140,444,628	ins 1	PNPLA7	ins T	3	1
94	12	75,804,262	del 1	GLIPR1L2	del T	3	1
94	14	39,703,406	del 1	MIA2	del A	4	0
94	16	711,429	C > T	WDR90	L1211F	4	0
94	16	2,296,877	C > T	DCI	G94S	4	0
94	16	29,756,110	G > A	C16orf54	R55C	4	0
265	1	15,642,964	G > A	FHAD1	splice	2	2
265	1	175,049,319	C > T	TNN	Q269X	3	1
265	2	68,805,143	del 1	APLF	del A	2	2
265	7	106,876,918	G > A	COG5	R700X	4	4
265	11	103,036,712	del 1	DYNC2H1	del A	2	2
265	12	57,861,159	ins 2	GLI1	ins TC	2	2
265	15	41,799,440	del 2	LTK	del GG	3	1
265	17	6,663,899	G > T	XAF1	E115X	3	1
448	1	12,837,658	C > A	PRAMEF12	C456X	1	1
448	1	55,252,738	C > T	TTC22	W250X	1	1
448	8	39,442,793	A > G	ADAM18	splice	1	1
448	8	110,493,660	A > G	PKHD1L1	splice	1	1
448	10	27,687,535	ins 1	PTCHD3	ins T	2	0

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
448	12	91,347,582	ins 3	C12orf12	ins TCT	2	0
448	16	89,169,034	G > A	ACSF3	W230X	2	0
448	19	57,956,741	ins 1	ZNF749	ins A	1	1
450	1	26,514,762	C > T	CNKSR1	Q503X	3	1
450	3	49,935,645	G > C	MST1R	splice	3	1
450	6	139,591,591	A > G	TXLNB	splice	3	2
450	6	150,343,126	G > T	RAET1L	Y113X	1	3
450	7	20,698,222	del 1	ABCB5	del A	3	1
450	15	23,006,504	del 2	NIPA2	del AG	2	1
450	19	11,917,219	C > T	ZNF491	R151X	3	1
450	20	20,269,334	C > T	C20orf26	R967X	1	2
450	22	39,482,295	del 2	APOBEC3G	del AC	3	0
470	2	1,440,061	del 1	TPO	del C	2	1
470	5	179,222,858	G > A	LTC4S	splice	3	0
470	8	1,626,461	ins 3	DLGAP2	ins CAC	3	0
470	8	39,581,344	del 1	ADAM18	del T	2	1
470	9	101,782,725	G > T	COL15A1	splice	3	0
470	16	29,821,424	del 6	MAZ	del GCGGCA	3	0
515	2	11,725,989	A > G	GREB1	N373D	5	0
515	5	41,909,884	A > C	C5orf51	L82I	5	0
515	5	149,375,886	C > T	TIGD6	R9H	5	0
515	7	6,054,834	C > T	AIMP2	R65C	5	0
515	11	64,535,270	A > C	SF1	G372V	5	0
515	12	51,693,021	C > T	BIN2	E190K	5	0
515	12	53,070,240	A > G	KRT1	R432C	5	0
515	14	92,439,151	G > T	TRIP11	P1877T	5	0
515	22	42,611,145	del 6	TCF20	del CCACTG	5	0
548	1	20,517,583	C > T	UBXN10	R177X	1	1
548	1	24,768,620	del 2	NIPAL3	del CT	2	0
548	7	141,366,203	A > G	KIAA1147	M235T	2	0
548	9	271,626	G > T	DOCK8	splice	2	0
548	14	103,390,316	G > A	AMN	W69X	2	0
548	X	151,999,254	T > C	CETN2	M1V	2	0
590	1	40,349,142	G > A	TRIT1	R8X	4	1
590	11	8,959,233	ins 2	ASCL3	ins TT	4	1
590	19	45,856,554	del 2	ERCC2	del AA	1	4

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
737	3	38,349,147	del 4	SLC22A14	del TTTG	1	3
737	3	108,672,558	C > A	GUCA1C	E18X	1	5
737	4	54,327,103	ins 4	LNX1	ins TAGT	3	3
737	5	148,989,259	G > A	FLJ41603	splice	1	4
737	10	91,177,230	C > T	IFIT5	R92X	4	2
737	13	32,910,477	del 5	BRCA2	del AAAAG	3	3
737	20	47,846,885	ins 4	DDX27	ins AGGT	1	5
737	22	29,085,155	del 1	CHEK2	1639 del C	1	5
758	1	100,681,587	A > G	DBT	S242P	3	1
758	2	179,539,084	del 3	TTN	del TTC	2	2
758	5	41,181,567	del 1	C6	del T	3	1
758	8	145,114,782	G > A	OPLAH	R52C	3	1
758	8	145,745,962	C > T	LRRC14	R224C	3	1
758	10	21,097,546	G > A	NEBL	S885F	3	1
758	11	86,123,537	G > A	CCDC81	E443K	3	1
758	11	124,789,743	del 3	HEPN1	del AGG	2	2
758	13	49,956,983	T > C	CAB39L	K22E	3	1
758	18	77,894,865	C > A	ADNP2	N523K	3	1
758	19	57,956,492	del 2	ZNF749	del AG	2	2
758	21	33,726,411	del 1	URB1	del A	2	2
779	2	99,172,078	del 15	INPP4A	del GCTGCA GAAGGA GCG	2	3
779	3	108,072,387	G > A	HHLA2	V60I	5	0
779	3	186,335,139	G > A	AHSG	W191X	2	3
779	4	174,235,304	C > A	GALNT7	P529T	5	0
779	5	115,426,819	C > T	COMMD10	A59V	4	1
779	11	33,080,554	G > C	TCP11L1	A230P	4	1
779	11	60,264,978	del 1	MS4A12	del A	1	5
779	12	10,124,287	G > A	CLEC12A	splice	1	5
779	14	91,409,528	A > C	RPS6KA5	W171L	4	1
779	15	90,245,192	G > A	WDR93	W72X	1	5
779	16	56,487,186	A > G	OGFOD1	M56V	4	1
779	16	56,867,318	C > T	NUP93	L513F	4	1
779	16	75,200,720	del 3	ZFP1	del AAC	4	1
779	16	84,213,211	G > A	TAF1C	A555V	4	1

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
779	16	84,476,117	C > T	ATP2C2	A438V	4	1
779	20	60,498,519	A > G	CDH4	Y462C	4	1
823	3	75,786,848	del 2	ZNF717	del TG	3	3
823	6	51,483,962	G > A	PKHD1	Q4048X	1	5
823	9	123,937,512	ins 1	CEP110	ins A	3	3
823	12	3,939,163	C > A	PARP11	E14X	2	5
823	13	37,441,424	del 1	SMAD9	del G	2	5
823	13	43,597,815	A > C	DNAJC15	E18A	5	1
823	15	49,620,801	del 21	C15orf33	del TGATGATGGTGATGATGATGA	2	5
823	19	48,631,184	C > T	LIG1	T639A	5	1
823	19	49,344,555	A > G	PLEKHA4	R586W	5	1
910	3	170,204,096	G > A	SLC7A14	T275I	4	4
910	5	36,972,087	C > A	NIPBL	P273H	5	3
910	9	74,324,240	G > A	TMEM2	R916C	3	4
910	9	86,292,683	C > A	UBQLN1	G355V	6	3
910	9	130,938,639	G > C	CIZ1	T565R	4	4
910	11	48,166,641	G > A	PTPRJ	R959H	5	3
910	15	89,862,208	C > T	POLG	G1076D	6	2
910	22	22,293,921	del 3	PPM1F	del TCC	4	5
923	17	995,067	G > A	T123T	ABR	9	0
923	17	4,009,040	G > A	F447F	ZZEF1	9	0
923	17	19,584,830	C > T	splice	SLC47A2	9	0
923	17	29,119,490	C > T	P309P	CRLF3	9	0
923	17	37,682,099	T > C	splice	CRKRS	9	0
944	2	101,652,507	G > A	TBC1D8	R511X	3	2
944	2	167,333,973	del 1	SCN7A	del A	3	2
944	2	234,237,188	C > T	SAG	R193X	2	3
944	2	142,012,092	T > C	LRP1B	splice	3	2
944	5	40,945,363	del 11	C7	del ACGTCGACAGA	3	2
944	12	133,331,690	G > A	ANKLE2	R71X	3	2
944	15	48,052,104	T > A	SEMA6D	splice	2	3
944	17	16,338,283	C > T	TRPV2	splice	3	2

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
944	20	52,775,647	del 8	CYP24A1	del TTAGACTG	3	2
963	1	145,075,809	G > T	PDE4DIP	C18X	1	3
963	1	92,446,372	A > G	BRDT	splice	2	2
963	2	20,870,706	ins 1	GDF7	ins G	2	2
963	3	51,970,273	C > T	RRP9	splice	3	1
963	4	72,897,699	del 1	NPFFR2	del G	2	2
963	9	78,848,442	ins 4	PCSK5	ins CCAT	2	2
963	12	56,091,287	del 8	ITGA7	del GAGCAATA	2	2
963	13	60,590,205	A > G	DIAPH3	splice	3	1
963	19	7,553,859	G > T	PEX11G	S13X	4	0
963	20	25,657,232	del 1	ZNF337	del G	3	1
963	20	2,464,109	G > A	ZNF343	R500X	3	1
985	1	62,740,565	G > A	KANK4	R71X	2	3
985	9	123,912,468	del 4	CEP110	del AGTC	2	2
997	5	40,852,813	del 1	CARD6	del T	1	3
997	7	150,171,600	del 35	GIMAP8	del AGATATAGTGCCTTCAACTACCGGGCAA	4	0
997	7	154,143,297	A > G	DPP6	splice	3	1
998	2	99,439,319	del 24	C2orf55	CCGCTCCTCTCTCGGGCTCCGTCT	2	1
998	12	118,016,956	del 1	KSR2	C	2	1
998	12	974,309	ins 1	WNK1	C	1	3
998	17	10,416,943	del 2	MYH1	AT	2	1
1026	5	40,841,572	del 2	CARD6	del TT	1	3
1026	7	2,373,120	del 1	C7orf46	del G	1	3
1026	7	116,339,295	C > T	MET	Q53X	1	3
1026	12	72,068,116	T > C	THAP2	R69X	1	3
1026	15	45,436,245	G > A	DUOX1	Q655X	1	3
1026	19	51,857,618	del 1	ETFB	del G	4	1
1036	2	220,432,786	ins 1	OBSL1	ins T	1	4

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
1036	4	145,629,407	C > A	HHIP	F415L	4	1
1036	7	120,773,875	A > T	C7orf58	splice	3	2
1036	11	6,588,881	C > T	DNHD1	R4048X	3	2
1036	14	94,962,708	A > T	SERPINA12	splice	3	2
1036	15	41,247,952	C > T	CHAC1	Q259X	3	2
1036	17	48,742,513	G > T	ABCC3	splice	3	2
1041	1	218,504,400	del 1	RRP15	del C	5	2
1041	2	25,972,836	C > A	ASXL2	S530I	3	3
1041	2	48,030,603	C > T	MSH6	P1073S	6	3
1041	2	54,120,878	G > A	PSME4	T1324I	6	3
1041	3	16,358,700	G > A	RFTN1	Q458X	4	2
1041	6	116,436,913	G > T	NT5DC1	V142L	4	2
1041	11	102,711,262	C > G	MMP3	A230P	7	2
1041	11	108,098,545	A > G	ATM	T39A	1	4
1085	1	55,277,538	G > A	C1orf177	W184X	2	2
1085	2	166,847,822	C > T	SCN1A	R1988Q	3	1
1085	2	220,424,161	del 15	OBSL1	del CAAGGT CACGGC GAT	3	1
1085	3	52,662,944	T > C	PBRM1	Y470C	3	1
1085	7	88,965,970	C > T	ZNF804B	S1225F	3	1
1085	7	151,860,796	G > A	MLL3	P3289L	3	1
1085	8	144,413,499	C > T	TOP1MT	E45K	2	1
1085	8	144,942,321	C > T	EPPK1	V1701M	2	2
1085	8	145,113,568	C > T	OPLAH	G205D	2	2
1085	10	75,335,403	C > A	USP54	R5I	2	2
1085	10	115,349,499	A > C	NRAP	S1672A	3	1
1085	18	2,929,103	G > A	LPIN2	L504F	3	1
1085	19	11,034,678	ins 1	YIPF2	ins G	4	0
1085	19	14,755,002	A > G	EMR3	L323P	3	1
1085	19	16,859,994	C > T	NWD1	R46W	3	1
1085	19	38,379,447	G > A	WDR87	R1583X	2	1
1085	19	49,407,626	G > T	NUCB1	R53L	3	1
1105	1	179,782,232	G > A	FAM163A	splice	3	3
1105	2	242,144,350	ins 1	ANO7	ins G	4	2
1105	4	159,136,408	del 1	TMEM144	del T	3	3
1105	8	61,693,943	ins 6	CHD7	ins	3	3

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
					AAAGCA		
1105	8	139,701,173	T > G	COL22A1	E950A	2	4
1105	11	108,199,929	T > G	ATM	V1076G	3	2
1105	11	111,594,604	del 3	SIK2	del CTC	3	3
1105	19	10,398,009	G > A	ICAM4	W107X	2	4
1105	X	69,420,288	C > T	DGAT2L6	R151X	4	2
1152	6	119,344,211	C > T	FAM184A	G230E	6	0
1152	7	130,354,024	C > T	TSGA13	splice	2	3
1152	7	107,434,307	T > C	SLC26A3	K51E	6	0
1152	10	129,914,266	T > A	MKI67	K136X	3	2
1152	15	59,910,679	T > C	GCNT3	V81A	6	1
1152	17	3,655,111	G > A	ITGAE	R576C	6	0
1152	17	3,656,597	C > T	ITGAE	R552H	6	0
1152	17	7,496,497	G > T	FXR2	P445T	6	0
1249	3	184,429,558	ins 1	MAGEF1	C	1	5
1249	3	186,938,920	A > C	MASP1	splice	3	2
1249	7	44,621,766	del 1	TMED4	G	1	5
1249	10	129,537,125	C>T	FOXI2	Q285X	2	4
1249	10	133,946,942	C > T	JAKMIP3	R254X	3	3
1249	11	8,959,233	ins 2	ASCL3	TT	5	1
1249	12	23,757,468	C > T	SOX5	splice	3	3
1249	14	74,970,172	C > G	LTBP2	splice	3	3
1249	16	79,245,682	C > T	WWOX	Q412X	3	3
1249	17	18,164,452	ins 1	SMCR7	G	4	2
1249	18	14,105,272	del 1	ZNF519	G	2	4
1249	19	56,473,544	ins 1	NLRP8	T	2	4
1249	20	3,845,089	del 2	MAVS	AG	2	4
1257	2	64,863,751	C > T	SERTAD2	M86I	4	0
1257	5	50,125,738	A > G	PARP8	M733V	3	1
1257	5	153,432,717	G > A	MFAP3	R33H	4	0
1257	9	115,456,447	A > C	SOSSC	M31R	4	0
1257	12	105,425,570	C > T	ALDH1L2	G803E	3	1
1257	14	74,325,603	G > C	PTGR2	splice	4	0
1257	17	3,975,988	A > T	ZZEF1	W1258R	4	0
1257	17	17,168,247	C > T	COPS3	V166M	4	0
1257	17	41,234,451	C > T	BRCA1	R1443X	2	1
1257	22	30,683,242	G > A	GATSL3	T137M	4	0

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
1292	1	78,399,087	C > T	NEXN	R328X	3	5
1292	2	17,946,238	ins 1	GEN1	ins T	4	1
1292	2	29,240,023	C > T	FAM179A	Q350X	4	3
1292	3	187,088,654	G > A	RTP4	W78X	3	4
1329	2	58,386,929	ins 4	FANCL	ins TAAT	6	3
1329	5	78,347,133	del 4	DMGDH	GCTC	5	4
1439	1	226,047,037	del 2	TMEM63A	del AG	3	1
1439	8	94,777,641	C > T	TMEM67	R91X	3	0
1439	9	6,329,011	del 1	TPD52L3	del G	2	1
1439	11	289,916	C > T	ATHL1	R34X	2	1
1439	12	112,194,135	A > G	ACAD10	splice	3	1
1439	19	44,932,911	ins 1	ZNF229	ins T	4	0
1604	1	32,098,150	G > A	PEF1	Q191X	5	0
1604	1	161,681,140	C > G	FCRLA	N142K	4	1
1604	3	128,706,510	G > A	KIAA1257	R206X	1	4
1604	5	180,057,098	G > A	FLT4	S174L	5	1
1604	5	131,706,028	G > T	SLC22A5	D122Y	4	1
1604	5	140,754,878	C > G	PCDHGA6	L410V	4	1
1604	11	118,886,015	A > G	CCDC84	splice	1	4
1604	12	57,178,801	G > A	HSD17B6	splice	1	4
1604	12	125,599,025	C > G	AACS	G306G	5	0
1604	16	67,961,489	C > T	PSKH1	R407W	4	1
1604	16	71,317,899	A >G	FTSJD1	V642A	4	1
1604	17	4,718,769	A > T	PLD2	splice	1	4
1604	17	1,657,676	C > T	SERPINF2	R442W	4	1
1604	17	60,061,640	T > C	MED13	Y927C	4	1
1604	X	129,149,813	T > C	BCORL1	I1022T	4	1
1760	1	153,314,171	ins 1	PGLYRP4	ins T	2	5
1760	1	213,046,107	C > A	FLVCR1	S324X	3	4
1760	4	104,104,413	G > A	CENPE	R260X	2	5
1760	5	159,854,837	del 1	PTTG1	del C	2	5
1760	6	88,273,900	del 2	RARS2	del TT	3	4
1760	6	117,113,940	del 1	GPRC6A	del C	6	1
1760	8	124,714,911	T > A	ANXA13	K94X	3	4
1760	9	139,338,286	G > A	SEC16A	Q2332X	2	5
1760	12	110,340,895	C > T	TCHP	R22X	3	4
1760	17	19,584,702	del 1	SLC47A2	del T	3	4

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
1760	19	15,132,683	del 1	CCDC105	del C	6	1
1760	19	45,855,809	del 1	ERCC2	del G	5	1
1760	20	44,054,446	C > T	PIGT	R471X	3	4
1778	9	114,842,440	G > A	SUSD1	Q494X	2	2
1778	9	134,394,332	C > T	POMT1	R375X	2	2
1778	12	30,792,531	G > A	IPO8	R598X	2	2
1778	11	27,695,732	del 2	BDNF	AT	1	3
1778	12	82,832,545	ins 1	C12orf26	A	1	3
1778	3	160,395,408	del 4	ARL14	ACAG	1	3
1790	1	62,740,121	G > A	KANK4	R219X	6	3
1790	18	29,645,866	ins 4	RNF125	ins TCTG	4	4
1830	1	159,785,362	C > T	FCRL6	Q406X	1	2
1830	1	44,438,170	C > T	DPH2	R249X	1	2
1830	6	160,240,107	C > T	PNLDC1	R452X	1	2
1830	8	27,168,350	G > A	TRIM35	Q135X	1	2
1830	13	33,679,848	C > G	STARD13	splice	1	2
1830	14	73,137,900	G > A	DPF3	R340X	1	2
1830	14	73,980,908	C > T	HEATR4	W414X	1	2
1830	16	29,830,874	A > G	PAGR1	splice	1	2
1830	16	68,371,361	G > C	PRMT7	splice	1	2
1830	19	51,411,852	C > T	KLK4	W153X	1	2
1830	20	25,656,642	del 2	ZNF337	CT	2	1
1831	3	171,443,808	A > G	PLD1	splice	2	4
1831	6	38,994,442	G > T	DNAH8	splice	3	2
1831	6	160,962,215	T > C	LPA	splice	5	2
1831	10	70,263,453	C > A	SLC25A16	splice	4	3
1859	1	1,115,937	G > A	TTLL10	W121X	4	3
1859	5	175,110,645	A > G	HRH2	I137V	4	2
1859	17	54,939,586	C > T	DGKE	R500X	5	2
1863	3	158,390,167	ins 1	LXN	ins A	1	4
1863	7	99,247,737	del 1	CYP3A5	del C	3	4
1888	3	48,506,433	del 1	ATRIP	del G	2	3
1888	7	47,849,094	A > G	PKD1L1	R2555X	2	3
1903	16	10,855,319	A > T	NUBP1	C235X	2	2
1903	19	56,162,745	del 2	CCDC106	del CC	2	2
1919	1	153,910,311	G > A	DENND4B	P665L	7	0
1919	1	214,820,463	T > G	CENPF	L2517R	4	2

CF	Chr	Position	Variant	Gene	Effect	Number of breast and ovarian cancer cases	
						With variant	Without variant
1919	3	14,547,261	G > A	GRIP2	P909S	5	2
1919	8	120,575,123	G > A	ENPP2	R851W	5	1
1919	12	86,373,459	T > C	MGAT4C	M349V	7	0
1919	16	81,164,184	del 1	PKD1L2	del G	4	3
1919	19	54,823,619	G > C	LILRA5	T37S	4	2
1919	22	26,165,065	G > C	MYO18B	K394N	5	1
JBC267	2	68,805,143	del 1	APLF	del A	1	
JBC267	4	87,735,675	C > T	PTPN13	Q2482X	2	
JBC267	7	105,662,771	C > G	CDHR3	Y651X	1	
JBC267	14	74,507,287	del 2	C14orf45	del AC	2	
JBC267	14	75,142,931	del 1	KIAA0317	del G	1	