

Variable Number Tandem Repeats (VNTRs) as modifiers of breast cancer risk in carriers of *BRCA1*

185delAG

Authors: Yuan Chun Ding¹, Aaron W. Adamson¹, Mehrdad Bakhtiari², Carmina Patrick¹, Jonghun Park², Yael Laitman⁴, Jeffrey N. Weitzel³, Vineet Bafna², Eitan Friedman^{4,5}, Susan L. Neuhausen^{1*}

¹ Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA 91010

² Department of Computer Science and Engineering, University of California San Diego, San Diego, CA

³ Latin American School of Oncology, Tuxla Gutierrez, Chiapas, MX and Natera, San Carlos, CA

⁴ Oncogenetics Unit, Institute of Human Genetics, Sheba Medical Center. Tel-Hashomer, Israel

⁵ The Sackler School of Medicine, Tel-Aviv University, Tel-Aviv Israel

* Corresponding author: email: sneuhausen@coh.org

Keywords Variable number tandem repeats (VNTRs), breast cancer, association analysis

Abstract

Despite substantial efforts in identifying both rare and common variants affecting disease risk, in the majority of diseases, a large proportion of unexplained genetic risk remains. We propose that variable number tandem repeats (VNTRs) may explain a proportion of the missing genetic risk. Herein, we tested whether VNTRs are causal modifiers of breast cancer risk in 347 female carriers of *BRCA1* 185delAG, an important group given their high risk of developing breast cancer. We performed targeted-capture to sequence VNTRs, called genotypes with adVNTR, and tested the association of VNTRs and breast cancer risk using Cox regression models. Of 303 VNTRs that passed quality control checks, 4 VNTRs were significantly associated with risk to develop breast cancer at false discovery rate [FDR] < 0.05 and an additional 4 VNTRs had FDR < 0.25. After determining the specific risk alleles, there was a significantly earlier age at development of breast cancer in carriers of the risk genotypes compared to those without the risk genotypes for seven of eight VNTRs. Results from this first systematic study of VNTRs demonstrate that VNTRs may explain a proportion of the unexplained genetic risk for disease and have larger effects than SNPs.

INTRODUCTION

For carriers of pathogenic variants (PVs) in *BRCA1*, the lifetime risk for developing breast cancer (up to 80% lifetime risk) is a six-fold increase over that of average risk women and ovarian cancer risk (up to a 44% lifetime risk) is up to a 30-fold increase [1]. Despite these substantially elevated risks, penetrance is incomplete (not all carriers will develop cancer) and age at cancer diagnosis varies. The limited understanding of factors that modify cancer risks in *BRCA1* carriers hampers clinical decision-making ability, including decisions about the appropriate type and timing of risk reducing surgeries. Therefore, there is a critical, clinically relevant need for more refined risk estimates.

The variation in risk, even in identical PVs carriers, suggests that modifier factors, both genetic and environmental, affect cancer risks [2]. Studies to identify “modifier genes” that govern the phenotypic expression of *BRCA* PV carriers have been ongoing since the early 2000's, conducted largely through the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) [3, 4]. Through genome-wide association studies (GWAS), single nucleotide polymorphisms (SNPs) have been identified that, when combined into a polygenic risk score (PRS), better define *BRCA1* carriers at higher and lower risk of developing breast cancer (e.g., [5-7]). However, these modifier variants explain only a portion of the variation in risk [8, 9]. Identifying additional genetic modifiers will facilitate better risk estimates for clinical decision-making on timing and options for risk reduction.

Variable number tandem repeats (VNTRs) may plausibly account for some of the missing genetic risk. They are known to modulate biologic processes, including gene expression and protein function [10-14]. These eVNTRs (VNTR expression Quantitative Trait Loci) also mediate risks of developing various cancers [15, 16] including breast cancer [17-20]. A genome-wide investigation of VNTRs as modifiers has been hampered by technical difficulties; however, adVNTR [10, 21] became available to genotype VNTRs (i.e., count repeat units) from next generation sequencing (NGS) data. This tool uses Hidden Markov models (HMM) to model each VNTR, count repeat units, and detect sequence variation.

We tested a new paradigm – that VNTRs are causal modifiers of breast cancer risk. They have not been systematically investigated as they are poorly [tagged by nearby SNPs](#) [12]. Previous GWAS conducted through CIMBA have demonstrated heterogeneity of breast cancer risk by *BRCA1/2* mutations, breast tumor subtypes, and race and ethnicity [22]. Therefore, to reduce potential confounding with unmeasured variables, we tested the association in carriers of a single recurring PV in *BRCA1*. We performed targeted-capture to sequence VNTRs, called genotypes with adVNTR, and explored the association of VNTRs and breast cancer in 327 women carrying the pathogenic *BRCA1* 185delAG mutation NM_007294.4(*BRCA1*):c.68_69del (p.Glu23fs) (rs80357914).

METHODS

Cases. Females carrying the pathogenic *BRCA1* variant 185delAG (NM_007294.3:c.66_67del) were eligible. Of the 347 participants with DNA, 250 were enrolled by Dr. E. Friedman from the Suzanne Levy-Gertner Oncogenetics Unit at the Sheba Medical Center (SMC) in Israel. All participants underwent oncogenetic counseling and genotyping of cancer susceptibility genes, including *BRCA1*. Referral to the oncogenetics services came from several sources: women who developed breast and/or ovarian cancer (consecutive women at the SMC), cancer free women with a significant family history of breast and /or ovarian cancer, and from population screens of the three predominant mutations in Ashkenazi Jewish (AJ) women in *BRCA1* and *BRCA2* [23], a procedure recently approved and included in the Israeli "health basket" for all AJ women as a screening procedure with no need for pre-test counseling. Another 95 participants were recruited and enrolled by Dr. J. Weitzel into the Clinical Cancer Genomics Community Research Network housed at the City of Hope and another 2 participants were recruited and enrolled in a research study led by Dr. S. Neuhausen. All participants provided written informed consent under IRB-approved protocols at their respective institutions. All participants were unrelated.

VNTR genotyping

VNTR selection. To get an initial list of VNTRs (of four or more base pair repeats), Tandem Repeat Finder (TRF) [24] was applied to the human reference genome [GRCh38], and 559,804 VNTRs were identified. To

focus on the most relevant candidates, we selected VNTRs that intersected with coding exons, promoters, or untranslated regions (UTRs) of genes in RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>). VNTRs were excluded if they were located in low-complexity sequence (e.g. close to a telomere) resulting in 8953 candidate VNTRs. Lastly, only candidate VNTRs with total length of 140 bp or shorter (n=6271) were included so that genotypes could be confidently assigned with Illumina short read sequencing data. We used the Agilent SureDesign software to design probes for 6271 VNTRs. Of these 6271 VNTRs, 1398 are in coding exons, 2000 are in promoter regions, and 2873 are in UTRs. Using the least stringent parameters to design probes for target enrichment, probes were designed to cover 6186 VNTRs; 85 VNTRs were excluded because they were in repetitive DNA regions where probes could not be designed. We also excluded 21 VNTRs on the Y chromosome.

Library preparation. Libraries were created from 500 ng DNA using KAPA Hyper (KAPA Biosystems) reagents along with our optimized protocols [25, 26] to maximize efficiency and minimize cost including bar-coding samples prior to hybridization in order to hybridize 24 samples per bait capture kit. Prior to bait capture, indexed samples were carefully quantified with both Picogreen and qPCR assays to ensure equal representation of each sample in the pool. The protocol has been optimized to minimize the number of PCR cycles, reducing duplicate reads to less than 20%. Sequencing was performed in the City of Hope Integrative Genomics Core (IGC) on a HiSeq2500 genetic analyzer (Illumina Inc, San Diego, CA).

Targeted capture DNA sequencing and processing of reads. For the first 192 samples, we obtained sequencing on 150 bp paired-end reads. The remaining samples were sequenced with 250 bp paired-end reads in order to provide more flanking sequence for better accuracy of the call.

Once sequencing was complete, the fastq files were uploaded to a shared folder for processing. Sequence reads were aligned to NCBI build GRCh38 using Burrows-Wheeler Aligner (BWA) to generate BAM files. Duplicate reads were detected by Picard MarkDuplicates (<http://broadinstitute.github.io/picard/>) and only unique reads were kept for subsequent analysis. The BAM files including unmapped reads were used for assigning VNTRs. From the BAM files, genotypes from VNTRs were assigned using adVNTR-NN adapted

from adVNTR [21] based on minimal total supporting reads ≥ 10 and minimal proportion of reads to support alternative allele ≥ 0.25 .

Confirmation of VNTR genotyping results from adVNTR. Using the unique flanking regions of the selected VNTRs, PCR primers were designed to amplify 50 ng DNA from up to 4 samples per VNTR genotype. PCR reactions were performed using TAQ polymerase (Qiagen) and amplification was confirmed using gel electrophoresis. Samples were then sequenced on an Applied Biosystems SeqStudio Genetic Analyzer (ThermoFisher Scientific).

VNTR sequences were visualized using Quality Check and Variant Analysis Modules on the ThermoFisher Cloud. The visualized sequence in conjunction with the product sizes from the post-PCR gel electrophoresis were used to verify genotyping calling made by adVNTR. For homozygotes, this was done by observing a single band of the correct size during gel electrophoresis and by quality sequence for the number of repeats called by adVNTR. Whereas heterozygotes were confirmed by observing multiple bands of expected size differentials on the gel and a poor-quality sanger sequence at the point of allele differences.

Statistical analysis. After genotypes were assigned for each VNTR, we tested for HWE. For those that were in HWE ($p > 0.001$), we tested the association of the VNTR and risk to develop breast cancer using Cox regression models in both the primary and secondary analyses (described below). In the model, women with a first breast cancer are considered as affected with time to breast cancer diagnosis as the end point; those unaffected with any cancer or diagnosed with ovarian cancer prior to breast cancer were censored at age of last follow-up and age at ovarian cancer diagnosis, respectively. There were too few cases of ovarian cancer for association analysis.

In the primary analysis, we tested the association between each VNTR marker as a continuous variable and disease risk. Three separate VNTR genotypes were constructed: 1) the average length of the two alleles; 2) the length of only the shorter allele; and 3) the length of only the longer allele [27]. Analyses were adjusted for sample collection site (US or Israel). Probability values were adjusted for multiple comparisons using the False

Discovery Rate (FDR) method of Benjamini and Hochberg [28]. For VNTRs with associations with $FDR < 0.25$ in the primary association analysis, a secondary association analysis was performed to identify the specific risk groups of repeat alleles using a sliding window method [27]. Specifically, a threshold T along the number of repeats from short to long was used to dichotomize allele lengths. An allele was denoted as ‘short’ if it had shorter than T repeat motifs, and ‘long’ otherwise. Multiple values of threshold T were chosen for association tests. For each specific threshold T, the VNTR genotype of an individual was converted to homozygous-short-allele genotype (S/S), heterozygous-short-and-long-allele genotype (S/L), or homozygous-long-allele genotype (L/L). The optimal threshold for each VNTR was determined by choosing T that provided the smallest p-value among the multiple association tests. This secondary analysis allowed us to identify critical cut points for risk alleles along the continuous repeat allele distribution in a VNTR and then to estimate the effect size of the association related to the risk genotypes. Kaplan-Meier curves and log-rank tests were used to graphically examine differences in the cumulative probability of breast cancer risk among VNTR genotype groups categorized using the critical cut points for risk alleles.

Luciferase assays. We conducted luciferase assays to test alleles of one VNTR in the promoter region to determine if it affected expression. We selected the VNTR with the lowest FDR that was in a promoter or 5’UTR region.

PCR amplification and subcloning. Primers, designed to flank the VNTR, also included restriction enzyme sites. The PCR products for each risk allele and reference repeat allele served as inserts for subcloning into plasmids. PCR reactions were performed using Herculase high fidelity polymerase (Agilent) and the products sizes were confirmed by gel electrophoresis. Because an individual with a rare VNTR allele is almost always heterozygote, the PCR reaction generated two different sized products. Following amplification, each PCR product was cloned separately into the pGEM-T plasmid (Promega) according to the manufacturer’s instructions followed by transformation into *E. coli* cells. Ampicillin resistant colonies were picked, grown overnight, followed by the isolation of plasmid DNA. Several plasmids were Sanger sequenced to ensure that clones contained inserts with the appropriate sequence and length corresponding to the desired repeat alleles.

Construction of luciferase reporter plasmids. Following sequence confirmation, the pGEM-T plasmids served as the template DNA for another PCR reaction using the same primers described above. The PCR products containing the VNTR alleles were digested with the appropriate restriction enzymes, purified, and cloned into restriction enzyme sites in the pGL3-Promoter luciferase gene-reporter plasmid (Promega) upstream of the SV40 promoter. Sequencing was performed to confirm the fidelity and orientation of all inserts.

Luciferase assay. We transfected the luciferase gene-reporter plasmids along with the pRL Renilla luciferase control plasmid into an MCF7 cell line (from ATCC) and performed luciferase assays using the Dual-Luciferase Reporter Assay System (Promega). After 48 hours, each group of cells was lysed and manipulated according to the manufacturer's instructions, and the fluorescence intensity of each group of cells was expressed as the ratio of firefly luciferase activity to Renilla luciferase activity. The relative luciferase activities of the plasmid constructs were determined by normalizing the standardized values to empty pGL3-Promoter plasmid. All transfections were performed in quadruplicate, and each construct was tested in three independent experiments. The average of the 12 relative luciferase measurements for each allele were expressed as the mean \pm standard error of mean (SEM). Difference in relative activity values between the risk repeat allele group and reference repeat allele group was tested by one-way ANOVA analysis. The P-value was adjusted for multiple testing using the Tukey's method [29]; adjusted p-values less than 0.05 were considered as statistically significant.

RESULTS

Participants. The cancer status and ages at diagnosis or enrollment (for non-cancer cases) are shown in Table 1. Of the 347 women, ages ranged from 18 to 77 years with 49% having been diagnosed with breast cancer, of which 4.9% also were also diagnosed with ovarian cancer. The median age at diagnosis was 55 years (95% CI: 50 - 58 years). The cumulative risk of breast cancer by age 70 years was 75% .

VNTR genotyping. In total, we sequenced 6165 VNTRs in 347 *BRCA1* 185delAG PV carriers. Genotypes were called using adVNTR-NN. In Figure 1, the flow diagram of steps for elimination of VNTRs and samples is

shown. Of 6165 VNTRs, 3847 (62.4%) VNTRs were removed due to missing more than 5% of genotypes, with the main reasons being VNTRs located in GC-rich regions which had poorer amplification during library generation, imperfect repeats, or flanked by other repetitive elements. Another 1622 VNTRs were removed because they were monomorphic (1588 VNTRs) or not in Hardy-Weinberg equilibrium (P value < 0.001 ; 34 VNTRs). Lastly 393 VNTRs had heterozygosity < 0.02 . Because this is a homogeneous dataset of Ashkenazi Jewish ancestry, it was expected that more VNTRs would be monomorphic and within VNTRs, not all alleles would be present. Twenty samples were removed that had more than 10% missing genotypes leaving 327 for analysis. The summary of repeat alleles in this dataset for the 303 VNTRs is shown in Table 2.

Association of VNTRs and risk of developing cancer. In the primary analysis, we used Cox proportional hazards models to evaluate the association between each VNTR and risk of developing breast cancer, considering the VNTR as a continuous variable. Of 303 VNTRs analyzed, 8 VNTRs had unadjusted p-values < 0.005 , of which four VNTRs had FDR < 0.05 , and an additional four had FDR < 0.25 (Table 3; Supplemental Table 1). The alleles for each of the eight VNTRs were accurately called, with 100% consistency among the adVNTR, agarose gel, and Sanger sequencing results (VNTR 558420 is shown as an example in Supplemental Figure 1). We then conducted the secondary analysis for the eight VNTRs to identify the specific risk repeat alleles contributing to the significant association. For 7 of 8 VNTRs, there was a significant ($P < 0.05$) difference in breast cancer risk between risk and reference genotype groups categorized using the critical cut point of risk allele (forward slash character in Table 3). For VNTR 47260, although breast cancer risk increased with repeat length based on the linear trend test (FDR = 0.035), there were too few long repeat alleles ($> 9R$) for a stable estimation of cut point in the categorical test. Kaplan-Meier (KM) curves were used to graphically show the difference in the cumulative probability of breast cancer risk for the VNTR genotype groups (Figure 2 and Supplemental Figure 2). Individuals with the risk genotypes had significantly earlier ages at developing breast cancer (log-rank p value < 0.05) (Figure 2 and Supplemental Figure 2). For example, the median ages at breast cancer diagnosis for carriers with the S/S genotype and the L/L genotype in VNTR357331 were 40 years and

56 years, respectively (log-rank p value of 0.0014, Figure 2), indicating the age-modifying effect of breast cancer diagnosis among carriers harboring risk genotype (S/S).

Effect of VNTR alleles on expression. For testing the effect on gene expression, we selected the VNTR with the lowest FDR that was located in a gene promoter or 5' UTR. We tested VNTR 558420 located in the 5' UTR of *ZNF501* (p-value = 0.0025 and FDR = 0.135) (Table 3 and Supplemental Figure 3) with repeats of 2R, 3R and 4R and 3 genotypes (3 samples with genotype 2/3, 314 with 3/3, and 5 with 3/4). In Figure 3, normalized luciferase activity is shown for the 2R, 3R, 4R repeats and the control (empty vector) with standard error bars on the top of each group mean. There was a significant (adjusted p value < 0.05) difference between the 2R and 4R groups with the 3R intermediate (Figure 3) and a significant linear trend of decreased luciferase activity with increasing number of repeats (p = 0.021) (Supplemental Figure 4).

DISCUSSION

Our study is the first to conduct a systematic study of VNTRs and association with risk to develop cancer in high-risk *BRCA1* PV carriers. We identified four VNTRs significantly associated with risk of developing breast cancer in women carrying the 185delAG *BRCA1* PV (FDR < 0.05) and another four VNTRs associated with FDR < 0.25.

None of the small number of previous association studies of risk of developing breast cancer and VNTRs at candidate genes had investigated the eight VNTRs we identified. Krontiris and coworkers reported an association of rare alleles in a *HRAS1* VNTR and development of cancers, including breast cancer [17], and a meta-analysis of 13 breast cancer studies found an association with breast cancer risk [30]. Functional analysis showed that this *HRAS* VNTR altered CpG DNA methylation [31]. In a meta-analysis of 17 studies of a CAG-repeat polymorphism in the androgen receptor, they found an association of longer CAG repeats with an increased risk of breast cancer in Caucasian women [32]. In a meta-analysis of two studies of the MNS16A VNTR in the *hTERT* promoter, they found a significant association with development of breast cancer. In a Japanese study of an 18-bp VNTR in the promoter of *PTTG1IP*, they found a significant association with risk of

estrogen-receptor positive breast cancer, with functional analysis showing that an increase in the number of repeats increased the binding affinity of ER-alpha [20]. In a study of a VNTR in the promoter of *XRCC5*, they found a significant association with age at breast cancer diagnosis [18]. Our study did not include the two trinucleotide repeats and the *XRCC5* VNTR, and the *MSNI* VNTR was monomorphic and the *PTTG1IP* VNTR was missing too many genotypes in our set and thus were excluded.

Of the eight VNTRs that we found to be associated with risk of developing breast cancer in this population, several warrant further investigation. VNTR 945060 is in the 5'UTR of *ERCC6L*, a DNA helicase. *ERCC6L* is highly expressed in breast tissue and higher levels of expression have been associated with worse survival [33]; silencing of *ERCC6L* in breast cell lines significantly inhibited cell proliferation [33, 34]. A second VNTR, 253688, is located 3' of *FLJ22447*, a lncRNA located near *HIF-1 α* . In a study of esophageal squamous cell carcinoma and gastric cancers to determine the effect of *FLJ22447* on *HIF-1 α* , they observed that low expression of lncRNA was associated with expression of *HIF-1 α* suggesting that *FLJ22447* may have a regulatory function on *HIF-1 α* expression [35]. High over-expression of *HIF-1 α* is common in breast cancers and is particularly common in *BRCA1* carriers [36-38]. This VNTR may alter risk to develop breast through affecting *HIF-1 α* .

Given the reports that there are shared genetic contributions between breast cancer and schizophrenia [39], it is interesting that three of the VNTRs are at or in genes (*SYN2*, *ZNF501*, *ZNF804A*) associated with risk to develop schizophrenia [40-44]; VNTR 549198 is in exon 12 of *SYN2*; VNTR 472060 is in exon 4 of *ZNF804A*; and VNTR558420 is in the 5'UTR of *ZNF501* and all are most commonly expressed in brain (proteatlas.org). From our luciferase assays, there was differential expression from varying alleles in the VNTR in the 5'UTR of *ZNF501*; expression differences for this VNTR were only associated with brain tissue in GTEX [10]. The exonic VNTRs in *SYN2* and in *ZNF804A* cause expansions of poly-serine (Supplemental Figure 5) and poly-alanine (Supplemental Figure 6) tracts, respectively. VNTR expansions in gene coding regions have been associated with multiple diseases [45]. Further investigation is needed to assess possible roles in development of breast cancer.

This was a pilot study to determine the feasibility of conducting targeted sequencing of VNTRs and investigating the association of VNTRs as modifiers of disease risk, similar to what has been accomplished with SNPs [8, 46]. We purposefully included women carrying the specific *BRCA1* 185delAG Ashkenazi Jewish founder PV to try to explain the known variation in risk in women carrying this PV and to reduce potential confounding with unmeasured variables; however, the consequence is that it reduced the number of VNTRs that were polymorphic and restricted the sample size. In hindsight, using targeted capture and sequencing of 250 bp reads limited the size of repeats and reduced the number of VNTRs that made it through all the quality control checks due to poor amplification of VNTRs in GC-rich regions, difficulty in aligning VNTRs with imperfect repeats and/or with low complexity/repetitive sequence in the flanking regions. As costs continue to decrease for whole-genome sequencing and for long-read sequences such as performed by PacBio, we will be able to obtain analyzable data on a larger number of VNTRs.

BRCA1 breast cancers are generally basal, triple-negative hormone receptor cancers (TNBC). We have seen from SNP studies of both *BRCA1* carriers and women with TNBC that there are fewer SNPs associated with risk than for estrogen-receptor positive breast cancers. Thus, identification of VNTRs significantly associated with risk of developing breast cancer in this genetically and ethnically homogeneous population is encouraging; several of which have been observed to play a role in breast cancer. HRs for these VNTRs ranged from 1.7 to 6.1 whereas HRs for SNPs generally range from 1.01 to 1.4 [8, 47], suggesting that VNTRs may have larger effects than SNPs. These results need to be validated in larger datasets that include women of diverse ethnicities, a wider spectrum of *BRCA1* PVs, and carriers of *BRCA2* PVs. Moreover, a larger genome-wide VNTR association study may identify additional VNTRs.

In summary, the results from this study demonstrate that VNTRs may explain a proportion of the unexplained genetic risk for disease. Similar to SNPs, VNTRs significantly associated with the disease of interest could be incorporated into polygenic risk scores (PRS) to test for improved risk assessment and clinical applicability.

Acknowledgments. This work was supported by the Basser Foundation (SLN and EF) and the Beckman Research Institute of City of Hope Excellence Awards Program. Sequencing of the VNTRs was performed in the Integrative Genomics Core supported by the National Cancer Institute of the National Institutes of Health under grant number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. SLN is partially supported by the Morris and Horowitz Families Professorship. VB and JP were supported in part by R01GM114362, R01HG010149 and RO1HG011558 from the NIH.

Author Contribution Statements. SLN and EF conceived of the idea, developed the design, and obtained funding to conduct the study. VB, MB, and JP did the analysis of the sequencing data to obtain genotypes. AA and CP prepared the DNA libraries for sequencing, conducted the luciferase assays, and the validation of the VNTR genotypes. YCD did quality control of the VNTR genotypes and conducted the association statistical analysis. SLN, EF, JW, and YL contributed DNA samples and data. All authors have contributed to, read, and approved the manuscript.

REFERENCES CITED

1. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, Jervis S, van Leeuwen FE, Milne RL, Andrieu N, Goldgar DE, Terry MB, Rookus MA, Easton DF, Antoniou AC, Brca, Consortium BC, McGuffog L, Evans DG, Barrowdale D, Frost D, Adlard J, Ong KR, Izatt L, Tischkowitz M, Eeles R, Davidson R, Hodgson S, Ellis S, Nogues C, Lasset C, Stoppa-Lyonnet D, Fricker JP, Faivre L, Berthet P, Hooning MJ, van der Kolk LE, Kets CM, Adank MA, John EM, Chung WK, Andrulis IL, Southey M, Daly MB, Buys SS, Osorio A, Engel C, Kast K, Schmutzler RK, Caldes T, Jakubowska A, Simard J, Friedlander ML, McLachlan SA, Machackova E, Foretova L, Tan YY, Singer CF, Olah E, Gerdes AM, Arver B, Olsson H. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* 2017; 317: p. 2402-2416.
2. Levy-Lahad E and Friedman E. Cancer risks among BRCA1 and BRCA2 mutation carriers. *Br J Cancer* 2007; 96: p. 11-5.
3. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T, Ghoussaini M, Barrowdale D, Embrace, Peock S, Cook M, Oliver C, Frost D, Eccles D, Evans DG, Eeles R, Izatt L, Chu C, Douglas F, Paterson J, Stoppa-Lyonnet D, Houdayer C, Mazoyer S, Giraud S, Lasset C, Remenieras A, Caron O, Hardouin A, Berthet P, Collaborators GS, Hogervorst FB, Rookus MA, Jager A, van den Ouweland A, Hoogerbrugge N, van der Luijt RB, Meijers-Heijboer H, Gomez Garcia EB, Hebon, Devilee P, Vreeswijk MP, Lubinski J, Jakubowska A, Gronwald J, Huzarski T, Byrski T, Gorski B, Cybulski C, Spurdle AB, Holland H, kConFab, Goldgar DE, John EM, Hopper JL, Southey M, Buys SS, Daly MB, Terry MB, Schmutzler RK, Wappenschmidt B, Engel C, Meindl A, Preisler-Adams S, Arnold N, Niederacher D, Sutter C, Domchek SM, Nathanson KL, Rebbeck T, Blum JL, Piedmonte M, Rodriguez GC, Wakeley K, Boggess JF, Basil J, Blank SV, Friedman E, Kaufman B, Laitman Y, Milgrom R, Andrulis IL, Glendon G, Ozcelik H, Kirchoff T, Vijai J, Gaudet MM, Altshuler D, Guiducci C, Swe B, Loman N, Harbst K, Rantala J, Ehrencrona H,

- Gerdes AM, Thomassen M, Sunde L, Peterlongo P, Manoukian S, Bonanni B, Viel A, Radice P, Caldes T, de la Hoya M, Singer CF, Fink-Retter A, Greene MH, Mai PL, Loud JT, Guidugli L, Lindor NM, Hansen TV, Nielsen FC, Blanco I, Lazaro C, Garber J, Ramus SJ, Gayther SA, Phelan C, Narod S, Szabo CI, Mod S, Benitez J, Osorio A, Nevanlinna H, Heikkinen T, Caligo MA, Beattie MS, Hamann U, Godwin AK, Montagna M, Casella C, Neuhausen SL, Karlan BY, Tung N, Toland AE, Weitzel J, Olopade O, Simard J, Soucy P, Rubinstein WS, Arason A, Rennert G, Martin NG, Montgomery GW, Chang-Claude J, Flesch-Janys D, Brauch H, Genica, Severi G, Baglietto L, Cox A, Cross SS, Miron P, Gerty SM, Tapper W, Yannoukakos D, Fountzilas G, Fasching PA, Beckmann MW, Dos Santos Silva I, Peto J, Lambrechts D, Paridaens R, Rudiger T, Forsti A, Winqvist R, Pylkas K, Diasio RB, Lee AM, Eckel-Passow J, Vachon C, Blows F, Driver K, Dunning A, Pharoah PP, Offit K, Pankratz VS, Hakonarson H, Chenevix-Trench G, Easton DF and Couch FJ. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* 2010; 42: p. 885-92.
4. Chenevix-Trench G, Milne RL, Antoniou AC, Couch FJ, Easton DF, Goldgar DE, Cimba. An international initiative to identify genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA). *Breast Cancer Res* 2007; 9: p. 104.
5. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, Soucy P, Fredericksen Z, Barrowdale D, Dennis J, Gaudet MM, Dicks E, Kosel M, Healey S, Sinilnikova OM, Lee A, Bacot F, Vincent D, Hogervorst FB, Peock S, Stoppa-Lyonnet D, Jakubowska A, kConFab I, Radice P, Schmutzler RK, Swe B, Domchek SM, Piedmonte M, Singer CF, Friedman E, Thomassen M, Ontario Cancer Genetics N, Hansen TV, Neuhausen SL, Szabo CI, Blanco I, Greene MH, Karlan BY, Garber J, Phelan CM, Weitzel JN, Montagna M, Olah E, Andrulis IL, Godwin AK, Yannoukakos D, Goldgar DE, Caldes T, Nevanlinna H, Osorio A, Terry MB, Daly MB, van Rensburg EJ, Hamann U, Ramus SJ, Toland AE, Caligo MA, Olopade OI, Tung N, Claes K, Beattie MS, Southey MC, Imyanitov EN,

Tischkowitz M, Janavicius R, John EM, Kwong A, Diez O, Balmana J, Barkardottir RB, Arun BK, Rennert G, Teo SH, Ganz PA, Campbell I, van der Hout AH, van Deurzen CH, Seynaeve C, Gomez Garcia EB, van Leeuwen FE, Meijers-Heijboer HE, Gille JJ, Ausems MG, Blok MJ, Ligtenberg MJ, Rookus MA, Devilee P, Verhoef S, van Os TA, Wijnen JT, Hebon, Embrace, Frost D, Ellis S, Fineberg E, Platte R, Evans DG, Izatt L, Eeles RA, Adlard J, Eccles DM, Cook J, Brewer C, Douglas F, Hodgson S, Morrison PJ, Side LE, Donaldson A, Houghton C, Rogers MT, Dorkins H, Eason J, Gregory H, McCann E, Murray A, Calender A, Hardouin A, Berthet P, Delnatte C, Nogues C, Lasset C, Houdayer C, Leroux D, Rouleau E, Prieur F, Damiola F, Sobol H, Coupier I, Venat-Bouvet L, Castera L, Gauthier-Villars M, Leone M, Pujol P, Mazoyer S, Bignon YJ, Collaborators GS, Zlowocka-Perlowska E, Gronwald J, Lubinski J, Durda K, Jaworska K, Huzarski T, Spurdle AB, Viel A, Peissel B, Bonanni B, Melloni G, Ottini L, Papi L, Varesco L, Tibiletti MG, Peterlongo P, Volorio S, Manoukian S, Pensotti V, Arnold N, Engel C, Deissler H, Gadzicki D, Gehrig A, Kast K, Rhiem K, Meindl A, Niederacher D, Ditsch N, Plendl H, Preisler-Adams S, Engert S, Sutter C, Varon-Mateeva R, Wappenschmidt B, Weber BH, Arver B, Stenmark-Askmal M, Loman N, Rosenquist R, Einbeigi Z, Nathanson KL, Rebbeck TR, Blank SV, Cohn DE, Rodriguez GC, Small L, Friedlander M, Bae-Jump VL, Fink-Retter A, Rappaport C, Gschwantler-Kaulich D, Pfeiler G, Tea MK, Lindor NM, Kaufman B, Shimon Paluch S, Laitman Y, Skytte AB, Gerdes AM, Pedersen IS, Moeller ST, Kruse TA, Jensen UB, Vijai J, Sarrel K, Robson M, Kauff N, Mulligan AM, Glendon G, Ozcelik H, Ejlersen B, Nielsen FC, Jonson L, Andersen MK, Ding YC, Steele L, Foretova L, Teule A, Lazaro C, Brunet J, Pujana MA, Mai PL, Loud JT, Walsh C, Lester J, Orsulic S, Narod SA, Herzog J, Sand SR, Tognazzo S, Agata S, Vaszko T, Weaver J, Stavropoulou AV, Buys SS, Romero A, de la Hoya M, Aittomaki K, Muranen TA, Duran M, Chung WK, Lasa A, Dorfling CM, Miron A, Bcfr, Benitez J, Senter L, Huo D, Chan SB, Sokolenko AP, Chiquette J, Tihomirova L, Friebel TM, Agnarsson BA, Lu KH, Lejbkowitz F, James PA, Hall P, Dunning AM, Tessier D, Cunningham J, Slager SL, Wang C, Hart S, Stevens K, Simard J, Pastinen T, Pankratz VS, Offit K, Easton DF, Chenevix-Trench G, Antoniou AC and Cimba. Genome-wide association study in

BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet* 2013; 9: p. e1003212.

6. Kuchenbaecker KB, McGuffog L, Barrowdale D, Lee A, Soucy P, Dennis J, Domchek SM, Robson M, Spurdle AB, Ramus SJ, Mavaddat N, Terry MB, Neuhausen SL, Schmutzler RK, Simard J, Pharoah PDP, Offit K, Couch FJ, Chenevix-Trench G, Easton DF, Antoniou AC. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst* 2017; 109.
7. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, Tyrer JP, Chen TH, Wang Q, Bolla MK, Yang X, Adank MA, Ahearn T, Aittomaki K, Allen J, Andrulis IL, Anton-Culver H, Antonenkova NN, Arndt V, Aronson KJ, Auer PL, Auvinen P, Barrdahl M, Beane Freeman LE, Beckmann MW, Behrens S, Benitez J, Bermisheva M, Bernstein L, Blomqvist C, Bogdanova NV, Bojesen SE, Bonanni B, Borresen-Dale AL, Brauch H, Bremer M, Brenner H, Brentnall A, Brock IW, Brooks-Wilson A, Brucker SY, Bruning T, Burwinkel B, Campa D, Carter BD, Castelao JE, Chanock SJ, Chlebowski R, Christiansen H, Clarke CL, Collee JM, Cordina-Duverger E, Cornelissen S, Couch FJ, Cox A, Cross SS, Czene K, Daly MB, Devilee P, Dork T, Dos-Santos-Silva I, Dumont M, Durcan L, Dwek M, Eccles DM, Ekici AB, Eliassen AH, Ellberg C, Engel C, Eriksson M, Evans DG, Fasching PA, Figueroa J, Fletcher O, Flyger H, Forsti A, Fritschi L, Gabrielson M, Gago-Dominguez M, Gapstur SM, Garcia-Saenz JA, Gaudet MM, Georgoulas V, Giles GG, Gilyazova IR, Glendon G, Goldberg MS, Goldgar DE, Gonzalez-Neira A, Grenaker Alnaes GI, Grip M, Gronwald J, Grundy A, Guenel P, Haeberle L, Hahnen E, Haiman CA, Hakansson N, Hamann U, Hankinson SE, Harkness EF, Hart SN, He W, Hein A, Heyworth J, Hillemanns P, Hollestelle A, Hooning MJ, Hoover RN, Hopper JL, Howell A, Huang G, Humphreys K, Hunter DJ, Jakimovska M, Jakubowska A, Janni W, John EM, Johnson N, Jones ME, Jukkola-Vuorinen A, Jung A, Kaaks R, Kaczmarek K, Kataja V, Keeman R, Kerin MJ, Khusnutdinova E, Kiiski JI, Knight JA, Ko YD, Kosma VM, Koutros S, Kristensen VN, Kruger U, Kuhl T, Lambrechts D, Le Marchand L, Lee E, Lejbkowitz F, Lilyquist J, Lindblom A, Lindstrom S, Lissowska J, Lo WY,

Loibl S, Long J, Lubinski J, Lux MP, MacInnis RJ, Maishman T, Makalic E, Maleva Kostovska I, Mannermaa A, Manoukian S, Margolin S, Martens JWM, Martinez ME, Mavroudis D, McLean C, Meindl A, Menon U, Middha P, Miller N, Moreno F, Mulligan AM, Mulot C, Munoz-Garzon VM, Neuhausen SL, Nevanlinna H, Neven P, Newman WG, Nielsen SF, Nordestgaard BG, Norman A, Offit K, Olson JE, Olsson H, Orr N, Pankratz VS, Park-Simon TW, Perez JIA, Perez-Barrios C, Peterlongo P, Peto J, Pinchev M, Plaseska-Karanfilska D, Polley EC, Prentice R, Presneau N, Prokofyeva D, Purrington K, Pylkas K, Rack B, Radice P, Rau-Murthy R, Rennert G, Rennert HS, Rhenius V, Robson M, Romero A, Ruddy KJ, Ruebner M, Saloustros E, Sandler DP, Sawyer EJ, Schmidt DF, Schmutzler RK, Schneeweiss A, Schoemaker MJ, Schumacher F, Schurmann P, Schwentner L, Scott C, Scott RJ, Seynaeve C, Shah M, Sherman ME, Shrubsole MJ, Shu XO, Slager S, Smeets A, Sohn C, Soucy P, Southey MC, Spinelli JJ, Stegmaier C, Stone J, Swerdlow AJ, Tamimi RM, Tapper WJ, Taylor JA, Terry MB, Thone K, Tollenaar R, Tomlinson I, Truong T, Tzardi M, Ulmer HU, Untch M, Vachon CM, van Veen EM, Vijai J, Weinberg CR, Wendt C, Whittemore AS, Wildiers H, Willett W, Winqvist R, Wolk A, Yang XR, Yannoukakos D, Zhang Y, Zheng W, Ziogas A, Investigators A, kConFab AI, Collaborators N, Dunning AM, Thompson DJ, Chenevix-Trench G, Chang-Claude J, Schmidt MK, Hall P, Milne RL, Pharoah PDP, Antoniou AC, Chatterjee N, Kraft P, Garcia-Closas M, Simard J and Easton DF. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019; 104: p. 21-34.

8. Barnes DR, Rookus MA, McGuffog L, Leslie G, Mooij TM, Dennis J, Mavaddat N, Adlard J, Ahmed M, Aittomaki K, Andrieu N, Andrulis IL, Arnold N, Arun BK, Azzollini J, Balmana J, Barkardottir RB, Barrowdale D, Benitez J, Berthet P, Bialkowska K, Blanco AM, Blok MJ, Bonanni B, Boonen SE, Borg A, Bozsik A, Bradbury AR, Brennan P, Brewer C, Brunet J, Buys SS, Caldes T, Caligo MA, Campbell I, Christensen LL, Chung WK, Claes KBM, Colas C, Collaborators GS, Collaborators E, Collonge-Rame MA, Cook J, Daly MB, Davidson R, de la Hoya M, de Putter R, Delnatte C, Devilee P, Diez O, Ding YC, Domchek SM, Dorfling CM, Dumont M, Eeles R, Ejlersen B, Engel C, Evans DG, Faivre L,

Foretova L, Fostira F, Friedlander M, Friedman E, Frost D, Ganz PA, Garber J, Gehrig A, Gerdes AM, Gesta P, Giraud S, Glendon G, Godwin AK, Goldgar DE, Gonzalez-Neira A, Greene MH, Gschwantler-Kaulich D, Hahnen E, Hamann U, Hanson H, Hentschel J, Hogervorst FBL, Hooning MJ, Horvath J, Hu C, Hulick PJ, Imyanitov EN, kConFab I, Investigators H, Investigators G, Isaacs C, Izatt L, Izquierdo A, Jakubowska A, James PA, Janavicius R, John EM, Joseph V, Karlan BY, Kast K, Koudijs M, Kruse TA, Kwong A, Laitman Y, Lasset C, Lazaro C, Lester J, Lesueur F, Liljegren A, Loud JT, Lubinski J, Mai PL, Manoukian S, Mari V, Mebirouk N, Meijers-Heijboer HEJ, Meindl A, Mensenkamp AR, Miller A, Montagna M, Mouret-Fourme E, Mukherjee S, Mulligan AM, Nathanson KL, Neuhausen SL, Nevanlinna H, Niederacher D, Nielsen FC, Nikitina-Zake L, Nogues C, Olah E, Olopade OI, Ong KR, O'Shaughnessy-Kirwan A, Osorio A, Ott CE, Papi L, Park SK, Parsons MT, Pedersen IS, Peissel B, Peixoto A, Peterlongo P, Pfeiler G, Phillips KA, Prajzandanc K, Pujana MA, Radice P, Ramser J, Ramus SJ, Rantala J, Rennert G, Risch HA, Robson M, Ronlund K, Salani R, Schuster H, Senter L, Shah PD, Sharma P, Side LE, Singer CF, Slavin TP, Soucy P, Southey MC, Spurdle AB, Steinemann D, Steinsnyder Z, Stoppa-Lyonnet D, Sutter C, Tan YY, Teixeira MR, Teo SH, Thull DL, Tischkowitz M, Tognazzo S, Toland AE, Trainer AH, Tung N, van Engelen K, van Rensburg EJ, Vega A, Vierstraete J, Wagner G, Walker L, Wang-Gohrke S, Wappenschmidt B, Weitzel JN, Yadav S, Yang X, Yannoukakos D, Zimbalatti D, Offit K, Thomassen M, Couch FJ, Schmutzler RK, Simard J, Easton DF, Chenevix-Trench G, Antoniou AC, Consortium of Investigators of Modifiers of B and Brca. Polygenic risk scores and breast and epithelial ovarian cancer risks for carriers of BRCA1 and BRCA2 pathogenic variants. *Genet Med* 2020.

9. Milne RL and Antoniou AC. Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. *Endocr Relat Cancer* 2016; 23: p. T69-84.
10. Bakhtiari M, Park J, Ding YC, Shleizer-Burko S, Neuhausen SL, Halldorsson BV, Stefansson K, Gymrek M, Bafna V. Variable number tandem repeats mediate the expression of proximal genes. *Nat Commun* 2021; 12: p. 2075.

11. Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, Goren A, Gymrek M. The impact of short tandem repeat variation on gene expression. *Nat Genet* 2019; 51: p. 1652-1659.
12. Quilez J, Guilmatre A, Garg P, Highnam G, Gymrek M, Erlich Y, Joshi RS, Mittelman D, Sharp AJ. Polymorphic tandem repeats within gene promoters act as modifiers of gene expression and DNA methylation in humans. *Nucleic Acids Res* 2016; 44: p. 3750-62.
13. Brookes KJ. The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward? *Genomics* 2013; 101: p. 273-81.
14. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, Daly MJ, Price AL, Pritchard JK, Sharp AJ, Erlich Y. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet* 2016; 48: p. 22-9.
15. Hofer P, Zochmeister C, Behm C, Brezina S, Baierl A, Doriguzzi A, Vanas V, Holzmann K, Sutterluty-Fall H, Gsur A. MNS16A tandem repeat minisatellite of human telomerase gene: functional studies in colorectal, lung and prostate cancer. *Oncotarget* 2017; 8: p. 28021-28027.
16. Rose AM, Krishan A, Chakarova CF, Moya L, Chambers SK, Hollands M, Illingworth JC, Williams SMG, McCabe HE, Shah AZ, Palmer CNA, Chakravarti A, Berg JN, Batra J, Bhattacharya SS. MSR1 repeats modulate gene expression and affect risk of breast and prostate cancer. *Ann Oncol* 2018; 29: p. 1292-1303.
17. Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N. An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N Engl J Med* 1993; 329: p. 517-23.
18. Rajaei M, Saadat I, Omidvari S, Saadat M. Association between polymorphisms at promoters of XRCC5 and XRCC6 genes and risk of breast cancer. *Med Oncol* 2014; 31: p. 885.
19. Wang Y, Hu Z, Liang J, Wang Z, Tang J, Wang S, Wang X, Qin J, Wang X, Shen H. A tandem repeat of human telomerase reverse transcriptase (hTERT) and risk of breast cancer development and metastasis in Chinese women. *Carcinogenesis* 2008; 29: p. 1197-201.

20. Xiang C, Gao H, Meng L, Qin Z, Ma R, Liu Y, Jiang Y, Dang C, Jin L, He F, Wang H. Functional variable number of tandem repeats variation in the promoter of proto-oncogene PTTG1IP is associated with risk of estrogen receptor-positive breast cancer. *Cancer Sci* 2012; 103: p. 1121-8.
21. Bakhtiari M, Shleizer-Burko S, Gymrek M, Bansal V, Bafna V. Targeted genotyping of variable number tandem repeats with adVNTR. *Genome Res* 2018; 28: p. 1709-1719.
22. Rebbeck TR, Mitra N, Wan F, Sinilnikova OM, Healey S, McGuffog L, Mazoyer S, Chenevix-Trench G, Easton DF, Antoniou AC, Nathanson KL, Consortium C, Laitman Y, Kushnir A, Paluch-Shimon S, Berger R, Zidan J, Friedman E, Ehrencrona H, Stenmark-Askmal M, Einbeigi Z, Loman N, Harbst K, Rantala J, Melin B, Huo D, Olopade OI, Seldon J, Ganz PA, Nussbaum RL, Chan SB, Odunsi K, Gayther SA, Domchek SM, Arun BK, Lu KH, Mitchell G, Karlan BY, Walsh C, Lester J, Godwin AK, Pathak H, Ross E, Daly MB, Whittemore AS, John EM, Miron A, Terry MB, Chung WK, Goldgar DE, Buys SS, Janavicius R, Tihomirova L, Tung N, Dorfling CM, van Rensburg EJ, Steele L, Neuhausen SL, Ding YC, Ejlersen B, Gerdes AM, Hansen T, Ramon y Cajal T, Osorio A, Benitez J, Godino J, Tejada MI, Duran M, Weitzel JN, Bobolis KA, Sand SR, Fontaine A, Savarese A, Pasini B, Peissel B, Bonanni B, Zaffaroni D, Vignolo-Lutati F, Scuvera G, Giannini G, Bernard L, Genuardi M, Radice P, Dolcetti R, Manoukian S, Pensotti V, Gismondi V, Yannoukakos D, Fostira F, Garber J, Torres D, Rashid MU, Hamann U, Peock S, Frost D, Platte R, Evans DG, Eeles R, Davidson R, Eccles D, Cole T, Cook J, Brewer C, Hodgson S, Morrison PJ, Walker L, Porteous ME, Kennedy MJ, Izatt L, Adlard J, Donaldson A, Ellis S, Sharma P, Schmutzler RK, Wappenschmidt B, Becker A, Rhiem K, Hahnen E, Engel C, Meindl A, Engert S, Ditsch N, Arnold N, Plendl HJ, Mundhenke C, Niederacher D, Fleisch M, Sutter C, Bartram CR, Dikow N, Wang-Gohrke S, Gadzicki D, Steinemann D, Kast K, Beer M, Varon-Mateeva R, Gehrig A, Weber BH, Stoppa-Lyonnet D, Sinilnikova OM, Mazoyer S, Houdayer C, Belotti M, Gauthier-Villars M, Damiola F, Boutry-Kryza N, Lasset C, Sobol H, Peyrat JP, Muller D, Fricker JP, Collonge-Rame MA, Mortemousque I, Nogues C, Rouleau E, Isaacs C, De Paepe A, Poppe B, Claes K, De Leeneer K, Piedmonte M, Rodriguez G, Wakely K, Boggess J, Blank SV, Basil J, Azodi M, Phillips

- KA, Caldes T, de la Hoya M, Romero A, Nevanlinna H, Aittomaki K, van der Hout AH, Hogervorst FB, Verhoef S, Collee JM, Seynaeve C, Oosterwijk JC, Gille JJ, Wijnen JT, Gomez Garcia EB, Kets CM, Ausems MG, Aalfs CM, Devilee P, Mensenkamp AR, Kwong A, Olah E, Papp J, Diez O, Lazaro C, Darder E, Blanco I, Salinas M, Jakubowska A, Lubinski J, Gronwald J, Jaworska-Bieniek K, Durda K, Sukiennicki G, Huzarski T, Byrski T, Cybulski C, Toloczko-Grabarek A, Zlowocka-Perlowska E, Menkiszak J, Arason A, Barkardottir RB, Simard J, Laframboise R, Montagna M, Agata S, Alducci E, Peixoto A, Teixeira MR, Spurdle AB, Lee MH, Park SK, Kim SW, Friebel TM, Couch FJ, Lindor NM, Pankratz VS, Guidugli L, Wang X, Tischkowitz M, Foretova L, Vijai J, Offit K, Robson M, Rau-Murthy R, Kauff N, Fink-Retter A, Singer CF, Rappaport C, Gschwantler-Kaulich D, Pfeiler G, Tea MK, Berger A, Greene MH, Mai PL, Imyanitov EN, Toland AE, Senter L, Bojesen A, Pedersen IS, Skytte AB, Sunde L, Thomassen M, Moeller ST, Kruse TA, Jensen UB, Caligo MA, Aretini P, Teo SH, Selkirk CG, Hulick PJ and Andrulis I. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA* 2015; 313: p. 1347-61.
23. Roa BB, Boyd AA, Volcik K, Richards CS. Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2. *Nat Genet* 1996; 14: p. 185-7.
24. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; 27: p. 573-80.
25. Borgstrom E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. *PLoS One* 2011; 6: p. e19119.
26. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shamas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol* 2011; 12: p. R1.

27. Rebbeck TR, Kantoff PW, Krithivas K, Neuhausen S, Blackwood MA, Godwin AK, Daly MB, Narod SA, Garber JE, Lynch HT, Weber BL, Brown M. Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. *Am J Hum Genet* 1999; 64: p. 1371-7.
28. Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 1995; 57: p. 289-300.
29. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979; 6: p. 65-70.
30. Zhang C, Lv GQ, Yu XM, Gu YL, Li JP, Du LF, Zhou P. Current evidence on the relationship between HRAS1 polymorphism and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat* 2011; 128: p. 467-72.
31. Clark J and Smith SS. Secondary structure at a hot spot for DNA methylation in DNA from human breast cancers. *Cancer Genomics Proteomics* 2008; 5: p. 241-51.
32. Mao Q, Qiu M, Dong G, Xia W, Zhang S, Xu Y, Wang J, Rong Y, Xu L, Jiang F. CAG repeat polymorphisms in the androgen receptor and breast cancer risk in women: a meta-analysis of 17 studies. *Onco Targets Ther* 2015; 8: p. 2111-20.
33. Pu SY, Yu Q, Wu H, Jiang JJ, Chen XQ, He YH, Kong QP. ERCC6L, a DNA helicase, is involved in cell proliferation and associated with survival and progress in breast and kidney cancers. *Oncotarget* 2017; 8: p. 42116-42124.
34. Liu J, Sun J, Zhang Q, Zeng Z. shRNA knockdown of DNA helicase ERCC6L expression inhibits human breast cancer growth. *Mol Med Rep* 2018; 18: p. 3490-3496.
35. Bahramian S, Sahebi R, Roohinejad Z, Delshad E, Javid N, Amini A, Razavi AE, Shafiee M, Shamsabadi FT. Low expression of LncRNA-CAF attributed to the high expression of HIF1A in esophageal squamous cell carcinoma and gastric cancer patients. *Mol Biol Rep* 2022; 49: p. 895-905.
36. Gilkes DM and Semenza GL. Role of hypoxia-inducible factors in breast cancer metastasis. *Future Oncol* 2013; 9: p. 1623-36.

37. van der Groep P, Bouter A, Menko FH, van der Wall E, van Diest PJ. High frequency of HIF-1alpha overexpression in BRCA1 related breast cancer. *Breast Cancer Res Treat* 2008; 111: p. 475-80.
38. van der Groep P, van Diest PJ, Smolders YH, Ausems MG, van der Luijt RB, Menko FH, Bart J, de Vries EG, van der Wall E. HIF-1alpha overexpression in ductal carcinoma in situ of the breast in BRCA1 and BRCA2 mutation carriers. *PLoS One* 2013; 8: p. e56055.
39. Lu D, Song J, Lu Y, Fall K, Chen X, Fang F, Landen M, Hultman CM, Czene K, Sullivan P, Tamimi RM, Valdimarsdottir UA. A shared genetic contribution to breast cancer and schizophrenia. *Nat Commun* 2020; 11: p. 4637.
40. Chang H, Xiao X, Li M. The schizophrenia risk gene ZNF804A: clinical associations, biological mechanisms and neuronal functions. *Mol Psychiatry* 2017; 22: p. 944-953.
41. Klockmeier K, Silva Ramos E, Rasko T, Marti Pastor A, Wanker EE. Schizophrenia risk candidate protein ZNF804A interacts with STAT2 and influences interferon-mediated gene transcription in mammalian cells. *J Mol Biol* 2021; 433: p. 167184.
42. Lee HJ, Song JY, Kim JW, Jin SY, Hong MS, Park JK, Chung JH, Shibata H, Fukumaki Y. Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia. *Behav Brain Funct* 2005; 1: p. 15.
43. Li X, Su X, Liu J, Li H, Li M, andMe Research T, Li W, Luo XJ. Transcriptome-wide association study identifies new susceptibility genes and pathways for depression. *Transl Psychiatry* 2021; 11: p. 306.
44. Saviouk V, Moreau MP, Tereshchenko IV, Brzustowicz LM. Association of synapsin 2 with schizophrenia in families of Northern European ancestry. *Schizophr Res* 2007; 96: p. 100-11.
45. Depienne C andMandel JL. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet* 2021; 108: p. 764-785.
46. Antoniou AC, Spurdle AB, Sinilnikova OM, Healey S, Pooley KA, Schmutzler RK, Versmold B, Engel C, Meindl A, Arnold N, Hofmann W, Sutter C, Niederacher D, Deissler H, Caldes T, Kampjarvi K, Nevanlinna H, Simard J, Beesley J, Chen X, Kathleen Cuningham Consortium for Research into

- Familial Breast C, Neuhausen SL, Rebbeck TR, Wagner T, Lynch HT, Isaacs C, Weitzel J, Ganz PA, Daly MB, Tomlinson G, Olopade OI, Blum JL, Couch FJ, Peterlongo P, Manoukian S, Barile M, Radice P, Szabo CI, Pereira LH, Greene MH, Rennert G, Lejbkowitz F, Barnett-Griness O, Andrulis IL, Ozelik H, Ocn, Gerdes AM, Caligo MA, Laitman Y, Kaufman B, Milgrom R, Friedman E, Swedish B, collaborators Bs, Domchek SM, Nathanson KL, Osorio A, Llort G, Milne RL, Benitez J, Hamann U, Hogervorst FB, Manders P, Ligtenberg MJ, van den Ouweland AM, collaborators D-H, Peock S, Cook M, Platte R, Evans DG, Eeles R, Pichert G, Chu C, Eccles D, Davidson R, Douglas F, Embrace, Godwin AK, Barjhoux L, Mazoyer S, Sobol H, Bourdon V, Eisinger F, Chompret A, Capoulade C, Bressac-de Paillerets B, Lenoir GM, Gauthier-Villars M, Houdayer C, Stoppa-Lyonnet D, Gemo, Chenevix-Trench G, Easton DF, Cimba. Common breast cancer-predisposition alleles are associated with breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Am J Hum Genet* 2008; 82: p. 937-48.
47. Antoniou AC, Hardy R, Walker L, Evans DG, Shenton A, Eeles R, Shanley S, Pichert G, Izatt L, Rose S, Douglas F, Eccles D, Morrison PJ, Scott J, Zimmern RL, Easton DF, Pharoah PD. Predicting the likelihood of carrying a BRCA1 or BRCA2 mutation: validation of BOADICEA, BRCAPRO, IBIS, Myriad and the Manchester scoring system using data from UK genetics clinics. *J Med Genet* 2008; 45: p. 425-31.

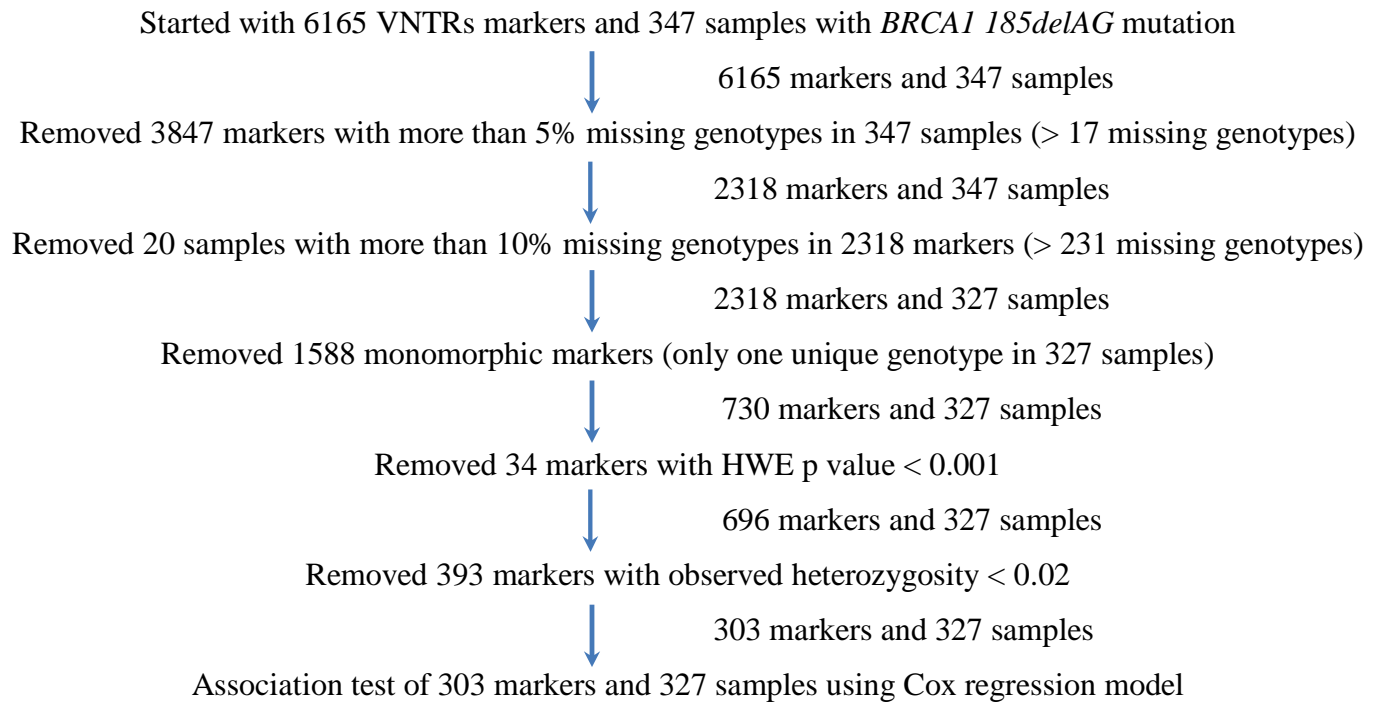


Figure 1. Flow diagram of process and result of VNTR marker and sample filtering.

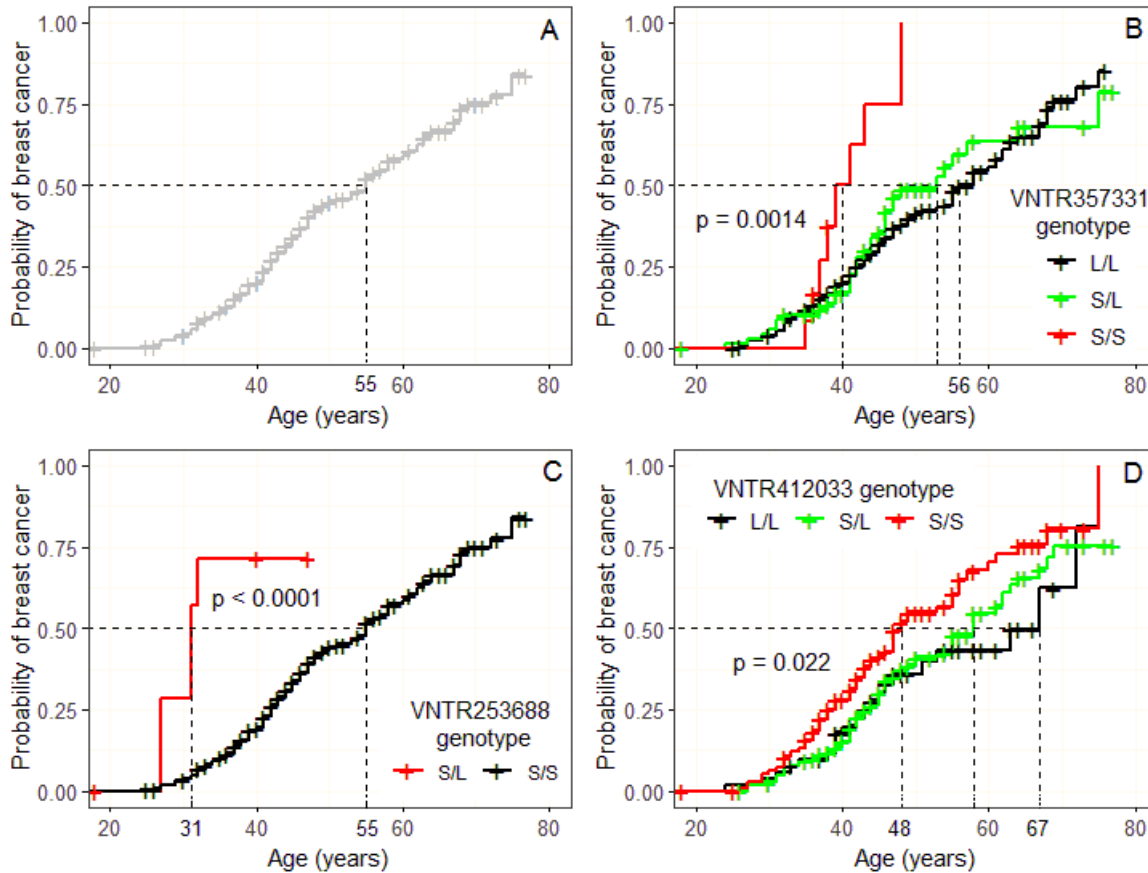


Figure 2. Kaplan-Meier estimates of the cumulative probability of breast cancer diagnosis. The age at breast cancer diagnosis is on the X-axis and proportion of participants diagnosed with breast cancer is on the Y-axis. The horizontal/vertical dash line is the median age at diagnosis of breast cancer. In this step function of breast cancer risk over age, in panel A, the cumulative incidence of breast cancer for all *BRCA1* 185 delAG carriers is shown. For panels B, C, and D, the Kaplan-Meier curves for each of the three VNTRs with FDR < 0.05 are shown. Panel B is VNTR 357331, Panel C is VNTR253688, and panel D is VNTR412033. The black line is the reference genotype, the green line (when present) is the heterozygote genotype, and the red is the risk genotype. For each of the VNTRs, there were significantly different risks by genotype (log-rank p value < 0.05) with earlier ages of developing breast cancer among participants carrying the risk genotypes.

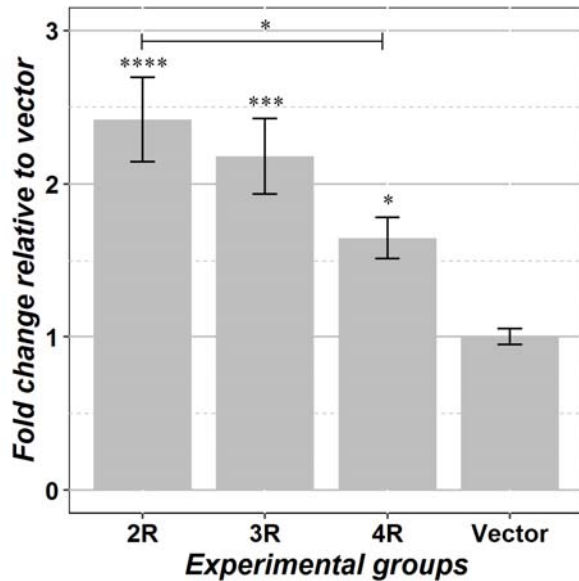


Figure 3. Association of VNTR558420 with *ZNF501* gene expression by luciferase assay. Each experimental group is composed of 12 data points. Data represent fold change in the repeat group relative to vector group, with standard error bar shown for each group. Significance was assessed by one-way ANOVA with pairwise t test and P-value adjusted by the Tukey's method. Asterisk above standard error bar indicates significance test between the repeat group and vector group; asterisk above the line indicates the significance between the 2R and 4R repeat groups; *P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

Table 1. Participant characteristics

	N (%)	Age in years (Mean, Range)
Sample origin		
US	97 (28.0)	43.9, 18 - 72
Israel	250 (72.0)	46.8, 24 - 77
Cancer status		
Breast	153 (44.1)	43.5, 24 - 75
Ovarian	60 (17.3)	54.0, 35 - 77
Breast & Ovarian	17 (4.9)	48.0, 27 - 68
Unaffected	117 (33.7)	44.9, 18 - 76

Table 2. Summary of repeat alleles in the 303 VNTRs in 327 female *BRCA1 185delAG* mutation carriers

# repeat alleles	# VNTRs	Heterozygosity Median(range)	Repeat motif length range (bp)	# VNTRs of a given repeat motif length			
				3 to 5 bp	6 to 10 bp	11 to 20 bp	≥ 21 bp
2	173	0.14 (0.02 - 0.52)	3 to 51	26	34	71	42
3	71	0.13 (0.02 - 0.64)	3 to 49	35	19	11	6
4	25	0.28 (0.02 - 0.59)	4 to 23	17	6	1	1
5	16	0.46 (0.12 - 0.72)	3 to 14	11	4	1	
6	7	0.56 (0.18 - 0.69)	4 to 6	6	1		
7	6	0.63 (0.52 - 0.82)	4 to 7	5	1		
8	2	0.72 (0.67 - 0.77)	4	2			
9	3	0.70 (0.47 - 0.78)	4	<u>3</u>			
				105	65	84	49

bp: base pair

Table 3. Association of VNTR with breast cancer risk in female carriers of *BRCA1 185delAG*

VNTR ID ^a	Trend test, VNTR as a continuous variable			Categorical test to determine cut point of risk repeats in a VNTR			
	P	FDR	HR (95% CI)	Repeat number / critical cut point ^b	Reference, risk genotype ^c	P	HR (CI)
253688	9.1E-05	0.024	6.12 (2.47 - 15.15)	5, 9 / 10	S/S, S/L	9.1E-05	6.12 (2.47 - 15.15)
357331	2.6E-04	0.035	3.57 (1.79 - 7.14)	4 / 5, 6	L/L, S/S	2.9E-04	3.59 (1.80 - 7.16)
472060	3.9E-04	0.035	2.54 (1.52 - 4.25)	4 / 9, 13	L/L, S/L	6.8E-01	0.81 (0.30 - 2.19)
412033	6.9E-04	0.046	1.35 (1.14 - 1.59)	7, 8 / 9, 10, 11	L/L, S/S	3.4E-02	1.69 (1.04 - 2.75)
558420	2.5E-03	0.135	6.06 (1.88 - 19.51)	2, 3 / 4	S/S, S/L	2.5E-03	6.06 (1.88 - 19.51)
735300	4.1E-03	0.183	3.08 (1.43 - 6.65)	5 / 6	S/S, S/L	4.1E-03	3.08 (1.43 - 6.65)
945060	4.8E-03	0.183	1.60 (1.15 - 2.23)	6, 7, 8, 9 / 10	S/S, S/L	1.1E-02	2.73 (1.26 - 5.94)
549198	2.0E-03	0.209	2.94 (1.49 - 5.92)	4 / 5	L/L, S/L	2.0E-03	2.97 (1.49 - 5.91)

a. VNTR IDs were assigned in this project and their genomic annotations are provided in supplemental table 1; **b:** in categorical test to determine cut point of risk repeat for a VNTR by a sliding window method, the forward slash “/” character is the final critical cut point based on smaller p values among the multiple tests of cut points, greater or equal to repeat number at right side of slash is defined as long allele (L), repeat numbers at left side of slash as short allele (S); **c:** in Cox regression model using categorical VNTR genotype as predict variable, hazard ratio was calculated by comparing risk genotype to reference genotype.