The relation of heterozygous *BRCA1* and *BRCA2* germline variants in the onset of pediatric cancer

By

Anne Sophie Blokland

Internship Report

Princess Máxima Center for Pediatric Oncology

MSc Bioinformatics and Biocomplexity

Utrecht University

SUPERVISORS

Michelle Kleisman
Prof. dr. Roland Kuiper
Nienke van Engelen
Princess Máxima Center

EXAMINERS

Prof. dr. Roland Kuiper
Dr. Patrick Kemmeren
Princess Máxima Center

DATE

7 February 2023

Table of Contents

Layman summary

Kanker in kinderen is een proces dat nog niet zo goed wordt begrepen. Ongeveer 10% van alle kinderkankers wordt veroorzaakt door een genetische aandoening, waardoor kinderen een hoger risico hebben om kanker te ontwikkelen. Mutaties in de *BRCA1* en *BRCA2* genen worden vaak geassocieerd met een hoger risico op het ontwikkelen van borst en eierstok kanker in volwassenen. Steeds vaker zien we ook mutaties in deze genen bij kinderen met kanker. Verschillende onderzoeken suggereren dat een pathogene kiembaanvariant in één allel in een van deze genen het risico op het ontwikkelen van hersen en solide tumoren bij kinderen verhoogt. Echter, dit is nog steeds niet bewezen.

Dit onderzoek ging voornamelijk over het aantonen van de aanwezigheid van een potentiële tweede genetische gebeurtenis die een rol zou kunnen hebben gespeeld bij het ontstaan van de kinderkanker bij patiënten die een pathogene *BRCA1* of *BRCA2* kiembaanvariant in één allel dragen. Voor dit onderzoek werden er twee verschillende analyses gedaan. Ten eerste werd er gekeken naar de somatische mutatiepatronen van patiënten met kinderkanker om te onderzoeken of er sprake was van een verlies van het homologe recombinatie reparatie mechanisme. Dit reparatie mechanisme is erg belangrijk voor het herstellen van fouten in het DNA, en een verlies hiervan kan zorgen voor onbalans van de allelen en in sommige gevolgen ook kanker. Er werd gekeken of bepaalde mutatiepatronen die vaak worden gezien bij het verlies van het reparatie mechanisme in volwassenen overeen kwamen met de mutatiepatronen van de kinderen met kanker. Ook werd er met een algoritme, genaamd CHORD, een score berekend die een voorspelling doet over hoe waarschijnlijk het is dat iemand een verlies van het reparatie mechanisme heeft. Daarnaast werd gekeken of varianten die veel voorkomen in de populatie en een hoge associatie hebben met een verlaagde expressie van het *BRCA2* gen, ook wel eQTLs genoemd, voorkomen op het kiembaan allel zonder de pathogene variant van de kinderen. De aanwezigheid van eQTLs in de kiembaan zouden eventueel kunnen leiden tot onbalans van de allelen.

De resultaten toonden aan dat een verlies van het reparatie mechanisme alleen werd waargenomen bij kinderkanker patiënten met twee pathogene kiembaanvarianten op beide allelen van het *BRCA2* gen of een pathogene *BRCA2* kiembaanvariant en een somatische tweede hit. Er werd nog geen verband gevonden tussen pathogene *BRCA1* of *BRCA2* kiembaanvarianten op één allel en het ontstaan van kinderkanker. Verder werden er wel eQTLs gevonden op de kiembaan allelen van de kinderen met kanker, alleen is het onduidelijk of deze eQTLs voorkomen op het allel zonder de pathogene variant en of deze eQTLs ook daadwerkelijk een effect hebben. Om de resultaten beter te begrijpen moet opvolgend onderzoek worden gedaan naar de potentiële tweede genetische gebeurtenis die de tumorontwikkeling zou kunnen hebben veroorzaakt. Ook zijn er meer patiënten nodig in deze studie om de resultaten beter te begrijpen.

Abstract

Childhood cancer is a complex and poorly understood process. It is estimated that around 10% of pediatric cancer cases are due to a genetic condition known as a cancer predisposition syndrome (CPS). Variants in adult cancer predisposing genes (CPGs), *BRCA1* and *BRCA2*, have been found to be enriched in pediatric cancer patients and are often associated with breast and ovarian cancer in adults. Suggestions are made that heterozygous germline variants in these genes can increase the risk of developing brain and solid tumors in children, although the relationship between heterozygous *BRCA1* and *BRCA2* germline variants and pediatric cancer is not yet demonstrated.

This study aimed to examine the presence of a potential second genetic event in pediatric patients with a heterozygous *BRCA1* or *BRCA2* germline variant. The study employed two approaches. First, somatic mutational pattern analysis was conducted to assess the presence of homologous recombination deficiency (HRD), which is associated with most cancers in adult patients with a *BRCA1* or *BRCA2* germline variant. Hence, a comparison of HRD-related mutational signatures and the use of CHORD aimed to assess the presence of HRD in pediatric patients with a heterozygous *BRCA1* or *BRCA2* germline variant. Next, the study aimed to investigate the relationship between expression quantitative trait loci (eQTLs) on the non-pathogenic allele of the *BRCA2* gene in pediatric patients and allelic imbalance by comparing germline variants of patients with the presence of eQTLs.

Results showed that HRD was only observed in pediatric patients with a biallelic *BRCA2* germline variant or heterozygous *BRCA2* germline variant and a somatic second hit. However, no relation was found between heterozygous *BRCA1* or *BRCA2* germline variants and HRD, and therefore the onset of pediatric cancer. Moreover, the study found a slightly higher percentage of cis-eQTLs within patients with a *BRCA2* germline variant compared to the healthy population, but not compared to the AYAs patients with bladder cancer. Additionally, seven clusters of co-occurring eQTLs were found. Accordingly, read-based phasing and RNA expression analysis are required to validate whether combinations of cis-eQTL clusters are present on the non-pathogenic allele and therefore could induce allelic imbalance leading to tumor development. Due to a limited sample size, further research is required to better understand the findings.

List of abbreviations

AYA	Adolescents and young adults
BAM	Binary alignment map
CNV	Copy number variation
COSMIC	Catalogue of Somatic Mutations in Cancer
CPG	Cancer predisposing gene
CPS	Cancer predisposing syndrome
CRAM	Compressed reference-oriented alignment map
eQTL	Expression quantitative trait loci
GATK	Genome analysis toolkit
GnomAD	Genome Aggregation Database
GoNL	Genome of the Netherlands
HMF	Hartwig Medical Foundation
HRR	Homologous recombination repair
HRD	Homologous recombination deficiency
ID	Insertions and deletions (indels)
SBS	Single base substitution
SNP	Single nucleotide polymorphisms
SNV	Single nucleotide variation
SV	Structural variant
TMB	Tumor mutational burden
VAF	Variant allele frequency
VCF	Variant call format
VEP	Variant Effect Predictor
WGS	Whole genome sequencing

1. Introduction

The onset of childhood cancer is a complex and poorly understood process. While most cases are due to coincidence and environmental factors, more cases are thought to be a genetic disorder and may be related to cancer predisposition syndromes (CPS) (Ripperger et al., 2017). CPS is the condition in which a genetic variant, inherited or de novo, increases the chance of developing cancer at a younger age compared to the general population. Approximately 10% of cancer among children is due to this syndrome (Kratz et al., 2021). Recent research has shown an enrichment of germline pathogenic variants (PVs) in adult cancer predisposing genes (CPGs) (e.g. *BRCA1, BRCA2, MLH1, MSH2* & *MSH6)* in pediatric cancer patients (Zhang et al., 2015; Kratz et al., 2022). However, the association between heterozygous PVs in adult CPGs and their role of the onset of pediatric cancer is not yet demonstrated.

A higher frequency of germline PVs in the adult CPGs *BRCA1* and *BRCA2* has been found in pediatric cancer patients (Kratz *et al.*, 2022). These genes are often associated with hereditary breast and ovarian cancer (HBOC), pancreatic and gastric cancer in adults (Petrucelli, Daly and Pal, 1993; Decker et al., 2016). Cancer development in these adults can be due to inherited *BRCA1* or *BRCA2* PVs which combined with a somatic second hit or loss-of-heterozygosity (LOH) can cause tumor formation. Another way is through having a biallelic germline variant, such as in the disease Fanconi anemia (Woodward and Meyer, 2021; Kratz et al., 2022). In both cases there is prominent evidence that there is complete loss of *BRCA1* or *BRCA2* gene function. Besides, an increased risk of breast and ovarian cancers is found in adults with a heterozygous *BRCA1* and *BRCA2* germline variant (Nguyen et al., 2020), while knowledge of the relationship between heterozygous *BRCA1* and *BRCA2* germline variants and the risk of childhood cancer is still lacking. Extensive research of germline variants in pediatric patients has been performed, which shows an enrichment of heterozygous variants in adult CPGs *BRCA1* and *BRCA2* associated with a higher risk of developing brain and solid tumors in children. This could suggest a causal role of heterozygous germline variants in these CPGs in the onset of cancer in a subset of pediatric cancer cases, which link seems particularly strong in the *BRCA2* gene (Zhang et al., 2015; Kratz et al., 2022).

In most (adult) cancer patients with a germline variant in the *BRCA1* and *BRCA2* genes, an association is found with a defective homologous recombination repair (HRR) pathway. The HRR pathway is responsible for repairing various genetic lesions in DNA (Creeden et al., 2021). The identification of homologous recombination deficiency (HRD) in pediatric cancer patients harbouring germline variants in the *BRCA1* and *BRCA2* genes is crucial to provide valuable information on the combined impact of a heterozygous germline PV in the *BRCA* genes and other genetic factors which may have contributed to the development of pediatric cancer.

Genetic aberrations caused by HRD can be identified as genomic signatures or footprints (Nguyen et al., 2020). A mutational signature is a unique pattern of mutations that was caused by specific processes that led to the development of a tumor. Signatures resulting from HRD exist dominantly as small insertions and deletions (indels) and structural variants (SVs). However, single base substitutions (SBS) can also display a characteristic pattern due to HRD, defined as mutational signature SBS3 (**Figure 1.A**) (Alexandrov *et al.*, 2013a). Deletions associated with HRD, which are equal or greater than 5 base pairs in size are correlated with indel signature ID6 (**Figure 1.B**)

(Nguyen et al., 2020). These deletions are characterized by having extended stretches of microhomology at the breakpoint junctions (Alexandrov *et al.*, 2013b). Hence, discovering signatures SBS3 and ID6 in somatic data can indicate a defect in the HRR pathway, along with increased genomic instability.

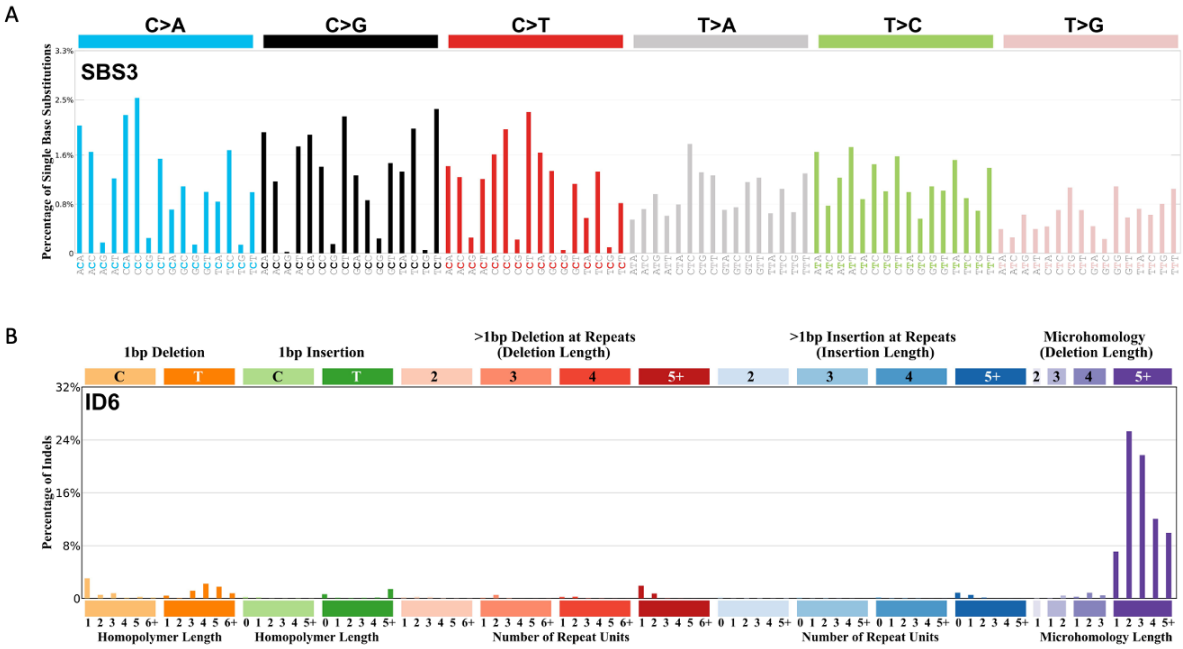


Figure 1. Mutational profiles of SNVs and indels which have been associated with a defective homologous recombination repair pathway. A. Signature SBS3 is a specific pattern of mutations in the DNA that is characterised by the presence of six specific substitution mutations: C>A, C>G, C>T, T>A, T>C, and T>G. The height of each bar represents what percentage of the total number of mutations in the signature is made by one specific mutation subtype. **B.** Deletions with a length of > 5 bp with microhomology is a characteristic for ID6 and an important feature in the relation to HRD.

Several whole genome sequencing (WGS) classifiers, based on machine-learning algorithms, are developed to give more insight in the role of *BRCA1* and *BRCA2* germline variants in the onset of cancer. These algorithms can predict *BRCA1* and *BRCA2* deficiencies based on somatic mutational patterns, and provide more insight into whether a patient has an impaired HRR pathway or not. Several algorithms were evaluated for the prediction of HRD in the *BRCA1* and *BRCA2* genes. CHORD (Classifier of HOmologous Recombination Deficiency), a random forest model, deemed most effective in the HRD prediction (Nguyen *et al.*, 2020; Štancl *et al.*, 2022). CHORD assigns a positive or negative HRD classification score based on the presence and frequency of somatic mutations that are associated with HRD. In addition, identification of SVs enables the distinction between *BRCA1*-type HRD and *BRCA2*-type HRD, which can be differentiated by the presence of structural duplications ranging in size from 1–100 kb (Nguyen et al., 2020). Although CHORD is based on adult tumor data, we aim to investigate if predictions made by CHORD can be used to determine potential occurrence of HRD in pediatric cancer patients with a heterozygous *BRCA1* or *BRCA2* germline variant.

The analysis of somatic genomic data can provide valuable information regarding the presence of HRD in pediatric cancer patients with germline PVs in the *BRCA1* and *BRCA2* genes. Although germline variants can contribute to the development of cancer in children, they are often not the sole cause (Zhang et al., 2015). For example, the functional copy of the gene may not be able to

compensate for the loss of the pathogenic allele. This causes a decrease in normal function of the gene, known as haploinsufficiency (Karaayvaz-Yildirim et al., 2020). The reduction in function of the wildtype allele of *BRCA1* and *BRCA2* can be attributed to the presence of certain single nucleotide polymorphisms (SNPs) that have an reducing impact on gene expression. The eQTLGen Consortium has compiled a database containing information about the effects of SNPs that are related to specific traits (Võsa et al., 2021). These variants, better known as expression quantitative trait loci (eQTLs), are located in regulatory regions and may have the ability to modulate gene expression (Li et al., 2013). There are two main types of eQTLs: cis-eQTLs and trans-eQTLs. Cis-eQTLs are located on or near (<1 Mb) the gene whose expression they affect, while trans-eQTLs affect gene expression through long-range interactions located on a different chromosomes (Võsa et al., 2021). Cis-eQTLs are extensively studied and utilized, from which the effect is represented in **Figure 2**. Determining eQTLs and deducing a pattern leading to haploinsufficiency might give insight into allelic imbalance and its possible role in the onset of childhood cancer.

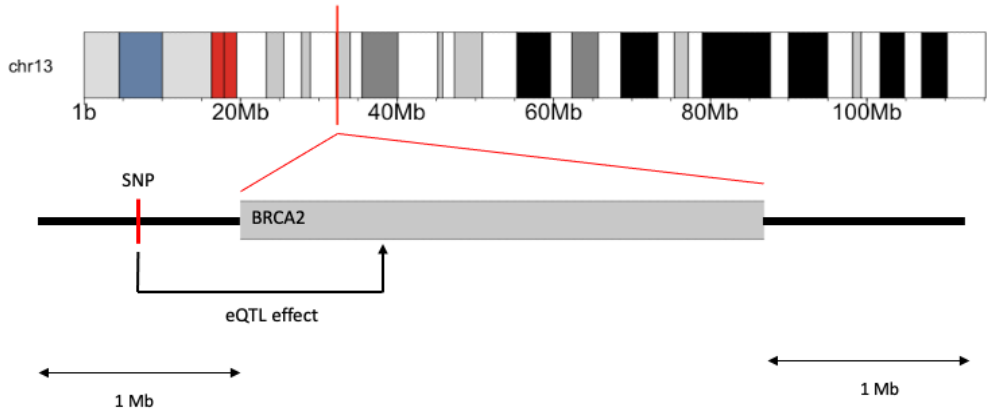


Figure 2. A typical effect of a cis-eQTL on the expression of the *BRCA2* gene. The *BRCA2* gene located on chromosome 13 is represented. Certain SNPs located within 1 Mb near or on the *BRCA2* gene can induce allelic imbalance and affect the *BRCA2* protein expression.

The aim of the project was to investigate the relationship between pediatric cancer patients with a germline *BRCA1* or *BRCA2* variant and the identification of a second genetic factor, such as a defective HRR pathway, that may have played a role in the onset of pediatric cancer. Additionally, the study aimed to determine the presence of eQTLs that could reduce the expression of the non-pathogenic allele, in order to gain a deeper understanding of the various genetic factors that could contribute to pediatric cancer.

2. Method

2.1. Workflow overview

In this section, the methods used to predict HRD in patients with a *BRCA1* or *BRCA2* germline variant and to compare the patient's germline data with the presence of eQTLs are explained in detail. A visual representation of these procedures is provided in **Figure 3**.

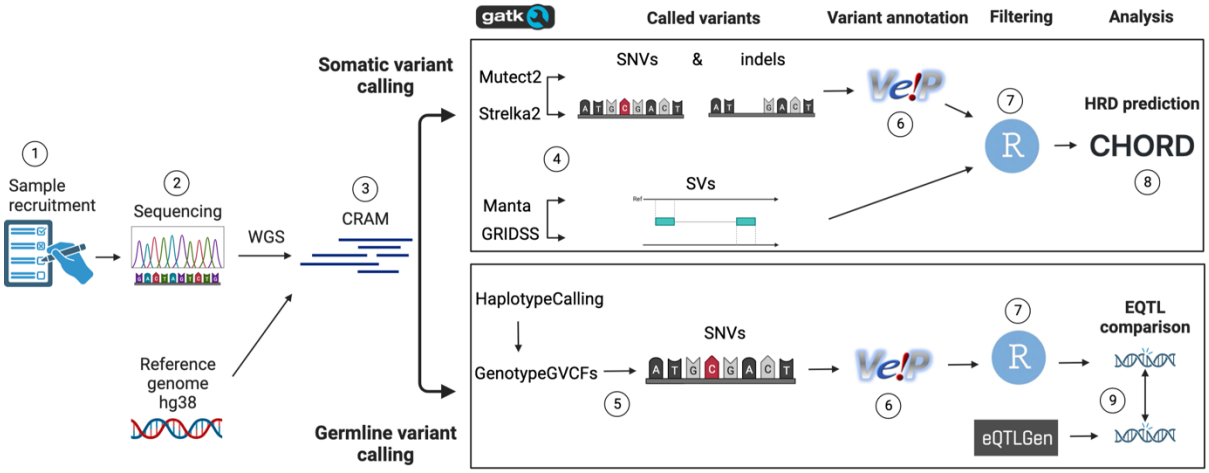


Figure 3. A workflow for *BRCA1* and *BRCA2* somatic and germline variant analysis: HRD Prediction and eQTL comparison. DNA samples from patients with cancer and a heterozygous *BRCA1* or *BRCA2* germline variant were collected (1). The samples were processed using high-throughput sequencing techniques (2). The data was aligned to the human reference genome and stored in CRAM format (3). Somatic variants were called using Mutect2 and Strelka2 for SNV and indel calling. Manta and GRIDSS were used for SV calling (4). Germline variants (SNVs) were called with HaplotypeCaller and GenotypeGVCFs (5). All variants, except SVs, were annotated with the VEP tool (6) and filtered based on various criteria (7). A HRD score based on somatic mutations was predicted with CHORD (8) and germline variants were compared with the presence of eQTLs (9).

2.2. Sample selection

Whole genome sequencing (WGS) data of germline and tumor samples was collected for patients with a germline *BRCA1* or *BRCA2* PV. Samples were retrieved from the Princess Máxima Center in Utrecht and from the INFORM study database. Seven patients were selected for further germline and somatic analysis (**Supplementary Table 1)**.

25 WGS samples of adult patients with a germline *BRCA2* PV and HRD were used as positive controls to reproduce and validate results (Priestley *et al.*, 2019; de Witte *et al.*, 2022). All samples were requested from the Hartwig Medical Foundation (HMF) databank and used for eQTL comparison analysis. Six samples were excluded for HRD prediction analysis due to a low tumor purity (≤0.3). An overview of the adult patient data and their characteristics is represented in **Supplementary Table 2**. Data access and patient consent had been obtained from all necessary parties before data analysis was conducted.

116 WGS samples of adolescent and young adult (AYAs) patients with bladder cancer were retrieved from the MOTIEF study to use as negative controls for the eQTL comparison analysis.

2.3. Sequencing information

WGS for all samples was performed using the Illumina NovaSeq 6000 platform sequencing technique. The sequencing depth was targeted at 30X for pediatric germline samples and 100X for pediatric tumor samples. For samples requested by the HMF databank, the sequencing depth was targeted at 50X for germline samples, and at 150X for tumor samples. CRAM files were received from external institutions and processed with the in-house pipeline of the Princess Máxima Center (**2.14.**)

2.4. Data analysis

Sequencing data of pediatric patients was stored in FASTQ files. These files were converted to unaligned BAM (uBAM) format using the PICARD v2.10.10 fastqToSam tool prior to read mapping (*Picard Tools - By Broad Institute*, no date). The sequencing reads were then aligned to version GRCh38 of the human reference genome using the Burrows-Wheeler Aligner (BWA) v0.7.13 (Li and Durbin, 2009). Mapped reads were stored in CRAM format and could be used for variant calling. HMF data was originally mapped to version GRCh37 of the human reference genome. Therefore, CRAM files were converted to uBAM format using the PICARD RevertSam tool, and remapped to version GRCh38 of the human reference genome.

2.5. Somatic variant calling and filtering

Somatic variant calling of SNVs and indels was conducted with The Genome Analysis Toolkit (GATK) version 4.1.1.0 Mutect2 and Strelka2, using CRAM files as input (McKenna *et al.*, 2010; Saunders *et al.*, 2012; Benjamin *et al.*, 2019). The Mutect2 pipeline allowed the exclusion of soft clipped bases, which helps to minimize noise in the analysis of the 3' and 5' untranslated regions. Variants with a PASS filter were selected using GATK SelectVariants and GATK FilterMutectCalls, to assure the probability of identifying true variants (Auwera and O'Connor, 2020). Variant calls by Strelka2 were filtered by extracting SNVs and indels with a PASS filter from the generated VCF file. All somatic variant information was stored in VCF files.

Variants were annotated with Ensembl Variant Effect Predictor (VEP) version 105 (McLaren *et al.*, 2016), which provided information on how each variant may affect the corresponding protein. Population frequency data from the Genome Aggregation Database (gnomAD) version 3.0 (Karczewski *et al.*, 2020) and the Genome of the Netherlands (GoNL) (Boomsma *et al.*, 2014) were added to provide information about the prevalence of the variants in healthy populations.

Somatic variant calls by Mutect2 and Strelka2 were filtered using R version 4.2.2 (RStudio Team, 2020). Variants in the centromeric regions were removed from the VCF file due to their repetitive and complex nature (Lamb and Birchler, 2003). Additionally, variants with reads in the germline sample and a population frequency higher than 1% in the gnomAD or GoNL database were filtered out to exclude common variants. SNVs were selected based on a minimum coverage of 20X, more than 5 supporting reads of the alternative allele and a variant allele frequency (VAF) of at least 0.25. Indels were selected based on a minimum coverage of 20X, more than 4 supporting reads of the alternative allele and a VAF of at least 0.2. The criteria differ because indels are less common compared to SNVs and CHORD requires a minimum of 50 indels to predict a HRD score.

2.6. Somatic SV calling

Manta and GRIDSS were used for detection of structural variants (SVs) (Chen *et al.*, 2016; Cameron *et al.*, 2017). Manta required a CRAM file as input whereas CRAM files had to be converted to BAM format for SV calling with GRIDSS. Somatic SV information was stored in a VCF file. R was used to annotate SVs by defining SV types and lengths, and to filter the data to select only variants with a PASS filter. To further reduce the number of false positive variants and improve accuracy of the results, a minimal of 5 paired and split alternative reads was set. However, the filtering parameters were adjusted if there were less then 30 SVs per sample left as CHORD required a minimum of 30 SVs per sample to calculate HRD scores.

2.7. Mutational profile analysis

The R package MutationalPatterns version 3.4.1 was used to reconstruct mutational profiles from somatic SNV and indel data (Blokzijl *et al.*, 2018). These profiles were then compared to the mutational signatures SBS3 and ID6 from the Cancer Genome Project's database (COSMIC v3.3) by calculating a cosine similarity score (Tate *et al.*, 2019). The cosine similarity score represents a measure of similarity between vectors that is based on ranges from 0 to 1 (Alexandrov *et al.*, 2013a). The standard cosine similarity threshold of at least 0.85 was used to indicate a high level of similarity between the pediatric profiles and ID6. However, due to a featureless pattern of SBS3, the standard threshold for similarity was adjusted to a score of at least 0.8 to show a high level of similarity with this signature. The R package BSgenome.Hsapiens.UCSC.hg38 from Bioconductor was used as a reference genome for this analysis.

2.8. Predicting HRD score

The random forest model CHORD v.2.0 was used as a R-package HRD score calculations (Nguyen *et al.*, 2020). The mutational context of the filtered somatic variants was analysed to predict a HR status in each patient, being either HRD-positive or HRD-negative. CHORD generated a HRD probability value per sample, a HR status, together with probability values for the distinguishing *BRCA1*-type HRD from *BRCA2*-type HRD. The cut-off value for indicating the presence of HRD was set at 0.5, the same threshold used in the study by Nguyen *et al.* (2020).

2.9. Copy number variations

Copy number variations (CNVs) were detected to identify all changes in the number of copies of a specific segment within each patient. WGS germline and somatic CRAM files were used as input and CNVs were calculated using GATK version 4.1.7.0. The segment files containing the copy numbers were subsequently annotated in R, where segments of neutral copy numbers between 0.6 and 1.2 were filtered out. Further processing generated CNV plots, which could be used for analysis.

2.10. Germline calling

Germline calling was performed to identify inherited genetic variants. The HaplotypeCaller tool of GATK4 was used to call SNVs and small indels (Poplin *et al.*, 2018) and processes CRAM files to GVCFs. Subsequently, genotype calling was done to determine the genotype at each position. The GenotypeGVCFs tool of GATK4 was used to perform genotype calling with a GVCF as input and VCF as output (Nielsen *et al.*, 2011). The VEP tool was used for variant annotation (McLaren *et al.*, 2016). Population frequencies were added from the gnomAD database and GoNL database

(Boomsma *et al.*, 2014; Karczewski *et al.*, 2020). Germline variants with a VAF of at least 0.1 were selected to increase the biological relevance of selected variants.

2.11. Variant validation in IGV

To ensure the accuracy of germline and somatic variants, a random selection of 75 called variants per sample were further verified with the tool Integrated Genomics Viewer (IGV). IGV was used to visualize the genomic data and to examine whether the called variants were true somatic or germline variants (Robinson *et al.*, 2011). The proportion of true variants per variant caller was calculated to evaluate the performance of the variant caller.

2.12. eQTL comparison analysis

All variants that significantly correlated with negative *BRCA2* gene expression were identified from the eQTLGen Consortium database, downloaded on October 28, 2022 (Võsa *et al.*, 2018). eQTLs were filtered in R based on the following selection criteria: a FDR-adjusted p-value of less than 0.05 to indicate statistical significance and a Z-score of less than -2 to identify variants associated with negative expression of the gene. Selected eQTLs were compared to the germline variants identified from the samples to see possible enrichment of these eQTLs compared to a healthy population.

2.13. Hierarchical clustering

A hierarchical clustering analysis was conducted based on the prevalence of eQTLs to examine the association between a *BRCA2* germline variant and the co-occurrence of eQTLs in each sample within its corresponding cohort. R-package pheatmap version 1.0.12 was used to draw a clustered heatmap (Kolde, 2018).

2.14. Code availability: bitbucket

The codes used in the study for the analysis can be accessed through the following link: https://bitbucket.org/princessmaximacenter/pmc_kuiper_pipelines/src/master/scripts/adultCancerPredisposingGenes/

3. Results

The results are presented in two main sections. First, the study looked at similarities between the patient's somatic data and mutational signatures SBS3 and ID6 and used CHORD to calculate the prediction score for HRD. Next, germline eQTLs located in *BRCA2* gene regions which could potentially reduce gene expression were identified. Hence, eQTL comparison was conducted. The goal was to understand if there was a connection or pattern between these eQTLs, their influence on reduced gene expression and the potential occurrence of allelic imbalance. Comparison was only done for the *BRCA2* gene region as samples with a *BRCA1* germline variant were limited.

Seven patients were selected for germline and somatic analysis, represented in **Table 1**. Additional patient data can be found in **Supplementary Table 1**. Patients P.20034 and P.20003 were selected due to the diagnosis Fanconi Anemia, characterised by biallelic PVs in the *BRCA2* gene. The other five patients carry a germline *BRCA1* or *BRCA2* variant. Only germline data was available for patient P.95988.

PATIENT	GENE	CANCER TYPE	MUTATION CHARACTERISTIC	ADDITIONAL INFORMATION	DATA AVAILABILITY
P.20034	*BRCA2*	High Grade Glioma	Compound heterozygous	Fanconi Anemia	G & T
P.20003	*BRCA2*	Medulloblastoma, Wilms tumor	Homozygous	Fanconi Anemia	G & T (MB)
P.20778	*BRCA2*	Acute Lymphocytic Leukemia	Heterozygous	-	G & T
P.67780	*BRCA2*	Low Grade Glioma	Heterozygous	-	G & T
P.51624	*BRCA2*	Osteosarcoma	Germline variant + somatic 2nd hit	LOH	G & T
P.23410	*BRCA1*	Neuroblastoma	Heterozygous	-	G & T
P.95988	*BRCA2*	Ewing sarcoma	Heterozygous	-	G

Table 1. An overview of all pediatric samples included in the study. For all included tumor data, matched germline control data was available. G = germline, T = Tumor.

The HMF cohort was used as a positive control to validate the pipelines that were used in this study for mutational signature comparison and HRD predictions.

3.1. HRD prediction

3.1.1. Mutational pattern analysis

Somatic mutational pattern characteristics of SNVs and indels of each patient are displayed in **Figure 4** & **Figure 5**, which were compared to the HRD-associated reference signatures SBS3 and ID6 (**Figure 1**). The mutational pattern analysis showed the presence of signature SBS3 in four out of six pediatric patients as a cosine similarity score of at least 0.8 to SBS3 was identified in these patients (**Figure 4**). These patients included three with a biallelic PV or second hit and one with a heterozygous *BRCA1* germline variant.

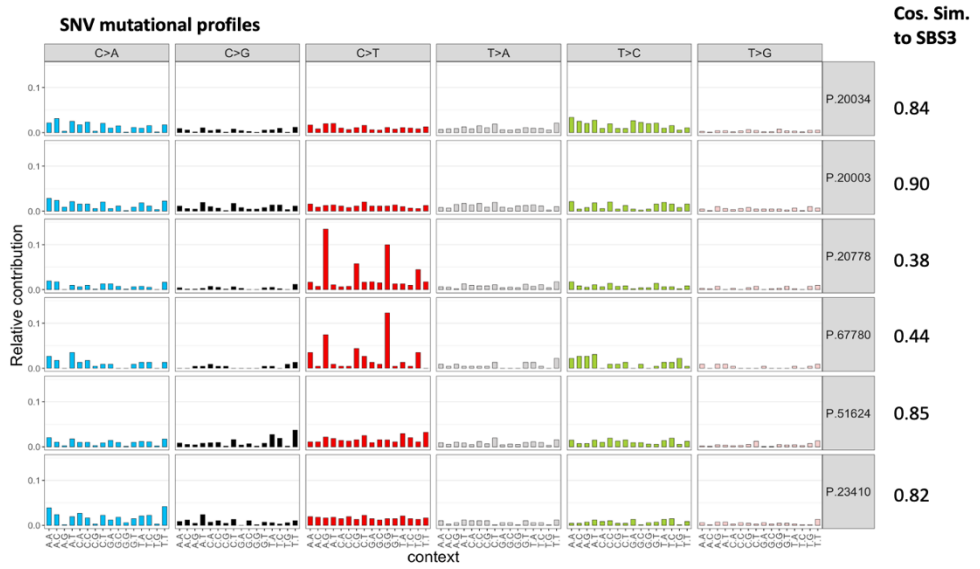


Figure 4. Four patients with a heterozygous *BRCA2* germline variant show similarity to SBS3. Mutational profile plot show the relative contribution of SNVs found in the tumor sample of the patients. A threshold of at least 0.8 is considered to indicate a high level of similarity to COSMIC SBS3.

No correlation was found between mutational signature ID6 and indel patterns of patients as cosine similarities were below 0.85. In addition to comparing the overall mutational profile of indels to ID6, a further analysis was conducted specifically looking at a subset of indel mutations. The subpart 5+ bp deletions with a microhomology pattern is an important feature for HRD prediction. A high similarity was found to the subpart of ID6 as the cosine similarity was equal or higher than 0.85 for the patients with Fanconi anemia and the patient with the combination of a *BRCA2* germline variant and a somatic second hit (**Figure 5)**.

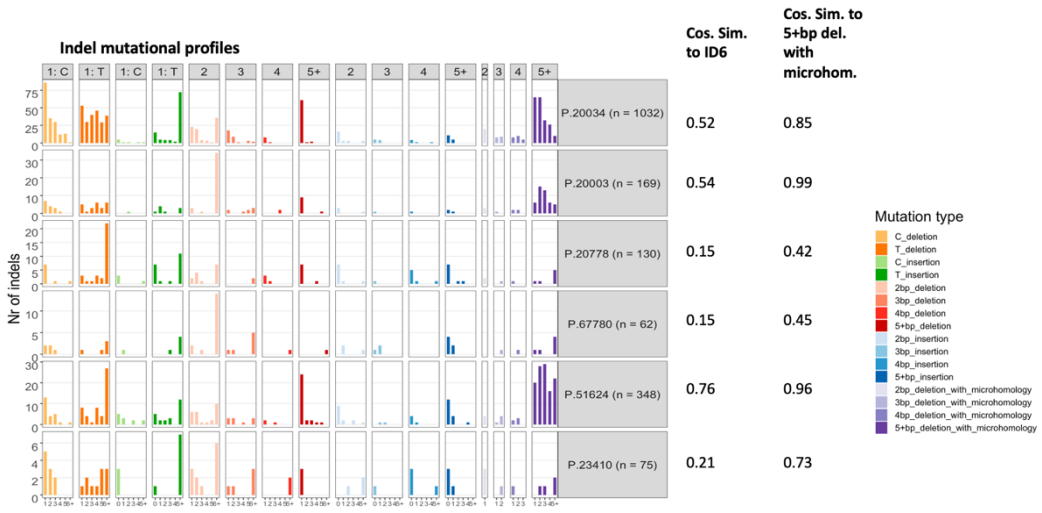


Figure 5. Three patients with a heterozygous *BRCA2* germline variant show the presence of the 5+bp deletion with microhomology subpart of ID6. No similarities are identified when the complete indel mutational profiles of the patients were compared to COSMIC ID6. Three patients show a high similarity to the 5+bp deletion with microhomology subpart of ID6 (≥0.8).

In summary, three patients showed a similarity with the HRD-related mutational profiles as expected by their homozygous variant: two patients with Fanconi Anemia (P.20034, P.20003) and one patient with a combination of a *BRCA2* germline variant and a somatic second hit (P.51624). Patient P.23410 with a heterozygous *BRCA1* germline variant presented a good

similarity to SBS3 but not to the subpart of the ID6, making it challenging to suggest a possible HRR defect. No HRD-association was found for patients with a heterozygous *BRCA2* germline variant.

In **Supplementary Figure 1**, a heatmap is displayed that illustrates the relationship between mutational characteristics of the pediatric patients and various signatures from COSMIC (Tate et al., 2019). Multiple signatures are displayed, but they have a relatively low similarity with the mutational profile of the patients. Many of these signatures are identified as "clock-like" and have no association to any specific trait.

Validation was done for the mutational signature comparison with the HMF control cohort. Mutational profiles were generated of somatic data of the 19 patients and compared to signatures SBS3 and ID6. Fourteen patients showed a cosine similarity above the threshold (≥0.8), indicating a good similarity to SBS3. Furthermore, a good similarity (≥0.85) with signature ID6 was found in 8 patients (**Supplementary Figure 2)** and all patients showed a high similarity to the subpart of ID6. These results confirms that most of the HMF patients with a germline *BRCA2* variant and HRD have a high similarity to the HRD-associated signatures. **Supplementary Figure 3** displayed a heatmap of the comparison between the mutational profiles of HMF patients and various signatures from the COSMIC. The comparison showed that there were only associations between the mutational profiles of HMF patients and signatures classified as "clock-like." No similarities were found between HMF patient profiles and any other trait related signatures.

3.1.2. Evaluation of the performance of CHORD

The performance of CHORD was evaluated to identify a possible difference in prediction score between our own implemented variant callers as well as variant callers used for training the CHORD classifier. CHORD predicted a HRD score, based on input from four different combinations of variant callers (Mutect2-Manta, Mutect2-GRIDSS, Strelka2-Manta & Strelka2-GRIDSS). **Figure 6** presents an evaluation of the performance of CHORD using six patients randomly selected from the HMF cohort.

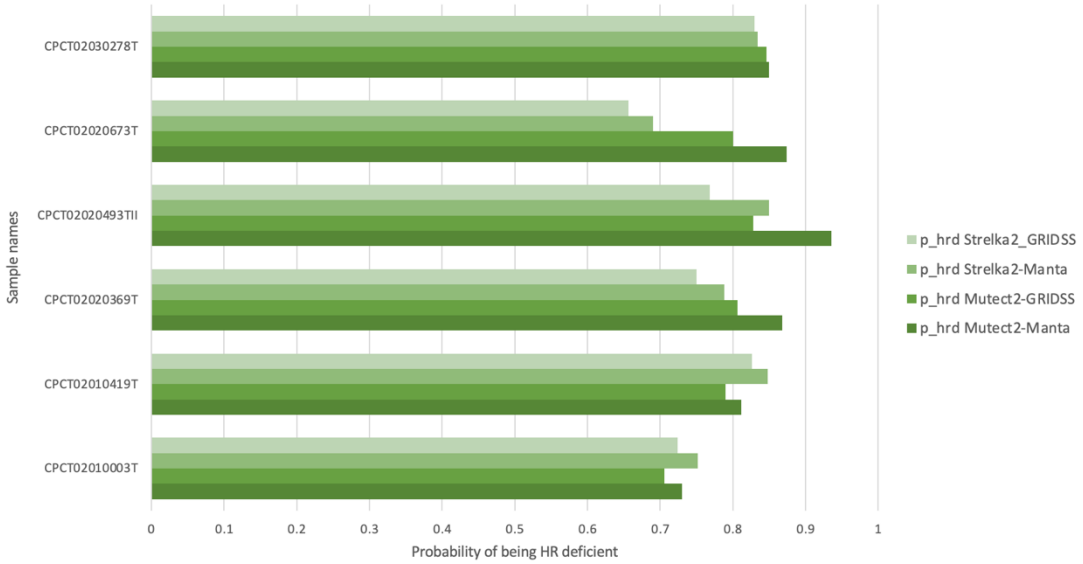


Figure 6. The evaluation of the performance of CHORD, using four variant calling combinations. The results suggest that the scores obtained from the four variant caller combinations are relatively similar for the six HMF patients.

The findings indicate that the scores generated by the four different combinations of variant callers are similar for the six patients selected from the HMF cohort, and therefore comparable. Accordingly, using the combination Mutect2 and Manta which is implemented in our own pipeline should be appropriate for predicting reliable HRD scores.

Additionally, the performance of CHORD was evaluated to identify a possible difference in predicting a HRD score for pediatric patients between the four different combinations of variant callers (**Supplementary Figure 4.A)**. CHORD generated higher HRD probability (p_hrd) scores when Mutect2 was used as variant caller for SNVs and indels compared to Strelka2**.** This difference mostly influenced the HRD prediction of patient P.20003. HRD was not detected in this patient when Strelka2 was used as SNV and indel caller (p_hrd = 0.004), whereas HRD was detected when Mutect2 was used (p_hrd = 0.730). In contrast, the type of SV caller did not have a substantial impact on the HRD results.

To further investigate the significant discrepancy in prediction between Mutect2 and Strelka2, a validation of the called variants was performed by comparing the similarity of identified SNVs and indels. A similarity of more than 80% was assessed in the detection of somatic SNVs, except for patient P.67780. However, less than 25% similarity was assessed for indel calling (**Supplementary Figure 4.B**). Thus, both tools had a similar performance in SNV calling whereas discrepancies in detected indels was assessed. Somatic indels from both variant callers were visualized in IGV to determine the accuracy of the identified indels. No clear distinction could be made between the percentage of true somatic indels called by either Mutect2 or Strelka2 (**Supplementary Figure 4.C**). Hence, a similar specificity was detected for both variant callers.

The evaluation results of the HMF cohort indicate comparable scores between the four combinations of variant callers for the six adult patients. A discrepancy in HRD prediction was observed between Mutect2 and Strelka2 in pediatric patients, but a similar accuracy was found for both variant callers. Therefore, the use of Mutect2 as SNV and indel caller and Manta as SV caller in our pipeline is appropriate for reliable HRD score prediction for pediatric patients.

3.1.3. HRD probability prediction

Probability scores of being HR deficient were calculated with the CHORD using the somatic variant data generated by Mutect2 and Manta. HRD scores of the pediatric patients are shown in **Figure 7**.

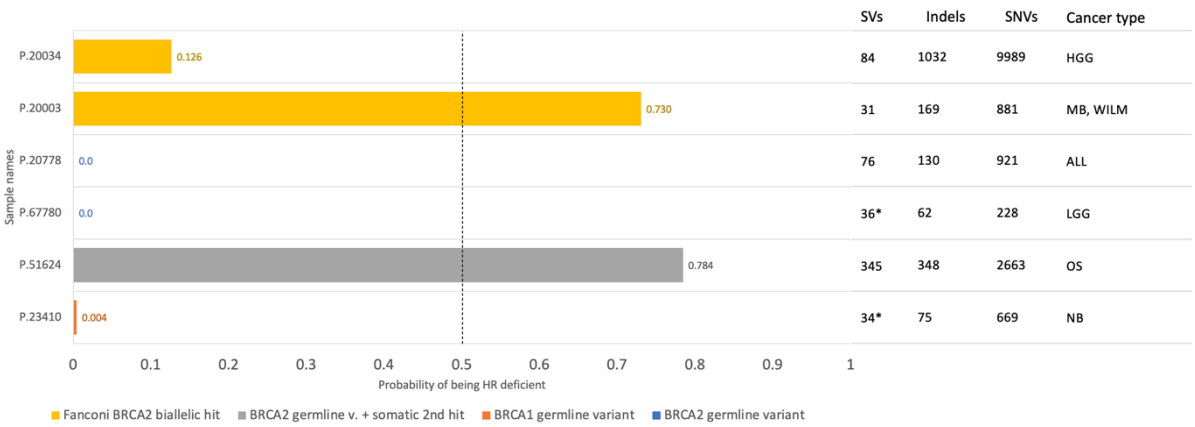


Figure 7. Probability score of being HR deficient generated with Mutect2 and Manta represented together with number of filtered variants and cancer type per patient. The probability scores of being HR deficient are plotted alongside the number of filtered variants and the type of cancer for each patient. Patients P.20003 and P.51624 have a high probability score for having HRD above the threshold of 0.5. The other four samples have a probability score below the threshold. The filtering criteria include a minimum variant allele frequency of 0.25 for SNVs and 0.2 for indels, and a minimum of 5 alternative paired & split reads for structural variants (SVs). However, to meet the minimum input requirements of the CHORD, the filtering criteria for SVs were adjusted to 4 alternative paired & split reads for sample P.67780, and 3 for sample P.23410.

One Fanconi patient with a *BRCA2* biallelic hit (P.20003) and one patient with a *BRCA2* germline variant and a somatic second hit (P.51624), displayed a significant HRD score above the threshold (≥0.5). Therefore, these two patients likely have a defective HRR pathway. The other four patients did not display HRD scores above the threshold, indicating no deficiency in the HRR pathway. Three out of four patients, all with a heterozygous *BRCA1* or *BRCA2* germline variant, displayed scores close to zero. A score of 0.126 was detected for a Fanconi patient with a compound heterozygous BRCA2 germline variant. It is notable that the two patients with positive HRD score had a relatively higher proportion of SNVs, indels, and SVs (see Table in **Figure 7**) and a higher tumor mutational burden (TMB) (**Supplementary Table 1**) compared to the other four patients. In addition, many copy numbers variations were found in patient P.51624, suggesting a possible contribution towards the identification of HRD and genomic instability (**Supplementary Figure 5**). Furthermore, a probability score was calculated by CHORD to distinguish *BRCA1*-type from *BRCA2*-type HRD. **Supplementary Table 1** shows A *BRCA2*-type HRD was correctly predicted for all patients with a *BRCA2* variant for which a positive HRD score was calculated. Hence, a positive HRD score was observed for two out of three patients with a homozygous variant, as expected for these two. Patients with a heterozygous *BRCA1* or *BRCA2* germline variant did not show indications of HRD, and therefore no second genetic event.

In addition, CHORD was used to determine the likelihood of a deficiency in the HRR pathway by analyzing somatic variant data of the HMF adult patients obtained from Mutect2 and Manta. A HRD-score higher than 0.5 was calculated for all 19 adult patients, indicating a deficient HR-status. Moreover, all patients with a *BRCA2*-type HRD score ≥0.5 was calculated for all patients, confirming HRD resulting from the *BRCA2* variant, depicted in **Supplementary Figure 6**.

3.2. eQTL analysis

3.2.1 Cis and trans eQTLs

eQTLs were selected that are related to a negative expression of the *BRCA2* gene. 65 cis-eQTLs were identified that are significantly associated with a reduced expression of the *BRCA2* gene. Characteristics of these cis-eQTLs can be found in **Supplementary Table 3**. However, no trans-eQTLs significantly reduced the expression of the *BRCA2* gene (**Supplementary Table 4**). Therefore, trans-eQTLs were excluded from the analysis.

3.2.2. eQTL comparison

The selected cis-eQTLs were compared to the filtered germline variants of all patients by evaluating the genomic coordinates (location, reference allele, and alternative allele) of the eQTLs and germline variants. In most pediatric patients and HMF adult patients, cis-eQTLs were found on the germline, displayed in **Figure 8**.

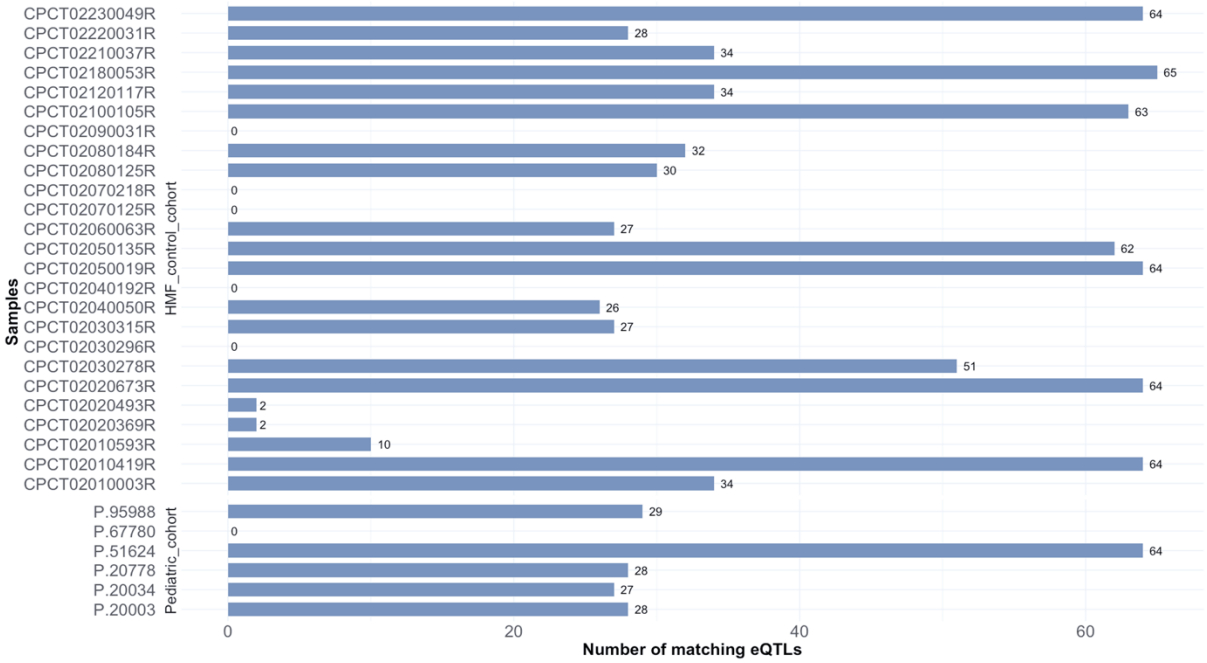


Figure 8. The number of matching eQTLs per patient for the pediatric cohort and HMF adult control cohort. Filtered variants found in the germline data of patients were compared to the 65 cis-eQTL variants that negatively influenced gene expression of *BRCA2*. For most patients, a number of eQTLs was found in the germline data.

64 out of 65 eQTLs that are associated with lower expression of the *BRCA2* gene were present in the germline data of patient P.51624. This patient showed a relatively high copy number variations in the germline compared to the other pediatric patients (**Supplementary Figure 5**). In contrast, no eQTLs were identified in one of the pediatric patients, P.67780. eQTL comparison was also conducted for the 25 adult patients of the HMF cohort with a *BRCA2* germline variant. eQTLs were identified in 20 adult patients, indicating that the occurrence of eQTLs that are linked to a reduced expression of *BRCA2* in the germline is not uncommon in patients with a germline *BRCA2* variant. Additionally, the frequency of these eQTLs in the healthy population was determined based on population frequencies present in the gnomAD database. This analysis

demonstrated that the occurrence of eQTLs is prevalent within the germline data of the general population, although with lower frequencies (as illustrated in **Figure 9**).

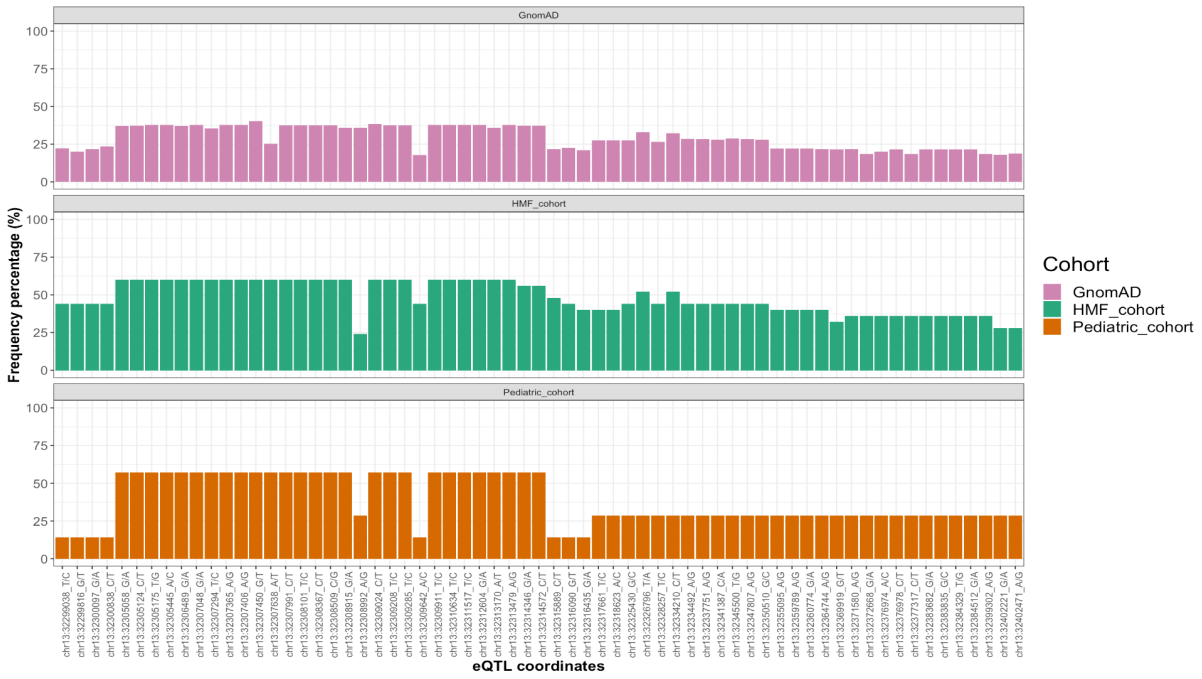


Figure 9. Population frequencies of eQTL coordinates between three different cohorts. The gnomAD cohort represents the presence of eQTLs related to negative *BRCA2* expression in the healthy population. Similar distributions were detected in the HMF cohort and the pediatric cohort, where a small increase of eQTL presence towards the healthy population is observed between positions 32305058-32314572 on chromosome 13.

SNPs located near and on the *BRCA2* gene have a population frequency between 12-60% in adult and pediatric patients with a heterozygous *BRCA2* germline variant. Frequencies of eQTLs in the healthy population are between 20-40% (Karczewski *et al.*, 2020). In general, a higher occurrence of eQTLs is observed between region 32305058-32314572 on chromosome 13. Patients with a germline *BRCA2* variant show a higher percentage of eQTLs compared to the healthy population. However, due to the similarities in the distribution of cis-eQTLs between pediatric patients and HMF adult patients with a *BRCA2* germline variant, it was not possible to determine a potential contributing factor of eQTLs in pediatric patients.

Seven clusters of co-occurring eQTLs were discovered after analysing the eQTL distribution throughout the genome (**Supplementary Figure 7**). A higher number of variants were identified with a lower Z-score near the start region of the *BRCA2* gene, suggesting that these eQTLs may have a greater effect on negative expression of the gene. It is necessary to determine the effect of these co-occurring clusters on expression of the *BRCA2* gene in order to interpret the results of haploinsufficiency.

Results were validated by analysing the presence of eQTLs in a negative control cohort of 116 AYAs patients with bladder cancer. The analysis revealed that 87 of these patients contained cis-eQTLs that were correlated with negative expression of the *BRCA2* gene. A similar population frequency distribution was found when compared to patients with a *BRCA2* germline variant (**Supplementary Figure 8**). Interestingly, the same seven clusters of co-occurring eQTLs were found in bladder cancer patients as well.

Hierarchical clustering was performed based on the prevalence of eQTLs in each sample to check the association of a *BRCA2* germline variant and the co-occurrence of eQTLs. The clustering indicated a lack of discernible grouping of eQTLs among the three distinct populations, suggesting an absence of eQTL patterns that are predominant within a particular cohort. (**Supplementary Figure 9**).

4. Discussion

This study focused on pediatric patients with a heterozygous *BRCA1* and *BRCA2* germline variant and the identification of a potential second genetic event that may have contributed to the onset of pediatric cancer. To investigate the presence of a potential second genetic event, two approaches were examined. First, somatic mutational patterns were analysed to identify a possible HRD, which could reveal the presence of a missed second genetic event. The study further investigated the potential relation between eQTLs in the germline of the patients and allelic imbalance contributing to the onset of pediatric cancer. It is important to note that the sample size was limited in this study, meaning that further research is required to fully understand the findings of this study.

4.1. HRD findings and limitations

The HRD prediction analysis aimed to investigate whether pediatric cancer patients with a heterozygous *BRCA1* or *BRCA2* germline variant had a deficiency in their HRR pathway through evincing HRD-related signatures and by calculating a HRD score. The HMF adult control samples together with current research shows that biallelic inactivation in the *BRCA1* and *BRCA2* genes has a relatively high correlation to HRD and similarity to signatures SBS3 and ID6 (Nguyen *et al.*, 2020). The cosine similarity was used as the measure of similarity between the HRD-related signatures and the somatic mutational profiles of the patients. Four out of six pediatric patients showed high similarity to signature SBS3 due to a cosine similarity higher than 0.8. The similarity score to SBS3 was adjusted compared to the commonly used threshold value of 0.85 (Blokzijl *et al.*, 2018; Levatić *et al.*, 2022), due to a featureless profile of SBS3. Three out of six patients, all with a biallelic BRCA2 germline variant or an identified second hit, had resemblance with the subset 5+ bp deletions with a microhomology of ID6, due to a cosine similarity score equal or greater than 0.85. No similarity to a HRD-related signature was found in patients without a biallelic variant or identified second hit. These results show the absence of HRD-related signatures in pediatric patients without a second genetic event. In addition, no HRD score above the threshold was predicted with CHORD in the pediatric patients with a heterozygous *BRCA1* or *BRCA2* germline variant, assuming no HRR pathway defect. A likely HRR pathway defect was found in patient P.51624 with a germline *BRCA2* variant and somatic 2nd hit and in patient P.20003 with a biallelic *BRCA2* variant. Although these results confirm the hypothesis that HRD-related signatures and HRD were expected in the patients with a second hit or biallelic germline variant, no positive HRD score was found in patient P.20034. Two suggestions could be made: a prediction score for HRD below 0.5 may suggest that the impact of HRD may be less pronounced in children and that the threshold for HRD prediction should be re-evaluated for pediatric patients. However, it is important to note that the model was trained on adult patient and according to existing literature, a HRD prediction score below the threshold should always be considered as HR proficient, based on significance of the trained model (Nguyen et al., 2020). On the other hand, children tend to have lower TMBs compared to adults which often do not meet CHORD's TMB requirements (Rahal *et al.*, 2018). Depending on tumor type, the number of SVs often do not meet the minimum requirement of 30 to generate a HRD prediction score. Hence, less strict selection criteria were used for the filtering of SVs. This approach may have led to a higher number of false positive SVs in the tumor data and ultimately, a less reliable HRD score.
Furthermore, evaluating CHORD with different variant callers produced comparable HRD scores for HMF adult controls, suggesting the in-house pipeline with Mutect2 and Manta seems to predict HRD accurately. However, the combinations with Mutect2 produced higher HRD probability

scores compared to combinations with Strelka2 in pediatric patients. This result, along with the fact that the indels identified by Mutect2 and Strelka had a very low similarity, suggest that there may be a contributing factor in pediatric patients that affects HRD prediction scores.

4.2. eQTL findings and limitations

In order to deduce if other genetic changes may affect the expression of the wildtype allele, cis-eQTLs affecting the expression of the *BRCA2* gene were compared to all filtered variants found in the germline of pediatric patients and the HMF adult control cohort. Germline variants were found in the patients on the same location as cis-eQTLs were identified. Comparing the frequencies of these germline variants to the healthy population showed a slightly higher percentage of a subpart of cis-eQTLs found around the *BRCA2* gene in patients with a *BRCA2* germline variant. Further analysis was done to distinguish the different patterns and numbers of cis-eQTLs found in the patient's germline. Ultimately, seven clusters of cis-eQTLs that could negatively influence gene expression were identified across the genome. These clusters contained varying numbers of cis-eQTLs, which were identified as the variants typically co-occurred within the same cluster. Despite the efforts to understand the germline cis-eQTL information contained in the data, no determination was done about which cis-eQTLs are present on the wildtype (non-pathogenic) allele. This means that more analysis is needed to draw a conclusion about whether cis-eQTLs on the wildtype (non-pathogenic) allele together with the pathogenic *BRCA2* germline variant could lead to haploinsufficiency, and therefore allelic imbalance.

Additionally, bladder cancer AYAs patients were used as a negative control for cis-eQTL analysis. Despite that these patients did not have pathogenic variants in the *BRCA2* gene, a slightly higher percentage of cis-eQTLs compared to the healthy population was found, with almost the same distribution as found in the pediatric and HMF adult patients with a BRCA2 germline variant (**Supplementary Figure 8**). In addition, the bladder cancer patients showed the same seven clusters of co-occuring cis-eQTLs. While the presence of cis-eQTLs is common in the population, the validation of these seven specific clusters could suggest a rare cooperation of specific combinations of clusters with a clear effect on reduced expression of the *BRCA2* gene. However, further analysis and expansion of the cohort is required to validate whether combinations of clusters could affect reduced gene expression.

4.3. Conclusion

The objective of this study was to investigate if heterozygous *BRCA1* or *BRCA2* germline variants in pediatric patients together with a second genetic event could have initiated tumor development. An association with HRD was only observed in pediatric patients with a biallelic *BRCA2* germline variant or heterozygous germline variant and a somatic second hit. Accordingly, HRD was not found in patients with a heterozygous *BRCA1* or *BRCA2* germline variant, and therefore no relation to the onset of pediatric cancer was found. In addition, a slightly higher percentage of cis-eQTLs within patients with a *BRCA2* germline variant was found compared to the healthy population, but not compared to the AYAs patients with bladder cancer. Seven clusters of co-occurring cis-eQTLs were found. Accordingly, splitting the alleles to identify clusters of cis-eQTLs on the non-pathogenic allele together with RNA expression analysis are recommended to identify the potential effect of cis-eQTL clusters on gene expression. Due to sample size limitations, further research is necessary to fully understand our findings.

4.4. Future perspectives and recommendations

It is known that the genetic cause of HRD varies depending on tumor type (Nguyen *et al.*, 2020). Research has shown that an association with a germline variant in the *BRCA2* gene and cancer is commonly found in pediatric patients who have a brain or solid tumor (Kratz *et al.*, 2022). To increase the likelihood of identifying patients with this specific association, it is recommended to focus recruitment efforts on pediatric patients with brain or solid tumors who have a heterozygous *BRCA2* germline variant. By specifically targeting this patient population, it will probably be easier to identify and better understand of the effects of HRD within pediatric cancer patients. Moreover, due to sample size limitation, three distinct groups were identified in this study: patients with biallelic germline variants associated with Fanconi Anemia, patients with a germline variant and a somatic alteration (e.g. LOH), and patients with only a heterozygous germline variant without a somatic alteration. However, it has already been established that the first two groups are associated with HRD, as supported by prior research (Woodward and Meyer, 2021). Therefore, the focus should shift to study the effect of heterozygous *BRCA1* or *BRCA2* germline variants in pediatric patients. The group with biallelic PVs or a second hit could be used as a positive control cohort.

As certain clusters of cis-eQTLs were identified as potential affecters, a next step in the cis-eQTL analysis would be allele phasing. Through the process of phasing, the paternally and maternally inherited copies of the patient's germline data can be distinguished, allowing for the determination of the phase of each cis-eQTL and the haplotype blocks to which they belong. Therefore, a full understanding of the cis-eQTLs per allele can be gained. This allows for further analysis to determine which clusters of the cis-eQTLs are present on the wildtype allele. WhatsHap is a tool that can be used to perform read-based phasing (Martin *et al.*, 2016). Additionally, it would be interesting to analyse RNAseq expression levels. By analysing the gene expression profiles, it is possible to identify clusters of co-occurring cis-eQTLs with a significant impact on reduced *BRCA1* and *BRCA2* gene expression. This impact is particularly noteworthy if the effect is found to be stronger in the pediatric cohort compared to control cohorts and healthy population. This can provide insights into the functional consequences of innocent germline variants and how it may contribute to cancer development in pediatric patients with a heterozygous *BRCA1* or *BRCA2* germline variant.

This study did not analyse the effect of cis-eQTLs on the *BRCA1* gene, due to sample size limitation of patients with a heterozygous *BRCA1* germline variant. Comparing germline variants with cis- or trans-eQTLs that could affect the expression of *BRCA1* would be interesting for future research. In addition, eQTLs on the wildtype allele can also result in overexpression of the *BRCA1* or *BRCA2* gene, what could possibly lead to allelic imbalance (Satyananda *et al.*, 2021). Therefore, it can be important to study the effect of eQTLs on the wildtype allele, resulting in gene overexpression, in future research.

This study focused on the *BRCA1* and *BRCA2* genes, which are known to be associated with the process of HRR. However, multiple other genes are involved in HRR and can lead to deficiencies in this process. Two examples of these genes are *PALB2* and *RAD51C*, which are among the most common genetic causes of HRR deficiencies. These genes, along with recombination mediators or co-mediators, would be interesting to consider in future studies for predicting HRD.

Lastly, the HRD status could provide predictive information on the expected degree of benefit from a class of drugs called PARP-inhibitors. Recent studies have shown sensitive responses to PARP-inhibitors therapy in adult patients with *BRCA1* or *BRCA2* germline variants and ovarian cancer (Zhang, Yuan and Hao, 2014). If it were to be determined that these types of variants are also a cause of HRD and the development of pediatric cancers, adjustments and recommendations to pediatric treatment plans could be made.

References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., *et al.* (2013a) 'Deciphering Signatures of Mutational Processes Operative in Human Cancer', *Cell Reports*, 3(1), pp. 246–259.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., *et al.* (2013b) 'Signatures of mutational processes in human cancer', *Nature*, 500(7463), pp. 415–421.

Auwera, G.V. der and O'Connor, B. (2020) *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly Media.

Benjamin, D. *et al.* (2019) 'Calling Somatic SNVs and Indels with Mutect2'. bioRxiv, p. 861054.

Blokzijl, F. *et al.* (2018) 'MutationalPatterns: comprehensive genome-wide analysis of mutational processes', *Genome Medicine*, 10(1), p. 33.

Boomsma, D.I. *et al.* (2014) 'The Genome of the Netherlands: design, and project goals', *European journal of human genetics: EJHG*, 22(2), pp. 221–227.

Cameron, D.L. *et al.* (2017) 'GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly', *Genome Research*, 27(12), pp. 2050–2060.

Chen, X. *et al.* (2016) 'Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications', *Bioinformatics*, 32(8), pp. 1220–1222.

Creeden, J.F. *et al.* (2021) 'Homologous recombination proficiency in ovarian and breast cancer patients', *BMC Cancer*, 21(1), p. 1154.

Decker, B. *et al.* (2016) 'Biallelic BRCA2 Mutations Shape the Somatic Mutational Landscape of Aggressive Prostate Tumors', *American Journal of Human Genetics*, 98(5), pp. 818–829.

Karaayvaz-Yildirim, M. *et al.* (2020) 'Aneuploidy and a deregulated DNA damage response suggest haploinsufficiency in breast tissues of BRCA2 mutation carriers', *Science Advances*, 6(5), p. eaay2611.

Karczewski, K.J. *et al.* (2020) 'The mutational constraint spectrum quantified from variation in 141,456 humans', *Nature*, 581(7809), pp. 434–443.

Kolde, R. (2018) 'pheatmap'. Available at: https://cran.microsoft.com/snapshot/2018-06-22/web/packages/pheatmap/pheatmap.pdf.

Kratz, C.P. *et al.* (2021) 'Predisposition to cancer in children and adolescents', *The Lancet Child & Adolescent Health*, 5(2), pp. 142–154.

Kratz, C.P. *et al.* (2022) 'Heterozygous BRCA1 and BRCA2 and Mismatch Repair Gene Pathogenic Variants in Children and Adolescents With Cancer', *Journal of the National Cancer Institute*, 114(11), pp. 1523–1532.

Lamb, J.C. and Birchler, J.A. (2003) 'The role of DNA sequence in centromere formation', *Genome Biology*, 4(5), p. 214.

Levatić, J. *et al.* (2022) 'Mutational signatures are markers of drug sensitivity of cancer cells', *Nature Communications*, 13, p. 2926.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760.

Li, Q. *et al.* (2013) 'Integrative eQTL-Based Analyses Reveal the Biology of Breast Cancer Risk Loci', *Cell*, 152(3), pp. 633–641.

Martin, M. *et al.* (2016) 'WhatsHap: fast and accurate read-based phasing'. bioRxiv, p. 085050.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, 20(9), pp. 1297–1303.

McLaren, W. *et al.* (2016) 'The Ensembl Variant Effect Predictor', *Genome Biology*, 17(1), p. 122.

Narzisi, G. *et al.* (2018) 'Genome-wide somatic variant calling using localized colored de Bruijn graphs', *Communications Biology*, 1, p. 20.

Nguyen, L. *et al.* (2020) 'Pan-cancer landscape of homologous recombination deficiency', *Nature Communications*, 11(1), p. 5584.

Nielsen, R. *et al.* (2011) 'Genotype and SNP calling from next-generation sequencing data', *Nature reviews. Genetics*, 12(6), pp. 443–451.

Petrucelli, N., Daly, M.B. and Pal, T. (1993) 'BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer', in M.P. Adam et al. (eds) *GeneReviews*®. Seattle (WA): University of Washington, Seattle.

Picard Tools - By Broad Institute (no date). Available at: http://broadinstitute.github.io/picard/

Poplin, R. *et al.* (2018) 'Scaling accurate genetic variant discovery to tens of thousands of samples'. bioRxiv, p. 201178.

Rahal, Z. *et al.* (2018) 'Genomics of adult and pediatric solid tumors', *American Journal of Cancer Research*, 8(8), pp. 1356–1386.

Ripperger, T. *et al.* (2017) 'Childhood cancer predisposition syndromes—A concise review and recommendations by the Cancer Predisposition Working Group of the Society for Pediatric Oncology and Hematology', *American Journal of Medical Genetics Part A*, 173(4), pp. 1017–1037.

Robinson, J.T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26.

Satyananda, V. *et al.* (2021) 'High BRCA2 Gene Expression is Associated with Aggressive and Highly Proliferative Breast Cancer', *Annals of Surgical Oncology*, 28(12), pp. 7356–7365.

Saunders, C.T. *et al.* (2012) 'Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs', *Bioinformatics*, 28(14), pp. 1811–1817.

Schild, D. and Wiese, C. (2010) 'Overexpression of RAD51 suppresses recombination defects: a possible mechanism to reverse genomic instability', *Nucleic Acids Research*, 38(4), pp. 1061–1070.

Štancl, P. *et al.* (2022) 'The Great Majority of Homologous Recombination Repair-Deficient Tumors Are Accounted for by Established Causes', *Frontiers in Genetics*, 13.

Tate, J.G. *et al.* (2019) 'COSMIC: the Catalogue Of Somatic Mutations In Cancer', *Nucleic Acids Research*, 47(D1), pp. D941–D947.

Võsa, U. *et al.* (2018) 'Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis'. bioRxiv, p. 447367.

Võsa, U. *et al.* (2021) 'Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression', *Nature Genetics*, 53(9), pp. 1300–1310.

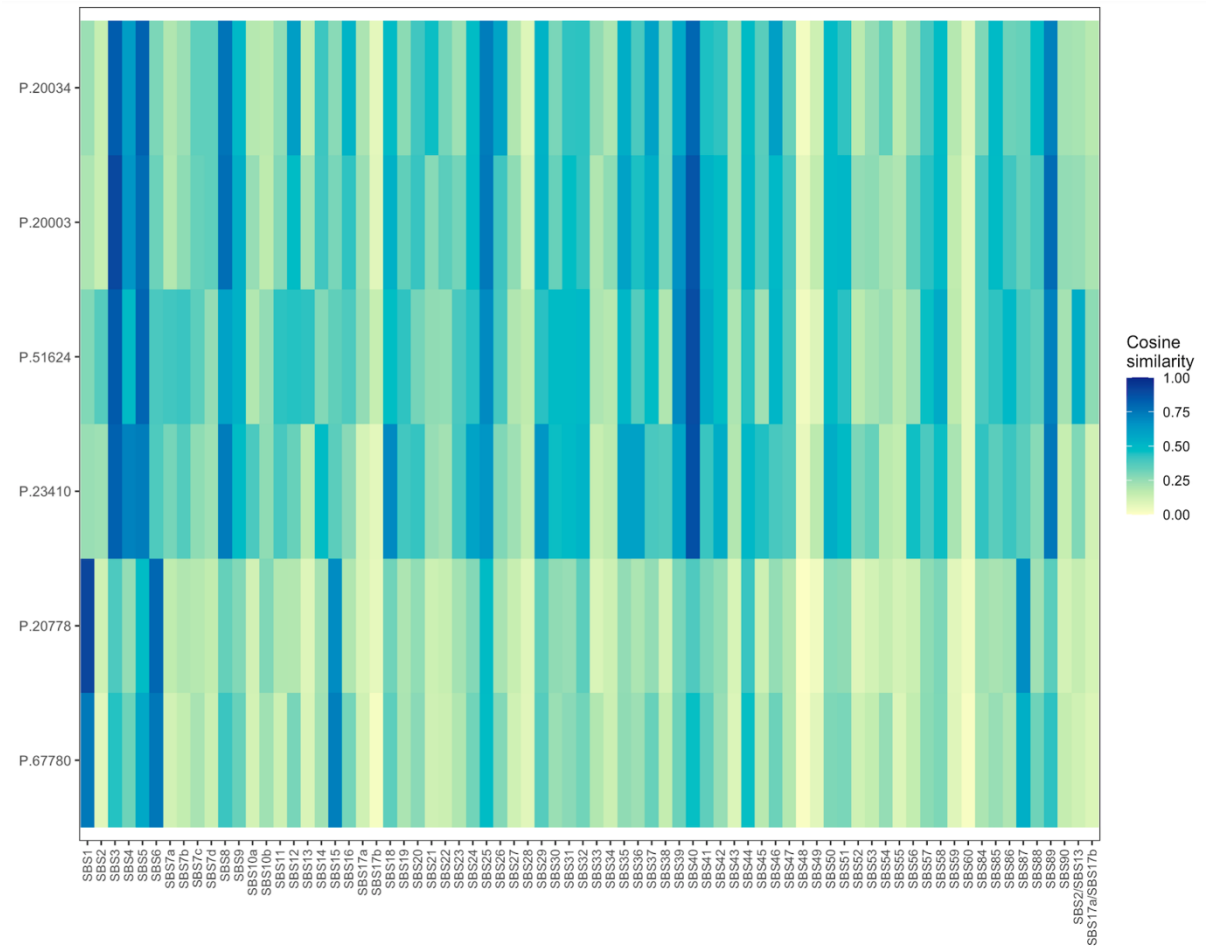
Woodward, E.R. and Meyer, S. (2021) 'Fanconi Anaemia, Childhood Cancer and the BRCA Genes', *Genes*, 12(10), p. 1520.

Zhang, J. *et al.* (2015) 'Germline Mutations in Predisposition Genes in Pediatric Cancer', *New England Journal of Medicine*, 373(24), pp. 2336–2346.

Zhang, S., Yuan, Y. and Hao, D. (2014) 'A Genomic Instability Score in Discriminating Nonequivalent Outcomes of BRCA1/2 Mutations and in Predicting Outcomes of Ovarian Cancer Treated with Platinum-Based Chemotherapy', *PLoS ONE*, 9(12), p. e113169.

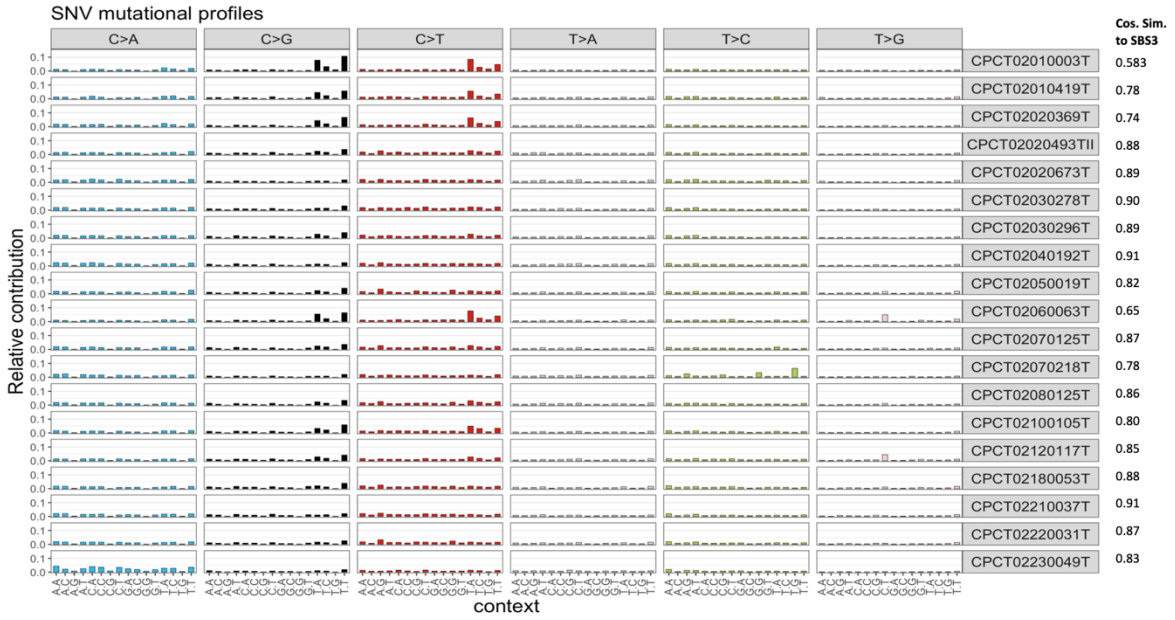
Supplementary Material

Supplementary Figures

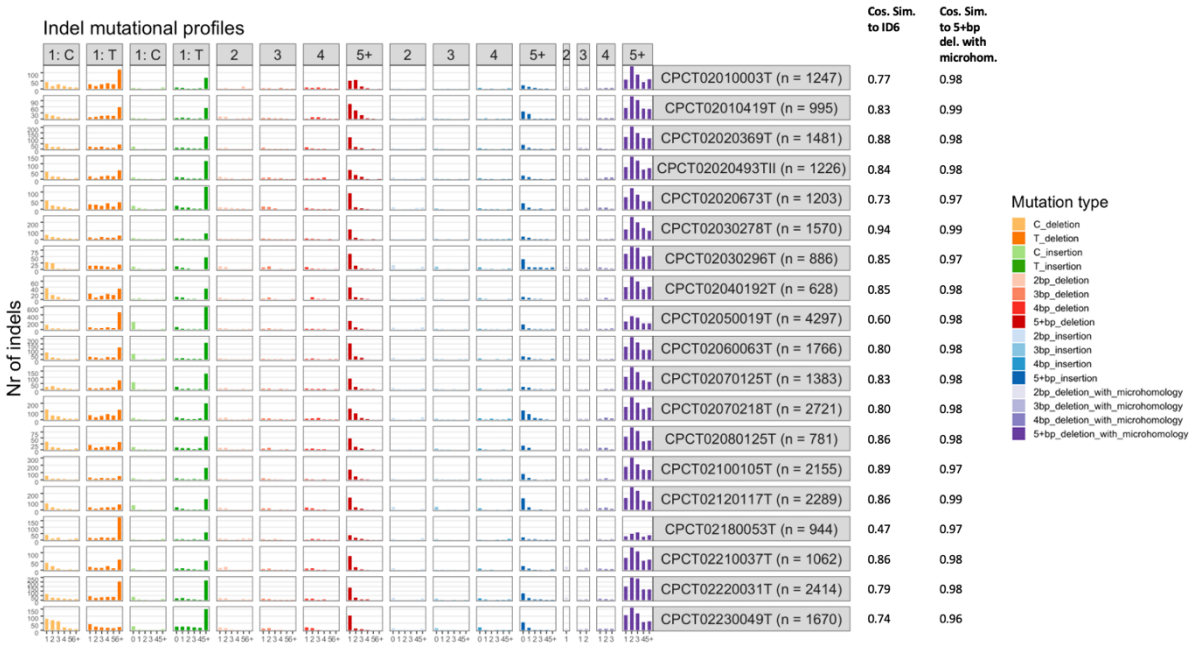


Supplementary Figure 1. Heatmap of cosine similarities between the signatures from COSMIC and the somatic mutational profiles of pediatric patients with *BRCA1* or *BRCA2* germline variant(s). 3 patients show a high similarity to the HRD-related signature SBS3 from COSMIC (≥0.8).

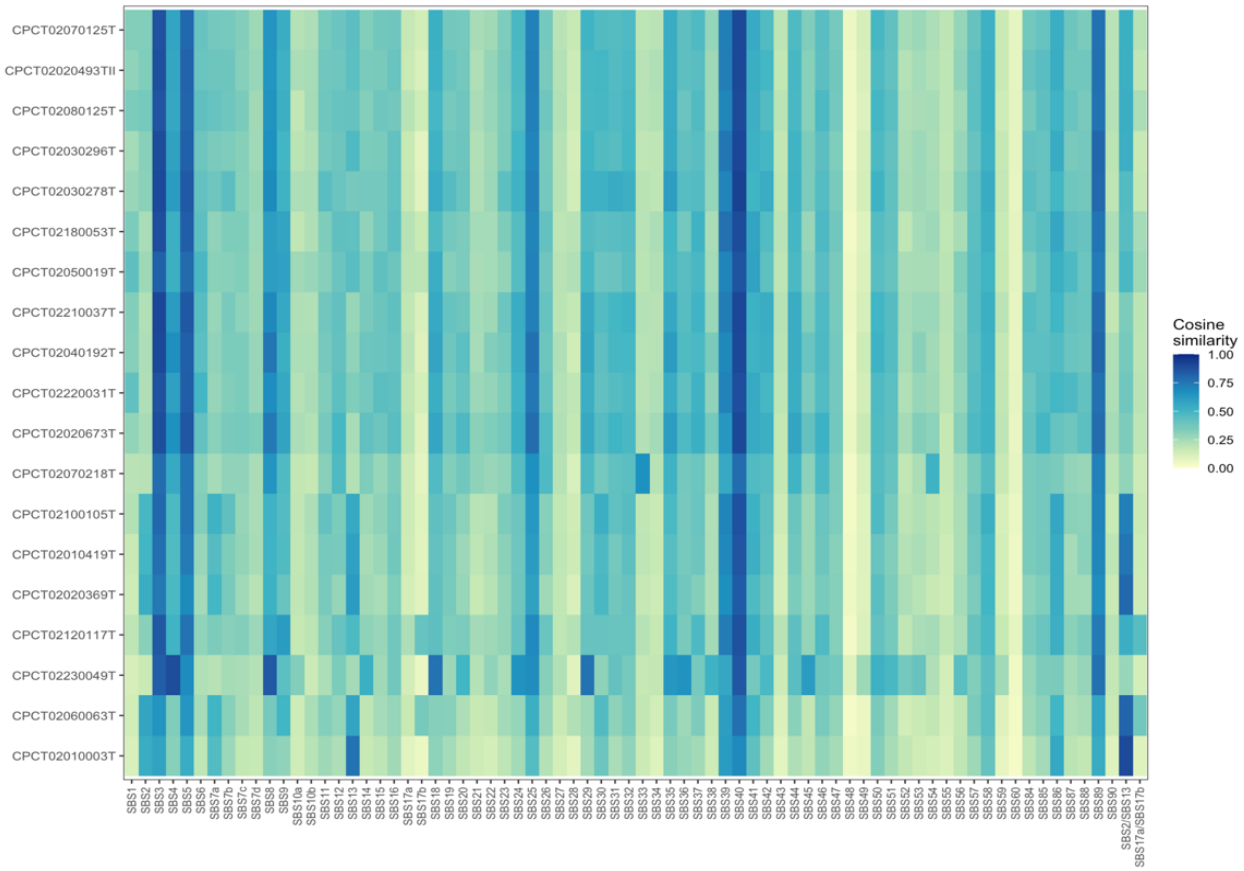
A



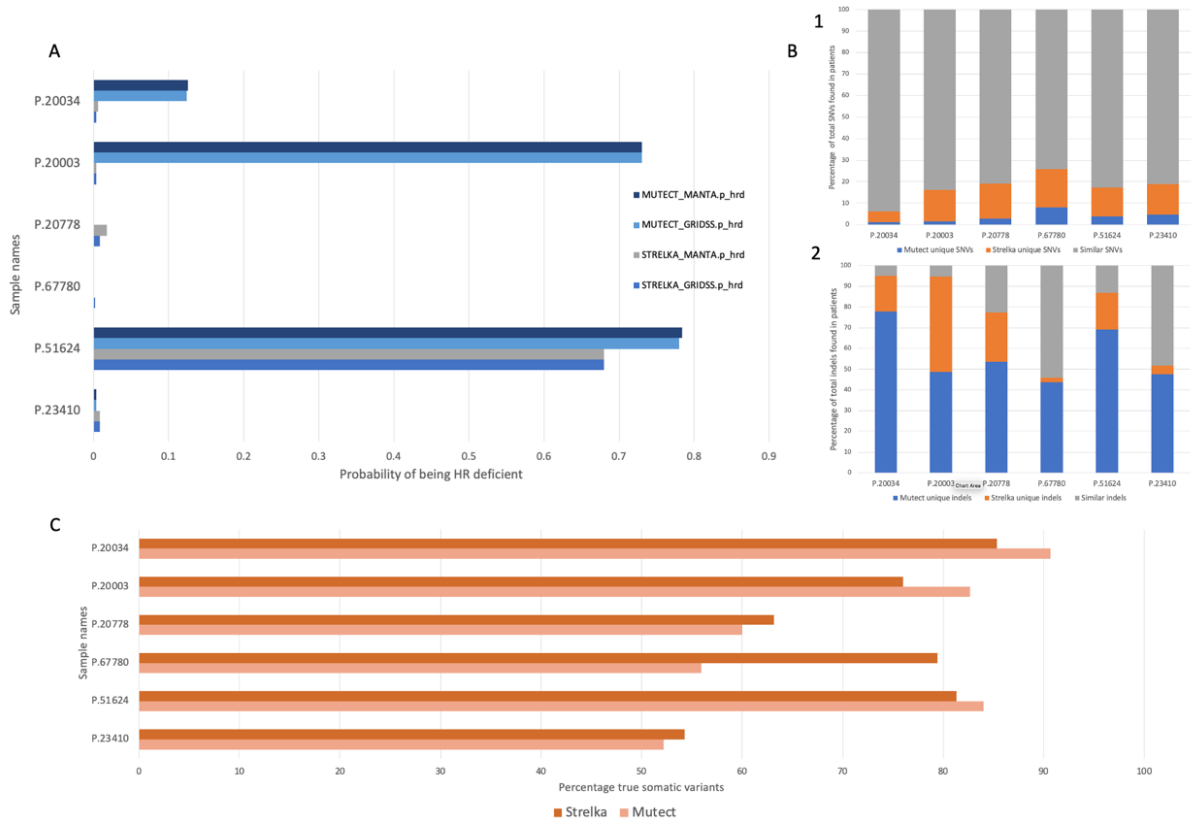
B



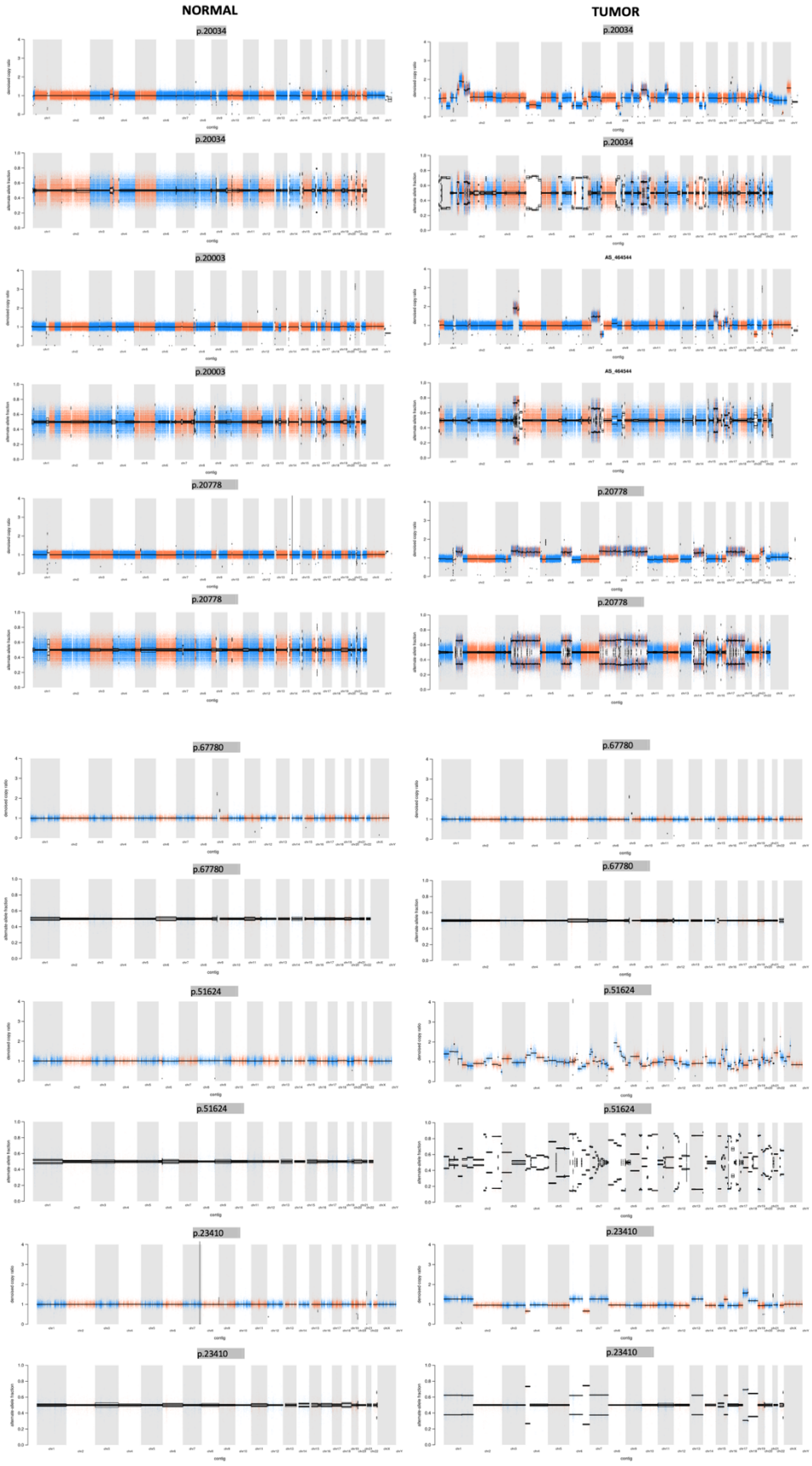
Supplementary Figure 2. Mutational profiles of the HMF patients. A. The SNV mutational profiles of the HMF patients. 14 patients show a high cosine similarity score (≥0.8), indicating a high similarity to HRD-related signature SBS3 from COSMIC. **B.** The indel mutational profiles of the HMF patients. 8 patients show a high cosine similarity to ID6 from COSMIC (≥0.85). All HMF adult patient show an extremely high cosine similarity with the 5+bp deletion with microhomology subpart of ID6.



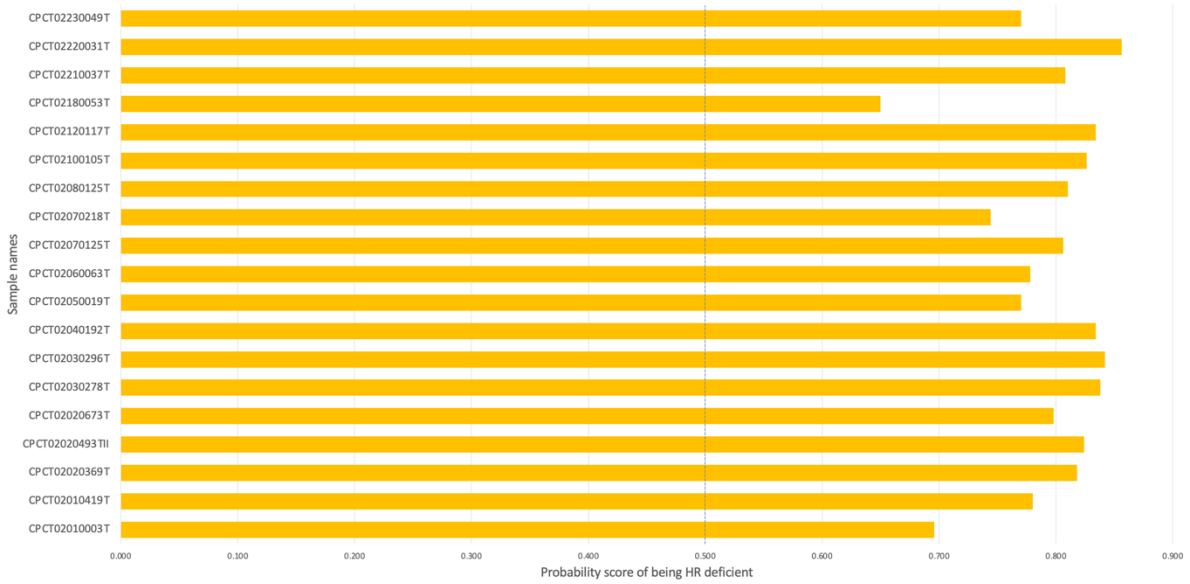
Supplementary Figure 3. Heatmap of cosine similarities between the signatures from COSMIC and the somatic mutational profiles of the HMF adult control patients with a *BRCA2* germline variant. Fourteen patients show a high similarity to the HRD-related signature SBS3 from COSMIC (≥0.8).



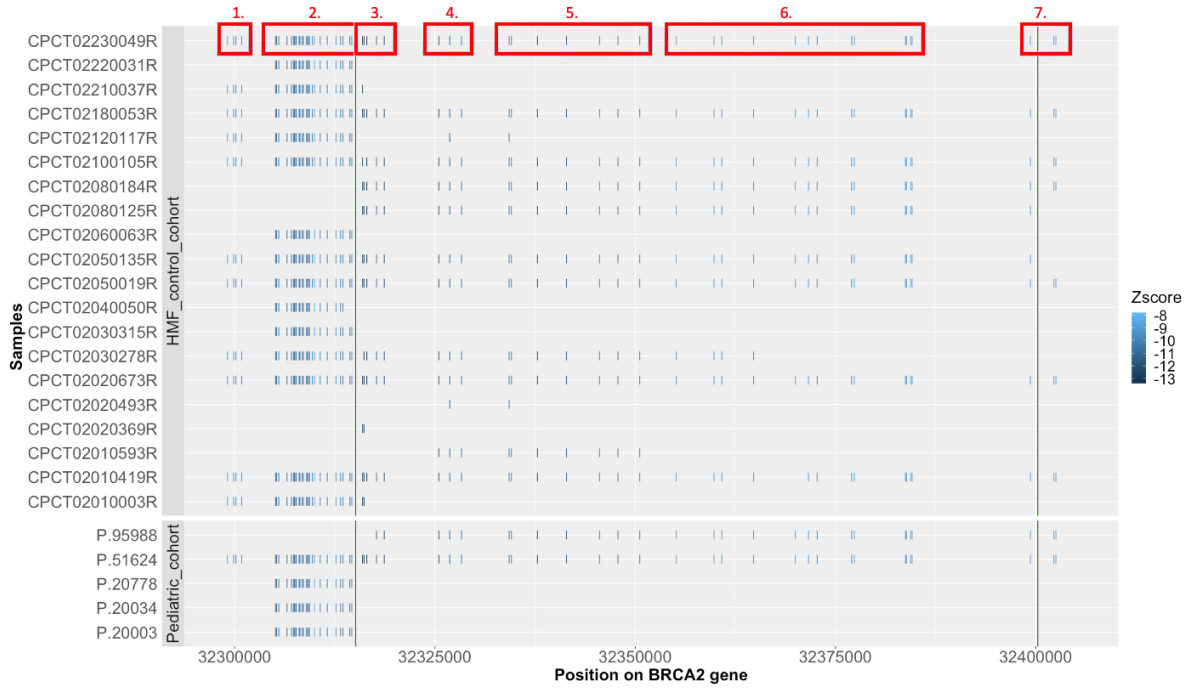
Supplementary Figure 4. Classier validation shows a significant contrast in the HRD prediction score between different variant caller combinations. **A.** Probability score of being HR deficient (p_hrd) between different combinations of variants callers shows a clear difference in prediction for sample P.20003 & P.51624. **B.1** A high degree of similarity between SNVs called by Mutect2 and Strelka2 is represented, with over 80% of all SNVs being identical across both variant callers after filtering. **B.2** The similarities and differences of indels called with Mutect2 and Strelka2 are shown. There is relatively little overlap (≤25%) in the indels identified by Mutect2 and Strelka2 **C.** An analysis of the true somatic variants identified by Mutect2 and Strelka2 through the use of the Integrative Genomics Viewer (IGV) revealed no significant difference in the reliability of performance of the two callers.



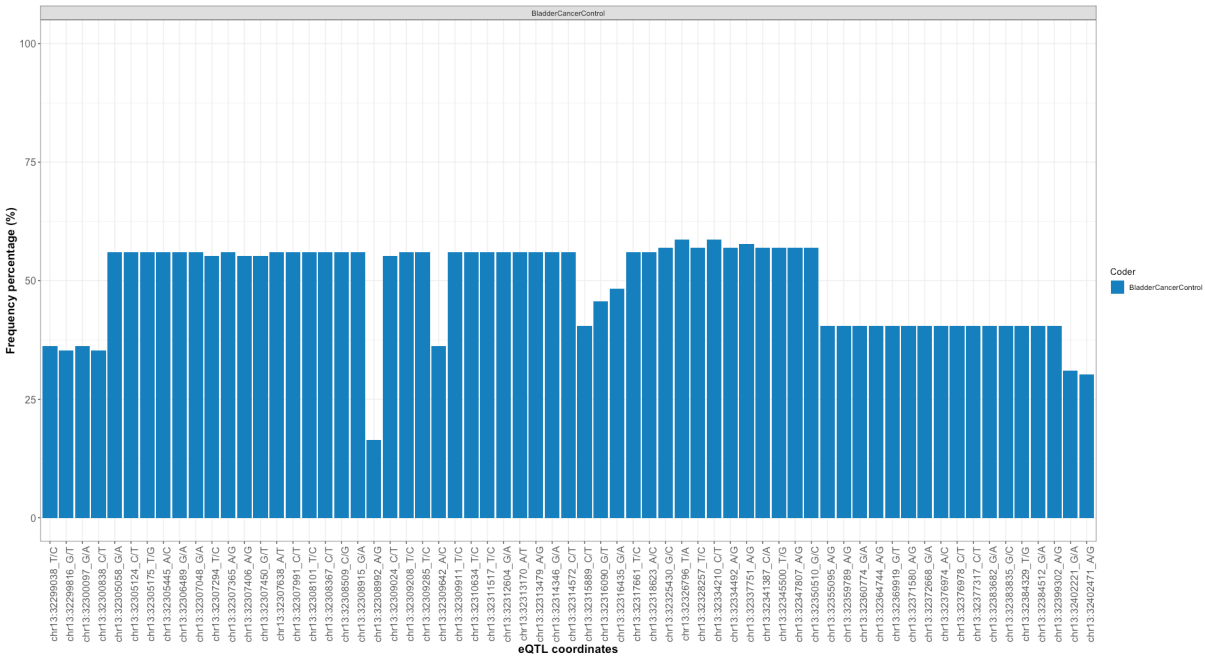
Supplementary Figure 5. CNV plots for all pediatric patient with a *BRCA1* or *BRCA2* germline variant. Patient P.51624 shows a high number of CNVs. No CNV analysis was done for patient P.95988.



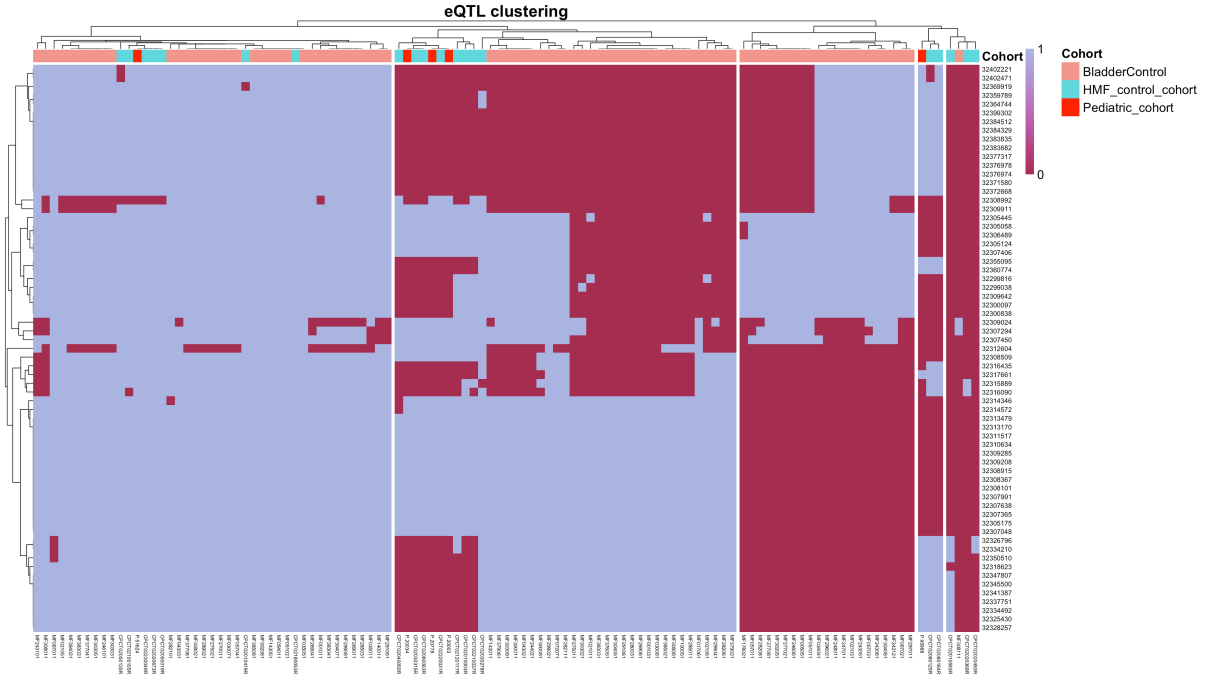
Supplementary Figure 6. Probability score of being HR deficient for the HMF patients generated with Mutect2 and Manta. All adult patients with a germline *BRCA2* variant show a HRD score above the threshold of 0.5, indicating a deficiency in the HRR pathway.



Supplementary Figure 7. Seven clusters of co-occurring eQTLs were discovered after analysing the eQTL distribution throughout the genome. The locations of the start and end of the *BRCA2* gene were highlighted in green. A Z-score was assigned to all eQTLs to measure the reliability of their impact on gene expression levels.



Supplementary Figure 8. Population frequencies of eQTL coordinates of the AYAs bladder cancer patient cohort. Similar distributions were detected in the HMF cohort and the pediatric cohort.



Supplementary Figure 9. Hierarchical clustering was performed based on the prevalence of eQTLs in each sample to check the association of a *BRCA2* germline variant and the co-occurrence of eQTLs. No grouping was observed, indicating an absence of eQTL patterns that are predominant within a particular cohort

Supplementary Tables

Supplementary Table 1. Supplementary information and variant characteristics of the pediatric patients.

SAMPLE NAME	GERMLINE PV	GENDER	TMB	HRD-TYPE
P.20034	c.658_659delGT (p.Val220fs) & c.5969del (p.Asp1990fs)	-	3.75	none
P.20003	c.1772_1775delTTTA (p.Ile591_Tyr592?fs)	-	0.36	*BRCA2*-type
P.20778	c.145G>T (p.Glu49Ter)	Girl	0.38	none
P.67780	c.3847_3848delGT (p.Val1283fs)	Boy	0.10	none
P.51624	c.5286T>A (p.Tyr1762Ter)	Girl	1.13	*BRCA2*-type
P.23410	c.2517_2518del (p.His839fs)	Girl	0.25	none
P.95988	c.9672dupA (p.Tyr3225fs)	-	-	-

Supplementary Table 2. Supplementary information and variant characteristics of the HMF adult control cohort. . G = germline, T = Tumor.

SAMPLE NAME	CANCER TYPE	GENDER	GENE	BIALLELIC IN TUMOR	GERMLINE PV	SNVS	INDEL	SV	TMB	ANALYSIS
CPCT02010003	Breast	female	*BRCA2*	TRUE	13:32340000C>A	3698	1247	842	1.95	G & T
CPCT02010419	Breast	female	*BRCA2*	TRUE	13:32319109G>T	2389	995	825	1.42	G & T
CPCT02020369	Breast	female	*BRCA2*	FALSE	13:32338200CTG>C	10909	1481	598	4.38	G & T
CPCT02020493	Breast	female	*BRCA2*	TRUE	13:32363369G>C	4866	1226	486	2.22	G & T
CPCT02020673	Prostate	male	*BRCA2*	FALSE	13:32341170GAA>G	4677	1203	233	2.06	G & T
CPCT02030278	Ovary	female	*BRCA2*	NA	NA	10411	1571	834	4.32	G & T
CPCT02030296	Ovary	female	*BRCA2*	TRUE	13:32340630CTT>C	5280	886	406	2.21	G & T
CPCT02040192	Ovary	female	*BRCA2*	TRUE	13:32326143TAA>T	5890	628	106	2.23	G & T
CPCT02050019	Skin	male	*BRCA2*	TRUE	13:32363269T>A	21839	4297	1072	9.17	G & T
CPCT02060063	Gallbladder	female	*BRCA2*	TRUE	13:32355268AGT>A	13478	1766	434	5.28	G & T
CPCT02070125	Prostate	male	*BRCA2*	TRUE	13:32330989CAG>C	4790	1383	736	2.33	G & T
CPCT02070218	Prostate	male	*BRCA2*	FALSE	13:32340985AGTTT>A	13966	2721	1477	6.12	G & T
CPCT02080125	Breast	female	*BRCA2*	TRUE	13:32379913G>A	4098	781	288	1.74	G & T
CPCT02100105	Breast	female	*BRCA2*	TRUE	13:32339568ATACT>A	13075	2155	606	5.34	G & T
CPCT02120117	Breast	female	*BRCA2*	TRUE	13:32340630CTT>C	12691	2290	758	5.30	G & T
CPCT02180053	Breast	female	*BRCA2*	TRUE	13:32340775T>TA	3994	944	1216	2.07	G & T
CPCT02210037	Breast	female	*BRCA2*	TRUE	13:32340300GT>G	6172	1062	390	2.57	G & T
CPCT02220031	Prostate	male	*BRCA2*	FALSE	13:32331032T>G	11922	2414	470	4.99	G & T
CPCT02230049	CUP	female	*BRCA2*	TRUE	13:32319123AG>A	16434	1670	434	6.25	G & T
CPCT02010593	Breast	female	*BRCA2*	TRUE	13:32363258CT>C	-	-	-	-	G
CPCT02030315	Hepatobiliary	male	*BRCA2*	TRUE	13:32337841TG>T	-	-	-	-	G
CPCT02040050	Ovary	female	*BRCA2*	TRUE	13:32340623CAT>C	-	-	-	-	G
CPCT02050135	Pancreas	male	*BRCA2*	TRUE	13:32379913G>A	-	-	-	-	G
CPCT02080184	Breast	female	*BRCA2*	TRUE	13:32371012AAAGG>A	-	-	-	-	G
CPCT02090031	Pancreas	male	*BRCA2*	TRUE	13:32340630CTT>C	-	-	-	-	G

Supplementary Table 3. An overview of all 65 cis-eQTLs with characteristics.

ROW.NAMES	PVALUE	SNP	ASSESSEDALLELE	OTHERALLELE	ZSCORE	GENESYMBOL	FDR	BONFERRONIP	CHR38	POS38	CADD_PHRED	POPMAX	CONSEQUENCE	IMPACT
CHR13:32315226_G/A	3.43E-13	rs3092989	A	G	7.2765	*BRCA2*	0	4.37E-05	chr13	32315226	3.964	0.199	intron_variant	MODIFIER
CHR13:32315655_A/G	0.00011611	rs206118	G	A	-3.8541	*BRCA2*	0.22922361	1	chr13	32315655	11.68	0.176	5_prime_UTR_variant	MODIFIER
CHR13:32315889_C/T	1.08E-34	rs9562605	T	C	-12.2859	*BRCA2*	0	1.37E-26	chr13	32315889	4.717	0.33	intron_variant	MODIFIER
CHR13:32316090_G/T	7.79E-36	rs9567552	T	G	-12.4966	*BRCA2*	0	9.92E-28	chr13	32316090	8.175	0.352	intron_variant	MODIFIER
CHR13:32316435_G/A	7.05E-41	rs1799943	A	G	-13.3884	*BRCA2*	0	8.98E-33	chr13	32316435	8.123	0.387	5_prime_UTR_variant	MODIFIER
CHR13:32317661_T/C	5.68E-32	rs11571579	C	T	-11.7684	*BRCA2*	0	7.23E-24	chr13	32317661	5.793	0.397	intron_variant	MODIFIER
CHR13:32318623_A/C	6.88E-31	rs11571583	C	A	-11.5558	*BRCA2*	0	8.77E-23	chr13	32318623	2.261	0.395	intron_variant	MODIFIER
CHR13:32325331_G/T	5.98E-07	rs4942423	T	G	4.9918	*BRCA2*	0.00181897	1	chr13	32325331	3.011	0.0653	intron_variant	MODIFIER
CHR13:32325430_G/C	6.11E-32	rs11571613	C	G	-11.7622	*BRCA2*	0	7.78E-24	chr13	32325430	6.926	0.397	intron_variant	MODIFIER
CHR13:32326796_T/A	7.31E-29	rs3752451	A	T	-11.1481	*BRCA2*	0	9.30E-21	chr13	32326796	2.578	0.491	intron_variant	MODIFIER
CHR13:32328257_T/C	8.72E-28	rs12869544	C	T	-10.9253	*BRCA2*	0	1.11E-19	chr13	32328257	3.927	0.396	intron_variant	MODIFIER
CHR13:32329548_C/T	6.10E-10	rs2126042	T	C	6.1879	*BRCA2*	0	0.077691231	chr13	32329548	7.398	0.248	intron_variant	MODIFIER
CHR13:32332592_A/C	7.11E-13	rs144848	C	A	7.1774	*BRCA2*	0	9.05E-05	chr13	32332592	17.21	0.358	missense_variant	MODERATE
CHR13:32334210_C/T	4.73E-29	rs1963505	T	C	-11.1866	*BRCA2*	0	6.03E-21	chr13	32334210	7.305	0.49	intron_variant	MODIFIER
CHR13:32334492_A/G	2.84E-29	rs1029304	G	A	-11.2318	*BRCA2*	0	3.62E-21	chr13	32334492	1.222	0.397	intron_variant	MODIFIER
CHR13:32336191_T/C	4.45E-10	rs2320236	C	T	6.2375	*BRCA2*	0	0.056654366	chr13	32336191	14.04	0.211	intron_variant	MODIFIER
CHR13:32337751_A/G	1.35E-30	rs1801406	G	A	-11.4982	*BRCA2*	0	1.72E-22	chr13	32337751	8.234	0.396	synonymous_variant	LOW
CHR13:32340099_C/T	0.04580346	rs4987117	T	C	-1.9971	*BRCA2*	0.99992203	1	chr13	32340099	0.185	0.0309	missense_variant	MODERATE
CHR13:32341387_C/A	3.46E-29	rs11571662	A	C	-11.2146	*BRCA2*	0	4.40E-21	chr13	32341387	0.499	0.396	intron_variant	MODIFIER
CHR13:32344804_A/G	7.90E-10	rs4942439	G	A	6.1472	*BRCA2*	1.32E-05	0.100551631	chr13	32344804	5.253	0.247	intron_variant	MODIFIER
CHR13:32344830_G/A	8.06E-10	rs4942440	A	G	6.1439	*BRCA2*	1.32E-05	0.102599287	chr13	32344830	3.488	0.247	intron_variant	MODIFIER
CHR13:32345389_A/T	8.51E-10	rs4942443	T	A	6.1353	*BRCA2*	1.32E-05	0.108306746	chr13	32345389	1.784	0.242	intron_variant	MODIFIER
CHR13:32345500_T/G	2.83E-29	rs9567576	G	T	-11.2323	*BRCA2*	0	3.60E-21	chr13	32345500	1.259	0.395	intron_variant	MODIFIER
CHR13:32346481_G/A	7.17E-09	rs206079	A	G	5.787	*BRCA2*	4.52E-05	0.912467654	chr13	32346481	7.317	0.46	intron_variant	MODIFIER
CHR13:32347807_A/G	6.86E-29	rs9567578	G	A	-11.1537	*BRCA2*	0	8.74E-21	chr13	32347807	3.428	0.394	intron_variant	MODIFIER
CHR13:32349553_C/T	7.83E-10	rs4942448	T	C	6.1485	*BRCA2*	1.32E-05	0.099667878	chr13	32349553	0.796	0.243	intron_variant	MODIFIER
CHR13:32350510_G/C	7.26E-29	rs9567582	C	G	-11.1487	*BRCA2*	0	9.24E-21	chr13	32350510	2.554	0.391	intron_variant	MODIFIER
CHR13:32350747_C/T	2.32E-13	rs559067	T	C	7.3291	*BRCA2*	0	2.95E-05	chr13	32350747	1.806	0.359	intron_variant	MODIFIER
CHR13:32351120_C/A	8.37E-10	rs1853521	A	C	6.138	*BRCA2*	1.32E-05	0.106549429	chr13	32351120	0.008	0.247	intron_variant	MODIFIER
CHR13:32352880_G/A	0.00406852	rs11571700	A	G	2.8729	*BRCA2*	0.99029283	1	chr13	32352880	1.261	0.104	intron_variant	MODIFIER
CHR13:32353757_C/T	3.58E-13	rs9943876	T	C	7.2708	*BRCA2*	0	4.56E-05	chr13	32353757	1.503	0.353	intron_variant	MODIFIER
CHR13:32353975_G/A	1.47E-14	rs9943890	A	G	7.6905	*BRCA2*	0	1.87E-06	chr13	32353975	7.084	0.268	intron_variant	MODIFIER
CHR13:32354065_A/G	1.26E-14	rs9943888	G	A	7.7097	*BRCA2*	0	1.61E-06	chr13	32354065	6.21	0.268	intron_variant	MODIFIER
CHR13:32354267_T/C	3.47E-14	rs1460817	C	T	7.5796	*BRCA2*	0	4.42E-06	chr13	32354267	8.663	0.267	intron_variant	MODIFIER
CHR13:32355095_A/G	3.86E-20	rs1799955	G	A	-9.1919	*BRCA2*	0	4.91E-12	chr13	32355095	7.146	0.394	synonymous_variant	LOW
CHR13:32359789_A/G	7.87E-20	rs9567600	G	A	-9.1149	*BRCA2*	0	1.00E-11	chr13	32359789	3.191	0.394	intron_variant	MODIFIER
CHR13:32360774_G/A	1.14E-20	rs11571717	A	G	-9.3226	*BRCA2*	0	1.45E-12	chr13	32360774	2.62	0.394	intron_variant	MODIFIER
CHR13:32361269_T/C	1.58E-14	rs9534259	C	T	7.6812	*BRCA2*	0	2.01E-06	chr13	32361269	1.303	0.268	intron_variant	MODIFIER
CHR13:32364744_A/G	9.92E-20	rs11571725	G	A	-9.0896	*BRCA2*	0	1.26E-11	chr13	32364744	3.726	0.394	intron_variant	MODIFIER

CHR13:32365149_T/G	2.89E-13	rs9534269	G	T	7.2997	*BRCA2*	0	3.68E-05	chr13	32365149	3.425	0.263	intron_variant	MODIFIER
CHR13:32366751_C/T	2.58E-14	rs11571734	T	C	7.618	*BRCA2*	0	3.28E-06	chr13	32366751	0.016	0.269	intron_variant	MODIFIER
CHR13:32367964_C/T	0.03172965	rs139834007	T	C	2.1479	*BRCA2*	0.99978687	1	chr13	32367964	0.426	0.0141	intron_variant	MODIFIER
CHR13:32369919_G/T	2.57E-19	rs9634672	T	G	-8.9855	*BRCA2*	0	3.28E-11	chr13	32369919	0.017	0.392	intron_variant	MODIFIER
CHR13:32369961_G/A	1.20E-14	rs11571742	A	G	7.7161	*BRCA2*	0	1.53E-06	chr13	32369961	0.21	0.265	intron_variant	MODIFIER
CHR13:32371400_A/T	2.07E-14	rs10492395	T	A	7.6465	*BRCA2*	0	2.63E-06	chr13	32371400	0.721	0.266	intron_variant	MODIFIER
CHR13:32371580_A/G	2.51E-19	rs9567609	G	A	-8.9883	*BRCA2*	0	3.19E-11	chr13	32371580	0.724	0.394	intron_variant	MODIFIER
CHR13:32372668_G/A	2.94E-19	rs61946969	A	G	-8.9706	*BRCA2*	0	3.75E-11	chr13	32372668	6.239	0.394	intron_variant	MODIFIER
CHR13:32376974_A/C	1.31E-19	rs3764791	C	A	-9.0595	*BRCA2*	0	1.67E-11	chr13	32376974	1.07	0.391	intron_variant	MODIFIER
CHR13:32376978_C/T	2.28E-19	rs3764792	T	C	-8.999	*BRCA2*	0	2.90E-11	chr13	32376978	2.852	0.393	intron_variant	MODIFIER
CHR13:32377317_C/T	2.29E-19	rs9567623	T	C	-8.9982	*BRCA2*	0	2.92E-11	chr13	32377317	0.506	0.394	intron_variant	MODIFIER
CHR13:32377459_A/C	1.28E-14	rs9526148	C	A	7.7077	*BRCA2*	0	1.63E-06	chr13	32377459	1.691	0.266	intron_variant	MODIFIER
CHR13:32381972_G/A	9.66E-15	rs10870659	A	G	7.7436	*BRCA2*	0	1.23E-06	chr13	32381972	8.782	0.266	intron_variant	MODIFIER
CHR13:32382146_G/A	1.15E-14	rs9534344	A	G	7.7221	*BRCA2*	0	1.46E-06	chr13	32382146	2.579	0.266	intron_variant	MODIFIER
CHR13:32383682_G/A	1.49E-19	rs7337574	A	G	-9.0451	*BRCA2*	0	1.90E-11	chr13	32383682	5.663	0.393	intron_variant	MODIFIER
CHR13:32383835_G/C	3.65E-19	rs7337784	C	G	-8.9471	*BRCA2*	0	4.64E-11	chr13	32383835	15.68	0.395	intron_variant	MODIFIER
CHR13:32384329_T/G	3.61E-19	rs11571787	G	T	-8.948	*BRCA2*	0	4.60E-11	chr13	32384329	5.396	0.394	intron_variant	MODIFIER
CHR13:32384512_G/A	3.79E-19	rs9567639	A	G	-8.9427	*BRCA2*	0	4.83E-11	chr13	32384512	0.136	0.394	intron_variant	MODIFIER
CHR13:32384750_G/A	1.14E-07	rs10492396	A	G	5.3033	*BRCA2*	0.00036453	1	chr13	32384750	14.65	0.0648	intron_variant	MODIFIER
CHR13:32389701_A/G	0.01598753	rs9526160	G	A	2.4092	*BRCA2*	0.99929288	1	chr13	32389701	2.432	0.0612	intron_variant	MODIFIER
CHR13:32395894_T/C	4.85E-14	rs1012130	C	T	7.536	*BRCA2*	0	6.18E-06	chr13	32395894	6.565	0.276	downstream_gene_variant	MODIFIER
CHR13:32399139_A/G	1.97E-14	rs7334543	G	A	7.6527	*BRCA2*	0	2.51E-06	chr13	32399139	9.257	0.306	downstream_gene_variant	MODIFIER
CHR13:32399302_A/G	8.97E-20	rs11571836	G	A	-9.1008	*BRCA2*	0	1.14E-11	chr13	32399302	9.088	0.393	downstream_gene_variant	MODIFIER

Supplementary Table 4. Number of eQTLs per filter step. After final filtering, 65 significant cis-eQTLs could be used for eQTL comparison.

	Trans-eQTLs	Cis-eQTLs
Total in database	**10293**	**6675**
Significant (p < 0.05)	**634**	**652**
Sig. & neg. exp. (Z-score > -2)	**282**	**342**
Sig. & neg. exp. & significant Bonferroni correction (p < 0.05)	**0**	**65**