Computational methods for improved cancer risk prediction based on multi-gene panel analysis in a routine diagnostic setting

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Corinna Ernst

aus Ludwigslust

Köln
2021

Berichterstatter: Prof. Dr. Andreas Beyer
Prof. Dr. Michael Nothnagel

Tag der mündlichen Prüfung: 22. Juni 2021

To my family

Contents

List of abbreviations

BC	Breast cancer
bp	Base pairs
CH	Clonal hematopoiesis
CI	Confidence interval
CNV	Copy number variant
ENIGMA	Evidence-based Network for the Interpretation of Germline Mutant Alleles
ER	Estrogen receptor
FC	Fold change
GATK	Genome Analysis Toolkit
GC-HBOC	German Consortium for Hereditay Breast and Ovarian Cancer
GWAS	Genome-wide association study
HER2	Human epidermal growth factor receptor 2
HR	Hazard ratio
LFS1	Li-Fraumeni syndrome 1
LR	Likelihood ratio
LTR	Lifetime risk
MAF	Minor allele frequency
MCC	Matthews correlation coefficient
MLPA	Multiplex ligation-dependent probe amplification
OC	Ovarian cancer
OR	Odds ratio
PCR	Polymerase chain reaction

PR	Progesterone receptor
PRS	Polygenic risk score
PTV	Protein-truncating variant
SD	Standard deviation
SNP	Single-nucleotide polymorphism
SNV	Single nucleotide variation
SO	Sequence Ontology
SV	Structural variant
TCGA	The Cancer Genome Atlas
VCF	Variant Call Format
VF	Variant allele fraction
VUS	Variant of uncertain signficance
WES	Whole exome sequencing
WGS	Whole genome sequencing

List of figures

List of tables

Preface

Main Publications (peer-reviewed)

C Ernst, E Hahnen, C Engel, M Nothnagel, J Weber, RK Schmutzler, and J Hauke. **Performance of *in silico* prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics.** BMC medical genomics, 11(1):35, 2018. [Impact Factor (2018): 2.568]

J Borde*, **C Ernst***, B Wappenschmidt, D Niederacher, K Weber-Lassalle, J Horváth, E Pohl-Rescigno, N Arnold, A Rump, A Gehrig, J Hentschel, U Faust, V Dutrannoy, A Meindl, M Kuzyakova, S Wang-Gohrke, BHF Weber, C Sutter, AE Volk, O Giannakopoulou, A Lee, C Engel, MK Schmidt, AC Antoniou, RK Schmutzler†, K Kuchenbaecker†, and E Hahnen†. **Performance of Breast Cancer Polygenic Risk Scores in 760 Female *CHEK2* Germline Mutation Carriers.** Journal of the National Cancer Institute, 2020. doi: 10.1093/jnci/djaa203. Online ahead of print. [Impact Factor (2019): 11.577]

*equal contribution as first authors
†equal contribution as last authors

L Lepkes, M Kayali, B Blümcke, J Weber, M Suszynska, S Schmidt, J Borde, K Klonowska, B Wappenschmidt, J Hauke, P Kozlowski, RK Schmutzler, E Hahnen, and **C Ernst**. **Performance of In Silico Prediction Tools for the Detection of Germline Copy Number Variations in Cancer Predisposition Genes in 4208 Female Index Patients with Familial Breast and Ovarian Cancer.** Cancers, 13(1):118, 2021. [Impact Factor (2019): 6.126]

Additional co-authored publications (peer-reviewed)

G Neidhardt, J Hauke, J Ramser, E Groß, A Gehrig, CR Müller, AK Kahlert, K Hackmann, E Honisch, D Niederacher, S Heilmann-Heimbach, A Franke, W Lieb, H Thiele, J Altmüller, P Nürnberg, K Klaschik, **C Ernst**, N Ditsch, F Jessen, A Ramirez, B Wappenschmidt, C Engel, K Rhiem, A Meindl, RK Schmutzler, and E Hahnen. **Association Between Loss-of-Function Mutations Within the *FANCM* Gene and Early-Onset Familial Breast Cancer.** JAMA Oncology, 3(9):1245-1248, 2017. [Impact Factor (2017): 20.871]

J Hauke, J Horváth, E Groß, A Gehrig, E Honisch, K Hackmann, G Schmidt, N Arnold, U Faust, C Sutter, J Hentschel, S Wang-Gohrke, M Smogavec, BHF Weber, N Weber-Lassalle, K Weber-Lassalle, J Borde, **C Ernst**, J Altmüller, AE Volk, H Thiele, V Hübbel, P Nürnberg, K Keupp, B Versmold, E Pohl, C Kubisch, S Grill, V Paul, N Herold, N Lichey, K Rhiem, N Ditsch, C Ruckert, B Wappenschmidt, B Auber, A Rump, D Niederacher, T Haaf, J Ramser, B Dworniczak, C Engel, A Meindl, RK Schmutzler and E Hahnen. **Gene panel testing of 5589 *BRCA1/2*-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer.** Cancer Medicine, 7(4):1349-1358, 2018. [Impact Factor (2018): 3.357]

N Weber-Lassalle, J Hauke, J Ramser, L Richters, E Gross, B Blümcke, A Gehrig, AK Kahlert, CR Müller, K Hackmann, E Honisch, K Weber-Lassalle, D Niederacher, J Borde, H Thiele, **C Ernst**, J Altmüller, G Neidhardt, P Nürnberg, K Klaschik, C Schroeder, K Platzer, AE Volk, S Wang-Gohrke, W Just, B Auber, C Kubisch, G Schmidt, J Horváth, B Wappenschmidt, C Engel, N Arnold, B Dworniczak, K Rhiem, A Meindl, RK Schmutzler, and E Hahnen. ***BRIP1* loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer.** Breast Cancer Research, 20(1):7, 2018. [Impact Factor (2018): 5.676]

K Weber-Lassalle, P Harter, J Hauke, **C Ernst**, S Kommoss, F Marmé, N Weber-Lassalle, K Prieske, D Dietrich, J Borde, E Pohl-Rescigno, A Reuss, B Ataseven, C Engel, JC Stingl, RK Schmutzler, and E Hahnen. **Diagnosis of Li-Fraumeni Syndrome: Differentiating *TP53* germline mutations from clonal hematopoiesis: Results of the observational AGO-TR1 trial.** Human Mutation, 39(12):2040-2046, 2018. [Impact Factor (2018): 4.453]

J Hauke, E Hahnen, S Schneider, A Reuss, L Richters, S Kommoss, A Heimbach, F Marmé, S Schmidt, K Prieske, H Gevensleben, A Burges, J Borde, N De Gregorio, P Nürnberg, A El-Balat, H Thiele, F Hilpert, J Altmüller, W Meier, D Dietrich, R Kimmig, B Schoemig-Markiefka, K Kast, E Braicu, K Baumann, C Jackisch, TW Park-Simon, **C Ernst**, L Hanker, J Pfisterer, A Schnelzer, A du Bois, RK Schmutzler and P Harter. **Deleterious somatic variants in 473 consecutive individuals with ovarian cancer: results of the observational AGO-TR1 study (NCT02222883).** Journal of Medical Genetics, 56(9):574-580, 2019. [Impact Factor (2019): 4.943]

K Klaschik, J Hauke, G Neidhardt, C Tränkle, HM Surowy, S Heilmann-Heimbach, G Rappl, E Mangold, N Arnold, D Niederacher, C Sutter, B Burwinkel, C Engel, B Wappenschmidt, A Meindl, **C Ernst**, K Weber-Lassalle, N Weber-Lassalle, S Schmidt, J Borde, RK Schmutzler, E Hahnen, and E Pohl-Rescigno. **The *GPRC5A* frameshift variant c.183del is not associated with increased breast cancer risk in *BRCA1* mutation carriers.** International Journal of Cancer, 144(7):1761-1763, 2019. [Impact Factor (2019): 5.145]

N Weber-Lassalle, J Borde, K Weber-Lassalle, J Horváth, D Niederacher, N Arnold, S Kaulfuß, **C Ernst**, VG Paul, E Honisch, K Klaschik, AE Volk, C Kubisch, S Rapp, N Lichey, J Altmüller, L Lepkes, E Pohl-Rescigno, H Thiele, P Nürnberg, M Larsen, L Richters, K Rhiem, B Wappenschmidt, C Engel, A Meindl, RK Schmutzler, E Hahnen and J Hauke. **Germline loss-of-function variants in the *BARD1* gene are associated with early-onset familial breast cancer but not ovarian cancer.** Breast Cancer Research, 21(1):55, 2019. [Impact Factor (2019): 4.988]

E Pohl-Rescigno, J Hauke, S Loibl, V Möbus, C Denkert, P Fasching, M Kayali, **C Ernst**, N Weber-Lassalle, C Hanusch, H Tesch, V Müller, J Altmüller, H Thiele, M Untch, K Lübbe, P Nürnberg, K Rhiem, J Furlanetto, B Lederer, C Jackisch, V Nekljudova, RK Schmutzler, A Schneeweiss and E Hahnen. **Association of Germline Variant Status With Therapy Response in High-risk Early-Stage Breast Cancer: A Secondary Analysis of the GeparOcto Randomized Clinical Trial.** JAMA Oncology, 6(5):1-5, 2020. [Impact Factor (2019): 24.799]

J Hauke, P Harter, **C Ernst**, A Burges, S Schmidt, A Reuss, J Borde, N de Gregorio, D Dietrich, A El-Balat, M Kayali, H Gevensleben, F Hilpert, J Altmüller, A Heimbach, W Meier, B Schoemig-Markiefka, H Thiele, R Kimmig, P Nürnberg, K Kast, L Richters, J Sehouli, RK Schmutzler, and E Hahnen. **Sensitivity and specificity of loss of heterozygosity analysis for the classification of rare germline variants in *BRCA1/2*: Results of the observational AGO-TR1 study (NCT02222883).** Journal of Medical Genetics, 2020. doi: 10.1136/jmedgenet-2020-107353. Online ahead of print. [Impact Factor (2019): 4.943]

1 Summary

Multi-gene panel approaches allow for screening for putative disease-causing genetic factors in several genes simultaneously, while keeping costs, storage requirements, and computational times comparatively low compared to whole genome or exome sequencing. Therefore, multi-gene panel sequencing has become a standard approach for the investigation and diagnosis of various diseases with a hereditary component in clinical labs wordwide, and several tools have become established for multi-gene panel data processing, providing automized, easy-to-use solutions for the detection and annotation of single nucleotide variants (SNVs) and short insertions and deletions (indels) in exons of well-established disease-associated genes.

This thesis demonstrates, in the context of personalized risk prediction for familial breast and ovarian cancer (BC/OC), how bioinformatic analyses that go beyond standard variant calling with automized workflows, can contribute to an improvement of genetic testing based on multi-gene panels in clinical diagnostics. These improvements include multi-gene panel design, variant detection, and variant interpretation.

Since only less than one third of BC/OC cases with a familial burden can be explained by a germline mutation in confirmed high- to moderate-risk genes, the search for further genetic risk factors is ongoing, but may be hindered by low mutation prevalences in the corresponding genes that demand for huge sample sizes in order to achieve sufficient statistical power. Here, the potential association of pathogenic variants in suspected risk genes *FANCM*, *BARD1*, and *BRIP1* with hereditary BC and/or OC was assessed in case-control studies including well-characterized index patients and geographically matched female controls. *FANCM* and *BARD1* were confirmed as risk genes for hereditary BC, and *BRIP1* was confirmed as a highly penetrant OC risk gene without pronounced effects on BC risk. Consequently, coding regions of *FANCM* and *BARD1* should be included in sequencing targets of multi-gene panels for diagnostic germline testing of individuals at risk for familial BC, and coding regions of *BRIP1* for individuals at risk for familial OC, respectively.

Recent studies revealed that BC/OC risks are modified by additional genetic factors, i.e., common SNVs and indels which are usually not even located in coding genomic regions. Investigation of these BC/OC-associated polymorphisms represents a paradigm shift in contrast to the analysis of rare (pathogenic) variants. Their effects are not sufficiently large to contribute individually to BC/OC risks, but they can be combined into polygenic risk scores (PRS), which could achieve clinically useful degrees of risk discrimination. In collaboration with Julika Borde, I assessed the utility of PRSs for BC risk prediction in a clinical cohort of females carrying a heterozygous protein-truncating variant (PTV) in *CHEK2* and were independent of former genome-wide association studies. We found that, based on PRSs, BC risk can be stratified such that *CHEK2* PTV

carriers may have both a BC risk equivalent to that of the general population but, on the other hand, may also fall into risk groups for which access to intensified prophylactic measures is recommended. The SNP sets employed in our study each comprise less than 100 loci, and hence, have the potential to be straightforwardly implemented into multi-gene panel analyses. In addition, SNPs can be used for ethnicity checks and quality assurance purposes.

Detection of large genomic insertions or deletions, so-called copy number variants (CNVs), from sequencing data requires read depth-based approaches that go beyond standard variant calling. In a joint work with Louisa Lepkes, the utility of these *in silico* CNV detection approaches for multi-gene panel data in clinical diagnostics was evaluated, and the prevalence of CNVs in cancer predisposition genes in individuals at risk for familial BC/OC was assessed. We showed that CNVs constitute a non-negligible fraction in the spectrum of putative BC/OC-causing variants, namely 1.81% in our study sample of 4208 female index patients. However, due to high proportions of false positive predictions, which primarily accumulated at the extremes of the length or GC content distribution of sequencing targets, wet lab verification of *in silico* predicted CNVs is required in the framework of clinical diagnostics.

The third part of the thesis deals with considerations regarding the interpretation of genetic testing outcomes in the context of genetic counseling. Interpretation of missense mutations is a particularly challenging task, as their impact to protein function are difficult to predict and they can therefore often only be classified as variants of uncertain significance (VUS). Thus, the use of *in silico* approaches for automated variant classification has become established in many laboratories. I evaluated the performance of four *in silico* prediction tools embedded in the widely-used, commercial Alamut™ Visual software (Interactive Biosoftware, Rouen, France) and found that all tools under investigation suffered from poor specificities, resulting in an unacceptable proportion of variants falsely classified as pathogenic, and that this shortcoming could not be bypassed by considering the predictions in combination. Thus, clinical consequences should never be based solely on *in silico* forecasts, but my findings indicate that *in silico* prediction tools provide clues to the benignity of variants.

In collaboration with Dr. Jan Hauke, the determination of variant pathogenicity based on the comparison of observed variant allele fractions (VFs) in paired blood- and tumor-derived samples was assessed, considering 208 rare *BRCA1/2* germline variants in 181 OC patients. Our results demonstrate that a significantly increased VF in tumor in comparison to the corresponding blood-derived sample are insufficient to infer pathogenicity, but decreased VFs may provide a suitable criterion for the assessment of *BRCA1/2* variants as benign.

In collaboration with Konstantin Weber-Lassalle, I investigated pairwise blood- and tumor-derived DNA samples of OC patients with the aim to prove the existence of pathogenic variants in *TP53* and *PPM1D* in blood cells arising from clonal hematopoiesis (CH) rather than from germline inheritance, and to evaluate the frequency of CH occurences in dependence to the exposure to chemotherapy. We found that CH represents a frequent event following chemotherapy, affecting 26 out of the 523 OC index patients

enrolled in our study sample. Therefore, the possibility of CH always has to be considered prior to a potential misdiagnosis of Li-Fraumeni syndrome 1, a cancer predisposition syndrome linked to pathogenic variants in *TP53*.

2 Introduction

2.1 Next-generation sequencing

2.1.1 Next-generation sequencing technology

Starting from the first decade of this century, a new class of sequencing methods became established, which are referred to as second- or next-generation sequencing (NGS). The advantage of NGS over the previously dominant Sanger sequencing method is its time and cost efficiency [50]. The term is commonly used in reference to implementations of cyclic-array sequencing, as provided by the widespread commercial NGS platform Illumina[1] [151]. The typical sequencing protocol can be divided into three steps, namely (1) random fragmentation of input DNA with subsequent ligation of artificial adapter sequences, (2) separation and clustered amplification of single DNA molecules via polymerase chain reaction (PCR), and (3) the sequencing process itself by alternating cycles of enzyme-driven biochemistry treatment and imaging-based data aquisition [151]. Each treatment cycle serves for the recognition of one additional nucleotide of the DNA sequence per template. Therefore, the starting point of NGS data processing typically appears as nucleotide sequences – so-called (sequencing) reads – of constant length (i.e., read length), corresponding to the number of applied treatment cycles.

Several variations of the sequencing protocol exist. Of particular relevance are paired-end sequencing and targeted sequencing. For paired-end sequencing, both ends of amplified DNA fragments are sequenced in opposite orientation. As the (expected) distance of the resulting paired reads in the donor genome sequence is approximately known, paired reads showing aberrating insert lengths or are in anomalous orientation can be of use for the detection of genomic rearrangements [114].

NGS is characterized by an increased base call error rate in comparison to Sanger sequencing [46, 52], i.e., depending on sequence patterns and read position, a notable amount of bases (>1 in 10^5 bases) is expected to be reported incorrectly [90, 127, 151].

A further challenge of NGS is the amount of data that is generated. Given that each genomic reference position is covered by several sequencing reads, data storage and processing have become the typical bottlenecks for NGS applications [37]. For many purposes, especially in a clinical setting, it is therefore advisable to limit sequencing to regions of particular interest, e.g., to coding regions or exons only. Thus, costs, storage requirements and computation times are decreased considerably, and more individuals can be sequenced in parallel, and with higher throughput [58].

[1] http://www.illumina.com

2.1.2 Next-generation sequencing data processing

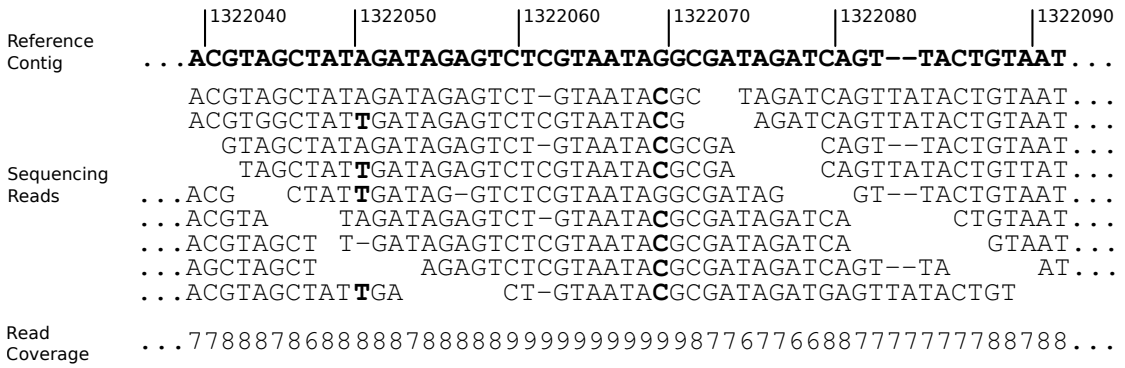
Sequencing reads are typically provided in FASTQ format as (compressed) text data accompanied by character-encoded base quality values referring to the probability of miscalling per observed nucleotide [30]. In a typical NGS approach, these files contain millions of reads, for which the information about the originating position in the genome of interest is lost, and must first be assigned based on the observed sequences. This analysis step is called read mapping. As the genomes of sequenced individuals are usually unknown, a so-called reference genome, a consensus among several individuals from the same species, serves as genome template in the mapping procedure. The first human reference genome was released in 2003 and covers 99% of the euchromatin, interrupted by several gaps in the assembly [71]. The reference has been improved several times since then, not least due to advances in sequencing technology, such as (ultra-)long read approaches, but also due to more efficient algorithms and enhanced computing capacities. The current version hg38 (equivalent to GRCh38) was released in 2013 [147], and offered for the first time centromere sequences, but also its predecessor, hg19 (equivalent to GRCh37) [28], is still frequently used, especially for practical sequencing applications, such as in clinical diagnostics.

An exact occurrence of its observed sequence (or its reverse complement) in the reference genome represents the most likely origin of a sequencing read. But, due to natural genetic variation and technical artifacts, many reads do not match exactly, and the read mapping has to be error-tolerant. Hence, the read mapping problem can be defined as the search for the best sequence alignment between each read and a given reference genome. Alignment means the arrangement of sequences, i.e., strings over an alphabet Σ including gaps `-`, such that the sequences remain unchanged (except for included gaps) and each element of the sequences can be assigned to an alignment column. Typically, the optimal alignment is interpreted as the one that would require the fewest number of edit operations to convert the sequence in one alignment row to the sequence in another alignment row. In the context of DNA sequencing, the edit operations between two strings over $\Sigma = \{\texttt{A}, \texttt{C}, \texttt{G}, \texttt{T}, \texttt{-}\}$ represent substitutions, insertions and deletions of bases compared to the reference sequence, and hence, if not caused by technical artifacts, genomic variation. See Figure 2.1 A for the visualization of mapped reads against an artificial reference sequence, including four putative genomic variants.

Due to the size of typical reference genomes (hg19: $> 3.1 \times 10^9$ base pairs (bp), hg38: $> 3.2 \times 10^9$ bp), and the number of reads to process, exact read mapping is unfeasible in practice, and heuristics were developed for practical applications. The popular read mapper Burrows-Wheeler Alignment tool (`bwa`) [96], and its extensions `bwa-mem` and `bwa-mem2`, are based on backward search with Burrows-Wheeler Transform [95, 176].

After read mapping, each base position in the reference genome can be assigned a read coverage (or synonymously read depth), i.e., the number of reads covering the respective position in the mapping. In addition, a mapping quality is generally reported for each read, referring to the confidence that the assigned genomic location is the true and only plausible read origin. A standard format for represenation, storage and processing of mapped reads is the SAM (Sequence Alignment Map) format, which is

A



B

```
##fileformat=VCFv4.2
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=RO,Number=1,Type=Integer,Description="Reference allele observation count">
##FORMAT=<ID=AO,Number=A,Type=Integer,Description="Alternate allele observation count">
##FORMAT=<ID=VF,Number=A,Type=Float,Description="Alternate variant allele fraction">
#CHROM POS     ID REF ALT QUAL FILTER INFO FORMAT         SAMPLE
chr3   1322050 .  A   T   .    .      .    GT:DP:RO:AO:VF 0/1:8:3:4:0.5
chr3   1322060 .  TC  T   .    .      .    GT:DP:RO:AO:VF 0/1:9:5:4:0.44
chr3   1322069 .  G   C   .    .      .    GT:DP:RO:AO:VF 1/1:9:1:8:0.89
chr3   1322082 .  T   TTA .    .      .    GT:DP:RO:AO:VF 0/1:7:3:4:0.57
```

Figure 2.1: (A) Schematic representation of mapped reads with read length 30 and (B) corresponding Variant Call Format (VCF) output, assuming the reference contig is called `chr3`. Mapped reads and VCF output represent 4 putative genomic variants, namely a heterozygous base substitution, a heterozygous deletion of one base `C`, a homozygous substitution, and a heterozygous insertion of `TA` in comparison to the reference sequence. Visualization of read directions (forward or reverse), base and mapping qualities was omitted.

called BAM (Binary Alignment Map) if binary compressed [97]. SAM/BAM files contain the obtained sequences for each sequencing read, accompanied by base qualities, as well as alignments to the reference genome and mapping qualities.

The process of finding genetic variation from mapped reads with respect to the reference genome is commonly referred to as *variant calling*, although the term typically refers only to the identification of single nucleotide variations (SNVs) and short insertions and deletions (indels). Standard variant calling approaches aim to distinguish true genetic variation from technical artifacts given the observed base calls in all reads covering a certain reference position, frequently under consideration of read directions, base and/or mapping qualities, and can roughly be distinguished into approaches for germline and somatic variant calling.

Germline variants are usually inherited from the previous generation via sperm cells

or oocytes, and thus, are present in every cell of the corresponding individual. *Somatic* variants arise *de novo* in the genome, they are passed on to daughter cells during somatic cell division, but they are not inherited to descendants. Germline variants appear either heterozygous or homozygous in autosomes. Hence, germline variant calling is essentially the task of deciding given the proportion of reads showing a putative alternative allele, i.e., the *variant allele fraction* or simply *variant fraction* (VF), between the three posssible genotypes homozygous reference allele (VF≈0), heterozygous alternative allele (VF≈0.5), and homozygous alternative allele (VF≈1). In the context of somatic variant calling, the assumption that VFs directly map zygotic status is not necessarily true because the processed DNA does not originate from a single nucleus (except in single-cell sequencing approaches). Thus, genomic alterations must be assumed to occur even at much lower VFs than 0.5, making them difficult to distinguish from technical artifacts [185]. A widespread approach to circumvent this drawback, is variant calling for matched pairs of DNA samples originating from the same individual but from different tissues, e.g., very often tumor and blood. This strategy allows somatic variants to be distinguished from germline variants and the joint analysis provide additional evidence in the determination of sequencing artifacts in case the reads of the associated sample originating from a different tissue also show partially deviating bases at the same genomic position.

A variety of tools, both commercial and publicly available, exist for the aim of SNV and indel calling, some of which can be applied for both germline and somatic (single sample as well as paired sample) variant calling. Common publicly available tools for germline variant calling are the Genome Analysis Toolkit (GATK) HaplotypeCaller [40] and FreeBayes [54]. Both approaches create local read *assemblies*, i.e. alignments of reads with each other (instead of alignments against a reference sequence), with subsequent derivation of the potential original genomic sequence, rather than examining each genomic locus independently. These local assemblies provide *phasing* information, i.e., assignment of reads and/or genomic alterations to haplotypes, which is incorporated into the calling procedure to provide additional statistical evidence.

Genetic variations are predominantly stored in standardized text files in Variant Call Format (VCF). An example for VCF encoding of four variants is shown in Figure 2.1 B.

Identification of genomic alterations affecting more than a few nucleotides, and therefore can not be fully covered by multiple overlapping reads in the mapping, requires different algorithms and the consideration of entities other than for the calling of SNVs and indels. Such variants are called *structural variants* (SVs) and are usually defined as genomic rearrangements, namely deletions, duplications, insertions, inversions, or translocations, affecting more than 50 bp [10, 105, 162].

Read depth-based approaches are adapted to the identifcation of copy number variants (CNVs), i.e. large deletions or duplications having significant effects on observed read counts or mean read coverages in the affected genomic region. Further, several approaches have been published for the detection of SVs by inference from discordant read pairs and/or so-called split reads [7, 26, 134, 162]. Read-pair methods use discordant read pairs with inconsistent orientation or insert sizes for SV calling, and split read-based

approaches identify start and stop positions of putative SVs by scanning for partially mapped reads. Furthermore, assembly-based approaches exist for SV calling, which are typically hampered by the short read lengths resulting from NGS approaches [162]. As all of the aforementioned strategies, i.e., read depth-based, read pair, split read and assembly-based approaches, imply their own intrinsic limitations concerning their ability for detection of all possible types of SVs, the majority of tools for SV calling make use of a combination of existing methods.

2.1.3 Targeted next-generation sequencing

Targeted NGS approaches narrow analyses to genomic regions of special interest, and hence, decrease costs, storage requirements, and computation times per sample significantly in comparison to *whole genome sequencing* (WGS). Thus, more samples can be sequenced per run and with higher read coverage, which in turn facilitates the reliable detection of genetic variants [58]. In case the relevant region should not or cannot be further restricted, sequencing is often limited to the exons of protein-coding genes, i.e., the *exome*. The exact definiton of the targeted region in *whole exome sequencing* (WES) approaches usually depends on the commercial kits offered. For humans, they comprise about 2% of the genome [164, 180].

In routine diagnostics, it is often sufficient to limit genetic tests to genes that are known or assumed to be implicated in a particular phenotype, as provided by so-called *multi-gene panel* approaches. This reduces the ressources required per sample even further compared to WES, and consequently multi-gene panel approaches have become an established tool for the investigation and diagnosis of a variety of diseases [12, 65, 81, 122, 142, 191].

Broadly, there exist two common techniques to target specific genomic regions: amplicon-based approaches and hybridization-based capture [82]. *Amplicon sequencing* relies on specifically designed paired PCR primers for the genomic regions of interest with a subsequent additional PCR reaction, and results in reads with identical genomic start and stop positions, but outstanding read depth [143,192]. For *hybridization-based capture sequencing*, customized synthetic oligonucleotide probes are used to capture previously randomly fragmented DNA and bind it to a solid surface. This allows for the elimination of dispensable genomic DNA regions prior to amplification and sequencing [82,143]. Hybridization-based capture sequencing results in sequencing reads with randomly distributed genomic start positions within targeted regions, and more uniformly distributed patterns of read coverage in comparison to amplicon-based approaches. This is also due to the fact that putative PCR duplicates, i.e. reads that begin and end at the same position in the genome, are typically identified and sorted out in NGS data analysis workflows. In amplicon sequencing, this step is prohibited, since reads originating from identical primers share start and stop positions by design [82]. The resulting low homogeneity of read coverage is challenging for reliable CNV detection with read depth-based approaches [18]. With regard to SNV and indel calling based on amplicon sequencing data, technical errors that occur in early PCR cycles, may be copied many times and can hardly be distinguished from real-world genetic variation, especially if no additional in-

formation can be derived from further overlapping and/or reverse amplicons. However, amplicon sequencing is generally more cost-efficient than hybridization-based capture sequencing (although prices are converging due to the general decline of sequencing costs [150]), with lower amounts of reads mapping outside target regions, and allows for the detection of (somatic) variants with extreme low VF [62, 143].

2.1.4 Variant annotation

Due to human diversity and the sheer size of the human genome, (targeted) NGS typically results in a vast number of genomic alterations detected, e.g., output of typical human WES is expected to consist of 30,000 to 50,000 variants per sample [126]. However, only the very smallest part of these variants is likely to be associated with a phenotype of interest (if at all). Therefore, automated filters are required to narrow down the set of detected variants to these of putative interest for further analysis.

When looking at diseases with a pronounced monogenic component, it can generally be assumed that causative mutations occur rarely in the general population. Consequently, a commonly applied filter in clinical diagnostics is based on observed *minor allele frequencies* (MAFs) in population-specific databases. For example, in the course of breast and ovarian cancer (BC/OC) risk prediction, the *BRCA1/2* Gene Variant Classification Criteria proposed by the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) recommend the classification of variants with MAF > 0.01 as `Not Pathogenic` per default [4, 125]. Publicly available databases that are suitable and widely-used sources for the extraction of population-specific MAFs are listed in Table 2.1. However, rareness is not a sufficient criterion to conclude on the effect of a genetic variant at the protein or even physiological level. The automated provision of additional information, i.e., a biological context gaining insights for a variant's characteristics and its putative effects, is referred to as *variant annotation*. Straightforward is the prediction on transcript level, i.e., annotation of affected transcripts (and thus, implicitly of the gene affected) and the expected consequences at protein level, which are commonly described using a standardized set of terms based on Sequence Ontology (SO) [5, 29, 112]. Without further knowledge, protein-truncating variants (PTVs), i.e., variants predicted to shorten the coding sequence, in corresponding known risk genes are most likely clinically relevant. This generally includes frameshift and nonsense variants, splice site-disrupting variants and deletions with lengths $>$ 50bp [138]. Although the terms PTV and loss-of-function variant are often used synonymously, possible effects at the protein level cannot necessarily be predicted at the transcript level alone. For example, missense variants may be localized in functional domains and thus also have significant effects on protein function. Common and publicly available annotation software for predicting the effects of genetic variants at the transcript and/or gene level are SnpEff [29] and Ensembl Variant Effect Predictor (VEP) [112]. Both tools also provide the translation from the genome-related VCF format into the transcript- and/or protein-related variant nomenclature proposed by the Human Genome Variation Society (HGVS nomenclature) [39], which is widely-used in clinical routine.

Even more challenging than predicting the effect of a variant at the sequence level is

Table 2.1: Commonly used databases for the extraction of population-specific minor allele frequencies (MAFs)

Database	Description	Reference
1000 Genomes Project	Low coverage and exome sequencing data for 2504 individuals from 26 populations	[165]
dbSNP	Human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publications, population-specific frequencies, molecular consequences, and genomic and RefSeq mapping information for both common variations and clinical mutations	[152]
ExAC	Gene- and transcript-centric displays of variation, including population-specific frequencies and functional annotation data as well as short read support for variants called in 60706 individuals	[92]
FLOSSIES	Frequencies and profiles of variants in 27 (putative) BC risk genes in 10000 women (≈7000 European American, ≈3000 African American) who are older than age 70 and have never had cancer	[3]
gnomAD v2	Includes genetic variation for 15708 whole genomes and 125748 exomes with respect to reference GRCh37/hg19	[73]
gnomAD v3	Includes genetic variation fo 71702 whole genomes with respect to reference GRCh38	[73]

the *in silico* prediction of its clinical significance, i.e., typically its categorization into the five pathogenicity classes `Pathogenic`, `Likely pathogenic`, `Uncertain significance`, `Likely benign`, and `Benign` [129]. Corresponding information can potentially be retrieved from the literatue or from disease variant databases such as ClinVar [88] or the Human Gene Mutation Database (HGMD) [158]. However, these classifications may be incorrect, based on insufficient evidence, or inconsistent across databases and/or submitters [132], if the variant of interest is listed at all. Therefore, numerous tools were developed aiming to predict functional effects and potential clinical significance *in silico* on the basis of additional information such as distances between the biochemical properties of amino acids involved, expected changes in protein 3D structure, and evolutionary conservation, and/or under employment of machine learning approaches trained on well-characterized variant sets. Hu and colleagues recently provided a comprehensive listing of available software [69]. However, the reliability of such *in silico* predictions remains elusive, and, in particular, their applicability in clinical diagnostics. Thus, the *multifactorial likelihood analysis* method [57, 125] has become the recommended standard for the assessment of variant pathogenicity in cancer syndrome genes. This approach aims to combine a variety of independent sources of evidence, e.g., *in silico* prediction, co-segregation, family cancer history, co-occurence with further pathogenic variants, case-control information, and tumor pathology, into a product of likelihood ratios (LRs), but the required input data may usually not be at hand, especially for rare variants.

In recent years, *genome-wide association studies* (GWASs) in large populations identified single-nucleotide polymorphisms (SNPs), i.e., variants that are present in a sufficiently large fraction (e.g., $> 1\%$) of a population, which can be used as markers for functional alleles in complex genetic traits [36,42,159,168]. This includes general characteristics such as body size, but also mental disorders such as depression and schizophrenia, metabolic diseases such as diabetes, or cancer diseases, among others [42]. Although significant associations exist, the observed effects are generally too small to contribute individually to the development of a particular phenotype, but can be combined to overall estimates of genetic propensity, so-called *polygenic risk scores* (PRSs) via

$$\mathrm{PRS}_i = \sum_{j=1}^{N} \beta_j g_{ij} \text{ with } g \in \{0, 1, 2\},$$

where β is the per-allele log OR and g_{ij} is the number of effect alleles in person i for locus j. Investigation of polymorphisms, which are often localized in non-coding genomic sequence, represents a paradigm shift in contrast to the analysis of rare (pathogenic) variants. Besides the improved prediction of genetic predispositions, SNPs can be utilized for several quality checks, e.g., checks for gender, duplicated samples and ethnicity background.

2.2 Hereditary breast and ovarian cancer

2.2.1 Hereditary breast cancer

BC is the most commonly diagnosed cancer among women and the leading cause of female cancer deaths. In 2018, the number of newly diagnosed BC cases worldwide was estimated to be about 2.1 million [24]. In Germany, about one in eight women develops BC in her lifetime [139]. BC is a heterogeneous disease whose molecular subtypes can be distinguished according to the expression of hormonal receptors, namely estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). These subtypes can be approximated clinically into luminal A (ER- and/or PR-positve, HER2-negative), luminal B (ER- and/or PR-positive, HER2-positive), HER2-enriched (ER- and PR-negative, HER2-positive), and triple-negative BC (ER-, PR- and HER2-negative) [67].

Most cases of BC in women occur sporadically and are caused by somatic mutations acquired in breast tissue with increasing age and depending on environmental influences, nutritional factors, hormone status and other factors [24, 91]. It is estimated that approximately 5–10% of all BC diagnoses can be attributed to causal germline mutations in established risk genes [24]. Among these BC risk genes, *BRCA1* and *BRCA2* are frequently affected: Kast and colleagues [74] reported a *BRCA1/2* mutation prevalence of 18.3% in German index patients with familial BC history[2]. These two genes are also among the BC risk genes with highest penetrance: Kuchenbaecker and colleagues ascertained lifetime risks (i.e., cumulative risks of developing BC until age 80 years, LTR) of 72% for *BRCA1* mutation carriers and of 69% for *BRCA2* mutation carriers [83]. The distinction between risk genes with high, moderate and low penetrance is usually based on relative risks, i.e., the ratio of LTRs for developing BC (or another disease of interest) in the presence and absence of a pathogenic variant (or another risk factor) [113]. Genes with high penetrance are associated with relative risks >4, whereas genes with low penetrance are associated with relative risks <2 [45, 156]. Data on relative risks for BC often differ widely between studies, due to differences in the composition of the study samples considered, e.g., in terms of defintion of BC family history, but also because odds ratios (ORs) are often mistakenly interpreted as relative risks [113, 145].

TP53, *PTEN*, and *CDH1* were identified as BC risk genes with high penetrance besides *BRCA1/2*, but mutations in these genes occur at much lower frequencies, e.g., *TP53* germline mutations are accounted for <1% of BC cases [65]. *TP53* and *PTEN* are associated with syndromes that predispose to cancer development in general: *TP53* germline mutations cause *Li-Fraumeni syndrome 1* (LFS1) and germline mutations in *PTEN* are the cause of *Cowden syndrome*. Short descriptions, associated LTRs and mutation prevalences for BC risk genes with high penetrance are summarized in Table 2.2. Reported prevalences of pathogenic germline variants in Europeans were estimated based on Non-Finnish Europeans under exclusion of The Cancer Genome Atlas (TCGA) data in ExAC [92] by a similar approach as proposed for the identification of population-

[2] ≥2 females with BC, of these ≥1 with BC before the age of 50 years, no relatives with OC and metachronous BC

specific frequencies of pathogenic *BRCA1/2* mutations by Maxwell et al. [110]. Variants were considered as pathogenic if they effect frame shifts or nonsense, or were located +/-2 bp around splice sites with respect to canonical transcripts NM_007294 (*BRCA1*), NM_000059 (*BRCA2*), NM_004360 (*CDH1*), NM_000314 (*PTEN*), and NM_000546 (*TP53*). Further, missense variants and inframe deletions were considered pathogenic if reported as `Pathogenic` and/or `Likely pathogenic` in the ClinVar database [88]. Variants were excluded if they were located upstream or downstream to the last known pathogenic variant in the corresponding transcript due to ClinVar. It must be emphasized that the resulting values are rough estimates, as structural variants were not considered and potentially pathogenic missense and inframe mutations may not have been adequately reported in ClinVar.

Further genes were shown to be associated with intermediately increased risks for BC, i.e., relative risks from 2 to 4 [45], namely *ATM*, *CHEK2*, *NF1*, *PALB2*, and *STK11*. Characteristics including associated LTRs for these BC risk genes with moderate penetrance, are summarized in Table 2.3. Reported prevalences of pathogenic variants are based on non-Finnish Europeans under exclusion of TCGA data in ExAC [92] considering RefSeq transcripts NM_000051 (*ATM*), NM_001005735 (*CHEK2*), NM_001042492 (*NF1*), NM_024675 (*PALB2*), and NM_000455 (*STK11*) as described above. *STK11* germline mutations cause *Peutz-Jeghers syndrome* and germline mutations in *NF1* are the cause of neurofibromatosis type 1 (NF1). Both syndromes result in a generally increased risk of cancer [136, 175].

The search for genes associated with an increased risk for BC is ongoing, as less than one fifth of BC cases in individuals with a familial burden of gynaecological tumors[3] can be explained by a germline mutation in high risk genes *BRCA1* and *BRCA2* [74], and approximately further 6% by a mutation in additional BC risk genes with high to moderate penetrance [65]. See Table 2.4 for an overview of BC risk genes with low penetrance and/or genes whose impact on BC risk is suspected. Reported prevalences of pathogenic variants are based on non-Finnish Europeans under exclusion of TCGA data in ExAC [92] considering RefSeq transcripts NM_000465 (*BARD1*), NM_032043 (*BRIP1*), NM_020937 (*FANCM*), NM_002485 (*NBN*), NM_058216 (*RAD51C*), and NM_002878 (*RAD51D*) as described above.

Determination of additional genes causing a familial burden for BC is complicated by low mutation prevalences in the corresponding genes and/or comparatively small associated effects that demand for huge sample sizes in order to achieve sufficient statistical power. In addition, there may be interactions between the genes of interest, or risks are increased exclusively for a particular BC subtype.

[3] Index patients fulfilling the inclusion criteria of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) for germline testing (see Table 2.7)

Table 2.2: Characterization, associated lifetime risks until the age of 80 years (LTR BC 80y) and estimated prevalences of pathogenic variants in non-Finnish Europeans (see section 2.2.1) of breast cancer (BC) risk genes with high penetrance.

Gene	Locus	Protein function	LTR BC 80y	Prevalence	References
BRCA1	17q21.31	Tumor suppression by maintaining genomic stability by mediating transcription, DNA repair of double-stranded breaks, and recombination	0.55–0.75	3.2×10^{-3}	[41, 45, 68, 83]
BRCA2	13q13.1	Involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair	0.45–0.76	3.9×10^{-3}	[41, 45, 68, 83]
CDH1	16q22.1	Calcium-dependent cell-cell adhesion glycoprotein, mediates cell-cell interaction	0.43–0.55	7.4×10^{-5}	[45, 140, 184]
PTEN	10q23.3	Negative regulation of intracellular phosphatidylinositol-3,4,5-trisphosphate levels, tumor suppression by negative regulation of AKT/PKB signaling pathway; germline mutations cause Cowden syndrome	0.85–0.91	3.7×10^{-5}	[1, 160]
TP53	17p13.1	Tumor supression, mediating cellular stress response; germline mutations cause Li-Fraumeni syndrome 1	0.61	7.7×10^{-4}	[1, 106]
General Population			0.12–0.15		[2, 111, 139]

Table 2.3: Characterization, associated lifetime risks until the age of 80 years (LTR BC 80y), and estimated prevalences of pathogenic variants in non-Finnish Europeans (see section 2.2.1) of breast cancer (BC) risk genes with moderate penetrance.

Gene	Locus	Protein function	LTR BC 80y	Prevalence	References
ATM	11q22.3	Phosphorylating cell cycle checkpoint kinase; regulator of a wide variety of downstream proteins, including tumor suppressor proteins TP53, BRCA1, CHEK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1	0.23–0.33	3.5×10^{-3}	[41, 45, 68, 107]
CHEK2	22q12.1	Required for checkpoint-mediated cell cycle arrest, activation of DNA repair and apoptosis in response to the presence of DNA double-strand breaks; interacts with BRCA1	0.22–0.29	6.7×10^{-3}	[1, 41, 45, 68, 146]
NF1	17q11.2	Negative regulator of the RAS signal transduction pathway; mutations are linked to neurofibromatosis type 1, juvenile myelomonocytic leukemia and Watson syndrome	0.17–0.35	8.5×10^{-4}	[1, 174, 179]
PALB2	16p12.2	Acts as a molecular scaffold that links BRCA1 and BRCA2 to form a complex that is essential for homologous recombination repair	0.42–0.45	1.1×10^{-3}	[1, 13, 41, 45, 68]
STK11	9p13.3	Serine-threonine protein kinase, participating in membrane bonding and apoptosis and negative regulator of the mTOR pathway; germline mutations cause Peutz-Jeghers syndrome	0.37	rare ($< 10^{-5}$)*	[1, 136]
General Population			0.12–0.15		[2, 111, 139]

*No pathogenic variants identified in ExAC Non-Finnish Europeans, estimator is based on the estimated incidence of Peutz-Jeghers syndrome [89].

Table 2.4: Characterization, asscoiated lifetime risks until the age of 80 years (LTR BC 80y), and estimated prevalences of pathogenic variants in non-Finnish Europeans (see section 2.2.1) of breast cancer (BC) risk genes with low penetrance and suspected BC risk genes.

Gene	Locus	Protein function	LTR BC 80y	Prevalence	References
BARD1	2q35	Interacts with the N-terminal region of BRCA1; plays a central role in the control of the cell cycle in response to DNA damage	not increased / moderately increased / 0.23	7.7×10^{-4}	[1, 35, 41, 86, 103, 153]
BRIP1	17q23.2	Involved in the repair of DNA double-strand breaks by homologous recombination in a bound complex with BRCA1	not increased / moderately increased	1.5×10^{-3}	[35, 44]
FANCM	14q21.2	DNA-dependent ATPase component required to anchor the multi-subunit Fanconi anemia core complex to chromatin after DNA damage	moderately increased*	1.7×10^{-3}	[77]
NBN	8q21.3	Member of the MRE11/RAD50 double-strand break repair complex	not increased / 0.23–0.33	1.4×10^{-3}	[1, 35, 45]
RAD51C	17q22	Involved in the homologous recombination and repair of DNA, interact with other RAD51 paralogs, reported to be important for Holliday junction resolution	not increased / 0.21–0.22	1.3×10^{-3}	[41, 86, 189]
RAD51D	17q12	Complex formed with other members of the RAD51 protein family catalyzes homologous pairing between single- and double-stranded DNA, and is thought to play a role in recombinational repair of DNA	not increased / 0.20	3.7×10^{-4}	[35, 41, 86, 189]
General Population			0.12–0.15		[2, 111, 139]

*considering the c.5101C>T nonsense mutation in Finnish individuals [77]

2.2.2 Hereditary ovarian cancer

OC is the eighth most common cancer occurring in women and the second most common cause of gynaecological-related cancer death worldwide [24]. In 2018, almost 3×10^5 women worldwide received a diagnosis of OC; in Germany, the annual number of diagnoses is approximately 7700, making OC approximately ten times less frequent than BC [24, 139]. The majority of OC cases originate in the epithelial tissue, and initially causes only unspecific symptoms, which is why the diagnosis is typically made only at an advanced, metastatic stage. Epithelial OC is currently divided into five main subtypes: high grade serous (≈70%), endometrioid (≈10%), clear cell (≈10%), low grade serous (<5%), and mucinous OC (<5%) [66]. With respect to histopathology, epidemiological and genetic risk factors, response to chemotherapy, and prognosis, these subtypes can be considered as distinct diseases [177].

Besides reproductive, demographic and lifestyle factors, an OC family history is assumed to be the strongest risk factor for the development of OC. Approximately 15 to 25% of unselected OC cases can be attributed to a mutation in a known OC risk gene [123,178]. *BRCA1/2*-associated disease is the most common form of hereditary OC. Estimated cumulative risks for OC development until the age of 70 years range from 0.35 to 0.59 for *BRCA1* germline mutation carriers and from 0.11 to 0.17 for *BRCA2* germline mutation carriers, respectively [14, 108]. Furthermore, germline mutations causing cancer predisposition syndromes, namely in *STK11* (Peutz-Jeghers syndrome) and *MLH1*, *MSH2* and *MSH6* (Lynch syndrome), are associated with relative risks for OC >4, as well as mutations in *RAD51C* and *RAD51D*. See Table 2.5 for a summarizing overview of confirmed OC risk genes with high to moderate penetrance. Reported prevalences of pathogenic variants in Lynch genes were estimated based on Non-Finnish Europeans under exclusion of TCGA data in ExAC [92] as described in section 2.2.1, considering RefSeq transcripts NM_000249 (*MLH1*), NM_000251 (*MSH2*), and NM_000179 (*MSH6*).

Similar to hereditary BC, the search for further risk genes causing OC is ongoing. Their determination is complicated mainly by the fact that OC occurs ten times rarer than BC, which means that the sample sizes used in gene association studies are often too small to achieve sufficient statistical power. In addition, different OC subtypes are typically pooled in these studies (in order to avoid unnecessarily reducing the sample size), although it can not be excluded that they have different genetic backgrounds [63], e.g., germline muations in *STK11* are primarily associated with non-epthelial OC [38]. Finally, associations for BC-only risk genes may arise due to individuals in the sample, who, in addition to OC, also display a family history of BC [115]. An overview of confirmed OC risk genes with low penetrance and suspected OC risk genes is given in Table 2.6.

Table 2.5: Characterization, associated lifetime risks until the age of 80 years (LTR OC 80y) and estimated prevalences of pathogenic variants in non-Finnish Europeans (see section 2.2.1) of ovarian cancer (OC) risk genes with high to moderate penetrance.

Gene	Locus	Protein function	LTR OC 80y	Prevalence	References
BRCA1	17q21.31	Tumor suppression by maintaining genomic stability by mediating transcription, DNA repair of double-stranded breaks, and recombination	0.44–0.60	3.2×10^{-3}	[1, 83, 86, 123]
BRCA2	13q13.1	Involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair	0.17–0.30	3.9×10^{-3}	[1, 83, 86, 123]
BRIP1	17q23.2	Involved in the repair of DNA double-strand breaks by homologous recombination in a bound complex with BRCA1	0.05–0.10	1.5×10^{-3}	[1, 86, 98, 123, 133]
MLH1	3p22.2	Components of the DNA mismatch repair system that	0.05-0.20*	2.6×10^{-4}	[1, 20, 47, 117]
MSH2	2p21-p16.3	recognizes and repairs errors that arise during DNA	0.11–0.24*	1.6×10^{-3}	[1, 20, 47, 117]
MSH6	2p16.3	replication, germline mutations cause a substantial risk of several cancers (Lynch syndrome)	0.01–0.14†	1.3×10^{-3}	[1, 20, 47, 117]
RAD51C	17q22	Involved in the homologous recombination and repair of DNA, interact with other RAD51 paralogs, reported to be important for Holliday junction resolution	0.05–0.11	1.3×10^{-3}	[1, 102, 189]
RAD51D	17q12	Complex formed with other members of the RAD51 protein family catalyzes homologous pairing between single- and double-stranded DNA, and is thought to play a role in recombinational repair of DNA	0.10–0.20	3.7×10^{-4}	[1, 101, 189]
General Population			0.01–0.02		[2, 139, 169]

*upper range bound estimated by Bonadona et al. [20] for the age of 70 years
†lower range bound estimated by Bonadona et al. [20] for the age of 70 years

Table 2.6: Characterization, associated lifetime risks until the age of 80 years (LTR OC 80y) and estimated prevalences of protein-truncating variants in non-Finnish Europeans (see section 2.2.1) of ovarian cancer (OC) risk genes with low penetrance and suspected OC risk genes.

Gene	Locus	Protein function	LTR OC 80y	Prevalence	References
ATM	11q22.3	Phosphorylating cell cycle checkpoint kinase; regulator of a wide variety of downstream proteins, including tumor suppressor proteins TP53, BRCA1, CHEK2, checkpoint proteins RAD17 and RAD9, and DNA repair protein NBS1	<0.05	3.5×10^{-3}	[86, 98]
BARD1	2q35	Interacts with the N-terminal region of BRCA1; plays a central role in the control of the cell cycle in response to DNA damage	uncertain	7.7×10^{-4}	[9, 133, 178]
NBN	8q21.3	Member of the MRE11/RAD50 double-strand break repair complex	uncertain	1.4×10^{-3}	[86, 133]
PALB2	16p12.2	Acts as a molecular scaffold that links BRCA1 and BRCA2 to form a complex that is essential for homologous recombination repair	uncertain / 0.05	1.1×10^{-3}	[133, 188]
STK11	9p13.3	Serine-threonine protein kinase, participating in membrane bonding and apoptosis and negative regulator of the mTOR pathway; germline mutations cause Peutz-Jeghers syndrome	increased	rare ($< 10^{-5}$)*	[86, 136]
TP53	17p13.1	Tumor supression, mediating cellular stress response; germline mutations cause Li-Fraumeni syndrome 1	uncertain / approx. 0.03	7.7×10^{-4}	[106, 170]
General Population			0.01–0.02		[2, 139, 169]

*No pathogenic variants identified in ExAC Non-Finnish Europeans, estimator is based on the estimated incidence of Peutz-Jeghers syndrome [89].

2.2.3 Genetic counseling of women at risk for familial breast and/or ovarian cancer

Genetic testing has become an integral part of the clinical care of BC/OC patients and their families since the identification of *BRCA1/2* as highly penetrant BC/OC risk genes in the 90s.

The inclusion criteria for diagnostic germline testing of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC), a multicenter consortium of interdisciplinary university centers specialized in providing counseling, genetic testing and preventive measures for persons at risk for familial BC/OC, are shown in Table 2.7.

Nowadays, advances in targeted NGS afford testing for further hereditary predisposition genes than *BRCA1/2*, and multi-gene panel approaches have become the standard in labs providing clincial management of persons with a familial BC/OC burden. In the Center for Familial Breast and Ovarian Cancer of the Uninversity Hospital of Cologne, the TruRisk® gene panel is applied for germline testing. The hybridization-based capture sequencing panel covers coding exons and multiple introns and non-coding exons of all confirmed BC/OC risk genes, as well as coding exons of additional genes whose association with BC/OC is under investigation (research genes). The composition of the TruRisk® gene panel is regularly adapted to the current state of research (see Table 2.8 for a densed overview of variations between the different versions), e.g., since TruRisk® v3, the panel also covers 306 loci of the BRIDGES PRS SNP set implemented in the CanRisk tool [15] for BC risk prediction.

However, the assessment of individual BC/OC risk is not based solely on genetic tests, but also includes family history, lifestyle factors and ethnicity [186]. Clinical management of individuals found to be at high risk for BC focuses on risk reduction and early diagnosis of cancer, and includes screening with annual mammography or breast magnetic resonance imaging (MRI), or risk-reducing mastectomy. Chemoprevention, more specifically treatment with tamoxifen in unaffected *BRCA1/2* germline mutations carriers [78], is under discussion. Concerning OC, regular screening for tumors at early stage, e.g., via transvaginal ultrasound, is not proven to decrease mortality, so women at high risk are recommended to undergo bilateral risk-reducing salpingo-oophorectomy around the age of 40 and after completion of childbearing, or earlier, dependent on the earliest age of OC diagnosed in their families [6, 11, 38].

Clinical decision-making whether preventive measures are offered or not is not necessarily consistent, and corresponding guidelines are subject to continuous revision. For example, the UK National Health and Care Excellence (NICE) guidelines generally consider an annual mammography for women with an estimated LTR >30%, starting at the age of 30 years [100, 119], whereas the US National Comprehensive Cancer Network (NCCN) recommends an annual mammography for patients with a LTR >20% starting earliest at the age of 30 years or at an age that is 10 years younger than the age at the earliest BC diagnosis in the family, whichever is later [17]. In Germany, intensified surveillance programs including magnetic resonance imaging of the breast are recommended for patients with a pathogenic *BRCA1/2* mutation and patients with a residual lifetime risk for BC of ≥30% [6]. In any case, before deciding on preventive measures,

Table 2.7: Inclusion criteria of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) for diagnostic germline testing, as of February 2021 [56]. At least one criterion has to be fullfilled within the family.

≥3 women with BC
≥2 women with BC, 1 with onset below 51 years of age
≥1 woman with BC and ≥1 woman with OC
≥2 women with OC
≥1 woman with OC before the age of 80
≥1 woman with BC before the age of 36
≥1 woman with bilateral BC with onset before the age of 51
≥1 woman with triple-negative BC before the age of 50
≥1 male with BC and 1 woman with BC or OC

Table 2.8: TruRisk® v1–v3 multi-gene panel characteristics, including a description of the genomic regions covered, number of sequencing targets, overall size of targeted regions and mean size of targeted regions (MST).

	Description	No. of Targets	Size [kbp]	MST [kbp] (min–max)
v1	coding exons of 10 BC/OC risk genes (*ATM*, *BRCA1/2*, *BRIP1*, *CDH1*, *CHEK2*, *PALB2*, *RAD51C/D*, *TP53*), 6 syndrome-associated (Lynch, Cowden and Peutz-Jehghers) and 17 research genes	576	172	0.30 (0.09–5.04)
v2	coding exons of 10 BC/OC risk genes (*ATM*, *BRCA1/2*, *BRIP1*, *CDH1*, *CHEK2*, *PALB2*, *RAD51C/D*, *TP53*), 7 syndrome-associated (Lynch, Cowden and Peutz-Jehghers), and 17 research genes; 66 identifier SNPs	674	200	0.30 (0.12–5.04)
v3	coding exons of 11 BC/OC risk genes (*ATM*, *BARD1*, *BRCA1/2*, *BRIP1*, *CDH1*, *CHEK2*, *PALB2*, *RAD51C/D*, *TP53*), 7 syndrome-associated (Lynch, Cowden and Peutz-Jehghers), and 16 research genes; 331 PRS/identifer SNPs	1023	267	0.26 (0.12–5.04)

the risks of potential interventions, e.g., X-ray exposure through regular mammograms, the risks and long-term consequences of surgical interventions or increased psychological stress, must be carefully considered in relation to expected BC/OC risks.

In addition to predicting BC/OC risk for individuals with a familial burden, genetic germline testing is also gaining importance for those already diagnosed with BC, as it has been shown that *BRCA1/2* germline mutation carriers with triple-negative BC benefit more from certain, i.e., platinum-based, treatment regimes than noncarriers [25, 173]. A recently published study by Pohl-Rescigno and colleagues suggests that comparable effects may also be seen for hormone receptor-positive tumors [130].

3 Aims of the thesis

This thesis aims to assess and improve existing methods for variant detection and variant classification using multi-gene panels in the framework of genetic counseling for persons at risk for hereditary BC/OC. There is a threefold objective, namely (1) optimization of multi-gene panel design, (2) comprehensive variant detection beyond SNVs and short indels, and (3) more accurate personal risk prediction based on improved variant interpretation.

3.1 Improvement of multi-gene panel design

As in 70% of all index cases of familial BC/OC no pathogenic variants can be identified in known moderate- to high-penetrant risk genes [65, 74], the search for further predisposing genetic modificators is ongoing. Multi-gene panels may include genes with insufficient evidence and/or unclear association in addition to confirmed BC/OC risk genes. However, the extent of regions covered by a multi-gene panel in a routine diagnostic setting should be as limited as possible with regard to the resources required in terms of sequencing capacity, computing time and storage space. In addition, associations with BC/OC of putative pathogenic variants in such genes, and hence, the utility of preventive measures derived from them, are largely unknown. As a result, people seeking advice are told of the existence of corresponding variants, but no consistent guidelines exist for further clinical managment. The achievement of sufficient statistical evidence for the detection of risk genes with moderate penetrance, which may even rarely be affected by pathogenic mutations, requires large sample sizes as achieved in routine genetic diagnostics. Here, the potential association of pathogenic variants in suspected risk genes *FANCM*, *BARD1*, and *BRIP1* with hereditary BC and/or OC was assessed in case-control studies including well-characterized index patients and geographically matched female controls.

In addition, in collaboration with Julika Borde, I examined the utility of PRSs in personalized BC risk prediction. Established sets of BC susceptibility loci used for PRS calculations in female *BRCA1/2* mutation carriers [84], as well as in women unselected for *BRCA1/2* germline mutation status [109], each comprise less than 100 loci, so that these PRSs have the potential to be straightforwardly implemented in routine diagnostic multi-gene panel analyses. However, the performance of PRSs in GWAS-independent clinical cohorts is poorly studied for moderate penetrant risk genes and thus the clinical implementation of PRSs is pending. Here, the performance of PRSs was assessed in a GWAS-independent clinical cohort of females carrying a pathogenic alteration in *CHEK2*, which is the third most frequently mutated BC risk gene in many European

countries [99].

3.2 Improvement of variant detection

With the spread of NGS in clinical diagnostics, especially multi-gene panels, commercial tools have become established that provide easy-to-use, automated solutions for NGS data processing. These tools, e.g., SOPHiA DDM® (SOPHiA GENETICS, Lausanne, Switzerland), NextGENe® (SoftGenetics, State College, PA, USA), or JSI SeqPilot software (Kippenheim, Germany), embed SNV and indel calling in an automated workflow, and provide starting from raw sequencing reads listings of variants of putative relevance in attractive table formats. Due to the typically high read depth in targeted regions, SNVs and indels can usually be predicted reliably from multi-gene panel data. Performance of out-of-the-box solutions for the prediction of CNVs, or SVs, which requires adapted computational methods, however, remains questionable [118, 131, 190]. Therefore, reliable detection of CNVs requires long-read sequencing [70] or additional wet lab approaches such as array comparative genomic hybridization (aCGH) or multiplex ligation-dependent probe amplification (MLPA) [32, 60, 75, 148]. As these additional analyses are costly and time-consuming, they are usually applied solely to genes known to be frequently affected by CNVs, i.e., *BRCA1/2* in the context of hereditary BC/OC [33,72,144], and the prevalence of CNVs in non-*BRCA1/2* cancer predisposition genes is poorly studied. A joint work with Louisa Lepkes therefore was aimed at two objectives: First, to evaluate the utility of *in silico* CNV detection approaches adapted to multi-gene panels in clinical diagnostics, and second, to assess the prevalence of CNVs in cancer predisposition genes in individuals at risk for familial BC/OC.

3.3 Improvement of variant interpretation

The third part of the thesis deals with considerations regarding the interpretation of genetic testing outcomes in the course of genetic counseling for hereditary BC/OC. In particular, the interpretation of missense mutations is a specially challenging task, as their impact to protein function are difficult to predict and they can therefore often only be classified as variant of uncertain significance (VUS). Thus, genetic testing results in uncertainty about the mutation status of corresponding risk genes, and advice seekers can neither be offered prophylactic measures nor relief. Reliable classifcation is further hindered by the fact that VUS are often extremely rare variants. As of September 2016, 64.4% of missense VUS in *BRCA1/2* reported in the patient registry of the GC-HBOC, were private, i.e., observed exclusively in one index patient. To circumvent the problem of missing information on rare genetic variants, the use of *in silico* approaches for automated variant classification has become established in many laboratories. However, studies revealed a diverse picture of the performance of these applications [59, 76, 93, 104, 141, 161]. Due to the known specific weaknesses of each prediction tool, a common strategy in clinical practice is to combine the results of various

approaches, e.g., assuming a disease-causing mutation when at least half of several approaches classify a variant as damaging. Therefore, I aimed to evaluate the performance of four *in silico* prediction tools embedded in the widely-used, commercial Alamut™ Visual software (Interactive Biosoftware, Rouen, France), namely Align-GVGD [163], SIFT [85], MutationTaster2 [149], and PolyPhen-2 [8], and to study how combinations of their results may influence the reliability of *in silico* predictions compared to standalone usage.

Furthermore, in a joint work with Dr. Jan Hauke, I aimed to investigate the implications of shifts in observed VFs between paired blood and tumor-derived samples for variant classification of rare *BRCA1/2* germline variants. Tumor development caused by heterozygous germline mutations is generally suggested to be initiated by inactivation of the wild-type allele in the corresponding risk gene, i.e., loss of heterozygosity (LOH) due to somatic mutations, deletion of the wild-type allele, or promotor methylation [80]. Loss of the wild-type allele is supposed to be indicated by signifcantly increased VFs in the tumor- vs blood-derived DNA. Thus, the potential benefit of VF comparison between blood- and tumor-derived samples for variant classification needs to be examined.

Pathogenic germline variants in *TP53*, which are mostly missense mutations, cause LFS1, and are therefore accompanied by extensive medical consequences [51]. However, *TP53* variants may arise *de novo*, either as *somatic mosaicism*, such that multiple tissues exhibit the variant at various level, or represent *clonal haematopoiesis* (CH), i.e., deleterious mutations affecting the hematopoietic stem and progenitor cells exclusively [16]. Occurence of CH is associated with increased age, tobacco use and exposure to radiotherapy [51]. Furthermore, CH in the *TP53* and *PPM1D* genes has been reported to be associated with prior exposure to chemotherapy [34]. Differentiation between germline variants, somatic mosaicism and CH is essential, as CH events in *TP53* cause increased risks of hematologic cancer, but do not cause LFS1 [55,116]. In collaboration with Konstantin Weber-Lassalle, I investigated pairwise blood- and tumor-derived DNA samples of OC patients with the aim to prove the existence of pathogenic variants in *TP53* and *PPM1D* in blood cells arising from CH rather than from germline inheritance, and to evaluate the frequency of CH occurences in dependence to the exposure to chemotherapy.

4 List of publications

4.1 Main publications

4.1.1 Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics.

Ernst et al. (2018) BMC Medical Genomics, 11(1):35
doi:10.1186/s12920-018-0353-y

4.1.1.1 Abstract (excerpt from the original publication)[1]

Background: The use of next-generation sequencing approaches in clinical diagnostics has led to a tremendous increase in data and a vast number of variants of uncertain significance that require interpretation. Therefore, prediction of the effects of missense mutations using *in silico* tools has become a frequently used approach. Aim of this study was to assess the reliability of *in silico* prediction as a basis for clinical decision making in the context of hereditary breast and/or ovarian cancer.
Methods: We tested the performance of four prediction tools (Align-GVGD, SIFT, PolyPhen-2, MutationTaster2) using a set of 236 *BRCA1/2* missense variants that had previously been classified by expert committees. However, a major pitfall in the creation of a reliable evaluation set for our purpose is the generally accepted classification of *BRCA1/2* missense variants using the multifactorial likelihood model, which is partially based on Align-GVGD results. To overcome this drawback we identified 161 variants whose classification is independent of any previous *in silico* prediction. In addition to the performance as stand-alone tools we examined the sensitivity, specificity, accuracy and Matthews correlation coefficient (MCC) of combined approaches.
Results: PolyPhen-2 achieved the lowest sensitivity (0.67), specificity (0.67), accuracy (0.67) and MCC (0.39). Align-GVGD achieved the highest values of specificity (0.92), accuracy (0.92) and MCC (0.73), but was outperformed regarding its sensitivity (0.90) by SIFT (1.00) and MutationTaster2 (1.00). All tools suffered from poor specificities, resulting in an unacceptable proportion of false positive results in a clinical setting. This shortcoming could not be bypassed by combination of these tools. In the best case scenario, 138 families would be affected by the misclassification of neutralvariants within the cohort of patients of the German Consortium for Hereditary Breast and Ovarian Cancer.
Conclusion: We show that due to low specificities state-of-the-art *in silico* prediction tools are not suitable to predict pathogenicity of variants of uncertain significance in *BRCA1/2*. Thus, clinical consequences should never be based solely on *in silico* forecasts.

However, our data suggests that SIFT and MutationTaster2 could be suitable to predict benignity, as both tools did not result in false negative predictions in our analysis.

4.1.1.2 Own contributions

I developed the statistical analysis approach in collaboration with Prof. Dr. M. Nothnagel (Department of Statistical Genetics and Bioinformatics, Cologne Center for Genomics (CCG), University of Cologne). I ran all *in silico* analyses and developed the approach for gaining an Align-GVGD-independent sample set. I conceptualized and prepared data visualization. I authored the inital manuscript draft, and I revised and approved the final version.

4.1.1.3 Contribution of co-authors

Prof. Dr. R. Schmutzler and PD Dr. E. Hahnen (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) conceptualized the study. PD Dr. C. Engel (Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig) and Dr. J. Hauke (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) collected and curated the input samples and defined the set of invesitgated VUS. Prof. Dr. M. Nothnagel (Department of Statistical Genetics and Bioinformatics, Cologne Center for Genomics (CCG), University of Cologne) was involved in the development of the the statistical analysis approach. J. Weber (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) derived the results for combined *in silico* approaches. Prof. Dr. R. Schmutzler, PD Dr. E. Hahnen and Dr. J. Hauke supported the preparation of the manuscript. All authors revised the manuscript and gave their approval to the final version.

4.1.2 Performance of Breast Cancer Polygenic Risk Scores in 760 Female CHEK2 Germline Mutation Carriers

Borde et al. (2020) Journal of the National Cancer Institute
doi:10.1093/jnci/djaa203

4.1.2.1 Abstract (excerpt from the original publication)

BACKGROUND: Genome-wide association studies (GWAS) suggest that the combined effects of breast cancer (BC)-associated single nucleotide polymorphisms (SNPs)can improve BC risk stratification using polygenic risk scores (PRSs). The performance of PRSs in GWAS-independent clinical cohorts is poorly studied in individuals carrying mutations in moderately penetrant BC predisposition genes such as *CHEK2*. **METHODS:** 760 female CHEK2mutation carriers were included; 561 women were affected with BC, of whom 74 developed metachronous contralateral BC (mCBC). For PRS calculations, two SNP sets covering 77 (SNP set 1, developed for BC risk stratification in women unselected for their *BRCA1/2* germline mutation status) and 88 (SNP set 2, developed for BC risk stratification in female *BRCA1/2* mutation carriers) BC-associated SNPs were used. **RESULTS:** Both SNP sets provided concordant PRS results at the individual level ($r = 0.91$, $p < 2.20 \times 10^{-16}$). Weighted cohort Cox regression analyses revealed significant associationsof PRSs with the risk for first BC. For SNP set1, a hazard ratio (HR) of 1.71 per standard deviation of the PRS was observed (95%confidence interval [CI]:1.36 to 2.15, $p = 3.87 \times 10^{-6}$). PRSs identify a subgroup of *CHEK2* mutation carriers with a predicted lifetime risk for first BC that exceeds the surveillance thresholds defined by international guidelines. Association of PRS with mCBC was examined via Cox regression analysis (SNP set 1 HR: 1.23, 95%CI:0.86 to 1.78, p=0.26). **CONCLUSION:** PRSs may be used to personalize risk-adapted preventive measures for women with *CHEK2* mutations. Larger studies are required to assess the role of PRSs in mCBC predisposition.

4.1.2.2 Own contributions

I performed the analysis of the amplicon-based Fluidigm Access Array data, starting from demultiplexing and mapping up to variant calling and conversion to input data in TPED format. In collaboration with Dr. K. Kuchenbaecker and Dr. O. Giannakopoulou (Division of Psychiatry & UCL Genetics Institute, University College of London (UCL)) I developed and implemented the sample and variant quality filters. I performed all statistical analyses and established the nearest shrunken centroid classifier and inclusion of a third dimension in multidimensional scaling for ethnicity checks. In collaboration with Dr. K. Kuchenbaecker, I developed and implemented PRS computations, approximation of theoretically expected mean values $\overline{PRS}$, and the weighted cohort Cox regression approach for analysis of the association of standardized PRSs with BC risk. I examined the proportional hazard assumption in age-stratified Cox regression models, computed age-specific HRs per SD of the PRS, and performed Cox regression for the evaluation of the association of PRSs and age at first BC diagnosis with mCBC. I conceptualized

and prepared data visualization. I contributed to the manuscript draft, and revised and approved the final version.

4.1.2.3 Contribution of co-authors

J. Borde (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) defined, provided and curated the input sample including the ascertainment of germline mutation status, the retrievement of missing phenotype data, and implemented and applied several consistency checks on patient data. J. Borde designed and applied customized target enrichment employing a 48.48 amplicon-based panel Access Array system (Fluidigm, San Francisco, CA, USA) and performed the entire DNA sequencing. J. Borde extracted data from public databases dbSNP and LDlink. J. Borde and I prepared the first manuscript draft in close collaboration. Dr. B. Wappenschmidt (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne), Dr. D. Niederacher (Department of Gynecology & Obstetrics, University Hospital Duesseldorf, Heinrich-Heine University Duesseldorf), Dr. G. Schmidt (Institute of Human Genetics, Hannover Medical School), PD Dr. A. Quante (Department of Gynecology and Obstetrics, Technical University Munich), Dr. J.Horvárth (Institute for Human Genetics, University Hospital Muenster), Prof. Dr. N. Arnold (Institute of Clinical Molecular Biology, Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Christian-Albrechts University Kiel), Dr. A. Rump (Institute of Clinical Genetics, Technische Universitaet Dresden), Dr. A. Gehrig (Institute of Human Genetics, Julius-Maximilians-Universität Würzburg), Dr. J. Hentschel (Institute of Human Genetics, University of Leipzig Hospitals and Clinics), Dr. U. Faust (Institute of Medical Genetics and Applied Genomics, University Hospital Tuebingen), V. Dutrannoy (Institute of Medical and Human Genetics, Charité Universitätsmedizin Berlin), Prof. Dr. A. Meindl (Department of Gynecology and Obstetrics, Ludwig-Maximilians-University Munich), M. Kuzyakova (Institute of Human Genetics, University Medical Center, Georg August University of Göttingen), Dr. S. Wang-Gohrke (Department of Gynaecology and Obstetrics, University Hospital Ulm), Prof. Dr. B. Weber (Institute of Human Genetics, University Hospital Regensburg), Dr. C. Sutter (Institute of Human Genetics, University of Heidelberg), PD Dr. A. Volk (Institute of Human Genetics, University Medical Center Hamburg-Eppendorf), and PD Dr. C. Engel (Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig) provided samples and resources. K. Weber-Lassalle, Dr. E. Pohl-Rescigno, and Dr. N. Weber-Lassalle (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) were involved in design and establishment of the customized panel sequencing. Dr. K. Kuchenbaecker (Division of Psychiatry & UCL Genetics Institute, University College of London (UCL)) supervised statistical analyses. Dr. O. Giannakopoulou (Division of Psychiatry & UCL Genetics Institute, University College of London (UCL)) provided support for analyses involving PLINK. Dr. M. Schmidt (Division of Molecular Pathology, Netherlands Cancer Institute) and Prof. Dr. A. Antoniou (Department of Public Health and Primary Care, Centre for Can-

cer Genetic Epidemiology, University of Cambridge) gave further advice regarding the analysis strategy. Dr. A. Lee (Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge) provided population-based incidence rates of *CHEK2* mutation carriers. The study was supervised and conceptualized by PD Dr. E. Hahnen (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne), Prof. Dr. R. Schmutzler (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne), and Dr. K. Kuchenbaecker. The writing team consisted of PD Dr. E. Hahnen, J. Borde and myself. All authors revised and approved the final manuscript.

4.1.3 Performance of In Silico Prediction Tools for the Detection of Germline Copy Number Variations in Cancer Predisposition Genes in 4208 Female Index Patients with Familial Breast and Ovarian Cancer

Lepkes et al. (2021) Cancers, 13(1):118
doi:10.3390/cancers13010118

4.1.3.1 Abstract (excerpt from the original publication)[2]

The identification of germline copy number variants (CNVs) by targeted next-generation sequencing (NGS) frequently relies on *in silico* CNV prediction tools with unknown sensitivities. We investigated the performances of four *in silico* CNV prediction tools, including one commercial (Sophia Genetics DDM) and three non-commercial tools (ExomeDepth, GATK gCNV, panelcn.MOPS) in 17 cancer predisposition genes in 4,208 female index patients with familial breast and/or ovarian cancer (BC/OC). CNV predictions were verified via multiplex ligation-dependent probe amplification. We identified 77 CNVs in 76 out of 4,208 patients (1.81%); 33 CNVs were identified in genes other than *BRCA1/2*, mostly in *ATM*, *CHEK2*, and *RAD51C* and less frequently in *BARD1*, *MLH1*, *MSH2*, *PALB2*, *PMS2*, *RAD51D*, and *TP53*. The Sophia Genetics DDM software showed the highest sensitivity; six CNVs were missed by at least one of the non-commercial tools. The positive predictive values ranged from 5.9% (74/1,249) for panelcn.MOPS to 79.1% (72/91) for ExomeDepth. Verification of *in silico* predicted CNVs is required due to high frequencies of false positive predictions, particularly affecting target regions at the extremes of the GC content or target length distributions. CNV detection should not be restricted to *BRCA1/2* due to the relevant proportion of CNVs in further BC/OC predisposition genes.

4.1.3.2 Own contributions

I was invovled in the conceptualization of the study by introducing the evaluation of non-commercial CNV detection tools and the evaluation of performance dependent on sequencing target characteristics. I extracted and processed the hybridization-based capture input data and ran CNV calling with ExomeDepth, GATK gCNV and panelcn.MOPs. I evaluated in collaboration with J. Weber (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) the CNV prediction performance. I extracted the sequencing target characteristics and did all statistical analyses. I wrote the first manuscript draft, revised and approved the final manuscript, and served as submitting and corresponding author.

4.1.3.3 Contribution of co-authors

L. Lepkes (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) provided and curated in collaboration with Dr. B. Blümcke (Center for Familial Breast & Ovarian Cancer, Center for

[2]License: `https://creativecommons.org/licenses/by/4.0/`

Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) the sample set, and extracted the results of the Sophia Genetics DDM (SOPHiA DDM®) software. L. Lepkes, M. Kayali, and S. Schmidt (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) ran MLPA analyses. J. Weber (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) was involved in evaluation of CNV prediction performance. MLPA analyses for *BARD1* were run by Dr. M. Suszynska and Dr. K. Klonowska under supervision of Prof. Dr. P. Kozlowski (Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan). Dr. B. Wappenschmidt (Center for Familial Breast & Ovarian Cancer, Center for Integated Oncology (CIO), Medical Faculty, University Hospital Cologne) provided samples. Prof. Dr. P. Kozlowski and Prof. Dr. R. Schmutzler provided resources. The study was supervised by PD Dr. E. Hahnen and Prof. Dr. R. Schmutzler. All authors revised and approved the final manuscript.

4.2 Additional co-authored publications

4.2.1 Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer

Neidhardt et al. (2017) JAMA Oncology, 3(9):1245-1248
doi:10.1001/jamaoncol.2016.5592

4.2.1.1 Abstract (excerpt from the original publication)

Importance Germline mutations in established moderately or highly penetrant risk genes for breast cancer (BC) and/or ovarian cancer (OC), including *BRCA1* and *BRCA2*, explain fewer than half of all familial BC and/or OC cases. Based on the genotyping of 2 loss-of-function (LoF) variants c.5101C>T (p.GIn1701Ter [rs147021911]) and c.5791C>T (p.Arg1931Ter [rs144567652]), the *FANCM* gene has been suggested as a novel BC predisposition gene, while the analysis of the entire coding region of the *FANCM* gene in familial index cases and geographically matched controls is pending.
Objectives To assess the mutational spectrum within the *FANCM* gene, and to determine a potential association of LoF germline mutations within the *FANCM* gene with BC and/or OC risk.
Design, Setting, and Participants For the purpose of identification and characterization of novel BC and/or OC predisposition genes, a total of 2047 well-characterized familial BC index cases, 628 OC cases, and 2187 geographically matched controls were screened for LoF mutations within the *FANCM* gene by next-generation sequencing. All patients previously tested negative for pathogenic *BRCA1* and *BRCA2* mutations. All data collection occurred between June 1, 2013, and April 30, 2016. Data analysis was performed from May 1, 2016, to July 1, 2016.
Main Outcomes and Measures *FANCM* LoF mutation frequencies in patients with BC and/or OC were compared with the *FANCM* LoF mutation frequencies in geographically matched controls by univariate logistic regression. Positive associations were stratified by age at onset and cancer family history.
Results In this case-control study, 2047 well-characterized familial female BC index cases, 628 OC cases, and 2187 geographically matched controls were screened for truncating *FANCM* alterations. Heterozygous LoF mutations within the *FANCM* gene were significantly associated with familial BC risk, with an overall odds ratio (OR) of 2.05 (95% CI, 0.94–4.54; P=.049) and a mutation frequency of 1.03% in index cases. In familial patients whose BC onset was before age 51 years, an elevated OR of 2.44 (95% CI, 1.08–5.59; P=.02) was observed. A more pronounced association was identified for patients with a triple-negative BC tumor phenotype (OR, 3.75; 95% CI, 1.00–12.85; P=.02). No significant association was detected for unselected OC cases (OR, 1.74; 95% CI, 0.57–5.08; P=.27).
Conclusions and Relevance Based on the significant associations of heterozygous LoF mutations with early-onset or triple-negative BC, *FANCM* should be included in diagnostic gene panel testing for individual risk assessment. Larger studies are required to determine age-dependent disease risks for BC and to assess a potential role of *FANCM*

mutations in OC pathogenesis.

4.2.1.2 Own contributions

I extracted allele frequencies of observed variants from the ExAC Non-Finnish Europeans under exclusion of TCGA data [92] as input for variant classification, and revised and approved the final manuscript.

4.2.2 Gene panel testing of 5589 BRCA1/2-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer

Hauke et al. (2018) Cancer Medicine, 7(4):1349-1358
doi:10.1002/cam4.1376

4.2.2.1 Abstract (excerpt from the original publication)

The prevalence of germ line mutations in non-*BRCA1/2* genes associated with hereditary breast cancer (BC) is low, and the role of some of these genes in BC predisposition and pathogenesis is conflicting. In this study, 5589 consecutive BC index patients negative for pathogenic *BRCA1/2* mutations and 2189 female controls were screened for germ line mutations in eight cancer predisposition genes (*ATM*, *CDH1*, *CHEK2*, *NBN*, *PALB2*, *RAD 51C*, *RAD51D*, and *TP53*). All patients met the inclusion criteria of the German Consortium for Hereditary Breast and Ovarian Cancer for germ line testing. The highest mutation prevalence was observed in the *CHEK2* gene (2.5%), followed by *ATM* (1.5%) and *PALB2* (1.2%). The mutation prevalence in each of the remaining genes was 0.3% or lower. Using Exome Aggregation Consortium control data, we confirm significant associations of heterozygous germ line mutations with BC for *ATM* (OR: 3.63, 95%CI: 2.67–4.94), *CDH1* (OR: 17.04, 95%CI: 3.5482), *CHEK2* (OR: 2.93, 95%CI: 2.29–3.75), *PALB2* (OR: 9.53, 95%CI: 6.25–14.51), and *TP53* (OR: 7.30, 95%CI: 1.22–43.68). *NBN* germ line mutations were not significantly associated with BC risk (OR: 1.39, 95%CI: 0.73–2.64). Due to their low mutation prevalence, the *RAD51C* and *RAD51D* genes require further investigation. Compared with control datasets, predicted damaging rare missense variants were significantly more prevalent in *CHEK2* and *TP53* in BC index patients. Compared with the overall sample, only *TP53* mutation carriers show a significantly younger age at first BC diagnosis. We demonstrate a significant association of deleterious variants in the *CHEK2*, *PALB2*, and *TP53* genes with bilateral BC. Both, *ATM* and *CHEK2*, were negatively associated with triple-negative breast cancer (TNBC) and estrogen receptor (ER)-negative tumor phenotypes. A particularly high *CHEK2* mutation prevalence (5.2%) was observed in patients with human epidermal growth factor receptor 2 (HER2)-positive tumors.

4.2.2.2 Own contributions

I supported the conceptualisation of variant filters and statistical data evaluation. I applied VF, MAF (based on ExAC Non-Finnish Europeans under exclusion of TCGA data [92]) and *in silico* prediction filters [85, 149] to the variant set, and I revised and approved the final manuscript.

4.2.3 BRIP1 Loss-Of-Function Mutations Confer High Risk for Familial Ovarian Cancer, but Not Familial Breast Cancer

Weber-Lassalle et al. (2018) Breast Cancer Research, 20(1):7
doi:10.1186/s13058-018-0935-9

4.2.3.1 Abstract (excerpt from the original publication)[3]

Background: Germline mutations in the *BRIP1* gene have been described as conferring a moderate risk for ovarian cancer (OC), while the role of *BRIP1* in breast cancer (BC) pathogenesis remains controversial.
Methods: To assess the role of deleterious *BRIP1* germline mutations in BC/OC predisposition, 6341 well-characterized index patients with BC, 706 index patients with OC, and 2189 geographically matched female controls were screened for loss-of-function (LoF) mutations and potentially damaging missense variants. All index patients met the inclusion criteria of the German Consortium for Hereditary Breast and Ovarian Cancer for germline testing and tested negative for pathogenic *BRCA1/2* variants.
Results: *BRIP1* LoF mutations confer a high OC risk in familial index patients (odds ratio (OR) = 20.97, 95% confidence interval (CI) = 12.02–36.57, P < 0.0001) and in the subgroup of index patients with late-onset OC (OR = 29.91, 95% CI = 14.99–59.66, P < 0.0001). No significant association of *BRIP1* LoF mutations with familial BC was observed (OR = 1.81 95% CI = 1.00–3.30, P = 0.0623). In the subgroup of familial BC index patients without a family history of OC there was also no apparent association (OR = 1.42, 95% CI = 0.70–2.90, P = 0.3030). In 1027 familial BC index patients with a family history of OC, the *BRIP1* mutation prevalence was significantly higher than that observed in controls (OR = 3.59, 95% CI = 1.43–9.01; P = 0.0168). Based on the negative association between *BRIP1* LoF mutations and familial BC in the absence of an OC family history, we conclude that the elevated mutation prevalence in the latter cohort was driven by the occurrence of OC in these families. Compared with controls, predicted damaging rare missense variants were significantly more prevalent in OC (P = 0.0014) but not in BC (P = 0.0693) patients.
Conclusions: To avoid ambiguous results, studies aimed at assessing the impact of candidate predisposition gene mutations on BC risk might differentiate between BC index patients with an OC family history and those without. In familial cases, we suggest that BRIP1 is a high-risk gene for late-onset OC but not a BC predisposition gene, though minor effects cannot be excluded.

4.2.3.2 Own contributions

I supported the conceptualisation of variant filters and statistical data evaluation. I applied VF, MAF (based on ExAC Non-Finnish Europeans under exclusion of TCGA data [92]), last exon and *in silico* prediction filters [85, 149] to the variant set, and I revised and approved the final manuscript.

[3]License: `https://creativecommons.org/licenses/by/4.0/`

4.2.4 Diagnosis of Li-Fraumeni Syndrome: Differentiating TP53 germline mutations from clonal hematopoiesis: Results of the observational AGO-TR1 trial

Weber-Lassalle et al. (2018) Human Mutation, 39(12):2040-2046
doi:10.1002/humu.23653

4.2.4.1 Abstract (excerpt from the original publication)

The Li-Fraumeni cancer predisposition syndrome (LFS1) presents with a variety of tumor types and the *TP53* gene is covered by most diagnostic cancer gene panels. We demonstrate that deleterious *TP53* variants identified in blood-derived DNA of 523 patients with ovarian cancer (AGO-TR1 trial) were not causal for the patients' ovarian cancer in three out of six *TP53*-positive cases. In three out of six patients, deleterious *TP53* mutations were identified with low variant fractions in blood-derived DNA but not in the tumor of the patient seeking advice. The analysis of the *TP53* and *PPM1D* genes, both intimately involved in chemotherapy-induced and/or age-related clonal hematopoiesis (CH), in 523 patients and 1,053 age-matched female control individuals revealed that CH represents a frequent event following chemotherapy, affecting 26 of the 523 patients enrolled (5.0%). Considering that *TP53* mutations may arise from chemotherapy-induced CH, our findings help to avoid false-positive genetic diagnoses of LFS1.

4.2.4.2 Own contributions

I was involved in the development of variant filters and applied the VF and annotation filters for the hybridization capture-based NGS-derived variants. I analyzed the amplicon-based Fluidigm Access Array data, starting with demultiplexing and mapping up to variant calling. I developed and applied the variant filter criteria, including the determination and application of quality thresholds concerning MAFs, minimum base and mapping qualities, minimum read depth, and minimum counts of observed alternate nucleotides. I supported the conceptualistaion of statistical data evaluation. I contributed to the manuscript draft and revised and approved the final manuscript.

4.2.5 Deleterious somatic variants in 473 consecutive individuals with ovarian cancer: results of the observational AGO-TR1 study (NCT02222883)

Hauke et al. (2019) Journal of Medical Genetics, 56(9):574-580
doi:10.1136/jmedgenet-2018-105930

4.2.5.1 Abstract (excerpt from the original publication)

Background For individuals with ovarian cancer (OC), therapy options mainly depend on *BRCA1/2* germline status. What is the prevalence of deleterious somatic variants, that is, does genetic tumour testing identify subgroups of individuals who also might benefit from targeted therapy?
Methods Paired analysis of tumour-derived versus blood-derived DNA to determine the prevalence of deleterious somatic variants in OC predisposition genes (*ATM*, *BRCA1/2*, *BRIP1*, *MSH2/6*, *PALB2*, *RAD51C/D* and *TP53*) and the *PIK3CA* and *PTEN* genes in individuals with OC (AGO-TR1 study, NCT02222883). Results were complemented by *BRCA1*, *PALB2* and *RAD51C* promoter methylation analyses and stratified by histological subtype; 473 individuals were included.
Results The combined analyses revealed that deleterious germline variants in established OC predisposition genes (all: 125/473, 26.4%; *BRCA1/2*: 97/473, 20.5%), deleterious somatic variants in established OC predisposition genes excluding *TP53* (all: 39/473, 8.2%; *BRCA1/2*: 30/473, 6.3%) and promoter methylation (all: 67/473, 14.2%; *BRCA1*: 57/473, 12.1%; *RAD51C*: 10/473, 2.1%; *PALB2*: 0/473) were mutually exclusive, with a few exceptions. The same holds true for deleterious somatic *PIK3CA* and/or *PTEN* variants (33/473, 7.0%) found to be enriched in endometrioid and clear cell OC (16/35, 45.7%); 84.3% of the deleterious single-nucleotide/indel germline variants in established OC predisposition genes showed significantly higher variant fractions (VFs) in the tumour-derived versus blood-derived DNA, indicating a loss of the wild-type alleles.
Conclusion Tumour sequencing of the *BRCA1*, *BRCA2*, *PIK3CA* and *PTEN* genes along with *BRCA1* and *RAD51C* promoter methylation analyses identified large subgroups of germline mutation-negative individuals who may be addressed in interventional studies using PARP or PI3K/AKT/mTOR inhibitors.

4.2.5.2 Own contributions

I supported the conceptualisation of data analysis and statistical data evaluation. I extracted VFs and generated variant annotations. I revised and approved the final manuscript.

4.2.6 The GPRC5A frameshift variant c.183del is not associated with increased breast cancer risk in BRCA1 mutation carriers

Klaschik et al. (2019) International Journal of Cancer, 144(7):1761-1763
doi:10.1002/ijc.32016

4.2.6.1 Summary

The above mentioned manuscript is a letter to the editor with respect to a publication of Sokolenko and colleagues [154]. The authors reported the heterozygous *GPRC5A* c.183del variant as a modifier of BC risk in *BRCA1* germline mutation carriers. Although GPRC5A protein expression is dysregulated in mammary tumors in *BRCA1* germline mutation carriers, no association with tumour stage, lymph node status, histological grading or histological tumour type, or with overall and recurrence-free survival could be confirmed. Furthermore, no significant accumulation of the heterozygous *GPRC5A* c.183del variant in 1707 BC index cases with *BRCA1* haploinsufficiency (15/1707; carrier frequency 0.88%) compared to 3451 *BRCA1/2*-negative BC index patients (21/3451; carrier frequency 0.61%; OR = 1.45, 95%CI:0.75–2.82, p = 0.273) or 3308 geographically matched controls (26/3308; carrier frequency 0.79%; OR = 1.12, 95%CI:0.59–2.12, p = 0.730) was observed. CRISPR/Cas9-induced *GPRC5A* knockout using the MDA-MB-231 cell line, did not reveal any notable differences in *BRCA1* expression compared to that in the *GPRC5A* proficient cell clones, and did not affect BRCA1 protein levels. Moreover, investigation of proliferation (PCNA) and apoptosis (caspase3) markers did not reveal any differences between the *GPRC5A* knockout and *GPRC5A* wild type cell clones.

Therefore, it is suggested that *GPRC5A* does not function as a modifier of BC risk in *BRCA1* germline mutation carriers.

4.2.6.2 Own contributions

I contributed to curation of the input samples for case-control studies and preparation of figures. I revised and approved the final manuscript.

4.2.7 Germline loss-of-function variants in the BARD1 gene are associated with early-onset familial breast cancer but not ovarian cancer

Weber-Lassalle et al. (2019) Breast Cancer Research, 21(1):55
doi:10.1186/s13058-019-1137-9

4.2.7.1 Abstract (excerpt from the original publication)[4]

Background: The role of the *BARD1* gene in breast cancer (BC) and ovarian cancer (OC) predisposition remains elusive, as published case-control investigations have revealed controversial results. We aimed to assess the role of deleterious *BARD1* germline variants in BC/OC predisposition in a sample of 4920 *BRCA1/2*-negative female BC/OC index patients of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC).
Methods: A total of 4469 female index patients with BC, 451 index patients with OC, and 2767 geographically matched female control individuals were screened for loss-of-function (LoF) mutations and potentially damaging rare missense variants in *BARD1*. All patients met the inclusion criteria of the GC-HBOC for germline testing and reported at least one relative with BC or OC. Additional control datasets (Exome Aggregation Consortium, ExAC; Fabulous Ladies Over Seventy, FLOSSIES) were included for the calculation of odds ratios (ORs).
Results: We identified LoF variants in 23 of 4469 BC index patients (0.51%) and in 36 of 37,265 control individuals (0.10%), resulting in an OR of 5.35 (95% confidence interval [CI]=3.17–9.04; P<0.00001). *BARD1*-mutated BC index patients showed a significantly younger mean age at first diagnosis (AAD; 42.3years, range 24–60years) compared with the overall study sample (48.6years, range 17–92years; P=0.00347). In the subgroup of BC index patients with an AAD <40years, an OR of 12.04 (95% CI=5.78–25.08; P<0.00001) was observed. An OR of 7.43 (95% CI=4.26–12.98; P<0.00001) was observed when stratified for an AAD <50years. LoF variants in *BARD1* were not significantly associated with BC in the subgroup of index patients with an AAD ≥50years (OR=2.29; 95% CI=0.82–6.45; P=0.11217). Overall, rare and predicted damaging *BARD1* missense variants were significantly more prevalent in BC index patients compared with control individuals (OR=2.15; 95% CI=1.26–3.67; P=0.00723). Neither LoF variants nor predicted damaging rare missense variants in *BARD1* were identified in 451 familial index patients with OC.
Conclusions: Due to the significant association of germline LoF variants in *BARD1* with early-onset BC, we suggest that intensified BC surveillance programs should be offered to women carrying pathogenic *BARD1* gene variants.

4.2.7.2 Own contributions

I applied variant annotation, and allele frequency (based on ExAC Non-Finnish Europeans under exclusion of TCGA data [92]) and *in silico* prediction filters [85, 149] to the

[4] License: `https://creativecommons.org/licenses/by/4.0/`

variant set. I was involved in the conceptualisation of statistical data analysis, and I revised and approved the final manuscript.

4.2.8 Association of Germline Variant Status With Therapy Response in High-Risk Early-Stage Breast Cancer: A Secondary Analysis of the GeparOcto Randomized Clinial Trial

Pohl-Rescigno et al. (2020) JAMA Oncology, 6(5):744-748
doi:10.1001/jamaoncol.2020.0007

4.2.8.1 Abstract (excerpt from the original publication)

Importance The GeparOcto randomized clinical trial compared the efficacy of 2 neoadjuvant breast cancer (BC) treatment regimens: sequential intense dose-dense epirubicin, paclitaxel, and cyclophosphamide (iddEPC) vs weekly paclitaxel and nonpegylated liposomal doxorubicin (PM) in patients with different biological BC subtypes. Patients with triple-negative BC (TNBC) randomized to the PM arm received additional carboplatin (PMCb). Overall, no difference in pathologic complete response (pCR) rates was observed between study arms. It remained elusive whether the germline variant status of *BRCA1/2* and further BC predisposition genes are associated with treatment outcome.
Objective To determine treatment outcome for BC according to germline variant status.
Design, Setting, and Participants This retrospective biomarker study is a secondary analysis of the GeparOcto multicenter prospective randomized clinical trial conducted between December 2014 and June 2016. Genetic analyses assessing for variants in *BRCA1/2* and 16 other BC predisposition genes in 914 of 945 women were performed at the Center for Familial Breast and Ovarian Cancer, Cologne, Germany, from August 2017 through December 2018.
Main Outcomes and Measures Proportion of patients who achieved pCR (ypT0/is ypN0 definition) after neoadjuvant treatment according to germline variant status.
Results In the study sample of 914 women with different BC subtypes with a mean (range) age at BC diagnosis of 48 (21–76) years, overall higher pCR rates were observed in patients with *BRCA1/2* variants than in patients without (60.4% vs 46.7%; odds ratio [OR], 1.74; 95% CI, 1.13–2.68; P=.01); variants in non-*BRCA1/2* BC predisposition genes were not associated with therapy response. Patients with TNBC with *BRCA1/2* variants achieved highest pCR rates. In the TNBC subgroup, a positive *BRCA1/2* variant status was associated with therapy response in both the PMCb arm (74.3% vs 47.0% without *BRCA1/2* variant; OR, 3.26; 95% CI, 1.44–7.39; P=.005) and the iddEPC arm (64.7% vs 45.0%; OR, 2.24; 95% CI, 1.04–4.84; P=.04). A positive *BRCA1/2* variant status was also associated with elevated pCR rates in patients with *ERBB2*-negative, hormone receptor-positive BC (31.8% vs 11.9%; OR, 3.44; 95% CI, 1.22-9.72; P=.02).
Conclusions and Relevance Effective chemotherapy for *BRCA1/2*-mutated TNBC is commonly suggested to be platinum-based. With a pCR rate of 64.7%, iddEPC may also be effective in these patients, though further prospective studies are needed. The elevated pCR rate in *BRCA1/2*-mutated *ERBB2*-negative, hormone receptorpositive BC suggests that germline *BRCA1/2* testing should be considered prior to treatment start.

4.2.8.2 Own contributions

I carried out *in silico* CNV prediction, including configuration and run-wise execution of panelcn.MOPS [131], ExomeDepth [128] and OPaCNAT[5] with subsequent filtering for putative true positive predictions based on sequencing target characteristics. I extracted allele frequencies of observed variants from the ExAC Non-Finnish Europeans under exclusion of TCGA data [92] as input for variant classification, and revised and approved the final manuscript.

[5] https://bitbucket.org/CorinnaErnst/opacnat/

4.2.9 Sensitivity and specificity of loss of heterozygosity analysis for the classification of rare germline variants in BRCA1/2: Results of the observational AGO-TR1 study (NCT02222883)

Hauke et al. (2020) Journal of Medical Genetics
doi:10.1136/jmedgenet-2020-107353

4.2.9.1 Abstract (excerpt from the original publication)[6]

Variant-specific loss of heterozygosity (LOH) analyses may be useful to classify *BRCA1/2* germline variants of unknown significance (VUS). The sensitivity and specificity of this approach, however, remains unknown. We performed comparative next-generation sequencing analyses of the *BRCA1/2* genes using blood- and tumour-derived DNA of 488 patients with ovarian cancer enrolled in the observational AGO-TR1 trial (NCT02222883). Overall, 94 pathogenic, 90 benign, and 24 VUS were identified in the germline. A significantly increased variant fraction (VF) of a germline variant in the tumour indicates loss of the wild-type allele; a decreased VF indicates loss of the variant allele. We demonstrate that significantly increased VFs predict pathogenicity with high sensitivity (0.84, 95%CI: 0.77-0.91), poor specificity (0.63, 95%CI: 0.53-0.73) and poor positive predictive value (PPV; 0.71, 95%CI: 0.62-0.79). Significantly decreased VFs predict benignity with low sensitivity (0.26, 95%CI: 0.17-0.35), high specificity (1.0, 95%CI: 0.96-1.00) and PPV (1.0, 95%CI: 0.85-1.00). Variant classification based on significantly increased VFs results in an inacceptable proportion of false-positive results. A significantly decreased VF in the tumour may be exploited as a reliable predictor for benignity, with no false-negative result observed. When applying the latter approach, VUS identified in four patients can now be considered benign.

4.2.9.2 Own contributions

I extracted allele frequencies of observed variants from the ExAC Non-Finnish Europeans under exclusion of TCGA data [92] as input for variant classification and extracted VF data. I conceptualized and carried out statistical data analysis and data visualization. I contributed to the manuscript draft, and revised and approved the final manuscript.

5 Results

5.1 Breast and/or ovarian cancer risk genes

To assess the association of putative risk genes *FANCM*, *BARD1*, and *BRIP1* with BC and/or OC, case-control studies were employed [121, 182, 183]. Heterozygous PTVs within the *FANCM* gene were associated with BC in individuals with familial BC history (OR = 2.05, 95% confidence interval (CI): 0.94–4.54, Pearson's χ^2 test $p = 0.049$) [121]. The ORs were increased for the subgroups of individuals with age at primary BC diagnosis of <51 years (OR = 2.44, 95% CI: 1.08–5.59, $p = 0.02$), and with triple-negative BC tumors (OR = 3.75, 95% CI: 1.00–12.85, $p = 0.02$). No significant association between OC and the observation of PTVs in *FANCM* could be observed considering a sample of 628 OC cases.

Considering 706 *BRCA1/2* negative OC cases and 2189 controls, PTVs in *BRIP1* were signifcantly associated with OC (OR = 20.97, 95% CI: 12.02–36.57, two-sided Fisher's Exact test $p < 10^{-4}$), and the association was even stronger in the subgroup of 255 women with OC diagnosis after the age of 60 (OR = 29.91, 95% CI: 14.99–59.66, $p < 10^{-4}$). No significant association of *BRIP1* PTVs with familial BC was observed.

Considering 4469 BC index patients and 2767 female control individuals, PTVs in *BARD1* were siginificantly associated with BC (OR = 5.35, 95% CI: 3.17–9.04, two-sided Fisher's Exact test $p < 10^{-5}$) [182]. *BARD1* PTV carriers showed a significantly younger mean age at primary diagnosis (42.3 years, range 24–60 years) compared with the overall study sample (48.6 years, range 17–92 years; two-sided Student's t-test $p = 3.5 \times 10^{-3}$). Consistent with this, stronger associations were found for the subgroups with lower age of disease (age at primary diagnosis <40 years: OR = 12.04, 95% CI: 5.78–25.08, two-sided Fisher's Exact test $p < 10^{-5}$; <50 years: OR = 7.43, 95% CI: 4.26–12.98, $p < 10^{-5}$). No *BARD1* PTVs were identified in 451 index patients with familial OC.

In summary, these case-control studies confirmed *FANCM* and *BARD1* as risk genes for hereditary BC, but additional studies including larger sample sets will be required to establish their potential role in OC pathogenesis. *BRIP1* could be shown to represent a high-risk gene for late-onset OC but not a BC predisposition gene, though minor effects cannot be excluded. *FANCM* and *BARD1* should be included in diagnostic multi-gene panel testing for individuals at risk for familial BC, and *BRIP1* for individuals at risk for familial OC.

5.2 Polygenic risk scores

Association of PRSs with BC risk was assessed with respect to two SNP sets, namely one set covering 77 SNPs developed for BC risk stratification in women unselected for their *BRCA1/2* germline mutation status [109] and one set covering 88 SNPs developed for BC risk stratification in female *BRCA1/2* mutation carriers [84]. Both SNP sets provided concordant PRS results at the individual level (Pearson's correlation coefficient $r = 0.91$, $p < 2.20 \times 10^{-16}$). Weighted cohort Cox regression analyses revealed statistically significant associations of standardized PRSs with the risk for BC. Considering the SNP set for BC risk stratification in women unselected for *BRCA1/2* germline mutation status, a hazard ratio (HR) of 1.71 per standard deviation (SD) of the PRS was observed (95% CI: 1.36–2.15, two-sided Wald test $p = 3.87 \times 10^{-6}$). The association of standardized PRSs with primary BC did not differ between carriers of the c.1100delC founder mutation and carriers of other PTVs in *CHEK2*, and the association attenuated with increasing age. Prediction of absolute cumulative BC risks revealed LTRs for primary BC at the age of 80 years of 0.13 at the 10th percentile versus 0.33 at the 90th percentile. It follows that, depending on the PRS, BC risk can be stratified in such a way that *CHEK2* PTV carriers may, on the one hand, have a BC risk corresponding to the general population but, on the other hand, may fall into risk groups for which access to intensified screening is recommended. Therefore, PRS calculation should be included in diagnostic multi-gene panel testing for individuals at risk for familial BC.

5.3 Copy number variant calling

In the study sample of 4208 female patients with familial BC/OC, the commercial SOPHiA DDM® software predicted 134 CNVs, of which 103 were classified as CNVs with high confidence and 31 were classified as CNVs with medium confidence [94]. Of the 134 predicted CNVs, 77 (57.46%) could be verified by MLPA. 33 CNVs were identified in genes other than *BRCA1/2*, mostly in *ATM*, *CHEK2*, and *RAD51C* and less frequently in *BARD1*, *MLH1*, *MSH2*, *PALB2*, *PMS2*, *RAD51D*, and *TP53*. The overall CNV prevalence of 1.81% and the CNV prevalence in non-*BRCA1/2* genes of 0.76% in the study sample, demonstrate the need for CNV prediction in genetic testing of individuals at risk for familial BC/OC, and that it should not be limited to *BRCA1/2*.

CNV calling with publicly available tools ExomeDepth, GATK gCNV, and panelcn.MOPS with subsequent MLPA verification did not result in the detection of additional CNVs, but 4 to 5 verified CNVs were missed by each of these methods. In conclusion, on multi-gene panel data the SOPHiA DDM® software outperforms the examined publicly available approaches for *in silico* CNV detection regarding its sensitivity. However, for 409 sequencing targets, no unambiguous copy number could be predicted by the SOPHiA DDM® software, and 57 out of 134 predicted CNVs (42.54%) could not be verified. These findings suggest that the need for improvement of read depth-based approaches for *in silico* CNV detection on multi-gene panel data persists, especially for the purpose of diagnostic gene testing. Wrong predictions and predictions

with low confidence of the SOPHiA DDM® software accumulate for sequencing targets with characteristics at the extremes of target length and GC content contributions, while moderately reduced mappability had no effect on the accuracy of CNV predictions. Therefore, sequencing targets of multi-gene panels should be optimized towards sequencing target lengths of >200 bp and average GC contents close to 0.5 with respect to CNV detection.

5.4 In silico prediction of variant pathogenicity

Considering 161 variants whose classification were independent of previous *in silico* prediction, sensitivities, i.e., the proportion of correctly predicted variants among all pathogenic variants, ranged from 0.67 (95% CI: 0.43–0.85) for PolyPhen-2 to 1.00 (95% CI: 0.84–1.00) for SIFT and MutationTaster2 [48]. Specificities, i.e., the proportion of correctly predicted variants among all benign variants, ranged from 0.67 (95% CI: 0.59–75) for PolyPhen-2 to 0.92 (95% CI: 0.86–0.96) for Align-GVGD. Consistently, PolyPhen-2 achieved the lowest MCC (0.39). Align-GVGD achieved the highest MCC (0.73), but was outperformed regarding its sensitivity (0.90, 95% CI: 0.70–0.99) by SIFT and MutationTaster2. All tools suffered predominantly from erroneous predictions of pathogenicity, and this shortcoming could not be bypassed by their combination. Thus, it has to be concluded that clinical consequences should never be based solely on *in silico* forecasts. However, due to observed negative predictive values (SIFT: 1.00 (95% CI: 0.96–1.00), MutationTaster2: 1.00 (95% CI: 0.97–1.00)), i.e., the proportion of correctly predicted variants among all variants predicted benign, it may be assumed that SIFT and MutationTaster2 could be suitable to predict benignity of VUS in *BRCA1/2*.

5.5 Implications of observed variant fractions for the assessment of pathogenicity

Germline variant calling and filtering revealed 208 rare variants in 181 of the 488 OC patients considered [64]. Of the 94 pathogenic germline variants identified, 79 showed significantly increased VFs in corresponding tumor samples, with fold changes (FCs) ranging from 1.15 to 2.05. The VFs between blood and tumor for the remaining 15 class 4/5 variants did not differ significantly (FC range 0.85–1.13). Of the 90 benign variants identified, 33 showed significantly increased VFs in the tumor with FCs ranging from 1.22 to 2.02, and 23 showed significantly decreased VFs with FCs ranging from 0.06 to 0.84. Variant classification, i.e., determination of pathogenicity based on the observation of significantly increased VFs, is hampered by the random distribution of VFs observed for benign variants. In contrast, observation of significantly decreased VFs were specific for benign variants and could serve as a suitable criterion for the classification of variants as benign. Such an approach resulted in a PPV of 1.0 (95% CI: 0.85–1.00) and a positive LR of 49.07 (95% CI: 3.02–795.93) in our sample.

5.6 Clonal hematopoiesis

In the sample of 523 OC patients, 7 potentially deleterious[1] missense mutations in *TP53* were observed in blood-derived DNA of 6 patients [181]. The VFs of 3 of these missense variants, each of which represented the only *TP53* mutation in the corresponding patient, were comparable or decreased in comparion to the VFs observed in the corresponding tumor samples (range 0.46–0.77), namely close to 0.5 (range 0.49–0.55). In contrast, the remaining 4 *TP53* variants, each of which had a VF considerably lower than 0.5 (namely, 0.34, 0.26, 0.17, and 0.07), were not or only barely detectable (VF<0.01) in the corresponding tumor samples. However, additional tumor-specific deleterious missense variants with VFs ranging from 0.39 to 0.63 were found in these patients in *TP53*, suggesting that the deleterious *TP53* variants identified in blood-derived DNA were not causal for the development of OC in these cases.

As in a sample of 1053 cancer-free female control individuals with a mean age at blood draw of 59.3 years (range 19–80 years), no deleterious *TP53* variants were identfied, it can be concluded that age-related CH affecting the *TP53* gene represents a rare event, and the observed CH events in the study sample were rather chemotherapy-induced. Also, CH events affecting the *PPM1D* gene were shown to be associated with prior exposure to chemotherapy [34]. In the study sample of 523 OC patients, 24 (4.6%) patients carried truncating variants in *PPM1D* with VF≤0.4, which were not or only barely detectable in the corresponding tumor-derived DNA. Of these 24 *PPM1D*-positive patients, 18 had completed first line platinum-based chemotherapy prior to blood draw, and in 5 cases, blood was drawn during 1st line chemotherapy. In the control sample of 1053 cancer-free female individuals, one 77-year-old woman carried a nonsense variant in *PPM1D* (VF=0.15). Thus, it can be concluded that CH events in *TP53* and *PPM1D* are primarily induced by exposure to chemotherapy, and less frequently by increased age. Consequently, to avoid misdiagnosis of LFS1 and resulting unnecessary medical interventions, non-homozygous deleterious *TP53* variants should be confirmed in an additional tissue, which is not part of the hematopoietic compartment, e.g. tumor, fingernails, or hair follicles.

[1] equivalent to the classification as `Non-functional` in the IARC TP53 database [22] and/or pathogenic according to the UMD TP53 database (`https://p53.fr/tp53-database`) and/or the Seshat TP53 variant classification tool [155], as of January to July 2018

6 Discussion

This thesis aimed to improve multi-gene panel testing in the framework of genetic counseling for persons at risk for familial BC/OC. The results contribute to an improvement in terms of gene panel design, variant calling and variant interpretation. Many findings, especially regarding variant detection and interpretation, are thereby also transferable to multi-gene panel testing with respect to a variety of other diseases with a genetic component, including other cancer entities [120, 122, 137], mental disorders [81], neurological diseases [12, 43], and metabolic diseases [157], among others.

6.1 Implications of case-control studies and polygenic risk score validation for optimized multi-gene panel design

With the introduction of multi-gene panel sequencing into clinical diagnostics, parallel screening of multiple disease-associated genes became feasible. Therefore, multi-gene panels usually cover the (coding) exons and at least 20bp of flanking intronic sequences of established risk genes. In addition, positions of common genetic variants, which may also be localized in non-coding genomic regions, can be sequencing targets, if significant correlations with disease development have been found in GWASs and the corresponding effects can be summarized into PRSs. Moreover, polymorphisms can be used for ethnicity checks and quality assurance purposes.

6.1.1 Gene selection for multi-gene panel analyses in genetic counseling of women at risk for familial breast and ovarian cancer

The genetic landscape of hereditary BC and OC, as far as known at the present time, is summarized in frequency-risk profiles in Fig. 6.1. Not coincidentally, *BRCA1* and *BRCA2* were the first genes to be shown to be associated with the development of BC/OC, as pathogenic variants within these genes have both high penetrance and prevalence, and hence, statistical significance can be achieved for comparatively small sample sizes. With the collaboration of the GC-HBOC, the search for predisposing genetic modificators of hereditary BC/OC is ongoing. In addition to the challenge of having sufficiently large sample sizes to detect risk genes in which pathogenic variants have low prevalence and/or penetrance, it is often problematic to clearly distinguish whether genes associated with BC are actually also associated with OC, and vice versa. During this thesis, *FANCM* and *BARD1* were confirmed as risk genes for hereditary BC, and *BRIP1* was confirmed as highly penetrant OC risk gene without pronounced effects on BC risk. Hence, coding regions of *FANCM* and *BARD1* should be included in sequenc-

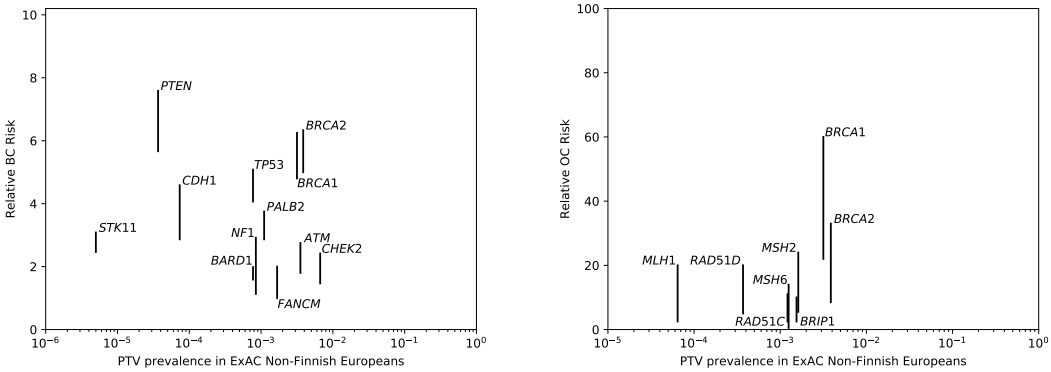


Figure 6.1: Risk penetrance profiles (adapted from Turnbull et al. [171]) for pathogenic variants in established risk genes for (left) breast and/or (right) ovarian cancer (BC/OC). Reported prevalences were estimated based on non-Finnish Europeans under exclusion of TCGA data in ExAC [92], as described in section 2.2.1.

ing targets of multi-gene panels for diagnostic germline testing of individuals at risk for familial BC, and coding regions of *BRIP1* for individuals at risk for familial OC, respectively. Our conclusions were drawn based on the numbers of observed PTVs, that were assumed to represent pathogenic variants without further evidence. In contrast, missense mutations, whose pathogenicity usually cannot be conclusively clarified without further experiments, were not integrated into the calculations, although a high penetrance cannot be excluded. These limitations are common for gene association studies, as well as the highly simplifying assumption that all pathogenic variants have a more or less similar effect on risk. This approach is no longer appropriate, particularly for well-explored genes, e.g., BC/OC risks of *BRCA1/2* germline mutation carriers are well-known to vary significantly in dependence to their localization in specific regions [83,135,166,167]. A further example of the limitations in the simple distinction of a variant into either pathogenic or non-pathogenic is the common variant NM_007194:c.470C>T in *CHEK2*. Although its significant association with BC has been consistently demonstrated in several studies and *CHEK2* is a confirmed BC risk gene with moderate penetrance, the variant contributes only slightly to an increase in BC risk: Han and colleagues reported OR = 1.58 (95% CI: 1.42–1.75) based on a meta-analysis including 15985 BC cases and 18609 control individuals from eight studies [61]. Consequently, NM_007194:c.470C>T is not treated as a pathogenic variant within the GC-HBOC [65], but has meanwhile been integrated into several SNP sets for BC PRS calculation [84, 109].

Inclusion of new genes in multi-gene panel analyses leads to the problem that risk assessment, i.e., determination of the magnitude of a corresponding variant's modification on cancer risk, that determines genetic counseling, is based on little evidence, and therefore uncertain. Even if several studies exist, and these are based on increas-

ingly larger case numbers, estimates of cancer risks vary widely between them, primarily due to biases based on the composition of the underlying study samples with regard to ethnical background, familial predisposition, BC/OC subtypes, etc. Furthermore, risks are age-dependent and cannot be directly derived from ORs or HRs. As part of the development of the ASK2ME web service, Braun and colleagues introduced a statistical framework for gaining age-specific penetrance risk estimates from such measures, which is based on the strongly simplifying assumption that these ratios remain constant during lifetime [23]. Therefore, our findings have been incorporated into their calculation of cancer risks associated with mutations in *BRIP1* [1]. The assumption of lifelong constancy was also made for relative risks for inclusion of *CHEK2* and *ATM* into the breast cancer risk prediction model BOADICEA [91].

In addition to confirming risk genes and improving the estimation of associated risks, the increasing amount of available data also facilitates the reliable exclusion of putative risk genes and, through the detection of double mutation carriers, also increasingly allows the investigation of the interplay of multiple risk genes in disease development. This is exemplified by the work of Klaschik et al. [79], excluding variant NM_003979:c.183del in *GPRC5A* as a modifier of BC risk in *BRCA1* germline mutation carriers.

6.1.2 Inclusion of polymorphisms for polygenic risk score computation and quality control

On the basis of a GWAS-independent study sample of female heterozygous *CHEK2* PTV carriers, my colleague Julika Borde and I were able to demonstrate that PRSs improve personalized BC risk prediction [21]. The SNP sets utilized for PRS calculations in our study, each comprise less than 100 loci, so that these PRSs have the potential to be straightforwardly implemented in routine diagnostic multi-gene panel analyses. The current version of the TruRisk® gene panel even covers slightly more than 300 loci of the BRIDGES PRS SNP set implemented in the CanRisk tool [15] for BC risk prediction. However, as of January 2021, the Polygenic Score (PGS) Catalog[1], a database for polygenic risk scores [87], contains 101 PRSs for BC with numbers of considered loci ranging from 9 to 6.4×10^6 and 7 PRSs for OC risk prediction with numbers of considered loci ranging from 11 to 36. Furthermore, some of the reported BC PRSs are specific to certain BC subtypes, raising the question of which SNP set has the greatest utility and informative value in the context of genetic counseling for individuals at risk for familial BC/OC. Although several studies confirmed that predictive power of PRSs generally increase with SNP set size [27], the extent of loci covered by a multi-gene panel should be as limited as possible with respect to consumed resources in a routine diagnostic setting. Clearly, genotyping of millions of loci is unfeasible. Moreover, Yanes and colleagues pointed out that the additional SNPs discovered only with increasing power of GWASs, tend to show smaller effects and therefore also contribute comparatively little to overall PRS relative to other SNPs [187]. The distributions of effect sizes dependent on SNP set size for the 101 BC PRSs listed in the PGS catalog are shown in Figure 6.2. In addition,

[1] https://www.pgscatalog.org

the performance improvement for larger SNP sets resultig from larger GWASs may also be caused by downward correction of already known associations (whose effects tend to be overestimated in smaller studies), rather than solely due to the addition of numerous loci with unexplained biological causality. Even if future research will have be determined the best-performing PRS set, its incorporation in BC/OC risk prediction cannot be readily generalized. A potential pitfall is the assumption of absolute independence of the PRS with other risk factors, such as germline mutation status and family history. Exemplary is the SNP rs132390, which is associated with the *CHEK2* founder mutation c.1100delC, and which we therefore had excluded in our study [21]. With regard to *BRCA1/2* mutation carriers, a recent study by Coignard et al. [31] reported that (1) several SNPs associated with BC risk in the general population are actually associated with the *BRCA1/2* mutation status, and hence do not have effects on BC risk in mutation carriers, (2) in line with Kuchenbaecker et al. [84], average effect sizes of individual SNPs tend to be smaller for mutation carriers than for the general population, and (3) SNPs exist that have an effect on BC risk exclusively in *BRCA1* or *BRCA2* mutation carriers, but not in the general population.

In addition, the potential of PRSs for risk stratification of PTV carriers is still poorly known for many risk genes. However, a recent study confirmed the utility of a 86-SNP PRS for carriers of PTVs in risk genes *ATM* and *PALB2*, besides *BRCA1/2* and *CHEK2* [53].

The recognition of the contribution of multiple traits to individual disease risks represents a paradigm shift in both personalized risk predictions as well as the sequencing strategy in clinical diagnostics. Up to now, multi-gene panel sequencing was applied to identify single pathogenic variants in index patients, which were then confirmed or excluded in unaffected relatives by Sanger sequencing. Inclusion of PRSs in risk prediction now requires sequencing of many loci in each individual.

Besides risk stratification for healthy individuals, polymorphisms have the potential to be utilized for quality controls, namely checks for duplicated samples and ethnicity backgrounds. The latter is especially crucial for the interpretation of PRSs, as GWASs have so far been conducted primarily for populations of European descent, and hence, can not be generalized to individuals of non-European descent [159]. Turner and colleagues pointed out that SNP sets of size $\geq 10^5$ are required to yield reliable kinship estimates [172]. Usually, so many SNPs are not covered in multi-gene panel analysis in a clinical diagnostic setting. However, the covered set of independent SNPs should at least be sufficient to identify duplicates, as exemplified by three monozygotic twin pairs in our study [21]. Further, we demonstrated how ethnicity stratification based on identity-by-state (IBS) values, i.e., for a pair of individuals i and j

$$\mathrm{IBS}_{i,j} = \frac{1}{N}\sum_{k=1}^{N}\frac{(g_{i,k}-p_k)(g_{j,k}-p_k)}{p_k(1-p_k)},$$

where N is the number of SNPs, $g_{i,k} \in \{0, 0.5, 1\}$ is the genotype of person i at SNP k, and p_k is the frequency of SNP k, can be adapted to smaller SNP set sizes by introducing nearest shrunken centroid classiers and an additional dimension to the analysis [21].

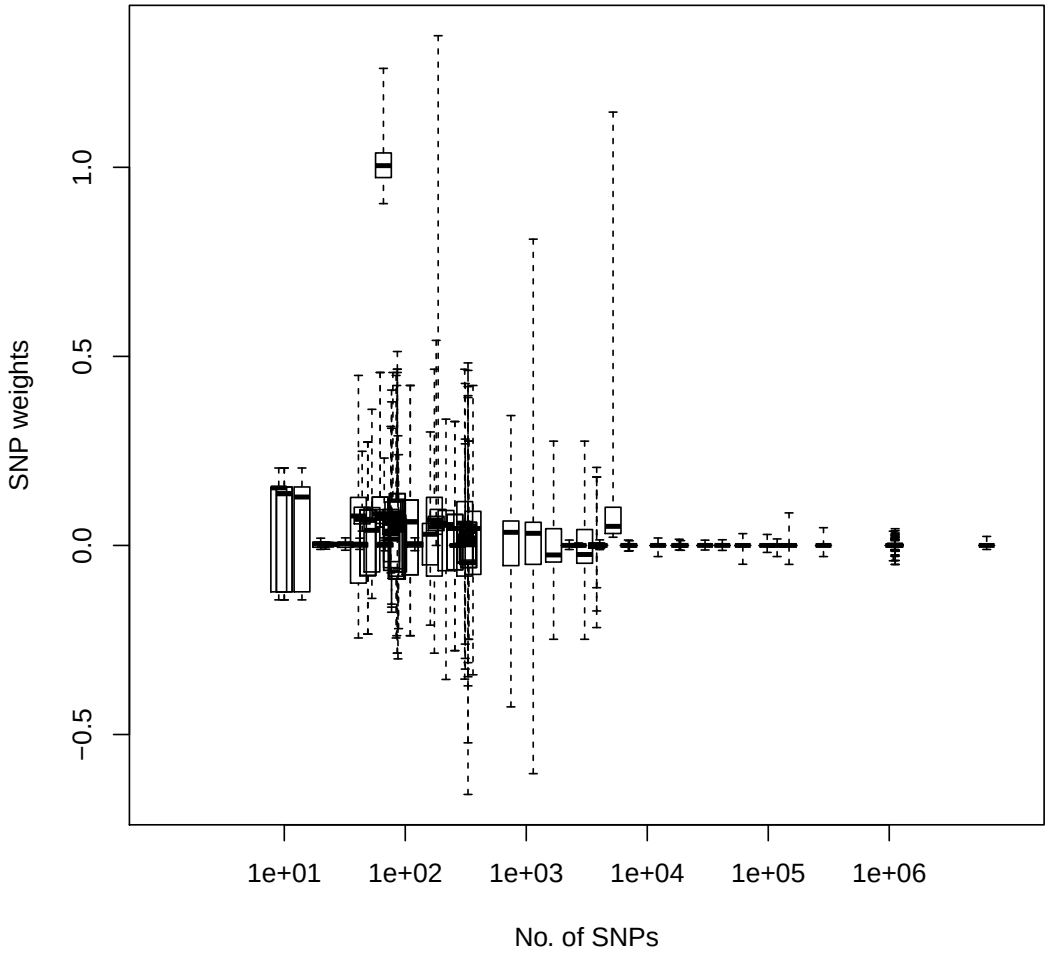


Figure 6.2: Effect sizes of single nucleotide polymorphisms (SNPs) in dependence to set size for 101 breast cancer (BC) polygenic risk scores (PRSs) as listed in the Polygenic Score (PGS) Catalog, a database for PRSs [87] as of January 2021. The smallest set consisted of 9 SNPs, while the largest consisted of 6390808 SNPs.

6.2 Enhancement of variant detection

Louisa Lepkes and I demonstrated in our study that CNVs constitute a non-negligible fraction in the spectrum of putative BC/OC-causing variants, namely 1.81% of 4208 GC-HBOC index patients were affected by at least one CNV in a cancer-predisposing gene [94]. More than two-fifths of these CNVs were localized in genes other than *BRCA1/2*. Therefore, CNV detection should become standard in genetic testing alongside calling of SNVs and indels, and should be applied to all genes of interest. However, consistent with many other studies we have also shown that CNV detection methods are limited in terms of their sensitivity, but also their specificity. The combined use of several tools does not sufficiently overcome these problems, as these are in principle based on the same assumptions and methods, namely that after normalization for inter- and intra-sample variance, CNVs determine observed read depths of mean read coverage per sequencing target in a way that these are approximately halfed or (at least) doubled in comparison to other samples. Therefore, missed or wrong CNV predictions frequently accumulate for the same sequencing targets among different approaches. An improvement of the reliability of *in silico* CNV detection approaches, as required for large-scale deployment in clinical diagnostics, depends on the development of more capable methods. I contribute to these efforts through the development of our inhouse CNV detection approach OPaCNAT, which is based on the usage of generalized additive model (GAMs) [49], and has already been applied, for example, in the GeparOcto study [130]. In a nutshell, my approach models centered read counts along the genomic positions of a sequencing target as a product of two smooth functions, namely, a generic background function that contributes to all samples under consideration and a sample-specific smooth function which is used for final CNV calling. Therefore, in contrast to other tools, the method takes FCs at several positions per sequencing target into account, instead of examining only single values per target and sample, such as read counts or mean read coverage. Due to this different form of input data, OPaCNAT has the potential to detect CNVs that are consistently missed by conventional approaches, as was the case in the GeparOcto analyses for a duplication of exon 8 in NM_000059.3 (*BRCA2*) that was missed by the established *in silico* detection approaches ExomeDepth and panelcn.MOPS (data not shown). The data generated in collaboration with Louisa Lepkes [94], comprising verified CNVs in 17 cancer predisposition genes of 4208 GC-HBOC index patients, will serve as a comprehensive benchmark for the continued development of OPaCNAT in the future. Furthermore, we have shown that the reliability of CNV detection may benefit, in addition to an improvement of the available tools, from an optimization of the sequencing target regions in terms of GC content and length. Accumulation of false positive CNV predictions at the extremes of the length or GC content distribution of sequencing targets, as observed for the SOPHiA DDM® software, can be determined consistently for all tools under investigation, as exemplified in Figure 6.3 for panelcn.MOPS and ExomeDepth. Further examination is required to determine if the performance of *in silico* CNV prediction can be noticeably improved by according re-definition of CNV target regions, e.g., in terms of optimizing GC contents.

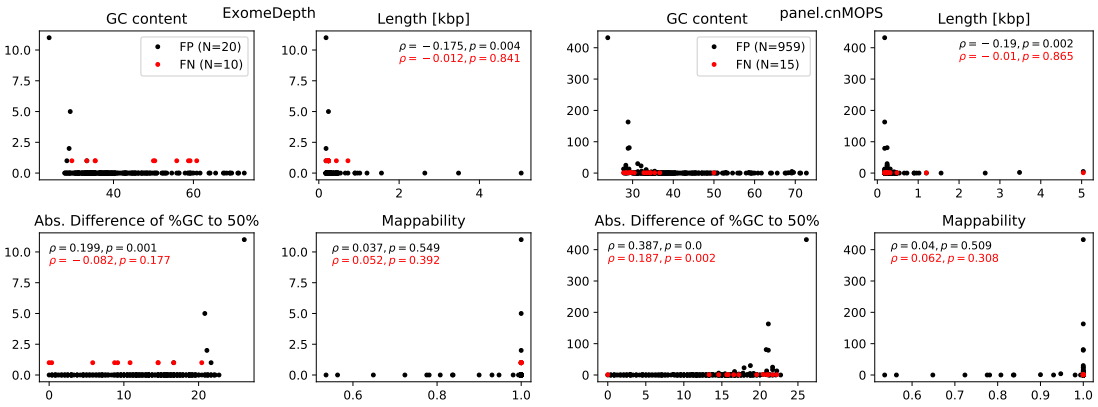


Figure 6.3: False positive (FP, black) and false negative (FN, red) *in silico* CNV predictions of ExomeDepth (left) and panelcn.MOPS (right) depending on sequencing target characteristics GC content, length and mappability. ρ: Spearman's rank correlation coefficient, p: asymptotic t approximation p value.

However, read depth-based CNV calling approaches are only applicable to deletions or duplications spanning at least one sequencing target, and thus, have significant effects on observed read counts, or mean read coverage. Identification of a sample's entire mutational spectrum requires the additional application of split read-, read pair- and/or assembly-based approaches, i.e., SV callers, beyond standard variant calling and read depth-based CNV detection. Their relevance and standard applicability in routine diagnostics, as well as adequate wet lab verification of *in silico* predicted SVs, e.g., via long read sequencing, is subject of ongoing research within the GC-HBOC.

6.3 Implications of observed variant fractions and in silico predictions for variant interpretation

The proportion of VUS, whose clinical relevance to disease risk remains unclear, increases proportionally with the number of genes tested in a multi-gene panel approach [124], and assessment of these variants via multifactorial likelihood analysis requires input data, e.g., on co-segregation, family cancer history, co-occurrence with further pathogenic variants in the same gene, tumor pathology or case-control ratios, which are usually not available. During this thesis, I contributed to the recognition that these difficulties cannot be circumvented by the application of *in silico* prediction approaches alone, even when their results are considered in combination [48]. Therefore, this practice should not be considered for clinical laboratories providing genetic counseling.

However, with the increase of genetic testing and the establishment and extension of comprehensive, systematic and centralized patient registries, preferably with follow-up and under continuous counseling, information on multiple individuals with identical VUS can be pooled. Thus, evidence for improved assessment of rare variants can be achieved.

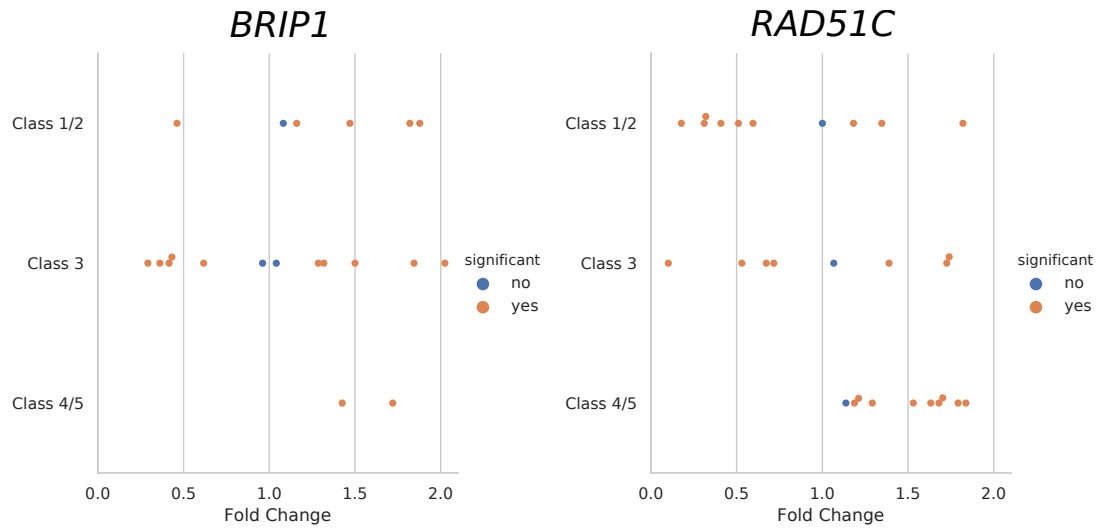


Figure 6.4: Log2 fold changes of variants fractions (VFs) observed in blood and corresponding ovary tumor sample for rare variants (MAF < 0.01 with respect to ExAC Non-Finnish Europeans under exclusion of TCGA data) in OC high risk genes *BRIP1* and *RAD51C* in 488 OC patients. Variants were previously classified according to the regulations of the ENIGMA consortium using the IARC 5-tier variant classification system into classes 1/2 (benign), 3 (VUS), or 4/5 (pathogenic). Statistically signifcant deviations in VF (two-sided Fisher $p < 0.05$ after Benjamini-Hochberg correction) are encoded in orange.

In collaboration with Dr. Jan Hauke, I assessed the determination of variant pathogenicity based on the comparison of VFs between blood and tumor samples considering 208 rare variants in 181 OC patients [64]. Our results demonstrate that significantly decreased VFs in tumor tissue may provide a suitable criterion for the assessment of *BRCA1/2* variants as benign. We observed that these findings are principally transferable to further OC risk genes besides *BRCA1/2*, as exemplified for *BRIP1* and *RAD51C* in Figure 6.4, although the prevalence of pathogenic germline mutations in these genes was too low to achieve statistical significance in our study sample. Further and larger studies are required to quantify the utility of these kind of LOH analysis for the classification of rare germline mutations affecting other genes and tumor entities, i.e., BC.

In collaboration with Konstantin Weber-Lassalle, I demonstrated that before interpreting variants in the *TP53* risk gene, it must first be clarified whether chemotherapy-induced and/or age-related CH might be present [181]. A recent study [19] also showed significant associations of CH with age and chemotherapy exposure in moderate penetrant BC risk gene *CHEK2*, and of CH with age in the also moderate penetrant BC risk gene *ATM*. These two genes were not included in our study design, although they are examined as part of multi-gene panel analyses in clinical diagnostics, because our study

sample consisted of OC patients. Another limitation of our study was that we could not associate putative CH events with subsequent occurrence of hematologic cancer. This was mainly due to the high mortality of ovarian cancer patients. Our results also show that CH is not exclusively manifested in significantly reduced VFs compared with 0.5, as we identified 3 putative CH events with VF $\geq$ 0.4 in the last exon of *PPM1D*. In addition, Bolton and colleagues found putative CH events with VFs up to 0.78 [19]. Therefore, confirmation of nonhomozygous deleteroius variants in *ATM*, *CHEK2*, and *TP53* in an additional tissue that is not part of the hematopoietic compartment is recommended regardless of observed VFs.

7 Conclusion

This thesis demonstrated that and how genetic counseling based on multi-gene panel sequencing can be improved by comprehensive bioinformatic analyses as they are currently not provided by the automated, preferably commercial, solutions for NGS data processing which are used as standard in many laboratories.

First, ethnicity checks are a prerequisite to be able to evaluate germline variants and the applicability of PRSs, which were demonstrated to be eligible for personalized risk prediction.

Second, standard variant calling must be supplemented by approaches for the detection of large genomic rearrangements to explore the entire spectrum of putative damaging variants. However, these *in silico* predictions must always be verified by additional experiments, since existing methods tend to false positive calls.

Third, current methods for *in silico* variant annotation are inappropriate for clinical decision making, but may give an indication of benignity, as well as a significant decrease of tumor-derived VF in comaprison to germline VF. Before communication of a putative disease-causing variant, the possibility of CH has to be considered.

With an increasing amount of available data and corresponding research by consortia such as the GC-HBOC, personalized BC/OC risk prediction will continue to improve in the coming years, and risk prediction will become more specific to variants than to affected genes. The PRS will become available for people of non-European ethnicity and its potential for risk prediction will be further improved by adjustment for tumor subtype, germline mutation status but also application to specific questions such as the risk of contralateral BC. With the recognition of the relevance of SVs in BC/OC development, the corresponding detection methods, both wet and dry, will become more prevalent and evolve in laboratories, and the contribution of such variants to BC/OC predisposition will be better understood through the availability of more data. Furthermore, genetic analyses are becoming increasingly important for personalized therapy decision-making, e.g., an association of germline mutation status with therapy-induced CH is suspected and is currently being investigated in one of our research projects at the Center for Familial Breast and Ovarian Cancer in Cologne.

8 References

[1] ASK2ME™. http://www.ask2me.org. Accessed: 2020-12-09.

[2] Cancer Research UK Homepage. https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/lifetime-risk. Accessed: 2020-08-19.

[3] A database of germline genomic variation in healthy older women. https://whi.color.com. Accessed: 2020-09-29.

[4] ENIGMA *BRCA1/2* gene variant classification criteria. https://enigmaconsortium.org/wp-content/uploads/2016/06/ENIGMA_Rules_2015-03-26.pdf. Accessed: 2016-09-13.

[5] Sequence Ontology terms for describing variant consequences. http://www.sequenceontology.org/browser/current_release. Accessed: 2020-09-30.

[6] Interdisziplinäre S3-Leitlinie für die Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms. Langversion 4.3, 2020.

[7] R. P. Abo, M. Ducar, E. P. Garcia, A. R. Thorner, V. Rojas-Rudilla, L. Lin, L. M. Sholl, W. C. Hahn, M. Meyerson, N. I. Lindeman, et al. BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic acids research*, 43(3):e19–e19, 2015.

[8] I. A. Adzhubei, S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. A method and server for predicting damaging missense mutations. *Nat Meth*, 7(4):248–249, 2010.

[9] W. M. Alenezi, C. T. Fierheller, N. Recio, and P. N. Tonin. Literature Review of *BARD1* as a Cancer Predisposing Gene with a Focus on Breast and Ovarian Cancers. *Genes*, 11(8):856, 2020.

[10] C. Alkan, B. P. Coe, and E. E. Eichler. Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, 12(5):363, 2011.

[11] L. Andrews and D. G. Mutch. Hereditary ovarian cancer and risk reduction. *Best practice & research Clinical obstetrics & gynaecology*, 41:31–48, 2017.

[12] T. Antoniadi, C. Buxton, G. Dennis, N. Forrester, D. Smith, P. Lunt, and S. Burton-Jones. Application of targeted multi-gene panel testing for the diagnosis of inherited peripheral neuropathy provides a high diagnostic yield with unexpected phenotype-genotype variability. *BMC medical genetics*, 16(1):84, 2015.

[13] A. C. Antoniou, S. Casadei, T. Heikkinen, D. Barrowdale, K. Pylkäs, J. Roberts, A. Lee, D. Subramanian, K. De Leeneer, F. Fostira, et al. Breast-cancer risk in families with mutations in *PALB2*. *New England Journal of Medicine*, 371(6):497–506, 2014.

[14] A. C. Antoniou, A. Cunningham, J. Peto, D. Evans, F. Lalloo, S. Narod, H. Risch, J. Eyfjord, J. Hopper, M. Southey, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *British journal of cancer*, 98(8):1457–1466, 2008.

[15] S. Archer, C. Babb de Villiers, F. Scheibl, T. Carver, S. Hartley, A. Lee, A. P. Cunningham, D. F. Easton, J. G. McIntosh, J. Emery, et al. Evaluating clinician acceptability of the prototype CanRisk tool for predicting risk of breast and ovarian cancer: A multi-methods study. *PloS one*, 15(3):e0229999, 2020.

[16] F. Batalini, E. G. Peacock, L. Stobie, A. Robertson, J. Garber, J. N. Weitzel, and N. M. Tung. Li-Fraumeni syndrome: not a straightforward diagnosis anymorethe interpretation of pathogenic variants of low allele frequency and the differences between germline PVs, mosaicism, and clonal hematopoiesis. *Breast Cancer Research*, 21(1):107, 2019.

[17] T. B. Bevers, M. Helvie, E. Bonaccio, K. E. Calhoun, M. B. Daly, W. B. Farrar, J. E. Garber, R. Gray, C. C. Greenberg, R. Greenup, et al. Breast cancer screening and diagnosis, version 3.2018, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, 16(11):1362–1389, 2018.

[18] V. Boeva, T. Popova, M. Lienard, S. Toffoli, M. Kamal, C. Le Tourneau, D. Gentien, N. Servant, P. Gestraud, T. Rio Frio, et al. Multi-factor data normalization enables the detection of copy number aberrations in amplicon sequencing data. *Bioinformatics*, 30(24):3443–3450, 2014.

[19] K. L. Bolton, R. N. Ptashkin, T. Gao, L. Braunstein, S. M. Devlin, D. Kelly, M. Patel, A. Berthon, A. Syed, M. Yabe, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nature genetics*, 52(11):1219–1226, 2020.

[20] V. Bonadona, B. Bonaïti, S. Olschwang, S. Grandjouan, L. Huiart, M. Longy, R. Guimbaud, B. Buecher, Y.-J. Bignon, O. Caron, et al. Cancer risks associated with germline mutations in *MLH1*, *MSH2*, and *MSH6* genes in Lynch syndrome. *Jama*, 305(22):2304–2310, 2011.

[21] J. Borde, C. Ernst, B. Wappenschmidt, D. Niederacher, K. Weber-Lassalle, G. Schmidt, J. Hauke, A. S. Quante, N. Weber-Lassalle, J. Horváth, et al. Performance of breast cancer polygenic risk scores in 760 female *CHEK2* germline mutation carriers. *JNCI: Journal of the National Cancer Institute*, 2020.

[22] L. Bouaoun, D. Sonkin, M. Ardin, M. Hollstein, G. Byrnes, J. Zavadil, and M. Olivier. *TP53* variations in human cancers: new lessons from the IARC *TP53* database and genomics data. *Human mutation*, 37(9):865–876, 2016.

[23] D. Braun, J. Yang, M. Griffin, G. Parmigiani, and K. S. Hughes. A clinical decision support tool to predict cancer risk for commonly tested cancer-related germline mutations. *Journal of genetic counseling*, 27(5):1187–1199, 2018.

[24] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[25] T. Byrski, T. Huzarski, R. Dent, E. Marczyk, M. Jasiówka, J. Gronwald, J. Jakubowicz, C. Cybulski, R. Wisniowski, D. Godlewski, et al. Pathologic complete response to neoadjuvant cisplatin in BRCA1-positive breast cancer patients. *Breast cancer research and treatment*, 147(2):401–405, 2014.

[26] X. Chen, O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, A. J. Cox, S. Kruglyak, and C. T. Saunders. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32(8):1220–1222, 2016.

[27] S. W. Choi, T. S.-H. Mak, and P. F. O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, 15(9):2759–2772, 2020.

[28] D. M. Church, V. A. Schneider, T. Graves, K. Auger, F. Cunningham, N. Bouk, H.-C. Chen, R. Agarwala, W. M. McLaren, G. R. Ritchie, et al. Modernizing reference genome assemblies. *PLoS Biol*, 9(7):e1001091, 2011.

[29] P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92, 2012.

[30] P. J. Cock, C. J. Fields, N. Goto, M. L. Heuer, and P. M. Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, 38(6):1767–1771, 2010.

[31] J. Coignard, M. Lush, J. Beesley, T. A. OMara, J. Dennis, J. P. Tyrer, D. R. Barnes, L. McGuffog, G. Leslie, M. K. Bolla, et al. A case-only study to identify genetic modifiers of breast cancer risk for *BRCA1/BRCA2* mutation carriers. *Nature communications*, 12(1):1–22, 2021.

[32] P. Concolino and E. Capoluongo. Detection of *BRCA1/2* large genomic rearrangements in breast and ovarian cancer patients: an overview of the current methods. *Expert review of molecular diagnostics*, 19(9):795–802, 2019.

[33] P. Concolino, R. Rizza, F. Mignone, A. Costella, D. Guarino, I. Carboni, E. Capoluongo, C. Santonocito, A. Urbani, and A. Minucci. A comprehensive

BRCA1/2 NGS pipeline for an immediate Copy Number Variation (CNV) detection in breast and ovarian cancer molecular diagnosis. *Clinica Chimica Acta*, 480:173–179, 2018.

[34] C. C. Coombs, A. Zehir, S. M. Devlin, A. Kishtagari, A. Syed, P. Jonsson, D. M. Hyman, D. B. Solit, M. E. Robson, J. Baselga, et al. Therapy-related clonal hematopoiesis in patients with non-hematologic cancers is common and associated with adverse clinical outcomes. *Cell stem cell*, 21(3):374–382, 2017.

[35] F. J. Couch, H. Shimelis, C. Hu, S. N. Hart, E. C. Polley, J. Na, E. Hallberg, R. Moore, A. Thomas, J. Lilyquist, et al. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA oncology*, 3(9):1190–1196, 2017.

[36] D. J. Crouch and W. F. Bodmer. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences*, 117(32):18924–18933, 2020.

[37] R. Daber, S. Sukhadia, and J. J. Morrissette. Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets. *Cancer genetics*, 206(12):441–448, 2013.

[38] M. B. Daly, R. Pilarski, M. Berry, S. S. Buys, M. Farmer, S. Friedman, J. E. Garber, N. D. Kauff, S. Khan, C. Klein, et al. NCCN guidelines insights: genetic/familial high-risk assessment: breast and ovarian, version 2.2017. *Journal of the National Comprehensive Cancer Network*, 15(1):9–20, 2017.

[39] J. T. den Dunnen, R. Dalgleish, D. R. Maglott, R. K. Hart, M. S. Greenblatt, J. McGowan-Jordan, A.-F. Roux, T. Smith, S. E. Antonarakis, P. E. Taschner, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Human mutation*, 37(6):564–569, 2016.

[40] M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5):491, 2011.

[41] L. Dorling, S. Carvalho, J. Allen, A. González-Neira, C. Luccarini, C. Wahlström, K. A. Pooley, M. T. Parsons, C. Fortuno, Q. Wang, et al. Breast Cancer Risk Genes-Association Analysis in More than 113,000 Women. *The New England Journal of Medicine*, 2021.

[42] L. Duncan, H. Shen, B. Gelaye, J. Meijsen, K. Ressler, M. Feldman, R. Peterson, and B. Domingue. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, 10(1):1–9, 2019.

[43] P. Dunn, C. L. Albury, N. Maksemous, M. C. Benton, H. G. Sutherland, R. A. Smith, L. M. Haupt, and L. R. Griffiths. Next generation sequencing methods for diagnosis of epilepsy syndromes. *Frontiers in Genetics*, 9:20, 2018.

[44] D. F. Easton, F. Lesueur, B. Decker, K. Michailidou, J. Li, J. Allen, C. Luccarini, K. A. Pooley, M. Shah, M. K. Bolla, et al. No evidence that protein truncating variants in *BRIP1* are associated with breast cancer risk: implications for gene panel testing. *Journal of medical genetics*, 53(5):298–309, 2016.

[45] D. F. Easton, P. D. Pharoah, A. C. Antoniou, M. Tischkowitz, S. V. Tavtigian, K. L. Nathanson, P. Devilee, A. Meindl, F. J. Couch, M. Southey, et al. Gene-panel sequencing and the prediction of breast-cancer risk. *New England Journal of Medicine*, 372(23):2243–2257, 2015.

[46] R. C. Edgar and H. Flyvbjerg. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, 31(21):3476–3482, 2015.

[47] C. Engel, M. Loeffler, V. Steinke, N. Rahner, E. Holinski-Feder, W. Dietmaier, H. K. Schackert, H. Goergens, M. von Knebel Doeberitz, T. O. Goecke, et al. Risks of Less Common Cancers in Proven Mutation Carriers With Lynch Syndrome . *Journal of Clinical Oncology*, 30(35):4409–4415, 2012.

[48] C. Ernst, E. Hahnen, C. Engel, M. Nothnagel, J. Weber, R. K. Schmutzler, and J. Hauke. Performance of *in silico* prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics. *BMC medical genomics*, 11(1):35, 2018.

[49] C. Ernst, R. K. Schmutzler, and E. Hahnen. Generalized Additive Models for the Detection of Copy Number Variations (CNVs) Using Multi-gene Panel Sequencing Data. In *Advanced Studies in Classification and Data Science*, pages 199–213. Springer, 2020.

[50] M. Escalona, S. Rocha, and D. Posada. A comparison of tools for the simulation of genomic next-generation sequencing data. *Nature Reviews Genetics*, 17(8):459, 2016.

[51] D. G. Evans, E. R. Woodward, S. Bajalica-Lagercrantz, C. Oliveira, and T. Frebourg. Germline *TP53* Testing in Breast Cancers: Why, When and How? *Cancers*, 12(12):3762, 2020.

[52] E. J. Fox, K. S. Reid-Bayliss, M. J. Emond, and L. A. Loeb. Accuracy of next generation sequencing platforms. *Next generation, sequencing & applications*, 1, 2014.

[53] S. Gallagher, E. Hughes, S. Wagner, P. Tshiaba, E. Rosenthal, B. B. Roa, A. W. Kurian, S. M. Domchek, J. Garber, J. Lancaster, et al. Association of a polygenic risk score with breast cancer among women carriers of high-and moderate-risk breast cancer genes. *JAMA network open*, 3(7):e208501–e208501, 2020.

[54] E. Garrison and G. Marth. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*, 2012.

[55] G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *New England Journal of Medicine*, 371(26):2477–2487, 2014.

[56] German Consortium for Hereditary Breast and Ovarian Cancer. Indikationen für einen Gentest. `https://www.konsortium-familiaerer-brustkrebs.de/betreuungskonzept/molekulare-diagnostik/indikationen-gentest/`. Accessed: 2021-02-10.

[57] D. E. Goldgar, D. F. Easton, A. M. Deffenbaugh, A. N. Monteiro, S. V. Tavtigian, F. J. Couch, and Breast Cancer Information Core (BIC) Steering Committee. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to *BRCA1* and *BRCA2*. *The American Journal of Human Genetics*, 75(4):535–544, 2004.

[58] S. Goodwin, J. D. McPherson, and W. R. McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333, 2016.

[59] D. G. Grimm, C.-A. Azencott, F. Aicheler, U. Gieraths, D. G. MacArthur, K. E. Samocha, D. N. Cooper, P. D. Stenson, M. J. Daly, J. W. Smoller, et al. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Human mutation*, 36(5):513–523, 2015.

[60] K. Hackmann, F. Kuhlee, E. Betcheva-Krajcir, A.-K. Kahlert, L. Mackenroth, B. Klink, N. Di Donato, A. Tzschach, K. Kast, P. Wimberger, et al. Ready to clone: CNV detection and breakpoint fine-mapping in breast and ovarian cancer susceptibility genes by high-resolution array CGH. *Breast cancer research and treatment*, 159(3):585–590, 2016.

[61] F. Han, C. Guo, and L. Liu. The effect of *CHEK2* variant I157T on cancer susceptibility: evidence from a meta-analysis. *DNA and cell biology*, 32(6):329–335, 2013.

[62] O. Harismendy, R. B. Schwab, L. Bao, J. Olson, S. Rozenzhak, S. K. Kotsopoulos, S. Pond, B. Crain, M. S. Chee, K. Messer, et al. Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome biology*, 12(12):1–13, 2011.

[63] P. Harter, J. Hauke, F. Heitz, A. Reuss, S. Kommoss, F. Marmé, A. Heimbach, K. Prieske, L. Richters, A. Burges, et al. Prevalence of deleterious germline variants in risk genes including *BRCA1/2* in consecutive ovarian cancer patients (AGO-TR-1). *PloS one*, 12(10):e0186043, 2017.

[64] J. Hauke, P. Harter, C. Ernst, A. Burges, S. Schmidt, A. Reuss, J. Borde, N. De Gregorio, D. Dietrich, A. El-Balat, et al. Sensitivity and specificity of loss of heterozygosity analysis for the classification of rare germline variants in *BRCA1/2*: results of the observational AGO-TR1 study (NCT02222883). *Journal of Medical Genetics*, 2020.

[65] J. Hauke, J. Horvath, E. Groß, A. Gehrig, E. Honisch, K. Hackmann, G. Schmidt, N. Arnold, U. Faust, C. Sutter, et al. Gene panel testing of 5589 *BRCA1/2*-negative index patients with breast cancer in a routine diagnostic setting: results of the german consortium for hereditary breast and ovarian cancer. *Cancer medicine*, 7(4):1349–1358, 2018.

[66] R. L. Hollis and C. Gourley. Genetic and molecular changes in ovarian cancer. *Cancer biology & medicine*, 13(2):236, 2016.

[67] N. Howlader, K. A. Cronin, A. W. Kurian, and R. Andridge. Differences in breast cancer survival by molecular subtypes in the United States. *Cancer Epidemiology and Prevention Biomarkers*, 27(6):619–626, 2018.

[68] C. Hu, S. N. Hart, R. Gnanaolivu, H. Huang, K. Y. Lee, J. Na, C. Gao, J. Lilyquist, S. Yadav, N. J. Boddicker, et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *New England Journal of Medicine*, 2021.

[69] Z. Hu, C. Yu, M. Furutsuki, G. Andreoletti, M. Ly, R. Hoskins, A. N. Adhikari, and S. E. Brenner. VIPdb, a genetic variant impact predictor database. *Human mutation*, 40(9):1202–1214, 2019.

[70] J. Huddleston, M. J. Chaisson, K. M. Steinberg, W. Warren, K. Hoekzema, D. Gordon, T. A. Graves-Lindsay, K. M. Munson, Z. N. Kronenberg, L. Vives, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome research*, 27(5):677–685, 2017.

[71] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931, 2004.

[72] T. Judkins, E. Rosenthal, C. Arnell, L. A. Burbidge, W. Geary, T. Barrus, J. Schoenberger, J. Trost, R. J. Wenstrup, and B. B. Roa. Clinical significance of large rearrangements in *BRCA1* and *BRCA2*. *Cancer*, 118(21):5210–5216, 2012.

[73] K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, 2020.

[74] K. Kast, K. Rhiem, B. Wappenschmidt, E. Hahnen, J. Hauke, B. Bluemcke, V. Zarghooni, N. Herold, N. Ditsch, M. Kiechle, et al. Prevalence of *BRCA1/2* germline mutations in 21 401 families with breast and ovarian cancer. *Journal of medical genetics*, 53(7):465–471, 2016.

[75] J. Kerkhof, L. C. Schenkel, J. Reilly, S. McRobbie, E. Aref-Eshghi, A. Stuart, C. A. Rupar, P. Adams, R. A. Hegele, H. Lin, et al. Clinical validation of copy number variant detection from targeted next-generation sequencing panels. *The Journal of Molecular Diagnostics*, 19(6):905–920, 2017.

[76] B. Khabou, A.-M. Durand-Schneider, J.-L. Delaunay, T. Aït-Slimane, V. Barbu, F. Fakhfakh, C. Housset, and M. Maurice. Comparison of *in silico* prediction and experimental assessment of ABCB4 variants identified in patients with biliary diseases. *The international journal of biochemistry & cell biology*, 89:101–109, 2017.

[77] J. I. Kiiski, L. M. Pelttari, S. Khan, E. S. Freysteinsdottir, I. Reynisdottir, S. N. Hart, H. Shimelis, S. Vilske, A. Kallioniemi, J. Schleutker, et al. Exome sequencing identifies *FANCM* as a susceptibility gene for triple-negative breast cancer. *Proceedings of the National Academy of Sciences*, 111(42):15172–15177, 2014.

[78] M.-C. King, S. Wieand, K. Hale, M. Lee, T. Walsh, K. Owens, J. Tait, L. Ford, B. K. Dunn, J. Costantino, et al. Tamoxifen and breast cancer incidence among women with inherited mutations in *BRCA1* and *BRCA2*: National Surgical Adjuvant Breast and Bowel Project (NSABP-P1) Breast Cancer Prevention Trial. *Jama*, 286(18):2251–2256, 2001.

[79] K. Klaschik, J. Hauke, G. Neidhardt, C. Tränkle, H. M. Surowy, S. Heilmann-Heimbach, G. Rappl, E. Mangold, N. Arnold, D. Niederacher, et al. The *GPRC5A* frameshift variant c.183del is not associated with increased breast cancer risk in *BRCA1* mutation carriers. *International journal of cancer*, 144(7):1761–1763, 2019.

[80] A. G. Knudson. Hereditary cancer: two hits revisited. *Journal of cancer research and clinical oncology*, 122(3):135–140, 1996.

[81] E. Koyama, C. C. Zai, L. Bryushkova, J. L. Kennedy, and J. H. Beitchman. Predicting risk of suicidal ideation in youth using a multigene panel for impulsive aggression. *Psychiatry research*, 285:112726, 2020.

[82] I. Kozarewa, J. Armisen, A. F. Gardner, B. E. Slatko, and C. Hendrickson. Overview of target enrichment strategies. *Current protocols in molecular biology*, 112(1):7–21, 2015.

[83] K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes, K.-A. Phillips, T. M. Mooij, M.-J. Roos-Blom, S. Jervis, F. E. van Leeuwen, R. L. Milne, N. Andrieu, D. E. Goldgar, M. B. Terry, M. A. Rookus, D. F. Easton, A. C. Antoniou, and the BRCA1 and BRCA2 Cohort Consortium. Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. *JAMA*, 317(23):2402–2416, 06 2017.

[84] K. B. Kuchenbaecker, L. McGuffog, D. Barrowdale, A. Lee, P. Soucy, J. Dennis, S. M. Domchek, M. Robson, A. B. Spurdle, S. J. Ramus, et al. Evaluation of polygenic risk scores for breast and ovarian cancer risk prediction in *BRCA1* and *BRCA2* mutation carriers. *JNCI: Journal of the National Cancer Institute*, 109(7), 2017.

[85] P. Kumar, S. Henikoff, and P. C. Ng. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protocols*, 4(7):1073–1081, 2009.

[86] A. W. Kurian, E. Hughes, E. A. Handorf, A. Gutin, B. Allen, A.-R. Hartman, and M. J. Hall. Breast and ovarian cancer penetrance estimates derived from germline multiple-gene sequencing results in women. *JCO Precision Oncology*, 1:1–12, 2017.

[87] S. A. Lambert, L. Gil, S. Jupp, S. C. Ritchie, Y. Xu, A. Buniello, G. Abraham, M. Chapman, H. Parkinson, J. Danesh, et al. The Polygenic Score Catalog: an open database for reproducibility and systematic evaluation. *medRxiv*, 2020.

[88] M. J. Landrum, J. M. Lee, M. Benson, G. R. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, W. Jang, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids research*, 46(D1):D1062–D1067, 2018.

[89] A. Latchford, S. Cohen, M. Auth, M. Scaillon, J. Viala, R. Daniels, C. Talbotec, T. Attard, C. Durno, and W. Hyer. Management of Peutz-Jeghers syndrome in children and adolescents: a position paper from the ESPGHAN Polyposis Working Group. *Journal of pediatric gastroenterology and nutrition*, 68(3):442–452, 2019.

[90] C. Ledergerber and C. Dessimoz. Base-calling for next-generation sequencing platforms. *Briefings in bioinformatics*, 12(5):489–497, 2011.

[91] A. Lee, N. Mavaddat, A. N. Wilcox, A. P. Cunningham, T. Carver, S. Hartley, C. B. de Villiers, A. Izquierdo, J. Simard, M. K. Schmidt, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine*, 21(8):1708–1718, 2019.

[92] M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. ODonnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616):285–291, 2016.

[93] I. U. Leong, A. Stuckey, D. Lai, J. R. Skinner, and D. R. Love. Assessment of the predictive accuracy of five *in silico* prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC medical genetics*, 16(1):1–13, 2015.

[94] L. Lepkes, M. Kayali, B. Blümcke, J. Weber, M. Suszynska, S. Schmidt, J. Borde, K. Klonowska, B. Wappenschmidt, J. Hauke, and et al. Performance of In Silico Prediction Tools for the Detection of Germline Copy Number Variations in Cancer Predisposition Genes in 4208 Female Index Patients with Familial Breast and Ovarian Cancer. *Cancers*, 13(1):118, Jan 2021.

[95] H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*, 2013.

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

[97] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.

[98] J. Lilyquist, H. LaDuca, E. Polley, B. T. Davis, H. Shimelis, C. Hu, S. N. Hart, J. S. Dolinsky, F. J. Couch, and D. E. Goldgar. Frequency of mutations in a large series of clinically ascertained ovarian cancer cases tested on multi-gene panels compared to reference controls. *Gynecologic oncology*, 147(2):375–380, 2017.

[99] J. Lilyquist, K. J. Ruddy, C. M. Vachon, and F. J. Couch. Common Genetic Variation and Breast Cancer Risk – Past, Present, and Future. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 27(4):380–394, 2018.

[100] L. A. Littlejohn, J. Gibbs, L. B. Jordan, Z. H. Miedzybrodzka, C. Bell, D. Goudie, J. Dunlop, and J. N. Berg. Assessing the effectiveness of NICE criteria for stratifying breast cancer risk in a UK cohort. *European Journal of Human Genetics*, 26(4):599–603, 2018.

[101] C. Loveday, C. Turnbull, E. Ramsay, D. Hughes, E. Ruark, J. R. Frankum, G. Bowden, B. Kalmyrzaev, M. Warren-Perry, K. Snape, et al. Germline mutations in *RAD51D* confer susceptibility to ovarian cancer. *Nature genetics*, 43(9):879–882, 2011.

[102] C. Loveday, C. Turnbull, E. Ruark, R. M. M. Xicola, E. Ramsay, D. Hughes, M. Warren-Perry, K. Snape, D. Eccles, D. G. Evans, et al. Germline *RAD51C* mutations confer susceptibility to ovarian cancer. *Nature genetics*, 44(5):475, 2012.

[103] H.-M. Lu, S. Li, M. H. Black, S. Lee, R. Hoiness, S. Wu, W. Mu, R. Huether, J. Chen, S. Sridhar, et al. Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA oncology*, 5(1):51–57, 2019.

[104] B. Luxembourg, D. Mathias, S. Körber, and E. Seifried. Prediction of the pathogenicity of antithrombin sequence variations by in silico methods. *Thrombosis research*, 135(2):404–409, 2015.

[105] M. Mahmoud, N. Gobet, D. I. Cruz-Dávalos, N. Mounier, C. Dessimoz, and F. J. Sedlazeck. Structural variant calling: the long and the short of it. *Genome biology*, 20(1):246, 2019.

[106] P. L. Mai, A. F. Best, J. A. Peters, R. M. DeCastro, P. P. Khincha, J. T. Loud, R. C. Bremer, P. S. Rosenberg, and S. A. Savage. Risks of first and subsequent cancers among *TP53* mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. *Cancer*, 122(23):3673–3681, 2016.

[107] M. Marabelli, S.-C. Cheng, and G. Parmigiani. Penetrance of *ATM* gene mutations in breast cancer: a meta-analysis of different measures of risk. *Genetic epidemiology*, 40(5):425–431, 2016.

[108] N. Mavaddat, S. Peock, D. Frost, S. Ellis, R. Platte, E. Fineberg, D. G. Evans, L. Izatt, R. A. Eeles, J. Adlard, et al. Cancer risks for *BRCA1* and *BRCA2* mutation carriers: results from prospective analysis of embrace. *JNCI: Journal of the National Cancer Institute*, 105(11):812–822, 2013.

[109] N. Mavaddat, P. D. Pharoah, K. Michailidou, J. Tyrer, M. N. Brook, M. K. Bolla, Q. Wang, J. Dennis, A. M. Dunning, M. Shah, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *JNCI: Journal of the National Cancer Institute*, 107(5), 2015.

[110] K. N. Maxwell, S. M. Domchek, K. L. Nathanson, and M. E. Robson. Population frequency of germline *BRCA1/2* mutations. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 34(34):4183, 2016.

[111] A. McIntosh, C. Shaw, G. Evans, N. Turnbull, N. Bahar, M. Barclay, D. Easton, J. Emery, J. Gray, J. Halpin, et al. Clinical guidelines and evidence review for the classification and care of women at risk of familial breast cancer. *London: National Collaborating Centre for Primary Care/University of Sheffield*, 2004.

[112] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. The Ensembl Variant Effect Predictor. *Genome biology*, 17(1):122, 2016.

[113] L.-A. McNutt, C. Wu, X. Xue, and J. P. Hafner. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American journal of epidemiology*, 157(10):940–943, 2003.

[114] P. Medvedev, M. Stanciu, and M. Brudno. Computational methods for discovering structural variation with next-generation sequencing. *Nature methods*, 6(11):S13–S20, 2009.

[115] A. Meindl, H. Hellebrand, C. Wiek, V. Erven, B. Wappenschmidt, D. Niederacher, M. Freund, P. Lichtner, L. Hartmann, H. Schaal, et al. Germline mutations in breast and ovarian cancer pedigrees establish *RAD51C* as a human cancer susceptibility gene. *Nature genetics*, 42(5):410–414, 2010.

[116] R. L. Mitchell, C. Kosche, K. Burgess, S. Wadhwa, L. Buckingham, R. Ghai, J. Rotmensch, O. Klapko, and L. Usha. Misdiagnosis of Li-Fraumeni syndrome in a patient with clonal hematopoiesis and a somatic *TP53* mutation. *Journal of the National Comprehensive Cancer Network*, 16(5):461–466, 2018.

[117] P. Møller, T. T. Seppälä, I. Bernstein, E. Holinski-Feder, P. Sala, D. G. Evans, A. Lindblom, F. Macrae, I. Blanco, R. H. Sijmons, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut*, 67(7):1306–1316, 2018.

[118] J. M. Moreno-Cabrera, J. Del Valle, E. Castellanos, L. Feliubadaló, M. Pineda, J. Brunet, E. Serra, G. Capellà, C. Lázaro, and B. Gel. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *European Journal of Human Genetics*, 28(12):1645–1655, 2020.

[119] National Institute for Health and Care Excellence. Familial breast cancer: classification, care and managing breast cancer and related risks in people with a family history of breast cancer. `https://www.nice.org.uk/guidance/cg164`. November 20, 2019.

[120] C. L. Neben, A. D. Zimmer, W. Stedden, J. van den Akker, R. O'Connor, R. C. Chan, E. Chen, Z. Tan, A. Leon, J. Ji, et al. Multi-gene panel testing of 23,179 individuals for hereditary cancer risk identifies pathogenic variant carriers missed by current genetic testing guidelines. *The Journal of Molecular Diagnostics*, 21(4):646–657, 2019.

[121] G. Neidhardt, J. Hauke, J. Ramser, E. Groß, A. Gehrig, C. R. Müller, A.-K. Kahlert, K. Hackmann, E. Honisch, D. Niederacher, et al. Association between loss-of-function mutations within the *FANCM* gene and early-onset familial breast cancer. *JAMA oncology*, 3(9):1245–1248, 2017.

[122] K. A. Nguyen, J. S. Syed, C. R. Espenschied, H. LaDuca, A. M. Bhagat, A. Suarez-Sarmiento, T. K. O'Rourke Jr, K. L. Brierley, E. W. Hofstatter, and B. Shuch. Advances in the diagnosis of hereditary kidney cancer: Initial results of a multigene panel test. *Cancer*, 123(22):4363–4371, 2017.

[123] B. M. Norquist, M. I. Harrell, M. F. Brady, T. Walsh, M. K. Lee, S. Gulsuner, S. S. Bernards, S. Casadei, Q. Yi, R. A. Burger, et al. Inherited mutations in women with ovarian carcinoma. *JAMA oncology*, 2(4):482–490, 2016.

[124] E. Oleary, D. Iacoboni, J. Holle, S. T. Michalski, E. D. Esplin, S. Yang, and K. Ouyang. Expanded gene panel use for women with breast cancer: identification and intervention beyond breast cancer risk. *Annals of surgical oncology*, 24(10):3060–3066, 2017.

[125] M. T. Parsons, E. Tudini, H. Li, E. Hahnen, B. Wappenschmidt, L. Feliubadaló, C. M. Aalfs, S. Agata, K. Aittomäki, E. Alducci, et al. Large scale multifactorial likelihood quantitative analysis of *BRCA1* and *BRCA2* variants: An ENIGMA resource to support clinical variant classification. *Human mutation*, 40(9):1557–1578, 2019.

[126] R. Pereira, J. Oliveira, and M. Sousa. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of clinical medicine*, 9(1):132, 2020.

[127] F. Pfeiffer, C. Gröber, M. Blank, K. Händler, M. Beyer, J. L. Schultze, and G. Mayer. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports*, 8(1):1–14, 2018.

[128] V. Plagnol, J. Curtis, M. Epstein, K. Y. Mok, E. Stebbings, S. Grigoriadou, N. W. Wood, S. Hambleton, S. O. Burns, A. J. Thrasher, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*, 28(21):2747–2754, 2012.

[129] S. E. Plon, D. M. Eccles, D. Easton, W. D. Foulkes, M. Genuardi, M. S. Greenblatt, F. B. Hogervorst, N. Hoogerbrugge, A. B. Spurdle, S. V. Tavtigian, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation*, 29(11):1282–1291, 2008.

[130] E. Pohl-Rescigno, J. Hauke, S. Loibl, V. Möbus, C. Denkert, P. A. Fasching, M. Kayali, C. Ernst, N. Weber-Lassalle, C. Hanusch, et al. Association of Germline Variant Status With Therapy Response in High-risk Early-Stage Breast Cancer: A Secondary Analysis of the GeparOcto Randomized Clinical Trial. *JAMA oncology*, 6(5):744–748, 2020.

[131] G. Povysil, A. Tzika, J. Vogt, V. Haunschmid, L. Messiaen, J. Zschocke, G. Klambauer, S. Hochreiter, and K. Wimmer. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Human mutation*, 38(7):889–897, 2017.

[132] H.-Q. Qu, X. Wang, L. Tian, and H. Hakonarson. Application of ACMG criteria to classify variants in the human gene mutation database. *Journal of human genetics*, 64(11):1091–1095, 2019.

[133] S. J. Ramus, H. Song, E. Dicks, J. P. Tyrer, A. N. Rosenthal, M. P. Intermaggio, L. Fraser, A. Gentry-Maharaj, J. Hayward, S. Philpott, et al. Germline mutations

in the *BRIP1*, *BARD1*, *PALB2*, and *NBN* genes in women with ovarian cancer. *JNCI: Journal of the National Cancer Institute*, 107(11), 2015.

[134] T. Rausch, T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, and J. O. Korbel. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18):i333–i339, 2012.

[135] T. R. Rebbeck, N. Mitra, F. Wan, O. M. Sinilnikova, S. Healey, L. McGuffog, S. Mazoyer, G. Chenevix-Trench, D. F. Easton, A. C. Antoniou, et al. Association of type and location of *BRCA1* and *BRCA2* mutations with risk of breast and ovarian cancer. *Jama*, 313(13):1347–1361, 2015.

[136] N. Resta, D. Pierannunzio, G. M. Lenato, A. Stella, R. Capocaccia, R. Bagnulo, P. Lastella, F. C. Susca, C. Bozzao, D. C. Loconte, et al. Cancer risk associated with *STK11/LKB1* germline mutations in Peutz–Jeghers syndrome patients: results of an Italian multicenter study. *Digestive and Liver Disease*, 45(7):606–611, 2013.

[137] C. Ricker, J. O. Culver, K. Lowstuter, D. Sturgeon, J. D. Sturgeon, C. R. Chanock, W. J. Gauderman, K. J. McDonnell, G. E. Idos, and S. B. Gruber. Increased yield of actionable mutations using multi-gene panels to assess hereditary cancer susceptibility in an ethnically diverse clinical cohort. *Cancer genetics*, 209(4):130–137, 2016.

[138] M. A. Rivas, M. Pirinen, D. F. Conrad, M. Lek, E. K. Tsang, K. J. Karczewski, J. B. Maller, K. R. Kukurba, D. S. DeLuca, M. Fromer, et al. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science*, 348(6235):666–669, 2015.

[139] Robert Koch-Institut and Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Krebs in Deutschland 2015/2016, 2019.

[140] M. E. Roberts, J. M. O. Ranola, M. L. Marshall, L. R. Susswein, S. Graceffo, K. Bohnert, G. Tsai, R. T. Klein, K. S. Hruska, and B. H. Shirts. Comparison of *CDH1* penetrance estimates in clinically ascertained families vs families ascertained for multiple gastric cancers. *JAMA oncology*, 5(9):1325–1331, 2019.

[141] C. Rodrigues, A. Santos-Silva, E. Costa, and E. Bronze-da Rocha. Performance of In Silico Tools for the Evaluation of *UGT1A1* Missense Variants. *Human mutation*, 36(12):1215–1225, 2015.

[142] R. Russo, I. Andolfo, F. Manna, A. Gambale, R. Marra, B. E. Rosato, P. Caforio, V. Pinto, P. Pignataro, K. Radhakrishnan, et al. Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. *American Journal of Hematology*, 93(5):672–682, 2018.

[143] E. Samorodnitsky, B. M. Jewell, R. Hagopian, J. Miya, M. R. Wing, E. Lyon, S. Damodaran, D. Bhatt, J. W. Reeser, J. Datta, et al. Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Human mutation*, 36(9):903–914, 2015.

[144] A. Y. Schmidt, T. v. Hansen, L. B. Ahlborn, L. Jønson, C. W. Yde, and F. C. Nielsen. Next-Generation Sequencing–Based Detection of Germline Copy Number Variations in *BRCA1/BRCA2*: Validation of a One-Step Diagnostic Workflow. *The Journal of Molecular Diagnostics*, 19(6):809–816, 2017.

[145] C. O. Schmidt and T. Kohlmann. When to use the odds ratio or the relative risk? *International journal of public health*, 53(3):165, 2008.

[146] M. K. Schmidt, F. Hogervorst, R. Van Hien, S. Cornelissen, A. Broeks, M. A. Adank, H. Meijers, Q. Waisfisz, A. Hollestelle, M. Schutte, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology*, 34(23):2750, 2016.

[147] V. A. Schneider, T. Graves-Lindsay, K. Howe, N. Bouk, H.-C. Chen, P. A. Kitts, T. D. Murphy, K. D. Pruitt, F. Thibaud-Nissen, D. Albracht, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*, 27(5):849–864, 2017.

[148] J. P. Schouten, C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic acids research*, 30(12):e57–e57, 2002.

[149] J. M. Schwarz, D. N. Cooper, M. Schuelke, and D. Seelow. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Meth*, 11(4):361–362, 2014.

[150] K. Schwarze, J. Buchanan, J. M. Fermont, H. Dreau, M. W. Tilley, J. M. Taylor, P. Antoniou, S. J. Knight, C. Camps, M. M. Pentony, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine*, 22(1):85–94, 2020.

[151] J. Shendure and H. Ji. Next-generation DNA sequencing. *Nature biotechnology*, 26(10):1135–1145, 2008.

[152] S. T. Sherry, M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29(1):308–311, 2001.

[153] T. P. Slavin, K. N. Maxwell, J. Lilyquist, J. Vijai, S. L. Neuhausen, S. N. Hart, V. Ravichandran, T. Thomas, A. Maria, D. Villano, et al. The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. *NPJ Breast Cancer*, 3(1):1–10, 2017.

[154] A. P. Sokolenko, D. R. Bulanova, A. G. Iyevleva, S. N. Aleksakhina, E. V. Preobrazhenskaya, A. O. Ivantsov, E. S. Kuligina, N. V. Mitiushkina, E. N. Suspitsin, G. A. Yanus, et al. High prevalence of *GPRC5A* germline mutations in *BRCA1*-mutant breast cancer patients. *International journal of cancer*, 134(10):2352–2358, 2014.

[155] T. Soussi, B. Leroy, and P. E. Taschner. Recommendations for Analyzing and Reporting *TP53* Gene Variants in the High-Throughput Sequencing Era. *Human mutation*, 35(6):766–778, 2014.

[156] A. B. Spurdle, S. Greville-Heygate, A. C. Antoniou, M. Brown, L. Burke, M. De La Hoya, S. Domchek, T. Dörk, H. V. Firth, A. N. Monteiro, et al. Towards controlled terminology for reporting germline cancer susceptibility variants: an ENIGMA report. *Journal of medical genetics*, 56(6):347–357, 2019.

[157] A. Stalke, B. Skawran, B. Auber, T. Illig, B. Schlegelberger, N. Junge, I. Goldschmidt, C. Leiskau, N. von Neuhoff, U. Baumann, et al. Diagnosis of monogenic liver diseases in childhood by next-generation sequencing. *Clinical genetics*, 93(3):665–670, 2018.

[158] P. D. Stenson, M. Mort, E. V. Ball, M. Chapman, K. Evans, L. Azevedo, M. Hayden, S. Heywood, D. S. Millar, A. D. Phillips, et al. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Human Genetics*, pages 1–11, 2020.

[159] L. P. Sugrue and R. S. Desikan. What are polygenic scores and why are they important? *Jama*, 321(18):1820–1821, 2019.

[160] M.-H. Tan, J. L. Mester, J. Ngeow, L. A. Rybicki, M. S. Orloff, and C. Eng. Lifetime cancer risks in individuals with germline *PTEN* mutations. *Clinical Cancer Research*, 18(2):400–407, 2012.

[161] H. Tang and P. D. Thomas. Tools for predicting the functional impact of nonsynonymous genetic variation. *Genetics*, 203(2):635–647, 2016.

[162] L. Tattini, R. D'Aurizio, and A. Magi. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Frontiers in Bioengineering and Biotechnology*, 3:92, 2015.

[163] S. V. Tavtigian, P. B. Samollow, D. d. Silva, and A. Thomas. An Analysis of Unclassified Missense Substitutions in Human BRCA1. *Familial Cancer*, 5(1):77–88, 2006.

[164] J. K. Teer and J. C. Mullikin. Exome sequencing: the sweet spot before whole genomes. *Human molecular genetics*, 19(R2):R145–R151, 2010.

[165] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.

[166] D. Thompson, D. Easton, and Breast Cancer Linkage Consortium. Variation in cancer risks, by mutation position, in *BRCA2* mutation carriers. *The American Journal of Human Genetics*, 68(2):410–419, 2001.

[167] D. Thompson, D. Easton, and Breast Cancer Linkage Consortium. Variation in *BRCA1* cancer risks by mutation position. *Cancer Epidemiology and Prevention Biomarkers*, 11(4):329–336, 2002.

[168] A. Torkamani, N. E. Wineinger, and E. J. Topol. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, 19(9):581–590, 2018.

[169] L. A. Torre, B. Trabert, C. E. DeSantis, K. D. Miller, G. Samimi, C. D. Runowicz, M. M. Gaudet, A. Jemal, and R. L. Siegel. Ovarian cancer statistics, 2018. *CA: a cancer journal for clinicians*, 68(4):284–296, 2018.

[170] A. Toss, C. Tomasello, E. Razzaboni, G. Contu, G. Grandi, A. Cagnacci, R. J. Schilder, and L. Cortesi. Hereditary ovarian cancer: not only *BRCA* 1 and 2 genes. *BioMed research international*, 2015, 2015.

[171] C. Turnbull, A. Sud, and R. S. Houlston. Cancer genetics, precision prevention and a call to action. *Nature genetics*, 50(9):1212–1218, 2018.

[172] S. Turner, L. L. Armstrong, Y. Bradford, C. S. Carlson, D. C. Crawford, A. T. Crenshaw, M. de Andrade, K. F. Doheny, J. L. Haines, G. Hayes, et al. Quality control procedures for genome-wide association studies. *Current protocols in human genetics*, 68(1):1–19, 2011.

[173] A. Tutt, H. Tovey, M. C. U. Cheang, S. Kernaghan, L. Kilburn, P. Gazinska, J. Owen, J. Abraham, S. Barrett, P. Barrett-Lee, et al. Carboplatin in *BRCA1/2*-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. *Nature medicine*, 24(5):628–637, 2018.

[174] E. Uusitalo, R. A. Kallionpää, S. Kurki, M. Rantanen, J. Pitkäniemi, P. Kronqvist, P. Härkönen, R. Huovinen, O. Carpen, M. Pöyhönen, et al. Breast cancer in neurofibromatosis type 1: overrepresentation of unfavourable prognostic factors. *British journal of cancer*, 116(2):211–217, 2017.

[175] E. Uusitalo, M. Rantanen, R. A. Kallionpää, M. Pöyhönen, J. Leppävirta, H. Ylä-Outinen, V. M. Riccardi, E. Pukkala, J. Pitkäniemi, S. Peltonen, et al. Distinctive cancer associations in patients with neurofibromatosis type 1. *Journal of Clinical Oncology*, 34(17):1978–1986, 2016.

[176] M. Vasimuddin, S. Misra, H. Li, and S. Aluru. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 314–324. IEEE, 2019.

[177] S. Vaughan, J. I. Coward, R. C. Bast, A. Berchuck, J. S. Berek, J. D. Brenton, G. Coukos, C. C. Crum, R. Drapkin, D. Etemadmoghadam, et al. Rethinking ovarian cancer: recommendations for improving outcomes. *Nature Reviews Cancer*, 11(10):719–725, 2011.

[178] T. Walsh, S. Casadei, M. K. Lee, C. C. Pennil, A. S. Nord, A. M. Thornton, W. Roeb, K. J. Agnew, S. M. Stray, A. Wickramanayake, et al. Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proceedings of the National Academy of Sciences*, 108(44):18032–18037, 2011.

[179] X. Wang, J. K. Teer, R. N. Tousignant, A. M. Levin, D. Boulware, D. A. Chitale, B. M. Shaw, Z. Chen, Y. Zhang, J. O. Blakeley, et al. Breast cancer risk and germline genomic profiling of women with neurofibromatosis type 1 who developed breast cancer. *Genes, Chromosomes and Cancer*, 57(1):19–27, 2018.

[180] A. Warr, C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson. Exome sequencing: current and future perspectives. *G3: Genes, Genomes, Genetics*, 5(8):1543–1550, 2015.

[181] K. Weber-Lassalle, P. Harter, J. Hauke, C. Ernst, S. Kommoss, F. Marmé, N. Weber-Lassalle, K. Prieske, D. Dietrich, J. Borde, et al. Diagnosis of Li-Fraumeni Syndrome: Differentiating *TP53* germline mutations from clonal hematopoiesis: Results of the observational AGO-TR1 trial. *Human mutation*, 39(12):2040–2046, 2018.

[182] N. Weber-Lassalle, J. Borde, K. Weber-Lassalle, J. Horváth, D. Niederacher, N. Arnold, S. Kaulfuß, C. Ernst, V. G. Paul, E. Honisch, et al. Germline loss-of-function variants in the *BARD1* gene are associated with early-onset familial breast cancer but not ovarian cancer. *Breast Cancer Research*, 21(1):55, 2019.

[183] N. Weber-Lassalle, J. Hauke, J. Ramser, L. Richters, E. Gross, B. Blümcke, A. Gehrig, A.-K. Kahlert, C. R. Müller, K. Hackmann, et al. *BRIP1* loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer. *Breast Cancer Research*, 20(1):7, 2018.

[184] R. M. Xicola, S. Li, N. Rodriguez, P. Reinecke, R. Karam, V. Speare, M. H. Black, H. LaDuca, and X. Llor. Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. *Journal of Medical Genetics*, 56(12):838–843, 2019.

[185] C. Xu. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Computational and structural biotechnology journal*, 16:15–24, 2018.

[186] S. Yadav and F. J. Couch. Germline genetic testing for breast cancer risk: The past, present, and future. *American Society of Clinical Oncology Educational Book*, 39:61–74, 2019.

[187] T. Yanes, M.-A. Young, B. Meiser, and P. A. James. Clinical applications of polygenic breast cancer risk: a critical review and perspectives of an emerging field. *Breast Cancer Research*, 22(1):1–10, 2020.

[188] X. Yang, G. Leslie, A. Doroszuk, K. Aittomäki, C. Blomqvist, T. Heikkinen, H. Nevanlinna, M. Tischkowitz, et al. Cancer risks associated with germline *PALB2* pathogenic variants: an international study of 524 families. *Journal of Clinical Oncology*, 2020.

[189] X. Yang, H. Song, G. Leslie, C. Engel, E. Hahnen, B. Auber, J. Horváth, K. Kast, D. Niederacher, C. Turnbull, et al. Ovarian and breast cancer risks associated with pathogenic variants in *RAD51C* and *RAD51D*. *JNCI: Journal of the National Cancer Institute*, 2020.

[190] R. Yao, C. Zhang, T. Yu, N. Li, X. Hu, X. Wang, J. Wang, and Y. Shen. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Molecular cytogenetics*, 10(1):1–7, 2017.

[191] S. Yohe and B. Thyagarajan. Review of clinical next-generation sequencing. *Archives of pathology & laboratory medicine*, 141(11):1544–1557, 2017.

[192] S. E. Yost, H. Alakus, H. Matsui, R. B. Schwab, K. Jepsen, K. A. Frazer, and O. Harismendy. Mutascope: sensitive detection of somatic mutations from deep amplicon sequencing. *Bioinformatics*, 29(15):1908–1909, 2013.

9 Acknowledgements/Danksagung

First of all, I would like to thank my supervisor PD Dr. Eric Hahnen, who has always guided me with his knowledge and advice, and has been extremely supportive while I have been working on this dissertation. I am especially grateful for his excellent and extremely helpful suggestions during the preparation of this thesis and whenever I got stuck with my data. I think that I will never get a better boss in my life. The same way I would like to thank Prof. Dr. Rita K. Schmutzler for her support, and for giving me the opportunity and confidence to contribute my expertise in an interesting and exciting research field with a lot of personal responsibility. The fact that I was able to contribute my own ideas and had the freedom to learn and try out new things has always motivated me in my work. I am very appreciative of this.

My sincere thanks go to Prof. Dr. Andreas Beyer and Prof. Dr. Michael Nothnagel for their willingness to serve as my PhD tutors, their support, and especially for their critical questions during our tutor meetings, which brought me one step further each time.

In the research group of the Center for Familial Breast and Ovarian Cancer, I have found colleagues who take me as I am and with whom I have been able to collaborate on a variety of interesting and challenging research projects. Both were certainly not always easy and for both I am extremely grateful. Together, we have torn down the boundaries between our areas of expertise and we have even brought most of our research projects to a conclusion. Awesome! First of all, Julika Borde, thank you for getting us to London and back twice as a team, and for always keeping calm, smart and dependable. ... and your help with any Windows applications. I am looking forward to our next projects!

Thanks to Dr. Jan Hauke for fruitful collaborations and sharing his BC/OC knowledge. I thank Mohamad Kayali for always doing the wet lab work for my projects so efficiently and reliably, and our interesting discussions. I thank Dr. Guido Neidhardt for his support during my first time at the Center for Familial Breast and Ovarian Cancer. I thank Dr. Esther Pohl-Rescigno for her enthusiasm, for always being willing to share her knowledge with me, for so often taking the time to explain and for her encouragements. I thank Sandra Schmidt for her helpfulness in every situation, and for allowing me to benefit from her laboratory experience. I thank Konstantin Weber-Lassalle for our fruitful collaboration in the CH projects and that he brought me a little closer to the Rhenish way of life. I would like to thank the colleagues from molecular diagnostics for our succesful cooperation across divisions, for helpful explanations and straightforward exchange, especially Dr. Britta Blümcke and Dr. Nana Weber-Lassalle.

I thank Prof. Dr. Andreas Beyer and Prof. Dr. Achim Tresch for their helpful suggestions during the initial development of OPaCNAT. Furthermore, I thank Dr. Karoline Kuchenbaecker, Dr. Olga Giannakopoulou and Dr. Stefanie Mueller for their advice

and for amazing and educational weeks at the UCL. Thanks also to Prof. Dr. Kathrin Möllenhoff for her advice regarding the statistical anaylsis of our CH data.

This thesis is a milestone for me that I have been working towards for a long time, so I would also like to thank all the people who shared their expertise with me on my career path before my time in Cologne and from whom I was able to learn so much, in particular Prof. Dr. Burkhard Morgenstern (University of Göttingen), Prof. Dr. Ellen Baake (Bielefeld University), and from the Genome Informatics group of the University of Duisburg-Essen Dr. Johannes Köster and Dr. Christopher Schröder.

Last but not least, I want to thank the most important people in life: family and friends. Liebe Mutti, ich werde nie verstehen, wie du es geschafft hast, mit zwei kleinen Kindern zu promovieren. Ich danke dir für deine Unterstützung und dafür, dass du mich immer bestärkt hast. Ingmar, I will never be able to make up for the fact that you moved to nasty NRW because of me. Thank you for sharing your life and chili cheese nuggets with me and for constantly reminding me that there is also a life outside of BC/OC research. I thank my sister Alexandra for believing in me and her encouragements, und ich danke meinem Vater und Fred für die lebenslange Unterstützung. Many thanks to Philipp, Janika, and Tobias for feeding my head with things other than bioinformatics, for always willingly listening to my stories from everyday research life and for always believing in me. Till, merci pour nos pauses déjeuner ensemble et vos commentaires et remarques utiles sur mon travail.

10 Erklärung zur Dissertation

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten – noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen:

J Borde, C Ernst, B Wappenschmidt, et al. **Performance of Breast Cancer Polygenic Risk Scores in 760 Female *CHEK2* Germline Mutation Carriers.** Journal of the National Cancer Institute, 2020. doi: 10.1093/jnci/djaa203. Online ahead of print.

C Ernst, E Hahnen, C Engel, et al. **Performance of *in silico* prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics.** BMC medical genomics, 11(1):35, 2018.

J Hauke, E Hahnen, S Schneider, et al. **Deleterious somatic variants in 473 consecutive individuals with ovarian cancer: results of the observational AGO-TR1 study (NCT02222883).** Journal of Medical Genetics, 56(9):574-580, 2019.

J Hauke, P Harter, C Ernst, et al. **Sensitivity and specificity of loss of heterozygosity analysis for the classification of rare germline variants in BRCA1/2: Results of the observational AGO-TR1 study (NCT02222883).** Journal of Medical Genetics, 2020. doi: 10.1136/jmedgenet-2020-107353. Online ahead of print.

J Hauke, J Horváth, E Groß, et al. **Gene panel testing of 5589 *BRCA1/2*-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer.** Cancer Medicine, 7(4):1349-1358, 2018.

K Klaschik, J Hauke, G Neidhardt, et al. **The *GPRC5A* frameshift variant c.183del is not associated with increased breast cancer risk in *BRCA1* mutation carriers.** International Journal of Cancer, 144(7):1761-1763, 2019.

L Lepkes, M Kayali, B Blümcke, et al. **Performance of In Silico Prediction Tools for the Detection of Germline Copy Number Variations in Cancer Predisposition Genes in 4208 Female Index Patients with Familial Breast and Ovarian Cancer** Cancers, 13(1):118, 2021.

G Neidhardt, J Hauke, J Ramser, et al. **Association Between Loss-of-Function Mutations Within the *FANCM* Gene and Early-Onset Familial Breast Cancer.** JAMA Oncology, 3(9):1245-1248, 2017.

E Pohl-Rescigno, J Hauke, S Loibl, et al. **Association of Germline Variant Status With Therapy Response in High-risk Early-Stage Breast Cancer: A Secondary Analysis of the GeparOcto Randomized Clinical Trial.** JAMA Oncology, 6(5):1-5, 2020.

K Weber-Lassalle, P Harter, J Hauke, et al. **Diagnosis of Li-Fraumeni Syndrome: Differentiating *TP53* germline mutations from clonal hematopoiesis: Results of the observational AGO-TR1 trial.** Human Mutation, 39(12):2040-2046, 2018.

N Weber-Lassalle, J Borde, K Weber-Lassalle, et al. **Germline loss-of-function variants in the *BARD1* gene are associated with early-onset familial breast cancer but not ovarian cancer.** Breast Cancer Research, 21(1):55, 2019.

N Weber-Lassalle, J Hauke, J Ramser, et al. ***BRIP1* loss-of-function mutations confer high risk for familial ovarian cancer, but not familial breast cancer.** Breast Cancer Research, 20(1):7, 2018.

Köln, April 2021

Corinna Ernst