

Abstract

In Germany, a familial predisposition is suspected in about 30% of all cases of breast cancer (BC) and/or ovarian cancer (OC). BC and/or OC are defined as familial if patients meet the inclusion criteria of the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) for genetic germline testing [1] (**Table 1**). Pathogenic variants in *BRCA1/2* genes (MIM*113705, MIM*600185) are identified in about 24% of familial BC and/or OC cases [2]. In about 6% of familial BC and/or OC cases, germline pathogenic variants can be detected in additional, mostly moderate penetrant risk genes such as *CHEK2* (MIM*604373) [3] or *RAD51C* (MIM*602774) [4] which predispose for BC and/or OC [5, 6]. However, about 70% of the familial BC and OC cases currently cannot be explained by pathogenic variants in the established high-risk genes *BRCA1/2* or the additional known risk genes. Therefore, the identification as well as the characterization of novel BC and/or OC susceptibility genes is essential. Since studies with smaller sample sizes were not able to identify/validate additional risk genes so far, we hypothesized that the prevalence of pathogenic variants of new candidate BC risk genes will be very low, precisely because most of the non-*BRCA1/2* risk genes have a pathogenic variant prevalence <1% in familial BC patients [5]. The establishment of genotype-phenotype correlations of new candidate BC risk genes is challenging due to low pathogenic variant prevalences in these genes. Therefore, large cohorts and international cooperations are required to achieve statistically significant results after Bonferroni correction for multiple testing.

We used two approaches to identify new candidate BC and/or OC risk genes. Firstly, we performed a hypothesis-driven candidate gene approach, for which we investigated familial BC and OC index patients for protein truncating variants (PTVs) and potentially damaging missense variants within *BARD1* (MIM*601593) and *BRIP1* (MIM*605882). The encoded proteins of these genes interact with *BRCA1* and play an important role in DNA double-strand break (DSB) repair as it is known for the high-risk BC/OC genes *BRCA1* and *BRCA2* [7] (**Figure 6**). The DNA samples of familial BC and OC index patients were provided by centers of the GC-HBOC and collected in our DNA-biobank in Cologne, while the samples of geographically-matched female control individuals (GMCs) were provided by an epidemiologic health care study group of PD Dr. A. Volk, the Lifestyle Interventions and Independence for Elders (LIFE) study group¹ and the Gutenberg study group².

In order to investigate the role of *BARD1* in BC development, we analyzed 4,469 female familial BC index patients and 2,767 GMCs. A significant association between heterozygous germline *BARD1* PTVs and early-onset familial BC (odds ratio (OR)=12.04; 95% confidence interval (CI)=5.78–25.08; P<0.00001) was observed. Furthermore, we tested 451 familial OC index

¹<http://life.uni-leipzig.de>

²<http://www.gutenberghealthstudy.org>

patients and did not identify any PTV within *BARD1*. Consequently, *BARD1* seems to be a high to moderate risk gene for early-onset BC but is not associated with OC [8].

Genetic analysis of *BRIP1* revealed contrary results. Investigation of 706 OC index patients and 2,189 GMCs resulted in a significant association between *BRIP1* PTVs and late-onset OC (OR=29.91; 95%CI=14.99–59.66; P<0.0001). The analysis of 5,668 familial BC index patients revealed a slight association of *BRIP1* PTVs with familial BC, but the results did not reach the required levels of statistical significance (OR=1.81; 95%CI=1.00–3.30; P=0.0623). Therefore, we were able to show that *BRIP1* is a high-risk gene for late-onset OC, while an association with BC is still a matter of debate [9].

Secondly, we performed an exome-wide approach for the identification of new candidate BC risk genes and conducted a large case-control study (PERSPECTIVE) in cooperation with international collaborators (from Canada, the U.S., the U.K.). Gene selection was based on whole exome sequencing (WES) data of ~1,500 BC patients, of which 870 were provided by the GC-HBOC. Bioinformatic analyses performed by work groups from Canada (Toronto, Quebec), the U.S. (Utah), the U.K. (Cambridge), and Germany (Munich, Cologne) led to the selection of 216 new candidate genes. A common targeted sequencing Agilent SureSelect gene panel (QXT protocol) was designed covering all 216 new candidate genes and 35 validated and suspected BC risk genes from the BRIDGES study (251 genes in total). For the PERSPECTIVE project, we analyzed 3,199 familial index cases with BC (1-11 further relatives with BC) from Germany and without a personal history of OC, which were provided by centers of the GC-HBOC and collected in our DNA-biobank in Cologne. Additionally, we collected and sequenced a total of 2,767 GMCs. Data from familial BC index patients, GMCs and the Exome Aggregation Consortium (ExAC) were compared to identify new candidate genes associated with BC phenotype. Within the PERSPECTIVE study we were able to identify five genes (*MARCO* (MIM*604870), *PIK3CG* (MIM*601232), *RNF181* (MIM*612490), *TMEM161A* (no MIM*), *ZWINT* (MIM*609177)), that showed a significantly higher PTV prevalence in familial BC index patients compared to controls. In comparison to ExAC, the ORs ranged from 4.26 to 8.02 and p-values ranged from 0.00002 to 0.00541, whereby *MARCO* reached a significance level after Bonferroni correction for multiple testing (p-treshold ≤ 0.0002) (**Table 4**). Additionally, we identified six genes (*BPTF* (MIM*601819), *C4ORF48* (MIM*614690), *CTBP1* (MIM*602618), *FHL3* (MIM*602790), *NTHL1* (MIM*602656), *SLC2A8* (MIM*8605245)), that showed a significantly higher number of potentially damaging missense variants in familial BC index patients compared to controls. Five of them reached a significance level after Bonferroni correction for multiple testing (p-treshold ≤ 0.0002) (**Table 4**). Some of the new candidate BC risk genes have been described to be associated with cancer, while some have not yet been investigated for an association with cancer. Segregation analyses and functional studies are

required to characterize the new candidate BC risk genes and to identify novel signaling pathways involved in cancer predisposition.

Additional investigations during the PhD thesis showed that *BRCA1/2* testing should be considered before starting BC treatment, because chemotherapy is effective for *BRCA1/2*-mutated triple-negative BC (TNBC) and human epidermal growth factor receptor 2 (HER2)-negative/hormone receptor (HR)-positive BC [10]. We excluded *GPRC5A* (MIM*604138) and also *NBN* (MIM*602667) as BC risk genes, but confirmed a significant role of *ATM* (MIM*607585), *CDH1* (MIM*192090), *CHEK2* (MIM*604373), *PALB2* (MIM*61035) and *TP53* (MIM*191170) in BC [11, 12]. Associations between these genes and BC subtypes were discovered [12]. Furthermore, in a study with 523 OC patients (AGO-TR1 trial, NCT02222883), we demonstrated that in three out of six *TP53*-positive cases the deleterious *TP53* variants were not the cause of the patients' cancer. Considering that *TP53* mutations may arise from chemotherapy-induced clonal hematopoiesis, we suggest that conspicuous *TP53* test results in patients who received chemotherapy prior to blood draw should be supplemented by additional tissue tests, excluding hematopoietic compartment [13].

Zusammenfassung

Bei etwa 30% aller Brust- und/oder Eierstockkrebsfälle (breast cancer, BC; ovarian cancer, OC) in Deutschland wird eine familiäre Prädisposition vermutet. BC- und/oder OC-Erkrankungen werden als familiär definiert, wenn PatientInnen die Einschlusskriterien des Deutschen Konsortiums für Erblichen Brust- und Eierstockkrebs (GC-HBOC) für genetische Keimbahntestung erfüllen [1] (**Tabelle 1**). In etwa 24% der familiären BC- und/oder OC-Fälle wurden pathogene Varianten in den *BRCA1/2*-Genen (MIM*113705, MIM*600185) identifiziert [2]. Pathogene Varianten in zusätzlichen, meist mäßig-penetranten Risikogenen wie *CHEK2* (MIM*604373) [3] oder *RAD51C* (MIM*602774) [4], die für BC und/oder OC prädisponieren, können in etwa 6% der familiären BC- und/oder OC-Fälle nachgewiesen werden [5, 6]. Jedoch lassen sich etwa 70% der familiären BC- und OC-Fälle derzeit nicht durch pathogene Varianten in den etablierten Hochrisikogenen *BRCA1/2* oder den zusätzlichen bekannten Risikogenen erklären. Daher ist die Identifizierung sowie die Charakterisierung neuer BC- und/oder OC-Risikogene unerlässlich. Da Studien mit kleineren Stichprobengrößen bisher nicht in der Lage waren, zusätzliche Risikogene zu identifizieren/zu validieren, haben wir die Hypothese aufgestellt, dass die Prävalenz pathogener Varianten neuer Kandidaten für BC-Risikogene sehr gering sein wird, gerade weil die meisten der Nicht-*BRCA1/2*-Risikogene eine Prävalenz von pathogenen Varianten <1% in familiären BC-PatientInnen aufweisen [5]. Die Etablierung von Genotyp-Phänotyp-Korrelationen neuer Kandidaten-BC-Risikogene ist aufgrund der niedrigen Anzahl an pathogenen Varianten in diesen Genen eine Herausforderung. Daher sind große Kohorten und internationale Kooperationen erforderlich, um statistisch signifikante Ergebnisse nach der Bonferroni-Korrektur für die Mehrfachtestung zu erzielen.

Wir haben zwei Ansätze angewendet, um neue BC- und/oder OC-Risikogene zu identifizieren. Zum einen verfolgten wir einen hypothesengetriebenen Kandidatengen-Ansatz, bei dem wir familiäre BC- und OC-IndexpatientInnen auf proteintrunkierende-Varianten (PTVs) und potentiell Proteinfunktion beeinträchtigende missense-Varianten innerhalb von *BARD1* (MIM*601593) und *BRIP1* (MIM*605882) untersucht haben. Die kodierten Proteine dieser Gene interagieren mit *BRCA1* und spielen wie die Hochrisikogene *BRCA1* und *BRCA2* eine wichtige Rolle bei der DNA-Doppelstrangbruch (DSB)-Reparatur [7] (**Abbildung 6**). Die DNA-Proben von familiären BC- und OC-IndexpatientInnen wurden von Zentren des GC-HBOC zur Verfügung gestellt und in unserer DNA-Biobank in Köln gesammelt, während die Proben von geographisch passenden weiblichen Kontrollpersonen (GMCs) von einer epidemiologischen Studiengruppe des Gesundheitswesens von PD Dr. A. Volk, der Lifestyle Interventions and

Independence for Elders (LIFE) Studiengruppe³ und der Gutenberg Studiengruppe⁴ zur Verfügung gestellt wurden.

Um die Rolle von *BARD1* in der Entwicklung des BC zu untersuchen, analysierten wir 4.469 weibliche familiäre BC-IndexpatientInnen und 2.767 GMCs. Dabei wurde eine signifikante Assoziation zwischen heterozygoten *BARD1*-PTVs und früh einsetzenden familiärem BC beobachtet (Odds Ratio (OR)=12,04; 95% Konfidenzintervall (CI)=5,78-25,08; P<0,00001). Darüber hinaus testeten wir 451 familiäre OC-IndexpatientInnen und identifizierten keine PTV innerhalb von *BARD1*. Folglich scheint es sich bei *BARD1* um ein Hochrisikogen für den frühen Beginn von familiären BC zu handeln, das keine Assoziation mit OC zeigt [8].

Gegenteilige Ergebnisse ergab die genetische Analyse von *BRIP1*. Die Untersuchung von 706 OC-IndexpatientInnen und 2.189 GMCs zeigte eine signifikante Assoziation zwischen *BRIP1*-PTVs und spät einsetzendem OC (OR=29,91; 95%CI=14,99-59,66; P<0,0001). Die Analyse von 5.668 familiären BC-IndexpatientInnen ergab eine leichte Assoziation von *BRIP1*-PTVs mit familiärem BC, allerdings erreichten die Ergebnisse nicht die erforderliche statistische Signifikanz (OR=1,81; 95%CI=1,00-3,30; P=0,0623). Daher konnten wir zeigen, dass *BRIP1* ein Hochrisikogen für spät einsetzenden OC ist, während eine Assoziation mit BC noch umstritten ist [9].

Zum anderen haben wir einen exomweiten Ansatz zur Identifizierung neuer Kandidaten für BC-Risikogene durchgeführt und in Zusammenarbeit mit internationalen Kollaborationen (aus Kanada, den USA, Großbritannien) eine große Fall-Kontroll-Studie (PERSPECTIVE) durchgeführt. Die Genauswahl für dieses Projekt basierte auf Exomsequenzierungsdaten von ~1.500 BrustkrebspatientInnen, von denen 870 vom GC-HBOC bereitgestellt wurden. Bioinformatische Analysen von Arbeitsgruppen aus Kanada (Toronto, Quebec), den USA (Utah), Großbritannien (Cambridge) und Deutschland (München, Köln) führten zur Auswahl von 216 neuen Kandidatengen. Ein geläufiges Agilent SureSelect Genpanel für die gezielte Sequenzierung (QXT-Protokoll) wurde entwickelt, das alle 216 neuen Kandidatengene und 35 validierte und bestätigte Brustkrebsrisikogene aus der BRIDGES-Studie abdeckt (insgesamt 251 Gene). Für das PERSPECTIVE Projekt haben wir 3.199 familiäre Indexfälle mit BC (1-11 weitere Verwandte mit BC) aus Deutschland und ohne persönliche Vorgeschichte von OC analysiert, die von Zentren des GC-HBOC zur Verfügung gestellt und in unserer DNA-Biobank in Köln gesammelt wurden. Zusätzlich haben wir insgesamt 2.767 GMCs gesammelt und sequenziert. Daten von familiären BC-IndexpatientInnen, GMCs und Exome Aggregation Consortium (ExAC) wurden verglichen, um neue Kandidatengene zu identifizieren, die mit dem BC-Phänotyp assoziiert sind. Im Rahmen der PERSPECTIVE-Studie konnten wir fünf Gene (*MARCO* (MIM*604870), *PIK3CG* (MIM*601232), *RNF181* (MIM*612490), *TMEM161A* (keine

³<http://life.uni-leipzig.de>

⁴<http://www.gutenberghealthstudy.org>

MIM*), *ZWINT* (MIM*609177)) identifizieren, die eine signifikant höhere PTV-Prävalenz bei familiären BC-IndexpatientInnen im Vergleich zu den Kontrollen zeigten. Im Vergleich zu ExAC reichten die ORs von 4,26 bis 8,02 und die p-Werte von 0,00002 bis 0,00541, wobei MARCO ein Signifikanzniveau nach Bonferroni-Korrektur für die Mehrfachtestung erzielte (Schwellenwert $P \leq 0,0002$) (**Tabelle 4**). Darüber hinaus identifizierten wir sechs Gene (*BPTF* (MIM*601819), *C4ORF48* (MIM*614690), *CTBP1* (MIM*602618), *FHL3* (MIM*602790), *NTHL1* (MIM*602656), *SLC2A8* (MIM*8605245)), die bei familiären BC-IndexpatientInnen eine signifikant erhöhte Anzahl an missense-Varianten aufwiesen verglichen mit den Kontrollen. Diese missense-Varianten beeinträchtigen potenziell die Proteinfunktion. Fünf von ihnen erreichten ein Signifikanzniveau nach Bonferroni-Korrektur für die Mehrfachtestung (Schwellenwert $P \leq 0,0002$) (**Tabelle 4**). Einige der Risikogenkandidaten wurden bisher schon mit Krebs assoziiert, während andere noch nicht auf eine Assoziation mit Krebs hin untersucht worden sind. Segregationsanalysen und funktionelle Studien sind erforderlich, um die neuen Brustkrebsrisikogenkandidaten zu charakterisieren und neue Signalwege zu identifizieren, die an der Brustkrebsprädisposition beteiligt sind.

Zusätzliche Untersuchungen während der Doktorarbeit zeigten, dass eine *BRCA1/2*-Testung vor Behandlungsbeginn von BC in Betracht gezogen werden sollte, da eine Chemotherapie bei *BRCA1/2*-mutiertem triple-negative BC (TNBC) und human epidermal growth factor receptor 2 (HER2) negativen/hormone receptor (HR) positiven BC wirksam ist [10]. Wir schlossen *GPRC5A* (MIM*604138) und auch *NBN* (MIM*602667) als Risikogene für BC aus, bestätigten aber eine signifikante Rolle von *ATM* (MIM*607585), *CDH1* (MIM*192090), *CHEK2* (MIM*604373), *PALB2* (MIM*61035) und *TP53* (MIM*191170) bei BC [11, 12]. Assoziationen zwischen diesen Genen und den BC-Subtypen wurden entdeckt [12]. Darüber hinaus konnten wir in einer Studie mit 523 OC-PatientInnen (AGO-TR1-Studie, NCT02222883) zeigen, dass in drei von sechs *TP53*-positiven Fällen die trunkierenden *TP53*-Varianten nicht ursächlich für den Krebs der PatientInnen waren. In Anbetracht der Tatsache, dass *TP53*-Mutationen durch chemotherapieinduzierte klonale Hämatopoese entstehen können, schlagen wir vor, dass auffällige *TP53*-Testergebnisse bei PatientInnen, die vor der Blutabnahme eine Chemotherapie erhielten, durch zusätzliche Gewebetests ergänzt werden sollten, wobei das hämatopoetische Kompartiment ausgeschlossen werden sollte [13].