University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Dumont, M;Weber-Lassalle, N;Joly-Beauparlant, C;Ernst, C;Droit, A;Feng, BJ;Dubois, S;Collin-Deschesnes, AC;Soucy, P;Vallée, M;Fournier, F;Lemaçon, A;Adank, MA;Allen, J;Altmüller, J;Arnold, N;Ausems, MGEM;Berutti, R;Bolla, MK;Bull, S;Carvalho, S;Cornelissen, S;Dufault, MR;Dunning, AM;Engel, C;Gehrig, A;Geurts-Giele, WRR;Gieger, C;Green, J;Hackmann, K;Helmy, M;Hentschel, J;Hogervorst, FBL;Hollestelle, A;Hooning, MJ;Horváth, J;Ikram, MA;Kaulfuß, S;Keeman, R;Kuang, D;Luccarini, C;Maier, W;Martens, JWM;Niederacher, D;Nürnberg, P;Ott, CE;Peters, A;Pharoah, PDP;Ramirez, A;Ramser, J;Riedel-Heller, S;Schmidt, G;Shah, M;Scherer, M;Stäbler, A;Strom, TM;Sutter, C;Thiele, H;van Asperen, CJ;van der Kolk, L;van der Luijt, RB;Volk, AE;Wagner, M;Waisfisz, Q;Wang, Q;Wang-Gohrke, S;Weber, BHF;Devilee, P;Tavtigian, S;Bader, GD;Meindl, A;Goldgar, DE;Andrulis, IL;Schmutzler, RK;Easton, DF;Schmidt, MK;Hahnen, E;Simard, J

Title:
Uncovering the Contribution of Moderate-Penetrance Susceptibility Genes to Breast Cancer by Whole-Exome Sequencing and Targeted Enrichment Sequencing of Candidate Genes in Women of European Ancestry

Date:
2022-07-01

Citation:
Dumont, M., Weber-Lassalle, N., Joly-Beauparlant, C., Ernst, C., Droit, A., Feng, B. J., Dubois, S., Collin-Deschesnes, A. C., Soucy, P., Vallée, M., Fournier, F., Lemaçon, A., Adank, M. A., Allen, J., Altmüller, J., Arnold, N., Ausems, M. G. E. M., Berutti, R., Bolla, M. K. ,... Simard, J. (2022). Uncovering the Contribution of Moderate-Penetrance Susceptibility Genes to Breast Cancer by Whole-Exome Sequencing and Targeted Enrichment Sequencing of Candidate Genes in Women of European Ancestry. Cancers, 14 (14), pp.3363-. https://doi.org/10.3390/cancers14143363.

Persistent Link:
https://hdl.handle.net/11343/317160

License:
CC BY

cancers

MDPI

Article

Uncovering the Contribution of Moderate-Penetrance Susceptibility Genes to Breast Cancer by Whole-Exome Sequencing and Targeted Enrichment Sequencing of Candidate Genes in Women of European Ancestry

Martine Dumont [1], Nana Weber-Lassalle [2], Charles Joly-Beauparlant [1], Corinna Ernst [2], Arnaud Droit [1], Bing-Jian Feng [3,4], Stéphane Dubois [1], Annie-Claude Collin-Deschesnes [1], Penny Soucy [1], Maxime Vallée [1], Frédéric Fournier [1], Audrey Lemaçon [1], Muriel A. Adank [5], Jamie Allen [6], Janine Altmüller [7], Norbert Arnold [8], Margreet G. E. M. Ausems [9], Riccardo Berutti [10], Manjeet K. Bolla [6], Shelley Bull [11,12], Sara Carvalho [6], Sten Cornelissen [13], Michael R. Dufault [14], Alison M. Dunning [15], Christoph Engel [16], Andrea Gehrig [17], Willemina R. R. Geurts-Giele [18], Christian Gieger [19,20], Jessica Green [11,21], Karl Hackmann [22], Mohamed Helmy [23,24,25], Julia Hentschel [26], Frans B. L. Hogervorst [5], Antoinette Hollestelle [27], Maartje J. Hooning [27], Judit Horváth [28], M. Arfan Ikram [29], Silke Kaulfuß [30], Renske Keeman [13], Da Kuang [21,23], Craig Luccarini [15], Wolfgang Maier [31], John W. M. Martens [27], Dieter Niederacher [32], Peter Nürnberg [33], Claus-Eric Ott [34], Annette Peters [19,35], Paul D. P. Pharoah [6,15], Alfredo Ramirez [36], Juliane Ramser [37], Steffi Riedel-Heller [38], Gunnar Schmidt [39], Mitul Shah [15], Martin Scherer [40], Antje Stäbler [41], Tim M. Strom [10], Christian Sutter [42], Holger Thiele [7], Christi J. van Asperen [43], Lizet van der Kolk [5], Rob B. van der Luijt [43,44], Alexander E. Volk [45], Michael Wagner [46], Quinten Waisfisz [47], Qin Wang [6], Shan Wang-Gohrke [48], Bernhard H. F. Weber [49,50], Genome of the Netherlands Project †, GHS Study Group †, Peter Devilee [51], Sean Tavtigian [4,52], Gary D. Bader [11,21,23,53,54], Alfons Meindl [37], David E. Goldgar [3,4], Irene L. Andrulis [11,21], Rita K. Schmutzler [2], Douglas F. Easton [6,15], Marjanka K. Schmidt [13,55], Eric Hahnen [2] and Jacques Simard [1,56,*]

1 Genomics Center, CHU de Québec-Université Laval Research Center, 2705 Laurier Boulevard, Quebec City, QC GIV 4G2, Canada; martine.dumont@crchudequebec.ulaval.ca (M.D.); charles.joly-beauparlant@crchudequebec.ulaval.ca (C.J.-B.); arnaud.droit@crchudequebec.ulaval.ca (A.D.); stephane.dubois@crchudequebec.ulaval.ca (S.D.); annie-c.c-deschesnes@crchudequebec.ulaval.ca (A.-C.C.-D.); penny.soucy@crchudequebec.ulaval.ca (P.S.); maxime.vallee@chu-lyon.fr (M.V.); frederic.fournier.4@ulaval.ca (F.F.); audrey@mims.ai (A.L.)

2 Center for Familial Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Faculty of Medicine and University Hospital Cologne, University of Cologne, 50937 Cologne, Germany; nana.weber-lassalle@uk-koeln.de (N.W.-L.); corinna.ernst@uk-koeln.de (C.E.); rita.schmutzler@uk-koeln.de (R.K.S.); eric.hahnen@uk-koeln.de (E.H.)

3 Department of Dermatology, University of Utah, Salt Lake City, UT 84103, USA; bingjian.feng@hsc.utah.edu (B.-J.F.); david.goldgar@hsc.utah.edu (D.E.G.)

4 Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA; sean.tavtigian@hci.utah.edu

5 Family Cancer Clinic, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, 1066 Amsterdam, The Netherlands; m.adank@nki.nl (M.A.A.); f.hogervorst@nki.nl (F.B.L.H.); l.vd.kolk@nki.nl (L.v.d.K.)

6 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; jma@ebi.ac.uk (J.A.); mkh39@medschl.cam.ac.uk (M.K.B.); sc2017@medschl.cam.ac.uk (S.C.); paul.pharoah@medschl.cam.ac.uk (P.D.P.P.); qw232@medschl.cam.ac.uk (Q.W.); dfe20@medschl.cam.ac.uk (D.F.E.)

7 Cologne Center for Genomics (CCG), Faculty of Medicine, University Hospital Cologne, University of Cologne, 50931 Cologne, Germany; janine.altmueller@mdc-berlin.de (J.A.); holger.thiele@uni-koeln.de (H.T.)

check for updates

Citation: Dumont, M.; Weber-Lassalle, N.; Joly-Beauparlant, C.; Ernst, C.; Droit, A.; Feng, B.-J.; Dubois, S.; Collin-Deschesnes, A.-C.; Soucy, P.; Vallée, M.; et al. Uncovering the Contribution of Moderate-Penetrance Susceptibility Genes to Breast Cancer by Whole-Exome Sequencing and Targeted Enrichment Sequencing of Candidate Genes in Women of European Ancestry. *Cancers* **2022**, *14*, 3363. https://doi.org/10.3390/cancers14143363

Academic Editor: Evgeny Yakirevich

Received: 12 April 2022
Accepted: 1 July 2022
Published: 11 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

[8] Institute of Clinical Molecular Biology, Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, 24105 Kiel, Germany; norbert.arnold@uksh.de
[9] Division Laboratories, Pharmacy and Biomedical Genetics, Department of Genetics University Medical Center Utrecht, 3584 Utrecht, The Netherlands; m.g.e.m.ausems@umcutrecht.nl
[10] Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany; riccardo.berutti@helmholtz-muenchen.de (R.B.); tim.strom@tum.de (T.M.S.)
[11] Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON M5G 1X5, Canada; bull@lunenfeld.ca (S.B.); jessica.green@sinaihealth.ca (J.G.); gary.bader@utoronto.ca (G.D.B.); andrulis@lunenfeld.ca (I.L.A.)
[12] Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada
[13] Division of Molecular Pathology, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, 1066 Amsterdam, The Netherlands; s.cornelissen@nki.nl (S.C.); r.keeman@nki.nl (R.K.); mk.schmidt@nki.nl (M.K.S.)
[14] Precision Medicine and Computational Biology, Sanofi Genzyme, Cambridge, MA 02142, USA; michael.dufault@sanofi.com
[15] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; amd24@medschl.cam.ac.uk (A.M.D.); craig@srl.cam.ac.uk (C.L.); ms483@medschl.cam.ac.uk (M.S.)
[16] Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, 04107 Leipzig, Germany; christoph.engel@imise.uni-leipzig.de
[17] Centre of Familial Breast and Ovarian Cancer, Department of Medical Genetics, Institute of Human Genetics, University of Würzburg, 97074 Würzburg, Germany; gehrig@biozentrum.uni-wuerzburg.de
[18] Department of Clinical Genetics, Erasmus University Medical Center, 3015 Rotterdam, The Netherlands; w.geurts-giele@erasmusmc.nl
[19] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; christian.gieger@helmholtz-muenchen.de (C.G.); peters@helmholtz-muenchen.de (A.P.)
[20] Research Unit Molecular Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, 85764 Neuherberg, Germany
[21] Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; kvn.kuang@mail.utoronto.ca
[22] Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, 01307 Dresden, Germany; karl.hackmann@uniklinikum-dresden.de
[23] The Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada; mohamed_helmy@bii.a-star.edu.sg
[24] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore
[25] Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada
[26] Institute of Human Genetics, University Leipzig, 04103 Leipzig, Germany; julia.hentschel@medizin.uni-leipzig.de
[27] Department of Medical Oncology, Erasmus MC Cancer Institute, 3015 Rotterdam, The Netherlands; a.hollestelle@erasmusmc.nl (A.H.); m.hooning@erasmusmc.nl (M.J.H.); j.martens@erasmusmc.nl (J.W.M.M.)
[28] Institute of Human Genetics, University of Münster, 48149 Münster, Germany; judit.horvath@ukmuenster.de
[29] Department of Epidemiology, Erasmus MC University Medical Center, 3015 Rotterdam, The Netherlands; m.a.ikram@erasmusmc.nl
[30] Institute of Human Genetics, University Medical Center Göttingen, 37075 Göttingen, Germany; silke.kaulfuss@med.uni-goettingen.de
[31] German Center for Neurodegenerative Diseases (DZNE), Department of Neurodegenerative Diseases and Geriatric Psychiatry, Medical Faculty, University Hospital Bonn, 53127 Bonn, Germany; wolfgang.maier@ukb.uni-bonn.de
[32] Department of Gynecology and Obstetrics, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany; niederac@med.uni-duesseldorf.de
[33] Center for Molecular Medicine Cologne (CMMC), Cologne Center for Genomics (CCG), Faculty of Medicine, University Hospital Cologne, University of Cologne, 50931 Cologne, Germany; nuernberg@uni-koeln.de
[34] Institute of Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, 13353 Berlin, Germany; claus-eric.ott@charite.de
[35] Department of Epidemiology, Institute for Medical Information Processing, Biometry and Epidemiology, Medical Faculty, Ludwig-Maximilians-Universität München, 80539 Munich, Germany
[36] Division for Neurogenetics and Molecular Psychiatry, Medical Faculty, University of Cologne, 50937 Cologne, Germany; alfredo.ramirez@uk-koeln.de

[37] Division of Gynaecology and Obstetrics, Klinikum Rechts der Isar der Technischen Universität München, 81675 Munich, Germany; juliane.ramser@mri.tum.de (J.R.); alfons.meindl@gmx.de (A.M.)
[38] Institute of Social Medicine, Occupational Health and Public Health, Faculty of Medicine, University of Leipzig, 04103 Leipzig, Germany; steffi.riedel-heller@medizin.uni-leipzig.de
[39] Institute of Human Genetics, Hannover Medical School, 30625 Hannover, Germany; schmidt.gunnar@mh-hannover.de
[40] Department of Primary Medical Care, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany; m.scherer@uke.de
[41] Institute of Medical Genetics and Applied Genomics, University of Tübingen, 72076 Tübingen, Germany; antje.staebler@med.uni-tuebingen.de
[42] Institute of Human Genetics, University Hospital Heidelberg, 69120 Heidelberg, Germany; christian.sutter@med.uni-heidelberg.de
[43] Department of Clinical Genetics, Leiden University Medical Center, 2333 Leiden, The Netherlands; c.j.van_asperen@lumc.nl (C.J.v.A.); r.b.van_der_luijt@lumc.nl (R.B.v.d.L.)
[44] Department of Medical Genetics, University Medical Center, 3584 Utrecht, The Netherlands
[45] Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany; a.volk@uke.de
[46] Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, 53127 Bonn, Germany; michael.wagner@ukbonn.de
[47] Department of Human Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, 1081 Amsterdam, The Netherlands; q.waisfisz@vumc.nl
[48] Department of Gynaecology and Obstetrics, University of Ulm, 89081 Ulm, Germany; shan.wang-gohrke@uniklinik-ulm.de
[49] Institute of Human Genetics, Regensburg University, 93053 Regensburg, Germany; bweb@klinik.uni-regensburg.de
[50] Institute of Clinical Human Genetics, University Hospital Regensburg, 93053 Regensburg, Germany
[51] Department of Pathology, Department of Human Genetics, Leiden University Medical Center, 2333 Leiden, The Netherlands; p.devilee@lumc.nl
[52] Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT 84132, USA
[53] Department of Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada
[54] Princess Margaret Research Institute, University Health Network, Toronto, ON M5G 0A3, Canada
[55] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute—Antoni van Leeuwenhoek Hospital, 1066 Amsterdam, The Netherlands
[56] Department of Molecular Medicine, Faculty of Medicine, Université Laval, Quebec, QC G1V 0A6, Canada
* Correspondence: jacques.simard@crchudequebec.ulaval.ca; Tel.: +418-654-2264
† Genome of the Netherlands Project and GHS Study Group are listed in Acknowledgments.

Simple Summary: Genetic variants explaining approximately 40% of familial breast cancer risk have been identified, thus leaving a significant fraction of the heritability of this disease still unexplained. The exact nature of this missing fraction is unknown; more extensive sequencing efforts could potentially identify new moderate-penetrance breast cancer risk alleles. The aim of this study was to perform a large-scale whole-exome sequencing study, followed by a targeted validation, in breast cancer patients and healthy women of European descent. We identified 20 novel genes with modest evidence of association (p-value < 0.05) for either overall or subtype-specific breast cancer; however, much larger studies are needed to confirm the exact role of these genes in susceptibility to breast cancer.

Abstract: Rare variants in at least 10 genes, including *BRCA1*, *BRCA2, PALB2, ATM,* and *CHEK2*, are associated with increased risk of breast cancer; however, these variants, in combination with common variants identified through genome-wide association studies, explain only a fraction of the familial aggregation of the disease. To identify further susceptibility genes, we performed a two-stage whole-exome sequencing study. In the discovery stage, samples from 1528 breast cancer cases enriched for breast cancer susceptibility and 3733 geographically matched unaffected controls were sequenced. Using five different filtering and gene prioritization strategies, 198 genes were selected for further validation. These genes, and a panel of 32 known or suspected breast cancer susceptibility genes, were assessed in a validation set of 6211 cases and 6019 controls for their association with risk of breast cancer overall, and by estrogen receptor (ER) disease subtypes, using gene burden tests applied to loss-of-function and rare missense variants. Twenty genes showed nominal evidence of association (p-value < 0.05) with either overall or subtype-specific breast cancer. Our study had the

statistical power to detect susceptibility genes with effect sizes similar to *ATM*, *CHEK2*, and *PALB2*, however, it was underpowered to identify genes in which susceptibility variants are rarer or confer smaller effect sizes. Larger sample sizes would be required in order to identify such genes.

Keywords: breast cancer; genetic susceptibility; whole-exome sequencing; moderate-penetrance genes

1. Introduction

Breast cancer is the most common cause of cancer-related death and the most frequently diagnosed cancer among women worldwide [1]. Currently, approximately 40% of familial breast cancer risk is explained by a combination of common low-penetrance variants [2,3], together with coding rarer variants in predisposition genes, such as *BRCA1*, *BRCA2, PALB2, ATM*, and *CHEK2* that confer higher risks [4]. Thus, a significant fraction of the heritability of this disease is still unexplained. A better understanding of the genetic risk factors contributing to breast cancer can improve risk prediction, and hence better inform screening and prevention strategies, and may also inform understanding of the biology underlying breast cancer predisposition.

While linkage studies have identified high-risk breast cancer susceptibility genes, and genome-wide association studies (GWAS) using single-nucleotide polymorphism (SNP) arrays have identified many common variants conferring modest risks, neither of these approaches are powerful enough to identify rarer variants conferring moderate risk. Several genes associated with moderate risk (two- to four fold) of breast cancer have been identified by candidate gene sequencing approaches. These targeted approaches are limited to our understanding of the pathways involved in breast cancer etiology, and have thus mainly identified new breast cancer susceptibility alleles within genes either coding for proteins that interact with BRCA1 or BRCA2 (e.g., *PALB2*), or other genes involved in DNA repair processes, such as *ATM* and *CHEK2*. Since most genes have not yet been subjected to large-scale sequencing studies, it is possible that other susceptibility genes exist, which would explain some of the residual heritability of breast cancer.

Recently, whole-exome sequencing (WES) has provided a comprehensive, agnostic approach to exploring associations between rarer coding variants and disease. Compared to whole-genome sequencing (WGS), WES has a substantially lower cost and generates results that are more easily interpretable [5]. In the last few years, a number of WES studies have been carried out for breast cancer [6,7]. These studies rely on the sequencing of cases enriched for family history of disease, and matched controls, sometimes in combination with analysis of the segregation of the variants associated with disease in families. However, very few novel susceptibility genes have been robustly identified by these studies. This lack of success may, in part, be due to the lack of statistical power: since deleterious variants in most genes are likely to be very rare, studies need to be very large to detect them. More recently, the largest WES study for overall breast cancer reported increased breast cancer risk associated with *ATM*, *CHEK2*, *PALB2*, and *MSH6* [8]. However, while the first three genes are already well established, the association with *MSH6* has not been definitively replicated in large, targeted sequencing studies.

In an attempt to overcome these shortcomings, we designed a large-scale two-stage WES study. The discovery step was carried out using WES data from 1528 breast cancer cases enriched for genetic susceptibility to the disease, based on early onset breast cancer, a family history of disease, or bilateral breast cancer, and 3733 geographically matched, unaffected controls. The validation step assessed the associations between the presence of loss-of-function (LoF) and rare missense variants in 198 selected candidate genes and the risk of breast cancer overall, as well as estrogen receptor (ER) disease subtypes, in a replication set of 6211 breast cancer cases and 6019 controls. In addition to these genes, a panel of 32 known or suspected breast cancer susceptibility genes was also included in this targeted enrichment sequencing study.

2. Materials and Methods

2.1. Discovery Stage—Whole-Exome Sequencing (WES) Analysis

2.1.1. Studies and Datasets

For the discovery stage, WES was performed on 1528 breast cancer cases, and the resulting data were compared with whole-exome data obtained from 3733 unaffected controls (3483 unaffected women from four studies, as described below, and 250 unaffected men from GoNL trios) (Table S1). Briefly, breast cancer cases were selected from two clinic-based studies: The German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC) (1021 cases) [9], with half of the samples originating from the Cologne region (490 samples) and the other half from the Munich area (531 samples), and The Dutch Familial Bilateral Breast Cancer Study (DFBBCS) (511 cases) [10–12]. All subjects had previously tested negative for the presence of *BRCA1* and *BRCA2* germline mutations in a clinical genetic setting. All subjects from the GC-HBOC had a family history of breast cancer, while all subjects from the DFBBCS had bilateral breast cancer. Control whole-exome data were obtained from four studies: The Rotterdam Study [13], The Genome of the Netherlands Project (GoNL) [14,15], The German Study on Ageing, Cognition, and Dementia (AgeCoDe) [16], and The KORA Study [17]. A more detailed description of these studies and datasets is included in Table S1 and in the Supplementary Methods. All study subjects were recruited based on protocols approved by the Institutional Review Boards at each participating institution, and all subjects provided written informed consent.

2.1.2. Selection of Breast Cancer Cases

Breast cancer cases were selected according to the following criteria: early onset (<50 years of age) and a family history of one or more first-degree relatives with breast cancer diagnosed before 50 years of age, or two relatives diagnosed under the age of 60. A scoring system was established for the selection of families based on the assumption that rare, moderate-risk variants might be expected to confer a higher relative risk at young ages, as is the case for most known susceptibility genes [10]. In addition, the number of affected relatives weighted by their degree of relationship to the case was taken into account. Briefly, families were scored according to the following: (a) For each case in the family, a score of 1 was given if the case was diagnosed at age 50 or older, a score of 1.5 was given if diagnosed between 40–49 years, and a score of 2 if diagnosed at <40 years of age. (b) Cases were then weighted by their degree of relationship to the index case, where a weight of 1 was attributed for the index case, 0.5 for first-degree relatives, and 0.25 for second-degree relatives. (c) For bilateral cases, both cancers were included (thus giving a score of 2). (d) The total score for the family was obtained by the sum of these scores. Carriers of deleterious germline mutations in *BRCA1* and *BRCA2*, and *CHEK2* c.1100delC carriers were excluded. Related samples were identified on the basis of genotype data, and the individual with the youngest age of onset was chosen. The score distribution among breast cancer cases was as follows: 0.6% with a score below 2; 16.7% with a score between 2 and <2.5; 33.6% with a score between 2.5 and <3; 25.5% with a score between 3 and <3.5; 14.1% with a score between 3.5 and <4; 9.5% with a score $\geq$4.

2.1.3. Library Preparation, High-Throughput Sequencing, and Bioinformatics Analysis

Libraries for samples from GC-HBOC cases were prepared using the Agilent SureSelectXT Human All Exon V5 capture kit, while libraries for samples from DFBBCS cases were prepared using NimbleGen SeqCap EZ Exome Library v3.0. (Table S1). Barcoded libraries were sequenced on an Illumina HiSeq2500 for paired-end 100 bp sequencing. Available sequencing data from controls was obtained from other studies. These data were generated using the following capture technologies: NimbleGen SeqCap EZ Exome V2, NimbleGen SeqCap EZ Exome V3, Agilent SureSelect Human All Exon V3 and V5 (Table S1). Whole-exome data from cases and controls were demultiplexed and aligned to the reference genome using BWA-MEM to create aligned BAM files, followed by base quality score recalibration. The analysis of sequencing data from cases and controls was

restricted to the region of 34.5 Mb common to the different capture technologies used. For breast cancer cases, a mean coverage of 100× was obtained with >97.5% of bases above 10×. Variant calling was performed using a pipeline based on the GATK HaplotypeCaller. Details of library preparation and bioinformatics analysis including variant calling and quality control procedures are given in Supplementary Methods.

2.1.4. Variant Filtering and Gene Prioritization

Five research groups involved in this project (CHU de Québec-Université Laval Research Centre, Huntsman Cancer Institute, University of Utah, Munich Technical University, University of Toronto, University of Cambridge) performed variant filtering and gene prioritization using different approaches according to their expertise with regard to such analyses (Supplementary Methods). Variant filtering and gene prioritization strategies are summarized in Figure 1 and in Supplementary Methods. Most variant filtering strategies are similar for all groups, reflecting current best practices.

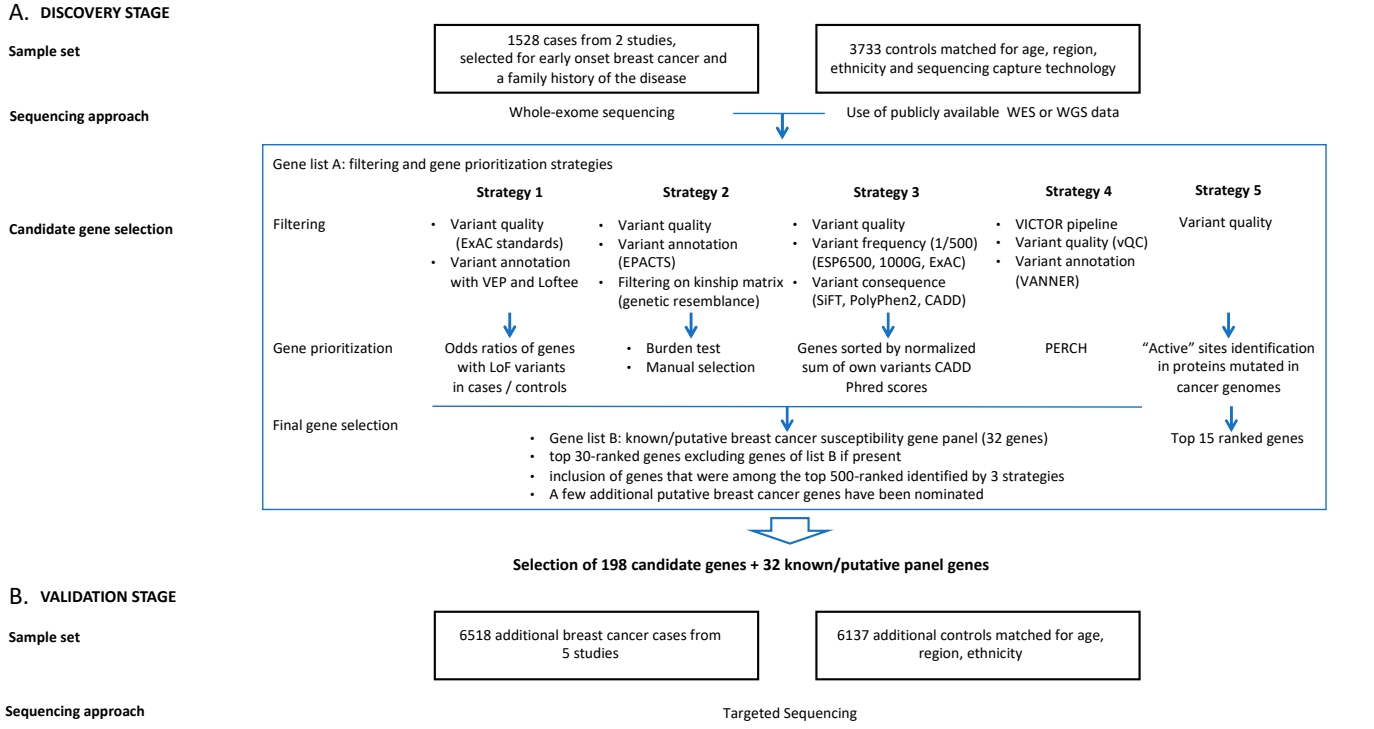


Figure 1. Study design for breast cancer susceptibility gene discovery by whole-exome sequencing (**A**) and validation by targeted enrichment sequencing (**B**).

2.1.5. Aggregation of Gene Lists

Following these gene prioritization analyses, the results from each group underwent manual scrutiny and were then merged into a final list of candidate genes for validation. The selection process first included the top 30-ranked genes from each group. Each group then nominated additional putative breast cancer susceptibility genes according to their analysis and expertise. Finally, the list was augmented by the inclusion of genes that were among the top 500-ranked by at least four groups. Through this selection process, 198 candidate genes identified in the WES were selected for targeted resequencing (Table S2). In addition to these candidate genes, this targeted sequencing enrichment stage also included 32 genes for which there was prior evidence of association with breast cancer risk [4,18]. These 32 genes were selected based on their *a priori* associations with the disease, and were not considered in the individual gene ranking process.

2.2. Validation Stage—Targeted Enrichment Sequencing

2.2.1. Sample Sets

A total of 12,655 samples, including 6518 women affected with breast cancer, and 6137 controls, were selected for the validation stage. These were drawn from five sources, namely: GC-HBOC [9] (n = 5966, 3199 cases and 2767 controls); a Dutch validation set comprising samples from two studies, the Amsterdam Breast Cancer Study-Familial (ABCS-F) [10] and the Rotterdam Breast Cancer Study (RBCS) [19] (n = 1941, 979 cases and 962 controls); Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) [20–22] (n = 2637, 1289 cases and 1348 controls); the Ontario Familial Breast Cancer Registry (OFBCR) [23] (n = 1191, 600 cases and 591 controls) and CARTaGENE (CaG) [24] (n = 920, 451 cases and 469 controls). The GC-HBOC samples did not overlap with those used for the discovery step. A more detailed description of these studies is included in Table S1 and in Supplementary Methods. All study subjects were recruited based on protocols approved by the Institutional Review Boards at each participating institution, and all subjects provided written informed consent.

2.2.2. Power Calculations

To perform power calculations for the replication stage, we assumed a significance level of p-value $< 10^{-4}$. The standard "exome-wide" significance level is $p < 2.5 \times 10^{-6}$; this is equivalent to assuming a prior probability 40-fold higher than an average gene. This is also the level proposed by Easton et al. [4] for assessing the evidence for candidate DNA repair genes included in gene panels. To allow for the oversampling of cases with a family history, bilaterality, or early age at onset, we assumed that the effect size (log odds ratio) was increased by a factor of 1.5. Based on these assumptions, the power exceeded 50% for genes with a carrier frequency >0.005 and an odds ratio (OR) in the population >2, or a carrier frequency >0.001 and an OR > 4. Based on the effect sizes and European allele frequencies [18], the power of our study to detect genes with the same characteristics as *ATM*, *CHEK2*, and *PALB2* was 0.44, 1.0 and 0.95, respectively. However, the power was low to detect genes with the characteristics of *RAD51C* (0.0039), *RAD51D* (0.0025), or *BARD1* (0.017).

2.2.3. Library Preparation and High-Throughput Sequencing

Library preparation was performed at the Cologne Center for Genomics, the Center for Familial Breast and Ovarian Cancer, and the Genomics Centre at CHU de Québec-Université Laval Research Center. Sequencing was performed at two centers: the Cologne Center for Genomics, Cologne, Germany, using Illumina HiSeq 4000 sequencers for paired-end 75 bp sequencing (mean coverage 140×); and the Genomics Centre at CHU de Québec-Université Laval Research Center, Quebec City, Canada, using Illumina HiSeq 2500 for paired-end 100 bp sequencing (mean coverage 185×). Additional details on library preparation and quality control procedures are given in the Supplementary Methods.

2.2.4. Bioinformatics and Statistical Analyses

Bioinformatics analyses were performed on the Compute Canada supercomputer, Graham. A pipeline was set up following GATK Best Practices [25]. GATK Unified Genotyper was used for variant calling, and variants were annotated and quality-controlled using VICTOR (variant interpretation for clinical testing or research), and both gene-level and variant-level significances were computed using PERCH (polymorphism evaluation, ranking and classification for heritable traits) [26]. PERCH quantitatively integrated deleteriousness prediction, allele frequency information, rare variant association analysis, biological relevance assessment, and the quality of variant calls. Additional details on the bioinformatics pipeline and variant calling procedure are given in the Supplementary Methods. Following these quality control and filtering steps, data from a total of 12,230 samples were further analyzed, including 6211 breast cancer cases and 6019 controls.

The principal association analysis was a gene burden analysis, in which the genotypes were collapsed into a 0/1 variable for each gene, according to whether or not a deleterious

variant was carried. For each gene, variant carriers were defined using the VICTOR software suite. LoF variants were defined as (1) StopGain or FrameShift variants, excluding those affecting only the last 5% of coding sequences (CDS); (2) SpliceSite variants that include coding and non-coding regions; and (3) variants that inhibit transcription or translation. Rare missense variants included in these analyses were those showing a deleteriousness score according to the BayesDel algorithm [26] (Tables S3–S5). A final list of 7437 variants was established, including LoF variants (n = 1696) and rare missense variants (n = 5741). Among these, 25 LoF variants had a minor allele frequency (MAF) between 0.01 and 0.001, and 98.5% of LoF variants had an MAF < 0.001; in contrast, 21 missense variants had an MAF between 0.01 and 0.001, and the remaining 99.6% of missense variants had an MAF < 0.001.

ORs and 95% confidence intervals (CI) associated with carrier status were computed using logistic regression, adjusting for study population as a covariate, and *p*-values were derived from Wald tests. For analyses performed by individual study, *p*-values were computed via Fisher's exact tests. All statistical analyses were performed in R v3.6.0.

3. Results

A final list of 7437 variants including LoF (n = 1696) and rare missense variants (n = 5741) were included in the gene burden analyses. Quantile–quantile (Q–Q) plots are shown in Figure S1A for LoF variants, and Figure S1B for missense variants. When known breast cancer susceptibility genes were excluded, there was no evidence of inflation in the test statistics.

3.1. Known or Suspected Breast Cancer Susceptibility Genes

Targeted enrichment sequencing of the panel of 32 known or suspected breast cancer susceptibility genes confirmed the associations of known breast cancer susceptibility genes *ATM, CHEK2*, and *PALB2* (Figure 2, Tables 1, S6 and S7). It should be noted that these analyses did not include *BRCA1* or *BRCA2*, as our original study design excluded samples from families that carried mutations in either of these genes.

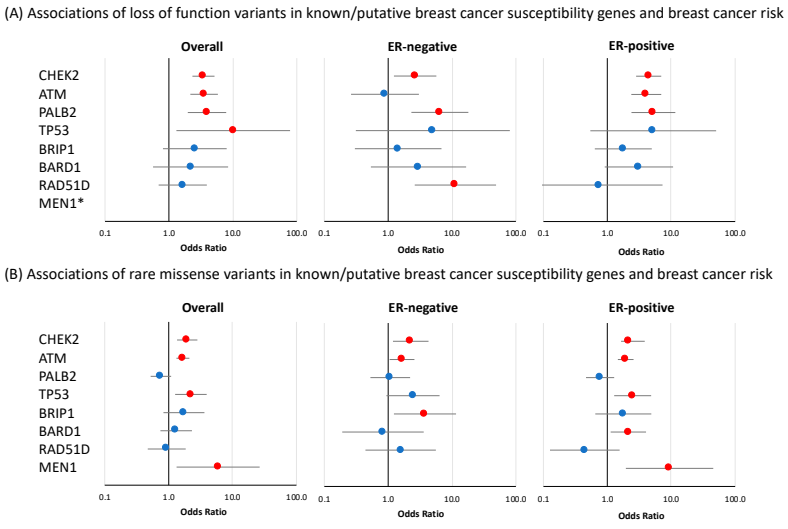


Figure 2. Risk of breast cancer overall and tumor subtypes associated with loss-of-function and missense variants in known or suspected breast cancer susceptibility genes included in breast cancer gene panels. Shown are odds ratios (ORs) and 95% confidence intervals (CIs) for breast cancer overall, estrogen receptor (ER)-negative breast cancer, and ER-positive breast cancer associated with loss-of-function variants (**A**) and missense variants (**B**) in genes for which evidence of association was observed in at least one of the analyses. Red circles indicate *p*-value < 0.05. * no loss-of-function variants were observed in *MEN1*.

Table 1. Associations of loss-of-function variants in 32 genes included in breast cancer gene panels with overall breast cancer risk and ER-negative and ER-positive breast cancer risk. Shown are odds ratios (ORs), 95% confidence intervals (CIs), and *p*-values. Associations with $p < 0.05$ are denoted in bold.

Gene	Overall						ER-Negative					ER-Positive				
	Number of Carriers	Number of Carriers	OR	95% CI		*p*-Value	Number of Carriers	OR	95% CI		*p*-Value	Number of Carriers	OR	95% CI		*p*-Value
	Controls (n = 6019)	Cases (n = 6211)					Cases (n = 808)					Cases (n = 2764)				
AKT1	1	1	0.89	0.06	14.25	0.935	0	0.00	0.00	Inf	0.997	1	1.74	0.11	27.90	0.694
ATM	**20**	**74**	**3.55**	**2.16**	**5.82**	**5.44×10^{-7}**	3	0.89	0.26	3.02	0.851	**42**	**4.06**	**2.37**	**6.97**	**3.39×10^{-7}**
BARD1	4	11	2.54	0.81	7.98	0.110	2	2.98	0.54	16.57	0.212	7	3.13	0.91	10.71	0.069
BRE/BABAM2	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
BRIP1	8	14	1.64	0.69	3.93	0.262	2	1.44	0.30	6.84	0.649	7	1.78	0.64	4.97	0.274
CDH1	0	4	0.00	0.00	Inf	0.906	2	0.00	0.00	Inf	0.994	1	0.00	0.00	Inf	0.994
CHEK2	**32**	**111**	**3.47**	**2.33**	**5.15**	**7.62×10^{-10}**	**9**	**2.65**	**1.24**	**5.70**	**0.012**	**56**	**4.44**	**2.82**	**6.99**	**1.23×10^{-10}**
EPCAM	3	1	0.33	0.04	3.22	0.343	0		0.00	Inf	0.995	1	0.60	0.06	5.77	0.658
FAM175A/ABRAXAS1	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
FANCC	15	13	0.86	0.41	1.81	0.692	1	0.63	0.08	4.91	0.662	6	1.00	0.37	2.66	0.995
FANCM	39	42	1.03	0.67	1.60	0.894	8	1.52	0.70	3.30	0.295	20	1.09	0.63	1.90	0.749
GEN1	2	3	1.54	0.26	9.26	0.636	1	7.68	0.45	131.19	0.159	2	3.56	0.32	39.34	0.300
MEN1	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
MLH1	1	1	0.89	0.06	14.25	0.935	0		0.00	Inf	0.997	1	1.74	0.11	27.90	0.694
MRE11A/MRE11	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
MSH2	0	2	0.00	0.00	Inf	0.934	0	-	-	-	-	1	0.00	0.00	Inf	0.994
MSH6	**47**	**19**	**0.36**	**0.21**	**0.62**	**2.07×10^{-4}**	5	0.55	0.22	1.39	0.204	**11**	**0.42**	**0.22**	**0.81**	**9.00×10^{-3}**
MUTYH	5	4	0.78	0.21	2.93	0.718	1	2.06	0.21	20.54	0.537	1	0.73	0.07	7.19	0.784
NBN	27	19	0.65	0.36	1.17	0.150	3	0.61	0.18	2.02	0.416	13	0.90	0.46	1.76	0.766
NF1	18	19	0.95	0.50	1.81	0.867	3	0.86	0.25	2.92	0.805	13	1.26	0.62	2.59	0.522
PALB2	**10**	**42**	**3.95**	**1.98**	**7.88**	**9.68×10^{-5}**	**7**	**6.45**	**2.31**	**17.98**	**3.72×10^{-4}**	**22**	**5.27**	**2.39**	**11.61**	**3.83×10^{-5}**
PIK3CA	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
PMS2	27	16	0.53	0.29	0.98	0.044	3	0.57	0.17	1.87	0.351	8	0.53	0.24	1.16	0.111
PTEN	1	2	1.78	0.16	19.65	0.637	0	0.00	0.00	Inf	0.997	1	1.74	0.11	27.90	0.694
RAD50	9	9	0.97	0.38	2.44	0.943	1	1.05	0.13	8.59	0.967	8	2.45	0.90	6.71	0.081
RAD51C	3	6	1.89	0.47	7.55	0.370	2	3.90	0.64	23.88	0.141	2	1.19	0.20	7.14	0.847
RAD51D	3	7	2.18	0.56	8.45	0.258	**5**	**11.26**	**2.60**	**48.68**	**1.19×10^{-3}**	1	0.74	0.08	7.31	0.795
RECQL	9	12	1.28	0.54	3.05	0.576	0	0.00	0.00	Inf	0.990	6	1.64	0.55	4.92	0.379
RINT1	6	5	0.76	0.23	2.50	0.655	0	0.00	0.00	Inf	0.992	2	0.59	0.12	2.90	0.512
STK11	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
TP53	**1**	**11**	**10.16**	**1.31**	**78.67**	**0.026**	1	5.00	0.31	80.03	0.255	3	5.24	0.54	50.39	0.152
XRCC2	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-

3.1.1. Loss-of-Function Variants

Significant associations for LoF variants with overall breast cancer risk were observed for *CHEK2* (OR = 3.47 (CI 95% 2.33–5.15), p-value = 7.62×10^{-10}); *ATM* (OR = 3.55 (CI 95% 2.16–5.82), p-value = 5.44×10^{-7}); and *PALB2* (OR = 3.95 (CI 95% 1.98–7.88), p-value = 9.68×10^{-5}), while more modest evidence of overall risk association was observed for *TP53* (OR = 10.16 (CI 95% 1.31–78.67, p-value = 0.026) (Figure 2 and Table 1). No nominally significant associations were observed for the remaining 28 genes in this panel. Of the four genes that were associated with breast cancer overall, *CHEK2* and *PALB2* also showed evidence of association with both ER-positive and ER-negative disease (*CHEK2* ER-positive: OR = 4.44 (CI 95% 2.82–6.99), p-value = 1.23×10^{-10} and ER-negative: OR = 2.65 (CI 95% 1.24–5.70), p-value = 0.012; *PALB2* ER-positive: OR = 5.27 (CI 95% 2.39–11.61), p-value = 3.83×10^{-5} and ER-negative: OR = 6.45 (CI 95% 2.31–17.98), p-value = 3.72×10^{-4}) while *ATM* showed evidence of an association with ER-positive disease (OR = 4.06 (CI 95% 2.37–6.97), p-value = 3.39×10^{-7}, p-diff = 0.026 for difference with ER-negative disease) (Figure 2, Table 1). Among the genes that had no evidence of association with overall breast cancer risk, *RAD51D* showed some evidence of association with ER-negative disease (OR = 11.26 (CI 95% 2.60–48.68), p-value = 1.19×10^{-3}, p-diff = 0.049 for difference with ER-positive disease) (Figure 2, Table 1). LoF variants in *MSH6* showed evidence of a negative association with breast cancer risk. This effect was observed for overall breast cancer risk (OR = 0.36 (CI 95% 0.21–0.62), p-value = 2.07×10^{-4}) as well as for ER-positive specific disease (OR = 0.42 (CI 95% 0.22–0.81), p-value = 9×10^{-3}) (Table 1).

3.1.2. Rare Missense Variants

As described in the Materials and Methods section, rare missense variants included in these analyses were those showing a deleteriousness score according to the BayesDel algorithm [26] (Tables S3–S5). In the panel of 32 known or suspected breast cancer susceptibility genes, there was evidence of an association with overall breast cancer risk for rare missense variants in four genes: *ATM* (OR = 1.67 (CI 95% 1.32–2.10), p-value = 1.56×10^{-5}); *CHEK2* (OR = 1.95 (CI 95% 1.35–2.82), p-value = 3.43×10^{-4}); *TP53* (OR = 2.22 (CI 95% 1.26–3.93), p-value = 5.86×10^{-3}) and *MEN1* (OR = 5.93 (CI 95% 1.33–26.50), p-value = 0.020) (Figure 2 and Table S7). The results of the tumor ER-subtype analyses are shown in Table S7. Among genes that were not associated with overall breast cancer risk, missense variants in *BARD1* showed some evidence of association with ER-positive disease (OR = 2.14 (CI 95% 1.13–4.04), p-value = 0.020), while an association with ER-negative breast cancer was observed for missense variants in *BRIP1* (OR = 3.78 (CI 95% 1.23–11.55), p-value = 0.020) (Figure 2, Table S7).

3.2. Validation of Candidate Genes Identified at the Discovery Stage

Analysis of the 198 genes selected from the WES analysis identified 20 potential candidate susceptibility genes with evidence of association with breast cancer (p-value < 0.05) (Figure 3, Tables 2 and S8–S10).

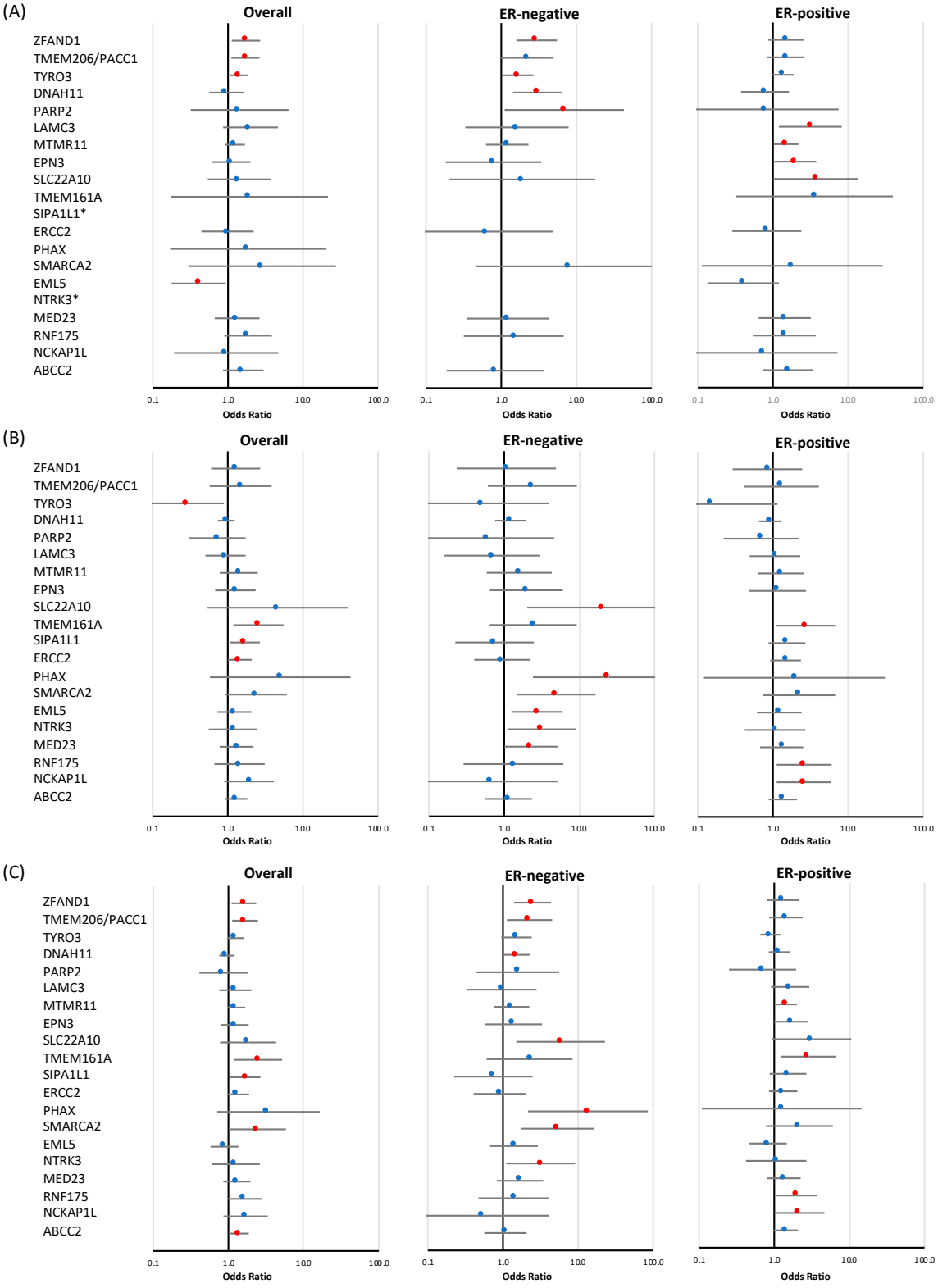


Figure 3. Risk of breast cancer overall and tumor subtypes associated with loss-of-function and missense variants in a subset of targeted genes identified at the discovery stage. Shown are odds ratios (ORs) and 95% confidence intervals (CIs) for breast cancer overall, estrogen receptor (ER)-negative breast cancer, and ER-positive breast cancer associated with loss-of-function variants (**A**), missense variants (**B**), and combined loss-of-function and missense variants (**C**) in genes for which evidence of association was observed in at least one of the analyses. Red circles indicate p-value < 0.05. * no loss-of-function variants were observed in *SIPA1L1* and *NTRK3*.

Table 2. Associations with breast cancer risk of loss-of-function and rare missense variants in top candidate genes identified at the discovery step. Shown are odds ratios (ORs), 95% confidence intervals (CIs), and *p*-values for breast cancer overall, estrogen receptor (ER)-negative breast cancer, and ER-positive breast cancer. Associations with $p < 0.05$ are denoted in bold.

	Overall						**ER-Negative**					**ER-Positive**				
	Number of Carriers		**OR**	**95% CI**		***p*-Value**	**Number of Carriers**	**OR**	**95% CI**		***p*-Value**	**Number of Carriers**	**OR**	**95% CI**		***p*-Value**
	Controls n = 6019	**Cases n = 6211**					**Cases n = 808**					**Cases n = 2764**				
(A) Loss-of-function variants																
ZFAND1	**31**	**59**	**1.73**	**1.12**	**2.68**	**0.014**	**16**	**2.96**	**1.59**	**5.50**	**6.37×10^{-4}**	24	1.50	0.87	2.60	0.146
TMEM206/PACC1	**33**	**56**	**1.70**	**1.11**	**2.62**	**0.016**	8	2.21	0.99	4.93	0.052	21	1.47	0.83	2.61	0.186
TYRO3	**89**	**131**	**1.40**	**1.06**	**1.83**	**0.016**	**22**	**1.66**	**1.03**	**2.68**	**0.038**	58	1.34	0.95	1.88	0.092
DNAH11	28	27	0.96	0.56	1.62	0.865	**10**	**3.00**	**1.42**	**6.32**	**3.88×10^{-3}**	10	0.78	0.37	1.63	0.507
PARP2	3	4	1.43	0.32	6.40	0.641	**2**	**6.89**	**1.11**	**42.78**	**0.038**	1	0.75	0.08	7.45	0.808
LAMC3	8	17	2.00	0.86	4.64	0.107	2	1.61	0.33	7.87	0.553	**11**	**3.14**	**1.20**	**8.20**	**0.020**
MTMR11	75	95	1.24	0.91	1.68	0.174	11	1.20	0.63	2.29	0.583	**49**	**1.49**	**1.03**	**2.17**	**0.037**
EPN3	21	24	1.10	0.61	1.99	0.742	2	0.79	0.18	3.40	0.747	**18**	**1.97**	**1.03**	**3.76**	**0.039**
SLC22A10	7	10	1.41	0.54	3.72	0.484	1	1.91	0.21	17.73	0.571	**6**	**3.70**	**1.01**	**13.57**	**0.048**
TMEM161A	1	2	1.94	0.18	21.43	0.590	0	0.00	0.00	Inf	0.997	2	3.56	0.32	39.34	0.300
SIPA1L1	0	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
ERCC2	12	12	0.99	0.44	2.20	0.972	1	0.62	0.08	4.83	0.646	5	0.83	0.29	2.38	0.724
PHAX	1	2	1.86	0.17	20.52	0.612	0	0.00	0.00	Inf	0.997	0		0.00	Inf	0.995
SMARCA2	1	3	2.85	0.30	27.42	0.365	1	7.68	0.45	131.19	0.159	1	1.80	0.11	28.81	0.678
EML5	18	8	0.41	0.18	0.94	0.035	0	0.00	0.00	Inf	0.986	4	0.40	0.14	1.19	0.100
NTRK3	0	1	0.00	0.00	inf	0.930	0	-	-	-	-	0	-	-	-	-
MED23	14	20	1.32	0.67	2.62	0.424	3	1.21	0.35	4.27	0.763	11	1.43	0.65	3.17	0.375
RNF175	11	21	1.85	0.89	3.85	0.099	2	1.46	0.32	6.75	0.628	7	1.42	0.54	3.72	0.478
NCKAP1L	3	3	0.95	0.19	4.71	0.948	0	0.00	0.00	Inf	0.995	1	0.73	0.07	7.19	0.784
ABCC2	16	26	1.60	0.86	2.99	0.140	2	0.83	0.19	3.67	0.804	12	1.59	0.74	3.42	0.240
(B) Rare missense variants																
ZFAND1	12	16	1.27	0.60	2.70	0.527	2	1.07	0.23	4.89	0.930	5	0.85	0.29	2.45	0.759
TMEM206/PACC1	7	11	1.48	0.57	3.83	0.416	3	2.37	0.61	9.27	0.214	5	1.29	0.41	4.06	0.669
TYRO3	13	4	0.29	0.10	0.89	0.031	1	0.51	0.07	3.93	0.516	1	0.15	0.02	1.15	0.068
DNAH11	116	115	0.95	0.74	1.24	0.726	21	1.22	0.76	1.97	0.406	52	0.91	0.65	1.28	0.603
PARP2	12	9	0.73	0.31	1.73	0.470	1	0.59	0.08	4.61	0.614	4	0.69	0.22	2.17	0.524
LAMC3	21	20	0.93	0.51	1.73	0.826	2	0.69	0.16	3.00	0.620	10	1.06	0.49	2.30	0.888
MTMR11	19	28	1.40	0.78	2.52	0.256	5	1.59	0.59	4.31	0.362	13	1.26	0.62	2.56	0.524
EPN3	18	23	1.26	0.68	2.35	0.460	4	1.98	0.65	6.05	0.233	8	1.15	0.48	2.74	0.751
SLC22A10	1	5	4.61	0.54	39.46	0.163	**3**	**20.76**	**2.03**	**211.95**	**0.011**	0	0.00	0.00	Inf	0.995
TMEM161A	**9**	**24**	**2.56**	**1.19**	**5.51**	**0.016**	3	2.44	0.64	9.26	0.190	**11**	**2.73**	**1.11**	**6.71**	**0.029**

Table 2. *Cont.*

	Overall						ER-Negative					ER-Positive				
	Number of Carriers		OR	95% CI		*p*-Value	Number of Carriers	OR	95% CI		*p*-Value	Number of Carriers	OR	95% CI		*p*-Value
	Controls n = 6019	Cases n = 6211					Cases n = 808					Cases n = 2764				
SIPA1L1	**29**	**50**	**1.68**	**1.06**	**2.67**	**0.026**	3	0.74	0.22	2.48	0.631	22	1.53	0.87	2.69	0.140
ERCC2	**49**	**73**	**1.45**	**1.01**	**2.08**	**0.047**	6	0.95	0.40	2.25	0.906	31	1.47	0.92	2.35	0.108
PHAX	1	5	4.99	0.58	42.78	0.143	**3**	**24.27**	**2.42**	**243.37**	**6.69×10^{-3}**	1	1.92	0.12	30.68	0.646
SMARCA2	6	15	2.36	0.92	6.09	0.076	**5**	**4.92**	**1.47**	**16.48**	**9.69×10^{-3}**	7	2.23	0.74	6.71	0.153
EML5	26	33	1.23	0.73	2.06	0.432	**9**	**2.74**	**1.25**	**6.01**	**0.012**	13	1.21	0.61	2.41	0.587
NTRK3	13	15	1.18	0.56	2.48	0.669	**5**	**3.17**	**1.11**	**9.11**	**0.032**	7	1.06	0.42	2.69	0.897
MED23	25	34	1.31	0.78	2.20	0.305	**8**	**2.29**	**1.01**	**5.18**	**0.048**	15	1.30	0.67	2.51	0.434
RNF175	11	16	1.43	0.66	3.09	0.361	2	1.34	0.29	6.14	0.710	**12**	**2.59**	**1.12**	**6.00**	**0.026**
NCKAP1L	10	20	1.92	0.90	4.10	0.093	1	0.65	0.08	5.16	0.686	**13**	**2.57**	**1.12**	**5.91**	**0.027**
ABCC2	57	76	1.29	0.91	1.82	0.150	9	1.14	0.56	2.34	0.714	36	1.36	0.88	2.09	0.163
(C) Combined analysis: loss-of-function and rare missense variants																
ZFAND1	**43**	**75**	**1.61**	**1.1**	**2.34**	**0.014**	**18**	**2.47**	**1.40**	**4.36**	**1.78×10^{-3}**	29	1.32	0.82	2.15	0.257
TMEM206/PACC1	**40**	**67**	**1.67**	**1.12**	**2.47**	**0.011**	**11**	**2.26**	**1.13**	**4.51**	**0.021**	26	1.43	0.86	2.39	0.169
TYRO3	102	135	1.25	0.97	1.63	0.090	23	1.51	0.95	2.41	0.082	59	1.18	0.85	1.64	0.324
DNAH11	144	142	0.95	0.75	1.21	0.696	**31**	**1.53**	**1.03**	**2.29**	**0.037**	62	0.89	0.65	1.21	0.449
PARP2	15	13	0.86	0.41	1.81	0.694	3	1.57	0.44	5.58	0.484	5	0.70	0.25	1.96	0.497
LAMC3	29	37	1.23	0.76	2.01	0.402	4	0.96	0.33	2.80	0.947	21	1.63	0.91	2.93	0.104
MTMR11	94	123	1.27	0.97	1.67	0.082	16	1.30	0.75	2.24	0.345	**62**	**1.44**	**1.03**	**2.01**	**0.032**
EPN3	38	47	1.21	0.79	1.86	0.385	6	1.37	0.57	3.30	0.484	26	1.67	1.00	2.81	0.051
SLC22A10	8	15	1.82	0.77	4.31	0.171	4	5.84	1.51	22.69	0.011	6	3.12	0.92	10.61	0.068
TMEM161A	**10**	**26**	**2.50**	**1.2**	**5.19**	**0.014**	3	2.26	0.61	8.44	0.224	**13**	**2.83**	**1.22**	**6.55**	**0.015**
SIPA1L1	**29**	**50**	**1.68**	**1.06**	**2.67**	**0.026**	3	0.74	0.22	2.48	0.631	22	1.53	0.87	2.69	0.140
ERCC2	60	83	1.34	0.96	1.88	0.083	7	0.90	0.41	2.01	0.802	35	1.32	0.86	2.04	0.210
PHAX	2	7	3.41	0.71	16.4	0.126	**3**	**13.49**	**2.15**	**84.71**	**5.50×10^{-3}**	1	1.26	0.11	14.53	0.854
SMARCA2	**7**	**18**	**2.43**	**1.01**	**5.83**	**0.046**	**6**	**5.27**	**1.73**	**16.05**	**3.40×10^{-3}**	8	2.17	0.78	6.03	0.138
EML5	44	41	0.89	0.58	1.36	0.581	9	1.40	0.67	2.91	0.366	17	0.83	0.47	1.47	0.519
NTRK3	13	16	1.25	0.6	2.6	0.552	**5**	**3.17**	**1.11**	**9.11**	**0.032**	7	1.06	0.42	2.69	0.897
MED23	39	53	1.29	0.85	1.96	0.225	10	1.69	0.83	3.43	0.148	26	1.35	0.82	2.25	0.242
RNF175	22	37	1.64	0.97	2.79	0.066	4	1.40	0.47	4.12	0.545	**19**	**2.00**	**1.07**	**3.75**	**0.031**
NCKAP1L	13	23	1.69	0.86	3.35	0.130	1	0.53	0.07	4.09	0.542	**14**	**2.16**	**1.01**	**4.65**	**0.048**
ABCC2	**72**	**102**	**1.38**	**1.02**	**1.87**	**0.039**	11	1.08	0.56	2.06	0.821	48	1.42	0.98	2.07	0.067

3.2.1. Loss-of-Function Variants

LoF variants in three genes (*ZFAND1*, *TYRO3*, and *TMEM206/PACC1*) displayed modest evidence of an association with breast cancer risk overall (p-value < 0.05) (Figure 3 and Table 2), with effect sizes for *ZFAND1* of 1.73 (CI 95% 1.12–2.68, p-value = 0.014), *TYRO3* of 1.40 (CI 95% 1.06–1.83, p-value = 0.016), and *TMEM206/PACC1* of 1.70 (CI 95% 1.11–2.62, p-value = 0.016). Of these genes, two were also associated with ER-negative breast cancer risk: *ZFAND1* (OR = 2.96 (CI 95% 1.59–5.50), p-value = 6.37×10^{-4}) and *TYRO3* (OR = 1.66 (CI 95% 1.03–2.68), p-value = 0.038). A negative overall association with breast cancer risk was observed for *EML5* (OR = 0.41 (CI 95% 0.18–0.94), p-value = 0.035). Tumor subtype specific analyses also revealed that, among the genes that showed no evidence of an association with overall breast cancer risk, two genes showed some evidence of an association with ER-negative breast cancer, namely *DNAH11* (OR = 3.00 (CI 95% 1.42–6.32), p-value = 3.88×10^{-3}, p-value for difference with ER-positive disease is 0.012) and *PARP2* (OR = 6.89 (CI 95% 1.11–42.78), p-value = 0.038), while *LAMC3* (OR = 3.14 (CI 95% 1.20–8.20), p-value = 0.020), *MTMR11* (OR = 1.49 (CI 95% 1.03–2.17), p-value = 0.037), *EPN3* (OR = 1.97 (CI 95% 1.03–3.76), p-value = 0.039), and *SLC22A10* (OR = 3.70 (CI 95% 1.01–13.57), p-value = 0.048) showed evidence of an association with ER-positive disease (Figure 3, Table 2).

3.2.2. Rare Missense Variants

There was modest evidence of an association with overall breast cancer risk for rare missense variants (Table 2) in three genes: *TMEM161A* (OR = 2.56 (CI 95% 1.19–5.51), p-value = 0.016), *SIPA1L1* (OR = 1.68 (CI 95% 1.06–2.67), p-value = 0.026), and *ERCC2* (OR = 1.45 (CI 95% 1.01–2.08), p-value = 0.047). Missense variants in *TYRO3* showed evidence of a negative association with risk (OR = 0.29 (CI 95% 0.10–0.89), p-value = 0.031), in contrast with the effects observed for LoF variants. Eleven missense variants were observed for this gene, each only once, either in cases or in controls; the exception was variant p.R750C rs145529129, which was predominantly observed in unaffected controls, and thus, appears to explain the majority of the observed protective association. Of the three genes associated with overall breast cancer risk, *TMEM161A* was also associated with ER-positive breast cancer risk (OR = 2.73 (CI 95% 1.11–6.71), p-value = 0.029). ER-subtype-specific analyses further revealed associations with ER-negative disease for *SLC22A10* (OR = 20.76 (CI 95% 2.03–211.95), p-value = 0.011), *PHAX* (OR = 24.27 (CI 95% 2.42–243.37), p-value = 6.69×10^{-3}), *SMARCA2* (OR = 4.92 (CI 95% 1.47–16.48), p-value = 9.69×10^{-3}), *EML5* (OR = 2.74 (CI 95% 1.25–6.01), p-value = 0.012), *NTRK3* (OR = 3.17 (CI 95% 1.11–9.11), p-value = 0.032), and *MED23* (OR = 2.29 (CI 95% 1.01–5.18), p-value = 0.048), while *RNF175* (OR = 2.59 (CI 95% 1.12–6.00), p-value = 0.026) and *NCKAP1L* (OR = 2.57 (CI 95% 1.12–5.91), p-value = 0.027) were associated with ER-positive breast cancer. As mentioned in the previous section, we observed modest evidence of a negative association with overall breast cancer risk for LoF variants in *EML5*, which is in contrast with the observed association of missense variants in this gene with ER-negative disease.

3.2.3. Combined Analysis of LoF and Rare Missense Variants

The combined analysis of LoF and rare missense variants (Table 2) identified a new association for *ABCC2* (p-value = 0.038) with overall breast cancer, and also identified additional distinct associations for two genes for which evidence of associations had been observed in either separate LoF or missense variants analyses, namely, a modest association of *SMARCA2* with overall breast cancer (OR = 2.43 (CI 95% 1.01–5.82), p-value = 0.046), and one for *TMEM206/PACC1* with ER-negative breast cancer (OR = 2.26 (CI 95% 1.13–4.51), p-value = 0.021). Moreover, the combined analysis also identified associations ($p < 0.05$), which had also been observed in separate LoF or missense variants analyses, for the following genes: *ZFAND1* (p-value = 0.014), *TMEM206/PACC1* (p-value = 0.011), *TMEM161A* (p-value = 0.014), and *SIPA1L1* (p-value = 0.026) with overall breast cancer; *ZFAND1* (p-value = 1.78×10^{-3}), *DNAH11* (p-value = 0.037), *SLC22A10* (p-value = 0.011), *PHAX*

(p-value = 5.50 × 10^{-3}), *SMARCA2* (p-value = 3.40 × 10^{-3}), and *NTRK3* (p-value = 0.032) with ER-negative breast cancer subtype; and *MTMR11* (p-value = 0.032), *TMEM161A* (p-value = 0.015), *RNF175* (p-value = 0.031), and *NCKAP1L* (p-value = 0.048) for ER-positive breast cancer subtype.

3.2.4. Individual LoF and Missense Variant Analysis

As mentioned in the Materials and Methods section, the main association analysis performed in our study was a gene burden analysis. This approach increases the power to detect an association, and has been used by a vast majority of association studies, including those recently published by Dorling et al. [18] and Hu et al. [27]. Nevertheless, it has been shown that single variants can show evidence of association when a weaker signal is obtained at the gene level [28,29]. We therefore examined whether certain individual LoF or missense variants with higher frequencies showed evidence of association, and whether they were driving the observed associations for a given gene. In this analysis, we only examined individual variants observed in at least three carriers, at least two of which were observed in breast cancer cases. As shown in Supplementary Table S11A, a total of seven LoF variants in six different genes showed evidence of association with overall breast cancer (nominal $p < 0.05$). Three of these variants were observed in known breast cancer susceptibility genes, one in *ATM* (c.1564_1565delGA, OR = 8.77 (CI 95% 1.11–69.29), p-value = 0.040), which is reported as pathogenic in the ClinVar database [30], and two in *CHEK2* (c.1100delC, OR = 3.08 (CI 95% 2.01–4.3), p-value = 2.83 × 10^{-7}; c.444 + 1G > A, OR = 8.12 (CI 95% 1.03–64.13), p-value = 0.047). Three other LoF variants showing evidence of association in our single variant analysis were located in the top three candidate genes identified through the gene burden test: *ZFAND1* (c.139-3_139delCAGG, OR = 1.76 (CI 95% 1.11–2.81), p-value = 0.017), *TMEM206/PACC1* (c.631delC, OR = 1.81 (CI 95% 1.15–2.86), p-value = 0.011), and *TYRO3* (c.308_308 + 1insCCTGAAGTCA, OR = 1.55 (CI 95% 1.15–2.09), p-value = 3.93 × 10^{-3}). These three variants have a much higher frequency than other such variants identified in these genes. They clearly contribute to the association observed at the gene level. Lastly, one LoF variant in the *TRPM4* gene had a very high OR, while no evidence of association was observed for this gene when all LoF variants were analyzed in aggregate (*TRPM4*, c.2254C > T, OR = 10.64 (CI 95% 1.37–82.46), p-value = 0.024). A similar analysis performed on rare missense variants (Table S11B) identified only three variants with nominally significant p-values: one in *ATM* (p.L2307F, OR = 5.94 (CI 95% 2.06–17.16), p-value = 9.85 × 10^{-4}), one in *PMS2* (p.I18T, OR = 4.74 (CI 95% 1.02–22.01), p-value = 0.047), and one in *RIC1* (p.R635H, OR = 6.07 (CI 95% 1.36–27.17), p-value = 0.018). With the exception of the *ATM* gene, *PMS2* and *RIC1* showed no evidence of association in the gene burden test. Even though these variants showed a deleteriousness score according to the BayesDel algorithm, caution should be taken when interpreting missense variants because their impact may differ greatly depending on their position in the gene (e.g., conserved or functional domains).

4. Discussion

Our study identified 20 new genes showing modest evidence of an association either with overall disease and/or with ER-positive breast cancer or ER-negative breast cancer (p-value < 0.05). These results are based on the analysis of LoF variants and rare missense variants. Of these twenty, three showed evidence of association with LoF variants for overall breast cancer (*ZFAND1, TYRO3, TMEM206/PACC1*), and a further six with breast cancer subtypes (*DNAH11, PARP2, LAMC3, MTMR11, EPN3, SLC22A10*). When missense variants were assessed, three further associations were observed for overall breast cancer (*TMEM161A, SIPA1L1, ERCC2)*, and a further seven were observed for subtypes (*RNF175, NCKAP1L, PHAX, SMARCA2, EML5, NTRK3, MED23)*. One additional association was observed when missense and LoF variants were combined (*ABCC2*). Detailed information on these 20 genes, including a description of the known functions and pathways in which they are involved, as well as the number of LoF and missense variants observed in the

different analyses, are provided in Supplementary Tables S2–S5. It should be noted that none of the genes reached the $p < 10^{-4}$ threshold in the replication stage, suggested for candidate genes [4]. Moreover, the number of associations observed at the $p < 0.05$ level for each analysis is actually less than the number that would be expected by chance (10, given 198 genes). Thus, larger replication studies will be required to confirm or refute these associations.

Among these genes, *ZFAND1, TYRO3*, and *TMEM206/PACC1* showed evidence of association with LoF variants for overall breast cancer. In *ZFAND1*, one variant (c.139-3_139delCAGG, rs1260212806) explained the majority of the observed association. It is located in a splice site junction (acceptor site). This variant was observed only in the German study, so it could be more frequent in this population. The *TMEM206/PACC1* gene is involved in the progression of colorectal cancer by accelerating cell proliferation and promoting cell migration and invasion [31]. Five LoF variants were identified in this gene, one of which was observed in 93% of the carriers (c.631delC, rs532279691). *TYRO3* has been found to be upregulated in various cancers, including AML, CML, multiple myeloma, melanoma, as well as uterine endometrial cancers [32–34]. Eight LoF variants were identified in *TYRO3*, including one recurrent variant c.308_308 + 1insCCTGAAGTCA (rs773930671), accounting for 85% of carriers of LoF variants in this gene. Our analysis of missense variants in *TYRO3* revealed evidence of a protective effect, which is opposite to that observed for LoF variants. Eleven rare missense variants were identified in our study, the majority only being observed once, and seven of them only observed in controls.

Tumor subtype-specific analyses were performed and, as described below, some evidence of ER-specific associations was observed. It should be noted that ER status was only available for 3572 (57.51%) breast cancer cases (808 ER-negative cases and 2764 ER-positive cases). Tumor subtype analyses of LoF variants revealed evidence of association for *ZFAND1, TYRO3, DNAH11*, and *PARP2* with ER-negative breast cancer, with *ZFAND1* showing the strongest association (OR = 2.96 (CI 95% 1.59–5.50), p-value = 6.37×10^{-4}). The observed association for this gene is mainly due to the variant rs1260212806, previously mentioned for breast cancer overall. A few reports have linked variants (other than those observed in the current study) in *DNAH11* with breast cancer risk, but these reports were based on comparatively small sample sizes (c.2081_2082del (p.Val694Glyfs*2)) [35] and studies involving specific populations (G > A (rs2494938-India)) [36]. More recently, large-scale GWAS studies performed by our group through the Breast Cancer Association Consortium (BCAC) identified a common variant located at the 3′-UTR region of this gene (rs7971) to be statistically significantly associated with breast cancer risk (p-value = 1.9×10^{-8}) [3]. Although further validation is needed, these observations seem to indicate that this locus may comprise both common and rare variants, with some evidence of association with breast cancer risk.

On the other hand, evidence of association with ER-positive disease with regard to LoF variants was observed for *MTMR11, LAMC3*, and *EPN3*. For *MTMR11*, the observed association is mainly due to one recurrent variant, c.14_15insG (rs587606143), which accounts for 94% of carriers of a LoF variant in this gene. Interestingly, a locus at 1q12-q21, which includes *MTMR11*, was identified by GWAS as being associated with mammographic density (rs11205303, OR = 0.73 (95% CI = 0.66–0.80), $p = 2.64 \times 10^{-11}$) [37]. Large-scale GWAS and fine-mapping studies performed in the BCAC have shown that common variants at 1q21.2 are significantly associated with overall and ER-positive breast cancer risk (overall breast cancer p-value = 9×10^{-14}; ER + breast cancer p-value = 1×10^{-12}) [3,38]. One study has also reported altered expression levels of *MTMR11* in breast cancer cells [39]. These observations make this locus an interesting candidate for further validation and characterization with regard to breast cancer. Although a modest association was observed in the current study for *SLC22A10*, the frequency of rare variants in cases and controls was very low for this gene (Tables S3–S5).

In addition to clinical estrogen receptor subtyping, analyses performed on the molecular subtypes of breast cancer, namely, luminal A, luminal B, HER2-enriched, basal-like, and

normal-like, would have been relevant in the context of this analysis, given the recognized heterogeneity of ER-positive and ER-negative tumors. However, the limited number of samples for which we had data did not allow us to perform a meaningful analysis of these molecular subtypes.

A number of genes showed an excess of missense variants (MAF < 0.01) classified as deleterious using bioinformatics in silico methods, without showing a corresponding excess of LoF variants. The best putative candidates in this group were *TMEM161A*, *ERCC2*, and *SIPA1L1* for overall breast cancer, *RNF175* and *NCKAP1L* for ER-positive disease, and *PHAX, SMARCA2, NTRK3, EML5*, and *MED23* for ER-negative disease. Because the observed effects did not reach the threshold significance level following the adjustment of probability for multiple testing, we cannot exclude the possibility that the excess of missense variants, in the absence of enrichment for LoF variants, may be a false-positive result. However, these findings may also suggest that some of these genes are intolerant of LoF variants. When looking at the best candidate genes for overall breast cancer, pLI (probability of being loss-of-function intolerant) scores and observed/expected (o/e) scores available on gnomAD seem to suggest that this may be the case for *SIPAL1*, with pLI = 1, o/e = 0.09 (0.05–0.17). Interestingly, we did not observe any LoF variants in this gene in any of our analyses. On the other hand, *ERCC2* (pLI = 0, o/e = 0.82 (0.62–1.09)) and *TMEM161A* (pLI = 0; o/e = 0.53 (0.35–0.82)) scores seem to indicate that these genes are not LoF intolerant. For ER-subtype-specific associations, LoF intolerant scores were observed for *NCKAP1L* (pL1 = 0.85, o/e = 0.2 (0.13–0.33)), *SMARCA2* (pL1 = 1, o/e = 0.12 (0.08–0.2)), and *NTRK3* (pL1 = 1, o/e = 0.12 (0.06–0.26)) (https://gnomad.broadinstitute.org, accessed on 9 February 2022). Among these genes, *ERCC2* is of particular interest, as it plays an important role in the nucleotide excision repair pathway, and common polymorphisms in this gene have been associated with risk of various types of cancers [40]. The evidence supporting *ERCC2* as a more penetrant breast cancer susceptibility gene, however, is not convincing [41].

Our study also included the analysis of 32 well-established or suspected susceptibility genes included in breast cancer gene panels. For many genes in such panels, the evidence of an association with cancer is often weak, and, therefore, further validation is required to confirm their contribution to breast cancer risk [4]. The current validation study successfully detected confirmed breast cancer susceptibility genes such as *ATM*, *CHEK2*, and *PALB2*. The ORs calculated in our study of these frequently mutated genes are similar to those calculated in recent publications [8,18,42–45] according to analyses of LoF and rare missense variants. Similar to the current analysis, the study by Dorling et al. [18] did not detect a statistically significant association of missense variants in *PALB2* (OR = 0.96 (95% CI 0.87–1.06)) with breast cancer risk. It should be noted, however, that there is a significant overlap of the current study with the Dorling et al. dataset, which may contribute towards explaining these observations. For the remaining suspected susceptibility genes, including genes confirmed to be associated with breast cancer risk or predisposition to specific cancer syndromes, as studied by Lee et al. [46], Dorling et al. [18], and/or Hu et al. [27] (*BARD1*, *CDH1*, *PTEN*, *RAD51C*, *RAD51D*, *STK11*, *TP53*), the frequencies of LoF and rare missense variants were too low in our study to detect or support an association with breast cancer risk.

Limitations and Weaknesses

It should be noted that this study has some limitations. Firstly, the discovery step involved WES of samples from breast cancer cases followed by analysis of the resulting data with available WES or WGS control data. This study design was motivated by the possibility of taking advantage of several available datasets. However, many factors resulting from this choice may have affected our ability to identify likely candidate genes at the discovery step, such as: (1) The matched control data were derived from different exome enrichment methods. These methods do not have the same efficiency in terms of probe hybridization and exon capture. This may have contributed to uneven exon sequencing

depth, which in turn could have affected the identification of variants [47]; (2) For the GoNL control dataset, exome data were extracted from WGS data. Although it is recognized that WGS outperforms WES, and provides a higher proportion of covered transcripts at a lower sequencing coverage and no strand bias [48], the 13× sequencing coverage for these samples may not have been sufficient for the purpose of identifying rare variants. In an attempt to minimize these impacts, several calling scenarios for the joint genotyping of case and control samples, taking into consideration ethnicity and capture technology, were performed in order to ensure that the most comprehensive variety of calling scenarios were covered. Five different variant filtering and prioritization strategies were also performed in an attempt to build the best candidate gene list for the validation step. Despite this, it cannot be excluded that insufficient or non-uniform sequencing coverage across exons between case and controls [49], batch effects, and population or cryptic stratification [50] could have had an impact on variant calling, which may not have been resolved at the bioinformatics level [48]. We also acknowledge the limitations of the currently available in silico tools for the prediction of pathogenic missense variants, and how these limitations could have impacted our gene selection.

Secondly, while the replication stage was quite large, important susceptibility may not have been selected at the discovery stage due to the limitations mentioned above. Moreover, although the validation stage was of a large enough scale to identify genes with effect sizes similar to those of *CHEK2*, *ATM*, and *PALB2*, the power was too low to detect genes with the characteristics of *RAD51C*, *RAD51D*, or *BARD1*. Indeed, the associations with LoF variants in these genes were all non-significant, despite the fact that the estimated ORs (~2) were comparable with those previously reported. Thus, if a susceptibility gene with the characteristics of the latter was among those identified at the discovery stage, the validation step would not have had sufficient power to confirm it. We therefore conclude that it is unlikely that any genes in the replication set with combined LoF frequencies comparable with *ATM* or *CHEK2* are associated with moderate risk, but that moderate-risk genes with lower LoF frequencies could be present.

Finally, our study was based on WES, which does not consider non-coding regulatory regions or large genomic rearrangements. Regulatory variants have been shown to contribute to the heritability of several diseases. In particular, the common susceptibility variants for breast cancer identified through GWAS are mostly located in regulatory regions [51].

5. Conclusions

We found nominal evidence of association with breast cancer overall for LoF variants in three genes (*ZFAND1*, *TMEM206/PACC1*, and *TYRO3),* with ER-negative breast cancer in two genes, *DNAH11* and *PARP2*, and with ER-positive disease in four genes *LAMC3*, *MTMR11, EPN3*, and *SLC22A10*. Analysis of rare missense variants identified evidence of associations of *TMEM161A*, *SIPA1L1*, and *ERCC2* with overall breast cancer, *RNF175* and *NCKAP1L* with ER-positive disease, and *SLC22A10*, *PHAX*, *SMARCA2*, *EML5*, *NTRK3*, and *MED23* with ER-negative disease. Our study design shows that we had the power to detect genes with effect sizes similar to some confirmed breast cancer susceptibility genes, such as *ATM* and *CHEK2*. The fact that we did not identify similar novel genes suggests that the remaining breast cancer susceptibility could be explained by lower frequency variants. The outcome of our study thus provides crucial information with regards to the planning of future sequencing efforts.

In order to efficiently identify novel lower frequency variants, larger collaborative sequencing efforts will be needed. The data generated in the current project will most certainly be useful for combining with other similar large-scale breast cancer studies in order to obtain better power to detect such variants in moderate-risk genes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14143363/s1. Supplementary Methods, Figure S1: Quantile–

quantile plot of the observed vs. expected chi-square statistics for (A) loss-of-function and (B) missense variants in the analysis of 230 target genes using logistic regression adjusted for the origin cohort. Each circle represents the chi-square statistic for a gene. The red diagonal line represents the predicted association statistics under the global null hypothesis of no association. Table S1: Description of studies included in the discovery and validation analyses. Table S2: Selection of genes for the validation stage according to gene prioritization strategies. Table S3: List of (A) loss-of-function variants and (B) rare missense variants identified in overall breast cancer at the validation stage. Table S4: List of (A) loss-of-function variants and (B) missense variants identified in ER-negative breast cancer cases at the validation stage. Table S5: List of (A) loss-of-function variants and (B) missense variants identified in ER-positive breast cancer cases at the validation stage. Table S6: Associations, by study, of loss-of-function and rare missense variants in 32 genes included in commercial breast cancer susceptibility gene panels and breast cancer risk. OR denotes odds ratio and CI the confidence interval. Table S7: Associations of rare missense variants in 32 genes included in commercial breast cancer gene panels and breast cancer risk overall (6211 cases, 6019 controls), ER-negative (808 cases, 6019 controls), and ER-positive breast cancer risk (2764 cases, 6019 controls). OR denotes odds ratio and CI the confidence interval. Table S8: Associations of (A) loss-of-function variants, (B) rare missense variants, and (C) combined loss-of-function and rare missense variants in 178 genes identified at the discovery stage (6211 cases, 6019 controls) with overall breast cancer risk and ER-negative and ER-positive breast cancer risk. OR denotes odds ratio and CI the confidence interval. Table S9: Associations with breast cancer risk of loss-of-function and rare missense variants in 178 genes identified at the discovery stage, by study. OR denotes odds ratio and CI the confidence interval. Table S10: Associations of loss-of-function and rare missense variants in 178 genes identified at the discovery stage with (A) ER-negative breast cancer risk (808 cases, 6019 controls) and (B) ER-positive breast cancer risk (2764 cases, 6019 controls). OR denotes odds ratio and CI the confidence interval. Table S11: Associations of individual (A) loss-of-function and (B) rare missense variants, observed in at least three carriers, among breast cancer cases and controls, that included at least two breast cancer cases, with overall breast cancer. References [52–65] are cited in the supplementary materials.

Author Contributions: Conceptualization: J.S., D.F.E., D.E.G., E.H., M.K.S., A.D., R.K.S., A.M., G.D.B., I.L.A. and S.T.; methodology: M.D., J.S., C.J.-B., M.V., B.-J.F. and H.T.; software: B.-J.F.; data curation: F.F., M.K.B., J.A. (Jamie Allen) and Q.W. (Qin Wang); investigation: N.W.-L., S.D., A.-C.C.-D., J.A. (Janine Altmüller) and P.N.; formal analysis: M.D., C.J.-B., N.W.-L., C.E. (Corinna Ernst), M.V., F.F., E.H., S.C. (Sara Carvalho), A.L., B.-J.F., M.H., D.F.E., D.E.G., R.B., A.M., G.D.B., M.R.D. and R.K.S.; resources: M.A.A., N.A., M.G.E.M.A., S.B., S.C. (Sten Cornelissen), A.M.D., C.E. (Christoph Engel), A.G., W.R.R.G.-G., C.G., J.G., K.H., F.B.L.H., A.H., M.J.H., J.H. (Judit Horváth), M.A.I., S.K., R.K., D.K., C.L., W.M., J.W.M.M., D.N., C.-E.O., A.P., P.D.P.P., A.R., J.R., S.R.-H., G.S., M.S. (Martin Scherer), M.S. (Mitul Shah), A.S., T.M.S., C.S., C.J.v.A., L.v.d.K., R.B.v.d.L., A.E.V., M.W., Q.W. (Quinten Waisfisz), S.W.-G., B.H.F.W., N.W.-L., P.D., I.L.A., M.K.S., A.M., E.H., R.K.S. and D.F.E.; writing—original draft preparation: M.D., P.S., J.S. and M.V.; writing—review and editing: all authors; funding acquisition: J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was part of: The PERSPECTIVE project, which was supported by the Government of Canada through Genome Canada (#13529) and the Canadian Institutes of Health Research (GPH-129344), the Ministère de l'Économie, de la Science et de l'Innovation du Québec through Genome Quebec, and the Quebec Breast Cancer Foundation; The PRE^{3}VENTION project, which was supported by a grant from the Ministère de l'Économie, de la Science et de l'Innovation du Québec through the PSR-SIIRI-949 program; The PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. Part of this work was also supported by NRNB (U.S. National Institutes of Health, National Center for Research Resources grant number P41 GM103504). Study specific funding information is provided in Supplementary Materials.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of CHU de Québec–Université Laval (project # 2014-1201, A13-10-1201 approved 24-10-2013; project # 2015-1866, A14-11-1866 approved 14-11-2014).

Informed Consent Statement: All study subjects were recruited on protocols approved by the Institutional Review Boards at each participating institution. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The complete dataset will not be made publicly available due to restraints imposed by the ethics committees of individual studies; requests for data can be made to the corresponding author or to the Data Access Coordination Committee (DACCs) of the Breast Cancer Association Consortium (https://bcac.ccge.medschl.cam.ac.uk/, accessed on 1 June 2022).

Acknowledgments: We thank all the individuals who took part in these studies, and all the researchers, clinicians, technicians, and administrative staff who have enabled this work to be carried out. Specific thanks: The authors would like to thank Martine Tranchant for sample management and skillful technical assistance. DFBBCS and GC-HBOC: The authors thank all study participants. DFBBCS would like to thank Hanne Meijers–Heijboer. The Rotterdam Study: The authors are grateful to the study participants, the staff from the Rotterdam Study, and the participating general practitioners and pharmacists. The authors would like to thank the initial principal investigator of this study, Albert Hofman. We thank the members of the Genomics Laboratory and the ERGO support team for their help in sampling the data and in creating the database. GoNL: This study makes use of data generated by the Genome of the Netherlands Project. A full list of the investigators is available from www.nlgenome.nl. Samples were contributed by LifeLines (http://lifelines.nl/lifelines-research/general, accessed on 1 June 2022), The Leiden Longevity Study (http://www.healthy-ageing.nl, accessed on 1 June 2022), The Netherlands Twin Registry (NTR: http://www.tweelingenregister.org, accessed on 1 June 2022), The Rotterdam study, (http://www.erasmus-epidemiology.nl/research/ergo.htm, accessed on 1 June 2022), and The Genetic Research in Isolated Populations program (http://www.epib.nl/research/geneticepi/research.html#gip, accessed on 1 June 2022). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI). AgeCoDe: Members of the AgeCoDe Study Group: Wolfgang Maier, Martin Scherer, Heinz-Harald Abholz, Cadja Bachmann, Horst Bickel, Wolfgang Blank, Hendrik van den Bussche, Sandra Eifflaender-Gorfer, Marion Eisele, Annette Ernst, Angela Fuchs, Kathrin Heser, Frank Jessen, Hanna Kaduszkiewicz, Teresa Kaufeler, Mirjam Köhler, Hans-Helmut König, Alexander Koppara, Carolin Lange, Hanna Leicht, Tobias Luck, Melanie Luppa, Manfred Mayer, Edelgard Mösch, Julia Olbrich, Michael Pentzek, Tina Posselt, Jana Prokein, Anna Schumacher, Steffi Riedel-Heller, Susanne Röhr, Janine Stein, Susanne Steinmann, Franziska Tebarth, Michael Wagner, Klaus Weckbecker, Dagmar Weeg, Jochen Werle, Siegfried Weyerer, Birgitt Wiese, Steffen Wolfsgruber, Thomas Zimmermann. Hendrik van den Bussche (2002–2011). AgeCoDe thanks all participating patients and their general practitioners for their collaboration. KORA Study: The KORA Study Group consists of A. Peters (speaker), L. Schwettmann, R. Leidl, M. Heier, B. Linkohr, H. Grallert, C. Gieger, and their co-workers, who are responsible for the design and conduct of the KORA studies. SEARCH: SEARCH thanks The SEARCH and EPIC teams. OFBCR: The Ontario Familial Breast Cancer Registry thanks Gord Glendon, Teresa Selander, and Nayana Weerasooriya. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does the mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The GHS Study Group is part of the Gutenberg Health Study. The authors would like to thank Sylvie Laboissière at McGill University and Genome Quebec Innovation Center for providing sequencing services and for helpful advice and technical support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Michailidou, K.; Hall, P.; Gonzalez-Neira, A.; Ghoussaini, M.; Dennis, J.; Milne, R.L.; Schmidt, M.K.; Chang-Claude, J.; Bojesen, S.E.; Bolla, M.K.; et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **2013**, *45*, 353–361. [CrossRef] [PubMed]
3. Michailidou, K.; Lindström, S.; Dennis, J.; Beesley, J.; Hui, S.; Kar, S.; Lemaçon, A.; Soucy, P.; Glubb, D.; Rostamianfar, A.; et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **2017**, *551*, 92–94. [CrossRef] [PubMed]

4. Easton, D.F.; Pharoah, P.D.; Antoniou, A.C.; Tischkowitz, M.; Tavtigian, S.V.; Nathanson, K.L.; Devilee, P.; Meindl, A.; Couch, F.J.; Southey, M.; et al. Gene-panel sequencing and the prediction of breast-cancer risk. *N. Engl. J. Med.* **2015**, *372*, 2243–2257. [CrossRef] [PubMed]
5. Kiezun, A.; Garimella, K.; Do, R.; Stitziel, N.O.; Neale, B.M.; McLaren, P.J.; Gupta, N.; Sklar, P.; Sullivan, P.F.; Moran, J.L.; et al. Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* **2012**, *44*, 623–630. [CrossRef]
6. Chandler, M.R.; Bilgili, E.P.; Merner, N.D. A Review of whole-exome sequencing efforts toward hereditary breast cancer susceptibility gene discovery. *Hum. Mutat.* **2016**, *37*, 835–846. [CrossRef]
7. Zelli, V.; Compagnoni, C.; Cannita, K.; Capelli, R.; Capalbo, C.; Di Vito Nolfi, M.; Alesse, E.; Zazzeroni, F.; Tessitore, A. Applications of Next Generation Sequencing to the Analysis of Familial Breast/Ovarian Cancer. *High Throughput* **2020**, *9*, 1. [CrossRef]
8. Lu, H.M.; Li, S.; Black, M.H.; Lee, S.; Hoiness, R.; Wu, S.; Mu, W.; Huether, R.; Chen, J.; Sridhar, S.; et al. Association of Breast and Ovarian Cancers with Predisposition Genes Identified by Large-Scale Sequencing. *JAMA Oncol.* **2019**, *5*, 51–57. [CrossRef]
9. Kast, K.; Rhiem, K.; Wappenschmidt, B.; Hahnen, E.; Hauke, J.; Bluemcke, B.; Zarghooni, V.; Herold, N.; Ditsch, N.; Kiechle, M.; et al. Prevalence of BRCA1/2 germline mutations in 21,401 families with breast and ovarian cancer. *J. Med. Genet.* **2016**, *53*, 465–471. [CrossRef]
10. Schmidt, M.K.; Hogervorst, F.; van Hien, R.; Cornelissen, S.; Broeks, A.; Adank, M.A.; Meijers, H.; Waisfisz, Q.; Hollestelle, A.; Schutte, M.; et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J. Clin. Oncol.* **2016**, *34*, 2750–2760. [CrossRef]
11. Kriege, M.; Hollestelle, A.; Jager, A.; Huijts, P.E.; Berns, E.M.; Sieuwerts, A.M.; Meijer-van Gelder, M.E.; Collée, J.M.; Devilee, P.; Hooning, M.J.; et al. Survival and contralateral breast cancer in CHEK2 1100delC breast cancer patients: Impact of adjuvant chemotherapy. *Br. J. Cancer* **2014**, *111*, 1004–1013. [CrossRef] [PubMed]
12. Schmidt, M.K.; Tollenaar, R.A.; de Kemp, S.R.; Broeks, A.; Cornelisse, C.J.; Smit, V.T.; Peterse, J.L.; van Leeuwen, F.E.; Van't Veer, L.J. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. *J. Clin. Oncol.* **2007**, *25*, 64–69. [CrossRef] [PubMed]
13. Ikram, M.A.; Brusselle, G.; Ghanbari, M.; Goedegebure, A.; Ikram, M.K.; Kavousi, M.; Kieboom, B.C.T.; Klaver, C.C.W.; de Knegt, R.J.; Luik, A.I.; et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur. J. Epidemiol.* **2020**, *35*, 483–517. [CrossRef] [PubMed]
14. Boomsma, D.I.; Wijmenga, C.; Slagboom, E.P.; Swertz, M.A.; Karssen, L.C.; Abdellaoui, A.; Ye, K.; Gurvey, V.; Vermaat, M.; Van Dijrk, F.; et al. The Genome of the Netherlands: Design, and project goals. *Eur. J. Hum. Genet.* **2013**, *22*, 221–227. [CrossRef]
15. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **2014**, *46*, 818–825. [CrossRef]
16. Jessen, F.; Wiese, B.; Bickel, H.; Eiffländer-Gorfer, S.; Fuchs, A.; Kaduszkiewicz, H.; Köhler, M.; Luck, T.; Mösch, E.; Pentzek, M.; et al. Prediction of dementia in primary care patients. *PLoS ONE* **2011**, *6*, e16852. [CrossRef]
17. Wichmann, H.E.; Gieger, C.; Illig, T.; MONICA/KORA Study Group. KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes. *Das Gesundh.* **2005**, *67*, S26–S30. [CrossRef]
18. Dorling, L.; Carvalho, S.; Allen, J.; González-Neira, A.; Luccarini, C.; Wahlström, C.; Pooley, K.A.; Parsons, M.T.; Fortuno, C.; Wang, Q.; et al. Breast Cancer Risk Genes—Association Analysis in More than 113,000 Women. *N. Engl. J. Med.* **2021**, *384*, 428–439. [CrossRef]
19. Liu, J.; Prager-van der Smissen, W.J.C.; Schmidt, M.K.; Collée, J.M.; Cornelissen, S.; Lamping, R.; Nieuwlaat, A.; Foekens, J.A.; Hooning, M.J.; Verhoef, S.; et al. Recurrent HOXB13 mutations in the Dutch population do not associate with increased breast cancer risk. *Sci. Rep.* **2016**, *6*, 30026. [CrossRef]
20. Dunning, A.M.; Healey, C.S.; Pharoah, P.D.; Teare, M.D.; Ponder, B.A.; Easton, D.F. A systematic review of genetic polymorphisms and breast Cancer risk. *Cancer Epidemiol. Biomark. Prev.* **1999**, *8*, 843–854.
21. Day, N.; Oakes, S.; Luben, R.; Khaw, K.T.; Bingham, S.; Welch, A.; Wareham, N. EPIC-Norfolk: Study design and characteristics of the cohort. european prospective investigation of Cancer. *Br. J. Cancer* **1999**, *80*, 95–103. [PubMed]
22. Kataoka, M.; Antoniou, A.; Warren, R.; Leyland, J.; Brown, J.; Audley, T.; Easton, D. Genetic models for the familial aggregation of mammographic breast density. *Cancer Epidemiol. Biomark. Prev.* **2009**, *18*, 1277–1284. [CrossRef] [PubMed]
23. John, E.M.; Hopper, J.L.; Beck, J.C.; Knight, J.A.; Neuhausen, S.L.; Senie, R.T.; Ziogas, A.; Andrulis, I.L.; Anton-Culver, H.; Boyd, N.; et al. The Breast Cancer Family Registry: An infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res.* **2004**, *6*, R375–R389. [CrossRef]
24. Awadalla, P.; Boileau, C.; Payette, Y.; Idaghdour, Y.; Goulet, J.P.; Knoppers, B.; Hamet, P.; Laberge, C.; CARTaGENE Project. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *Int. J. Epidemiol.* **2013**, *42*, 1285–1299. [CrossRef]
25. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From Fastq Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]
26. Feng, B.J. PERCH: A unified framework for disease gene prioritization. *Hum. Mutat.* **2017**, *38*, 243–251. [CrossRef] [PubMed]

27. Hu, C.; Hart, S.N.; Gnanaolivu, R.; Huang, H.; Lee, K.Y.; Na, J.; Gao, C.; Lilyquist, J.; Yadav, S.; Boddicker, N.J.; et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N. Engl. J. Med.* **2021**, *384*, 440–451. [CrossRef] [PubMed]
28. Lee, S.; Abecasis, G.R.; Boehnke, M.; Lin, X. Rare-variant association analysis: Study designs and statistical tests. *Am. J. Hum. Genet.* **2014**, *95*, 5–23. [CrossRef]
29. Liu, D.J.; Peloso, G.M.; Zhan, X.; Holmen, O.L.; Zawistowski, M.; Shuang, F.; Nikpay, M.; Auer, P.L.; Goel, A.; Zhang, H.; et al. Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **2014**, *46*, 200–204. [CrossRef]
30. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic. Acids Res.* **2018**, *46*, D1062–D1067. [CrossRef]
31. Zhao, J.; Zhu, D.; Zhang, X.; Zhang, Y.; Zhou, J.; Dong, M. TMEM206 promotes the malignancy of colorectal cancer cells by interacting with AKT and extracellular signal-regulated kinase signaling pathways. *J. Cell. Physiol.* **2019**, *234*, 10888–10898. [CrossRef] [PubMed]
32. Sun, W.S.; Fujimoto, J.; Tamaya, T. Clinical implications of coexpression of growth arrest-specific gene 6 and receptor tyrosine kinases Axl and Sky in human uterine leiomyoma. *Mol. Hum. Reprod.* **2003**, *9*, 701–707. [CrossRef] [PubMed]
33. Linger, R.M.; Keating, A.K.; Earp, H.S.; Graham, D.K. TAM receptor tyrosine kinases: Biologic functions, signaling, and potential therapeutic targeting in human cancer. *Adv. Cancer Res.* **2008**, *100*, 35–83. [CrossRef]
34. Zhu, S.; Wurdak, H.; Wang, Y.; Galkin, A.; Tao, H.; Li, J.; Lyssiotis, C.A.; Yan, F.; Tu, B.P.; Miraglia, L.; et al. A genomic screen identifies TYRO3 as a MITF regulator in melanoma. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17025–17030. [CrossRef]
35. Shahi, R.B.; De Brakeleer, S.; Caljon, B.; Pauwels, I.; Bonduelle, M.; Joris, S.; Fontaine, C.; Vanhoeij, M.; Van Dooren, S.; Teugels, E.; et al. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. *BMC Cancer* **2019**, *19*, 313. [CrossRef] [PubMed]
36. Verma, S.; Bakshi, D.; Sharma, V.; Sharma, I.; Shah, R.; Bhat, A.; Bhat, G.R.; Bhanu Sharma, B.; Wakhloo, A.; Kaul, S.; et al. Genetic variants of DNAH11 and LRFN2 genes and their association with ovarian and breast cancer. *Int. J. Gynaecol. Obs.* **2020**, *148*, 118–122. [CrossRef]
37. Fernandez-Navarro, P.; González-Neira, A.; Pita, G.; Díaz-Uriarte, R.; Tais Moreno, L.; Ederra, M.; Pedraz-Pingarrón, C.; Sánchez-Contador, C.; Antonio Vázquez-Carrete, J.; Moreo, P.; et al. Genome wide association study identifies a novel putative mammographic density locus at 1q12-q21. *Int. J. Cancer* **2015**, *136*, 2427–2436. [CrossRef] [PubMed]
38. Fachal, L.; Aschard, H.; Beesley, J.; Barnes, D.R.; Allen, J.; Kar, S.; Pooley, K.A.; Dennis, J.; Michailidou, K.; Turman, C.; et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat. Genet.* **2020**, *52*, 56–73. [CrossRef]
39. Lucci, M.A.; Orlandi, R.; Triulzi, T.; Tagliabue, E.; Balsari, A.; Villa-Moruzzi, E. Expression profile of tyrosine phosphatases in HER2 breast cancer cells and tumors. *Cell. Oncol.* **2010**, *32*, 361–372. [CrossRef]
40. Wu, K.-G.; He, X.-F.; Li, Y.-H.; Xie, W.-B.; Huang, X. Association between the XPD/ERCC2 Lys751Gln polymorphism and risk of cancer: Evidence from 224 case-control studies. *Tumour Biol.* **2014**, *35*, 11243–11259. [CrossRef]
41. Rump, A.; Benet-Pages, A.; Schubert, S.; Kuhlmann, J.D.; Janavičius, R.; Macháčková, E.; Foretová, L.; Kleibl, Z.; Lhota, F.; Zemankova, P.; et al. Identification and Functional Testing of ERCC2 Mutations in a Multi-national Cohort of Patients with Familial Breast- and Ovarian Cancer. *PLoS Genet.* **2016**, *12*, e1006248. [CrossRef] [PubMed]
42. Hauke, J.; Horvath, J.; Groß, E.; Gehrig, A.; Honisch, E.; Hackmann, K.; Schmidt, G.; Arnold, N.; Faust, U.; Sutter, C.; 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 Med.* **2018**, *7*, 1349–1358. [CrossRef] [PubMed]
43. Couch, F.J.; Shimelis, H.; Hu, C.; Hart, S.N.; Polley, E.C.; Na, J.; Hallberg, E.; Moore, R.; Thomas, A.; Lilyquist, J.; et al. Associations between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol.* **2017**, *3*, 1190–1196. [CrossRef] [PubMed]
44. Thompson, E.R.; Rowley, S.M.; Li, N.; McInerny, S.; Devereux, L.; Wong-Brown, M.W.; Trainer, A.H.; Mitchell, G.; Scott, R.J.; James, P.A.; et al. Panel Testing for Familial Breast Cancer: Calibrating the Tension between Research and Clinical Care. *J. Clin. Oncol.* **2016**, *34*, 1455–1459. [CrossRef] [PubMed]
45. Aloraifi, F.; McCartan, D.; McDevitt, T.; Green, A.J.; Bracken, A.; Geraghty, J. Protein-truncating variants in moderate-risk breast cancer susceptibility genes: A meta-analysis of high-risk case-control screening studies. *Cancer Genet.* **2015**, *208*, 455–463. [CrossRef]
46. Lee, K.; Seifert, B.A.; Shimelis, H.; Ghosh, R.; Crowley, S.B.; Carter, N.J.; Doonanco, K.; Foreman, A.K.; Ritter, D.I.; Jimenez, S.; et al. Clinical validity assessment of genes frequently tested on hereditary breast and ovarian cancer susceptibility sequencing panels. *Genet. Med.* **2019**, *21*, 1497–1506. [CrossRef]
47. Ku, C.S.; Cooper, D.N.; Patrinos, G.P. The Rise and Rise of Exome Sequencing. *Public Health Genom.* **2016**, *19*, 315–324. [CrossRef]
48. Lelieveld, S.H.; Spielmann, M.; Mundlos, S.; Veltman, J.A.; Gilissen, C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum. Mutat.* **2015**, *36*, 815–822. [CrossRef]
49. Sims, D.; Sudbery, I.; Ilott, N.E.; Heger, A.; Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* **2014**, *15*, 121–132. [CrossRef]
50. Wu, L.; Schaid, D.J.; Sicotte, H.; Wieben, E.D.; Li, H.; Petersen, G.M. Case-only exome sequencing and complex disease susceptibility gene discovery: Study design considerations. *J. Med. Genet.* **2015**, *52*, 10–16. [CrossRef]

51. Gusev, A.; Lee, S.H.; Trynka, G.; Finucane, H.; Vilhjálmsson, B.J.; Xu, H.; Zang, C.; Ripke, S.; Bulik-Sullivan, B.; Stahl, E.; et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **2014**, *95*, 535–552. [CrossRef]
52. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
53. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics* **2013**, *1303*, 3997.
54. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, A.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef] [PubMed]
55. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef] [PubMed]
56. 1000 Genomes Project Consortium; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [CrossRef]
57. Liu, X.; Wu, C.; Li, C.; Boerwinkle, E. dbNSFP V3.0: A One-Stop Database of Functional Predictions and Annotations for Hu-man Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.* **2016**, *37*, 235–241. [CrossRef]
58. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* **2016**, *536*, 285–291. [CrossRef]
59. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef]
60. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alföldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [CrossRef]
61. Forbes, S.A.; Beare, D.; Gunasekaran, P.; Leung, K.; Bindal, N.; Boutselakis, H.; Ding, M.; Bamford, S.; Cole, C.; Ward, S.; et al. COSMIC: Exploring the World's Knowledge of Somatic Mutations in Human Cancer. *Nucleic Acids Res.* **2015**, *43*, D805–D811. [CrossRef]
62. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the Sift Algorithm. *Nat. Protoc.* **2009**, *4*, 1073–1081. [CrossRef] [PubMed]
63. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting Functional Effect of Human Missense Mutations Using Polyphen-2. *Curr. Protoc. Hum. Genet.* **2013**, *7*, 7–20. [CrossRef] [PubMed]
64. Kircher, M.; Witten, D.M.; Jain, P.; O'Roak, B.J.; Cooper, G.M.; Shendure, J. A General Framework for Estimating the Relative Pathogenicity of Human Genetic Variants. *Nat. Genet.* **2014**, *46*, 310–315. [CrossRef] [PubMed]
65. Reimand, J.; Bader, G.D. Systematic Analysis of Somatic Mutations in Phosphorylation Signaling Predicts Novel Cancer Driv-ers. *Mol. Syst. Biol.* **2013**, *9*, 637. [CrossRef]