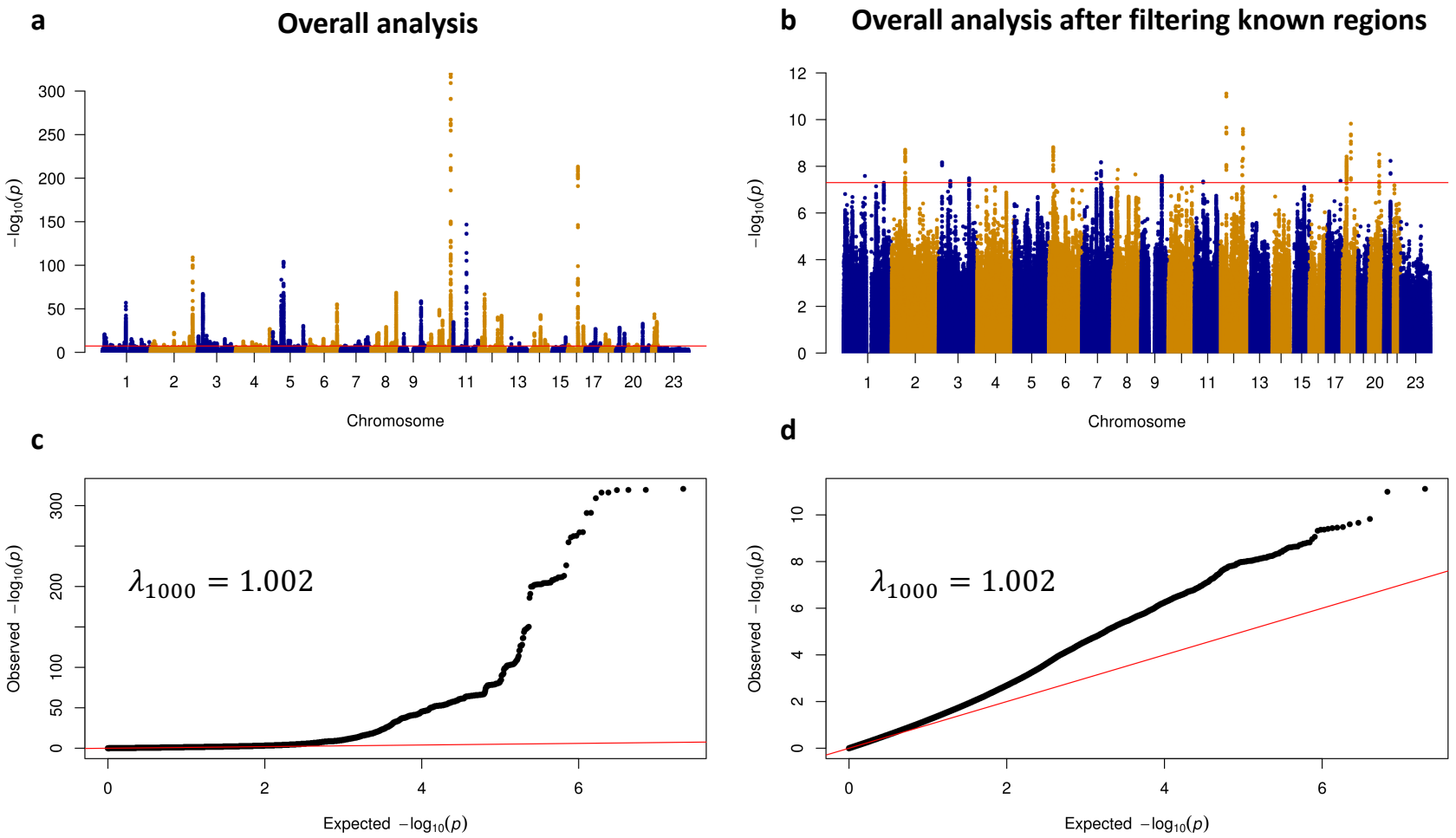
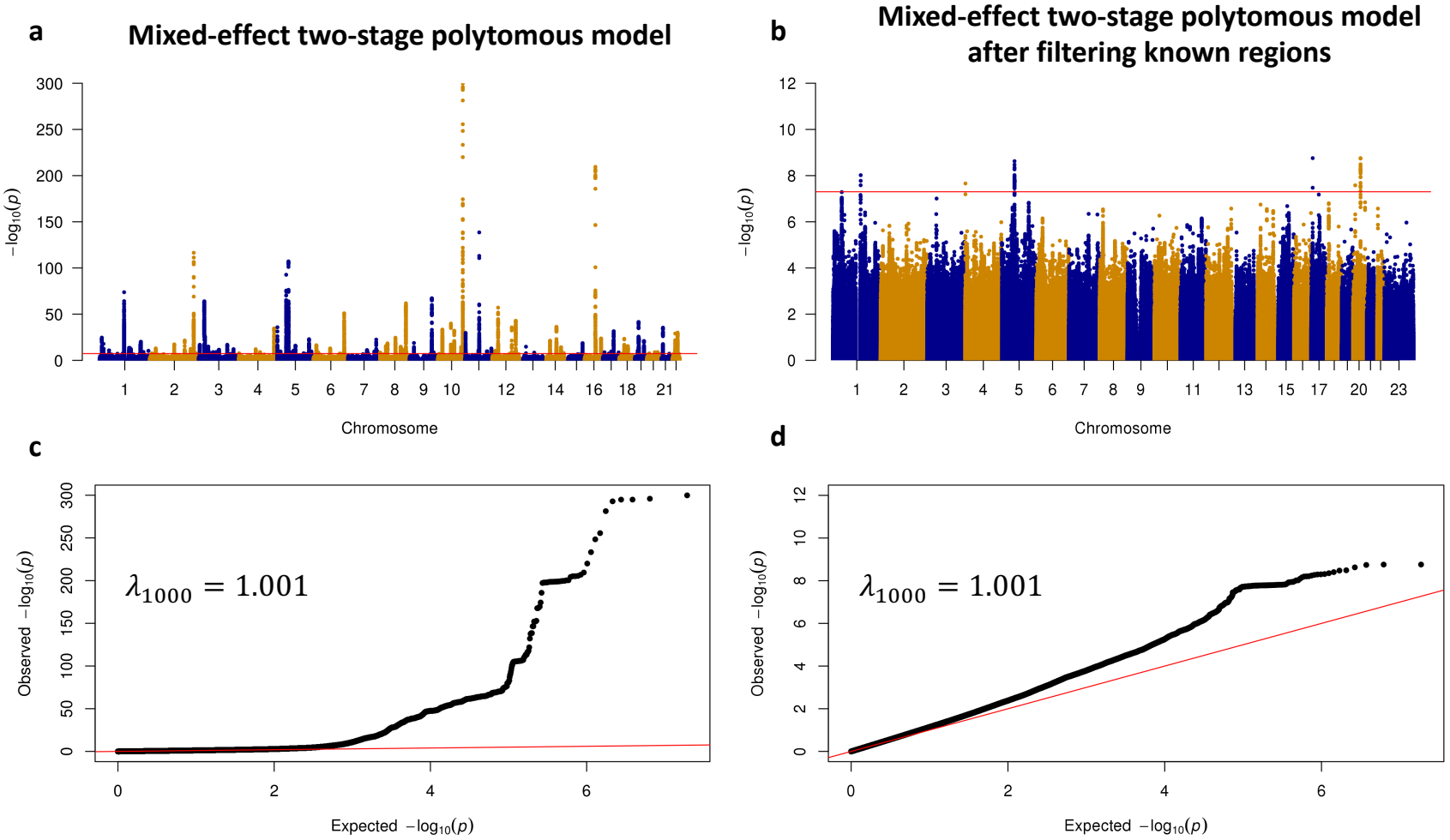
Supplementary figure 1. Variants associations with overall breast cancer risk identified using standard logistic regression (n = 133,384 cases, n = 113,789 controls). **a)** Manhattan plot showing -$\log_{10}P$ values for variant associations with breast cancer risk. **b)** Manhattan plot after excluding previous known regions (Online Methods) **c)** Quantile-Quantile (Q-Q) plot of observed P-values versus expected P-values for all variants. **d)** QQ plot[1] after excluding previous known regions. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5x 10^{-8}).



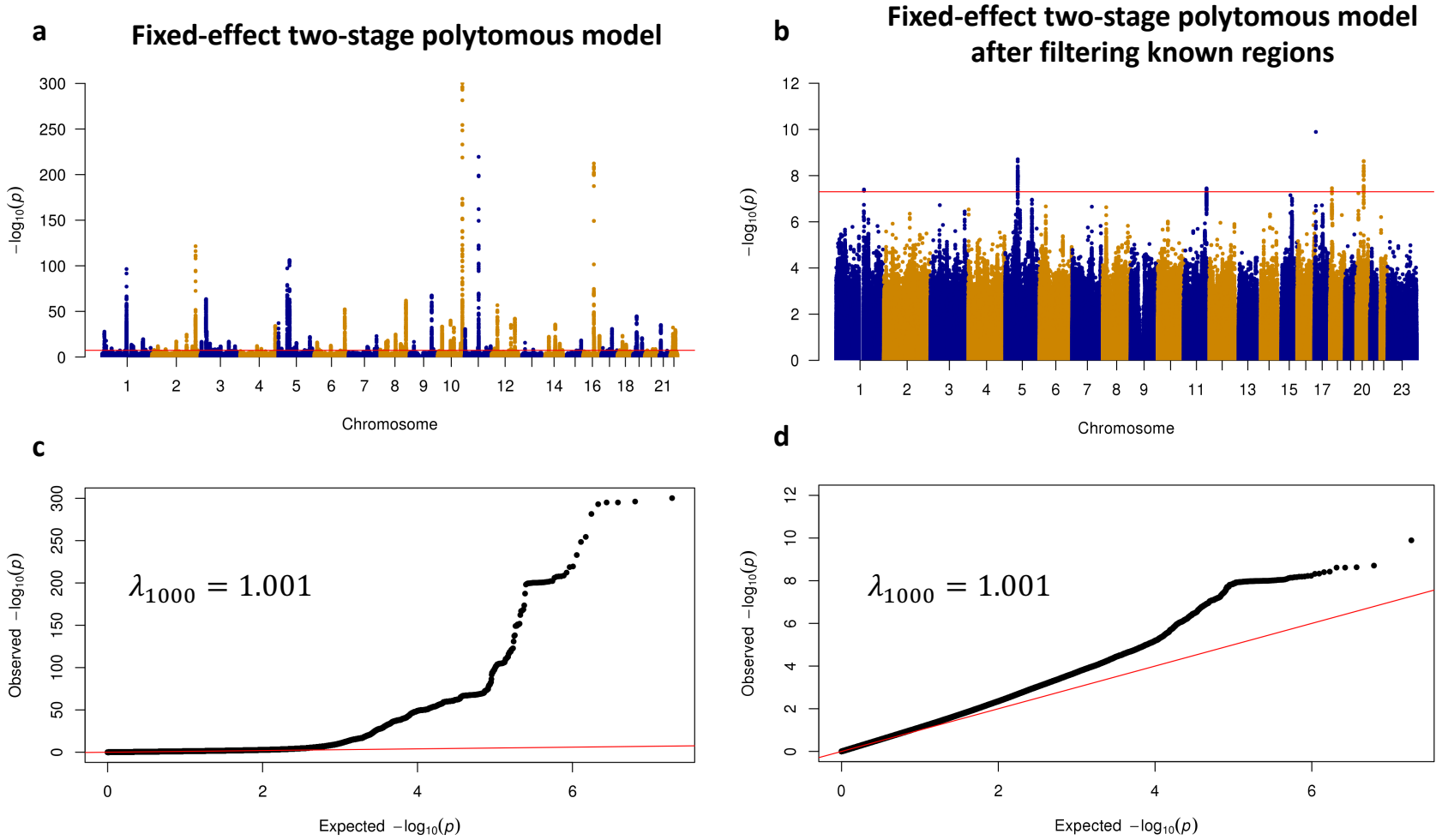
1) λ_{1000} scale the genomic inflation factor λ to a study with sample size of 1000 cases and 1000 controls using the formula $\lambda_{1000} = 1 + 500 * (\lambda - 1)/(\frac{1}{n_{cases}} + \frac{1}{n_{control}})$

Supplementary figure 2. Variant associations with breast cancer risk using a mixed-effect two-stage model (**Oline Methods**) accounting for tumor heterogeneity according to the ER, PR, HER2, and grade (n = 106,278 invasive cases, n = 91,477 controls). **a)** Manhattan plot showing -$\log_{10}P$ values for variant associations with breast cancer risk. **b)** Manhattan plot showing -$\log_{10}P$ values for variant associations with breast cancer risk after excluding previously known regions (Online Methods) and 22 loci identified through standard logistic regression analysis (**Supplementary Figure 2**). **c)** QQ plot[1] of observed P-values versus expected P-values for all variants. **d)** QQ plot of observed P-values versus expected P-values for remaining variants after excluding previously known regions and 22 loci identified through standard logistic regression analysis. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5 x 10^{-8}).



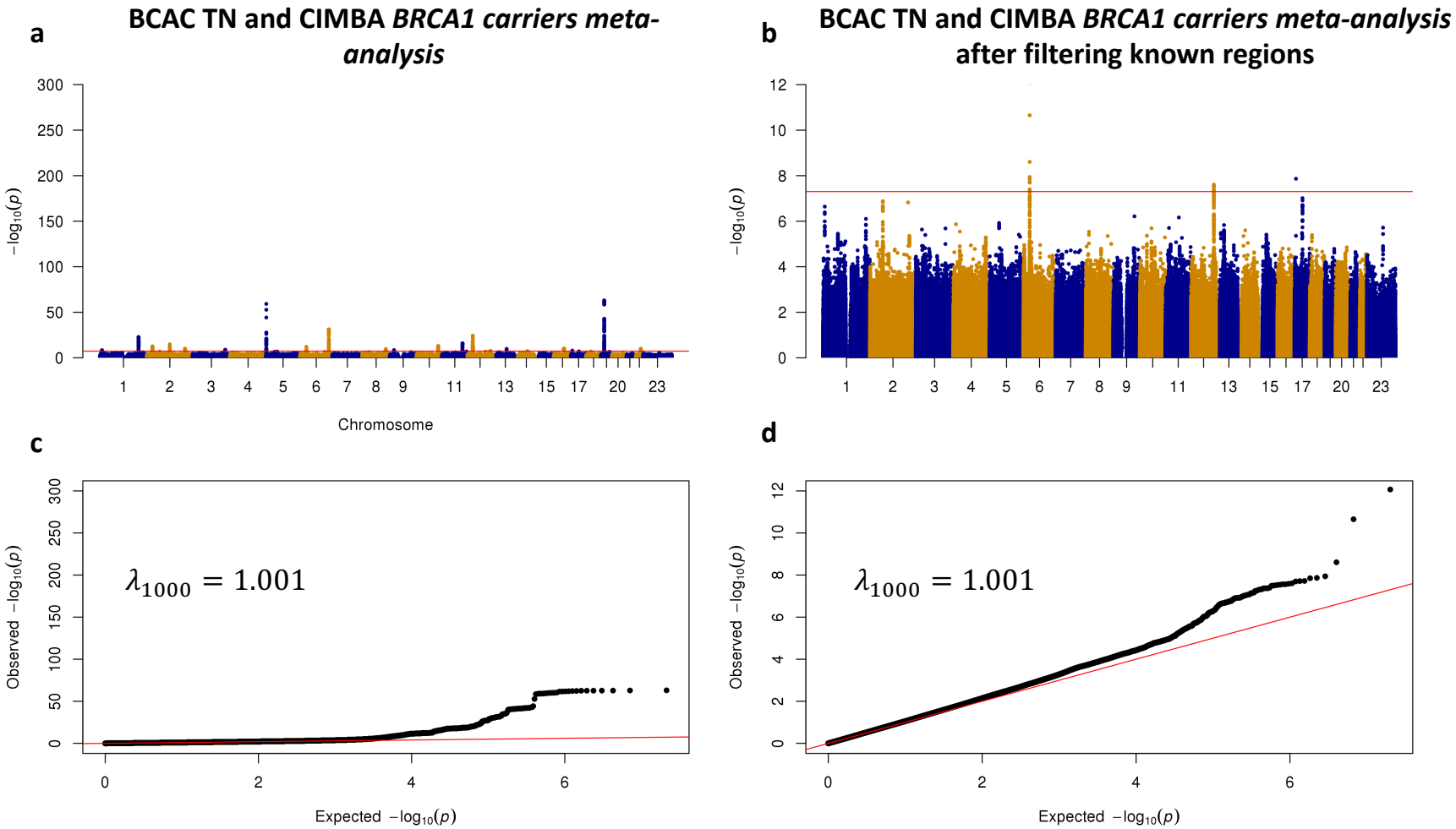
1) λ_{1000} scale the genomic inflation factor λ to a study with sample size of 1000 cases and 1000 controls using the formula $\lambda_{1000} = 1 + 500 * (\lambda - 1)/(\frac{1}{n_{cases}} + \frac{1}{n_{control}})$

Supplementary figure 3. Variant associations with breast cancer risk using a fixed-effect two-stage model (**Oline Methods**) accounting for tumor heterogeneity according to the ER, PR, HER2, and grade (n = 106,278 invasive cases, n = 91,477 controls). **a)** Manhattan plot showing -$\log_{10}$P values for variant associations with breast cancer risk. **b)** Manhattan plot showing -$\log_{10}$P values for variant associations with breast cancer risk after excluding previously known regions (Online Methods) and 22 loci identified through standard logistic regression analysis (**Supplementary Figure 2**). **c)** QQ plot[1] of observed P-values versus expected P-values for all variants. **d)** QQ plot of observed P-values versus expected P-values for remaining variants after excluding previously known regions and 22 loci identified through standard analysis. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5x 10^{-8}).



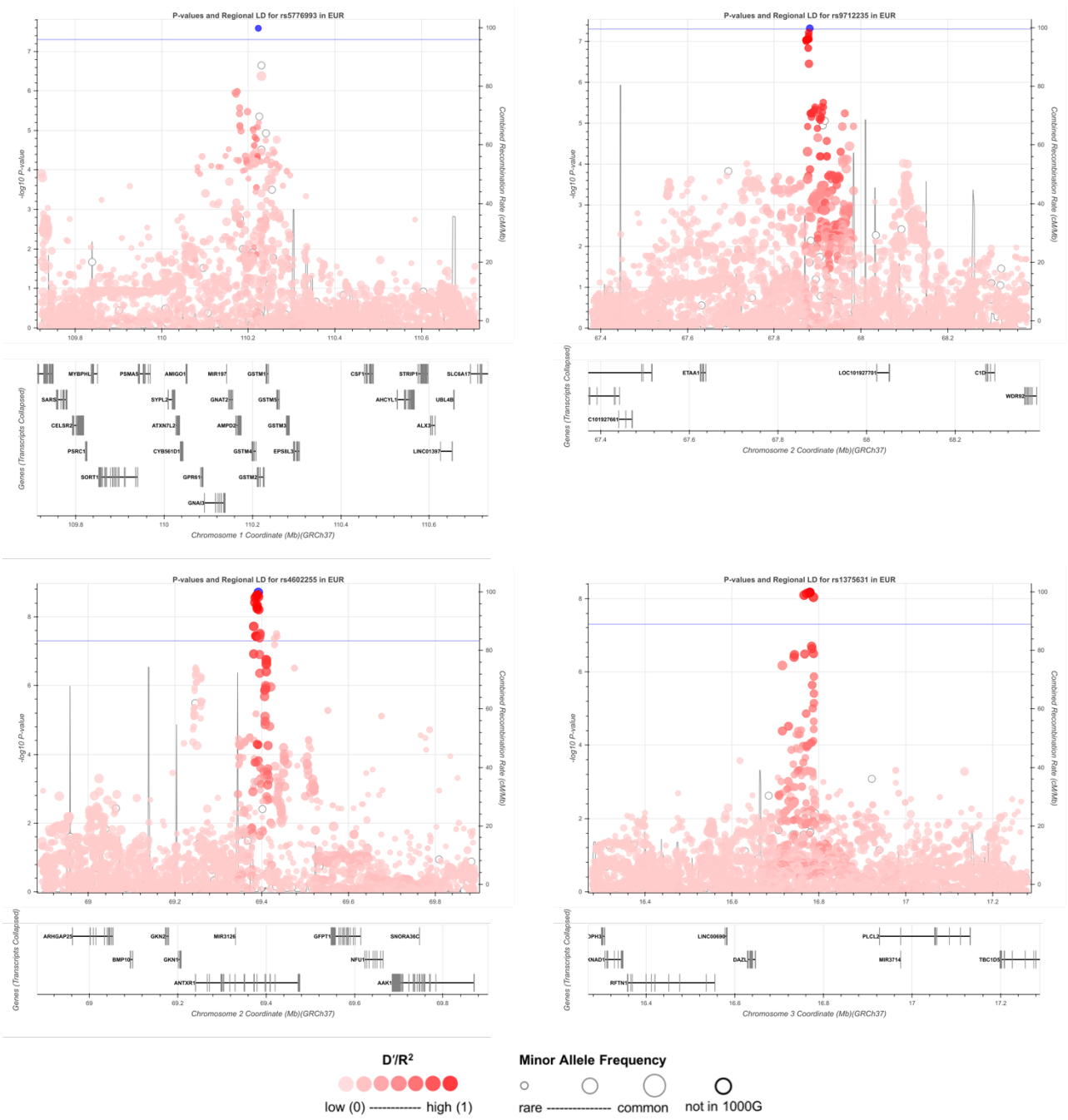
1) λ_{1000} scale the genomic inflation factor λ to a study with sample size of 1000 cases and 1000 controls using the formula $\lambda_{1000} = 1 + 500 * (\lambda - 1)/(\frac{1}{n_{cases}} + \frac{1}{n_{control}})$

Supplementary figure 4. Variant association with triple-negative breast cancer risk using a fixed-effect meta-analysis of results between BCAC TN and CIMBA *BRCA1 carriers* (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). **a)** Manhattan plot showing -$\log_{10}$P values for variant associations with triple-negative breast cancer risk. **b)** Manhattan plot showing -$\log_{10}$P values for variant associations with triple-negative breast cancer risk after excluding previously known regions (Online Methods). **c)** QQ plot[1] of observed P-values versus expected P-values for all variants **d)** QQ plot of observed P-values versus expected P-values for remaining variants after excluding previously known regions. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5x 10^{-8}).

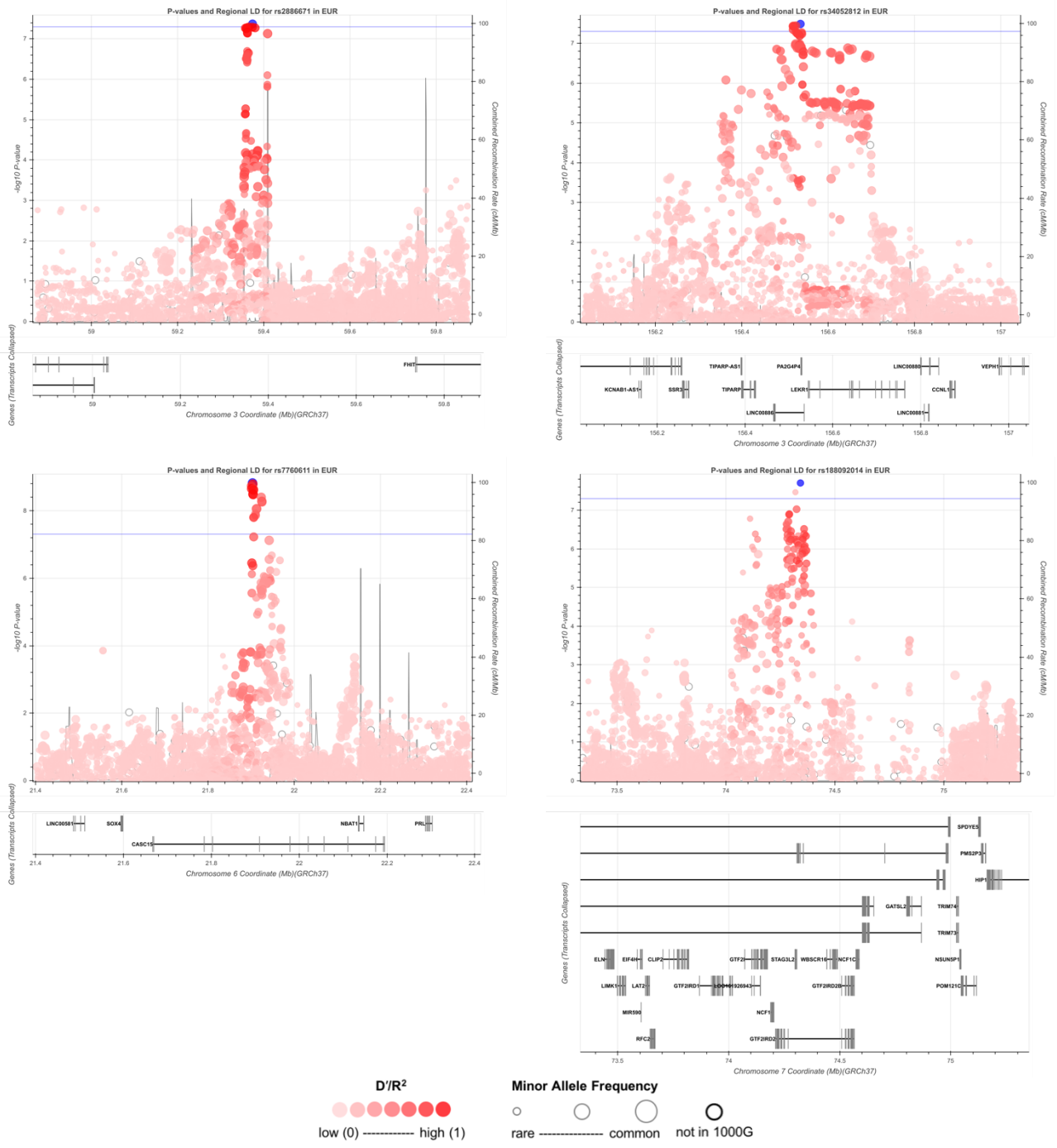


1) λ_{1000} scale the genomic inflation factor λ to a study with sample size of 1000 cases and 1000 controls using the formula $\lambda_{1000} = 1 + 500 * (\lambda - 1)/(\frac{1}{n_{cases}} + \frac{1}{n_{control}})$

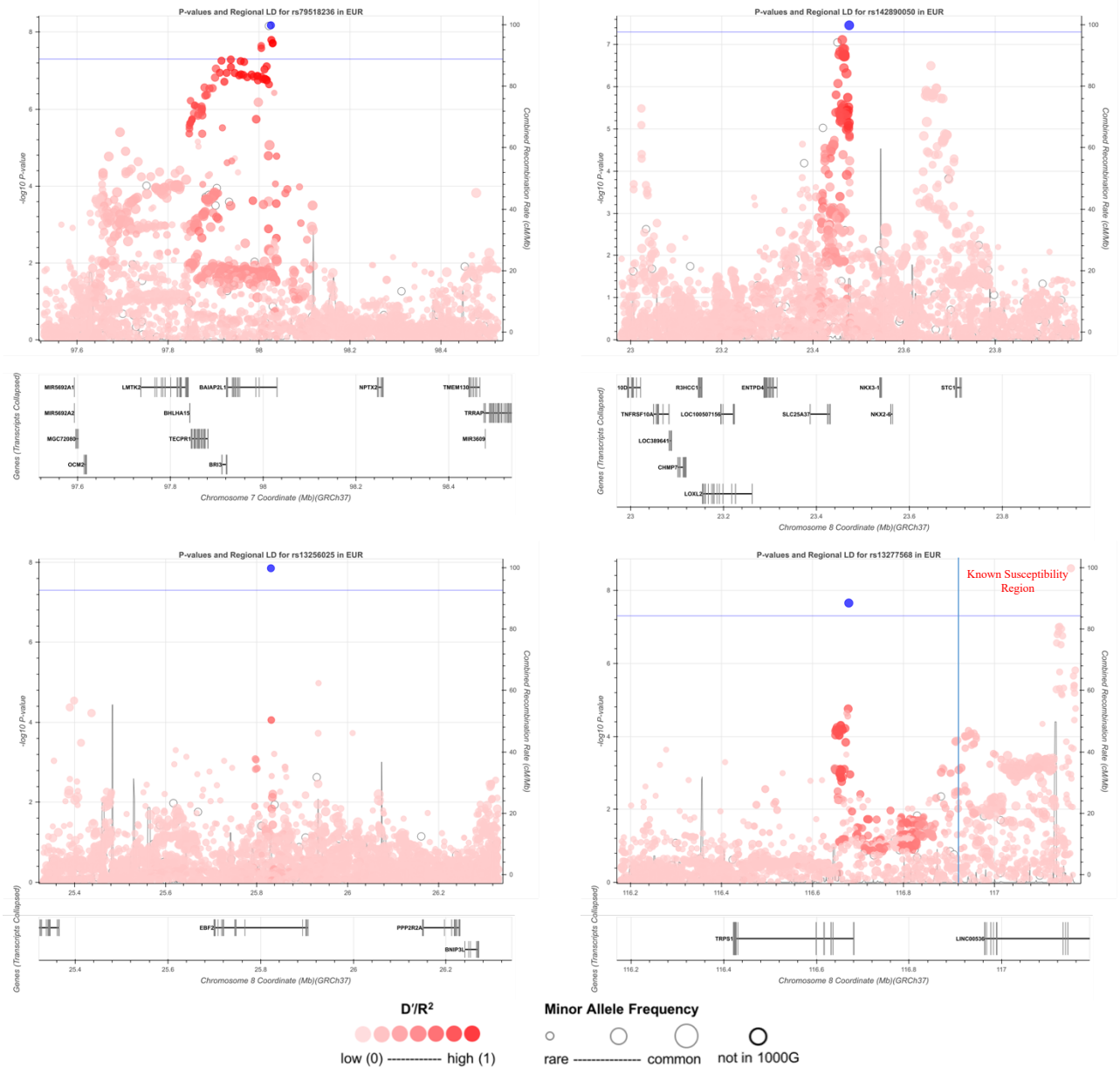
Supplementary figure 5. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



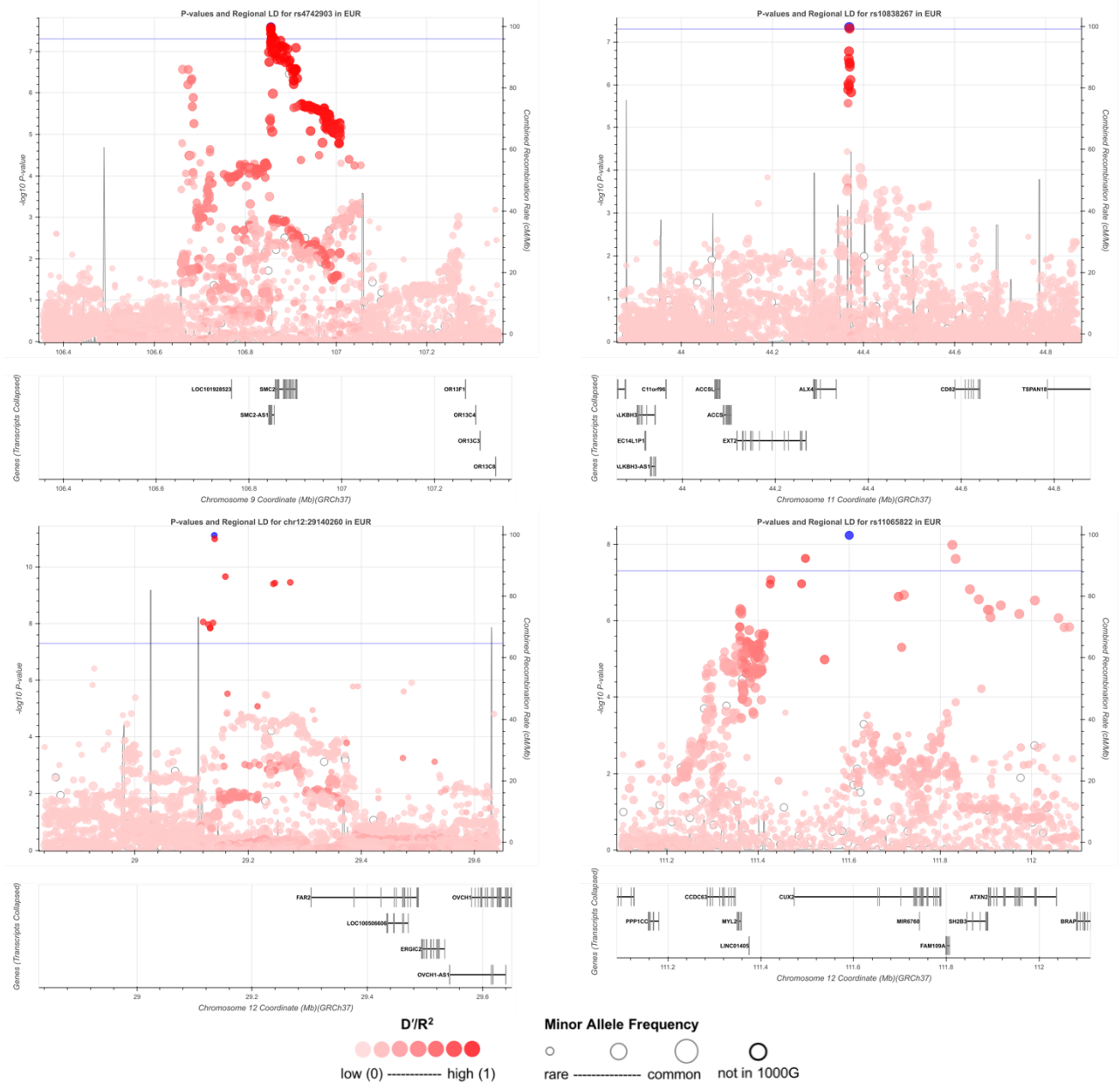
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



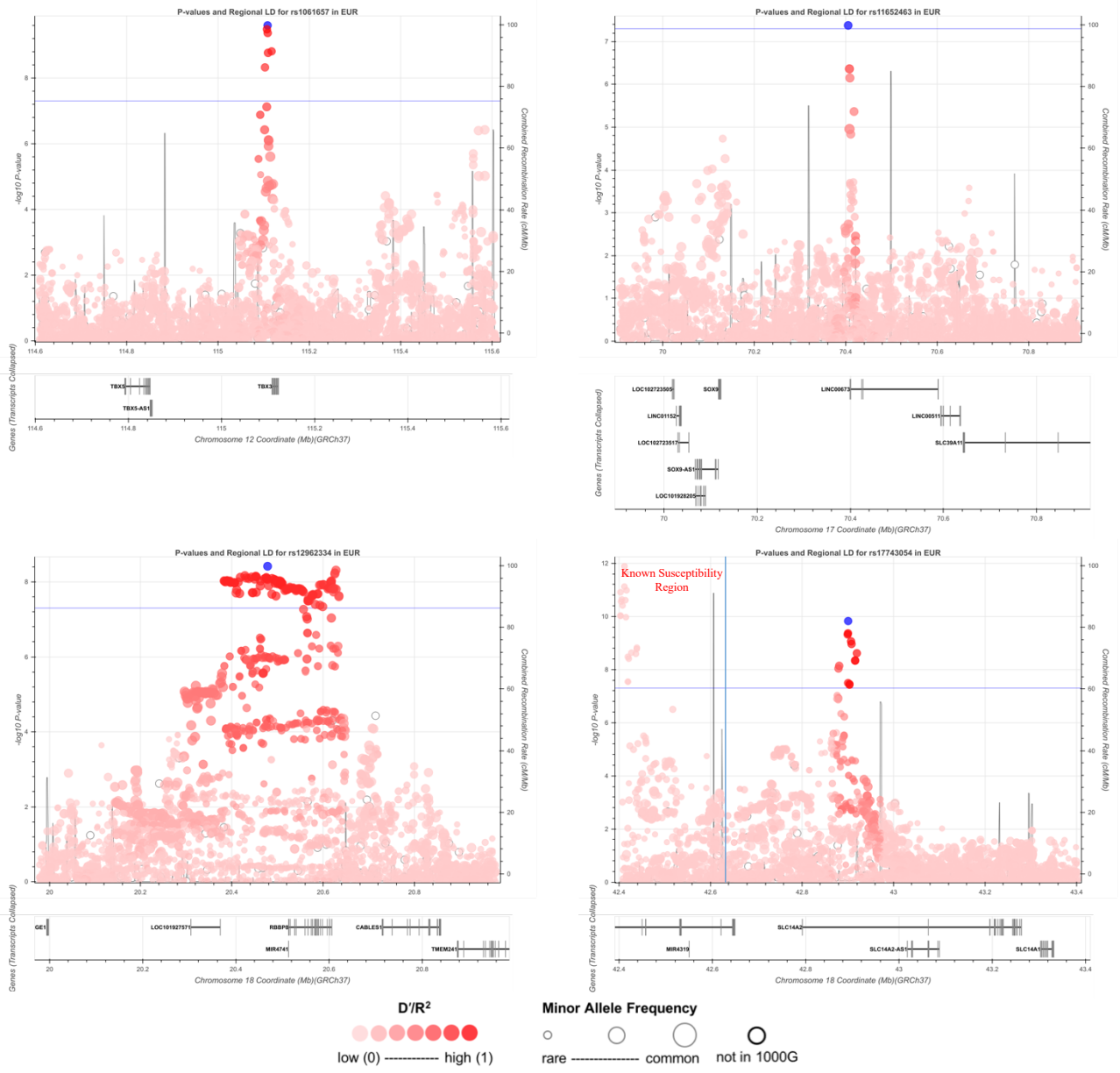
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



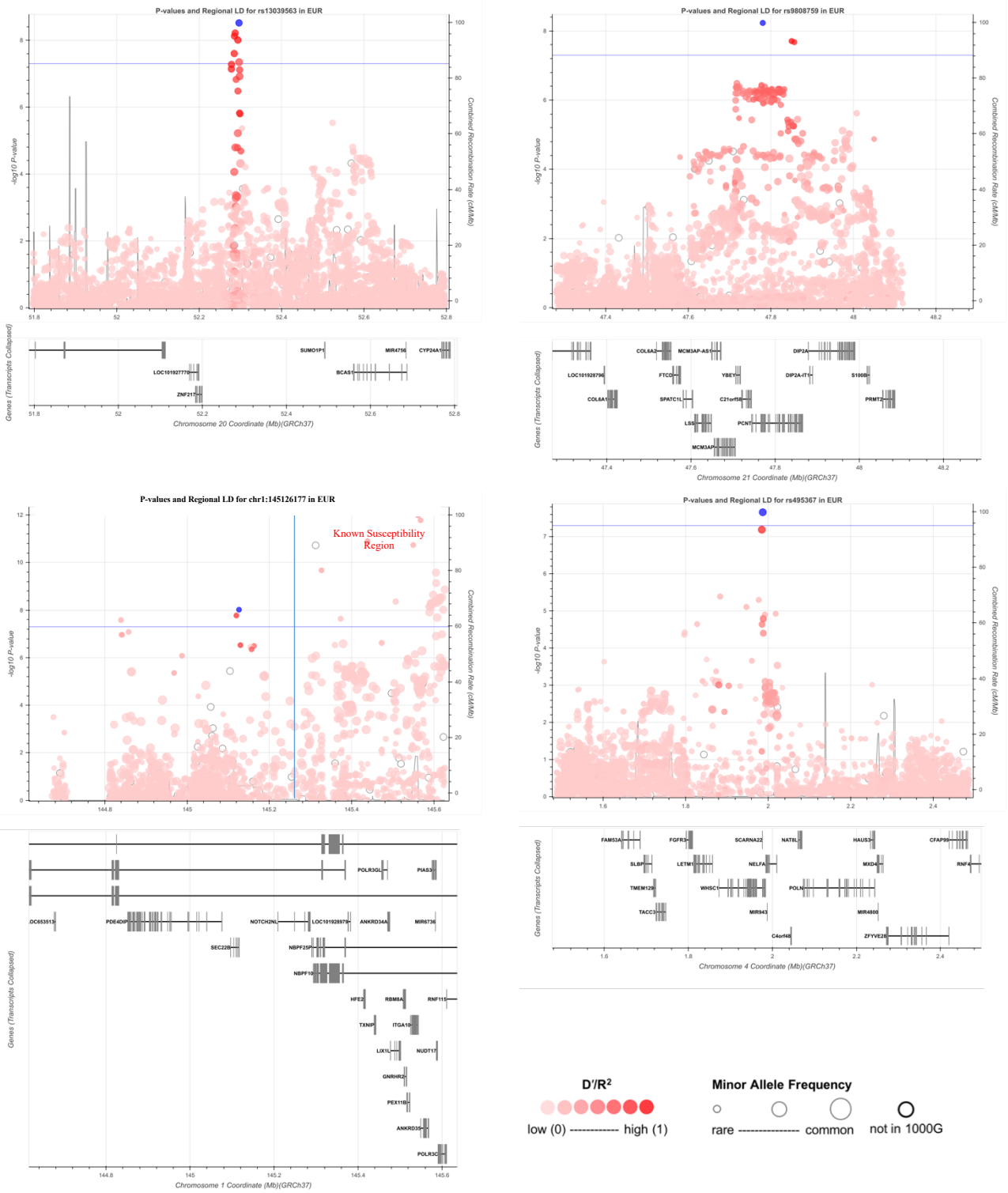
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



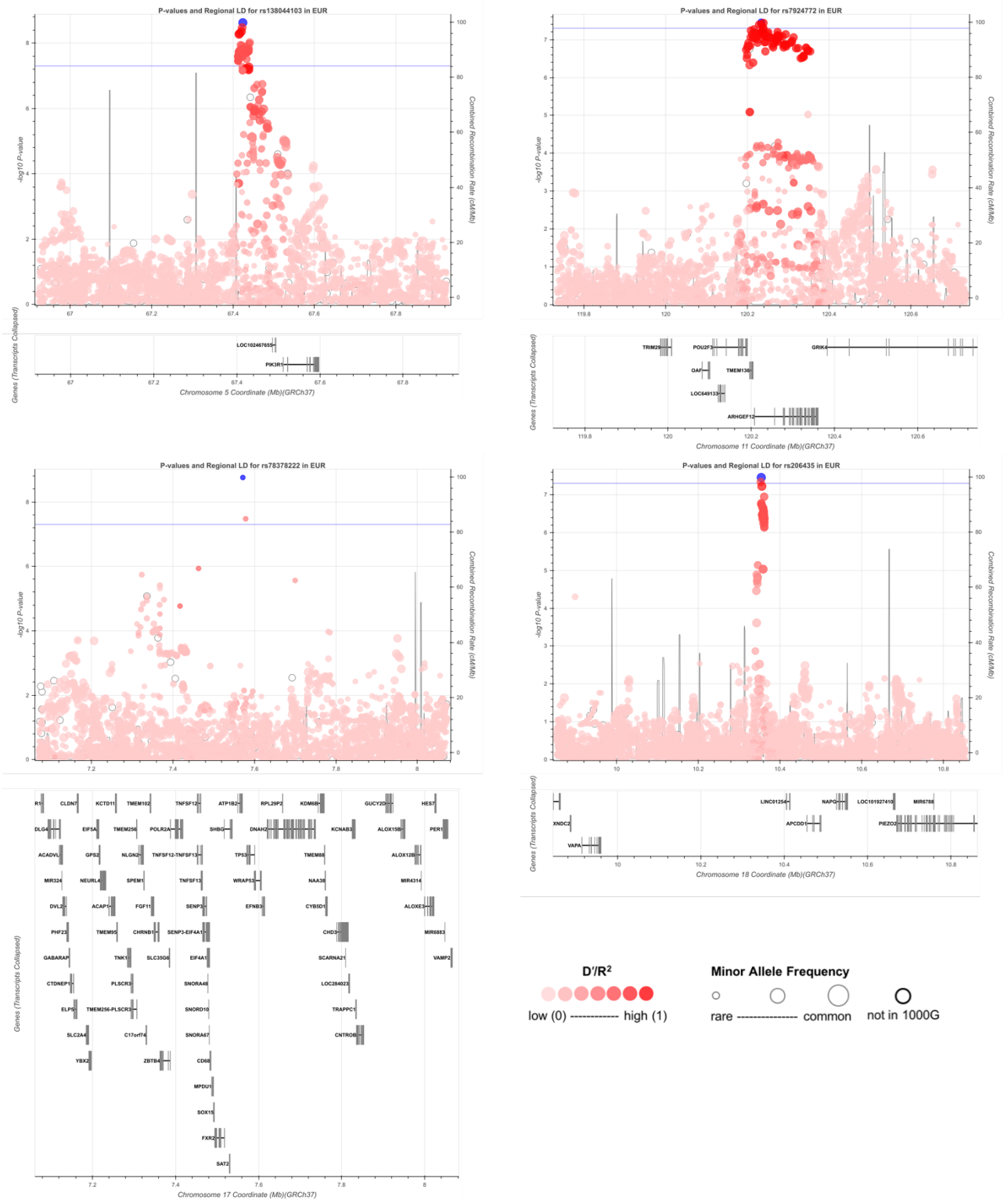
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



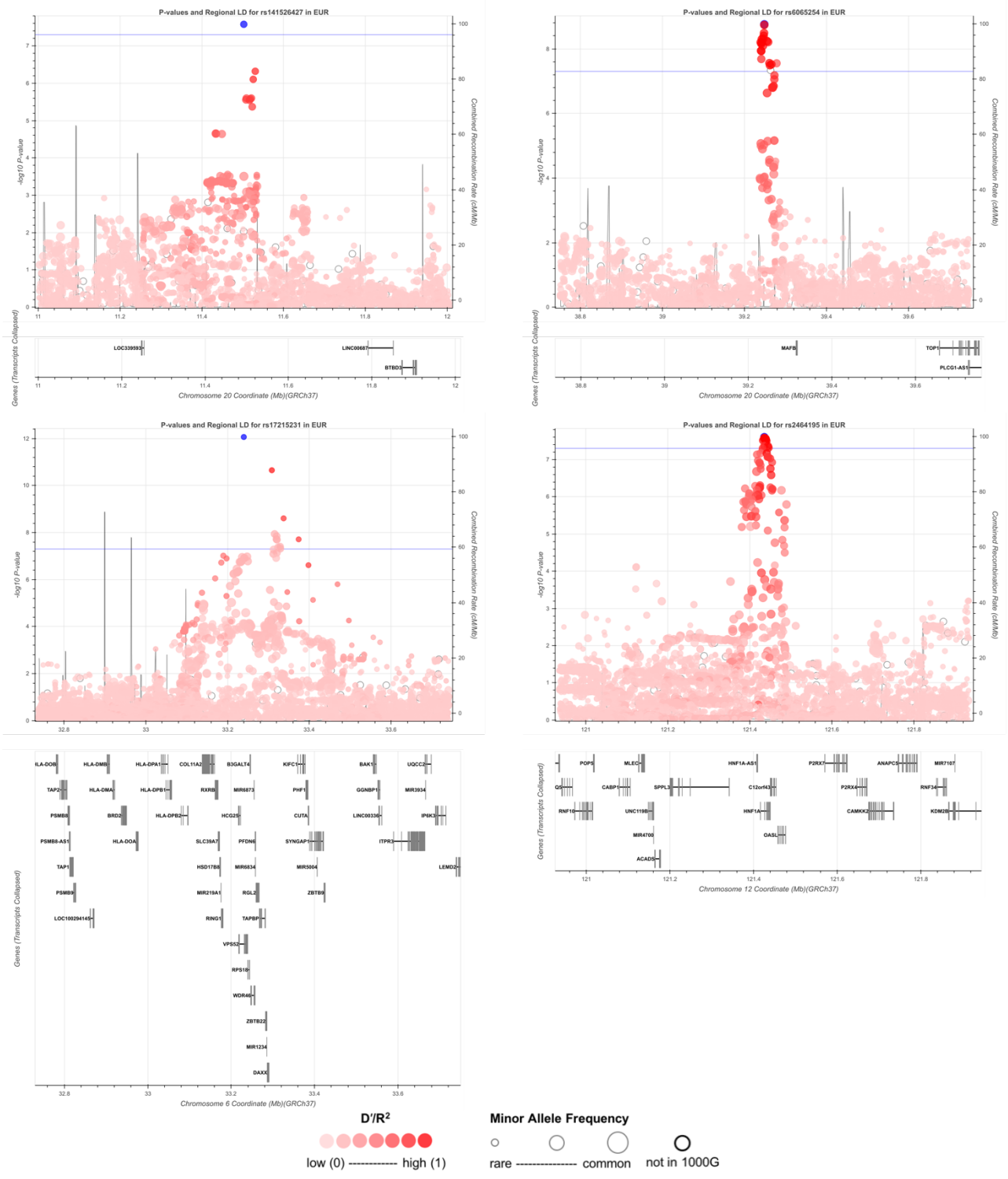
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).



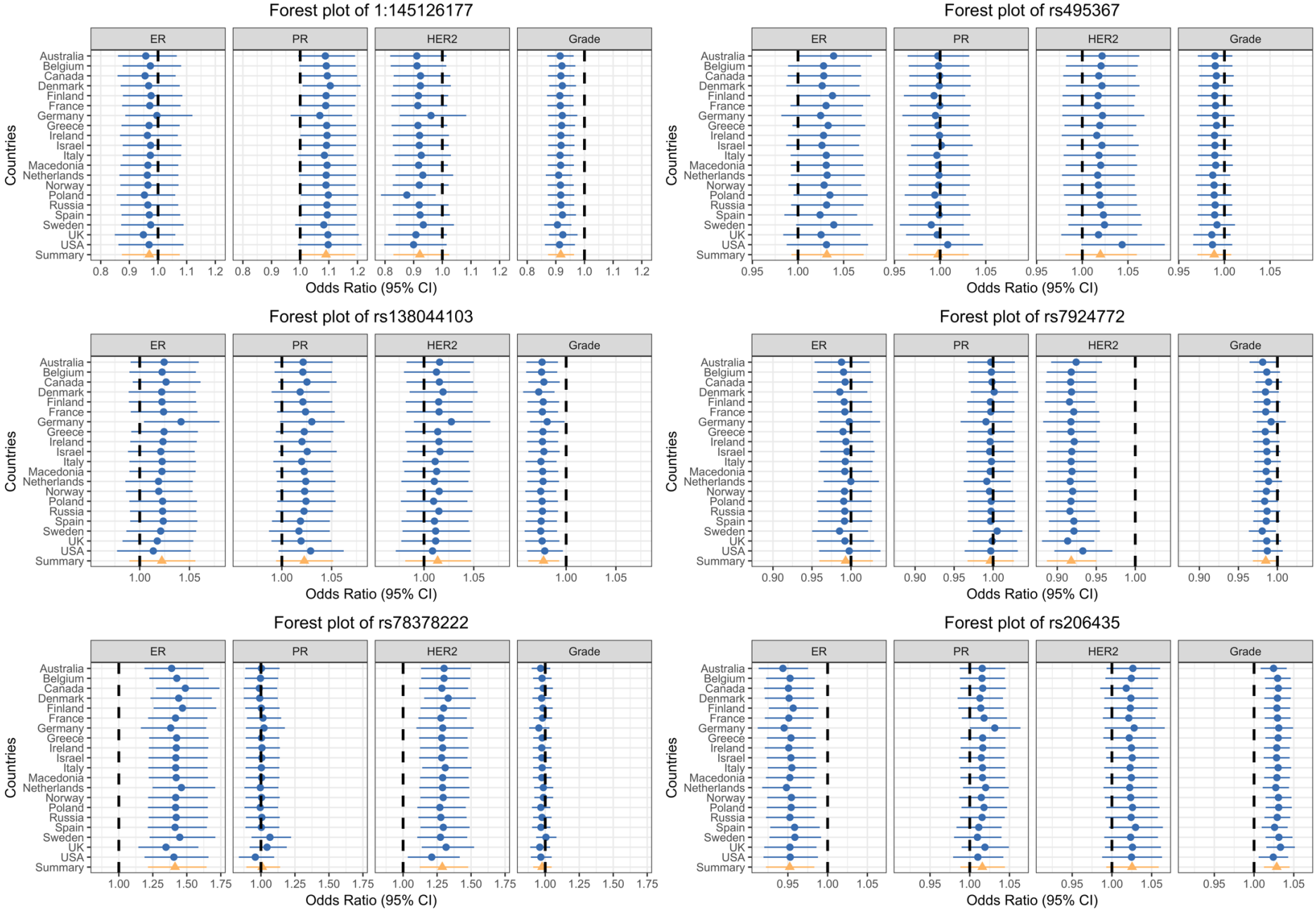
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).

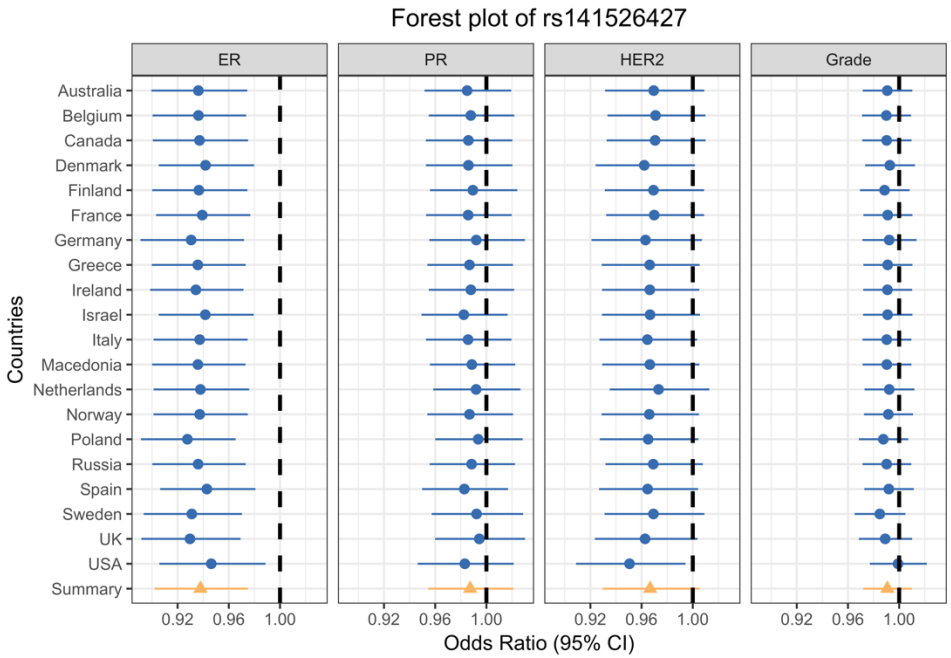


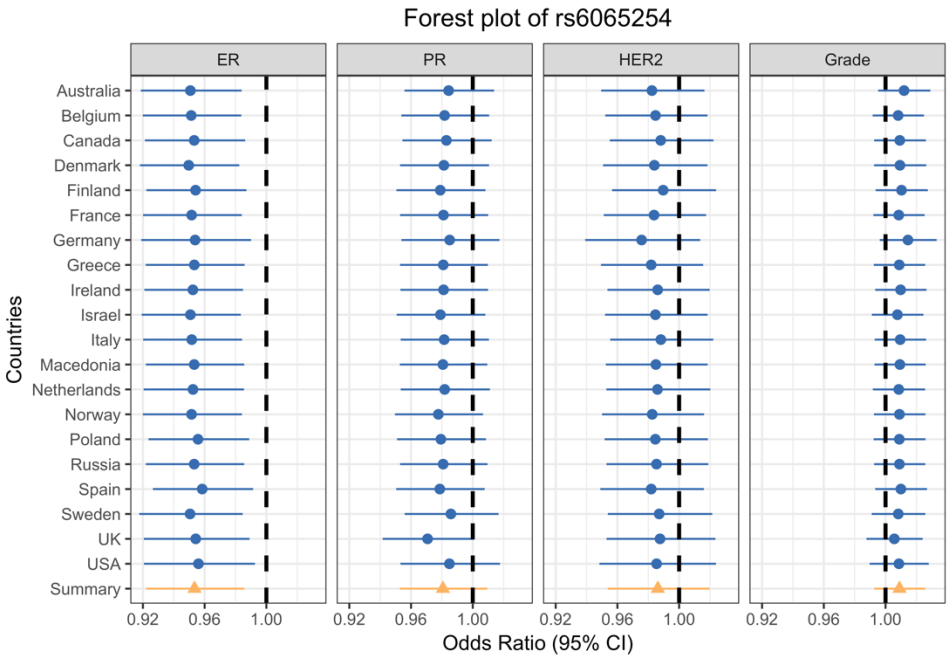
Supplementary figure 5 continued. Regional plots of the 32 identified breast cancer variants. The first 22 variants were identified through standard logistic regression (n = 133,384 cases, n = 113,789 controls), the following eight variants were identified through two-stage polytomous regression (n = 106,278 invasive cases, n = 91,477 controls), the last two variants were identified through meta-analysis of BCAC triple-negative and CIMBA *BRCA1* carriers (BCAC: n = 8,602 effective triple-negative cases, n = 91,477 controls; CIMBA *BRCA1* carriers: n = 9,414 cases, n = 9,494 controls). Plotted area is showing ±500 KB region around the identified susceptibility variant. P-values are raw p-values from two-tailed z-test statistics. Bonferroni correction was used to account for multiple testing (cut off P-value = 5×10^{-8}).

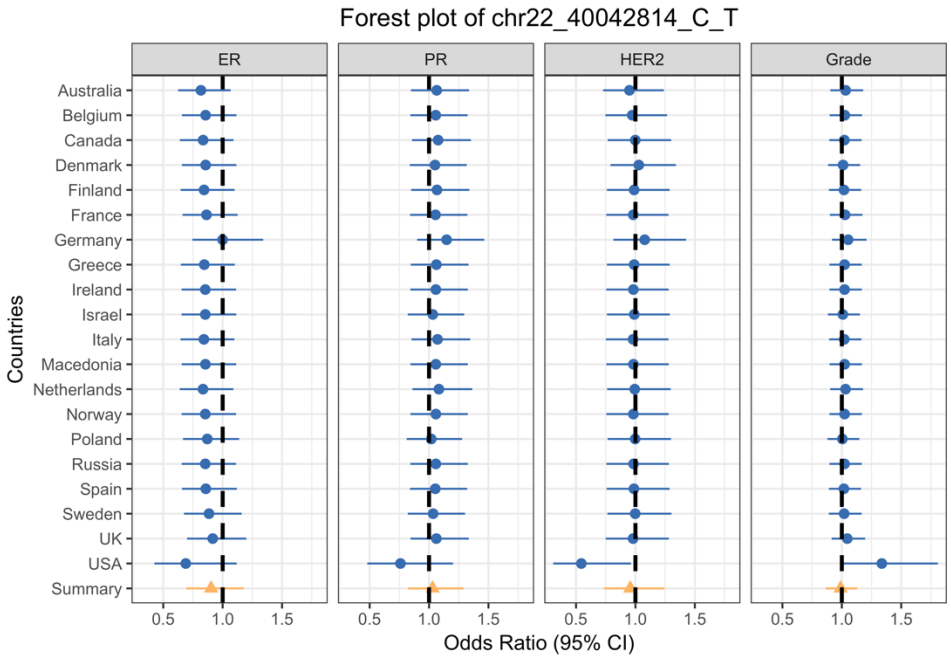


Supplementary figure 6. Country Specific sensitivity analysis of eight novel genome-wide significant loci identified using the two-stage regression models (n = 106,278 invasive cases, n = 91,477 controls), and chr22:40042814 which was dropped since the signal was observed only in studies from the USA.



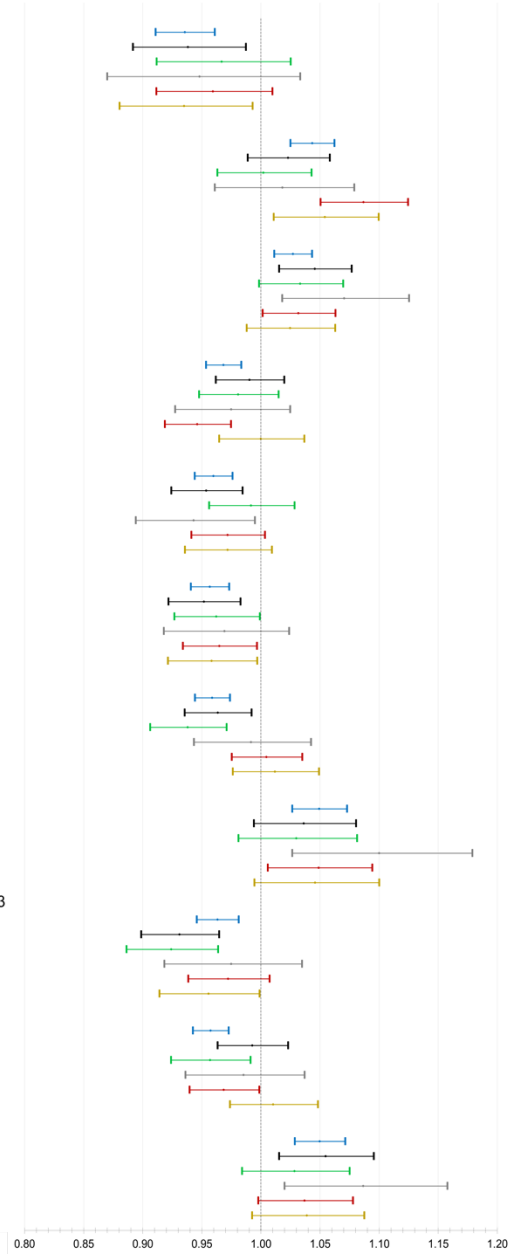

Forest plot of rs141526427
ER
PR
HER2
Grade
Australia
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Ireland
Israel
Italy
Macedonia
Netherlands
Norway
Poland
Russia
Spain
Sweden
UK
USA
Summary
Countries
0.92
0.96
1.00
Odds Ratio (95% CI)


Forest plot of rs6065254
ER
PR
HER2
Grade
Australia
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Ireland
Israel
Italy
Macedonia
Netherlands
Norway
Poland
Russia
Spain
Sweden
UK
USA
Summary
Countries
0.92
0.96
1.00
Odds Ratio (95% CI)


Forest plot of chr22_40042814_C_T
ER
PR
HER2
Grade
Australia
Belgium
Canada
Denmark
Finland
France
Germany
Greece
Ireland
Israel
Italy
Macedonia
Netherlands
Norway
Poland
Russia
Spain
Sweden
UK
USA
Summary
Countries
0.5
1.0
1.5
Odds Ratio (95% CI)

Supplementary Figure 7. Associations[1] between novel susceptibility variants identified using standard logistic regression with intrinsic-like breast cancer subtypes[2] (right panel, n = 106,278 invasive cases, n = 91,477 controls) and the second-stage heterogeneity p-values from the two-stage polytomous logistic regression model (left panel, n = 106,278 invasive cases, n = 91,477 controls).

				Global etiologic	Tumor characteristic heterogeneity *P*[4]			
rs5776993	1	110,222,901	0.12	0.50	0.29	0.54	0.35	0.29
rs9712235	2	67,881,757	0.26	6.70E-03	3.68E-04	0.02	0.03	0.85
rs4602255	2	69,392,128	0.45	0.17	0.69	0.67	0.62	0.11
rs1375631	3	16,778,867	0.5	0.40	0.09	0.21	0.57	0.55
rs2886671	3	59,373,745	0.43	0.67	0.75	0.52	0.43	0.39
rs34052812	3	156,535,958	0.33	0.98	0.74	0.72	0.62	0.97
rs7760611	6	21,903,533	0.47	1.40E-03	0.03	0.28	0.17	0.87
rs188092014	7	74,341,926	0.19	0.19	0.05	0.30	0.78	0.19
rs79518236	7	98,026,554	0.23	1.57E-03	0.37	0.02	0.30	1.94E-03
rs142890050	8	23,480,253	0.46	0.48	0.60	0.69	0.91	0.26
rs13256025	8	25,831,778	0.2	0.69	0.32	0.59	0.81	0.80



1 Per-minor allele odds ratio (95% confidence limits)

2. Luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); HER2-enriched-like (ER- and PR-, HER2+); triple-negative (ER-, PR-, HER2-)

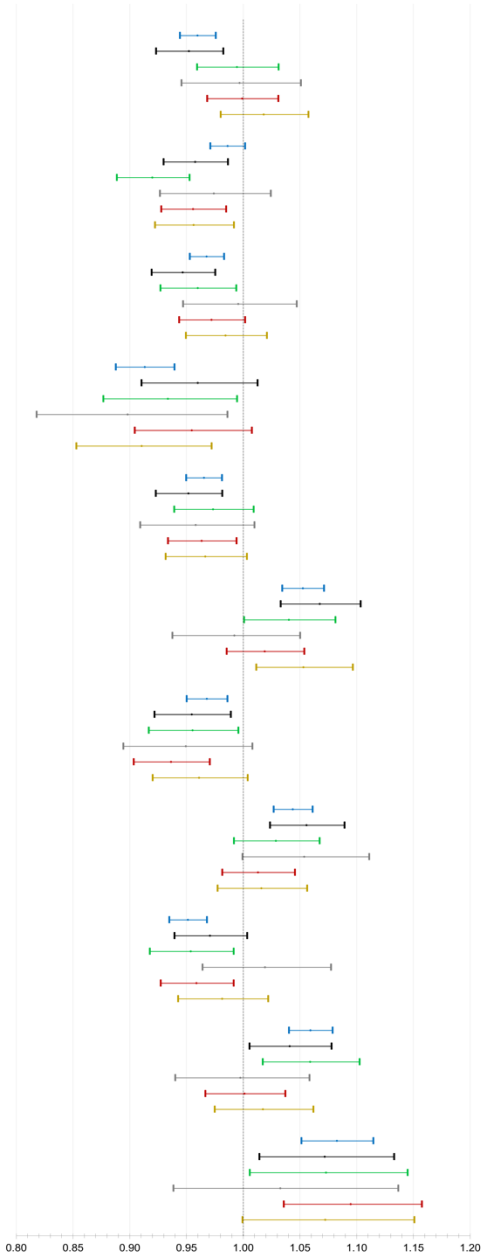
3. Based on a mixed-effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER was entered into the model as a fixed-effect term and PR, HER2, and grade were entered into the model as random-effect terms.

4. Results from second stage case-case parameters from a fixed effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER, PR, HER2, and grade are mutually adjusted for each other

5. Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)

Supplementary Figure 7 continued. Associations[1] between novel susceptibility variants identified using standard logistic regression with intrinsic-like breast cancer subtypes[2] (right panel, n = 106,278 invasive cases, n = 91,477 controls) and the second-stage heterogeneity p-values from the two-stage polytomous logistic regression model (left panel, n = 106,278 invasive cases, n = 91,477 controls).

				Global etiologic	Tumor characteristic heterogeneity *P*[4]			
rs13277568	8	116,679,547	0.37	0.06	0.90	0.19	0.17	0.34
rs4742903	9	106,856,793	0.44	2.80E-03	0.34	0.37	0.05	0.01
rs10838267	11	44,368,892	0.45	0.85	0.39	0.54	0.69	0.81
12:29140260	12	29,140,260	0.09	0.39	0.31	0.63	0.67	0.37
rs11065822	12	111,600,134	0.37	0.87	0.52	0.70	0.69	0.78
rs1061657	12	115,108,136	0.26	0.04	0.07	0.45	0.35	0.38
rs11652463	17	70,405,095	0.31	0.21	0.63	0.17	0.90	0.92
rs12962334	18	20,477,934	0.32	0.06	0.02	0.28	0.70	0.15
rs17743054	18	42,900,892	0.28	2.40E-06	0.04	0.25	0.01	0.01
rs13039563	20	52,296,849	0.24	4.88E-03	0.14	0.26	0.81	0.44
rs9808759	21	47,780,223	0.07	0.71	0.50	0.29	0.44	0.29



1 Per-minor allele odds ratio (95% confidence limits)

2. Luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); HER2-enriched-like (ER- and PR-, HER2+); triple-negative (ER-, PR-, HER2-)

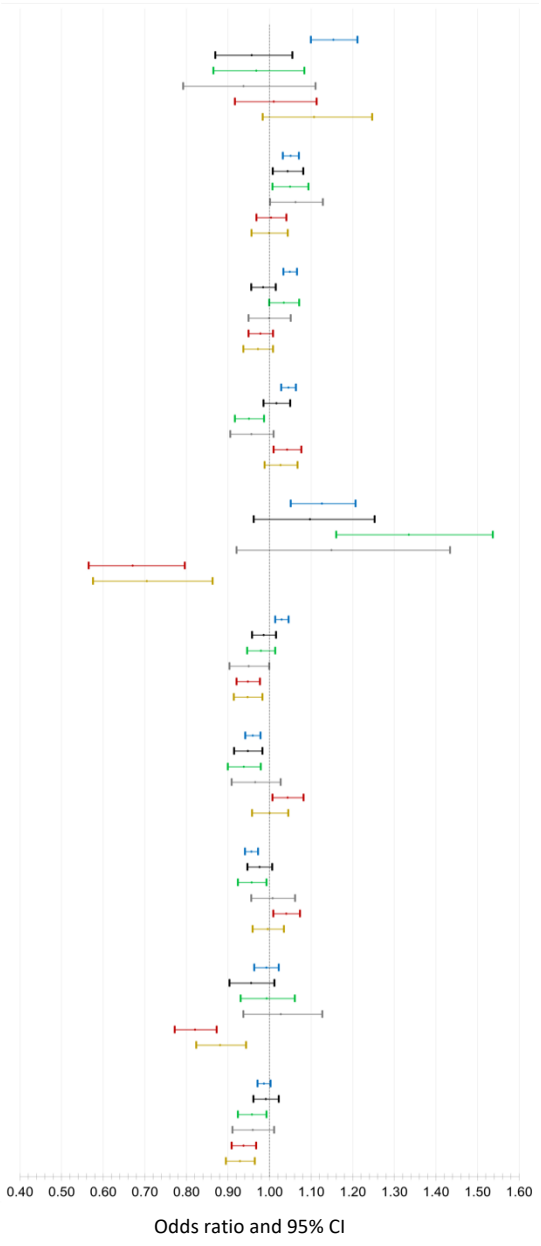
3. Based on a mixed-effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER was entered into the model as a fixed-effect term and PR, HER2, and grade were entered into the model as random-effect terms.

4. Results from second stage case-case parameters from a fixed effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER, PR, HER2, and grade are mutually adjusted for each other

5. Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)

Supplementary Figure 8 Risk[1] of breast cancer subtypes defined by intrinsic-like subtypes[2] (n = 106,278 invasive cases, n = 91,477 controls) among loci identified using the two-stage polytomous logistic regression model and the CIMBA / BCAC triple-negative meta-analysis.

Variant	chromosome	Position	MAF	Global etiologic heterogeneity *P*[3]	Tumor characteristic heterogeneity *P*[4]			
					ER[5]	PR[5]	HER2[5]	grade
chr1:145126177	1	145,126,177	0.04	2.8E-06	5.04E-01	5.35E-02	1.18E-01	6.87E-04
rs495367	4	1,986,972	0.35	5.8E-02	1.35E-01	9.22E-01	3.09E-01	2.42E-01
chr5:67424121	5	67,424,121	0.45	5.2E-07	1.70E-01	1.31E-01	4.20E-01	2.79E-03
rs7924772	11	120,233,626	0.39	1.4E-03	6.69E-01	8.31E-01	1.41E-06	9.95E-02
rs78378222	17	7,571,752	0.01	9.1E-08	7.01E-06	8.96E-01	2.67E-04	5.15E-01
rs206435	18	10,354,649	0.5	1.1E-09	2.79E-03	2.51E-01	1.44E-01	2.83E-04
rs141526427	20	11,502,618	0.25	6.2E-05	1.26E-03	4.44E-01	8.88E-02	3.22E-01
rs6065254	20	39,248,265	0.39	7.3E-07	4.34E-03	1.98E-01	3.92E-01	2.74E-01
rs17215231	6	33,239,869	0.08	2.4E-06	4.40E-02	2.46E-01	9.12E-03	7.87E-03
rs2464195	12	121,435,475	0.37	1.0E-02	8.92E-02	9.99E-01	4.05E-01	3.41E-01



1 Per-minor allele odds ratio (95% confidence limits)

2. Luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); luminal B-like (ER+ and/or PR+, HER2+); HER2-enriched-like (ER- and PR-, HER2+); triple-negative (ER-, PR-, HER2-)

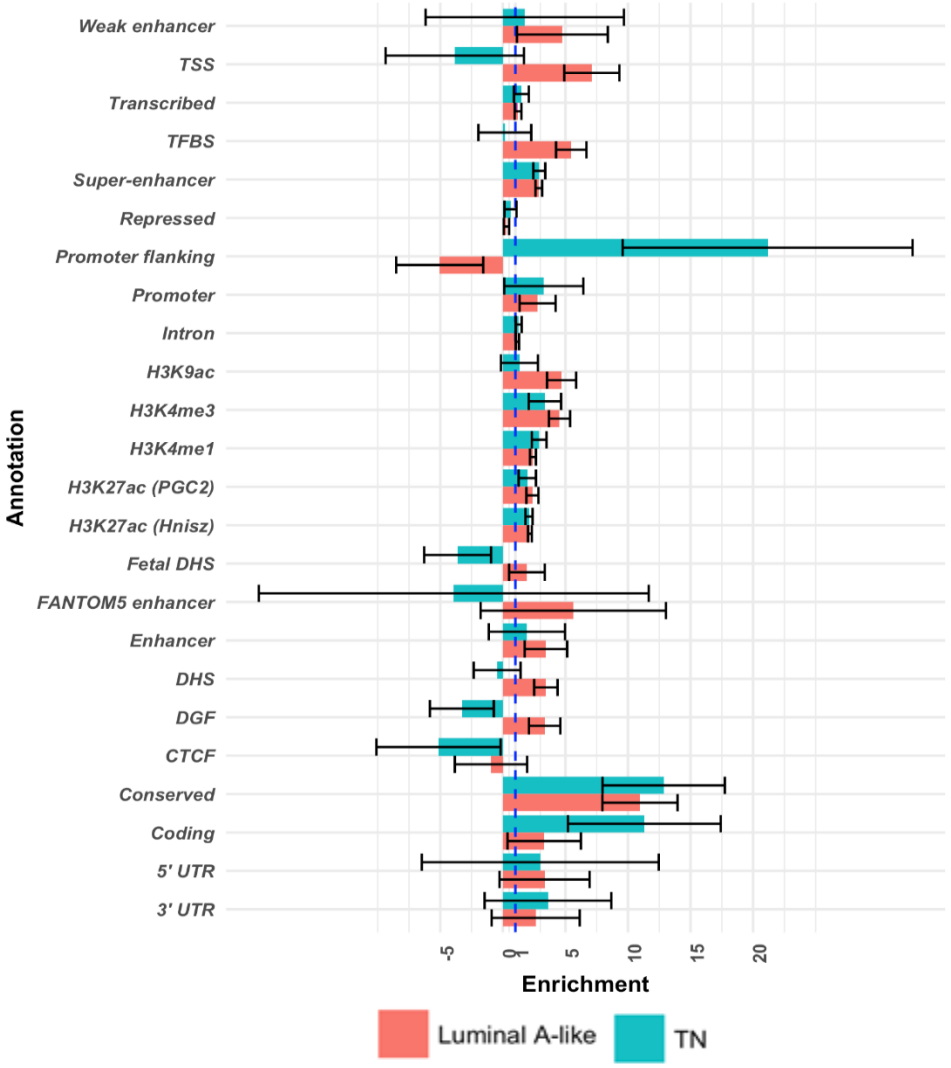
3. Based on a mixed-effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER was entered into the model as a fixed-effect term and PR, HER2, and grade were entered into the model as random-effect terms.

4. Results from second stage case-case parameters from a fixed effect two-stage polytomous model testing for heterogeneity between susceptibility variants and ER, PR, HER2, and grade, where ER, PR, HER2, and grade are mutually adjusted for each other

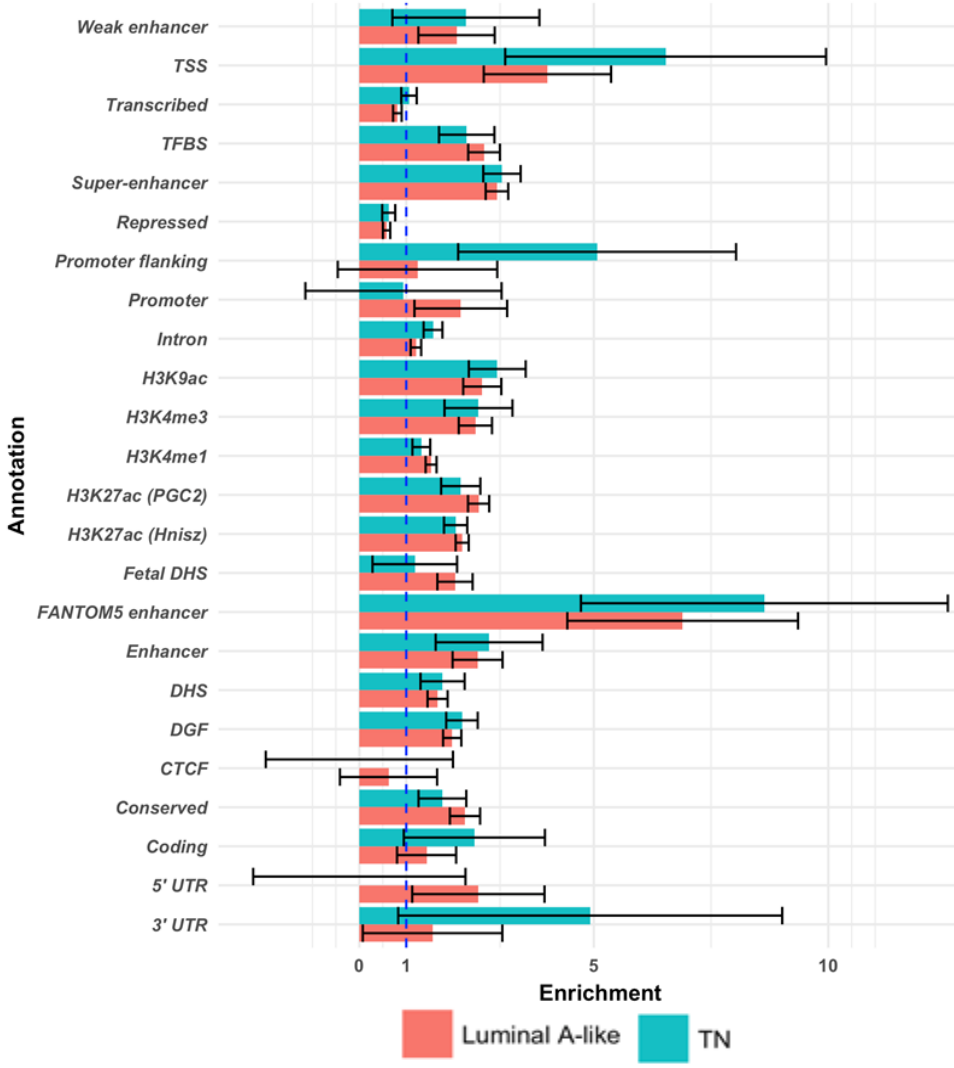
5. Estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2)

Supplementary figure 9. **a)** Enrichment analysis[1] results for 24 non-cell-type-specific, publicly available annotations for luminal A-like subtypes and triple-negative subtypes (n = 45,253 effective luminal A-like cases, n = 8,602 effective triple-negative cases, n = 91,477 controls). **b)** Enrichment analysis[1] results for 24 main annotations with ±500 bp extension for luminal A-like subtypes and triple-negative subtypes. No significant differences were found between luminal A-like and triple-negative after adjusting for multiple testing.

a)



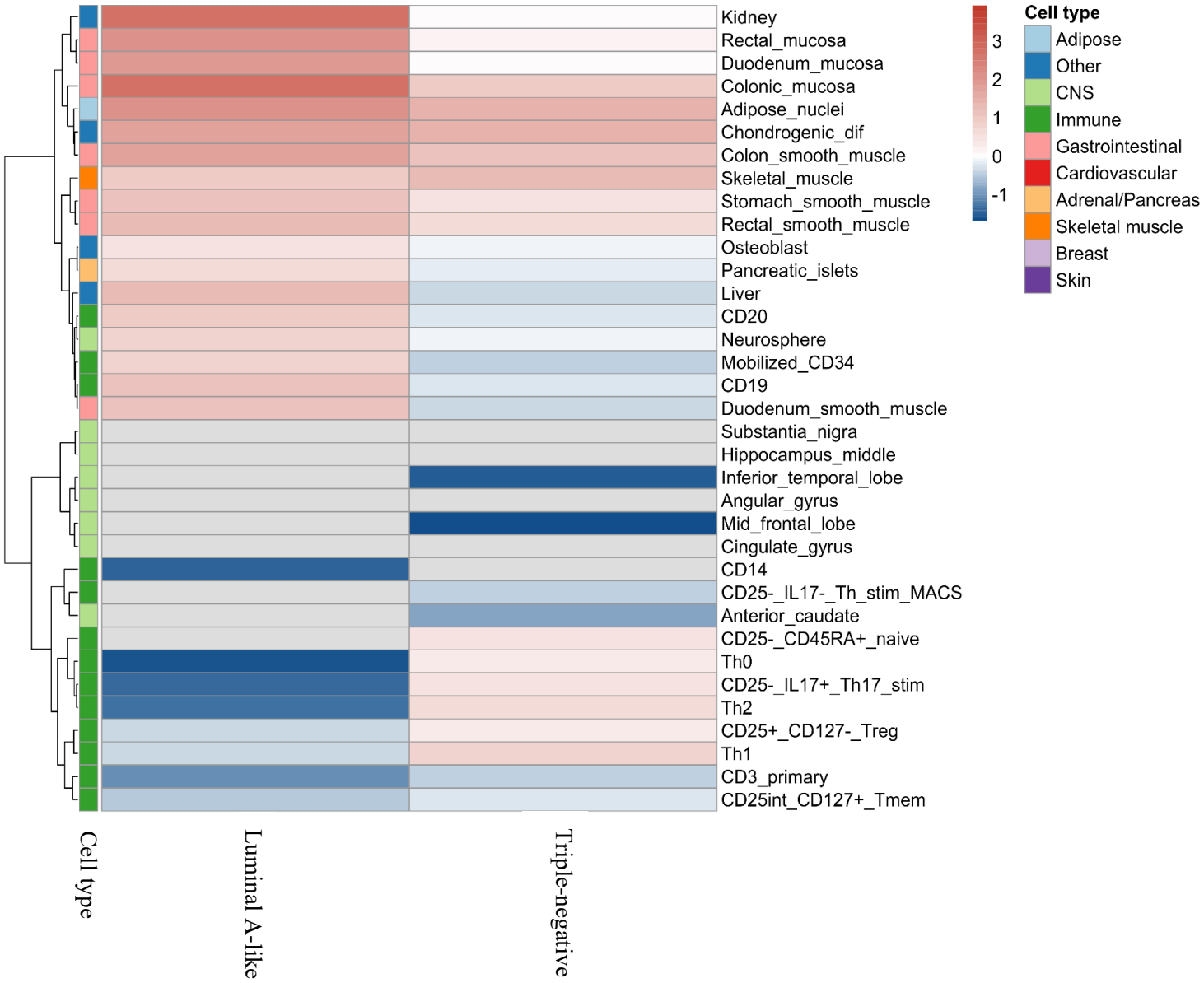
b)



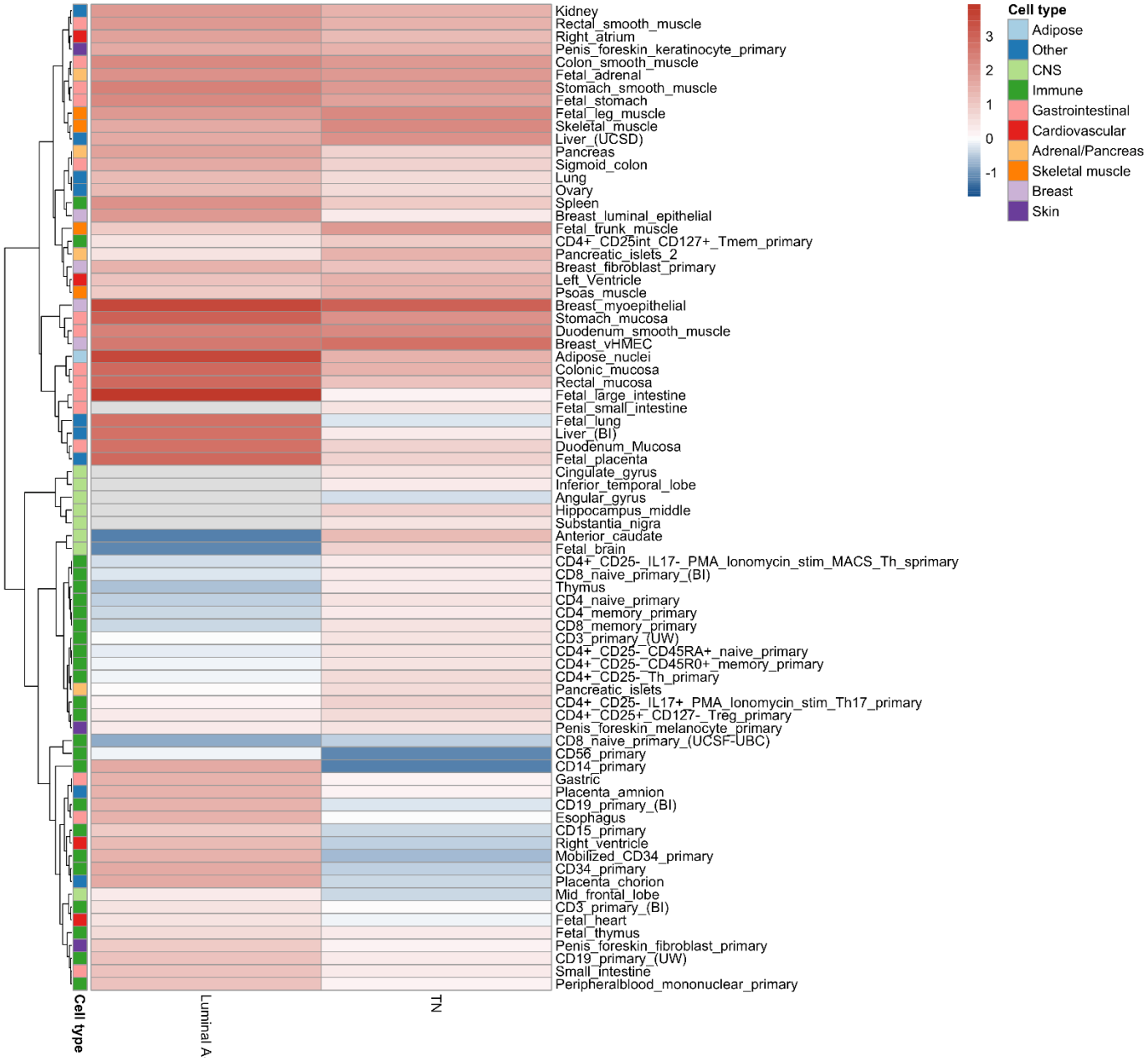
[1] Error bars represent Jackknife standard errors around the estimates of enrichment.

Supplementary figure 10. Enrichment analysis results for 220 cell-type-specific annotations of four histone marks - H3K4me1, H3K4me3, H3K9ac and H3K27ac – in the luminal A-like and triple-negative subtypes. Both luminal A-like and triple-negative subtypes were enriched for gastrointestinal cell types and suppression of central nervous system cells.

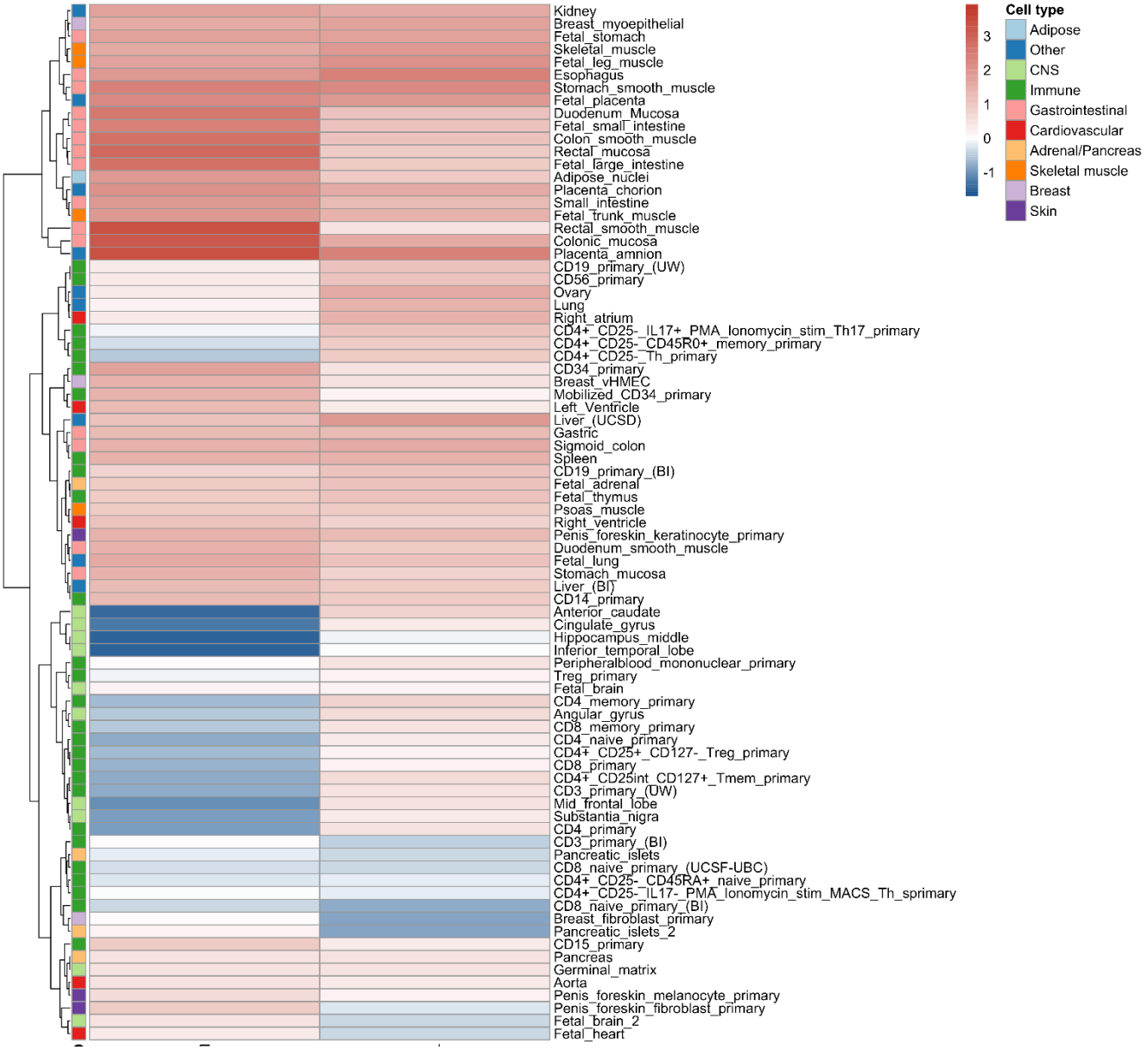
a) Heatmap showing patterns of cell-type specific enrichment for histone marks H3K27ac in luminal A-like tumors and TN tumors



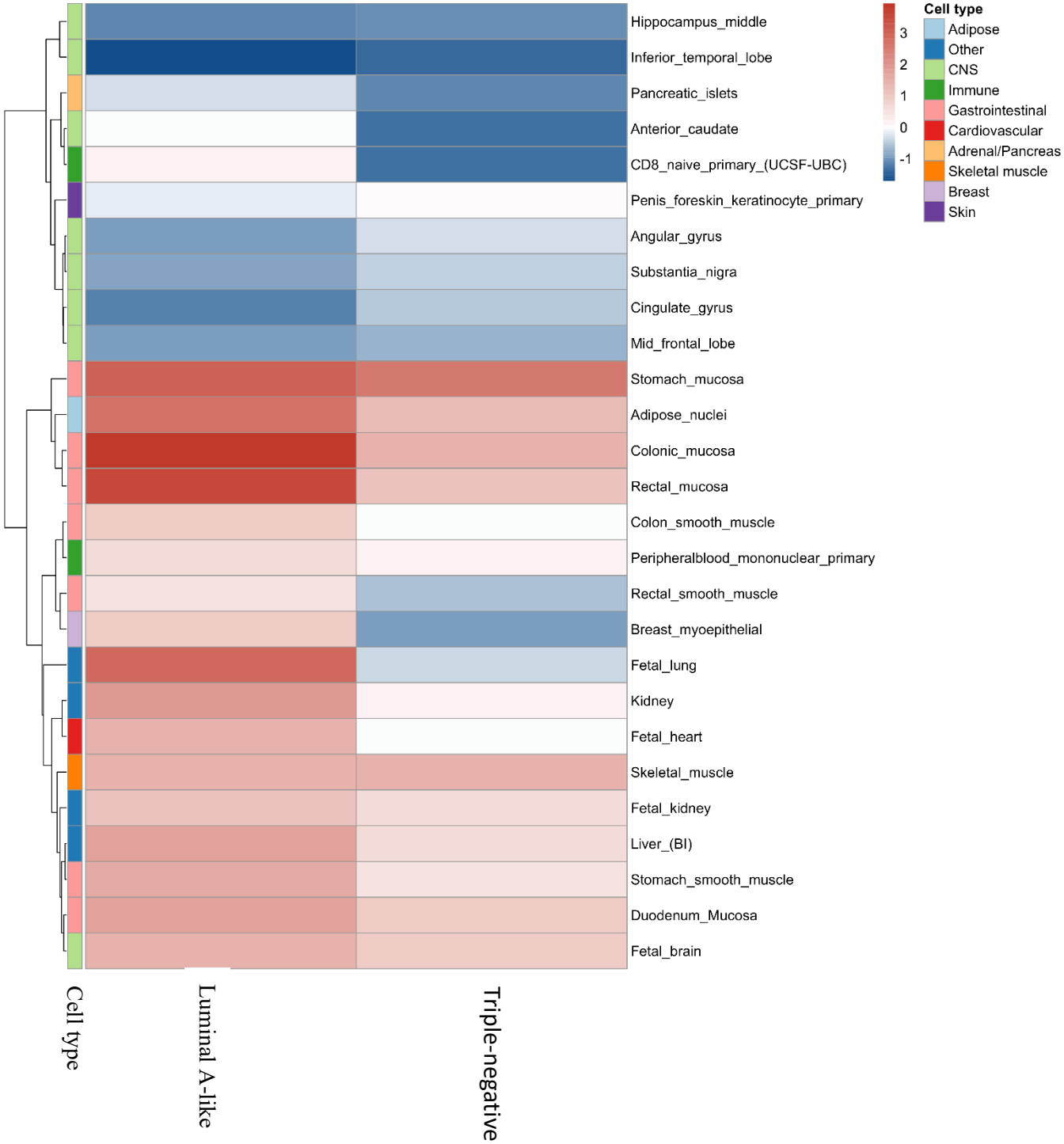
b) Heatmap showing patterns of cell-type specific enrichment for histone marks H3K4me1 in luminal A-like tumors and triple-negative tumors



c) Heatmap showing patterns of cell-type specific enrichment for histone marks H3K4me3 in luminal A-like tumors and triple-negative tumors



d) Heatmap showing patterns of cell-type specific enrichment for histone marks H3K9ac in luminal A-like tumors and triple-negative tumors



Supplementary Note

eQTL Analysis

Data from breast cancer tumors and adjacent normal breast tissue were accessed from The Cancer Genome Atlas (TCGA)[1]. Germline variant genotypes (Affymetrix 6.0 arrays) were processed and imputed to the 1000 Genomes reference panel (October 2014) and European ancestry ascertained as previously described[2]. Tumor tissue copy number was estimated from the Affymetrix 6.0 and called using the GISTIC2 algorithm[3]. Complete genotype, RNA-seq and copy number data were available for 679 genetically European patients (78 with adjacent normal tissue). Further, RNA-seq for normal breast tissue and imputed germline genotype data were available from 80 females from the GTEx Consortium[4]. Genes with a median expression level of 0 RPKM across samples were removed, and RPKM values of each gene were log2 transformed. Expression values of samples were quantile normalized. Genetic variants were evaluated for association with the expression of genes located within ±2Mb of the lead variant at each risk region using linear regression models, adjusting for ESR1 expression. Tumor tissue was also adjusted for copy number variation, as previously described[5]. eQTL analyses were performed using the MatrixEQTL program in R[6].

INQUISIT target gene analysis

Logic underlying INQUISIT predictions: Details of the INQUISIT pipeline have been previously described[1]. Briefly, genes were evaluated as potential targets of candidate causal variants through effects on: (1) distal gene regulation, (2) proximal regulation, or (3) a gene's coding sequence. We intersected CCV positions with multiple sources of genomic information, chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)[7] in MCF7 cells, and genome-wide chromosome conformation capture (Hi-C) in HMECs[8]. We used breast cell line computational enhancer–promoter correlations (PreSTIGE[9], IM-PET[10], FANTOM5[11]) breast cell super-enhancer[12], breast tissue-specific expression variants (eQTL) from multiple independent studies (TCGA (normal

breast and breast tumor) and GTEx breast, **See eQTL Methods**), transcription factor and histone modification chromatin immunoprecipitation followed by sequencing (ChIP-seq) from the ENCODE and Roadmap Epigenomics Projects together with the genomic features found to be significantly enriched for all known breast cancer CCVs[13], gene expression RNA-seq from several breast cancer lines and normal samples (ENCODE) and topologically associated domain (TAD) boundaries from T47D cells (ENCODE[14]). To assess the impact of intragenic variants, we evaluated their potential to alter primary protein coding sequence and splicing using Ensembl Variant Effect Predictor[15] using MaxEntScan and dbscSNV modules for splicing alterations based on “ada” and “rf” scores. Nonsense and missense changes were assessed with the REVEL ensemble algorithm, with CCVs displaying REVEL scores > 0.5 deemed deleterious.

Scoring hierarchy: Each target gene prediction category (distal, promoter or coding) was scored according to different criteria. Genes predicted to be distally-regulated targets of CCVs were awarded two points based on physical links (for example ChIA-PET), and one point for computational prediction methods, or eQTL associations. All CCVs were considered as potentially involved in distal regulation and all CCVs (including coding variants) were scored in this category. Intersection of a putative distal enhancer with genomic features found to be significantly enriched[20] were further upweighted with an additional point. In the case of multiple, independent interactions, an additional point was awarded. CCVs in gene proximal regulatory regions were intersected with histone ChIP-Seq peaks characteristic of promoters and assigned to the overlapping transcription start sites (defined as -1.0 kb - +0.1 kb). Further points were awarded to such genes if there was evidence for an eQTL association, while a lack of expression resulted in down-weighting as potential targets. Potential coding changes including missense, nonsense and predicted splicing alterations resulted in addition of one point to the encoded gene for each type of change, while lack of expression reduced the score. We added an additional point for predicted target genes that were also breast cancer drivers (278 genes[1,20]). For each category, scores potentially ranged from 0-8 (distal); 0-4 (promoter) or 0-3 (coding). We converted these scores into 'confidence levels': Level 1 (highest confidence)

when distal score >4, promoter score ≥3 or coding score >1; Level 2 when distal score ≤4 and ≥1, promoter score=1 or=2, coding score=1; and Level 3 when distal score <1 and >0, promoter score <1 and >0, and coding <1 and >0. For genes with multiple scores (for example, predicted as targets from multiple independent risk signals or predicted to be impacted in several categories), we recorded the highest score.

Global genomic enrichment analyses

We performed stratified LD score regression analyses[16-18] as previously described[2] for two major intrinsic-like subtypes, luminal A-like and triple-negative, using the summary statistics from the meta-analyses of OncoArray, iCOGs, and CIMBA. The analysis included all variants in the 1000 Genome Project Phase 1v3 release with MAF>1% and imputation quality score R2>0.3 in the OncoArray data. We restricted analysis to all variants present on the HapMap version 3 dataset. We first fit a model that included 24 non-cell-type-specific, publicly available annotations as well as 24 additional annotations that included a 500-bp window around each of the 24 main annotations. We also included 100-bp windows around ChIP–seq peaks and one annotation containing all variants, leading to a total of 53 overlapping annotations. In addition to the baseline model using 24 main annotations, we also performed cell-type-specific analyses using annotations of the four histone marks (H3K4me1, H3K4me3, H3K9ac and H3K27ac). Each cell-type-specific annotation corresponds to a histone mark in a single cell type (for example, H3K27ac in adipose nuclei tissues)[16]. There was a total of 220 such annotations. We further subdivided these 220 cell-type-specific annotations into 10 categories by aggregating the cell-type-specific annotations within each group (for example, variants related with any of the four histone modifications in any hematopoietic and immune cells were considered as one category). To estimate the enrichment of each marker, we ran 220 LD score regressions after adding each different histone mark to the baseline model. We used a Wald test to evaluate the differences in the functional enrichment between the luminal A-like and triple-negative subtypes, using the regression coefficients and standard error based on the models above. After

Bonferroni correction none of the differences were significant. Notably, the Wald test assumes that the enrichment estimates of luminal A-like and triple-negative subtypes were independent, but this assumption was violated by the sharing of controls between the subtypes. Under this scenario, our Wald test statistics were less conservative than had we adjusted for the correlation between estimates. However, given the lack of significant differences observed between luminal A-like and triple-negative subtypes we had no concern about a type one error.

Two-stage polytomous model

The two-stage polytomous logistic regression model allows us to efficiently test for genetic associations while accounting for tumor marker correlations and large amounts of missing tumor data [19]. We used this method to detect breast cancer susceptibility variants while taking account of four tumor characteristics: estrogen receptor (ER; ER-positive vs ER-negative), progesterone receptor (PR; PR-positive vs PR-negative), human epidermal growth factor receptor 2 (HER2; HER2-positive vs HER2-negative), and grade (defined as grade 1, grade 2, and grade 3). Below we describe in greater detail how we applied this method

In our study, we investigated for underlying heterogenous associations according to ER, PR, HER2, and grade; however, we will first start the discussion of fitting a two-stage polytomous model by first focusing on ER, PR, and HER2, and then discuss including grade in the model. The cross combination of ER, PR, and HER2 results in eight distinct breast cancer subtypes (8 = 2x2x2). Let N denote the total sample size and let $\mathrm{D_i}$ denote the disease status of ith subject which can take values from $\{0,1,2,\dots,8\}$ and $\mathrm{i} = 1,2,\dots,\mathrm{N}$. $\mathrm{D_i} = 0$ represent a control, and $\mathrm{D_i} = m$ represent the ith subject with the breast cancer subtypes M. Let $\mathrm{G_i}$ denote the genotype of a variant for ith subject, taking values from $\{0,1,2\}$. Let $\mathbf{X_i}$ denote the other covariates for the ith subject, for example principal components or age. In the first stage of the model, we fit a standard “saturated” polytomous logistic regression model:

$$\Pr(\mathrm{D_i} = \mathrm{m}|\mathrm{G_i}, \mathrm{X_i}) = \frac{\exp(\beta_m G_i + \boldsymbol{\eta_m^T X_i})}{1 + \sum_{m=1}^{8} \exp(\beta_m G_i + \boldsymbol{\eta_m^T X_i})}, \quad (1)$$

where β_{m} is the regression coefficient for a variant (G) associated with the mth subtype and η_m is the vector of regression coefficients for the other covariate (X) associated with mth subtype.

Each cancer subtype m is defined through a unique combination of ER, PR, and HER2; therefore, we can alternatively index the parameters β_{m} as $\beta_{s_1 s_2 s_3}$, where $s_1, s_2, s_3 \in \{0, 1\}$ for the three binary tumor characteristics. Originally, β_1 represented the regression coefficient of the ER-, PR-, HER2- subtype. With this indexing, β_1 can be alternatively written as β_{000} and, thus with this reparameterization we can represent the log odds ratio of the eight subtypes as:

$$\beta_{s_1 s_2 s_3} = \theta^{(0)} + \theta_1^{(1)} s_1 + \theta_2^{(1)} s_2 + \theta_3^{(1)} s_3 + \theta_{12}^{(2)} s_1 s_2 + \theta_{13}^{(2)} s_1 s_3 + \theta_{23}^{(2)} s_1 s_3 + \theta_{123}^{(3)} s_1 s_2 s_3, \quad (2)$$

where $\theta_0^{(0)}$ represents the case-control log odds ratio for a reference subtypes versus the controls. We have chosen ER-, PR-, HER2- as the reference subtype, but any subtype can be chosen as the reference subtype. $\theta_{\mathrm{k}}^{(1)}$ represents the case-case log odds ratio for the kth tumor characteristic after adjusting for the other tumor characteristics. We also refer $\theta_{\mathrm{k}}^{(1)}$ as the main effect of the kth tumor characteristic. $\theta_{\mathrm{k_1 k_2}}^{(2)}$ represents how the case-case log odds ratio associated with k_1th tumor characteristic is modified by levels of the k_2th tumor characteristic and vice versa. We also refer to $\theta_{\mathrm{k_1 k_2}}^{(2)}$ as the pairwise interaction between the k_1th tumor characteristic and the k_2th tumor characteristic. $\theta_{123}^{(3)}$ represents the third order interaction of the three tumor characteristics. This decomposition is equivalent to the first stage polytomous logistic regression since both the first stage and second stage have eight parameters. We can specify different two stage models by assuming different second stage parameters to be equal to 0. For example, the baseline two-stage model is represented by:

$$\beta_{s_1 s_2 s_3} = \theta^{(0)}. \quad (3)$$

This baseline model assumes all of the subtypes have the same log odds ratio and is equivalent to a standard case-control logistic regression testing the association between an exposure and breast cancer, irrespective of tumor subtypes. We can also constrain all of the second stage pairwise interactions and higher order interactions to be 0:

$$\beta_{s_1 s_2 s_3} = \theta^{(0)} + \theta_1^{(1)} s_1 + \theta_2^{(1)} s_2 + \theta_3^{(1)} s_3. \quad (4)$$

This additive two-stage model assumes the case-case log odds ratio of a tumor characteristic are not affected by interactions with the other tumor characteristics.

By adding the second stage pairwise interactions parameters into the model, we can also construct the pairwise interaction two-stage polytomous model:

$$\beta_{s_1 s_2 s_3} = \theta^{(0)} + \theta_1^{(1)} s_1 + \theta_2^{(1)} s_2 + \theta_3^{(1)} s_3 + \theta_{12}^{(2)} s_1 s_2 + \theta_{13}^{(2)} s_1 s_3 + \theta_{23}^{(2)} s_1 s_3. \quad (5)$$

This model evaluates how two tumor characteristics are modified by each other. For example, $\theta_{12}^{(2)}$ measures how the case-case log odds ratio associated of ER is modified by the status of PR and vice versa. If we further add the three-way interaction term between ER, PR, and HER2, then this model becomes saturated (as shown in in Equation 2) and is equivalent to the polytomous logistic regression.

When we add the three-level ordinal variable tumor grade into the model, we can define 24 (2x2x2x3) breast cancer subtypes. We can apply the same decomposition as implemented with three tumor characteristics to provide the following additive two-stage model:

$$\beta_{s_1 s_2 s_3 s_4} = \theta^{(0)} + \theta_1^{(1)} s_1 + \theta_2^{(1)} s_2 + \theta_3^{(1)} s_3 + \theta_4^{(1)} s_4, \quad (6)$$

where $\theta_4^{(1)}$ is the main effect of grade and s_4 can take the values from $\{1, 2, 3\}$. In this model, we assume the grade main effect linearly changes, meaning the average log odds ratios difference between grade 3 versus grade2 is the same the as the difference between grade 2 versus grade1. We can always describe the link between the first stage parameters and second stage parameters in Equation (6) in matrix form:

$$\begin{matrix}\mathrm{ER}-\mathrm{PR}-\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}-\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}+\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}-\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}-\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}+\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}+\mathrm{grade1}\\ \ldots\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}+\mathrm{grade3}\end{matrix}\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\\ \beta_6\\ \beta_7\\ \beta_8\\ \ldots\\ \beta_{24}\end{bmatrix}=\begin{bmatrix}1&0&0&0&1\\ 1&1&0&0&1\\ 1&0&1&0&1\\ 1&1&1&0&1\\ 1&0&0&1&1\\ 1&1&0&1&1\\ 1&0&1&1&1\\ 1&1&1&1&1\\ \ldots&\ldots&\ldots&\ldots&\ldots\\ 1&1&1&1&3\end{bmatrix}\begin{bmatrix}\theta^{(0)}\\ \theta_1^{(1)}\\ \theta_2^{(1)}\\ \theta_3^{(1)}\\ \theta_4^{(1)}\end{bmatrix}=\mathbf{Z}\begin{bmatrix}\theta^{(0)}\\ \boldsymbol{\theta^H}\end{bmatrix}=\mathbf{Z}\boldsymbol{\theta}, \tag{7}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients of the first stage parameters, $\boldsymbol{\theta}$ is the vector of all the second stage parameters, and $\boldsymbol{\theta^H}$ is a vector of second stage main effects.

Hypothesis testing of two-stage polytomous logistic regression

Under the two-stage model framework, there are three different tests we can construct. The first is the global association test:

$$H_0\colon \theta^{(0)}=0 \; and \; \boldsymbol{\theta^H}=\mathbf{0} \; versus \; H_1\colon either \; \theta^{(0)}\neq 0 \text{ or } \boldsymbol{\theta^H}\neq \mathbf{0}\,. \tag{8}$$

This test is designed to test whether a variant is associated with any of the 24 breast cancer subtypes. If the null hypothesis is rejected under this setting, then at least one of the first stage subtype case-control log odds ratios β_m is significantly not equal to 0. The second test is the global heterogeneity test:

$$H_0\colon \boldsymbol{\theta^H}=\mathbf{0} \; versus \; H_1\colon \boldsymbol{\theta^H}\neq \mathbf{0}\,. \tag{8}$$

This test is designed to test whether the associations between a variant and any two breast cancer subtypes are significantly different from each other. If the null hypothesis is rejected under this setting, then we can conclude that at least two of the first stage subtypes case-control log odds ratios are significantly different with each other ($\beta_{m_1}\neq\beta_{m_2}$).

If the global heterogeneity test is significant, then we can construct the third hypothesis tests, the specific tumor marker heterogeneity test:

$$H_0: \boldsymbol{\theta}^{\boldsymbol{H}}_{(k)} = 0 \; versus \; H_1: \boldsymbol{\theta}^{\boldsymbol{H}}_{(k)} \neq \; 0 \, . \tag{9}$$

This test is designed to test which tumor character is the source of the observed heterogeneity in the global heterogeneity test. Under the additive two-stage model in Equation (6), for example, we can test $H_0: \theta_1^{(1)} = 0 \; versus \; H_0: \theta_1^{(1)} \neq \; 0$. This is designed to test whether the case-case log odds ratio of ER is significant not equaling to 0 after adjusting for the effects of PR, HER2 and grade.

Mixed effect two-stage polytomous model

Although the additive two-stage model decreases the degrees of freedoms compared to the first stage polytomous logistic regression, the degrees of freedom of the two-stage model are still penalized when additional tumor characteristics are included into the model. To address this issue, we developed the mixed effect two-stage polytomous model to enter tumor characteristic variables into the model as either fixed- or random-effect terms. In this model, we keep the second stage main effect of ER ($\theta_1^{(1)}$) as a fixed effect since there is strong *a priori* evidence that ER is a common source of heterogeneity [20]. On the other hand, as there is minimal evidence suggesting that tumor characteristics such as PR, HER2, and grade are sources of heterogeneity, we assume the case-case parameter of PR ($\theta_2^{(1)}$), HER2 ($\theta_3^{(1)}$) and grade ($\theta_4^{(1)}$) as random effects. These random parameters have an assumed arbitrary distribution with mean 0 and variance σ^2. We always keep the baseline effect $\theta^{(0)}$ as fixed since it captures the overall association between a variant and breast cancer. Under the mixed effect two stage model, the global test for association is:

$$H_0: \theta^{(0)} = 0, \theta_1^{(1)} = 0, \sigma^2 = 0 \; versus \; H_1: either \; \theta^{(0)}, \theta_1^{(1)}, or \; \sigma^2 \neq 0 \tag{10}$$

The rejection of the null hypothesis implies that the variant is significantly associated with at least one of the 24 breast cancer subtypes. The global heterogeneity test under the mixed effect two-stage model would be:

$$H_0: \theta_1^{(1)} = 0 \; and \; \sigma^2 = 0 \; versus \; H_1: either \; \theta_1^{(1)} \text{or} \; \sigma^2 \neq 0 \, . \qquad (11)$$

The rejection of the null hypothesis would imply that the variant's associations between at least two breast cancer subtypes are significantly different. However, the specific tumor marker heterogeneity test for a specific tumor marker is not applied in the mixed effect two-stage model because it requires the estimate of case-case log odds ratio of PR, HER2 and grade which are note estimated when modeled as random effects.

Two-stage model for intrinsic subtypes of breast cancer

In previous sections, we showed how the first stage case control log odds ratios of breast cancer subtypes are decomposed to the case control log odds ratio of a reference subtype and the into case-case parameters of tumor characteristics. Using the hierarchical second stage decomposition, the two-stage model can also estimate the case control log odds ratio of specific breast cancer subtypes of interest. In our study we defined five intrinsic-like breast cancer subtypes based on tumor status of ER, PR, HER2 and grade: (1) luminal A-like (ER+ and/or PR+, HER2-, grade 1 & 2); (2) luminal B/HER2-negative-like (ER+ and/or PR+, HER2-, grade 3); (3) luminal B-like (ER+ and/or PR+, HER2+); (4) HER2-enriched-like (ER- and PR-, HER2+), and (5) triple-negative (ER-, PR-, HER2-). To estimate the case-control log odds ratios of these five intrinsic subtypes we can construct the two-stage model as:

$$\begin{matrix}\mathrm{ER}-\mathrm{PR}-\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}-\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}+\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}-\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}-\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}-\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}-\mathrm{PR}+\mathrm{HER2}+\mathrm{grade1}\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}+\mathrm{grade1}\\ \dots\\ \mathrm{ER}+\mathrm{PR}+\mathrm{HER2}+\mathrm{grade3}\end{matrix}\quad \boldsymbol{\beta}=\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\\ \beta_6\\ \beta_7\\ \beta_8\\ \dots\\ \beta_{24}\end{bmatrix}=\begin{bmatrix}0&0&0&0&1\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 0&0&0&1&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ \dots&\dots&\dots&\dots&\dots\\ 0&1&0&0&0\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\\ \theta_3\\ \theta_4\\ \theta_5\end{bmatrix}\quad \begin{matrix}\mathrm{Luminal\ A}-\mathrm{like,\ low\ grade}\\ \mathrm{Luminal\ B}-\mathrm{like}\\ \mathrm{Luminal\ B/HER2}-\mathrm{negative}-\mathrm{like}(12)\\ \mathrm{HER2\ enriched}-\mathrm{like}\\ \mathrm{Triple}-\mathrm{negative}\end{matrix}$$

Under this model, the second stage parameters provide estimates of case-control log odds ratios for the five tumor subtypes. This model is similar to directly fitting a polytomous logistic regression. However, we have incorporated into the two-stage model an efficient missing data algorithm that allows to take advantage of subjects with incomplete tumor characteristic data. The missing data algorithm has been described in detail elsewhere [1].

Modified LD score regression

Since the two-stage polytomous logistic regression implements an EM algorithm to account for missing tumor characteristics data, the effective sample size is not equivalent to the sample size of cases with complete tumor characteristic data. In this case the sample size is not available, but the log odds ratio for each variant $\hat{\beta}_j$ and the standard error s_j are given.

Under a case-control study, we consider the logistic regression model

$$log\left(\frac{p}{1-p}\right)=\alpha+\left(\boldsymbol{\beta}^{(J)}\right)^{\boldsymbol{T}}\boldsymbol{X},$$

where $\boldsymbol{\beta}^{(J)}=(\beta_1^{(J)},\beta_2^{(J)},\dots,\beta_M^{(J)})$ are the joint effect sizes. We define the heritability as $h^2=var(\left(\boldsymbol{\beta}^{(J)}\right)^T\boldsymbol{X})$, assuming X is standardized with mean 0 variance 1. If X is in the original 0, 1, 2 scale, we multiply the $\hat{\beta}_j$ and s_j by $\sqrt{2p_j(1-p_j)}$ to standardize, where p_j is the minor allele frequency for the jth variant. Therefore, the expected chi-square statistics (z_j^2) of variant j is

$$\mathrm{E}\left(z_j^2|l_j\right) = \frac{E\left(\hat{\beta}_j^2|l_j\right)}{s_j^2} = \frac{\left[E\left\{\left(\hat{\beta}_j - \beta_j\right)^2\middle|l_j\right\} + 2E\left[\left(\hat{\beta}_j - \beta_j\right)\beta_j|l_j\right] + E\left(\beta_j^2|l_j\right)\right]}{s_j^2} \tag{13}$$

$$= \frac{\left[E\left\{\left(\hat{\beta}_j - \beta_j\right)^2\middle|l_j\right\} + E\left(\beta_j^2|l_j\right)\right]}{s_j^2}$$

$$= 1 + \frac{E\{(\sum_k r_{jk}\beta_k^{(J)})^2\}}{s_j^2}$$

$$= 1 + \frac{h^2}{M}\frac{l_j}{s_j^2},$$

where $l_j = \sum_k r_{jk}^2$ is the LD score of the variant j and $1/s_j^2$ is the effective sample size for variant j. The modified LD score regression formula is:

$$\mathrm{E}\left(z_j^2|l_j\right) = 1 + \frac{h^2}{M}\frac{l_j}{s_j^2}. \tag{14}$$

To estimate the genetic correlation between two traits, the expected value of $z_{1j}z_{2j}$ for a variant j is

$$\mathrm{E}\left(\mathrm{z_{1j}}z_{2j}|l_j\right) = \frac{E\left(\hat{\beta}_{1j}\,\hat{\beta}_{2j}|l_j\right)}{s_{1j}s_{2j}} \tag{15}$$

$$= \frac{\left[E\left\{\left(\hat{\beta}_{1j} - \beta_{1j}\right)\left(\hat{\beta}_{2j} - \beta_{2j}\right)|l_j\right\} + E\left(\beta_{1j}\beta_{2j}|l_j\right)\right]}{s_{1j}s_{2j}}$$

$$= \frac{s_{12j}}{s_{1j}s_{2j}} + \frac{E\left(\sum_k r_{jk}\beta_{1k}^{(J)}\sum_k r_{jk}\beta_{2k}^{(J)}\,|l_j\right)}{s_{1j}s_{2j}}$$

$$= \frac{s_{12j}}{s_{1j}s_{2j}} + \frac{\rho_g}{M}\frac{l_j}{s_{1j}s_{2j}},$$

where ρ_g is the genetic covariance between the two different traits. Under this case, $1/s_{1j}^2$ and $1/s_{2j}^2$ are the effective sample size for variant j for the two traits respectively. The modified LD score regression for genetic covariance is

$$\mathrm{E}\left(\mathrm{z_{1j}}z_{2j}|l_j\right) = \frac{s_{12j}}{s_{1j}s_{2j}} + \frac{\rho_g}{M}\frac{l_j}{s_{1j}s_{2j}}. \tag{16}$$

The genetic correlation is given by $\frac{\rho_g}{\sqrt{h_1^2 h_2^2}}$.

Effective sample size of cases of two-stage polytomous model

The two-stage polytomous model implements the EM algorithm to impute missing tumor characteristics; therefore, the effective sample size of cases is not equivalent to the actual number of cases with available tumor characteristic data. We estimated the effective sample sizes to help demonstrate the benefit of using the EM algorithm to impute missing tumor characteristics and to aid comparability with previous studies (**Supplementary Table 4**). To estimate the effective sample size, suppose we have a complete dataset with no missing tumor characteristics, the sample size is n_k for the kth subtype and n_0 for the control. If we fit a two-stage polytomous model for the jth variant, the corresponding log odds ratio for kth subtype is $\hat{\beta}_{jk}$ and the standard error is s_{jk}. Then, approximately:

$$var\left(\hat{\beta}_{jk}|p_j\right) \approx \frac{n_0 + n_k}{2 * p_j\left(1 - p_j\right)(n_0 n_k)},$$

where p_j is the MAF of the jth variant. Now we consider fitting a two-stage polytomous model with missing tumor characteristics. Given the standard error s_{jk} of the log odds ratio and the control sample size, we have the estimate of effective number of cases as,

$$\hat{n}_k = \left(\frac{1}{n_0} - 2s_{jk}^2 p_j(1 - p_j)\right)^{-1}.$$

We used the median estimates of effective sample size of cases for all variants as the final estimate.

References Supplementary Note Methods

1. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
2. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92-94 (2017).
3. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
4. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
5. Li, Q. *et al.* Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* **152**, 633-41 (2013).
6. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353-8 (2012).
7. Fullwood, M.J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64 (2009).
8. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
9. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**, 1-13 (2014).
10. He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* **111**, E2191-9 (2014).
11. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).
12. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
13. Fachal, L. *et al.* Fine-mapping of 150 breast cancer risk regions identifies 178 high confidence target genes. *Nat Genet* **52**, 56-73 (2020).
14. Dixon, J.R. *et al.* Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet* **50**, 1388-1398 (2018).
15. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
16. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
17. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
18. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
19. Zhang, H. *et al.* A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. *bioRxiv*, 446039 (2018).
20. Milne, R.L. *et al.* Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* **advance online publication**(2017).

BCAC Funding and Acknowledgments

Funding

BCAC is funded by Cancer Research UK [C1287/A16563, C1287/A10118], the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and by the European Community´s Seventh Framework Programme under grant agreement number 223175 (grant number HEALTH-F2-2009-223175) (COGS). The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The DRIVE Consortium was funded by U19 CA148065.

The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 mention of trade names, commercial products, or

organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363]. The Australian Breast Cancer Tissue Bank (ABCTB) was supported by the National Health and Medical Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The AHS study is supported by the intramural research program of the National Institutes of Health, the National Cancer Institute (grant number Z01-CP010119), and the National Institute of Environmental Health Sciences (grant number Z01-ES049030). The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). The BCEES was funded by the National Health and Medical Research Council, Australia and the Cancer Council Western Australia and acknowledges funding from the National Breast Cancer Foundation (JS). For the BCFR-NY, BCFR-PA, BCFR-UT this work was supported by grant UM1 CA164920 from the National Cancer Institute. 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 mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. For BIGGS, ES is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, United Kingdom. IT is supported by the Oxford Biomedical Research Centre. The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de

Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. The BSUCH study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research. CCGP is supported by funding from the University of Crete. The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The CNIO-BCS was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). COLBCCC is supported by the German Cancer Research Center (DKFZ), Heidelberg, Germany. Diana Torres was in part supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation. The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The CTS was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97-10500) and is currently funded through the National Institutes of Health (R01 CA77398, UM1 CA164917, and U01 CA199277). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. HAC receives support from the Lon V Smith Foundation (LVS39420). The University of Westminster curates the DietCompLyf database funded by

Against Breast Cancer Registered Charity No. 1121258 and the NCRN. The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). FHRISK is funded from NIHR grant PGfAR 0707-10031. The GC-HBOC (German Consortium of Hereditary Breast and Ovarian Cancer) is supported by the German Cancer Aid (grant no 110837 and 113049, coordinator: Rita K. Schmutzler, Cologne). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). The GENICA was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute

of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. Generation Scotland (GENSCOT) received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “STratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z). Funding for identification of cases and contribution to BCAC funded in part by the Wellcome Trust Seed Award “Temporal trends in incidence and mortality of molecular subtypes of breast cancer to inform public health, policy and prevention” Reference 207800/Z/17/Z. The GEPARSIXTO study was conducted by the German Breast Group GmbH. The GESBC was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The HABCS study was supported by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The HEBCS was financially supported by the Helsinki University Hospital Research Fund, the Finnish Cancer Society, and the Sigrid Juselius Foundation..The HMBCS was supported by a grant from the Friends of Hannover Medical School and by the Rudolf Bartling Foundation. The HUBCS was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017), B.M. was supported by grant 17-44-020498, 17-29-06014 of the Russian Foundation for Basic Research, D.P. was supported by grant 18-29-09129 of the Russian Foundation for Basic Research, E.K was supported by the program for support the bioresource collections №007-030164/2, and the study was performed as part of the assignment of the Ministry of Science and Higher Education of the Russian Federation (№AAAA-A16-116020350032-1). Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The KARMA study was

supported by Märit and Hans Rausings Initiative Against Breast Cancer. The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow. LAABC is supported by grants (1RB-0287, 3PB-0102, 5PB-0018, 10PB-0098) from the California Breast Cancer Research Program. Incident breast cancer cases were collected by the USC Cancer Surveillance Program (CSP) which is supported under subcontract by the California Department of Health. The CSP is also part of the National Cancer Institute's Division of Cancer Prevention and Control Surveillance, Epidemiology, and End Results Program, under contract number N01CN25403. LMBC is supported by the 'Stichting tegen Kanker'. DL is supported by the FWO. The MABCS study is funded by the Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", MASA. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. MBCSG is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects “5x1000”). The MCBCS was supported by the NIH grants

CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative. The work of MTLGEBCS was supported by the Quebec Breast Cancer Foundation, the Canadian Institutes of Health Research for the “CIHR Team in Familial Risks of Breast Cancer” program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701. The NBCS has received funding from the K.G. Jebsen Centre for Breast Cancer Research; the Research Council of Norway grant 193387/V50 (to A-L Børresen-Dale and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen), South Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale) and the Norwegian Cancer Society (to A-L Børresen-Dale and V.N. Kristensen). The NBHS was supported by NIH grant R01CA100374. Biological sample preparation was conducted the Survey and Biospecimen Shared Resource, which is supported by P30 CA68485. The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grant UM1 CA164920 from the National Cancer Institute (USA). 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 mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The Carolina Breast Cancer Study was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The NHS was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The NHS2 was supported by NIH grants UM1 CA176726 and U19 CA148065. The OBCS was supported by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. The ORIGO study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The POSH study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. PROCAS is funded from NIHR grant PGfAR 0707-10031. The RBCS was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research

Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES049033). The Two Sister Study (2SISTER) was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-ES102245), and, also by a grant from Susan G. Komen for the Cure, grant FAS0703856. SKKDKFZS is supported by the DKFZ. The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council/Infrastructure gran. The SZBCS and IHCC were supported by Grant PBZ_KBN_122/P05/2004 and the program of the Minister of Science and Higher Education under the name "Regional Initiative of Excellence" in 2019-2022 project number 002/RID/2018/19 amount of financing 12 000 000 PLN. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The UKOPS study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The WHI program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01. The BCINIS was supported by the BCRF (breast cancer research foundation), NY. Nilanjan Chatterjee received supported from the National Human Genome Research Institute (1 R01 HG010480-01).

Acknowledgements

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. The COGS study would not have been possible without the contributions of the following: Rosalind A. Eeles, Ali Amin

Al Olama, Zsofia Kote-Jarai, Lesley McGuffog, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology Laboratory, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility. ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. ABCS thanks the Blood bank Sanquin, The Netherlands. ABCTB Investigators: Christine Clarke, Rosemary Balleine, Robert Baxter, Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. Samples are made available to researchers on a non-exclusive basis. BBCS thanks Eileen Williams, Elaine Ryder-Mills, Kara Sargus. BCEES thanks Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen Western Australia, Elizabeth Wylie, Rachel Lloyd. The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. BIGGS thanks Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones. The BREOGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. BSUCH thanks Peter Bugert, Medical

Faculty Mannheim. CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CCGP thanks Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, Konstantinos Pompodakis. CGPS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. CNIO-BCS thanks Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, Primitiva Menendez, the Human Genotyping-CEGEN Unit (CNIO). Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. The CTS Steering Committee includes Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, and Jessica Clague DeHart at the Beckman Research Institute of City of Hope, Dennis Deapen, Rich Pinder, and Eunjung Lee at the University of Southern California, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson at the Cancer Prevention Institute of California, Hoda Anton-Culver, Argyrios Ziogas, and Hannah Park at the University of California Irvine, and Fred Schumacher at Case Western University. DIETCOMPLYF thanks the patients, nurses and clinical staff involved in the study. The DietCompLyf study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). ESTHER thanks Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. FHRISK thanks NIHR for funding. GC-HBOC thanks Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber). The GENICA Network: Dr. Margarete Fischer-

Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen, Germany [HB, WYL], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2180 - 390900677 [HB], Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany [YDK, Christian Baisch], Institute of Pathology, University of Bonn, Germany [Hans-Peter Fischer], Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [Ute Hamann], Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany [TB, Beate Pesch, Sylvia Rabstein, Anne Lotz]; and Institute of Occupational Medicine and Maritime Medicine, University Medical Center Hamburg-Eppendorf, Germany [Volker Harth]. HABCS thanks Michael Bremer. HEBCS thanks Johanna Kiiski, Rainer Fagerholm, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä. HMBCS thanks Peter Hillemanns, Hans Christiansen and Johann H. Karstens. HUBCS thanks Shamil Gantsev. KARMA and SASBAC thank the Swedish Medical Research Counsel. KBCP thanks Eija Myöhänen, Helena Kemiläinen. kConFab/AOCS wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. LMBC thanks Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. MABCS thanks Emilija Lazarova (Clinic of Radiotherapy and Oncology), Dzengis Jasar, Mitko Karadjozov (Adzibadem-Sistina” Hospital), Andrej Arsovski and Liljana Stojanovska (Re-Medika” Hospital) for their contributions and commitment to this study. MARIE thanks Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. MBCSG (Milan Breast Cancer Study Group): Irene Feroce, Aliana Guerrieri Gonzaga, Monica Marabelli and and the personnel of the Cogentech Cancer Genetic Test Laboratory. The MCCS was made possible by the

contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. MSKCC thanks Marina Corines, Lauren Jacobs. MTLGEBCS would like to thank Martine Tranchant (CHU de Québec – Université Laval Research Center), Marie-France Valois, Annie Turgeon and Lea Heguy (McGill University Health Center, Royal Victoria Hospital; McGill University) for DNA extraction, sample management and skilful technical assistance. J.S. is Chair holder of the Canada Research Chair in Oncogenetics. The following are NBCS Collaborators: Kristine K. Sahlberg (PhD), Lars Ottestad (MD), Rolf Kåresen (Prof. Em.) Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD), Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geisler (MD), OSBREAC and Grethe I. Grenaker Alnæs (MSc). NBHS and For NHS and NHS2 the study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We would like to thank the participants and staff of the NHS and NHS2 for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data. OBCS thanks Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo and Kari Mononen for their contributions to this study. OFBCR thanks Teresa Selander, Nayana Weerasooriya. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information. The LUMC survival data were retrieved from the Leiden hospital-based cancer registry system (ONCDOC) with the help of Dr. J. Molenaar. PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The ethical approval for the POSH

study is MREC /00/6/69, UKCRN ID: 1137. We thank staff in the Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. PREFACE thanks Sonja Oeser and Silke Landrith. PROCAS thanks NIHR for funding. RBCS thanks Jannet Blom, Saskia Pelders, Annette Heemskerk and the Erasmus MC Family Cancer Clinic. SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. We thank the SEARCH and EPIC teams. SGBCC thanks the participants and research coordinator Ms Tan Siew Li. SKKDKFZS thanks all study participants, clinicians, family doctors, researchers and technicians for their contributions and commitment to this study. We thank the SUCCESS Study teams in Munich, Duessldorf, Erlangen and Ulm. SZBCS thanks Ewa Putresza. UCIBCS thanks Irene Masunaka. UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. DGE, and AH, are supported by the all Manchester NIHR Biomedical Research Centre (IS-BRC-1215-20007). The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible. Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute.

Members of consortia listed as authors

kConFab/AOCS Investigators

Stephen Fox, Ian Campbell (Peter MacCallum Cancer Centre, Melbourne, Australia); Georgia Chenevix-Trench, Amanda Spurdle, Penny Webb (QIMR Berghofer Medical Research Institute, Brisbane, Australia); Anna de Fazio (Westmead Millenium Institute, Sydney, Australia); Margaret Tassell (BCNA delegate, Community Representative); Judy Kirk (Westmead Hospital, Sydney, Australia); Geoff Lindeman (Walter and Eliza Hall Institute, Melbourne, Australia); Melanie Price

(University of Sydney, Sydney, Australia); Melissa Southey (University of Melbourne, Melbourne, Australia); Roger Milne (Cancer Council Victoria, Melbourne, Australia); Sid Deb (Melbourne Health, Melbourne, Australia); David Bowtell (Garvan Institute of Medical Research, Sydney, Australia).

ABCTB Investigators

Christine Clarke (Westmead Institute for Medical Research, University of Sydney, NSW, Australia); Rosemary Balleine (Pathology West ICPMR, Westmead, NSW, Australia); Robert Baxter (Kolling Institute of Medical Research, University of Sydney, Royal North Shore Hospital, NSW, Australia); Stephen Braye (Pathology North, John Hunter Hospital, Newcastle, NSW, 2305, Australia); Jane Carpenter (Westmead Institute for Medical Research, University of Sydney); Jane Dahlstrom (Department of Anatomical Pathology, ACT Pathology, Canberra Hospital, ACT, Australia; ANU Medical School, Australian National University, ACT, Australia); John Forbes (Department of Surgical Oncology, Calvary Mater Newcastle Hospital, Australian New Zealand Breast Cancer Trials Group, and School of Medicine and Public Health, University of Newcastle, NSW, Australia); C Soon Lee (School of Science and Health, The University of Western Sydney, Sydney, Australia); Deborah Marsh (Hormones and Cancer Group, Kolling Institute of Medical Research, Royal North Shore Hospital, University of Sydney, NSW, Australia); Adrienne Morey (SydPath St Vincent's Hospital, Sydney, NSW, Australia); Nirmala Pathmanathan (Department of Tissue Pathology and Diagnostic Oncology, Pathology West; Westmead Breast Cancer Institute, Westmead Hospital, NSW, Australia); Rodney Scott (Centre for Information Based Medicine, Hunter Medical Research Institute, NSW, 2305, Australia; Priority Research Centre for Cancer, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, NSW, Australia); Peter Simpson (The University of Queensland: UQ Centre for Clinical Research and School of Medicine, QLD, Australia); Allan Spigelman (Hereditary Cancer Clinic, St Vincent's Hospital, The Kinghorn Cancer Centre, Sydney, New South Wales, 2010, Australia); Nicholas Wilcken (Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, Australia; Sydney Medical School - Westmead, University of

Sydney, NSW, Australia); Desmond Yip (Department of Medical Oncology, The Canberra Hospital, ACT, Australia; ANU Medical School, Australian National University, ACT, Australia); Nikolajs Zeps (St John of God Perth Northern Hospitals, Perth, WA, Australia).

CIMBA Funding and Acknowledgements

Funding

CIMBA: The CIMBA data management and data analysis were supported by Cancer Research – UK grants C12292/A20861, C12292/A11174. ACA is a Cancer Research -UK Senior Cancer Research Fellow. GCT and ABS are NHMRC Research Fellows. iCOGS: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer (CRN-87521), and the Ministry of Economic Development, Innovation and Export Trade (PSR-SIIRI-701), Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. The PERSPECTIVE project was supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministry of Economy, Science and Innovation through Genome Québec, and The Quebec Breast Cancer Foundation. BCFR: UM1 CA164920 from the National Cancer Institute. 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 mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. BFBOCC: Lithuania (BFBOCC-LT): Research Council of Lithuania grant SEN-18/2015. BIDMC: Breast Cancer Research Foundation. BMBSA: Cancer Association of South Africa (PI Elizabeth J. van Rensburg). CNIO: Spanish Ministry of Health PI16/00440 supported by FEDER funds, the Spanish Ministry of Economy and Competitiveness (MINECO) SAF2014-57680-R and the Spanish Research Network on Rare diseases (CIBERER). COH-CCGCRN: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under grant number

R25CA112486, and RC4CA153828 (PI: J. Weitzel) from the National Cancer Institute and the Office of the Director, National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. CONSIT TEAM: Associazione Italiana Ricerca sul Cancro (AIRC; IG2015 no.16732) to P. Peterlongo**.** DEMOKRITOS: European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program of the General Secretariat for Research & Technology: SYN11_10_19 NBCA. Investing in knowledge society through the European Social Fund. DFKZ: German Cancer Research Center. EMBRACE: Cancer Research UK Grants C1287/A10118 and C1287/A11990. D. Gareth Evans and Fiona Lalloo are supported by an NIHR grant to the Biomedical Research Centre, Manchester. The Investigators at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust are supported by an NIHR grant to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. Ros Eeles and Elizabeth Bancroft are supported by Cancer Research UK Grant C5047/A8385. Ros Eeles is also supported by NIHR support to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. FCCC: The University of Kansas Cancer Center (P30 CA168524) and the Kansas Bioscience Authority Eminent Scholar Program. A.K.G. was funded by R0 1CA140323, R01 CA214545, and by the Chancellors Distinguished Chair in Biomedical Sciences Professorship. A.Vega is supported by the Spanish Health Research Foundation, Instituto de Salud Carlos III (ISCIII), partially supported by FEDER funds through Research Activity Intensification Program (contract grant numbers: INT15/00070, INT16/00154, INT17/00133), and through Centro de Investigación Biomédica en Red de Enferemdades Raras CIBERER (ACCI 2016: ER17P1AC7112/2018); Autonomous Government of Galicia (Consolidation and structuring program: IN607B), and by the Fundación Mutua Madrileña. GC-HBOC: German Cancer Aid (grant no 110837, Rita K. Schmutzler) and the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202,

713-241202, 14505/2470, 14575/2470). GEMO: Ligue Nationale Contre le Cancer; the Association “Le cancer du sein, parlons-en!” Award, the Canadian Institutes of Health Research for the "CIHR Team in Familial Risks of Breast Cancer" program, the Fondation ARC pour la recherche sur le cancer (grant PJA 20151203365) and the French National Institute of Cancer (INCa grants AOR 01 082, 2001-2003, 2013-1-BCB-01-ICH-1 and SHS-E-SP 18-015). GEORGETOWN: the Non-Therapeutic Subject Registry Shared Resource at Georgetown University (NIH/NCI grant P30-CA051008), the Fisher Center for Hereditary Cancer and Clinical Genomics Research, and Swing Fore the Cure. G-FAST: Bruce Poppe is a senior clinical investigator of FWO. Mattias Van Heetvelde obtained funding from IWT. HCSC: Spanish Ministry of Health PI15/00059, PI16/01292, and CB-161200301 CIBERONC from ISCIII (Spain), partially supported by European Regional Development FEDER funds. HEBCS: Helsinki University Hospital Research Fund, the Finnish Cancer Society and the Sigrid Juselius Foundation. HEBON: the Dutch Cancer Society grants NKI1998-1854, NKI2004-3088, NKI2007-3756, the Netherlands Organization of Scientific Research grant NWO 91109024, the Pink Ribbon grants 110005 and 2014-187.WO76, the BBMRI grant NWO 184.021.007/CP46 and the Transcan grant JTC 2012 Cancer 12-054. HEBON thanks the registration teams of Dutch Cancer Registry (IKNL; S. Siesling, J. Verloop) and the Dutch Pathology database (PALGA; L. Overbeek) for part of the data collection. HUNBOCS: Hungarian Research Grants KTIA-OTKA CK-80745 and NKFI_OTKA K-112228. ICO: The authors would like to particularly acknowledge the support of the Asociación Española Contra el Cáncer (AECC), the Instituto de Salud Carlos III (organismo adscrito al Ministerio de Economía y Competitividad) and “Fondo Europeo de Desarrollo Regional (FEDER), una manera de hacer Europa” (PI10/01422, PI13/00285, PIE13/00022, PI15/00854, PI16/00563 and CIBERONC) and the Institut Català de la Salut and Autonomous Government of Catalonia (2009SGR290, 2014SGR338 and PERIS Project MedPerCan). IHCC: PBZ_KBN_122/P05/2004. INHERIT: Canadian Institutes of Health Research for the “CIHR Team in Familial Risks of Breast Cancer” program – grant # CRN-87521 and the Ministry of Economic Development, Innovation and Export Trade – grant # PSR-SIIRI-701. IOVHBOCS: Ministero della Salute and “5x1000” Istituto

Oncologico Veneto grant. IPOBCS: Liga Portuguesa Contra o Cancro. kConFab: The National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. MAYO: NIH grants CA116167, CA192393 and CA176785, an NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201),and a grant from the Breast Cancer Research Foundation. MCGILL: Jewish General Hospital Weekend to End Breast Cancer, Quebec Ministry of Economic Development, Innovation and Export Trade. Marc Tischkowitz is supported by the funded by the European Union Seventh Framework Program (2007Y2013)/European Research Council (Grant No. 310018). MODSQUAD: MH CZ - DRO (MMCI, 00209805), **MEYS - NPS I - LO1413** to LF, and by Charles University in Prague project UNCE204024 (MZ). MSKCC: the Breast Cancer Research Foundation, the Robert and Kate Niehaus Clinical Cancer Genetics Initiative, the Andrew Sabin Research Fund and a Cancer Center Support Grant/Core Grant (P30 CA008748). NAROD: 1R01 CA149429-01. NCI: the Intramural Research Program of the US National Cancer Institute, NIH, and by support services contracts NO2-CP-11019-50, N02-CP-21013-63 and N02-CP-65504 with Westat, Inc, Rockville, MD. NICCC: Clalit Health Services in Israel, the Israel Cancer Association and the Breast Cancer Research Foundation (BCRF), NY. NNPIO: the Russian Foundation for Basic Research (grants 17-00-00171, 18-515-45012 and 19-515-25001). NRG Oncology: U10 CA180868, NRG SDMC grant U10 CA180822, NRG Administrative Office and the NRG Tissue Bank (CA 27469), the NRG Statistical and Data Center (CA 37517) and the Intramural Research Program, NCI, KAP is an Australian National Breast Cancer Foundation Fellow. OSUCCG: Ohio State University Comprehensive Cancer Center. PBCS: Italian Association of Cancer Research (AIRC) [IG 2013 N.14477] and Tuscany Institute for Tumors (ITT) grant 2014-2015-2016. SMC: the Israeli Cancer Association. SWE-BRCA: the Swedish Cancer Society. UCHICAGO: NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA125183), R01 CA142996, 1U01CA161032 and by the Ralph and Marion Falk Medical Research Trust, the Entertainment Industry Fund National

Women's Cancer Research Alliance and the Breast Cancer research Foundation. OIO is an ACS Clinical Research Professor. UCSF: UCSF Cancer Risk Program and Helen Diller Family Comprehensive Cancer Center. UKFOCR: Cancer Researc h UK. UPENN: Breast Cancer Research Foundation (to SMD, KLN); Susan G. Komen Foundation for the cure (SMD), Basser Research Center for BRCA (SMD, KLN). UPITT/MWH: Hackers for Hope Pittsburgh. VFCTG: Victorian Cancer Agency, Cancer Australia, National Breast Cancer Foundation. WCP: Dr Karlan is funded by the American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000124. Tracy A. O'Mara was supported by NHMRC Early Career Research Fellow.

Acknowledgements

All the families and clinicians who contribute to the studies; Catherine M. Phelan for her contribution to CIMBA until she passed away on 22 September 2017; Sue Healey, in particular taking on the task of mutation classification with the late Olga Sinilnikova; Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis; members and participants in the New York site of the Breast Cancer Family Registry; members and participants in the Ontario Familial Breast Cancer Registry; Vilius Rudaitis and Laimonas Griškevičius; Drs Janis Eglitis, Anna Krilova and Aivars Stengrevics; Yuan Chun Ding and Linda Steele for their work in participant enrollment and biospecimen and data management; Bent Ejlertsen and Anne-Marie Gerdes for the recruitment and genetic counseling of participants; Alicia Barroso, Rosario Alonso and Guillermo Pita; all the individuals and the researchers who took part in CONSIT TEAM (Consorzio Italiano Tumori Ereditari Alla Mammella), in particular: Dario Zimbalatti, Daniela Zaffaroni, Laura Ottini, Giuseppe Giannini, Liliana Varesco, Viviana Gismondi, Maria Grazia Tibiletti, Daniela Furlan, Antonella Savarese, Aline Martayan, Stefania Tommasi, Brunella Pilato and the personnel of the Cogentech Cancer Genetic Test Laboratory, Milan, Italy. Ms. JoEllen Weaver and Dr. Betsy Bove; FPGMX: members of the Cancer Genetics group (IDIS): Ana Blanco, Miguel Aguado, Uxía Esperón and Belinda Rodríguez.; IFE - Leipzig Research Centre for Civilization

Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber); We thank all participants, clinicians, family doctors, researchers, and technicians for their contributions and commitment to the DKFZ study and the collaborating groups in Lahore, Pakistan (Noor Muhammad, Sidra Gull, Seerat Bajwa, Faiz Ali Khan, Humaira Naeemi, Saima Faisal, Asif Loya, Mohammed Aasim Yusuf) and Bogota, Colombia (Diana Torres, Ignacio Briceno, Fabian Gil). Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) study is a study from the National Cancer Genetics Network UNICANCER Genetic Group, France. We wish to pay a tribute to Olga M. Sinilnikova, who with Dominique Stoppa-Lyonnet initiated and coordinated GEMO until she sadly passed away on the 30th June 2014. The team in Lyon (Olga Sinilnikova, Mélanie Léoné, Laure Barjhoux, Carole Verny-Pierre, Sylvie Mazoyer, Francesca Damiola, Valérie Sornin) managed the GEMO samples until the biological resource centre was transferred to Paris in December 2015 (Noura Mebirouk, Fabienne Lesueur, Dominique Stoppa-Lyonnet). We want to thank all the GEMO collaborating groups for their contribution to this study: Coordinating Centre, Service de Génétique, Institut Curie, Paris, France: Muriel Belotti, Ophélie Bertrand, Anne-Marie Birot, Bruno Buecher, Sandrine Caputo, Anaïs Dupré, Emmanuelle Fourme, Marion Gauthier-Villars, Lisa Golmard, Claude Houdayer, Marine Le Mentec, Virginie Moncoutier, Antoine de Pauw, Claire Saule, Dominique Stoppa-Lyonnet, and Inserm U900, Institut Curie, Paris, France: Fabienne Lesueur, Noura Mebirouk.Contributing Centres : Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, Lyon, France: Nadia Boutry-Kryza, Alain Calender, Sophie Giraud, Mélanie Léone. Institut Gustave Roussy, Villejuif, France: Brigitte Bressac-de-Paillerets, Olivier Caron, Marine Guillaud-Bataille. Centre Jean Perrin, Clermont–Ferrand, France: Yves-Jean Bignon, Nancy Uhrhammer. Centre Léon Bérard, Lyon, France: Valérie Bonadona, Christine Lasset. Centre François Baclesse, Caen, France: Pascaline Berthet, Laurent Castera, Dominique Vaur. Institut Paoli Calmettes, Marseille, France: Violaine Bourdon, Catherine Noguès, Tetsuro Noguchi, Cornel Popovici, Audrey Remenieras, Hagay Sobol. CHU Arnaud-de-Villeneuve, Montpellier, France: Isabelle Coupier, Pascal Pujol. Centre Oscar Lambret, Lille, France: Claude

Adenis, Aurélie Dumont, Françoise Révillion. Centre Paul Strauss, Strasbourg, France: Danièle Muller. Institut Bergonié, Bordeaux, France: Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubien, Michel Longy, Nicolas Sevenet, Institut Claudius Regaud, Toulouse, France: Laurence Gladieff, Rosine Guimbaud, Viviane Feillel, Christine Toulas. CHU Grenoble, France: Hélène Dreyfus, Christine Dominique Leroux, Magalie Peysselon, Rebischung. CHU Dijon, France: Amandine Baurand, Geoffrey Bertolone, Fanny Coron, Laurence Faivre, Caroline Jacquot, Sarab Lizard. CHU St-Etienne, France: Caroline Kientz, Marine Lebrun, Fabienne Prieur. Hôtel Dieu Centre Hospitalier, Chambéry, France: Sandra Fert Ferrer. Centre Antoine Lacassagne, Nice, France: Véronique Mari. CHU Limoges, France: Laurence Vénat-Bouvet. CHU Nantes, France: Stéphane Bézieau, Capucine Delnatte. CHU Bretonneau, Tours and Centre Hospitalier de Bourges France: Isabelle Mortemousque. Groupe Hospitalier Pitié-Salpétrière, Paris, France: Chrystelle Colas, Florence Coulet, Florent Soubrier, Mathilde Warcoin. CHU Vandoeuvre-les-Nancy, France: Myriam Bronner, Johanna Sokolowska. CHU Besançon, France: Marie-Agnès Collonge-Rame, Alexandre Damette. CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, France: Paul Gesta. Centre Hospitalier de La Rochelle : Hakima Lallaoui. CHU Nîmes Carémeau, France : Jean Chiesa. CHI Poissy, France: Denise Molina-Gomes. CHU Angers, France : Olivier Ingster; Ilse Coene en Brecht Crombez; Ilse Coene and Brecht Crombez; Alicia Tosar and Paula Diaque; Irja Erkkilä and Virpi Palola; The Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON) consists of the following Collaborating Centers: Coordinating center: Netherlands Cancer Institute, Amsterdam, NL: M.A. Rookus, F.B.L. Hogervorst, F.E. van Leeuwen, S. Verhoef, M.K. Schmidt, N.S. Russell, D.J. Jenner; Erasmus Medical Center, Rotterdam, NL: J.M. Collée, A.M.W. van den Ouweland, M.J. Hooning, C. Seynaeve, C.H.M. van Deurzen, I.M. Obdeijn; Leiden University Medical Center, NL: C.J. van Asperen, J.T. Wijnen, R.A.E.M. Tollenaar, P. Devilee, T.C.T.E.F. van Cronenburg; Radboud University Nijmegen Medical Center, NL: C.M. Kets, A.R. Mensenkamp; University Medical Center Utrecht, NL: M.G.E.M. Ausems, R.B. van der Luijt, C.C. van der Pol; Amsterdam Medical Center, NL: C.M. Aalfs, T.A.M. van Os; VU University Medical Center,

Amsterdam, NL: J.J.P. Gille, Q. Waisfisz, H.E.J. Meijers-Heijboer; University Hospital Maastricht, NL: E.B. Gómez-Garcia, M.J. Blok; University Medical Center Groningen, NL: J.C. Oosterwijk, A.H. van der Hout, M.J. Mourits, G.H. de Bock; The Netherlands Foundation for the detection of hereditary tumours, Leiden, NL: H.F. Vasen; The Netherlands Comprehensive Cancer Organization (IKNL): S. Siesling, J.Verloop; The Dutch Pathology Registry (PALGA): L.I.H. Overbeek; Hong Kong Sanatorium and Hospital; the Hungarian Breast and Ovarian Cancer Study Group members (Janos Papp, Aniko Bozsik, Timea Pocza, Zoltan Matrai, Miklos Kasler, Judit Franko, Maria Balogh, Gabriella Domokos, Judit Ferenczi, Department of Molecular Genetics, National Institute of Oncology, Budapest, Hungary) and the clinicians and patients for their contributions to this study; the Oncogenetics Group (VHIO) and the High Risk and Cancer Prevention Unit of the University Hospital Vall d'Hebron, Miguel Servet Progam (CP10/00617), and the Cellex Foundation for providing research facilities and equipment; the ICO Hereditary Cancer Program team led by Dr. Gabriel Capella; the ICO Hereditary Cancer Program team led by Dr. Gabriel Capella; Dr Martine Dumont for sample management and skillful assistance; Catarina Santos and Pedro Pinto; members of the Center of Molecular Diagnosis, Oncogenetics Department and Molecular Oncology Research Center of Barretos Cancer Hospital; Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab; the KOBRA Study Group; (National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA); Eva Machackova (Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute and MF MU, Brno, Czech Republic); and Michal Zikan, Petr Pohlreich and Zdenek Kleibl (Oncogynecologic Center and Department of Biochemistry and Experimental Oncology, First Faculty of Medicine, Charles University, Prague, Czech Republic); Anne Lincoln, Lauren Jacobs; the participants in Hereditary Breast/Ovarian Cancer Study and Breast Imaging Study for their selfless contributions to our research; the NICCC National Familial Cancer

Consultation Service team led by Sara Dishon, the lab team led by Dr. Flavio Lejbkowicz, and the research field operations team led by Dr. Mila Pinchev; the investigators of the Australia New Zealand NRG Oncology group; members and participants in the Ontario Cancer Genetics Network; Kevin Sweet, Caroline Craven, Julia Cooper, Amber Aielts, and Michelle O'Conor; Yip Cheng Har, Nur Aishah Mohd Taib, Phuah Sze Yee, Norhashimah Hassan and all the research nurses, research assistants and doctors involved in the MyBrCa Study for assistance in patient recruitment, data collection and sample preparation, Philip Iau, Sng Jen-Hwei and Sharifah Nor Akmal for contributing samples from the Singapore Breast Cancer Study and the HUKM-HKL Study respectively; the Meirav Comprehensive breast cancer center team at the Sheba Medical Center; Christina Selkirk; Helena Jernström, Karin Henriksson, Katja Harbst, Maria Soller, Ulf Kristoffersson; from Gothenburg Sahlgrenska University Hospital: Anna Öfverholm, Margareta Nordling, Per Karlsson, Zakaria Einbeigi; from Stockholm and Karolinska University Hospital: Anna von Wachenfeldt, Annelie Liljegren, Brita Arver, Gisela Barbany Bustinza; from Umeå University Hospital: Beatrice Melin, Christina Edwinsdotter Ardnor, Monica Emanuelsson; from Uppsala University: Hans Ehrencrona, Maritta Hellström Pigg, Richard Rosenquist; from Linköping University Hospital: Marie Stenmark-Askmalm, Sigrun Liedgren; Cecilia Zvocec, Qun Niu; Joyce Seldon and Lorna Kwan; Dr. Robert Nussbaum, Beth Crawford, Kate Loranger, Julie Mak, Nicola Stewart, Robin Lee, Amie Blanco and Peggy Conrad and Salina Chan; Carole Pye, Patricia Harrington and Eva Wozniak; Geoffrey Lindeman, Marion Harris, Martin Delatycki, Sarah Sawyer, Rebecca Driessen, and Ella Thompson for performing all DNA amplification.

Members of consortia listed as authors

EMBRACE

Helen Gregory (North of Scotland Regional Genetics Service, NHS Grampian & University of Aberdeen, Foresterhill, Aberdeen, UK); Zosia Miedzybrodzka (North of Scotland Regional Genetics Service, NHS Grampian & University of Aberdeen, Foresterhill, Aberdeen, UK); Patrick J. Morrison

(Northern Ireland Regional Genetics Centre, Belfast Health and Social Care Trust, and Department of Medical Genetics, Queens University Belfast, Belfast, UK); Kai-ren Ong (West Midlands Regional Genetics Service, Birmingham Women's Hospital Healthcare NHS Trust, Edgbaston, Birmingham, UK); Alan Donaldson (Clinical Genetics Department, St Michael's Hospital, Bristol, UK); Marc Tischkowitz (Department of Medical Genetics, University of Cambridge, UK); Mark T. Rogers (All Wales Medical Genetics Services, University Hospital of Wales, Cardiff, UK); M. John Kennedy (Academic Unit of Clinical and Molecular Oncology, Trinity College Dublin and St James's Hospital, Dublin, Eire); Mary E. Porteous (South East of Scotland Regional Genetics Service, Western General Hospital, Edinburgh, UK); Carole Brewer (Department of Clinical Genetics, Royal Devon & Exeter Hospital, Exeter, UK); Rosemarie Davidson (Clinical Genetics, Southern General Hospital, Glasgow, UK); Louise Izatt (Clinical Genetics, Guy's and St. Thomas' NHS Foundation Trust, London, UK); Angela Brady (North West Thames Regional Genetics Service, Kennedy-Galton Centre, Harrow, UK); Julian Barwell (Leicestershire Clinical Genetics Service, University Hospitals of Leicester NHS Trust, UK); Julian Adlard (Yorkshire Regional Genetics Service, Leeds, UK); Claire Foo (Department of Clinical Genetics, Alder Hey Hospital, Eaton Road, Liverpool, UK); D. Gareth Evans (Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK); Fiona Lalloo (Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK); Lucy E. Side (North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK); Jacqueline Eason (Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, UK); Alex Henderson (Institute of Genetic Medicine, Centre for Life, Newcastle Upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK); Lisa Walker (Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK); Rosalind A. Eeles (Oncogenetics Team, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, UK); Jackie Cook (Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield, UK); Katie Snape (South West Thames Regional Genetics Service, St.Georges Hospital, Cranmer

Terrace, Tooting, London, UK); Diana Eccles (University of Southampton Faculty of Medicine, Southampton University Hospitals NHS Trust, Southampton, UK); Alex Murray (All Wales Medical Genetics Services, Singleton Hospital, Swansea, UK); Emma McCann (All Wales Medical Genetics Service, Glan Clwyd Hospital, Rhyl, UK).

GEMO Study Collaborators

Dominique Stoppa-Lyonnet, Muriel Belotti, Anne-Marie Birot, Bruno Buecher, Emmanuelle Fourme, Marion Gauthier-Villars, Lisa Golmard, Claude Houdayer, Virginie Moncoutier, Antoine de Pauw, Claire Saule (Service de Génétique, Institut Curie, Paris, France); Fabienne Lesueur, Noura Mebirouk (Inserm U900, Institut Curie, Paris, France); Olga Sinilnikova†, Sylvie Mazoyer, Francesca Damiola, Laure Barjhoux, Carole Verny-Pierre, Mélanie Léone, Nadia Boutry-Kryza, Alain Calender, Sophie Giraud (Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, Lyon, France); Olivier Caron, Marine GuillaudBataille (Institut Gustave Roussy, Villejuif, France: Brigitte Bressac-de-Paillerets); YvesJean Bignon, Nancy Uhrhammer (Centre Jean Perrin, Clermont–Ferrand, France); Christine Lasset, Valérie Bonadona (Centre Léon Bérard, Lyon, France); Pascaline Berthet, Dominique Vaur, Laurent Castera (Centre François Baclesse, Caen, France); Hagay Sobol, Violaine Bourdon, Tetsuro Noguchi, Audrey Remenieras, François Eisinger, Catherine Noguès (Institut Paoli Calmettes, Marseille, France); Isabelle Coupier, Pascal Pujol (CHU Arnaud-de-Villeneuve, Montpellier, France); Jean-Philippe Peyrat, Joëlle Fournier, Françoise Révillion, Claude Adenis (Centre Oscar Lambret, Lille, France); Danièle Muller, Jean-Pierre Fricker (Centre Paul Strauss, Strasbourg, France); Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubien, Nicolas Sevenet, Michel Longy (Institut Bergonié, Bordeaux, France); Christine Toulas, Rosine Guimbaud, Laurence Gladieff, Viviane Feillel (Institut Claudius Regaud, Toulouse, France); Dominique Leroux, Hélène Dreyfus, Christine Rebischung, Magalie Peysselon (CHU Grenoble, France); Fanny Coron, Laurence Faivre, Amandine Baurand, Caroline Jacquot, Geoffrey, Bertolone, Sarab Lizard (CHU Dijon, France); Fabienne Prieur, Marine Lebrun, Caroline Kientz (CHU St-

Etienne, France); Sandra Fert Ferrer (Hôtel Dieu Centre Hospitalier, Chambéry, France); Véronique Mari (Centre Antoine Lacassagne, Nice, France); Laurence Vénat-Bouvet (CHU Limoges, France); Capucine Delnatte, Stéphane Bézieau (CHU Nantes, France); Isabelle Mortemousque (CHU Bretonneau, Tours and Centre Hospitalier de Bourges France); Florence Coulet, Chrystelle Colas, Florent Soubrier, Mathilde Warcoin (Groupe Hospitalier PitiéSalpétrière, Paris, France); Johanna Sokolowska, Myriam Bronner (CHU Vandoeuvreles-Nancy, France); Marie-Agnès Collonge-Rame, Alexandre Damette (CHU Besançon, France); Paul Gesta (CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, France); Hakima Lallaoui (Centre Hospitalier de La Rochelle); Jean Chiesa (CHU Nîmes Carémeau, France); Denise Molina-Gomes (CHI Poissy, France); Olivier Ingster (CHU Angers, France).