nature COMMUNICATIONS

ARTICLE

2q36.3 is associated with prognosis for oestrogen receptor-negative breast cancer patients treated with chemotherapy

Jingmei Li[1,2,*], Linda S. Lindström[3,4,*] et al.[#]

Large population-based registry studies have shown that breast cancer prognosis is inherited. Here we analyse single-nucleotide polymorphisms (SNPs) of genes implicated in human immunology and inflammation as candidates for prognostic markers of breast cancer survival involving 1,804 oestrogen receptor (ER)-negative patients treated with chemotherapy (279 events) from 14 European studies in a prior large-scale genotyping experiment, which is part of the Collaborative Oncological Gene-environment Study (COGS) initiative. We carry out replication using Asian COGS samples ($n = 522$, 53 events) and the Prospective Study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) study ($n = 315$, 108 events). Rs4458204_A near CCL20 (2q36.3) is found to be associated with breast cancer-specific death at a genome-wide significant level ($n = 2,641$, 440 events, combined allelic hazard ratio (HR) = 1.81 (1.49–2.19); P for trend = 1.90×10^{-9}). Such survival-associated variants can represent ideal targets for tailored therapeutics, and may also enhance our current prognostic prediction capabilities.

Correspondence and requests for materials should be addressed to K.C.K. (email: Kamila.Czene@ki.se).
[#]A full list of authors and their affiliations appears at the end of the paper.

W e have previously shown, through large population-based registry studies, that survival from breast cancer is correlated among relatives, consistent with an inherited cancer prognosis[1–4]. A potential explanation for the heritability of survival would be that family members are predisposed to developing a breast cancer tumour of predefined aetiology and predetermined tumour characteristics. This is plausible given the observation that carriers of high- and moderate-risk germline mutations in genes such as *BRCA1*, *BRCA2*, *CHEK2* and *PALB2*, are predisposed to specific subtypes of breast cancer[5–8], and that many common variants identified through genome-wide association studies (GWAS) tend to be associated with specific subtypes, with some variants more strongly associated with oestrogen receptor (ER)-negative or triple-negative breast cancer[9–11], while others more strongly associated with ER-positive breast cancer[12–14].

It is also possible that the inherited predeterminants of survival lie not in the biology of the tumour but rather the milieu in which the tumour arises. The tumour microenvironment is composed of tumour cells, fibroblasts, endothelial cells and infiltrating immune cells, which may inhibit or promote tumour growth and progression. There is empirical support for the concept that a host immune response might enhance the effects of conventional chemotherapy, conceivably having an influence on breast cancer outcome. For example, the presence of tumour-associated lymphocytes in a breast tumour has been suggested to be an independent predictor of neoadjuvant chemotherapy response[15]. Other studies have shown the host immune system to be involved in the elimination of tumour cells to control cancer growth[16,17].

In this candidate pathway study, we investigate the pre-specified hypothesis that the germline common variants of genes involved in immune response and inflammation can predict the response to breast cancer survival for ER-negative, chemotherapy-treated patients. We identify a single-nucleotide polymorphism (SNP) near the *CCL20* gene (2q36.3), which is associated with a difference in the clinical outcome of ER-negative breast cancer treated with chemotherapy independent of known tumour prognostic features.

Results

Individual patient-level genetic and phenotypic data were extracted from European studies in a prior large-scale genotyping experiment conducted in the Breast Cancer Association Consortium (BCAC), part of the Collaborative Oncological Gene-environment Study (COGS) initiative[18]. For this study, we selected women of European descent inferred from genetic ancestry with invasive breast cancer, who have had no previous diagnosis of the disease. Subjects missing follow-up information on vital status, time to vital status, date of study entry and cause of death data were excluded.

The selection of only ER-negative patients in this study was strongly motivated by prior insight. A Swedish study of the breast cancer prognosis of 834 sister pairs in which both were affected showed that younger sisters with poor older sister survival had worse survival than younger sisters with good older sister survival (number of breast cancer deaths within 5 years from diagnosis in younger sisters, $n_{event} = 65$, $P = 0.02$ in a multivariate proportional hazard (Cox) analysis)[3]. When stratified by ER subtypes, the increased risk of death from ER-negative breast cancer for younger sisters with poor older sister survival compared with younger sisters with good older sister survival was found to be almost sevenfold ($n = 139$ sister pairs, $n_{event} = 28$, hazard ratio (HR) = 6.69 (1.36–32.91), $P = 0.02$) in contrast to sister pairs with the ER-positive disease ($n = 584$ sister pairs, $n_{event} = 28$, HR = 1.54 (0.48–4.98), $P = 0.50$) (unpublished data). In addition,
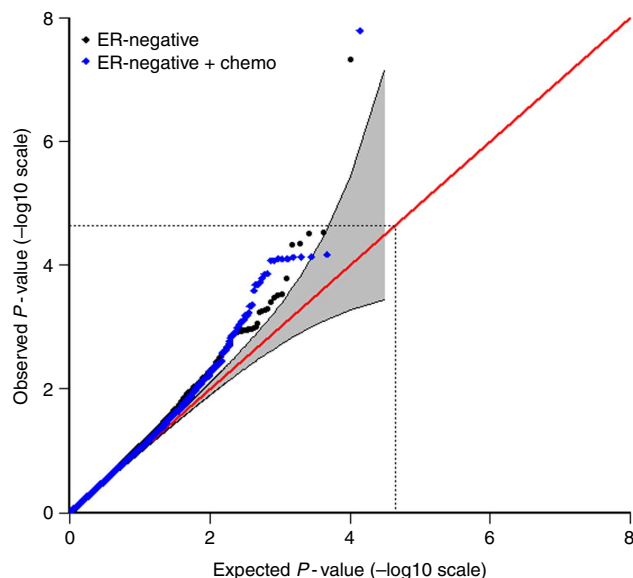

Figure 1 | Quantile–quantile (QQ) plots of the observed *P*-values for association in the discovery stage. QQ plot of the observed $-\log10$ *P*-values (*y* axis) versus the 'expected' $-\log10$ rank *P*-values (*x* axis) for trend tests of association of 7,020 human immunology and inflammation SNPs, with the risk of dying from breast cancer for all ER-negative breast cancer patients (black/below) and ER-negative patients treated with adjuvant chemotherapy (blue/above) (genomic inflation factor, $\lambda = 1.16$ and 1.14, respectively) in the discovery phase. The grey region indicates bootstrapped 95% confidence intervals. The diagonal red line indicates expected results under null hypothesis. The dotted lines indicate Bonferroni threshold for multiple-testing correction (2,184 independent tests with $r^2 < 0.2$).

in a recent Breast International Group phase III trial, increasing lymphocytic infiltration was found to be associated with excellent prognosis only for patients with node-positive, ER-negative/HER2-negative disease[19]. Twenty studies with ER-negative cases and at least one event (breast cancer-specific death) were eligible for the combined analysis (Supplementary Table 1). As we were primarily interested in response to chemotherapy, patients missing information on chemotherapy were not considered in our analyses. The 14 studies ($n = 1,804$) included in the combined analysis for the chemotherapy-treated subgroup are summarized in Supplementary Table 2. A total of 279 breast cancer-specific deaths were recorded in a 15-year follow-up.

For the replication phase, four iCOGS Asian studies with ER-negative breast cancer cases treated with chemotherapy and at least one death due to breast cancer in a 15-year follow-up were analysed ($n = 522$, 53 events, Supplementary Table 3). Early-onset breast cancer patients from the independent Prospective Study of Outcomes in Sporadic versus Hereditary breast cancer (POSH) study[20–21] were used as a second replication data set. In particular, we performed our replication using ER-negative breast cancer patients treated with chemotherapy in the POSH study's Stage 1 discovery data set samples ($n = 315$, 108 events) selected to facilitate studies on breast cancer prognosis[22]. The breast cancer-specific death rate is thus particularly high and there were few cases that drop out due to lack of phenotype information.

All women in participating studies had provided written consent for the research and approval for each study was obtained from their local ethical review board (Supplementary Tables 1 and 3). Collection of blood samples and clinical data from subjects was performed in accordance with local guidelines and regulations.

Table 1 | Summary of results for association of rs4458204_A with risk of dying from breast cancer.

Patients	n	Breast cancer-specific deaths	Per-allele HR (95% CI)*	P-value
Discovery				
ER-negative	2,218	332	1.83 (1.47–2.27)	4.68×10^{-8}
ER-negative not treated with chemotherapy	411	53	1.39 (0.69–2.81)	0.36
ER-negative and treated with chemotherapy	1,804	279	1.96 (1.55–2.47)	1.60×10^{-8}
			$I^2 = 0\%$; Phet = 0.84	
Replication				
ER-negative and treated with chemotherapy				
iCOGS Asian studies	522	53	1.97 (0.94–4.17)	0.07
POSH	315	108	1.41 (0.95–2.09)	0.08
Combined replication				
ER-negative and treated with chemotherapy	837	161	1.52 (1.07–2.15)	0.02
			$I^2 = 0\%$; Phet = 0.44	
Combined overall				
ER-negative and treated with chemotherapy	2641	440	1.81 (1.49–2.19)	1.90×10^{-9}
			$I^2 = 1.4\%$; Phet = 0.36	

CI, confidence interval; COGS, Collaborative Oncological Gene-environment Study; ER, oestrogen receptor; HR, hazard ratio; I^2, I^2 metric; Phet, P for heterogeneity; POSH, Prospective Study of Outcomes in Sporadic versus Hereditary breast cancer.
*Fifteen-year breast cancer-specific survival, delayed-entry Cox proportional hazards model stratified by study and adjusted for population stratification, age at diagnosis, tumour size, presence of distant metastasis, lymph node status, tumour grade as well as surgery, chemotherapy, hormone therapy and radiotherapy.
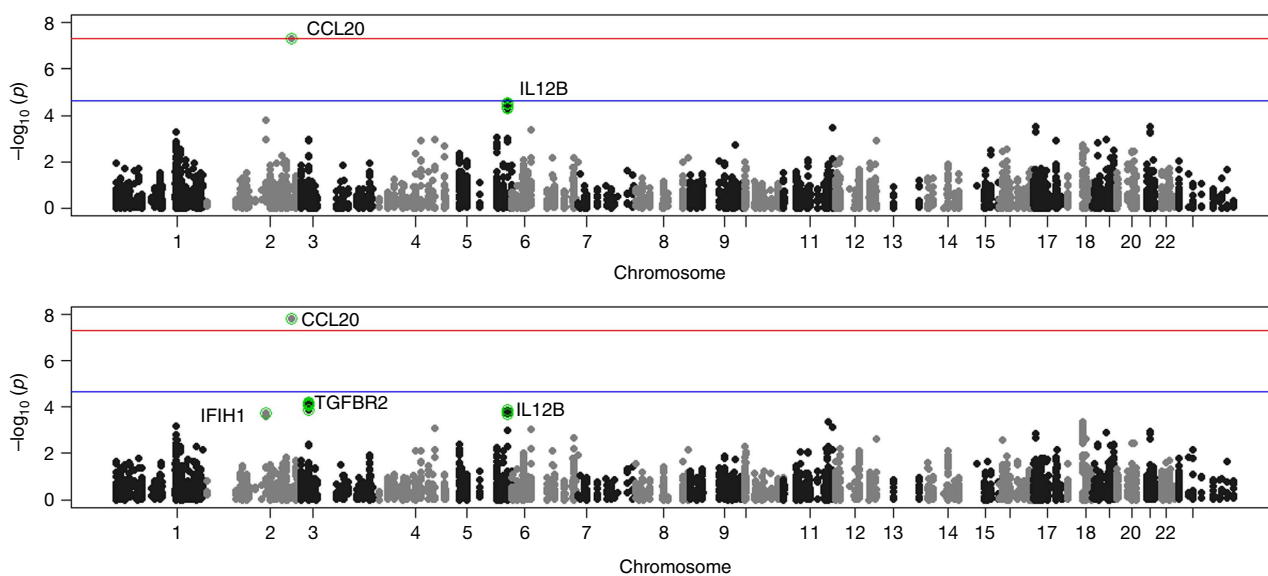
Figure 2 | Manhattan plot for association in the discovery stage. Manhattan plot showing directly genotyped SNPs plotted according to chromosomal location (x axis), with $-\log10$ P-values (y axis) derived from trend tests of association of 7,020 human immunology and inflammation SNPs with the risk of dying from breast cancer for all ER-negative patients (above) and ER-negative patients treated with chemotherapy (below) in the discovery phase. Blue and red lines indicate the Bonferroni threshold for multiple-testing correction for 2,184 ($r^2 < 0.2$) and genome-wide significance level (5×10^{-8}), respectively. SNPs with FDRs of <10% are additionally encircled and denoted in green. Chromosomal positions are based on NCBI build 36.

Genotyping was conducted using a custom Illumina iSelect genotyping array (iCOGS), comprising 211,115 SNPs. Details of quality control of the iCOGS data are described in detail elsewhere[18]. Briefly, individuals were excluded for any of the following reasons: genotypically not female XX (XY, XXY or XO), overall call rate <95%, low or high heterozygosity ($P < 1 \times 10^{-6}$, determined separately for individuals of European and East Asian ancestry), genotypes discordant with those determined in previous genotyping such that the individual appeared to be different, genotypes for the duplicate sample that seemed to be from a different individual and cryptic duplicates. SNPs with call rates of <95%, SNPs that deviated from Hardy–Weinberg equilibrium in controls at $P < 1 \times 10^{-7}$ and SNPs for which the genotypes were discrepant in >2% of duplicate samples across all COGS consortia were excluded. The final analyses in the parent COGS study were based on data from 199,961 SNPs.

Key genes related to human immunology and inflammation were identified from two comprehensive and highly curated gene panels (nCounter GX Human Immunology Kit and nCounter GX Human Inflammation Kit, NanoString Technologies, Seattle, WA, USA), which are commercially available (Supplementary Data 1). We identified all SNPs on the iCOGS within a 50-kb window of any gene on the panel. Out of 8,237 unique SNPs extracted from COGS, we further removed SNPs with low minor allele frequency
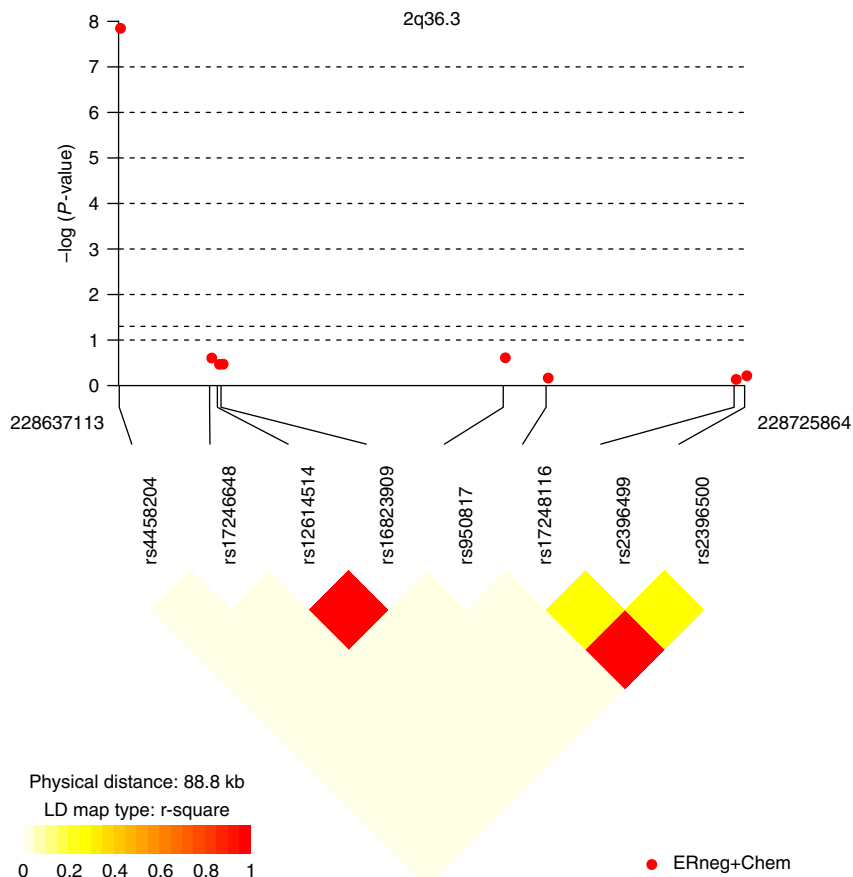
Figure 3 | Linkage disequilibrium plot of SNPs within a 100-kb window flanking rs4458204 in the discovery phase. The closest SNP flanking the left of rs4458204 is > 9.5 Mb away. Chromosomal positions are based on NCBI build 36. *P*-values are derived from trend tests of association. Plotted using 'snp.plotter' package in R.
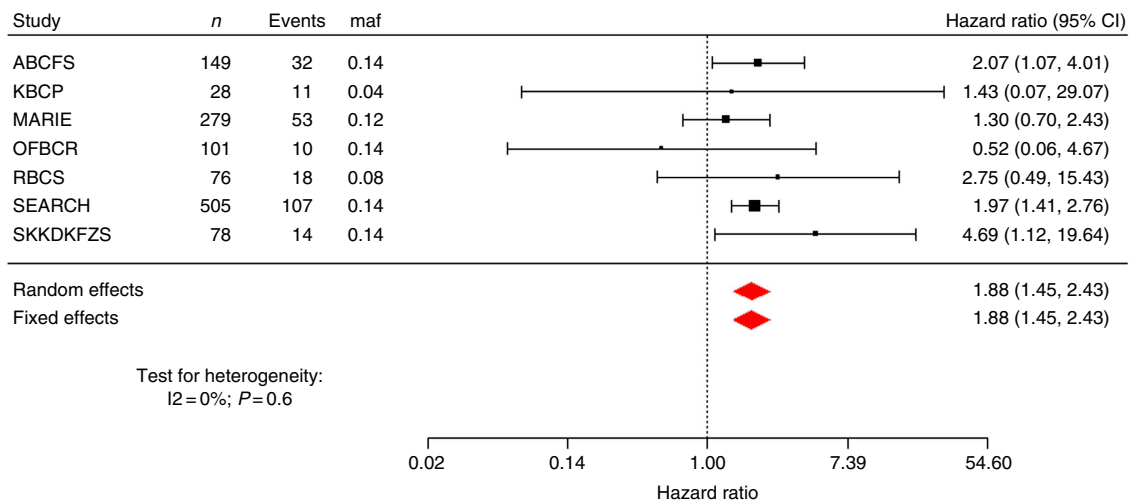

Figure 4 | Forest plot of a subset of studies in the discovery phase with at least ten events for rs4458204_A annotated to the *CCL20* gene. We found no evidence of heterogeneity in the per-allele HR across 14 studies ($I^2 = 0\%$, *P* for heterogeneity = 0.84). *P*-value for both fixed and random effects meta-analyses on all 14 studies was 3.93×10^{-7}, whereas on this reduced data set (studies with <10 events excluded for clarity of presentation) it was 1.76×10^{-6}, which passes the preset Bonferroni threshold of 2.29×10^{-5} for 2,184 independent tests. The 95% confidence interval for each study is given by a horizontal line, and the point estimate is given by a square whose height is inversely proportional to the s.e. of the estimate. The summary odds ratio is drawn as a diamond with horizontal limits at the confidence limits and width inversely proportional to itss.e.

(<0.05) and low call rate (<0.95). After quality-control exclusions, we analysed 7,020 non-overlapping SNPs in 557 unique gene regions (from 597 genes on the original nCounter panels).

In the POSH study, rs4458204 was genotyped on the Illumina 660 W-Quad SNP array. Details can be found in the parent POSH article[22]. Briefly, genotyping for the samples was conducted in
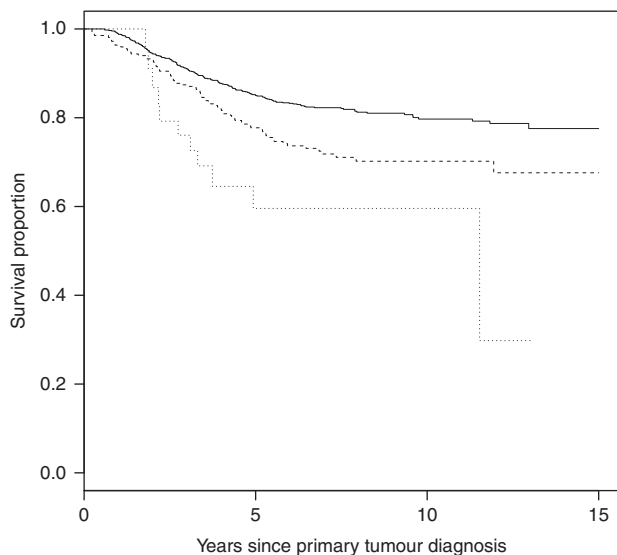
Figure 5 | Kaplan-Meier survival curves of breast cancer-specific survival in estrogen receptor-negative patients treated with chemotherapy for rs4458204 in the discovery phase. Analysis were adjusted only for time of blood draw and stratified by genotype. The p-value shown is based on the log-rank test. The number of events/n for each genotype are in parenthesis as follows rs4458204_GG (195/1415, single continuous line), rs4458204_AG (73/357, broken line) and rs4458204_AA (11/32, dotted line). The log rank P-value for this analysis was 3.18×10^{-6}.

two separate batches in two locations (Mayo Clinic and the Genome Institute of Singapore). To ensure harmonization of the genotype calling, the intensity data were combined and used to generate genotypes based on the algorithm available in the genotyping module of Illumina's Genome Studio software.

Breast cancer survival, right-truncated at 15 years after diagnosis, was modelled by using multivariate Cox proportional hazard analyses, treating each SNP as an ordinal variable (that is, 0, 1 and 2 copies of minor allele). Analyses were partially adjusted for age at diagnosis (years), study and seven principal components (as recommended by COGS) as covariates. As comparisons of survival are often confounded by differences in the patients, their tumours or the treatments, we further included covariates on tumour characteristics and treatment in a fully adjusted model, which is presented as the main analysis in this study. The fully adjusted model was additionally adjusted for tumour size (≤ 2, > 2 and ≤ 5, or > 5 cm), presence of distant metastasis (M from the Tumour, Nodes and Metastasis (TNM) staging system), lymph node status (negative/positive), histopathological grade (well, moderately or poorly differentiated), surgery (no surgery, breast-saving or mastectomy with or without axillary), hormone therapy (Yes/No) and radiotherapy (no radiation, breast only, breast and lymph nodes or lymph nodes only). Missing values were coded separately as missing. Separate baseline hazard functions were fitted for each study. Between-study heterogeneity was evaluated by using the Q statistic and the I^2 metric[23]. Estimated HRs and confidence limits are presented for heterozygotes and minor allele homozygotes, relative to the major allele homozygotes. Delayed entry (left truncation) was allowed for all models to adjust for the timing of blood draw. The proportional hazards assumption for each SNP was assessed using Schoenfeld's test statistics[24]. The Kaplan–Meier estimator for delayed-entry data was computed using the survfit function from the survival package in R. The Nagelkerke pseudo R-squared statistic was used to assess variance explained[25].

To adjust for multiple testing without overly penalizing the tests, we determined the number of 'independent' SNPs.

SNPs were thinned using the '—indep-pairwise' option in PLINK[26] such that all SNPs within a window size of 50 SNPs (step size of 10) were required to have $r^2 < 0.2$. This procedure resulted in a set of 2,184 independent SNPs pruned by linkage disequilibrium. The Bonferroni-adjusted threshold for 2,184 independent tests is 2.29×10^{-5}. In addition to standard Bonferroni adjustment, a 10% false discovery rate (FDR) threshold was applied to try to identify more candidate SNPs associated with breast cancer outcome. An FDR-adjusted P-value of 0.10 implies that 10% of significant tests will result in false positives.

The results for tests of association between 7,020 human immunology and inflammation SNPs and risk of death from ER-negative breast cancer are summarized in Supplementary Data 2 and 3. The deviation of the smaller observed P-values from those expected ($\lambda = 1.16$) is consistent with multiple weak associations between these SNPs and survival for ER-negative breast cancer patients (Fig. 1). In particular, for a single SNP rs4458204_A located on chromosome 2:228637113 (minor allele frequency = 0.12), the χ^2 (1df) association test statistic was much higher than for the other SNPs and was close to surpassing the threshold for experiment-wide significance after Bonferroni adjustment ($P < 2.29 \times 10^{-5}$) in the partially adjusted analysis stratified by study and adjusted for only population stratification and age ($n = 2,218$, 332 events, per-allele HR = 1.54 (1.26–1.90), P for trend = 3.62×10^{-5}, Supplementary Data 3). However, after further adjusting for appropriate patient tumour and treatment characteristics, the SNP association surpassed the threshold for genome-wide significance ($P < 5 \times 10^{-8}$) (per-allele HR = 1.83 (1.47–2.27), P for trend = 4.68×10^{-8}, Table 1 and Fig. 2), a conservative threshold which is likely to be overly stringent[27]. The lack of an association signal tower could be because the iCOGS was designed to have minimum linkage disequilibrium across SNPs. No SNP within a 100-kb window is correlated to rs4458204 with r2 > 0.2 (Fig. 3). The association was stronger for a subset of ER-negative patients who had been treated with chemotherapy ($n = 1,804$, 279 events, per-allele HR = 1.96 (1.55–2.47), P for trend = 1.60×10^{-8}). We found no evidence of heterogeneity in the per-allele HR across 14 studies ($I^2 = 0\%$, P for heterogeneity = 0.84; forest plot in Fig. 4). Univariate Kaplan–Meier survival curves of breast cancer-specific survival for ER-negative patients treated with chemotherapy by rs4458204 genotypes are presented in Fig. 5 (log-rank $P = 3.18 \times 10^{-6}$). The median survival time for the AA genotype at rs4458204 was 11.5 years. SNPs in three other loci corresponding to regions around the transforming growth factor beta receptor II (TGFBR2), interleukin 12B (IL12B) and interferon induced with helicase C domain 1 (IFIH1) genes were found to be associated with breast cancer-specific death with FDR-adjusted $P < 0.10$ (Fig. 2).

From our replication study of rs4458204_A using multi-ethnic iCOGS Asian samples (522 ER-negative patients treated with chemotherapy, 53 events; see Supplementary Table 3), the per-allele HR after controlling for tumour characteristics and treatment was 1.97 (0.94–4.17); P for trend = 0.07, Table 1). Together with multivariable-adjusted results from a second replication of the SNP using early-onset breast cancer patients POSH study, significant evidence of replication was observed (combined per-allele HR = 1.52 (1.07–2.15), P for trend = 0.02, Table 1). From a meta-analysis of both discovery and replication stages, the association of the SNP with risk of dying from breast cancer was found to be 1.81 (1.49–2.19; P for trend = 1.90×10^{-9}) with no observed heterogeneity ($I^2 = 1.4\%$, P for heterogeneity = 0.36; Table 1).

The cluster plots for the most significant SNP in our analysis, rs4458204 (CCL20), and three other index SNPs of loci for which
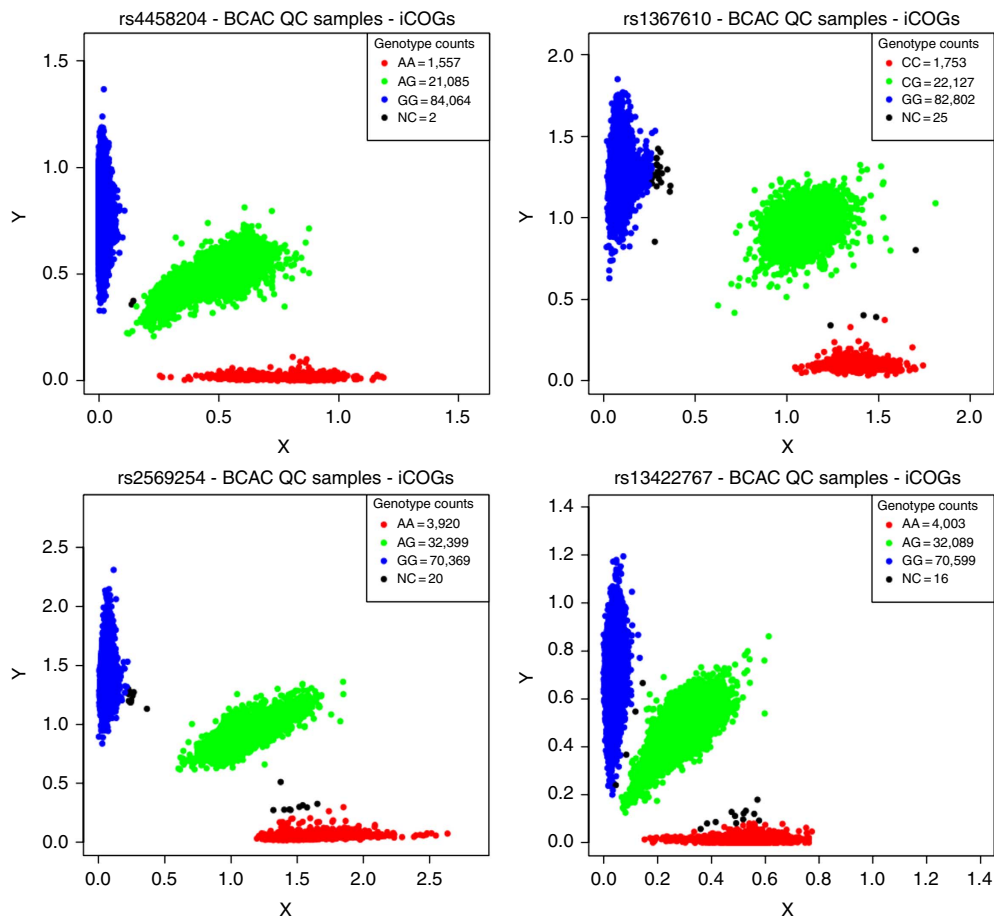
Figure 6 | Cluster plots of noteworthy SNPs. Cluster plots are shown for rs4458204 (*CCL20*), rs1367610 (*TGFBR2*), rs2569254 (*IL12B*) and rs13422767 (*IFIH1*) for the BCAC samples that passed quality control in the parent COGS study[18]. The SNP genotypes have been assigned based on cluster formation in scatter plots of normalized allele intensities X and Y. Each circle represents one individual's genotype. Blue and red clouds indicate homozygote genotypes for the SNP (AA/aa), green heterozygote (Aa) and black undetermined. Three distinct, tight clusters exhibited by all four representative SNPs indicate good discrimination of the three genotypes.

the associated test statistic passed FDR < 0.1, namely rs1367610 (*TGFBR2*), rs2569254 (*IL12B*) and rs13422767 (*IFIH1*), were examined. All SNPs showed good discrimination of the three genotypes in cluster plots for the BCAC samples that passed quality control in the parent COGS study (Fig. 6).

Discussion

rs4458204 is located ∼41.5 kb upstream of the chemokine (C–C motif) ligand 20 (*CCL20*) gene. Chemokines are important mediators of immune response, and *CCL20* has previously been shown to induce migration and proliferation of breast epithelial cells[28]. *CCL20* has also been reported to be strongly chemotactic for lymphocytes and weakly attracts neutrophils[29]. However, rs4458204 was not found to be a significant (*P* for trend > 0.05) expression trait quantitative locus in any of the tissues (that is, adipose subcutaneous, artery tibial, blood, heart, lung, muscle skeletal, nerve tibial, skin and thyroid) reported on the publicly available Genotype-Tissue Expression Portal[30].

It is of note that the association of rs4458204_A with the survival of ER-negative breast cancer patients treated with chemotherapy increased and the strength of the association became stronger after adjustment for tumour characteristics and type of treatment (per-allele HR (95% confidence interval) from 1.64 (1.31–2.05) to 1.96 (1.55–2.47), *P* for trend from 1.27×10^{-5} to 1.60×10^{-8}). This suggests that tumour

characteristics and treatment covariates are likely to be confounders and thus it is desirable to include them in the fully adjusted model to obtain a more accurate effect size of the genetic factor. Moreover, it has also been shown that adjustment for prognostic factors will lead to a gain in power for statistical analyses. Genes in other regions indentified by the less stringent FDR threshold (*TGFBR2*, *IL12B* and *IFIH1*) have been implicated to play a role in breast cancer disease progression, suggesting that there are potentially more variants in immune response and inflammation genes that are associated with breast cancer prognosis. Although *TGFBR2* is a breast cancer susceptibility locus[18], none of the SNPs annotated to this gene was significantly associated with breast cancer risk (*P* > 0.05) in the parent COGS study.

Although several GWAS have aimed to find genetic markers associated with breast cancer survival to date[22,31–33], few credible variants have been robustly identified. The threefold greater breast cancer mortality for affected sisters is comparable in magnitude to the familial relative risk for breast cancer incidence, for which close to 100 independent susceptibility loci based on common variants (SNPs) have been identified, and these explain only a small proportion of familial aggregation of risk[18]. The failure to identify a similar number of survival-associated loci influencing survival may reflect the much lower statistical power for survival analyses to date, but may also reflect the substantial heterogeneity in tumour characteristics and treatment. As such, it

has been suggested that sufficiently powered studies investigating specific cancer subtypes or treatment subgroups would need to be much larger to discover more regions in the genome associated with breast cancer prognosis[33]. In agreement, the association between rs4458204 and breast cancer survival for this study was found to be more pronounced (larger HR) for women with ER-negative disease treated with chemotherapy (Table 1). However, as we did not study the association for women with ER-positive disease, the impact of this SNP on survival for those women remains unclear. One of the strengths of our study is that we have based our gene selection on commercially pre-designed panels of genes known to be differentially expressed in immunology and inflammation, which covers a comprehensive and validated list of relevant genes. The use of the iCOGS array in the BCAC consortium allowed us to investigate genetic variation across >500 immune response genes and provided an unprecedented large sample size with detailed clinical information to examine their associations with breast cancer survival. The results were also replicated by the POSH study, which is not part of the COGS consortium. However, SNPs related to immune response and inflammation were not specifically selected to be put on the iCOGS panel to give comprehensive coverage of these genes; only 557 of the 597 genes ($\sim$93%) were represented. The proportion of total phenotypic variance (Nagelkerke pseudo R-squared) explained by this SNP alone was also small, at $\sim$1.3%, suggesting that many more variants will need to be discovered for such genetic data to be useful in a clinical setting.

Our findings suggest that host factors affecting the ability to respond to systemic treatment or to mount an effective immunologic response contribute to the heritability of prognosis. Such survival-associated variants can represent ideal targets for tailored therapeutics and may also enhance our current prognostic prediction capabilities.

References

1. Lindstrom, L. S. et al. Familial concordance in cancer survival: a Swedish population-based study. Lancet Oncol. 8, 1001–1006 (2007).
2. Hartman, M. et al. Is breast cancer prognosis inherited? Breast Cancer Res. 9, R39 (2007).
3. Lindström, L. et al. Association of breast cancer survival among sisters is not explained by shared primary tumour characteristics. Submitted (2013).
4. Verkooijen, H. M. et al. Breast cancer prognosis is inherited independently of patient, tumor and treatment characteristics. Int. J. Cancer 130, 2103–2110 (2012).
5. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat. Rev. Cancer 12, 68–78 (2012).
6. Honrado, E., Osorio, A., Palacios, J. & Benitez, J. Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations. Oncogene 25, 5837–5845 (2006).
7. Heikkinen, T. et al. The breast cancer susceptibility mutation PALB2 1592delT is associated with an aggressive tumor phenotype. Clin. Cancer Res. 15, 3214–3222 (2009).
8. Teo, Z. L. et al. Tumour morphology predicts PALB2 germline mutation status. Br. J. Cancer 109, 154–163 (2013).
9. Stevens, K. N. et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. Cancer Res. 71, 6240–6249 (2011).
10. Antoniou, A. C. et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat. Genet. 42, 885–892 (2010).
11. Haiman, C. A. et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat. Genet. 43, 1210–1214 (2011).
12. Broeks, A. et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum. Mol. Genet. 20, 3289–3303 (2011).
13. Stacey, S. N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 39, 865–869 (2007).
14. Stacey, S. N. et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 40, 703–706 (2008).
15. Denkert, C. et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. 28, 105–113 (2010).
16. Apetoh, L. et al. The interaction between HMGB1 and TLR4 dictates the outcome of anticancer chemotherapy and radiotherapy. Immunol. Rev. 220, 47–59 (2007).
17. Zitvogel, L. & Kroemer, G. The immune response against dying tumor cells: avoid disaster, achieve cure. Cell Death Differ. 15, 1–2 (2008).
18. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45, 353–361 (2013).
19. Loi, S. et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. J. Clin. Oncol. 31, 860–867 (2013).
20. Eccles, D. et al. Prospective study of Outcomes in Sporadic versus Hereditary breast cancer (POSH): study protocol. BMC Cancer 7, 160 (2007).
21. Copson, E. et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study. J. Natl Cancer Inst. 105, 978–988 (2013).
22. Rafiq, S. et al. Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. Cancer Res. 73, 1883–1891 (2013).
23. Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F. & Botella, J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol. Methods 11, 193–206 (2006).
24. Schoenfeld, D. Partial residuals for the proportional hazards regression model. Biometrika 69, 239–241 (1982).
25. Nagelkerke, N. J. D. A note on a general definition of the coefficient of determination. Biometrika 78, 691–692 (1991).
26. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
27. Milestone in Anhui. Nat. Genet. 43, 613 (2011).
28. Marsigliante, S., Vetrugno, C. & Muscella, A. CCL20 induces migration and proliferation on breast epithelial cells. J. Cell. Physiol. 228, 1873–1883 (2013).
29. Hieshima, K. et al. Molecular cloning of a novel human CC chemokine liver and activation-regulated chemokine (LARC) expressed in liver. Chemotactic activity for lymphocytes and gene localization on chromosome 2. J. Biol. Chem. 272, 5846–5853 (1997).
30. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
31. Shu, X. O. et al. Novel genetic markers of breast cancer survival identified by a genome-wide association study. Cancer Res. 72, 1182–1189 (2012).
32. Kiyotani, K. et al. A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. Hum. Mol. Genet. 21, 1665–1672 (2012).
33. Azzato, E. M. et al. A genome-wide association study of prognosis in breast cancer. Cancer Epidemiol. Biomarkers Prev. 19, 1140–1143 (2010).

Author contributions

Additional information

Jia N. Foo[1], Sajjad Rafiq[5], Marjanka K. Schmidt[6,7], Paul D.P. Pharoah[8,9], Kyriaki Michailidou[8], Joe Dennis[8], Manjeet K. Bolla[8], Qin Wang[8], Laura J. Van 't Veer[7], Sten Cornelissen[7], Emiel Rutgers[10], Melissa C. Southey[11], Carmel Apicella[12], Gillian S. Dite[12], John L. Hopper[12,13], Peter A. Fasching[14,15], Lothar Haeberle[14], Arif B. Ekici[16], Matthias W. Beckmann[14], Carl Blomqvist[17], Taru A. Muranen[18], Kristiina Aittomäki[19], Annika Lindblom[20], Sara Margolin[21], Arto Mannermaa[22,23,24], Veli-Matti Kosma[22,23,24], Jaana M. Hartikainen[22,23,24], Vesa Kataja[22,24,25], Georgia Chenevix-Trench[26], kConFab Investigators[27], Kelly-Anne Phillips[12,28], Sue-Anne McLachlan[29], Diether Lambrechts[30,31], Bernard Thienpont[30,31], Ann Smeets[32], Hans Wildiers[32], Jenny Chang-Claude[33], Dieter Flesch-Janys[33,34,35], Petra Seibold[33], Anja Rudolph[33], Graham G. Giles[12,36], Laura Baglietto[12,36], Gianluca Severi[12,36], Christopher A. Haiman[37], Brian E. Henderson[38], Fredrick Schumacher[37], Loic Le Marchand[38], Vessela Kristensen[39,40,41], Grethe I. Grenaker Alnæs[39,40], Anne-Lise Borresen-Dale[39,40], Silje Nord[39,40], Robert Winqvist[42], Katri Pylkäs[42], Arja Jukkola-Vuorinen[43], Mervi Grip[43], Irene L. Andrulis[44,45], Julia A. Knight[46,47], Gord Glendon[44], Sandrine Tchatchou[48], Peter Devilee[49,50], Robert Tollenaar[51], Caroline Seynaeve[52], Maartje Hooning[52], Mieke Kriege[52], Antoinette Hollestelle[52], Ans van den Ouweland[53], Yi Li[1],

Ute Hamann[54], Diana Torres[54,55], Hans U. Ulmer[56], Thomas Rüdiger[57], Chen-Yang Shen[58,59], Chia-Ni Hsiung[58], Pei-Ei Wu[60], Shou-Tung Chen[61], Soo Hwang Teo[62,63], Nur Aishah Mohd Taib[63], Cheng Har Yip[63], Gwo Fuang Ho[64], Keitaro Matsuo[65], Hidemi Ito[66], Hiroji Iwata[67], Kazuo Tajima[66], Daehee Kang[68,69], Ji-Yeob Choi[69], Sue K. Park[68,69], Keun-Young Yoo[70], Tom Maishman[5], William J. Tapper[5], Alison Dunning[9], Mitul Shah[9], Robert Luben[8], Judith Brown[8], Chiea Chuen Khor[1], Diana M. Eccles[5], Heli Nevanlinna[18], Douglas Easton[8,9], Keith Humphreys[71], Jianjun Liu[1], Per Hall[71] & Kamila Czene[71]

[1] Human Genetics, Genome Institute of Singapore, 60 Biopolis Street 02-01, Singapore 138672, Singapore. [2] Saw Swee Hock School of Public Health, National University of Singapore, MD3, 16 Medical Drive, Singapore 117597, Singapore. [3] Department of Surgery, University of California at San Francisco (UCSF), San Francisco, California 94143, USA. [4] Department of Biosciences and Nutrition, Karolinska Institutet and University Hospital, SE-171 77 Stockholm, Sweden. [5] Faculty of Medicine, University of Southampton, University Hospital Southampton, Southampton SO16 6YD, UK. [6] Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Postbus 90203, 1006 BE Amsterdam, The Netherlands. [7] Division of Molecular Pathology, Netherlands Cancer Institute, Postbus 90203, 1006 BE Amsterdam, The Netherlands. [8] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [9] Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Worts Causeway, Cambridge CB1 8RN, UK. [10] Department of Surgery, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [11] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, 5th floor, West Wing, Medical Building 181, Melbourne, Victoria 3010, Australia. [12] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Level 1, 723 Swanston Street, Melbourne, Victoria 3010, Australia. [13] Department of Health Science and Services, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea. [14] University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen–Nuremberg, Universitätsstrasse 21-23, 91054 Erlangen, Germany. [15] Division of Hematology and Oncology, Department of Medicine, David Geffen School of Medicine, University of California at Los Angeles, 10833 Le Conte Avenue, Los Angeles, California 90095, USA. [16] Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Schlossplatz 4, 91054 Erlangen, Germany. [17] Department of Oncology, University of Helsinki and Helsinki University Central Hospital, FI-00029 HUS Helsinki, Finland. [18] Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Haartmaninkatu 8, FI-00029 Helsinki, Finland. [19] Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Haartmaninkatu 8, FI-00029 Helsinki, Finland. [20] Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm SE-17177, Sweden. [21] Department of Oncology-Pathology, Karolinska Institutet, Stockholm SE-17177, Sweden. [22] School of Medicine, Institute of Clinical Medicine, Oncology, University of Eastern Finland, Yliopistonranta 1C, FI-70211 Kuopio, Finland. [23] Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, FI-70211 Kuopio, Finland. [24] Cancer Center, Kuopio University Hospital, Puijonlaaksontie 2, 70210 Kuopio, Finland. [25] Jyväskylä Central Hospital, Keskussairaalantie 19, 40620 Jyväskylä, Finland. [26] Department of Genetics, QIMR Berghofer Medical Research Institute, 300 Herston Road, Brisbane, Queensland 4006, Australia. [27] Research Department, Peter MacCallum Cancer Center, East Melbourne, Victoria 3002, Australia. [28] Sir Peter MacCallum Department of Oncology, University of Melbourne, St Andrews Place East, East Melbourne, Victoria 3002, Australia. [29] Department of Oncology, St Vincent's Hospital Melbourne, 9 Princes Street, Fitzroy, Victoria 3065, Australia. [30] Vesalius Research Center (VRC), VIB, Herestraat 49, 3000 Leuven, Belgium. [31] Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Herestraat 49, 3000 Leuven, Belgium. [32] Multidisciplinary Breast Center, University Hospitals Leuven, Herestraat 49, 3000 Leuven, Belgium. [33] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [34] Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. [35] Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Martinistrasse 52, 20246 Hamburg, Germany. [36] Cancer Epidemiology Centre, Cancer Council Victoria, 1 Rathdowne Street, Melbourne, Victoria 3053, Australia. [37] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 1975 San Pablo Street, Los Angeles, California 90033, USA. [38] University of Hawaii Cancer Center, 701 Ilalo Street, Honolulu, Hawaii 96813, USA. [39] Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, Ullernchausseen 70, N-0310 Oslo, Norway. [40] K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Kirkeveien 166, 0450 Oslo, Norway. [41] Department of Clinical Molecular Biology (EpiGen), Medical Division, Akershus University Hospital, Sykehusveien 25, 1478 Lørenskog, Norway. [42] Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu University Hospital, Aapistie 5A, FI-90220 Oulu, Finland. [43] Laboratory of Oncology, Oulu University Hospital, University of Oulu, Kajaanintie 50, FI-90220 Oulu, Finland. [44] Ontario Cancer Genetics Network, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1 × 5. [45] Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario, Canada M5S 1A8. [46] Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1 × 5. [47] Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, Ontario, Canada M5T 3M7. [48] Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1 × 5. [49] Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. [50] Department of Pathology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. [51] Department of Surgical Oncology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC Leiden, The Netherlands. [52] Department of Medical Oncology, Erasmus MC Cancer Institute, Groene Hilledijk 301, 3075EA Rotterdam, The Netherlands. [53] Department of Clinical Genetics, Erasmus University Medical Center, Groene Hilledijk 301, 3075EA Rotterdam, The Netherlands. [54] Molecular Genetics of Breast Cancer, DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [55] Institute of Human Genetics, Pontificia University Javeriana, Carrera 7, Bogotá 11001000, Colombia. [56] Frauenklinik der Stadtklinik Baden-Baden, D-7570 Baden-Baden, Germany. [57] Institute of Pathology, Städtisches Klinikum Karlsruhe, Moltkestrasse 90, 76133 Karlsruhe, Germany. [58] Institute of Biomedical Sciences, Academia Sinica, Academia Road Nankang, Taipei 115, Taiwan. [59] College of Public Health, China Medical University, No. 91, Hsueh-Shih Road, Taichung 40402, Taiwan. [60] Taiwan Biobank, Academia Sinica, Academia Road Nankang, Taipei 115, Taiwan. [61] Department of Surgery, Changhua Christian Hospital, No. 135, Nanhsiao Street, Changhua 500, Taiwan. [62] Cancer Research Initiatives Foundation, Sime Darby Medical Centre, 1 Jalan SS 12/1A, Subang Jaya47500, Selangor, Malaysia. [63] Breast Cancer Research Unit, Faculty of Medicine, University Malaya Cancer Research Institute, University Malaya, Lembah Pantai, 59100 Kuala Lumpur, Malaysia. [64] Department of Oncology, Faculty of Medicine, University Malaya, Lembah Pantai, 59100 Kuala Lumpur, Malaysia. [65] Department of Preventive Medicine, Faculty of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan. [66] Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, 1-1 Kanokoden Chikusa-ku, Nagoya 464-8681, Aichi, Japan. [67] Department of Breast Oncology, Aichi Cancer Center Hospital, 1-1 Kanokoden Chikusa-ku, Nagoya 464-8681, Aichi, Japan. [68] Department of Preventive Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea. [69] Department of Biomedical Sciences, Seoul National University Graduate School, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea. [70] Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea. [71] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, 171 77 Stockholm, Sweden. * These authors contributed equally to this work.