An investigation of the association of genetic susceptibility risk with somatic mutation burden in breast cancer

Bin Zhu*,[1], Anwesha Mukherjee[2], Mitchell J Machiela[1], Lei Song[1], Xing Hua[1], Jianxin Shi[1], Montserrat Garcia-Closas[1], Stephen J Chanock[1] and Nilanjan Chatterjee*,[1,3]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institute of Health, Bethesda, MD 20892, USA; [2]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA and [3]Department of Biostatistics and Department of Oncology, Johns Hopkins University, Baltimore, MD 21205, USA

Background: Genome-wide association studies have reported nearly 100 common germline susceptibility loci associated with the risk for breast cancer. Tumour sequencing studies have characterised somatic mutation profiles in breast cancer patients. The relationship between breast cancer susceptibility loci and somatic mutation patterns in breast cancer remains largely unexplored.

Methods: We used single-nucleotide polymorphism (SNP) genotyping array data and tumour exome sequencing data available from 638 breast cancer patients of European ancestry from The Cancer Genome Atlas (TCGA) project. We analysed both genotype data and, when necessary, imputed genotypes for 90 known breast cancer susceptibility loci. We performed linear regression models to investigate possible associations between germline risk variants with total somatic mutation count (TSMC), as well as specific mutation types. We examined individual SNP genotypes, as well as a multi-SNP polygenic risk score (PRS). Models were statistically adjusted for age at diagnosis, stage, oestrogen-receptor (ER) and progesterone-receptor (PR) status of breast cancer. We also performed stratified analyses by ER and PR status.

Results: We observed a significant inverse association ($P = 8.75 \times 10^{-6}$; FDR = 0.001) between the risk allele in rs2588809 of the gene *RAD51B* and TSMC across all breast cancer patients, for both ER$^+$ and ER$^-$ tumours. This association was also evident for different types of mutations. The PRS analysis for all patients, with or without rs2588809, showed a significant inverse association ($P = 0.01$ and 0.04, respectively) with TSMC. This inverse association was significant in ER$^+$ patients with the ER$^+$-specific PRS ($P = 0.02$), but not among ER$^-$ patients for the ER$^-$-specific PRS ($P = 0.39$).

Conclusions: We observed an inverse association between common germline risk variants and TSMC, which, if confirmed, could provide new insights into how germline variation informs our understanding of somatic mutation patterns in breast cancer.

Breast cancer is a complex disease, in which the characteristics of germline susceptibility loci as well as the spectrum of somatic alterations have begun to emerge, largely due to the capacity to conduct large-scale genome-wide studies. Genetic susceptibility loci implicated in breast cancer include highly penetrant rare variants in genes such as *BRCA1* and *BRCA2,* moderately penetrant low-frequency variants in genes such as *ATM, PALB2* and *CHEK2,* and multiple low-penetrant common variants

identified more recently through genome-wide association studies (GWAS; Mavaddat *et al,* 2010). Specifically, GWAS have reported nearly 100 common susceptibility loci for breast cancer, marked by common single-nucleotide polymorphisms (SNPs; Long *et al,* 2012; Siddiq *et al,* 2012; Bojesen *et al,* 2013; Garcia-Closas *et al,* 2013; Michailidou *et al,* 2013; Cai *et al,* 2014; Milne *et al,* 2014; Michailidou *et al,* 2015). Although each SNP individually exhibits small effect size, in combination they can explain a substantial

proportion of the variation of the familial risk for breast cancer, as well as risk in the general population (Easton *et al*, 2015; Mavaddat *et al*, 2015; Michailidou *et al*, 2015).

In parallel, whole-exome and -genome tumour sequencing studies have been conducted to define the landscape of somatic mutations of breast cancer (Banerji *et al*, 2012; Curtis *et al*, 2012; Ellis *et al*, 2012; Koboldt *et al*, 2012; Nik-Zainal *et al*, 2012; Shah *et al*, 2012; Stephens *et al*, 2012; Nik-Zainal *et al*, 2016), particularly focusing on identifying candidate driver genes, namely the genes harbouring mutations that confer selective growth advantage. In addition, studies of somatic mutation patterns for breast cancer have identified distinct mutational signatures across the genome. The number of somatic mutations (or mutational burden) in coding exons varies widely between breast tumours, and it has been related to age at diagnosis and to tumour grade (Stephens *et al*, 2012). Although genetic susceptibility variants show distinct associations with different pathological subtypes of breast cancer, particularly those defined by ER status (Mavaddat *et al*, 2010; Siddiq *et al*, 2012; Garcia-Closas *et al*, 2013), little is known of the relationship between the inherited and somatic genetic components. A recent survey of cancer predisposition genes, defined as genes harbouring high- or moderate-penetrant risk variants, has suggested that a large fraction of these genes could be oncogenic when mutates somatically (Rahman, 2014). In contrast, a study has reported no evidence that cancer susceptibility regions that harbour common low-penetrant susceptibility are preferentially selected for altered somatic mutation frequencies in cancer patients (Machiela *et al*, 2015).

In this report, using the breast cancer data from The Cancer Genome Atlas (TCGA) study (Koboldt *et al*, 2012), we examined associations between established breast cancer susceptibility loci and exome-wide single-nucleotide substitution counts observed in tumour tissues. As mutations carrying the aetiologic signature of tumours are expected to be present across the whole genome, and not only in the specific genes, examination of association of cancer risk factors and total somatic mutation count (TSMC) of substitutions in the whole exome, might be a powerful approach to explore relationships between germline variants and somatic substitutions. For example, biologic age, which is the strongest risk factor for many cancers and is often a surrogate for cumulative carcinogenic events, has been shown to be directly associated with TSMC across many cancers (DePinho, 2000). For other major risk factors, such as smoking for lung cancer and sun exposure for melanoma, specific signature mutations, as hallmarks of exposure, are observed across the whole genome (Pfeifer *et al*, 2002; 2005).

METHODS

Data were extracted from germline genotypes generated using the Affymetrix Genome-Wide Human SNP Array 6.0 on circulating leucocyte DNA drawn from 638 breast cancer cases of European ancestry. To remove subjects who may not be Caucasians, cases were selected on the basis of principal component analysis, which combined common SNP genotypes with ones from HapMap (Altshuler *et al*, 2010b) reference samples (Supplementary Figure 1). We performed genotype imputation for restricted European ancestry samples using IMPUTE2 (Howie *et al*, 2009) with haplotypes generated by the 1000 Genome Project (Phase3; Altshuler *et al*, 2010a) as the reference.

Genotyped or imputed dosage data were available for 90 established SNPs representing common susceptibility loci with minor-allele frequencies (MAFs) >0.01. All loci had reported breast cancer risk associations below the threshold for genome-wide significance ($P<5\times10^{-8}$). We initially selected 94 SNPs from the study by Michailidou *et al* (2015) and removed two SNPs

(rs7726159 and rs2380205) that were not genome-wide significant in the study by Michailidou *et al* (2013), and one SNP that is rare (rs17879961, MAF = 0.0049). Except for one SNP not present in the 1000 Genomes Project reference panel, the remaining 90 breast cancer susceptibility SNPs passed quality filter with IMPUTE2 info score >0.8 (Supplementary Table 1).

Somatic mutation data were obtained from whole-exome sequencing of TCGA breast cancer tumour samples. Mutation counts were extracted from the Mutation Annotation File (version 2.1.1.0 curated) generated by the Washington University Genome Institute. Details about sample preparation, sequencing protocol, and mutation calling pipeline are described elsewhere (Koboldt *et al*, 2012). For clinical information, we retrieved age at diagnosis, oestrogen-receptor (ER) status, progesterone-receptor (PR) status and tumour stage.

We used the somatic mutation burden, overall as TSMC or by mutation-specific types, as the outcome variable to perform linear regression analysis of association with SNP genotypes, individually or collectively as a polygenic risk score (PRS). Analyses were adjusted for subject and tumour characteristics, including age at diagnosis, ER status, PR status and tumour stage. Subjects with extremely low or high numbers of TSMC (bottom 1% and top 3% of subjects, respectively) were excluded as outliers. The mutation counts were log_{10} transformed and results were presented after standardising each type of log-transformed mutation count to have unit standard deviation so that effect-sizes are comparable across TSMC of different mutation types. At each locus, the genotype was coded based on the number of risk alleles (0, 1 or 2). For the current sample size, it was estimated that the study has 80% power to detect effect size of 0.5-s.d. unit change in mutation count (in log_{10} scale) per copy of a SNP allele with a population frequency of 0.33 at 5% type I error. Power curves for additional effect size and risk allele frequency combinations are illustrated in Supplementary Figure 2.

For each subject, the PRS reflected the total genetic susceptibility burden based on the 90 independent SNPs, and was defined as the weighted combination of each SNP genotype with the weights defined by previously reported log-odds-ratio of association of the SNPs with breast cancer (Supplementary Table 1). We used log-odds-ratio estimates for overall breast cancer and subtypes defined by ER status reported in the study by Michailidou *et al* (2015) for all non-correlated SNPs, and estimates in the study by Mavaddat *et al* (2015) for three correlated SNPs with conditional independent signals in 11q13 (rs554219, rs75915166 and rs78540526). In addition, we examined mutation burden with the following specific types: mutations from thymine (or adenine on the other strand) to other nucleotides, from cytosine (or guanine on the other strand) to other nucleotides, transition mutations, transversion mutations and APOBEC-mediated mutations, defined as cytosine to thymine and cytosine to guanine substitutions in the TCW motifs (W is either adenine or thymine; Roberts *et al*, 2013).

We further considered the somatic copy-number burden in the analysis. The processed segments of copy-number variation (CNV) were downloaded (in November 2015). Following previous work (Laddha *et al*, 2014), we used the magnitude 0.2 as the threshold to identify amplifications and deletions, and required at least 10 markers included in the CNV segment. The total number of CNV segments across the genome was calculated and treated as a covariate in the linear regression model.

RESULTS

The age at diagnosis of 638 breast cancer patients ranged from 26 to 90 years old with a median of 59.5 years. We first examined each characteristic without adjustment for other characteristics

(Table 1). Specifically, TSMC was higher for patients with older age at diagnosis ($P = 4.02 \times 10^{-4}$ for age groups, Table 1; $P = 5.04 \times 10^{-4}$ for the trend of age, Supplementary Figure 3), low PRS ($P = 0.04$ for PRS groups, Table 1; $P = 0.01$ for the trend of PRS, Figure 1), negative vs positive ER status ($P = 1.74 \times 10^{-10}$),

negative vs positive PR status ($P = 2.87 \times 10^{-15}$), and late vs early stages ($P = 4.08 \times 10^{-3}$). In addition, TMSC were significantly associated with patient group defined by both ER and PR status ($P = 2.95 \times 10^{-14}$). In an analysis of all characteristics (PRS, age at diagnosis, ER status, PR status and stage) simultaneously fitted in a linear regression model (Supplementary Table 4), we observed that the TMSC was associated with age at diagnosis ($P = 2.3 \times 10^{-6}$), tumour stage ($P = 3.05 \times 10^{-3}$ and 1.68×10^{-3} for stage II vs stage I and for stage III/IV vs stage I, respectively) and PR status ($P = 3.96 \times 10^{-7}$), but not with ER status ($P = 0.18$). Further, in stratified analyses by mutation type, we observed that both ER^{+} and PR^{+} tumours were significantly associated with a lower mutation count of thymine to other nucleotides, particularly ER^{+} tumours; whereas PR^{+}, but not ER^{+}, tumours were associated with lower mutation count of cytosine to other nucleotides (comparisons of positive vs negative ER and PR status; Supplementary Tables 5 and 6).

In an analysis of association of the mutation count and individual breast cancer susceptibility SNPs, rs2588809 in *RAD51B* was inversely associated with TSMC ($P = 8.75 \times 10^{-6}$, Table 2) with *P*-value of 0.001 adjusted for multiple comparisons using the Benjamini–Hochberg false-discovery rate (FDR; Benjamini and Hochberg, 1995). Statistical significance of the association was evident across all types of mutations and breast cancer subtypes (Table 2). Two other SNPs, rs11814448 in *DNAJC1* and rs13387042, which localises to a gene-poor region of chromosome 2q35, also showed possible inverse associations (FDR = 0.25) with TSMC (Supplementary Table 1 for all subjects, Supplementary Table 2 for ER^{+} subjects and Supplementary Table 3 for ER^{-} subjects).

PRS for overall breast cancer was inversely associated with TSMC ($P = 1.34 \times 10^{-2}$, Figure 1) as well as for all different types of mutations (Table 3). We observed a significant trend ($P = 2.28 \times 10^{-3}$, Figure 1 right panel) in that the strength of association of the individual SNPs with TSMC (measured by the regression coefficient) tended to be larger for those with larger reported odds ratio of association with breast cancer risk, but this trend was

Table 1. Total in 638 breast tumours stratified by ER and PR status, PRS (by tertile), age at diagnosis and tumour stage

			Statistics for log₁₀ (TSMC)		
	Status	**N**	**Mean (± s.d.)**	**P-value[a]**	
All		638	1.57 (± 0.31)		
ER	−	122	1.71 (± 0.32)	1.74e-10	
	+	516	1.54 (± 0.31)		
PR	−	186	1.71 (± 0.33)	2.87e-15	
	+	452	1.51 (± 0.29)		
ER and PR	− −	110	1.71 (± 0.31)	2.95e-14	
	− +	12	1.66 (± 0.38)		
	+ −	76	1.71 (± 0.35)		
	+ +	440	1.51 (± 0.29)		
PRS[b,c]	<0.53	211	1.61 (± 0.31)	4.17e-02	
	0.53–0.94	210	1.58 (± 0.33)		
	≥0.94	217	1.52 (± 0.29)		
Age (years)	<45	80	1.52 (± 0.34)	4.02e-04	
	45–55	153	1.50 (± 0.32)		
	≥55	405	1.61 (± 0.30)		
Stage	I	118	1.49 (± 0.30)	4.08e-03	
	II	360	1.58 (± 0.32)		
	III/IV	160	1.60 (± 0.30)		

Abbreviations: ER = oestrogen receptor; PR = progestrogen receptor; PRS = polygenic risk score; TSMC = total somatic mutation count.
[a]By Kruskal–Wallis test without adjustment for other characteristics.
[b]Grouped by tertiles of PRS, which is not standardised.
[c]PRS is based on 90 SNPs with reported breast cancer overall odds ratios (Supplementary Table 1).
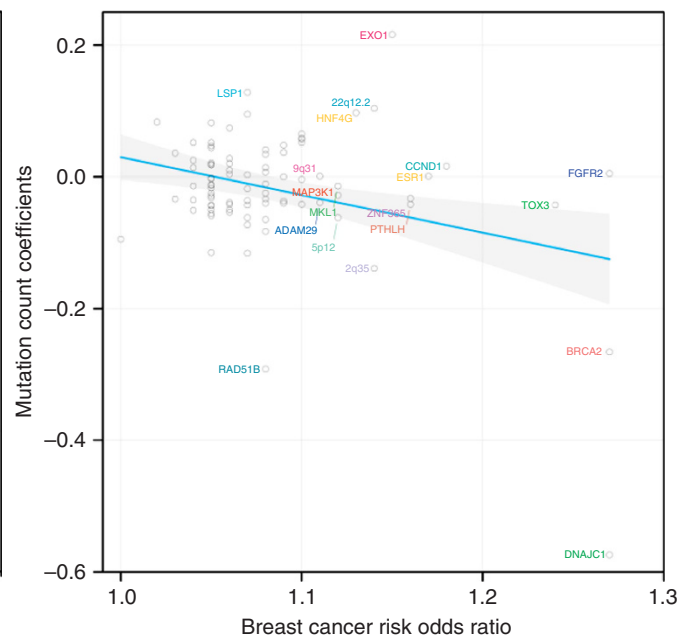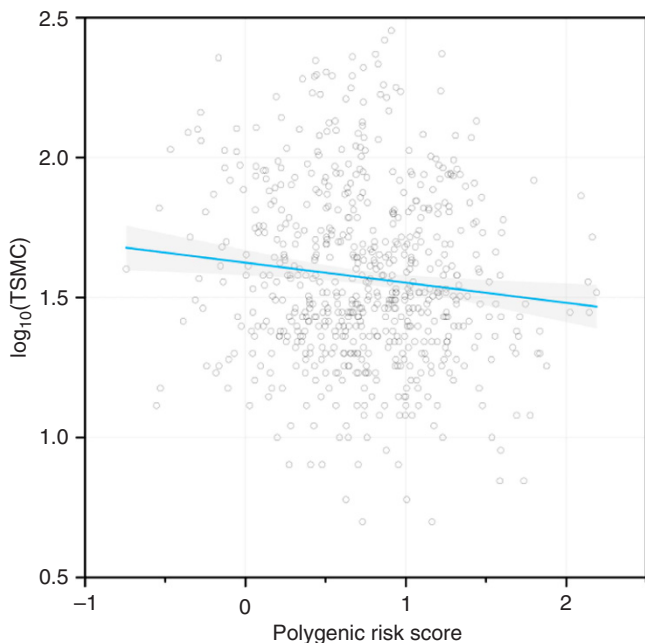
The $P = 0.01$ under the null hypothesis that the regression line is flat (left). Scatterplot for mutation count coefficient versus breast cancer risk overall odds ratio with $P = 2.28 \times 10^{-3}$ for the flat line hypothesis (right). The least squares regression lines are shown by the solid lines with 95% confidence intervals in gray shades.

Figure 1. Scatterplot for TSMC vs PRS.

Table 2. Association between somatic mutation phenotypes and SNP rs2588809 at RAD51B

Phenotype[a]	Group	Estimate[b]	S.e.	P-value
TSMC	All subjects	− 0.29	0.07	8.75e − 06
TSMC	ER[+]	− 0.25	0.07	6.24e − 04
TSMC	ER[−]	− 0.52[c]	0.17	2.46e − 03
TSMC	ER[+] PR[+]	− 0.25	0.08	1.77e − 03
TSMC	ER[+] PR[−]	− 0.27	0.22	2.13e − 01
TSMC	ER[−] PR[−]	− 0.46	0.18	1.07e − 02
T counts[d]	All subjects	− 0.27	0.07	6.59e − 05
C counts[e]	All subjects	− 0.28	0.07	3.12e − 05
Transitions[f]	All subjects	− 0.29	0.07	1.02e − 05
Transversions[g]	All subjects	− 0.25	0.07	1.61e − 04
APOBEC counts[h]	All subjects	− 0.20	0.07	2.64e − 03

Abbreviations: ER = oestrogen receptor; PR = progestrogen receptor; SNP = single-nucleotide polymorphism; TSMC = total somatic mutation count.
[a]The counts are $\log_{10}$ transformed and standardised with mean zero and variance one.
[b]The estimates are adjusted by age, tumour stage, ER (not for subgroup stratified by ER) and PR (not for subgroup stratified by PR) status.
[c]Although present in both subgroups, the strength of association appeared to be stronger for ER[−] than ER[+] subgroup ($P = 0.31$ for heterogeneity by testing interaction between genotype of rs2588809 and ER status).
[d]Counts of TA, TC and TG. TA, somatic mutation from thymine to adenine or somatic mutation on the other strand, that is, from adenine to thymine. The same notation is used for the following somatic mutation counts. TC, somatic mutation from thymine to cytosine. TG, counts of somatic mutation from thymine to guanine.
[e]Counts of CA, CG and CT. CA, somatic mutation from cytosine to adenine. CG, somatic mutation from cytosine to guanine. CT, somatic mutation from cytosine to thymine.
[f]Counts of mutations between purines or between pyrimidine.
[g]Counts of mutations from purines to pyrimidine or vise versa.
[h]Counts of APOBEC mutation signatures.

Table 3. Association between somatic mutation phenotypes and PRS

Phenotype[a]	Group	Estimate[b]	S.e.	P-value
TSMC	All subjects	− 0.19	0.08	1.34e − 02
TSMC	All subjects	− 0.16[c]	0.08	3.81e − 02
TSMC	ER[+]	− 0.19[d,e]	0.08	1.97e − 02
TSMC	ER[−]	0.18[f]	0.21	3.91e − 01
TSMC	ER[+] PR[+]	− 0.24[d]	0.09	7.24e − 03
TSMC	ER[+] PR[−]	0.01[d]	0.25	9.59e − 01
TSMC	ER[−] PR[−]	0.17[f]	0.22	4.36e − 01
T counts[g]	All subjects	− 0.21	0.08	8.18e − 03
T counts[g]	All subjects	− 0.18[c]	0.08	2.21e − 02
C counts[h]	All subjects	− 0.18	0.08	2.47e − 02
C counts[h]	All subjects	− 0.15[c]	0.08	6.11e − 02
Transitions[i]	All subjects	− 0.18	0.08	2.44e − 02
Transversions[j]	All subjects	− 0.19	0.08	1.54e − 02
APOBEC counts[k]	All subjects	− 0.23	0.08	4.47e − 03

Abbreviations: ER = oestrogen receptor; PR = progestrogen receptor; PRS = polygenic risk score; TSMC = total somatic mutation count.
[a]The counts were $\log_{10}$ transformed and standardised with mean zero and variance one.
[b]If not specified otherwise, PRS is based on 90 SNPs with reported breast cancer overall odds ratios; ER+ specific RPS was used for stratified analysis of ER[+], ER[+]PR[+] and ER[+]PR[−] groups, ER[−]-specific PRS for ER[−] and ER[−]PR[−] groups and overall PRS for all subjects. The estimates are adjusted by age, tumour stage, ER (not for subgroup stratified by ER) and PR (not for subgroup stratified by PR) statuses.
[c]PRS is based on 89 SNPs excluding rs2588809 at RAD51B with reported breast cancer overall odds ratios.
[d]ER[+]-specific PRS, which is based on 90 SNPs with reported breast cancer ER[+] odds ratios.
[e]The association appeared to be presented in ER[+] but not in ER- subgroup ($P = 0.01$ for heterogeneity by testing the equivalence of ER[+] and ER[−] Z-statistics).
[f]ER[−]-specific PRS, which is based on 90 SNPs with reported breast cancer ER[−] odds ratios.
[g]Counts of TA, TC and TG. TA, somatic mutation from thymine to adenine or somatic mutation on the other strand, that is, from adenine to thymine. The same notation is used for the following somatic mutation counts. TC, somatic mutation from thymine to cytosine. TG, counts of somatic mutation from thymine to guanine.
[h]Counts of CA, CG and CT. CA, somatic mutation from cytosine to adenine. CG, somatic mutation from cytosine to guanine. CT, somatic mutation from cytosine to thymine.
[i]Counts of mutations between purines or between pyrimidine.
[j]Counts of mutations from purines to pyrimidine or vise versa.
[k]Counts of APOBEC mutation signatures.

largely influenced by SNP rs11814448 in DNAJC1 ($P = 0.37$ when excluding SNP rs11814448). The association between PRS and TSMC remained significant ($P = 3.81 \times 10^{-2}$) after excluding rs2588809 in RAD51B in the calculation of PRS. Analysis of ER[+]-specific PRS in ER[+] tumours and ER[−]-specific PRS in ER[−] tumours, showed a significant inverse association for ER[+] but not ER[−] tumours ($P = 0.01$ for heterogeneity, Supplementary Figure 4). Further, the association observed for ER[+] tumour appears to be to be present only for ER[+] PR[+] tumours ($P = 7.24 \times 10^{-3}$) and not for ER[+] PR[−] tumour ($P = 0.95$). The association patterns for TSMC with PRS and RAD51B SNP did not change when we additionally adjusted for total number of CNV segments observed for the patients as covariate in the respective regression models (Supplementary Table 7 for PRS and Supplementary Table 8 for RAD51B SNP).

DISCUSSION

We reported an inverse association between TSMC in breast tumours and genetic predisposition conferred by common breast cancer susceptibility SNPs. In particular, a highly significant inverse association was observed with respect to the germline risk variant defined by the SNP rs2588809 in the DNA-repair gene RAD51B. Moreover, a significant inverse association was also observed for a PRS for breast cancer that includes genetic predisposition of 90 breast cancer associated loci but this association was only evident among ER[+] tumours with respect to ER[+]-specific PRS.

The reported inverse association may provide insight into links between germline risk variants and somatic mutations in breast cancer development. There are several possible underlying mechanisms by which the inverse association could arise.

It has been previously shown that genetic susceptibility loci differentially influence distinct subtypes of breast cancer (Stephens et al, 2012), which, in turn, can be related to the number of somatic mutations. For example, many SNPs that have been reported to date from GWAS of breast cancer show differential associations with the risk of ER[+] and ER[−] breast cancer. Because the number of ER[+] tumours included in GWAS to date has been substantially more than the number for ER[−] tumours (Michailidou et al, 2015), GWAS discovery has preferentially identified SNPs related to ER[+] tumours. As the number of total mutations tends to be larger in ER[−] than ER[+] tumours (Table 1), an inverse association between the SNPs and mutation counts may be observed if the analysis is not adjusted for ER status (Stephens et al, 2012).

We observed an inverse association between germline risk and TSMC after adjustment for age at diagnosis and tumour characteristics, including ER/PR status and stage. The association with the RAD51B SNP was present for both ER[+] and ER[−] tumours, although this SNP is only associated with the risk of ER[+] tumours (Michailidou et al, 2013). In contrast, the association of TSMC with PRS was present only for ER[+]-specific PRS in ER[+] tumours, and this association appeared to be strongest for the ER[+]PR[+] subtype. Further tumour characteristics, such as the grade that could not be evaluated in this report due to lack of available data, could explain the reported inverse associations between PRS and TSMC. However, the distinct pattern of

association seen for the *RAD51B* SNP and PRS are unlikely to be both explained by subtype heterogeneity.

We observed an association between higher TSMC with older age at diagnosis, ER$^-$, PR$^-$ and higher stage. Older age at diagnosis was previously reported to be associated with cytosine to thymine substitution in ER$^-$, but not with TSMC, across breast cancer patients, whereas the observation of higher TSMC associated with higher stage is consistent with the previous finding (Stephens *et al*, 2012). To our best knowledge, we are not aware of other studies reporting an association between TSMC, overall and by mutation type, with respect to joint status ER, PR, stage and age at diagnosis. Our results suggested that although PR status, but not ER, is strongly predictive of overall TSMC, distinct mutation signatures could be associated with ER (thymine to other nucleotides) and PR status (cytosine to other nucleotides) when the characteristics were analysed jointly.

It is possible that the inverse association we observe between genetic risk and TSMC is a broader phenomenon that cannot be explained by subtype heterogeneity alone. The best known example of interaction of germline and somatic mutation is the 'two-hit' model for carcinogenesis in retinoblastoma (Knudson, 1971). Under this model, the first-hit could be either a germline susceptibility variant or a somatic mutation in an important cancer predisposition gene. Thus subjects with elevated genetic predisposition may require fewer stages to develop a malignancy of the breast than subjects at lower genetic risk. Therefore, it is possible that the observed inverse association is the result of an underlying continuous process of cancer development, in which both germline variants and somatic mutations contribute and perhaps overlap with respect to their relative contributions to development of breast cancer. To further understand the biological basis of our observations, it will be necessary to understand the causal mechanisms that underpin the relationship between common susceptibility alleles and TSMC, as a marker of mutational events critical for development of distinct subtypes of breast cancer. In addition, the current study only recorded presence or absence of ER or PR. By quantifying the magnitudes of ER or PR as quantitative traits in future studies, it may be possible to delineate the relationships between TSMC with ER or PR levels more precisely.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Altshuler D, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De la Vega FM, Donnelly P, Egholm M, Flicek P, Gabriel SB, Gibbs RA, Knoppers BM, Lander ES, Lehrach H, Mardis ER, McVean GA, Nickerson D, Peltonen L, Schafer AJ, Sherry ST, Wang J, Wilson RK, Gibbs RA, Deiros D, Metzker M, Muzny D, Reid J, Wheeler D, Wang J, Li JX, Jian M, Li G, Li RQ, Liang HQ, Tian G, Wang B, Wang J, Wang W, Yang HM, Zhang XQ, Zheng HS, Lander ES, Altshuler DL, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL, Bentley DR, Gormley N, Humphray S, Kingsbury Z, Koko-Gonzales P, Stone J, McKernan KJ, Costa GL, Ichikawa JK, Lee CC, Sudbrak R, Lehrach H, Borodina TA, Dahl A, Davydov AN, Marquardt P, Mertes F, Nietfeld W, Rosenstiel P, Schreiber S, Soldatov AV,

Timmermann B, Tolzmann M, Egholm M, Affourtit J, Ashworth D, Attiya S, Bachorski M, Buglione E, Burke A, Caprio A, Celone C, Clark S, Conners D, Desany B, Gu L, Guccione L, Kao K, Kebbel A, Knowlton J, Labrecque M, McDade L, Mealmaker C, Minderman M, Nawrocki A, Niazi F, Pareja K, Ramenani R, Riches D, Song W, Turcotte C, Wang S, Mardis ER, Dooling D, Fulton L, Fulton R, Weinstock G, Durbin RM, Burton J, Carter DM, Churcher C, Coffey A, Cox A, Palotie A, Quail M, Skelly T, Stalker J, Swerdlow HP, Turner D, De Witte A, Giles S, Gibbs RA, Wheeler D, Bainbridge M, Challis D, Sabo A, Yu F, Yu J, Wang J, Fang XD, Guo XS, Li RQ, Li YR, Luo RB, Tai S, Wu HL, Zheng HC, Zheng XL, Zhou Y, Yang HM, Marth GT, Garrison EP, Huang W, Indap A, Kural D, Lee WP, Leong WF, Huang WC, Indap A, Kural D, Lee WP, Leong WF, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, Daly MJ, DePristo MA, Altshuler DL, Ball AD, Banks E, Bloom T, Browning BL, Cibulskis K, Fennell TJ, Garimella KV, Grossman SR, Handsaker RE, Hanna M, Hartl C, Jaffe DB, Kernytsky AM, Korn JM, Li H, Maguire JR, McCarroll SA, McKenna A, Nemesh JC, Philippakis AA, Poplin RE, Price A, Rivas MA, Sabeti PC, Schaffner SF, Shefler E, Shlyakhter IA, Cooper DN, Ball EV, Mort M, Phillips AD, Stenson PD, Sebat J, Makarov V, Ye K, Yoon SC, Bustamante CD, Clark AG, Boyko A, Degenhardt J, Gravel S, Gutenkunst RN, Kaganovich M, Keinan A, Lacroute P, Ma X, Reynolds A, Clarke L, Flicek P, Cunningham F, Herrero J, Keenen S, Kulesha E, Leinonen R, McLaren W, Radhakrishnan R, Smith RE, Zalunin V, Zheng-Bradley XQ, Korbel JO, Stutz AM, Humphray S, Bauer M, Cheetham RK, Cox T, Eberle M, James T, Kahn S, Murray L, Ye K, De La Vega FM, Fu YT, Hyland FCL, Manning JM, McLaughlin SF, Peckham HE, Sakarya O, Sun YA, Tsung EF, Batzer MA, Konkel MK, Walker JA, Sudbrak R, Albrecht MW, Amstislavskiy VS, Herwig R, Parkhomchuk DV, Sherry ST, Agarwala R, Khouri H, Morgulis AO, Paschall JE, Phan LD, Rotmistrovsky KE, Sanders RD, Shumway MF, Xiao CL, McVean GA, Auton A, Iqbal Z, Lunter G, Marchini JL, Moutsianas L, Myers S, Tumian A, Desany B, Knight J, Winer R, Craig DW, Beckstrom-Sternberg SM, Christoforides A, Kurdoglu AA, Pearson J, Sinari SA, Tembe WD, Haussler D, Hinrichs AS, Katzman SJ, Kern A, Kuhn RM, Przeworski M, Hernandez RD, Howie B, Kelley JL, Melton SC, Abecasis GR, Li Y, Anderson P, Blackwell T, Chen W, Cookson WO, Ding J, Kang HM, Lathrop M, Liang LM, Moffatt MF, Scheet P, Sidore C, Snyder M, Zhan XW, Zollner S, Awadalla P, Casals F, Idaghdour Y, Keebler J, Stone EA, Zilversmit M, Jorde L, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Sahinalp SC, Sudmant PH, Mardis ER, Chen K, Chinwalla A, Ding L, Koboldt DC, McLellan MD, Dooling D, Weinstock G, Wallis JW, Wendl MC, Zhang QY, Durbin RM, Albers CA, Ayub Q, Balasubramaniam S, Barrett JC, Carter DM, Chen YA, Conrad DF, Danecek P, Dermitzakis ET, Hu M, Huang N, Hurles ME, Jin HJ, Jostins L, Keane TM, Keane TM, Le SQ, Lindsay S, Long QA, MacArthur DG, Montgomery SB, Parts L, Stalker J, Tyler-Smith C, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Abyzov A, Balasubramanian S, Bjornson R, Du JA, Grubert F, Habegger L, Haraksingh R, Jee J, Khurana E, HYK Lam, Leng J, Mu XJ, Urban AE, Zhang ZD, Li YR, Luo RB, Marth GT, Garrison EP, Kural D, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, McCarroll SA, Banks E, DePristo MA, Handsaker RE, Hartl C, Korn JM, Li H, Nemesh JC, Sebat J, Makarov V, Ye K, Yoon SC, Degenhardt J, Kaganovich M, Clarke L, Smith RE, Zheng-Bradley XQ, Korbel JO, Humphray S, Cheetham RK, Eberle M, Kahn S, Murray L, Ye K, De la Vega FM, Fu YT, Peckham HE, Sun YA, Batzer MA, Konkel MK, Xiao CL, Iqbal Z, Desany B, Blackwell T, Snyder M, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Chen K, Chinwalla A, Ding L, McLellan MD, Wallis JW, Hurles ME, Conrad DF, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Du JA, Grubert F, Haraksingh R, Jee J, Khurana E, HYK Lam, Leng J, Mu XJ, Urban AE, Zhang ZD, Gibbs RA, Bainbridge M, Challis D, Coafra C, Dinh H, Kovar C, Lee S, Muzny D, Nazareth L, Reid J, Sabo A, Yu FL, Yu J, Marth GT, Garrison EP, Indap A, Leong WF, Quinlan AR, Stewart C, Ward AN, Wu JT, Cibulskis K, Fennell TJ, Gabriel SB, Garimella KV, Hartl C, Shefler E, Sougnez CL, Wilkinson J, Clark AG, Gravel S, Grubert F, Clarke L, Flicek P, Smith RE, Zheng-Bradley XQ, Sherry ST, Khouri HM, Paschall JE, Shumway MF, Xiao CL, McVean GA, Katzman SJ, Abecasis GR, Blackwell T, Mardis ER, Dooling D, Fulton L, Fulton R, Koboldt DC, Durbin RM, Balasubramaniam S, Coffey A, Keane TM, MacArthur DG, Palotie A, Scott C, Stalker J, Tyler-Smith C, Gerstein MB, Balasubramanian S, Chakravarti A, Knoppers BM,

Peltonen L, Abecasis GR, Bustamante CD, Gharani N, Gibbs RA, Jorde L, Kaye JS, Kent A, Li T, McGuire AL, McVean GA, Ossorio PN, Rotimi CN, Su YY, Toji LH, Tyler-Smith C, Brooks LD, Felsenfeld AL, McEwen JE, Abdallah A, Christopher R, Clemm NC, Collins FS, Duncanson A, Green ED, Guyer MS, Peterson JL, Schafer AJ, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 1000 Genomes Project Consortium (2010a) A map of human genome variation from population-scale sequencing. *Nature* 467(7319): 1061–1073.

Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu FL, Bonnen PE, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia XM, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren YR, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang QR, Ghori MJR, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsud I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng CQ, Brooks LD, McEwen JE. International HapMap 3 Consortium (2010b) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311): 52–58.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L, Cortes ML, Fernandez-Lopez JC, Peng S, Ardlie KG, Auclair D, Bautista-Pina V, Duke F, Francis J, Jung J, Maffuz-Aziz A, Onofrio RC, Parkin M, Pho NH, Quintanar-Jurado V, Ramos AH, Rebollar-Vega R, Rodriguez-Cuevas S, Romero-Cordoba SL, Schumacher SE, Stransky N, Thompson KM, Uribe-Figueroa L, Baselga J, Beroukhim R, Polyak K, Sgroi DC, Richardson AL, Jimenez-Sanchez G, Lander ES, Gabriel SB, Garraway LA, Golub TR, Melendez-Zajgla J, Toker A, Getz G, Hidalgo-Miranda A, Meyerson M (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486(7403): 405–409.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B Met* 57(1): 289–300.

Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, Edwards SL, Pickett HA, Shen HC, Smart CE, Hillman KM, Mai PL, Lawrenson K, Stutz MD, Lu Y, Karevan R, Woods N, Johnston RL, French JD, Chen X, Weischer M, Nielsen SF, Maranian MJ, Ghoussaini M, Ahmed S, Baynes C, Bolla MK, Wang Q, Dennis J, McGuffog L, Barrowdale D, Lee A, Healey S, Lush M, Tessier DC, Vincent D, Bacot F. AuAustralian Cancer Study; Australian Ovarian Cancer Study; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab); Gene Environment Interaction and Breast Cancer (GENICA); Swedish Breast Cancer Study (SWE-BRCA); Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); Epidemiological study of BRCA1 & BRCA2 Mutation Carriers (EMBRACE); Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO)stralian Cancer S, Australian Ovarian Cancer S, Kathleen Cuningham Foundation Consortium for Research into Familial Breast C, Gene Environment I, Breast C, Swedish Breast Cancer S, Hereditary B, Ovarian Cancer Research Group N, Epidemiological study of B, Carriers BM, Genetic Modifiers of Cancer Risk in BMCVergote I, Lambrechts E, Despierre E, Risch HA, Gonzalez-Neira A, Rossing MA, Pita G, Doherty JA, Alvarez N, Larson MC, Fridley BL, Schoof N, Chang-Claude J, Cicek MS, Peto J, Kalli KR, Broeks A, Armasu SM, Schmidt MK, Braaf LM, Winterhoff B, Nevanlinna H, Konecny GE, Lambrechts D, Rogmann L, Guenel P, Teoman A, Milne RL, Garcia JJ, Cox A, Shridhar V, Burwinkel B, Marme F, Hein R, Sawyer EJ, Haiman CA, Wang-Gohrke S, Andrulis IL, Moysich KB, Hopper JL, Odunsi K, Lindblom A, Giles GG, Brenner H, Simard J, Lurie G, Fasching PA, Carney ME, Radice P, Wilkens LR, Swerdlow A, Goodman MT, Brauch H, Garcia-Closas M, Hillemanns P, Winqvist R, Durst M, Devilee P, Runnebaum I, Jakubowska A, Lubinski J, Mannermaa A, Butzow R, Bogdanova NV, Dork T, Pelttari LM, Zheng W, Leminen A, Anton-Culver H, Bunker CH, Kristensen V, Ness RB, Muir K, Edwards R, Meindl A, Heitz F, Matsuo K, du Bois A, Wu AH, Harter P, Teo SH, Schwaab I, Shu XO, Blot W, Hosono S, Kang D, Nakanishi T, Hartman M, Yatabe Y, Hamann U, Karlan BY, Sangrajrang S, Kjaer SK, Gaborieau V,

Jensen A, Eccles D, Hogdall E, Shen CY, Brown J, Woo YL, Shah M, Azmi MA, Luben R, Omar SZ, Czene K, Vierkant RA, Nordestgaard BG, Flyger H, Vachon C, Olson JE, Wang X, Levine DA, Rudolph A, Weber RP, Flesch-Janys D, Iversen E, Nickels S, Schildkraut JM, Silva Idos S, Cramer DW, Gibson L, Terry KL, Fletcher O, Vitonis AF, van der Schoot CE, Poole EM, Hogervorst FB, Tworoger SS, Liu J, Bandera EV, Li J, Olson SH, Humphreys K, Orlow I, Blomqvist C, Rodriguez-Rodriguez L, Aittomaki K, Salvesen HB, Muranen TA, Wik E, Brouwers B, Krakstad C, Wauters E, Halle MK, Wildiers H, Kiemeney LA, Mulot C, Aben KK, Laurent-Puig P, Altena AM, Truong T, Massuger LF, Benitez J, Pejovic T, Perez JI, Hoatlin M, Zamora MP, Cook LS, Balasubramanian SP, Kelemen LE, Schneeweiss A, Le ND, Sohn C, Brooks-Wilson A, Tomlinson I, Kerin MJ, Miller N, Cybulski C, Henderson BE, Menkiszak J, Schumacher F, Wentzensen N, Le Marchand L, Yang HP, Mulligan AM, Glendon G, Engelholm SA, Knight JA, Hogdall CK, Apicella C, Gore M, Tsimiklis H, Song H, Southey MC, Jager A, den Ouweland AM, Brown R, Martens JW, Flanagan JM, Kriege M, Paul J, Margolin S, Siddiqui N, Severi G, Whittemore AS, Baglietto L, McGuire V, Stegmaier C, Sieh W, Muller H, Arndt V, Labreche F, Gao YT, Goldberg MS, Yang G, Dumont M, McLaughlin JR, Hartmann A, Ekici AB, Beckmann MW, Phelan CM, Lux MP, Permuth-Wey J, Peissel B, Sellers TA, Ficarazzi F, Barile M, Ziogas A, Ashworth A, Gentry-Maharaj A, Jones M, Ramus SJ, Orr N, Menon U, Pearce CL, Bruning T, Pike MC, Ko YD, Lissowska J, Figueroa J, Kupryjanczyk J, Chanock SJ, Dansonka-Mieszkowska A, Jukkola-Vuorinen A, Rzepecka IK, Pylkas K, Bidzinski M, Kauppila S, Hollestelle A, Seynaeve C, Tollenaar RA, Durda K, Jaworska K, Hartikainen JM, Kosma VM, Kataja V, Antonenkova NN, Long J, Shrubsole M, Deming-Halverson S, Lophatananon A, Siriwanarangsan P, Stewart-Brown S, Ditsch N, Lichtner P, Schmutzler RK, Ito H, Iwata H, Tajima K, Tseng CC, Stram DO, van den Berg D, Yip CH, Ikram MK, Teh YC, Cai H, Lu W, Signorello LB, Cai Q, Noh DY, Yoo KY, Miao H, Iau PT, Teo YY, McKay J, Shapiro C, Ademuyiwa F, Fountzilas G, Hsiung CN, Yu JC, Hou MF, Healey CS, Luccarini C, Peock S, Stoppa-Lyonnet D, Peterlongo P, Rebbeck TR, Piedmonte M, Singer CF, Friedman E, Thomassen M, Offit K, Hansen TV, Neuhausen SL, Szabo CI, Blanco I, Garber J, Narod SA, Weitzel JN, Montagna M, Olah E, Godwin AK, Yannoukakos D, Goldgar DE, Caldes T, Imyanitov EN, Tihomirova L, Arun BK, Campbell I, Mensenkamp AR, van Asperen CJ, van Roozendaal KE, Meijers-Heijboer H, Collee JM, Oosterwijk JC, Hooning MJ, Rookus MA, van der Luijt RB, Os TA, Evans DG, Frost D, Fineberg E, Barwell J, Walker L, Kennedy MJ, Platte R, Davidson R, Ellis SD, Cole T, Bressac-de Paillerets B, Buecher B, Damiola F, Faivre L, Frenay M, Sinilnikova OM, Caron O, Giraud S, Mazoyer S, Bonadona V, Caux-Moncoutier V, Toloczko-Grabarek A, Gronwald J, Byrski T, Spurdle AB, Bonanni B, Zaffaroni D, Giannini G, Bernard L, Dolcetti R, Manoukian S, Arnold N, Engel C, Deissler H, Rhiem K, Niederacher D, Plendl H, Sutter C, Wappenschmidt B, Borg A, Melin B, Rantala J, Soller M, Nathanson KL, Domchek SM, Rodriguez GC, Salani R, Kaulich DG, Tea MK, Paluch SS, Laitman Y, Skytte AB, Kruse TA, Jensen UB, Robson M, Gerdes AM, Ejlertsen B, Foretova L, Savage SA, Lester J, Soucy P, Kuchenbaecker KB, Olswold C, Cunningham JM, Slager S, Pankratz VS, Dicks E, Lakhani SR, Couch FJ, Hall P, Monteiro AN, Gayther SA, Pharoah PD, Reddel RR, Goode EL, Greene MH, Easton DF, Berchuck A, Antoniou AC, Chenevix-Trench G, Dunning AM (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet* 45(4): 371–384, 384e1–2.

Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, Shi J, Long J, Wen W, Choi JY, Noh DY, Shen CY, Matsuo K, Teo SH, Kim MK, Khoo US, Iwasaki M, Hartman M, Takahashi A, Ashikawa K, Matsuda K, Shin MH, Park MH, Zheng Y, Xiang YB, Ji BT, Park SK, Wu PE, Hsiung CN, Ito H, Kasuga Y, Kang P, Mariapun S, Ahn SH, Kang HS, Chan KY, Man EP, Iwata H, Tsugane S, Miao H, Liao J, Nakamura Y, Kubo M. DRIVE GAME-ON ConsortiumDelahanty RJ, Zhang Y, Li B, Li C, Gao YT, Shu XO, Kang D, Zheng W (2014) Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet* 46(8): 886–890.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S. GrouMETABRIC Groupp MLangerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S,

Caldas C, Aparicio S (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**(7403): 346–352.

DePinho RA (2000) The age of cancer. *Nature* **408**(6809): 248–254.

Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans GR, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD (2015) Gene-panel sequencing and the prediction of breast-cancer risk. *N Engl J Med* **372**(23): 2243–2257.

Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, Van Tine BA, Hoog J, Goiffon RJ, Goldstein TC, Ng S, Lin L, Crowder R, Snider J, Ballman K, Weber J, Chen K, Koboldt DC, Kandoth C, Schierding WS, McMichael JF, Miller CA, Lu C, Harris CC, McLellan MD, Wendl MC, DeSchryver K, Allred DC, Esserman L, Unzeitig G, Margenthaler J, Babiera GV, Marcom PK, Guenther JM, Leitch M, Hunt K, Olson J, Tao Y, Maher CA, Fulton LL, Fulton RS, Harrison M, Oberkfell B, Du F, Demeter R, Vickery TL, Elhammali A, Piwnica-Worms H, McDonald S, Watson M, Dooling DJ, Ota D, Chang LW, Bose R, Ley TJ, Piwnica-Worms D, Stuart JM, Wilson RK, Mardis ER (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**(7403): 353–360.

Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, Miron P, Fasching PA, Brauch H, Chang-Claude J, Carpenter J, Godwin AK, Nevanlinna H, Giles GG, Cox A, Hopper JL, Bolla MK, Wang Q, Dennis J, Dicks E, Howat WJ, Schoof N, Bojesen SE, Lambrechts D, Broeks A, Andrulis IL, Guenel P, Burwinkel B, Sawyer EJ, Hollestelle A, Fletcher O, Winqvist R, Brenner H, Mannermaa A, Hamann U, Meindl A, Lindblom A, Zheng W, Devillee P, Goldberg MS, Lubinski J, Kristensen V, Swerdlow A, Anton-Culver H, Dork T, Muir K, Matsuo K, Wu AH, Radice P, Teo SH, Shu XO, Blot W, Kang D, Hartman M, Sangrajrang S, Shen CY, Southey MC, Park DJ, Hammet F, Stone J, Veer LJ, Rutgers EJ, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Peto J, Schrauder MG, Ekici AB, Beckmann MW, Dos Santos Silva I, Johnson N, Warren H, Tomlinson I, Kerin MJ, Miller N, Marme F, Schneeweiss A, Sohn C, Truong T, Laurent-Puig P, Kerbrat P, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Perez JI, Menendez P, Muller H, Arndt V, Stegmaier C, Lichtner P, Lochmann M, Justenhoven C, Ko YD, Gene Environmental Interaction and breast Cancer (GENICA) Network, Muranen TA, Aittomaki K, Blomqvist C, Greco D, Heikkinen T, Ito H, Iwata H, Yatabe Y, Antonenkova NN, Margolin S, Kataja V, Kosma VM, Hartikainen JM, Balleine R, kConFab I, Tseng CC, Berg DV, Stram DO, Neven P, Dieudonne AS, Leunen K, Rudolph A, Nickels S, Flesch-Janys D, Peterlongo P, Peissel B, Bernard L, Olson JE, Wang X, Stevens K, Severi G, Baglietto L, McLean C, Coetzee GA, Feng Y, Henderson BE, Schumacher F, Bogdanova NV, Labreche F, Dumont M, Yip CH, Taib NA, Cheng CY, Shrubsole M, Long J, Pylkas K, Jukkola-Vuorinen A, Kauppila S, Knight JA, Glendon G, Mulligan AM, Tollenaar RA, Seynaeve CM, Kriege M, Hooning MJ, van den Ouweland AM, van Deurzen CH, Lu W, Gao YT, Cai H, Balasubramanian SP, Cross SS, Reed MW, Signorello L, Cai Q, Shah M, Miao H, Chan CW, Chia KS, Jakubowska A, Jaworska K, Durda K, Hsiung CN, Wu PE, Yu JC, Ashworth A, Jones M, Tessier DC, Gonzalez-Neira A, Pita G, Alonso MR, Vincent D, Bacot F, Ambrosone CB, Bandera EV, John EM, Chen GK, Hu JJ, Rodriguez-Gil JL, Bernstein L, Press MF, Ziegler RG, Millikan RM, Deming-Halverson SL, Nyante S, Ingles SA, Waisfisz Q, Tsimiklis H, Makalic E, Schmidt D, Bui M, Gibson L, Muller-Myhsok B, Schmutzler RK, Hein R, Dahmen N, Beckmann L, Aaltonen K, Czene K, Irwanto A, Liu J, Turnbull C. FamFamilial Breast Cancer Study (FBCS)ilial Breast Cancer SRahman N, Meijers-Heijboer H, Uiterlinden AG, Rivadeneira F. Australian Breast Cancer Tissue Bank (ABCTB) InvestigatorsOlswold C, Slager S, Pilarski R, Ademuyiwa F, Konstantopoulou I, Martin NG, Montgomery GW, Slamon DJ, Rauh C, Lux MP, Jud SM, Bruning T, Weaver J, Sharma P, Pathak H, Tapper W, Gerty S, Durcan L, Trichopoulos D, Tumino R, Peeters PH, Kaaks R, Campa D, Canzian F, Weiderpass E, Johansson M, Khaw KT, Travis R, Clavel-Chapelon F, Kolonel LN, Chen C, Beck A, Hankinson SE, Berg CD, Hoover RN, Lissowska J, Figueroa JD, Chasman DI, Gaudet MM, Diver WR, Willett WC, Hunter DJ, Simard J, Benitez J, Dunning AM, Sherman ME, Chenevix-Trench G, Chanock SJ, Hall P, Pharoah PD, Vachon C, Easton DF, Haiman CA, Kraft P (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* **45**(4): 392–398, 384e1–2.

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**: e1000529.

Knudson Jr. AG (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* **68**(4): 820–823.

Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, Wilson RK, Ally A, Balasundaram M, Butterfield YSN, Carlsen R, Carter C, Chu A, Chuah E, Chun HJE, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HYI, Mayo M, Moore RA, Mungall AJ, Pleasance ERobertson AG, Schein JE, Shafiei A, Sipahimalani P, Slobodan JR, Stoll D, Tam A, Thiessen N, Varhol RJ, Wye N, Zeng T, Zhao YJ, Birol I, Jones SJM, Marra MA, Cherniack AD, Saksena G, Onofrio RC, Pho NH, Carter SL, Schumacher SE, Tabak B, Hernandez B, Gentry J, Nguyen H, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Getz G, Gabriel SB, Meyerson M, Chin L, Park PJ, Kucherlapati R, Hoadley KA, Auman JT, Fan C, Turman YJ, Shi Y, Li L, Topal MD, He XP, Chao HH, Prat A, Silva GO, Iglesia MD, Zhao W, Usary J, Berg JS, Adams M, Booker J, Wu JY, Gulabani A, Bodenheimer T, Hoyle AP, Simons JV, Soloway MG, Mose LE, Jefferys SR, Balu S, Parker JS, Hayes DN, Perou CM, Malik S, Mahurkar S, Shen H, Weisenberger DJ, Triche T, Lai PH, Bootwalla MS, Maglinte DT, Berman BP, Van den Berg DJ, Baylin SB, Laird PW, Creighton CJ, Donehower LA, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang JH, Zhang HL, Wu CJ, Liu SY, Lawrence MS, Zou LH, Sivachenko A, Lin P, Stojanov P, Jing R, Cho J, Sinha R, Park RW, Nazaire MD, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Reynolds S, Kreisberg RB, Bernard B, Bressler R, Erkkila T, Lin J, Thorsson V, Zhang W, Shmulevich I, Ciriello G, Weinhold N, Schultz N, Gao JJ, Cerami E, Gross B, Jacobsen A, Sinha R, Aksoy BA, Antipin Y, Reva B, Shen RL, Taylor BS, Ladanyi M, Sander C, Anur P, Spellman PT, Lu YL, Liu WB, Verhaak RRG, Mills GB, Akbani R, Zhang NX, Broom BM, Casasent TD, Wakefield C, Unruh AK, Baggerly K, Coombes K, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Zhu JC, Szeto CC, Scott GK, Yau C, Paul EO, Carlin D, Wong C, Sokolov A, Thusberg J, Mooney S, Ng S, Goldstein TC, Ellrott K, Grifford M, Wilks C, Ma S, Craft B, Yan CH, Hu Y, Meerzaman D, Gastier-Foster JM, Bowen J, Ramirez NC, Black AD, Pyatt RE, White P, Zmuda EJ, Frick J, Lichtenberg T, Brookens R, George MM, Gerken MA, Harper HA, Leraas KM, Wise LJ, Tabler TR, McAllister C, Barr T, Hart-Kothari M, Tarvin K, Saller C, Sandusky G, Mitchell C, Iacocca MV, Brown J, Rabeno B, Czerwinski C, Petrelli N, Dolzhansky O, Abramov M, Voronina O, Potapova O, Marks JR, Suchorska WM, Murawa D, Kycler W, Ibbs M, Korski K, Spychala A, Murawa P, Brzezinski JJ, Perz H, Lazniak R, Teresiak M, Tatka H, Leporowska E, Bogusz-Czerniewicz M, Malicki J, Mackiewicz A, Wiznerowicz M, Le XV, Kohl B, Tien NV, Thorp R, Bang NV, Sussman H, Phu BD, Hajek R, Hung NP, VTP Tran, Thang HQ, Khan KZ, Penny R, Mallery D, Curley E, Shelton C, Yena P, Ingle JN, Couch FJ, Lingle WL, King TA, Gonzalez-Angulo AM, Mills GB, Dyer MD, Liu SY, Meng XL, Patangan M, Waldman F, Stoppler H, Rathmell WK, Thorne L, Huang M, Boice L, Hill A, Morrison C, Gaudioso C, Bshara W, Daily K, Egea SC, Pegram MD, Gomez-Fernandez C, Dhir R, Bhargava R, Brufsky A, Shriver CD, Hooke JA, Campbell JL, Mural RJ, Hu H, Somiari S, Larson C, Deyarmin B, Kvecher L, Kovatich AJ, Ellis MJ, King TA, Hu H, Couch FJ, Mural RJ, Stricker T, White K, Olopade O, Ingle JN, Luo CQ, Chen YQ, Marks JR, Waldman F, Wiznerowicz M, Bose R, Chang LW, Beck AH, Gonzalez-Angulo AM, Pihl T, Jensen M, Sfeir R, Kahn A, Chu A, Kothiyal P, Wang ZN, Snyder E, Pontius J, Ayala B, Backus M, Walton J, Baboud J, Berton D, Nicholls M, Srinivasan D, Raman R, Girshik S, Kigonya P, Alonso S, Sanbhadti R, Barletta S, Pot D, Sheth M, Demchok JA, KRM Shaw, Yang LM, Eley G, Ferguson ML, Tarnuzzer RW, Zhang JS, Dillon LAL, Buetow K, Fielding P, Ozenberger BA, Guyer MS, Sofia HJ, Palchik JD. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418): 61–70.

Laddha SV, Ganesan S, Chan CS, White E (2014) Mutational landscape of the essential autophagy gene BECN1 in human cancers. *Mol Cancer Res* **12**(4): 485–490.

Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, Wen W, Delahanty RJ, Lu W, Gao YT, Shen H, Park SK, Chen K, Shen CY, Ren Z, Haiman CA, Matsuo K, Kim MK, Khoo US, Iwasaki M, Zheng Y, Xiang YB, Gu K,

Rothman N, Wang W, Hu Z, Liu Y, Yoo KY, Noh DY, Han BG, Lee MH, Zheng H, Zhang L, Wu PE, Shieh YL, Chan SY, Wang S, Xie X, Kim SW, Henderson BE, Le Marchand L, Ito H, Kasuga Y, Ahn SH, Kang HS, Chan KY, Iwata H, Tsugane S, Li C, Shu XO, Kang DH, Zheng W (2012) Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *Plos Genet* 8(2): e1002532.

Machiela MJ, Ho BM, Fisher VA, Hua X, Chanock SJ (2015) Limited evidence that cancer susceptibility regions are preferential targets for somatic mutation. *Genome Biol* 16: 1–11.

Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010) Genetic susceptibility to breast cancer. *Mol Oncol* 4(3): 174–191.

Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, Wang Q, Dennis J, Dunning AM, Shah M, Luben R, Brown J, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Czene K, Darabi H, Eriksson M, Peto J, dos-Santos-Silva I, Dudbridge F, Johnson N, Schmidt MK, Broeks A, Verhoef S, Rutgers EJ, Swerdlow A, Ashworth A, Orr N, Schoemaker MJ, Figueroa J, Chanock SJ, Brinton L, Lissowska J, Couch FJ, Olson JE, Vachon C, Pankratz VS, Lambrechts D, Wildiers H, Van Ongeval C, Van Limbergen E, Kristensen V, Alnaes GG, Nord S, Borresen-Dale AL, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Trentham-Dietz A, Newcomb P, Titus L, Egan KM, Hunter DJ, Lindstrom S, Tamimi RM, Kraft P, Rahman N, Turnbull C, Renwick A, Seal S, Li JM, Liu JJ, Humphreys K, Benitez J, Zamora MP, Perez JIA, Menendez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dork T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, RAEM Tollenaar, Seynaeve C, van Asperen CJ, Cox A, Cross SS, Reed MWR, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang RX, Schurmann P, Bremer M, Christiansen H, Park-Simon TW, Hillemanns P, Guenel P, Truong T, Menegaux F, Sanchez M, Radice P, Peterlongo P, Manoukian S, Pensotti V, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Brauch H, Bruning T, Ko YD, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer HU, Forsti A, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Chenevix-Trench G, Balleine R, Giles GG, Milne RL, McLean C, Lindblom A, Margolin S, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Eilber U, Wang-Gohrke S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Dieffenbach AK, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Offit K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Perkins KA, Goldberg MS, Labreche F, Dumont M, Eccles DM, Tapper WJ, Rafiq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, Gonzalez-Neira A, Pita G, Alonso MR, Alvarez N, Herrero D, Tessier DC, Vincent D, Bacot F, Luccarini C, Baynes C, Ahmed S, Maranian M, Healey CS, Simard J, Hall P, Easton DF, Garcia-Closas M (2015) Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst* 107(5). pii: djv036.

Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, Darabi H, Brand JS, Bojesen SE, Nordestgaard BG, Flyger H, Nielsen SF, Rahman N, Turnbull C, Fletcher O, Peto J, Gibson L, dos-Santos-Silva I, Chang-Claude J, Flesch-Janys D, Rudolph A, Eilber U, Behrens S, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Khan S, Aaltonen K, Ahsan H, Kibriya MG, Whittemore AS, John EM, Malone KE, Gammon MD, Santella RM, Ursin G, Makalic E, Schmidt DF, Casey G, Hunter DJ, Gapstur SM, Gaudet MM, Diver WR, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Berg CD, Chanock SJ, Figueroa J, Hoover RN, Lambrechts D, Neven P, Wildiers H, van Limbergen E, Schmidt MK, Broeks A, Verhoef S, Cornelissen S, Couch FJ, Olson JE, Hallberg E, Vachon C, Waisfisz Q, Meijers-Heijboer H, Adank MA, van der Luijt RB, Li JM, Liu JJ, Humphreys K, Kang D, Choi JY, Park SK, Yoo KY, Matsuo K, Ito H, Iwata H, Tajima K, Guenel P, Truong T, Mulot C, Sanchez M, Burwinkel B, Marme F, Surowy H, Sohn C, Wu AH, Tseng CC, Van den Berg D, Stram DO, Gonzalez-Neira A, Benitez J, Zamora MP, Perez JIA, Shu XO, Lu W, Gao YT, Cai H, Cox A, Cross SS, Reed MWR, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Lindblom A, Margolin S, Teo SH, Yip CH, Taib NAM, Tan GH, Hooning MJ, Hollestelle A,

Martens JWM, Collee JM, Blot W, Signorello LB, Cai QY, Hopper JL, Southey MC, Tsimiklis H, Apicella C, Shen CY, Hsiung CN, Wu PE, Hou MF, Kristensen VN, Nord S, Alnaes GIG, Giles GG, Milne RL, McLean C, Canzian F, Trichopoulos D, Peeters P, Lund E, Sund M, Khaw KT, Gunter MJ, Palli D, Mortensen LM, Dossus L, Huerta JM, Meindl A, Schmutzler RK, Sutter C, Yang R, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Hartman M, Miao H, Chia KS, Chan CW, Fasching PA, Hein A, Beckmann MW, Haeberle L, Brenner H, Dieffenbach AK, Arndt V, Stegmaier C, Ashworth A, Orr N, Schoemaker MJ, Swerdlow AJ, Brinton L, Garcia-Closas M, Zheng W, Halverson SL, Shrubsole M, Long J, Goldberg MS, Labreche F, Dumont M, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Bruning T, Radice P, Peterlongo P, Manoukian S, Bernard L, Bogdanova NV, Dork T, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Devilee P, Tollenaar RA, Seynaeve C, Van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Huzarski T, Sangrajrang S, Gaborieau V, Brennan P, Mckay J, Slager S, Toland AE, Ambrosone CB, Yannoukakos D, Kabisch M, Torres D, Neuhausen SL, Anton-Culver H, Luccarini C, Baynes C, Ahmed S, Healey CS, Tessier DC, Vincent D, Bacot F, Pita G, Alonso MR, Alvarez N, Herrero D, Simard J, Pharoah PP, Kraft P, Dunning AM, Chenevix-Trench G, Hall P, Easton DF. BOCS, kConFab Investigators; AOCS Group, GENICA Network (2015) Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* 47(4): 373–380.

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Silva ID, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Czene K, Irwanto A, Liu JJ, Waisfisz Q, Meijers-Heijboer H, Adank M, Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Muller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LFA, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li JM, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang XS, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens MR, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Van't Veer LJ, van der Schoot CE, Guenel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinke B, Zamora MP, Perez JIA, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MWR, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, kConFab Investigators; Australian Ovarian Cancer Study Group, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AMW, Jager A, Bui QM, Stone J, Dite GS, Apicella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Muller H, Arndt V, Stegmaier C, Swerdlown A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labreche F, Dumont M, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Bruning T, GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Radice P, Peterlongo P, Manouldan S, Bonanni B, Devilee P, RAEM Tollenaar, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dork T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo KY, Noh DY, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng CC, Van den Berg D, Stram DO, Shu XO, Lu W, Gao YT, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng JH, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen CY, Hsiung CN, Wu PE, Ding SL, Sangrajrang S, Gaborieau V, Brennan P, Mckay J, Blot WJ, Signorello LB, Cai QY, Zheng W, Deming-Halverson S, Shrubsole M, Long JR, Simard J, Garcia-Closas M, Pharoah PDP, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45(4): 353–361.

Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menendez-Rodriguez P, Hardisson D, Mendiola M, Gonzalez-Neira A, Pita G, Alonso MR, Dennis J, Wang Q, Bolla MK, Swerdlow A, Ashworth A, Orr N, Schoemaker M, Ko YD, Brauch H, Hamann U.

GENICA Network, Andrulis IL, Knight JA, Glendon G, Tchatchou S, kConFab Investigators; Australian Ovarian Cancer Study Group, Matsuo K, Ito H, Iwata H, Tajima K, Li J, Brand JS, Brenner H, Dieffenbach AK, Arndt V, Stegmaier C, Lambrechts D, Peuteman G, Christiaens MR, Smeets A, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Hartman M, Hui M, Yen Lim W, Wan Chan C, Marme F, Yang R, Bugert P, Lindblom A, Margolin S, Garcia-Closas M, Chanock SJ, Lissowska J, Figueroa JD, Bojesen SE, Nordestgaard BG, Flyger H, Hooning MJ, Kriege M, van den Ouweland AM, Koppert LB, Fletcher O, Johnson N, dos-Santos-Silva I, Peto J, Zheng W, Deming-Halverson S, Shrubsole MJ, Long J, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Cox A, Cross SS, Reed MW, Schmidt MK, Broeks A, Cornelissen S, Braaf L, Kang D, Choi JY, Park SK, Noh DY, Simard J, Dumont M, Goldberg MS, Labreche F, Fasching PA, Hein A, Ekici AB, Beckmann MW, Radice P, Peterlongo P, Azzollini J, Barile M, Sawyer E, Tomlinson I, Kerin M, Miller N, Hopper JL, Schmidt DF, Makalic E, Southey MC, Hwang Teo S, Har Yip C, Sivanandan K, Tay WT, Shen CY, Hsiung CN, Yu JC, Hou MF, Guenel P, Truong T, Sanchez M, Mulot C, Blot W, Cai Q, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Bogdanova N, Dork T, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Shu XO, Lu W, Gao YT, Zhang B, Couch FJ, Toland AE. TNBCCYannoukakos D, Sangrajrang S, McKay J, Wang X, Olson JE, Vachon C, Purrington K, Severi G, Baglietto L, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Devilee P, Tollenaar RA, Seynaeve C, Czene K, Eriksson M, Humphreys K, Darabi H, Ahmed S, Shah M, Pharoah PD, Hall P, Giles GG, Benitez J, Dunning AM, Chenevix-Trench G, Easton DF (2014) Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* **23**(22): 6096–6111.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jonsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerod A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Borresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR (2012) Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**(5): 979–993.

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn S-M, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GKJ, Jang SJ, Jones DR, Kim H-Y, King TA, Krishnamurthy S, Lee HJ, Lee J-Y, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, Lvt Veer L, Tutt A, Knappskog S, BKT Tan, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JWM, Børresen-Dale A-L, Richardson AL, Kong G, Thomas G, Stratton MR (2016) Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**(7605): 47–54.

Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P (2002) Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**(48): 7435–7451.

Pfeifer GP, You YH, Besaratinia A (2005) Mutations induced by ultraviolet light. *Mutat Res* **571**(1-2): 19–31.

Rahman N (2014) Realizing the promise of cancer predisposition genes. *Nature* **505**(7483): 302–308.

Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G, Gordenin DA (2013) An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**(9): 970.

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, McPherson A, Shumansky K, Crisan A, Giuliany R, Heravi-Moussavi A, Rosner J, Lai D, Birol I, Varhol R, Tam A, Dhalla N, Zeng T, Ma K, Chan SK, Griffith M, Moradian A, Cheng SW, Morin GB, Watson P, Gelmon K, Chia S, Chin SF, Curtis C, Rueda OM, Pharoah PD, Damaraju S, Mackey J, Hoon K, Harkins T, Tadigotla V, Sigaroudinia M, Gascard P, Tlsty T, Costello JF, Meyer IM, Eaves CJ, Wasserman WW, Jones S, Huntsman D, Hirst M, Caldas C, Marra MA, Aparicio S (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**(7403): 395–399.

Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, Michailidou K, Stram DO, Beckmann L, Rhie SK, Ambrosone CB, Aittomaki K, Amiano P, Apicella C. Australian Breast Cancer Tissue Bank InvestigatorsBaglietto L, Bandera EV, Beckmann MW, Berg CD, Bernstein L, Blomqvist C, Brauch H, Brinton L, Bui QM, Buring JE, Buys SS, Campa D, Carpenter JE, Chasman DI, Chang-Claude J, Chen C, Clavel-Chapelon F, Cox A, Cross SS, Czene K, Deming SL, Diasio RB, Diver WR, Dunning AM, Durcan L, Ekici AB, Fasching PA. Familial Breast Cancer StudyFeigelson HS, Fejerman L, Figueroa JD, Fletcher O, Flesch-Janys D, Gaudet MM. GENICA consortiumGerty SM, Rodriguez-Gil JL, Giles GG, van Gils CH, Godwin AK, Graham N, Greco D, Hall P, Hankinson SE, Hartmann A, Hein R, Heinz J, Hoover RN, Hopper JL, Hu JJ, Huntsman S, Ingles SA, Irwanto A, Isaacs C, Jacobs KB, John EM, Justenhoven C, Kaaks R, Kolonel LN, Coetzee GA, Lathrop M, Le Marchand L, Lee AM, Lee IM, Lesnick T, Lichtner P, Liu J, Lund E, Makalic E, Martin NG, McLean CA, Meijers-Heijboer H, Meindl A, Miron P, Monroe KR, Montgomery GW, Muller-Myhsok B, Nickels S, Nyante SJ, Olswold C, Overvad K, Palli D, Park DJ, Palmer JR, Pathak H, Peto J, Pharoah P, Rahman N, Rivadeneira F, Schmidt DF, Schmutzler RK, Slager S, Southey MC, Stevens KN, Sinn HP, Press MF, Ross E, Riboli E, Ridker PM, Schumacher FR, Severi G, Dos Santos Silva I, Stone J, Sund M, Tapper WJ, Thun MJ, Travis RC, Turnbull C, Uitterlinden AG, Waisfisz Q, Wang X, Wang Z, Weaver J, Schulz-Wendtland R, Wilkens LR, Van Den Berg D, Zheng W, Ziegler RG, Ziv E, Nevanlinna H, Easton DF, Hunter DJ, Henderson BE, Chanock SJ, Garcia-Closas M, Kraft P, Haiman CA, Vachon CM (2012) A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* **21**(24): 5373–5384.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerod A. Oslo Breast Cancer Consortium (OSBREAC)Lee MT, Shen CY, Tee BT, Huimin BW, Broeks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, van 't Veer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis-Filho JS, Aparicio SA, Salomon AV, Borresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, Stratton MR (2012) The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**(7403): 400–404.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)