



Published in final edited form as:

Breast Cancer Res Treat. 2017 August ; 164(3): 707–717. doi:10.1007/s10549-017-4287-4.

Limited influence of germline genetic variation on all-cause mortality in women with early onset breast cancer: Evidence from gene-based tests, single-marker regression, and whole-genome prediction

Molly Scannell Bryan^{1,2}, Maria Argos², Irene L. Andrulis³, John L. Hopper⁴, Jenny Chang-Claude^{5,6}, Kathleen Malone⁷, Esther M. John^{8,9}, Marilie D. Gammon¹⁰, Mary Daly¹¹, Mary Beth Terry¹², Sandra S. Buys¹³, Dezheng Huo¹, Olofunmilayo Olopade¹, Jeanine M Genkinger¹², Farzana Jasmine¹, Muhammad G. Kibriya¹, Lin Chen¹, and Habibul Ahsan¹

¹University of Chicago, Chicago, Illinois

²University of Illinois at Chicago, Chicago, Illinois

³Lunefeld-Tanenbaum Research Institute, Sinai Health System and Department of Molecular Genetics, University of Toronto, Toronto, Ontario

⁴University of Melbourne, Melbourne, Australia

⁵Deutsches Krebsforschungszentrum in der Helmholtz-Gemeinschaft, Hamburg, Germany

⁶University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁷Fred Hutchinson Cancer Research Center, Seattle, Washington

⁸Cancer Prevention Institute of California, Fremont, California

⁹Stanford University School of Medicine, Stanford, California

¹⁰University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

¹¹Fox Chase Cancer Center, Philadelphia, Pennsylvania

¹²Columbia University, New York, New York

¹³University of Utah, Salt Lake City, Utah

Abstract

Purpose—Women diagnosed with breast cancer have heterogeneous survival outcomes that cannot be fully explained by known prognostic factors, and germline variation is a plausible but unconfirmed risk factor for poor survival outcomes.

CORRESPONDING AUTHOR CONTACT INFORMATION: Molly Scannell Bryan, University of Illinois at Chicago, 1747 W. Roosevelt Rd., Chicago, IL 60608-1264, (312)355-2328, fax: (312)355-2328, scannemo@uic.edu.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee, and with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study.

Methods—We used three approaches to test the hypothesis that germline variation drives some differences in survival: mortality loci identification, tumor aggressiveness loci identification, and whole-genome prediction. The 2954 study participants were women diagnosed with breast cancer before age 50, with a median follow-up of fifteen years. We first aimed to identify loci in gene regions that were associated with all-cause mortality. We next aimed to identify loci in gene regions associated with five histopathological characteristics related to tumor aggressiveness. We also predicted ten-year all-cause mortality on a subset of 1903 participants genotyped on genome-wide arrays (3,245,343 variants after imputation) using whole genome prediction methods.

Results—No risk loci for mortality or tumor aggressiveness were identified. This null result persisted when restricting to women with estrogen receptor positive tumors, when examining suggestive loci in an independent study, and when restricting to previously published risk loci. Additionally, the whole-genome prediction model also found no evidence to support an association.

Conclusion—Despite multiple complementary approaches, our study found no evidence that mortality in women with early onset breast cancer is influenced by germline variation.

Keywords

early onset breast cancer; single nucleotide polymorphisms; whole-genome prediction; skat-o; gene-based tests; survival

INTRODUCTION

While treatment and survival rates for women diagnosed with breast cancer have improved over time, [1] almost twenty five percent of the 250,000 American women who are diagnosed with breast cancer annually eventually die of the disease. [2,3] Mortality in women with breast cancer is associated with several factors, including stage at detection, socioeconomic factors, tumor characteristics, and treatment decisions. [4,5] However, after taking these into account, differences in survival persist. [4] Several lines of evidence, including animal and epidemiological studies, [6–8] suggest that germline genetic variation may influence the unexplained heterogeneity in mortality.

Identifying locations in the genome in which variation is associated with mortality can implicate cellular processes that are involved in oncogenesis, and may suggest targets for future pharmaceutical interventions. However, since there are many biologically plausible pathways that connect germline variation and mortality, and any single method may not be optimal to identify all regions of the genome that influence mortality trajectories. If the putative mortality-associated variant is common in the general population and has a moderately strong effect, it can often be identified through single marker regression approaches common in genome-wide association studies (GWAS). Several GWASs have been published that investigated associations with breast cancer mortality (summarized in supplemental text), but the results have largely been null or poorly replicated. Most studies were small (all but one meta-analysis recruited fewer than 2000 participants), and were only able to follow women for a short period of time relative to the expected median survival of women after breast cancer diagnosis (none were followed longer than seven years). In

contrast to GWASs, gene-based tests such as the sequence kernel association test-optimal (SKAT-O [9]) use a different approach and allow variants within a gene region to collectively contribute evidence for association, even if the putative variant is too rare to be tested individually, or has an effect that is too small to be detected with the strict significance thresholds necessitated by GWASs. No breast cancer mortality study has reported the results of gene-based tests genome-wide.

In addition to identifying individual loci that are associated with mortality, it is also of clinical interest to predict mortality in women with breast cancer to better identify those who may want to choose aggressive treatment, and also women who may more confidently choose to forgo toxic therapies. [10] Previous studies have attempted to incorporate germline variation into prediction models of breast cancer mortality using both candidate gene and polygenic risk score approaches; these models have produced mixed results. [11,12] If mortality in women with breast cancer is a polygenic trait, genetic restricted maximum likelihood prediction methods, [13] which allow all measured genetic variation to contribute to prediction without excluding variants that meet effect size thresholds, may improve upon these other method's ability to predict mortality, but this has not yet been applied to studies of breast cancer mortality.

The largely null and inconsistent results from previous studies suggest the hypothesis that germline variation is not strongly associated with mortality in women who have been diagnosed with breast cancer. The objective of this manuscript is to look for evidence that germline variation influences mortality in women with breast cancer by applying single-marker regressions, gene-based tests, and whole-genome prediction to a sample of women diagnosed with breast cancer before the age of fifty. The primary outcome of interest is all-cause mortality, and we also considered histopathological markers of tumor aggressiveness as secondary outcomes.

MATERIALS AND METHODS

Participants

The participants were enrolled in one of five studies originally designed to assess the factors associated with early onset breast cancer incidence and prognosis, summarized in supplemental Table 3. The participants were women of European descent without pathogenic mutations in the *BRCA1* or *BRCA2* genes. Ninety-eight percent were younger than 50 years at the time of their diagnosis. Details of recruitment and data collection are found in the supplemental text.

Any suggestive findings from the primary analyses were examined for evidence of replication in data provided by the participants of The Cancer Genome Atlas (TCGA [14]) breast cancer study. Single nucleotide variant (SNV) data were downloaded from the TCGA data portal in June 2015, and clinical data were downloaded in September 2016.

Genotyping

Details of the genotyping, imputation, and quality control are found in the supplemental text and in supplemental Figures 1 and 2.

Briefly, for the 3232 participants enrolled in the primary studies, germline DNA was extracted from peripheral blood. All were genotyped on the Illumina HumanExome 12 v1 exome array (238,524 variants). If the variant passed variant-level quality control (4335 variants excluded), was polymorphic in this population (135,931 included), and located in a gene region (ie: excluding intergenic variants, as annotated by ANNOVAR [15]), it was included in the mortality loci identification analyses (114,206 included). The quality control also excluded 278 women, resulting in 2954 women available for these analyses.

Of the 3232 women genotyped on the exome array, a subset of 2323 was additionally genotyped using the Illumina 610-Quad and Cyto12 v2 BeadChips, (555,259 variants interrogated; 3,310,158 variants after imputation to the HapMap3 [16]). The quality control process excluded 325 women with either no survival information or who did not pass QC (1998 women available). The exome and genome-wide variants were combined (3,245,343 variants).

The TCGA data included the 768 participants of the TCGA breast cancer study who were females of European ancestry. After quality control and imputation (details in supplemental text) to the 1000 Genomes phase 3[17], 711 women and 15,121,555 variants were available for replication.

Primary outcome: All-cause mortality

All-cause mortality was determined within each study through chart review, telephone interviews, linkage to cancer registries, and linkage mortality databases.

Statistical approach for identifying mortality risk loci—To identify variants or gene regions associated with the hazard of all-cause mortality, both single marker regression GWASs and gene-based SKAT-O Cox proportional hazards models were applied to the 2954 women who were assayed on the exome array. The GWASs were implemented using the GenABEL [18,19] package for the R software. [20] Only common variants on the exome array were included in the GWASs (minor allele frequency above the threshold of $(1/2n)^{1/2} = 0.0130$ following Wu et al. [21]; 25,938 variants). The SKAT-O analyses were implemented using the skatMeta R package, [22] after annotation to 16,317 genes by ANNOVAR. Each variant was weighted by the combined annotation dependent depletion (CADD) [23] scaled score of the predicted deleteriousness of the minor allele. The twenty variants or genes with the smallest p-value in the primary analyses were examined for evidence of replication in the TCGA data.

The relationship between germline variation and mortality in women with estrogen receptor positive (ER+) tumors was of particular interest due to a plausible but unconfirmed pharmacogenomic pathway to mortality. During the period that the women in this study were diagnosed, approximately 80% of those with ER+ tumors would have been treated with tamoxifen, [24,25] whose active metabolites are known to be influenced by germline variation. It is unclear whether this influences survival, and there has been conflicting evidence specifically on the effect of variation within the tamoxifen-metabolizing gene *CYP2D6* on survival. [26–28] To investigate this, both analyses were repeated after restricting to the women known to have ER+ tumors (n=1066).

Given the low rates of replication of previously identified mortality risk loci, it was also of interest to examine whether our analyses provided evidence of replication. The National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS catalog [29] was searched in January 2017 and variants identified with breast cancer mortality were identified. Genes containing these variants were examined in our analysis.

Additionally, it is not known whether risk loci that are associated with other breast cancer phenotypes may also be associated with mortality. If that were the case, it would suggest a shared genetic etiology between the other breast cancer phenotypes (primarily breast cancer incidence, but also mammographic density, and other breast tissue characteristics) and mortality. To look for evidence of this, the variants listed in the NHGRI-EBI catalog as being associated with any breast cancer phenotype were examined in our analysis if the p-value of that variant was genome-wide significant ($p < 5 \cdot 10^{-8}$).

Statistical approach for whole genome prediction of ten-year mortality—Of the 1998 women with exome and genome-wide assays, 1903 had complete survival information for ten years after diagnosis and were included. Of these, 400 (27%) died within ten years of diagnosis. This ten-year mortality status was predicted using the R package OmicKriging, [13] and the genetic relatedness matrix was created using all 3,245,343 variants by the GCTA software. [30] Using ten-fold cross validation, the Kriging formula estimated a linear predicted probability of mortality for each woman. These predictions were compared to that woman's actual mortality status to compute an area under the receiver operating characteristic (AUC). This procedure was repeated two hundred times to produce valid confidence intervals for the AUC.

Secondary outcome: Tumor characteristics

Molecular and histopathological properties of the tumor are evaluated at diagnosis, and women with more aggressive tumor profiles generally have worse survival trajectories. [31,32] To identify loci that are associated with any of the more aggressive tumor profiles, five logistic regressions were run to examine whether germline variation in gene regions could predict ER status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status, high tumor grade (three or higher), or high tumor stage (three or higher) using the SKAT package for R. [9,20] ER status was available for 1785 women, PR status for 1769, HER2 status for 658, tumor grade for 1790, and tumor stage for 1640.

Covariates used in the mortality risk loci identification analyses

To counter the potential confounding between genetic ancestry and mortality, principal components constructed from both common and rare variants were also included as covariates in each single marker regression and gene-based analysis. Details of the selection of the principal components are in the supplemental text. Study center was additionally included as a covariate.

RESULTS

The characteristics of the women that were included in each investigation after quality control are found in Table 1.

Common variants in gene regions and all-cause mortality: Single marker regression

The single marker regressions identified no common variants that were associated with mortality with a p-value smaller than the Bonferroni-corrected significance level of $1.8 \cdot 10^{-6}$ either in all women with breast cancer, or those with ER+ tumors. Quantile-quantile (QQ) plots of the association results are found in Figure 1A and 1B. None of the twenty most significant variants in the primary analyses were associated with mortality in the TCGA analysis with a p-value smaller than Bonferroni-corrected significance level of 0.0025 (supplemental Tables 4 and 5).

Twenty-three SNVs listed in the NHGRI-EBI GWAS catalog as significantly associated with any breast cancer phenotype were directly interrogated by the exome array. None had a p-value less than the Bonferroni-corrected level of $0.05/23 = 0.0022$ in either the overall or ER + only analysis, as shown in supplemental Table 6.

Gene regions and all-cause mortality: Gene-based tests

The gene-based tests identified no gene regions in which variation was associated with the hazard of mortality either in all women, or those with ER+ tumors with a p-value smaller than the Bonferroni-corrected significance level of $3.06 \cdot 10^{-6}$. QQ plots are shown in Figures 2A and 2B. None of the twenty gene regions with the smallest p-values in the primary analyses were associated with mortality in the TCGA population with a p-value smaller 0.0025 in the TCGA analysis (supplemental Tables 7 and 8). To determine if the results were sensitive to the weighting method, the analyses were repeated using weights that were a beta transformation of the minor allele frequency (as suggested by the SKAT authors⁹), and equal weights. These two additional weighting methods produced substantively similar null results.

Our mortality analyses measured four nonsynonymous SNVs within the gene *CYP2D6* on chromosome 22 (positions in the HG19 assembly: 42523844, 42523855, 42523975, 42524817), and found no evidence of association with mortality either in all women ($p = 7.81 \cdot 10^{-2}$) or in women with ER+ tumors ($p = 7.75 \cdot 10^{-1}$).

The NHGRI-EBI GWAS catalog lists variants in five gene regions that were also interrogated by the exome array as associated with breast cancer mortality. None of the genes were statistically significant in either mortality analysis (supplemental Table 9).

The NHGRI-EBI GWAS catalog lists variants in fifty-three assayed gene regions that were significantly associated with any breast cancer phenotype. Of these, none were associated with mortality with a p-value smaller than Bonferroni-corrected significance level for replication of previously identified loci of $p < 0.05/53 = 9.43 \cdot 10^{-4}$ (supplemental Tables 10 and 11).

Gene regions and tumor characteristics: Gene-based tests

No associations with any of the five characteristics of tumor aggressiveness reached the level of genome-wide significance, and the TCGA data did not provide any evidence of replication for twenty gene regions with the smallest p-value in the primary analyses (Figure 2C–G, and supplemental Tables 12 through 16).

The p-values of the NHGRI-EBI breast cancer gene regions for each tumor characteristic are summarized in supplemental Tables 17 and 18). Only one gene in the tumor characteristic analyses met a Bonferroni corrected threshold of significance for replication of the previous loci: *SLC4A7* ($p = 8.35 \cdot 10^{-4}$ for PR status). A variant located at chromosome 3:27416013 (the three prime untranslated region of *SLC4A7*) was reported in one previous study ($p = 2 \cdot 10^{-8}$), [33] and one meta-analysis which included that original study ($p = 2 \cdot 10^{-30}$) [34] as being associated with breast cancer incidence in women of European descent. Variants in the adjacent gene of *NEK10* were previously associated with incidence, both in our own analysis (under review) and previous studies. [35,36]

Ten-year mortality: Whole genome prediction

The AUC of the whole genome prediction was not significantly different than chance (AUC: 0.493, 95% CI: 0.479–0.510). The performance of Kriging prediction has been shown to improve if separate genetic relatedness matrices are constructed that group variants with similar association strengths together. [13] For this reason, the Kriging was repeated with multiple matrices three times: separating variants into two matrices based on rareness (minor allele frequency above and below 0.0127), three matrices based on their predicted functionality (annotation by ANNOVAR into (1) intergenic, (2) gene regions, but no change to amino acid translation, and (3) predicted to change amino acid translation), and the six matrices that represented the cross product of variant rareness and predicted functionality. In each of these models, the null result persisted.

For comparison, GCTA was used to estimate the proportion of variation in ten-year mortality that was associated with genotypic variation (“heritability”) using a single genetic relatedness matrix with all 3,245,343 variants. The heritability estimate was near zero with a large standard error (estimate: 0.000002, standard error: 0.2). Also for comparison, a polygenic risk score prediction was undertaken using the 81 SNVs that were measured in either the exome array or genome-wide array and listed in the NHGRI-EBI GWAS catalog as associated with any breast cancer phenotype. The predictive power of this polygenic risk score was also not different from chance (AUC 0.484; 95% CI from 2000 bootstrap replications: 0.453–0.516).

DISCUSSION

Multiple complementary analyses found no evidence that mortality in women diagnosed with breast cancer is influenced by germline variation. Our analysis included single marker regressions and gene-based tests of mortality and markers of tumor aggressiveness, and also whole genome prediction, heritability estimates, and polygenic risk score analyses of mortality. Each has its own limitations, but each returned null results. When combined with

our inability to confirm the mortality risk loci identified by previous GWASs, this suggests that unlike breast cancer incidence, [37] mortality in women diagnosed with breast cancer may not be strongly influenced by germline variation.

Our study represents the first use of several methods to investigate mortality in women with breast cancer, which we hypothesized would capture associations that may have been missed by previous research. We directly investigated for the first time the role of rare exonic variants, and were the first to use gene-based tests. While restricted to variants within gene regions, the results of the single marker regressions in the primary analyses was the longest follow-up in a GWAS of mortality in breast cancer. Our prediction model is the first to incorporate the influence of all germline variation.

Our sample included women who were diagnosed with breast cancer before age 50, and is the first genome-wide investigation of mortality that specifically targeted this young age group. This provided a unique insight into the survival trajectory of the one in five American women with breast cancer who are diagnosed that early. [3] Although recent work suggested that the genetic etiology of breast cancer incidence may be similar across ages of diagnosis, [35] the full relationship between age and genetic etiology of breast cancer is still not well-characterized. Therefore, the generalizability of these null results to women who were diagnosed later in life may be limited.

Our analysis did not find that gene regions that had been reproducibly associated with breast cancer incidence were also associated with mortality, suggesting a separate genetic etiology of incidence and mortality in breast cancer. To a large extent, the analysis did not support the hypothesis that incidence risk loci are associated with characteristics of tumor aggressiveness. One gene that contained a previous incidence risk locus, *SLC4A7/NEK10*, was associated with the PR status of the tumor in our analysis with a p-value that was slightly smaller the Bonferroni-corrected threshold, but the borderline significance of this association should be interpreted cautiously in light of the multiple hypotheses tested.

Our analysis focused on all-cause mortality. While this decision allowed for germline variation to beyond breast cancer-specific mortality (for example, susceptibility to cardiotoxic side effects from chemotherapy), it is possible that our analysis obscured associations with cause-specific mortality. While a sensitivity analysis would have been preferable, in our sample, breast cancer-specific mortality was not available for all participants.

Previous epidemiologic and animal studies suggested that germline variation does influence mortality in women with breast cancer. [6–8] While the discrepancy between past work and ours may be due to study design (many of the epidemiologic studies were carried out using a family-based design, which may be prone to bias from shared environment), our results do not rule out the possibility that genetic variation influences mortality. However, our results do suggest possible constraints on the genetic architecture of that association. If mortality were driven by variants that were not in strong linkage disequilibrium with any variants measured by the two array-based methods used here, these analyses would have limited ability to detect their association. While whole-genome sequencing would have been

preferable to comprehensively interrogate all variation, the exome and genome-wide arrays selected for use in this study were designed to well-interrogate much of the genome in people of European descent, and therefore only variants with a low frequency would have gone unmeasured. Thus, our results suggest that if germline variation were associated with survival in women with breast cancer, it would likely be characterized by one of two descriptions: large effect size, but so rare as to not have much effect on overall mortality, or small effect size.

If the putative variants were rare variants of large effect size, this suggests that family-based studies would be more appropriate than GWASs to identify them. If the putative variants were of weak effect (SKAT power calculations estimated we achieved 80% power in our sample size when detecting an association with a hazard ratio of 2.1, although our study may have lost power due to our replication data not fully matching our primary analyses in), they would be very difficult to detect. The identification of such variants may be unlikely to highlight mechanisms that are necessary or sufficient for mortality in women with breast cancer, and even the cumulative influence of multiple such variants would be unlikely to add to our ability to identify high-risk women with breast cancer at the population level.

Our findings suggest that future investment into studying the inherited nature of mortality in women with breast cancer may be most fruitful if it focuses on interaction analyses. This approach may identify germline variants that affect survival trajectories for a subset of women, by focusing on women where this association is already suggested, or focusing on women who have not been well-represented in survival studies to date. For example, recent research has suggested that other germline genetic variants may be associated with mortality in women with pathogenic mutations in the genes *BRCA2* and *TP53*. [38,39] Beyond gene-by-gene interactions, germline variation may have a larger influence on breast cancer mortality in populations that have different background risk factors such as ancestry (this sample was of European ancestry), or country of origin (this sample was recruited from affluent countries). Germline variation may also influence mortality by way of an interaction with treatment [40], which was not consistently available for our participants. Given the young age at diagnosis, the participants may have been treated more aggressively, compared to those diagnosed at a later age who may have had more comorbidities, and it is possible that germline variation has less of an impact on survival in the presence of aggressive surgery or treatment.

In conclusion, while germline genetic variation may still be associated with survival trajectories in women with early onset breast cancer, our multi-approach study was unable to find evidence of this.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

FUNDING

This study was supported by the National Institute of Health grants R01CA122171, RC1CA145506, U01CA122171, R01CA094069, UM1CA164920, RFA-CA-95-011, UO1CA/ES66572, UO1CA66572, U19CA148065, and R25-CA057699.

Portions of this manuscript have been included in the doctoral thesis of Molly Scannell Bryan at the University of Chicago (under embargo until 2019). The authors would like to thank Regina M. Santella of Columbia University. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. Samples from the CPIC were processed and distributed by the Coriell Institute for Medical Research. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

ABBREVIATIONS USED

AUC	area under the receiver operating characteristic
CADD	combined annotation dependent depletion
EBI	European Bioinformatics Institute
ER	estrogen receptor
GWAS	genome wide association study
HER2	human epidermal growth factor receptor 2
NHGRI	National Human Genome Research Institute
PR	progesterone receptor
QQ	quantile-quantile
SKAT	sequence kernel association test
SNV	single nucleotide variant
TCGA	The Cancer Genome Atlas

References

1. Nakano M, Fujisue M, Tashima R, Okumura Y, Nishiyama Y, Ohsako T, Toyozumi Y, Arima N, Nishimura R. Survival time according to the year of recurrence and subtype in recurrent breast cancer. *Breast*. Oct; 2015 24(5):588–593. [PubMed: 26099804]
2. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians*. Feb; 2016 66(1):31–42. [PubMed: 26513636]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*. Jan; 2017 67(1):7–30. [PubMed: 28055103]
4. Tao L, Gomez SL, Keegan THM, Kurian AW, Clarke CA. Breast cancer mortality in African-American and non-hispanic white women by molecular subtype and stage at diagnosis: A population based study. *Cancer Epidemiology, Biomarkers & Prevention*. Jul; 2015 24(7):1039–1045.
5. Menashe I, Anderson WF, Jatoi I, Rosenberg PS. Underlying causes of the black-white racial disparity in breast cancer mortality: a population-based analysis. *Journal of the National Cancer Institute*. Jul; 2009 101(14):993–1000. [PubMed: 19584327]

6. Bai L, Yang HH, Hu Y, Shukla A, Ha NH, Doran A, Faraji F, Goldberger N, Lee MP, Keane T, Hunter KW. An integrated genome-wide systems genetics screen for breast cancer metastasis susceptibility genes. *PLOS Genet.* Apr.2016 12(4):e1005989. [PubMed: 27074153]
7. Lindström LS, Hall P, Hartman M, Wiklund F, Grönberg H, Czene K. Familial concordance in cancer survival: a Swedish population-based study. *The Lancet. Oncology.* Nov; 2007 8(11):1001–1006. [PubMed: 17921068]
8. Hartman M, Lindström L, Dickman PW, Adami HO, Hall P, Czene K. Is breast cancer prognosis inherited? *Breast Cancer Research.* 2007; 9(3):R39. 13. [PubMed: 17598882]
9. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani DC, Wurfel MM, Lin X. NHLBI GO Exome Sequencing Project ESP Lung Project Team. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics.* Aug; 2012 91(2):224–237. [PubMed: 22863193]
10. Schmidt M, Victor A, Bratzel D, Boehm D, Cotarelo C, Lebrecht A, Siggelkow W, Hengstler JG, Elsässer A, Gehrman M, Lehr HA, Koelbl H, Minckwitz Gv, Harbeck N, Thomssen C. Longterm outcome prediction by clinicopathological risk classification algorithms in node-negative breast cancercomparison between Adjuvant!, St Gallen, and a novel risk algorithm used in the prospective randomized Node-Negative-Breast Cancer-3 (NNBC-3) trial. *Annals of Oncology.* Feb; 2009 20(2):258–264. [PubMed: 18824499]
11. Maas P, Barndahl M, Joshi AD, et al. Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. *JAMA Oncology.* May.2016
12. 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, Ongeval CV, Limbergen Ev, Kristensen V, Alnæs GG, Nord S, Borresen-Dale AL, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Fasching PA, Haerle 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 J, Liu J, Humphreys K, Benitez J, Zamora MP, Perez JIA, Menéndez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Bogdanova NV, Antonenkova NN, Dörk T, Anton-Culver H, Neuhausen SL, Ziogas A, Bernstein L, Devilee P, Tollenaar RAEM, Seynaeve C, Asperen CJv, Cox A, Cross SS, Reed MWR, Khusnutdinova E, Bermisheva M, Prokofyeva D, Takhirova Z, Meindl A, Schmutzler RK, Sutter C, Yang R, Schürmann P, Bremer M, Christiansen H, Park-Simon TW, Hillemanns P, Guénel 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, Brüning T, Ko YD, Sigurdson AJ, Doody MM, Hamann U, Torres D, Ulmer HU, Försti 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, Marchand LL, Eilber U, Wang-Gohrke S, Hoening MJ, Hollestelle A, Ouweland AMWvd, Koppert LB, Carpenter J, Clarke C, Scott R, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Brenner H, Arndt V, Stegmaier C, Diebnach AK, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, O-t K, Vijai J, Robson M, Rau-Murthy R, Dwek M, Swann R, Perkins KA, Goldberg MS, Labrèche F, Dumont M, Eccles DM, Tapper WJ, Raq S, John EM, Whittemore AS, Slager S, Yannoukakos D, Toland AE, Yao S, Zheng W, Halverson SL, González-Neira A, Pita G, Alonso MR, Álvarez 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. Prediction of breast cancer risk based on profiling with common genetic variants. *Journal of the National Cancer Institute.* May.2015 107(5):djv036. [PubMed: 25855707]
13. Wheeler HE, Aquino-Michaels K, Gamazon ER, Trubetskoy VV, Dolan ME, Huang RS, Cox NJ, Im HK. Poly-omic prediction of complex traits: OmicKriging. *Genetic Epidemiology.* Jul.2014 38(5):402415. arXiv: 1303.1788.
14. [Accessed 5 April 2017] The Cancer Genome Atlas Home Page. <https://cancergenome.nih.gov/>

15. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*. Oct.2015 10(10):15561566.
16. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang LY, Huang W, Liu B, Shen Y, Tam PKH, Tsui LC, Waye MMY, Wong JTF, Zeng C, Zhang Q, Chee MS, Galver LM, Kruglyak S, Murray SS, Oliphant AR, Montpetit A, Hudson TJ, Chagnon F, Ferretti V, Leboeuf M, Phillips MS, Verner A, Kwok PY, Duan S, Lind DL, Miller RD, Rice JP, Saccone NL, Taillon-Miller P, Xiao M, Nakamura Y, Sekine A, Sorimachi K, Tanaka T, Tanaka Y, Tsunoda T, Yoshino E, Bentley DR, Deloukas P, Hunt S, Powell D, Altshuler D, Gabriel SB, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Aniagwu T, Marshall PA, Matthew O, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M, Stein LD, Cunningham F, Kanani A, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Donnelly P, Marchini J, McVean GAT, Myers SR, Cardon LR, Abecasis GR, Morris A, Weir BS, Mullikin JC, Sherry ST, Feolo M, Daly MJ, Schaner SF, Qiu R, Kent A, Dunston GM, Kato K, Niikawa N, Knoppers BM, Foster MW, Clayton EW, Wang VO, Watkin J, Sodergren E, Weinstock GM, Wilson RK, Fulton LL, Rogers J, Birren BW, Han H, Wang H, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Todani K, Fujita T, Tanaka S, Holden AL, Lai EH, Collins FS, Brooks LD, McEwen JE, Guyer MS, Jordan E, Peterson JL, Spiegel J, Sung LM, Zacharia LF, Kennedy K, Dunn MG, Seabrook R, Shillito M, Skene B, Stewart JG, DLV, EWC, LBJ, Cho MK, Duster T, Jaspere M, Licinio J, Long JC, Ossorio PN, Spallone P, Terry SF, (chair) ESL, EHL, DAN, Boehnke M, Douglas JA, Hudson RR, Kruglyak L, Nussbaum RL. The International HapMap Project. *Nature*. Dec; 2003 426(6968):789–796. [PubMed: 14685227]
17. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. Oct; 2015 526(7571):68–74. [PubMed: 26432245]
18. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. May; 2007 23(10):1294–1296. [PubMed: 17384015]
19. Karssen LC, van Duijn CM, Aulchenko YS. The GenABEL Project for statistical genomics. *F1000Research*. May.2016 5:914. [PubMed: 27347381]
21. R Core Team. R: A language and environment for statistical computing, version3.2.2. 2015.
22. Voorman, A., Brody, J., Lumley, T. skatMeta: Efficient meta analysis for the SKAT test. Jun. 2013
23. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. Mar; 2014 46(3): 310–315. [PubMed: 24487276]
24. Hu Q, Luo T, Zhong X, He P, Tian T, Zheng H. Application status of tamoxifen in endocrine therapy for early breast cancer. *Experimental and Therapeutic Medicine*. Jun; 2015 9(6):2207–2212. [PubMed: 26136961]
25. Muss HB. Factors used to select adjuvant therapy of breast cancer in the United States: an overview of age, race, and socioeconomic status. *JNCI Monographs*. Dec; 2001 2001(30):52–55.
26. Hertz DL, Rae JM. One step at a time: CYP2D6 guided tamoxifen treatment awaits convincing evidence of clinical validity. *Pharmacogenomics*. Jun; 2016 17(8):823–826. [PubMed: 27249031]
27. Zembutsu H. Pharmacogenomics toward personalized tamoxifen therapy for breast cancer. *Pharmacogenomics*. 2015; 16(3):287–296. [PubMed: 25712191]
28. Kiyotani K, Mushiroda T, Tsunoda T, Morizono T, Hosono N, Kubo M, Tanigawara Y, Imamura CK, Flockhart DA, Aki F, Hirata K, Takatsuka Y, Okazaki M, Ohsumi S, Yamakawa T, Sasa M, Nakamura Y, Zembutsu H. A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Human Molecular Genetics*. Apr; 2012 21(7):1665–1672. [PubMed: 22180457]
29. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*. Jun.2009 106(23):93629367.
30. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*. Jan.2011 88(1):7682.
31. DeAbreu F, Schwartz G, Wells W, Tsongalis G. Personalized therapy for breast cancer. *Clinical Genetics*. Jul; 2014 86(1):62–67. [PubMed: 24635704]

32. Bagaria SP, Ray PS, Sim MS, Ye X, Shamonki JM, Cui X, Giuliano AE. Personalizing breast cancer staging by the inclusion of ER, PR, and HER2. *JAMA surgery*. Feb; 2014 149(2):125–129. [PubMed: 24306257]
33. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tomczyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos Santos Silva I, Lathrop M, Houlston RS, Peto J. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *Journal of the National Cancer Institute*. Mar; 2011 103(5):425–435. [PubMed: 21263130]
34. 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, Fletcher O, Peto J, Gibson L, dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waissz Q, Meijers-Heijboer H, Adank M, van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Müller-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 J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, 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, Veer LJV, van der Schoot CE, Guénel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel 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, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, Ouweland AMWvd, 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, Müller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lissowska J, Brinton L, Goldberg MS, Labrèche F, Dumont M, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Brüning T, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RAEM, Seynaeve C, Asperen CJv, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dörk 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, Berg DVD, 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, Siriwanarangsana P, Shen CY, Hsiung CN, Wu PE, Ding SL, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PDP, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF. Network TGGELA BCiG, Investigators k, Group AOCs, Hereditary Breast and Ovarian Cancer Research Group Netherlands (hebon), The Breast and Ovarian Cancer Susceptibility Collaboration. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*. Apr; 2013 45(4):353–361. [PubMed: 23535729]
35. Ahsan H, Halpern J, Kibriya MG, Pierce BL, Tong L, Gamazon E, McGuire V, Felberg A, Shi J, Jasmine F, Roy S, Brutus R, Argos M, Melkonian S, Chang-Claude J, Andrulis I, Hopper JL, John EM, Malone K, Ursin G, Gammon MD, Thomas DC, Seminara D, Casey G, Knight JA, Southey MC, Giles GG, Santella RM, Lee E, Conti D, Duggan D, Gallinger S, Haile R, Jenkins M, Lindor NM, Newcomb P, Michailidou K, Apicella C, Park DJ, Peto J, Fletcher O, Silva IdS, Lathrop M, Hunter DJ, Chanock SJ, Meindl A, Schmutzler RK, Müller-Myhsok B, Lochmann M, Beckmann L, Hein R, Makalic E, Schmidt DF, Bui QM, Stone J, Flesch-Janys D, Dahmen N, Nevanlinna H, Aittomäki K, Blomqvist C, Hall P, Czene K, Irwanto A, Liu J, Rahman N, Turnbull C, Dunning AM, Pharoah P, Waissz Q, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Nicolae D, Easton DF, Cox NJ, Whittemore AS. Study ftFBC. A genome-wide association study of early-onset breast cancer identifies PFKM as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. *Cancer Epidemiology Biomarkers & Prevention*. Apr; 2014 23(4):658–669.
36. Fehrer G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher FR, Schildkraut JM, Lindström S, Brennan P, Bickeböller H, Houlston RS, Landi MT, Caporaso N, Risch A, Amin Al Olama A, Berndt SI, Giovannucci EL, Grönberg H, Kote-Jarai Z, Ma J, Muir K, Stampfer MJ,

Stevens VL, Wiklund F, Willett WC, Goode EL, Permuth JB, Risch HA, Reid BM, Bezieau S, Brenner H, Chan AT, Chang-Claude J, Hudson TJ, Kocarnik JK, Newcomb PA, Schoen RE, Slattery ML, White E, Adank MA, Ahsan H, Aittomäki K, Baglietto L, Blomquist C, Canzian F, Czene K, Dos-Santos-Silva I, Eliassen AH, Figueroa JD, Flesch-Janys D, Fletcher O, Garcia-Closas M, Gaudet MM, Johnson N, Hall P, Hazra A, Hein R, Hofman A, Hopper JL, Irwanto A, Johansson M, Kaaks R, Kibriya MG, Lichtner P, Liu J, Lund E, Makalic E, Meindl A, Müller-Myhok B, Muranen TA, Nevanlinna H, Peeters PH, Peto J, Prentice RL, Rahman N, Sanchez MJ, Schmidt DF, Schmutzler RK, Southey MC, Tamimi R, Travis RC, Turnbull C, Uitterlinden AG, Wang Z, Whittemore AS, Yang XR, Zheng W, Buchanan DD, Casey G, Conti DV, Edlund CK, Gallinger S, Haile RW, Jenkins M, Le Marchand L, Li L, Lindor NM, Schmit SL, Thibodeau SN, Woods MO, Rafnar T, Gudmundsson J, Stacey SN, Stefansson K, Sulem P, Chen YA, Tyrer JP, Christiani DC, Wei Y, Shen H, Hu Z, Shu XO, Shiraishi K, Takahashi A, Bossé Y, Obeidat M, Nickle D, Timens W, Freedman ML, Li Q, Seminara D, Chanock SJ, Gong J, Peters U, Gruber SB, Amos CI, Sellers TA, Easton DF, Hunter DJ, Haiman CA, Henderson BE, Hung RJ. Ovarian Cancer Association Consortium (OCAC), PRACTICAL Consortium, Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), Colorectal Transdisciplinary (CORECT) Study, and African American Breast Cancer Consortium (AABC) and African Ancestry Prostate Cancer Consortium (AAPC). Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations. *Cancer Research*. Sep; 2016 76(17): 5103–5114. [PubMed: 27197191]

37. Bahcall O. Common variation and heritability estimates for breast, ovarian and prostate cancers. *Nature Genetics*. 2013
38. Evans, DG., Astley, S., Stavrinou, P., Harkness, E., Donnelly, LS., Dawe, S., Jacob, I., Harvie, M., Cuzick, J., Brentnall, A., Wilson, M., Harrison, F., Payne, K., Howell, A. Programme Grants for Applied Research. NIHR Journals Library; Southampton (UK): 2016. Improvement in risk prediction, early detection and prevention of breast cancer in the NHS Breast Screening Programme and family history clinics: a dual cohort study.
39. Fagerholm R, Khan S, Schmidt MK, García-Closas M, Heikkilä P, Saarela J, Beesley J, Jamshidi M, Aittomäki K, Liu J, Ali HR, Andrulis IL, Beckmann MW, Behrens S, Blows FM, Brenner H, Chang-Claude J, Couch FJ, Czene K, Fasching PA, Figueroa J, Floris G, Glendon G, Guo Q, Hall P, Hallberg E, Hamann U, Holleczeck B, Hooning MJ, Hopper JL, Jager A, Kabisch M, Keeman R, Kosma VM, Lambrechts D, Lindblom A, Mannermaa A, Margolin S, Provenzano E, Shah M, Southey MC, Dennis J, Lush M, Michailidou K, Wang Q, Bolla MK, Dunning AM, Easton DF, Pharoah PDP, Chenevix-Trench G, Blomqvist C, Nevanlinna H. kConFab/AOCS Investigators. TP53-based interaction analysis identifies cis-eQTL variants for TP53BP2, FBXO28, and FAM53A that associate with survival and treatment outcome in breast cancer. *Oncotarget*. Feb. 2017
40. Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *Journal of Clinical Oncology*. May; 2003 21(10):1973–1979. [PubMed: 12743151]

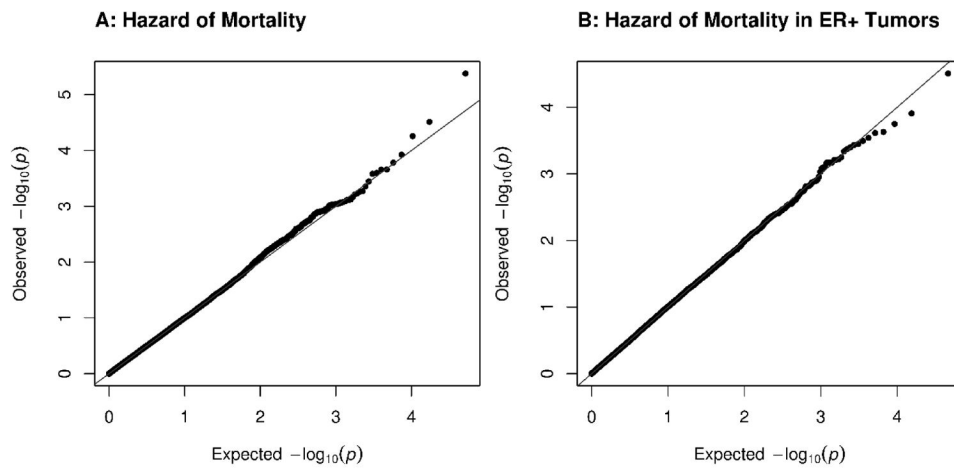


Figure 1. Single Marker Regression Results for Hazard of All-Cause Mortality
QQ plots of the single marker regression results for hazard of all-cause mortality using a Cox regression show no evidence of association between common variants in gene regions and mortality. The mortality analysis was repeated using all cases (A; N=2954) and cases with ER+ tumors (B; n=1066).

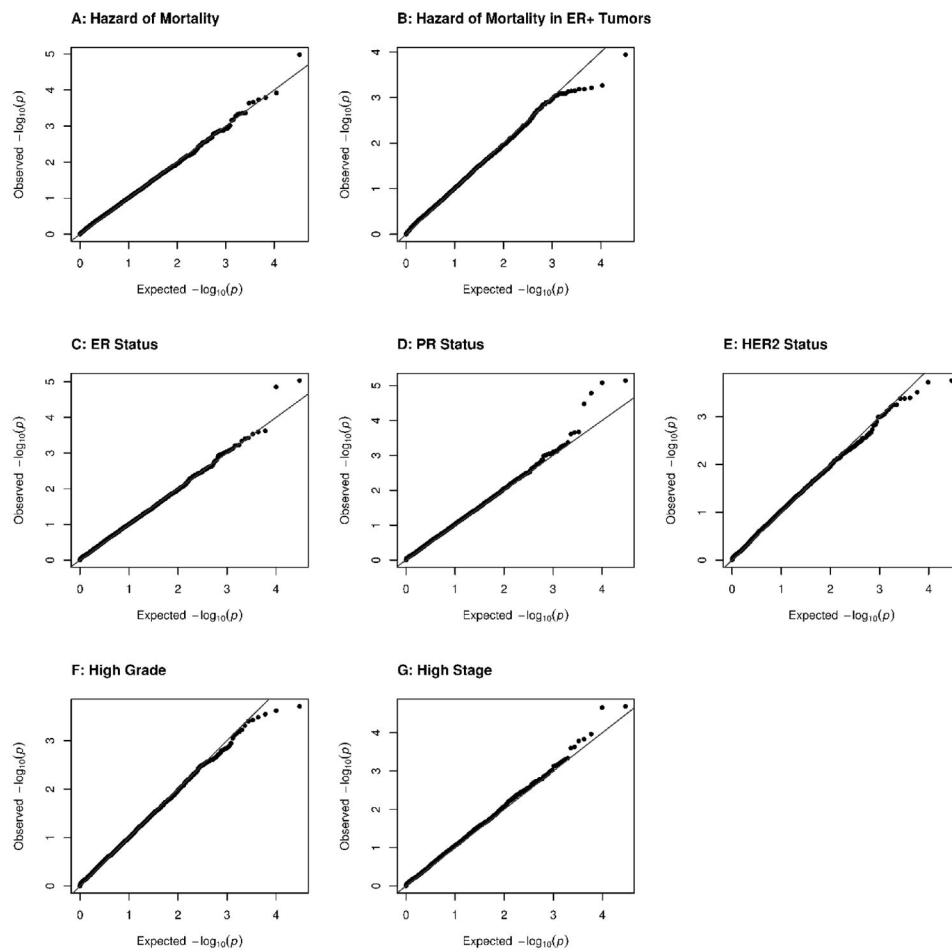


Figure 2. Gene-Based SKAT-O Results for Hazard of All-Cause Mortality and Tumor Characteristics. QQ plots of the gene-based results for hazard of all-cause mortality using a Cox regression (A–B) and logistic regression of tumor characteristics (C–G) show no evidence of association between variation in gene regions and mortality or tumor characteristics. The mortality analysis was repeated using all cases (A; $N=2954$) and cases with ER+ tumors (B; $n=1066$). The tumor characteristics investigated were ER status (C; $n=1785$), PR status (D; $n=1769$), HER2 status (E; $n=658$), high tumor grade (F; $n=1790$), and high tumor grade (G; $n=1640$).

Table 1

Demographic and Clinical Characteristics of Study Participants

	Risk Loci Identification (Primary)	Whole Genome Prediction (Primary)	Risk Loci Identification (Replication-TCGA)
Number of Participants	2954	1903	711
Number of Polymorphic Variants	114,206	3,245,343	15,121,555
Mean Age at Diagnosis (sd)	41 (6.2)	41 (5.7)	59 (13)
Median Years of Follow Up (IQR)	15 (9.5–17)	15 (12–17)	1.2 (0.4–3.4)
Number of Deaths Observed	728	547 ^a	73
Estrogen Receptor Status			
Positive	1066	758	538
Negative	719	557	132
Missing	1169	588	41
Progesterone Receptor Status			
Positive	1015	741	469
Negative	754	571	198
Missing	1185	599	44
HER2 Status			
Positive	280	238	106
Negative	378	255	386
Missing	2296	1410	219
Tumor Grade			
High Tumor Grade	865	659	-
Low Tumor Grade	925	690	-
Missing	1164	554	-
Tumor Stage			
High Tumor Stage	351	257	183
Low Tumor Stage	1289	993	516
Missing	1314	653	12

^a547 cases in the prediction analysis died during the follow up period, however, the prediction analysis used ten-year mortality as an outcome. 400 of the deaths occurred prior to ten years post-diagnosis