Study population details

Six of the study sites (Australia, Ontario, Philadelphia, Northern California, and New York) are members of the Breast Cancer Family Registry (BCFR), whose recruiting methods are described elsewhere. [1] Briefly, two of the BCFR centers (Northern California and Canada) recruited index cases through population-based registries, three (Utah, Philadelphia, and New York) recruited through clinic- and community-based outreach, and one (Australia) recruited through a mix of population- and clinic-based outreach. Participants were also included from four studies not included in the BCFR consortium. Three of these, the German Genetic Epidemiologic Study of Breast Cancer;[2] and Long Island Breast Cancer Study Project;[3] and the Seattle Study,[4] are population-based case-control studies whose recruiting methods are described elsewhere. The Chicago participants were enrolled from the Chicago Multi-ethnic Breast Cancer Epidemiologic Cohort, a hospital-based study of breast cancer at the University of Chicago.[5;6] The Chicago participants were identified through a clinic-based recruitment. Their demographic, clinical, and pathological data were gathered from medical chart review and cancer registry records, epidemiologic risk factors were collected via structured questionnaire, and mortality outcomes were ascertained via medical records and linkages with the national death index and cancer registry records. A schematic of the variants used in the primary analysis is shown in Supplemental Figure 1.

Exome-wide genotyping

Two versions of the exome array were used: The Illumina HumanExome 12v1.0 chip was used on 2527 cases, and the Illumina HumanExome 12v1.1 chip was used on 480 cases. The samples were processed using 49 plates in two batches, and the process was carried out according to the manufacturer's protocol. To improve the quantity and quality of available genomic DNA, the samples were whole genome amplified using the Qiagen Repli-G mini kit,[7] and were processed using 49 plates in two batches, following the manufacturer's protocol. TeCan Evo was used for automation. Raw data was processed by Genome Studio on 2010.3 software, and the no-call threshold was set at 0.15, per Illumina's recommendation for Infinium chips. Clustering was done using the Illumina supplied cluster files. After keeping only variants that were on both chips, 238,524 variants were interrogated. The quality control followed the protocol outline by Guo et al.[8] Participants were excluded for low genotyping rate (rate $< 95\%$; 219 excluded), high heterozygosity (F statistic greater than three standard deviations from the mean, or heterozygosity greater than four standard deviations from the mean; 31 excluded), and one of each pair of duplicated genotypes (eight samples excluded: three replicates; five duplicates from the same center). Additionally, due to the family-

based case ascertainment of some of the studies, seven participants were excluded whose genotypes were highly correlated (estimated relatedness from a GCTA-created genetic relatedness matrix greater than 0.4). [9] Variants were excluded from the analysis if they had a low call rate (rate $< 95\%$; 4335 excluded), or if they were common variants (defined below) with Hardey-Weinberg equilibrium p-values of less than $2.5 \cdot 10^{-7}$ in controls (p = 0.05 Bonferroni corrected for 200,000 tests; 39 excluded). The final variant-level exclusions were the result of evidence that on some plates variants were unreliably assigned (a plate-by-plate single marker regression analysis found that in some cases genotype could predict plate). For these variant-plate combinations, variants were excluded for all participants on that plate if this GWAS p-value was smaller than $2.5 \cdot 10^{-7}$. As a result of this quality control step, 100 variant-plate combinations were set to missing.

Principal component estimation

EIGENSTRAT [10] constructed two sets of principal components from the exome variants. One set was constructed using common variants assayed by the array ($MAF > 0.0130$) (PC_c), and one was constructed using rare variants (PC_r). To determine which of these principal components had a high chance of confounding the relationship, a series of non-genetic Cox regressions was run that included each additional PC in a step-wise fashion in order of their univariate significance. Likelihood ratio tests determined that the first three PC_c and the second PC_r were associated with the hazard of mortality, and including additional principal components did not improve the model fit. These four principal components were included in all subsequent mortality analyses. Similar analyses were done to determine the optimal number of PC_c and PC_r to include in each of the five tumor characteristic logistic analyses (ER status analysis included $PC_{c,1}$, $PC_{c,2}$, $PC_{c,3}$, and $PC_{r,2}$; PR status analysis included $PC_{c,1}$, $PC_{c,2}$, $PC_{c,6}$, and $PC_{r,2}$; HER2 status included $PC_{c,1}$, $PC_{c,2}$, $PC_{c,3}$, $PC_{c,4}$, $PC_{c,5}$, $PC_{r,2}$ and $PC_{r,4}$; high tumor grade analysis included $PC_{c,1}$; and high tumor stage analysis included $PC_{c,1}$, $PC_{c,4}$, and $PC_{r,2}$).

Genome-wide genotpying

Details of the BCFR genotyping are found elsewhere. [11] Briefly, the DNA was genotyped using the Illumina 610-Quad and Cyto12 v2 BeadChips, and standard laboratory quality control procedures were applied. After quality control, 555,254 variants and 3333 participants were brought forward to imputation, which was implemented by the Michigan imputation server, [12] employing ShapeIt [13] to pre-phase the variants and minimac3 to impute [14]. In order to best impute rare variants, [15;16] the entire 1000 Genomes phase 3 release [17] was used for a reference panel. Imputed variants with an imputation r^2greater than 0.8 were kept.

TCGA genome-wide genotyping, quality control, and imputation

The germline SNV data were measured using the Affymetrix Genome-Wide Human SNP 6.0 array, and the intensities were converted to genotype calls using the Broad Birdsuite.[18] To be comparable with the primary analysis, the analysis was restricted to female cases of European ancestry with mortality information available, excluding any participants or samples annotated "DNU" (Do Not Use) (900,380 variants from 768 participants). These samples were then subjected to the same quality control steps outlined above, resulting in the following exclusions: two participants were excluded for high levels of missingness; thirty nine were excluded for high heterozygosity; sixteen for outlying principal components, and none were highly related or duplicates. 42,150 variants were excluded due to their low call rate, and 60 were excluded for failing Hardey-Weinberg equilibrium. After these quality control procedures, 711 cases and 858,170 variants were brought forward for imputation. Imputation was implemented by the Michigan imputation server,[12] employing ShapeIt[13] and minimac3[14] to impute variants that were not measured. In order to best impute rare variants,[15;16] the entire 1000 Genomes phase 3 release[17] was used for a reference panel. Variants with an imputation r^2 greater than 0.3 were kept (15,121,555 variants). A schematic of the variants used in the replication analysis is in Supplemental Figure 2

References

[1] John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara D, and Breast Cancer Family Registry. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast cancer research: BCR*, 6(4):R375–389, 2004.

[2] Chang-Claude J, Eby N, Kiechle M, Bastert G, and Becher H. Breastfeeding and breast cancer risk by age 50 among women in Germany. *Cancer causes & control: CCC*, 11(8):687–695, September 2000.

[3] Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, Eng SM, Wolff MS, Stellman SD, Kabat GC, Levin B, Bradlow HL, Hatch M, Beyea J, Camann D, Trent M, Senie RT, Garbowski GC, Maffeo C, Montalvan P, Berkowitz GS, Kemeny M, Citron M, Schnabe F, Schuss A, Hajdu S, Vincguerra V, Collman GW, and Obrams GI. The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. *Breast Cancer Research and Treatment*, 74(3):235–254, June 2002.

[4] Friedrichsen DM, Malone KE, Doody DR, Daling JR, and Ostrander EA. Frequency of CHEK2 mutations in a population based, case-control study of breast cancer in young women. *Breast cancer research: BCR*, 6(6):R629–635, 2004.

[5] Nanda R, Schumm LP, Cummings S, Fackenthal JD, Sveen L, Ademuyiwa F, Cobleigh M, Esserman L, Lindor NM, Neuhausen SL, and Olopade OI. Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *JAMA*, 294(15):1925–1933, October 2005.

[6] Huo D, Senie RT, Daly M, Buys SS, Cummings S, Ogutha J, Hope K, and Olopade OI. Prediction of BRCA Mutations Using the BRCAPRO Model in Clinic-Based African American, Hispanic, and Other Minority Families in the United States. *Journal of Clinical Oncology*, 27(8):1184–1190, March 2009.

[7] Jasmine F, Ahsan H, Andrulis IL, John EM, Chang-Claude J, and Kibriya MG. Whole-genome amplification enables accurate genotyping for microarray-based high-density single nucleotide polymorphism array. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 17(12):3499–3508, December 2008.

[8] Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, Samuels DC, Shyr Y, and Long J. Illumina human exome genotyping array clustering and quality control. *Nature protocols*, 9(11):2643–2662, November 2014.

[9] Yang J, Lee SH, Goddard ME, and Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1):76–82, January 2011.

[10] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, and Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, August 2006.

[11] 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, Study ftFBC, Dunning AM, Pharoah P, Waisfisz Q, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Nicolae D, Easton DF, Cox NJ, and Whittemore AS. 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*, 23(4):658–669, April 2014.

[12] Michigan Imputation Server. *https://imputationserver.sph.umich.edu/index.html*, August 2016.

[13] Delaneau O, Marchini J, and Zagury JF. A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2):179–181, February 2012.

[14] Fuchsberger C, Abecasis GR, and Hinds DA. minimac2: faster genotype imputation. *Bioinformatics*, 31(5):782–784, March 2015.

[15] Wood AR, Perry JRB, Tanaka T, Hernandez DG, Zheng HF, Melzer D, Gibbs JR, Nalls MA, Weedon MN, Spector TD, Richards JB, Bandinelli S, Ferrucci L, Singleton AB, and Frayling TM. Imputation of variants from the 1000 Genomes Project modestly improves known associations and can identify low-frequency variant-phenotype associations undetected by HapMap based imputation. *PloS One*, 8(5):e64343, 2013.

[16] Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, Robinson MR, Perry JRB, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, The LifeLines Cohort Study, Esko T, Milani L, Mägi R, Metspalu A, Hamsten A, Magnusson PKE, Pedersen NL, Ingelsson E, Soranzo N, Keller MC, Wray NR, Goddard ME, and Visscher PM. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, 47(10):1114–1120, October 2015.

[17] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, October 2015.

[18] Broad Institute Birdsuite. *http://archive.broadinstitute.org/mpg/birdsuite/birdseed.html*.

[19] Azzato EM, Driver KE, Lesueur F, Shah M, Greenberg D, Easton DF, Teschendorff AE, Caldas C, Caporaso NE, and Pharoah PD. Effects of common germline genetic variation in cell cycle control genes on breast cancer survival: results from a population-based cohort. *Breast Cancer Research*, 10:R47, 2008.

[20] Azzato EM, Pharoah PDP, Harrington P, Easton DF, Greenberg D, Caporaso NE, Chanock SJ, Hoover RN, Thomas G, Hunter DJ, and Kraft P. A genome-wide association study of prognosis in breast cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 19(4):1140–1143, April 2010.

[21] Rafiq S, Khan S, Tapper W, Collins A, Upstill-Goddard R, Gerty S, Blomqvist C, Aittomäki K, Couch FJ, Liu J, Nevanlinna H, and Eccles D. A Genome Wide Meta-Analysis Study for Identification of Common Variation Associated with Breast Cancer Prognosis. *PLOS ONE*, 9(12):e101488, December 2014.

[22] Guo Q, Schmidt MK, Kraft P, Canisius S, Chen C, Khan S, Tyrer J, Bolla MK, Wang Q, Dennis J, Michailidou K, Lush M, Kar S, Beesley J, Dunning AM, Shah M, Czene K, Darabi H, Eriksson M, Lambrechts D, Weltens C, Leunen K, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Blomqvist C, Aittomäki K, Fagerholm R, Muranen TA, Couch FJ, Olson JE, Vachon C, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Broeks A, Hogervorst FB, Haiman CA, Henderson BE, Schumacher F, Marchand LL, Hopper JL, Tsimiklis H, Apicella C, Southey MC, Cox A, Cross SS, Reed MWR, Giles GG, Milne RL, McLean C, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Hooning MJ, Hollestelle A, Martens JWM, Ouweland AMWvd, Marme F, Schneeweiss A, Yang R, Burwinkel B, Figueroa J, Chanock SJ, Lissowska J, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Brenner H, Dieffenbach AK, Arndt V, Holleczek B, Mannermaa A,

Kataja V, Kosma VM, Hartikainen JM, Li J, Brand JS, Humphreys K, Devilee P, Tollenaar RAEM, Seynaeve C, Radice P, Peterlongo P, Bonanni B, Mariani P, Fasching PA, Beckmann MW, Hein A, Ekici AB, Chenevix-Trench G, Balleine R, Investigators k, Phillips KA, Benitez J, Zamora MP, Perez JIA, Menéndez P, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Hamann U, Kabisch M, Ulmer HU, Rüdiger T, Margolin S, Kristensen V, Nord S, Evans DG, Abraham JE, Earl HM, Hiller L, Dunn JA, Bowden S, Berg C, Campa D, Diver WR, Gapstur SM, Gaudet MM, Hankinson SE, Hoover RN, Hüsing A, Kaaks R, Machiela MJ, Willett W, Barrdahl M, Canzian F, Chin SF, Caldas C, Hunter DJ, Lindstrom S, García-Closas M, Hall P, Easton DF, Eccles DM, Rahman N, Nevanlinna H, and Pharoah PDP. Identification of Novel Genetic Markers of Breast Cancer Survival. *Journal of the National Cancer Institute*, 107(5):djv081, May 2015.

[23] Shu XO, Long J, Lu W, Li C, Chen WY, Delahanty R, Cheng J, Cai H, Zheng Y, Shi J, Gu K, Wang WJ, Kraft P, Gao YT, Cai Q, and Zheng W. Novel Genetic Markers of Breast Cancer Survival Identified by a Genome-Wide Association Study. *Cancer Research*, 72(5):1182–1189, March 2012.

[24] Khan S, Fagerholm R, Rafiq S, Tapper W, Aittomäki K, Liu J, Blomqvist C, Eccles D, and Nevanlinna H. Polymorphism at 19q13.41 Predicts Breast Cancer Survival Specifically after Endocrine Therapy. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 21(18):4086–4096, September 2015.

[25] Fagerholm R, Schmidt MK, Khan S, Rafiq S, Tapper W, Aittomäki K, Greco D, Heikkinen T, Muranen TA, Fasching PA, Janni W, Weinshilboum R, Loehberg CR, Hopper JL, Southey MC, Keeman R, Lindblom A, Margolin S, Mannermaa A, Kataja V, Chenevix-Trench G, kConFab Investigators, Lambrechts D, Wildiers H, Chang-Claude J, Seibold P, Couch FJ, Olson JE, Andrulis IL, Knight JA, García-Closas M, Figueroa J, Hooning MJ, Jager A, Shah M, Perkins BJ, Luben R, Hamann U, Kabisch M, Czene K, Hall P, Easton DF, Pharoah PDP, Liu J, Eccles D, Blomqvist C, and Nevanlinna H. The SNP rs6500843 in 16p13.3 is associated with survival specifically among chemotherapy-treated breast cancer patients. *Oncotarget*, 6(10):7390–7407, April 2015.

[26] Rafiq S, Tapper W, Collins A, Khan S, Politopoulos I, Gerty S, Blomqvist C, Couch FJ, Nevanlinna H, Liu J, and Eccles D. Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. *Cancer Research*, 73(6):1883–1891, March 2013.

[27] 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, and 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*, 21(7):1665–1672, April 2012.

[28] Song N, Choi JY, Sung H, Jeon S, Chung S, Park SK, Han W, Lee JW, Kim MK, Lee JY, Yoo KY, Han BG, Ahn SH, Noh DY, and Kang D. Prediction of breast cancer survival using clinical and genetic markers by tumor subtypes. *PloS One*, 10(4):e0122413, 2015.

Supplemental Figures and Tables

Figure 1: Schematic of Variants Used in Primary Analysis

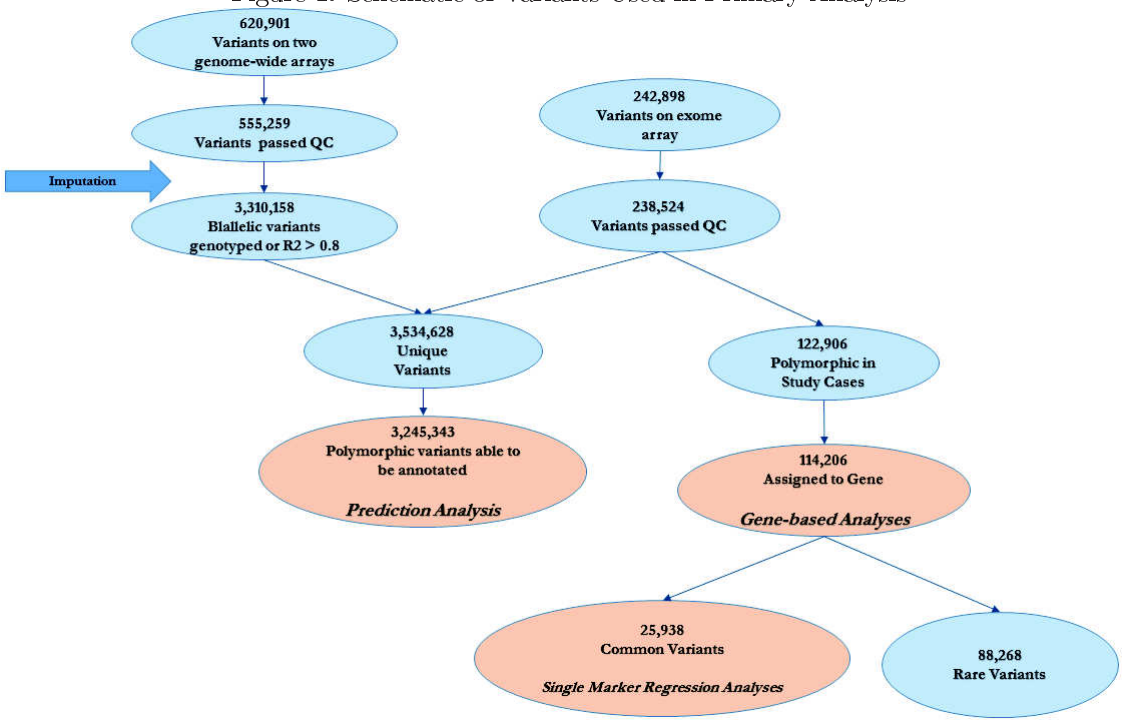

Figure 2: Variants Used in Replication Analysis

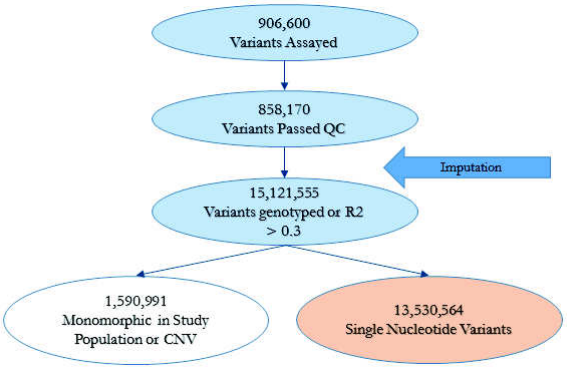

Table 1: Genome-wide Studies of the Association between Germline Genetic Variation and Breast Cancer Mortality

Study Title	Year	Population Description	Outcome	N	Median Follow Up Time	Events	Variants	Replication Description	Findings
A Genome-Wide Association Study of Prognosis in Breast Cancer [20]	2010	Postmenopausal women with invasive breast cancer	Breast cancer specific survival	1145	6 years	93	528,252	Top 10 genotyped in 4335 women with invasive breast cancer with 38,148 years at risk	Nothing genome-wide significant
A Genome-wide Association Study Identifies Locus at 10q22 Associated with Clinical Outcomes of Adjuvant Tamoxifen Therapy for Breast Cancer Patients in Japanese [27]	2011	Japanese patients with hormone receptor-positive, invasive breast cancer receiving adjuvant tamoxifen therapy	Recurrence-free survival	240	7 years	30	470,796	Two independent sets of 105 and 117 cases	15 SNVs in the primary analysis; rs10509373 (chr10:76397814) replicated (combined $p = 1.26 \cdot 10^{-10}$)
Novel Genetic Markers of Breast Cancer Survival Identified by a Genome-Wide Association Study [23]	2012	Shanghai-resident Chinese women	Total mortality	1950	6 years	299	613,031	Top 49 associations replicated in 4160 Shanghai women with breast cancer; Top association examined in Nurses Health Study	rs3784099 (chr14:68283210; $p = 1.44 \cdot 10^{-8}$ in discovery only)
Identification of Inherited Genetic Variations Influencing Prognosis in Early-onset Breast Cancer [26]	2013	UK women aged 40 or younger at diagnosis	Breast cancer specific survival	536	4 years	236	487,496	Top 35 associations genotyped in 1,516 independent cases from the same early-onset cohort	Nothing genome-wide significant
Genome Wide Meta-Analysis Study for Identification of Common Variation Associated with Breast Cancer Prognosis [21]	2014	UK women aged 40 or younger at diagnosis, and Finish women of all ages	Breast cancer specific survival	1341	6 years	237	475,141, imputed to 7.5 million	1523 additional participants of the POSH study	Nothing genome-wide significant

positions refer to the HG19 assembly
Studies that published both single-study results and contributed to a meta analysis will be represented twice

Table 2: Genome-wide Studies of the Association between Germline Genetic Variation and Breast Cancer Mortality(Continued)

Study Title	Year	Population Description	Outcome	N	Median Follow Up Time	Events	Variants	Replication Description	Findings
Identification of Novel Genetic Markers of Breast Cancer Survival[22]	2015	Meta-analysis of studies in populations of European ancestry	Breast cancer specific survival	37,954	5 years	2900	200,000-700,000; imputed to 9 million	N/A	rs148760487 (chr2:162922103; $p = 1.5 \cdot 10^{-8}$) and 27 others in high LD; rs2059614 (chr11:125389528; $p = 1.3 \cdot 10^{-9}$ in ER-cases
Polymorphism at 19q13.41 Predicts Breast Cancer Survival Specifically after Endocrine Therapy[24]	2015	Meta analysis of UK women aged 40 or younger at diagnosis, and Finish women of all ages	Breast cancer specific survival	1341	7 years	547	486,478	Two independent data sets with 5011 patients	Nothing genome-wide significant
Prediction of Breast Cancer Survival Using Clinical and Genetic Markers by Tumor Subtypes[28]	2015	Incident breast cancer cases in Seoul, South Korea	Recurrence-free survival	1732	4 years	214	2,210,580 genotyped and imputed	Any SNVs identified with $p < 10^{-6}$ and MAF $> .1$, and any common variants in high ($r^2 > 0.4$) LD with them were genotyped in 1494 additional women from South Korea	Nothing genome-wide significant

positions refer to the HG19 assembly
Studies that published both single-study results and contributed to a meta analysis will be represented twice

Table 3: Studies Included in Primary Analysis

Study Name	Study Location	Years Recruiting	Case Criteria	n with exome and genome-wide array	n with exome only
Breast Cancer Family Registry	Australia	1992-2000	Living in the Melbourne and Sydney metro areas, family recruited from the Victoria and NSW cancer registries	452	24
Breast Cancer Family Registry	New York, NY	1996-2000	Living in New York, New Jersey, or Connecticut	0	380
Breast Cancer Family Registry	Northern California	1996-2003	SEER Cancer registry in the San Francisco metro area	163	9
Breast Cancer Family Registry	Ontario	2001-2010	Ontario Cancer Registry	482	77
Breast Cancer Family Registry	Philadelphia, PA	1996-2000	Living in Philadelphia	0	265
Breast Cancer Family Registry	Utah	1996-2012	Living in Salt Lake City	0	96
Genetic Epidemiologic Study of Breast Cancer by Age 50	Germany	1992-1995	38 clinics in the Rhein-Neckar-Odenwald and Freiburg regions	377	5
Long Island Breast Cancer Study Project	New York	1996-1999	Nassau and Suffolk counties	144	1
Seattle	Seattle, WA	1990-1992	King, Pierce, and Snohomish counties; age less than 45 at diagnosis	285	3
University of Chicago	Chicago, IL	1998-2010	Treated at the University of Chicago Cancer Center	0	191

Participants are those included in the analysis after QC

Table 4: Most Significant Variants Identified by Single Marker Cox Regression with Hazard of All-Cause Mortality

					Primary Analysis			TCGA		
Chr:Position[a]	Gene	Function [b]	Minor Allele	Major Allele	MAF	HR(se)	p-value	MAF	HR(se)	p-value
6:35430686	FANCE	nonsynonymous SNV	G	A	0.055	1.576 (0.156)	4.177E-06	0.056	1.169 (0.413)	6.593E-01
1:116310967	CASQ2	nonsynonymous SNV	T	C	0.287	1.265 (0.071)	3.061E-05	0.256	0.831 (0.166)	3.544E-01
20:1551564	SIRPB1	nonsynonymous SNV	T	C	0.018	1.903 (0.304)	5.508E-05	0.028	0.486 (0.285)	2.186E-01
8:19819724	LPL	stopgain	C	G	0.100	1.367 (0.111)	1.190E-04	0.087	1.128 (0.306)	6.582E-01
17:66364691	ARSG	nonsynonymous SNV	C	G	0.429	0.821 (0.043)	1.651E-04	0.435	1.165 (0.203)	3.799E-01
2:87044316	CD8B	intronic	A	G	0.222	0.780 (0.052)	2.189E-04	0.212	1.033 (0.210)	8.741E-01
6:35423886	FANCE	nonsynonymous SNV	C	T	0.014	1.915 (0.337)	2.216E-04	0.015	2.155 (1.296)	2.017E-01
10:108543337	SORCS1	intronic	A	G	0.413	0.821 (0.044)	2.535E-04	0.398	1.076 (0.178)	6.588E-01
3:52833219	ITIH3	intronic	G	A	0.347	0.812 (0.046)	2.613E-04	0.363	0.762 (0.142)	1.449E-01
6:34730395	SNRPC	synonymous SNV	C	T	0.014	1.915 (0.349)	3.569E-04	0.015	1.830 (1.098)	3.137E-01
2:179545859	TTN	nonsynonymous SNV	C	T	0.289	1.218 (0.068)	4.411E-04	0.342	1.152 (0.189)	3.888E-01
2:179432185	TTN	nonsynonymous SNV	A	G	0.288	1.215 (0.068)	5.382E-04	0.342	1.165 (0.191)	3.498E-01
13:52343391	DHRS12	nonsynonymous SNV	C	T	0.064	0.646 (0.082)	5.490E-04	0.080	1.255 (0.294)	3.324E-01
14:24458162	DHRS4L2	nonsynonymous SNV	G	C	0.332	1.203 (0.065)	5.882E-04	0.384	0.659 (0.132)	3.687E-02
3:52874288	TMEM110	UTR3	T	C	0.243	0.801 (0.052)	6.107E-04	0.264	0.866 (0.167)	4.550E-01
1:205318321	KLHDC8A	intronic	C	T	0.387	0.831 (0.045)	6.879E-04	0.410	0.884 (0.142)	4.431E-01
7:123599845	SPAM1	nonsynonymous SNV	A	T	0.021	1.685 (0.261)	7.551E-04	0.016	1.996 (1.204)	2.521E-01
18:31645379	NOL4	intronic	G	A	0.421	0.834 (0.045)	7.745E-04	0.420	1.198 (0.203)	2.865E-01
5:131995964	IL13	synonymous SNV	G	A	0.196	1.236 (0.078)	7.746E-04	0.205	1.160 (0.270)	5.230E-01
11:124294703	OR8B4	nonsynonymous SNV	T	C	0.303	1.200 (0.065)	8.169E-04	0.307	1.127 (0.213)	5.260E-01

a: positions refer to the HG19 assembly
b: annotation from ANNOVAR
HR(se) = Hazard Ratio(Standard Error)

Table 5: Most Significant Variants Identified by Single Marker Cox Regression with Hazard of All-Cause Mortality in Cases with ER+ Tumors

					Primary Analysis			TCGA		
Chr:Position[a]	Gene	Function [b]	Minor Allele	Major Allele	MAF	HR(se)	p-value	MAF	HR(se)	p-value
6:26017542	HIST1H1A	nonsynonymous SNV	T	C	0.102	1.877 (0.284)	3.119E-05	0.090	0.577 (0.249)	2.022E-01
21:45503121	TRAPPC10	nonsynonymous SNV	G	A	0.175	1.510 (0.162)	1.239E-04	0.162	1.165 (0.330)	5.889E-01
1:248059476	OR2W3	synonymous SNV	C	A	0.204	0.615 (0.080)	1.775E-04	0.175	1.080 (0.265)	7.550E-01
6:133119564	SLC18B1	nonsynonymous SNV	C	A	0.031	2.094 (0.421)	2.343E-04	0.033	1.719 (1.038)	3.701E-01
20:34218673	CPNE1	nonsynonymous SNV	G	C	0.109	1.552 (0.186)	2.440E-04	0.112	1.082 (0.320)	7.895E-01
7:45148667	TBRG4	nonsynonymous SNV	G	A	0.170	1.473 (0.157)	2.882E-04	0.186	0.774 (0.175)	2.569E-01
1:161751741	ATF6	nonsynonymous SNV	A	G	0.277	0.676 (0.074)	3.193E-04	0.282	1.029 (0.255)	9.097E-01
15:83254708	CPEB1	intronic	A	G	0.384	0.705 (0.069)	3.559E-04	0.361	1.117 (0.230)	5.902E-01
18:23866185	TAF4B	nonsynonymous SNV	G	C	0.108	0.520 (0.095)	3.692E-04	0.095	0.602 (0.266)	2.506E-01
21:45506819	TRAPPC10	synonymous SNV	T	C	0.192	1.457 (0.155)	4.007E-04	0.173	1.261 (0.335)	3.815E-01
22:18611223	TUBA8	intronic	A	G	0.188	1.458 (0.156)	4.238E-04	0.195	0.813 (0.247)	4.960E-01
3:9871030	ARPC4-TTLL3	synonymous SNV	G	T	0.289	1.387 (0.129)	4.596E-04	0.290	1.223 (0.265)	3.531E-01
19:54603419	OSCAR	intronic	C	T	0.448	1.368 (0.124)	5.670E-04	0.422	1.021 (0.220)	9.234E-01
14:24805463	RIPK3	nonsynonymous SNV	G	T	0.063	1.685 (0.256)	6.057E-04	0.070	0.485 (0.255)	1.687E-01
1:161915501	ATF6	intronic	G	A	0.402	1.356 (0.121)	6.125E-04	0.401	1.206 (0.261)	3.870E-01
14:39736680	CTAGE5	synonymous SNV	C	T	0.063	1.751 (0.287)	6.184E-04	0.054	1.237 (0.439)	5.480E-01
20:34502107	PHF20	nonsynonymous SNV	G	A	0.105	1.519 (0.187)	6.761E-04	0.115	0.993 (0.294)	9.823E-01
17:66289041	ARSG	intronic	G	T	0.418	1.354 (0.121)	6.775E-04	0.428	1.164 (0.265)	5.052E-01
6:31237124	HLA-C	synonymous SNV	C	T	0.292	0.616 (0.088)	6.776E-04	0.311	1.143 (0.273)	5.769E-01
11:18586122	UEVLD	intronic	C	T	0.332	0.714 (0.071)	6.853E-04	0.352	0.938 (0.199)	7.628E-01

a: positions refer to the HG19 assembly
b: annotation from ANNOVAR
HR(se) = Hazard Ratio(Standard Error)

Table 6: Association with Mortality of Variants Previously Identified with Breast Cancer Phenotypes

chr:position[a]	first author	journal	year	reported trait	reported p-value	p-value in primary analysis	p-value in primary analysis of ER+ cases
chr1:114448389	Michailidou K	Nat Genet	2013	breast cancer	2.00E-08	5.09E-01	1.97E-01
chr1:121280613	Thomas G	Nat Genet	2009	breast cancer	7.00E-10	9.72E-02	3.42E-01
chr1:121280613	Michailidou K	Nat Genet	2013	breast cancer	2.00E-26	9.72E-02	3.42E-01
chr3:27416013	Fletcher O	J Natl Cancer Inst	2011	breast cancer	2.00E-08	1.34E-01	1.69E-01
chr3:27416013	Michailidou K	Nat Genet	2013	breast cancer	2.00E-30	1.34E-01	1.69E-01
chr6:127600630	Gold B	Proc Natl Acad Sci U S A	2008	breast cancer	3.00E-08	1.30E-01	9.86E-01
chr6:151914113	Michailidou K	Nat Genet	2013	breast cancer	2.00E-21	2.91E-01	8.34E-01
chr8:128355618	Michailidou K	Nat Genet	2013	breast cancer	1.00E-27	5.02E-01	9.12E-01
chr8:128355618	Easton DF	Nature	2007	breast cancer	5.00E-12	5.02E-01	9.12E-01
chr8:128387852	Fletcher O	J Natl Cancer Inst	2011	breast cancer	3.00E-11	9.35E-02	2.04E-01
chr9:22062134	Turnbull C	Nat Genet	2010	breast cancer	3.00E-08	5.60E-01	2.61E-01
chr10:64278682	Lindstrom S	Nat Genet	2011	mammographic density	1.00E-09	4.92E-01	4.17E-01
chr10:64278682	Michailidou K	Nat Genet	2013	breast cancer	1.00E-36	4.92E-01	4.17E-01
chr10:64278682	Turnbull C	Nat Genet	2010	breast cancer	5.00E-15	4.92E-01	4.17E-01
chr10:64278682	Lindstrom S	Nat Commun	2014	mammographic density	1.00E-16	4.92E-01	4.17E-01
chr10:80841148	Michailidou K	Nat Genet	2013	breast cancer	7.00E-22	2.83E-01	8.93E-01
chr10:80841148	Turnbull C	Nat Genet	2010	breast cancer	4.00E-09	2.83E-01	8.93E-01
chr10:123337335	Michailidou K	Nat Genet	2013	breast cancer	2.00E-170	1.63E-01	5.41E-01
chr10:123337335	Thomas G	Nat Genet	2009	breast cancer	2.00E-10	1.63E-01	5.41E-01
chr10:123337335	Turnbull C	Nat Genet	2010	breast cancer	4.00E-31	1.63E-01	5.41E-01
chr10:123346116	Gaudet MM	PLoS Genet	2010	breast cancer	1.00E-08	1.64E-01	2.84E-01
chr10:123346190	Li J	Breast Cancer Res Treat	2010	breast cancer	2.00E-13	1.75E-01	2.71E-01
chr10:123346190	Fletcher O	J Natl Cancer Inst	2011	breast cancer	1.00E-30	1.75E-01	2.71E-01
chr10:123346190	Hunter DJ	Nat Genet	2007	breast cancer	1.00E-10	1.75E-01	2.71E-01
chr10:123352317	Easton DF	Nature	2007	breast cancer	2.00E-76	2.25E-01	3.77E-01
chr11:1909006	Lindstrom S	Nat Commun	2014	mammographic density	1.00E-10	2.62E-01	4.84E-01
chr11:1909006	Easton DF	Nature	2007	breast cancer	3.00E-09	2.62E-01	4.84E-01
chr11:1909006	Michailidou K	Nat Genet	2013	breast cancer	2.00E-11	2.62E-01	4.84E-01
chr14:69034682	Michailidou K	Nat Genet	2013	breast cancer	3.00E-19	9.03E-01	1.66E-01
chr16:52586341	Low SK	PLoS One	2013	breast cancer	3.00E-11	9.85E-01	8.44E-01
chr16:52586341	Thomas G	Nat Genet	2009	breast cancer	1.00E-09	9.85E-01	8.44E-01
chr16:52586341	Michailidou K	Nat Genet	2013	breast cancer	2.00E-114	9.85E-01	8.44E-01
chr16:52586341	Turnbull C	Nat Genet	2010	breast cancer	3.00E-15	9.85E-01	8.44E-01
chr16:52586341	Stacey SN	Nat Genet	2007	breast cancer	6.00E-19	9.85E-01	8.44E-01
chr16:52586341	Garcia-Closas M	Nat Genet	2013	breast cancer	6.00E-13	9.85E-01	8.44E-01
chr16:52586341	Orr N	Nat Genet	2012	breast cancer	4.00E-15	9.85E-01	8.44E-01
chr16:52586341	Easton DF	Nature	2007	breast cancer	1.00E-36	9.85E-01	8.44E-01
chr16:52599188	Fejerman L	Nat Commun	2014	breast cancer	3.00E-09	9.04E-01	9.62E-01
chr16:52599188	Long J	PLoS Genet	2010	breast cancer	1.00E-28	9.04E-01	9.62E-01
chr16:52635164	Fletcher O	J Natl Cancer Inst	2011	breast cancer	4.00E-10	2.96E-01	7.03E-02
chr16:53813367	Michailidou K	Nat Genet	2013	breast cancer	6.00E-14	1.47E-01	8.37E-02
chr19:17389704	Antoniou AC	Nat Genet	2010	breast cancer	2.00E-09	7.78E-01	2.57E-01
chr19:17389704	Garcia-Closas M	Nat Genet	2013	breast cancer	9.00E-13	7.78E-01	2.57E-01
chr19:17389704	Couch FJ	PLoS Genet	2013	breast cancer	4.00E-13	7.78E-01	2.57E-01
chr19:17392894	Siddiq A	Hum Mol Genet	2012	breast cancer	4.00E-08	5.18E-01	1.04E-01
chr19:17394124	Purrington KS	Carcinogenesis	2013	breast cancer	2.00E-08	5.26E-01	1.09E-01

a: positions refer to the HG19 assembly

Table 7: Most Significant Genes Identified by SKAT-O Cox Regression with Hazard of All-Cause Mortality

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
ANP32D	218	2	1.21e-04	5181	16	5.93e-01
ARSG	4731	15	1.86e-04	225451	810	5.97e-01
CHAF1A	884	14	1.19e-03	66951	97	4.37e-01
CLCF1	7	3	4.72e-04	147	3	1.85e-01
DHRS12	525	4	2.16e-04	26783	153	4.38e-01
FANCE	215	6	1.05e-05	16131	97	5.01e-01
FPR1	2654	10	1.33e-03	18939	56	9.59e-02
HOXD12	72	4	4.30e-04	3207	9	1.02e-01
HOXD13	714	3	1.34e-03	3600	17	6.82e-01
KIAA1683	10879	33	9.63e-04	71392	149	1.52e-01
KLHDC8A	1143	2	6.62e-04	79783	157	8.27e-01
LPL	387	7	2.16e-04	49468	181	9.15e-01
NEK4	1307	9	1.19e-03	87145	224	8.04e-01
PNPLA3	2542	10	1.08e-03	58831	153	3.74e-02
SNRPC	45	2	1.17e-03	37183	103	3.71e-01
TM9SF3	2	1	4.36e-04	74752	358	6.46e-01
TMEM110	757	3	4.53e-04	78067	249	5.50e-01
UBE2Q2	5	2	1.51e-04	38709	180	3.98e-01
VSX1	590	5	6.85e-04	19808	57	4.36e-01
WDR12	364	4	5.35e-04	27693	127	8.09e-01

Table 8: Most Significant Genes Identified by SKAT-O Cox Regression with Hazard of All-Cause Mortality in ER+ Participants

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
ANKRD34C	8	2	8.82e-04	5206	45	5.77e-01
ARPC4-TTLL3	313	9	6.12e-04	23500	181	2.26e-01
ATF6	492	6	4.61e-04	73059	829	2.70e-01
BTN2A2	149	12	7.23e-04	7245	99	9.90e-01
CCDC67	556	9	1.11e-03	85332	491	3.98e-01
CPNE1	507	10	9.12e-04	10498	110	9.47e-01
E2F5	0	1	1.27e-03	10883	101	7.19e-01
FGFR1OP2	1	1	8.21e-04	10593	117	6.16e-02
GBA	11	3	6.53e-04	3951	29	4.20e-01
ISLR2	20	4	1.19e-03	71	10	7.35e-02
MAPK8IP1	15	4	6.53e-04	3718	73	1.54e-02
MARK2	3	2	1.24e-03	25839	180	5.03e-01
MED20	1	1	8.23e-04	3079	50	7.95e-01
SDC2	266	2	1.12e-03	71831	601	2.15e-01
SLC18B1	105	6	7.09e-04	8425	175	2.99e-01
STXBP4	284	2	9.49e-04	131435	842	2.87e-01
TAF4B	409	6	1.14e-04	72213	721	5.11e-01
TRPC1	1	1	8.20e-04	28994	264	4.50e-01
UEVLD	402	5	8.20e-04	25835	239	2.54e-01
YPEL3	9	2	1.25e-03	1049	9	8.73e-01

Table 9: Association with Mortality of Genes Previously Associated with Breast Cancer Mortality

first author	journal	year	chromosome	gene	p-value in primary analysis	p-value in primary analysis of ER+ cases
Rafiq S	Cancer Res	2013	1	PBX1	6.45E-01	5.01E-01
Rafiq S	Cancer Res	2013	1	SYT6	6.35E-01	9.56E-01
Guo Q	J Natl Cancer Inst	2015	11	PKNOX2	9.00E-01	4.09E-01
Shu XO	Cancer Res	2012	14	RAD51B	1.31E-01	7.97E-01
Khan S	Clin Cancer Res	2015	19	ZNF613	5.14E-01	7.42E-01

Table 10: Association with Mortality of Genes Previously Associated with Breast Cancer Phenotypes

first author	journal	year	chromo-some	gene	p-value in primary analysis	p-value in primary analysis of ER+ cases
Michailidou K	Nat Genet	2013	1	DCLRE1B	5.15E-01	1.98E-01
Michailidou K	Nat Genet	2013	1	EMBP1	9.58E-02	3.41E-01
Thomas G	Nat Genet	2009	1	EMBP1	9.58E-02	3.41E-01
Barnett GC	Radiother Oncol	2014	1	KCND3	7.60E-01	4.54E-01
Garcia-Closas M	Nat Genet	2013	1	LGR6	8.10E-01	4.44E-01
Garcia-Closas M	Nat Genet	2013	1	MDM4	4.10E-01	4.28E-01
Michailidou K	Nat Genet	2013	1	PEX14	9.89E-01	8.36E-02
Garcia-Closas M	Nat Genet	2013	1	PEX14	9.89E-01	8.36E-02
Cai Q	Nat Genet	2014	1	ZC3H11A	4.36E-01	5.42E-01
Michailidou K	Nat Genet	2013	2	DIRC3	5.64E-01	6.29E-01
Michailidou K	Nat Genet	2013	2	DLX2	6.87E-01	
Kim HC	Breast Cancer Res	2012	2	ERBB4	6.80E-01	6.66E-01
Michailidou K	Nat Genet	2013	3	ITPR1	8.65E-02	1.64E-02
Ahsan H	Cancer Epidemiol Biomarkers Prev	2014	3	NEK10	4.01E-01	4.09E-01
Fletcher O	J Natl Cancer Inst	2011	3	SLC4A7	2.61E-01	3.11E-01
Michailidou K	Nat Genet	2013	3	SLC4A7	2.61E-01	3.11E-01
Michailidou K	Nat Genet	2013	3	TGFBR2	1.14E-01	1.34E-02
Michailidou K	Nat Genet	2013	4	ADAM29	3.64E-01	3.66E-01
Michailidou K	Nat Genet	2013	4	TET2	1.79E-01	2.00E-01
Michailidou K	Nat Genet	2013	5	EBF1	7.40E-01	7.23E-01
Ahsan H	Cancer Epidemiol Biomarkers Prev	2014	5	MAP3K1	6.88E-01	1.45E-01
Michailidou K	Nat Genet	2013	5	PDE4D	4.52E-01	9.09E-01
Lindstrom S	Nat Commun	2014	5	PRDM6	2.37E-01	6.41E-01
Haiman CA	Nat Genet	2011	5	TERT	9.20E-01	3.86E-01
Garcia-Closas M	Nat Genet	2013	5	TERT	9.20E-01	3.86E-01
Michailidou K	Nat Genet	2013	5	TERT	9.20E-01	3.86E-01
Michailidou K	Nat Genet	2013	6	CCDC170	6.65E-01	6.81E-01
Gold B	Proc Natl Acad Sci U S A	2008	6	RNF146	1.29E-01	9.77E-01
Long J	PLoS Genet	2012	6	TAB2	4.39E-01	3.85E-01
Fletcher O	J Natl Cancer Inst	2011	8	CASC8	5.03E-01	6.51E-01
Easton DF	Nature	2007	8	CASC8	5.03E-01	6.51E-01
Ahsan H	Cancer Epidemiol Biomarkers Prev	2014	8	CASC8	5.03E-01	6.51E-01
Michailidou K	Nat Genet	2013	8	CASC8	5.03E-01	6.51E-01
Liu M	Mol Endocrinol	2013	8	LOC101927066	9.65E-02	2.04E-02
Turnbull C	Nat Genet	2010	9	CDKN2B-AS1	1.88E-01	9.36E-01
Kiyotani K	Hum Mol Genet	2011	10	C10orf11	8.15E-01	2.67E-01
Ahsan H	Cancer Epidemiol Biomarkers Prev	2014	10	FGFR2	2.09E-01	3.44E-01
Low SK	PLoS One	2013	10	FGFR2	2.09E-01	3.44E-01
Fletcher O	J Natl Cancer Inst	2011	10	FGFR2	2.09E-01	3.44E-01
Easton DF	Nature	2007	10	FGFR2	2.09E-01	3.44E-01
Turnbull C	Nat Genet	2010	10	FGFR2	2.09E-01	3.44E-01
Michailidou K	Nat Genet	2013	10	FGFR2	2.09E-01	3.44E-01
Thomas G	Nat Genet	2009	10	FGFR2	2.09E-01	3.44E-01
Gaudet MM	PLoS Genet	2010	10	FGFR2	2.09E-01	3.44E-01
Li J	Breast Cancer Res Treat	2010	10	FGFR2	2.09E-01	3.44E-01
Hunter DJ	Nat Genet	2007	10	FGFR2	2.09E-01	3.44E-01

Table 11: Association with Mortality of Genes Previously Associated with Breast Cancer Phenotypes (Continued)

first author	journal	year	chromo-some	gene	p-value in primary analysis	p-value in primary analysis of ER+ cases
Li J	Breast Cancer Res Treat	2010	10	FGFR2	2.09E-01	3.44E-01
Hunter DJ	Nat Genet	2007	10	FGFR2	2.09E-01	3.44E-01
Michailidou K	Nat Genet	2013	10	MLLT10	7.15E-01	5.22E-01
Michailidou K	Nat Genet	2013	10	TCF7L2	6.15E-01	8.82E-01
Michailidou K	Nat Genet	2013	10	ZMIZ1	3.53E-01	2.17E-01
Turnbull C	Nat Genet	2010	10	ZMIZ1	3.53E-01	2.17E-01
Michailidou K	Nat Genet	2013	10	ZNF365	8.88E-01	6.15E-01
Eriksson N	BMC Med Genet	2012	10	ZNF365	8.88E-01	6.15E-01
Cai Q	Hum Mol Genet	2011	10	ZNF365	8.88E-01	6.15E-01
Lindstrom S	Nat Commun	2014	10	ZNF365	8.88E-01	6.15E-01
Turnbull C	Nat Genet	2010	10	ZNF365	8.88E-01	6.15E-01
Lindstrom S	Nat Genet	2011	10	ZNF365	8.88E-01	6.15E-01
Easton DF	Nature	2007	11	LSP1	3.12E-01	5.98E-01
Lindstrom S	Nat Commun	2014	11	LSP1	3.12E-01	5.98E-01
Michailidou K	Nat Genet	2013	11	LSP1	3.12E-01	5.98E-01
Guo Q	J Natl Cancer Inst	2015	11	PKNOX2	9.00E-01	4.09E-01
Michailidou K	Nat Genet	2013	14	CCDC88C	2.16E-01	9.51E-01
Michailidou K	Nat Genet	2013	14	PAX9	5.83E-01	4.97E-01
Michailidou K	Nat Genet	2013	14	RAD51B	1.31E-01	7.97E-01
Orr N	Nat Genet	2012	14	RAD51B	1.31E-01	7.97E-01
Cai Q	Nat Genet	2014	15	PRC1-AS1	8.10E-01	8.69E-01
Michailidou K	Nat Genet	2013	16	CASC16	7.88E-01	5.78E-01
Turnbull C	Nat Genet	2010	16	CASC16	7.88E-01	5.78E-01
Thomas G	Nat Genet	2009	16	CASC16	7.88E-01	5.78E-01
Fejerman L	Nat Commun	2014	16	CASC16	7.88E-01	5.78E-01
Stacey SN	Nat Genet	2007	16	CASC16	7.88E-01	5.78E-01
Low SK	PLoS One	2013	16	CASC16	7.88E-01	5.78E-01
Garcia-Closas M	Nat Genet	2013	16	CASC16	7.88E-01	5.78E-01
Long J	PLoS Genet	2010	16	CASC16	7.88E-01	5.78E-01
Easton DF	Nature	2007	16	CASC16	7.88E-01	5.78E-01
Orr N	Nat Genet	2012	16	CASC16	7.88E-01	5.78E-01
Fletcher O	J Natl Cancer Inst	2011	16	CASC16	7.88E-01	5.78E-01
Michailidou K	Nat Genet	2013	16	CDYL2	9.40E-01	5.34E-02
Garcia-Closas M	Nat Genet	2013	16	FTO	6.31E-02	4.26E-02
Michailidou K	Nat Genet	2013	16	FTO	6.31E-02	4.26E-02
Ahsan H	Cancer Epidemiol Biomarkers Prev	2014	16	TOX3	8.30E-01	3.39E-01
Michailidou K	Nat Genet	2013	17	STXBP4	8.08E-01	9.49E-04
Michailidou K	Nat Genet	2013	18	CHST9	5.25E-01	6.86E-01
Purrington KS	Carcinogenesis	2013	19	ANKLE1	8.06E-01	1.43E-01
Siddiq A	Hum Mol Genet	2012	19	ANKLE1	8.06E-01	1.43E-01
Antoniou AC	Nat Genet	2010	19	BABAM1	7.62E-01	2.55E-01
Garcia-Closas M	Nat Genet	2013	19	BABAM1	7.62E-01	2.55E-01
Couch FJ	PLoS Genet	2013	19	BABAM1	7.62E-01	2.55E-01
Michailidou K	Nat Genet	2013	19	ELL	3.94E-02	8.10E-02
Siddiq A	Hum Mol Genet	2012	20	RALY	2.71E-01	5.48E-01
Michailidou K	Nat Genet	2013	22	EMID1	4.55E-01	2.23E-01
Michailidou K	Nat Genet	2013	22	MKL1	1.31E-01	3.16E-01
Lindstrom S	Nat Commun	2014	22	SGSM3	6.17E-01	2.63E-01
Lindstrom S	Nat Commun	2014	22	TMEM184B	6.21E-01	8.11E-01

Table 12: Most Significant Genes Identified by Logistic Regression with ER+ Tumor

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
ARHGAP29	23	9	9.47e-04	21631	216	2.25e-01
ARL10	14	2	2.94e-04	5050	33	1.66e-01
C12orf60	1666	5	4.53e-04	12083	80	2.37e-01
C9orf47	1661	7	7.96e-04	4675	33	5.66e-01
IL1RAP	976	5	9.24e-04	95561	761	1.42e-01
KATNA1	10	4	5.88e-04	42989	200	5.19e-01
KCNJ8	9	2	1.00e-03	397	14	5.37e-01
KLF10	13	4	7.49e-04	3667	38	1.00e+00
LRRK1	2511	23	1.08e-03	141047	954	8.03e-01
LY6G5B	1936	5	6.06e-04	1607	25	2.97e-01
MEIS3	17	4	1.40e-05	4214	36	7.19e-03
P3H3	2884	15	2.37e-04	12076	61	1.07e-01
POU5F1	5215	6	1.05e-03	17916	95	6.13e-01
PRDM5	2531	5	2.56e-04	205478	1257	8.89e-01
QRSL1	507	6	3.74e-04	37784	242	5.62e-02
SLC25A39	1676	2	3.99e-04	4748	30	1.00e+00
SLC38A4	210	3	8.98e-04	39639	356	5.01e-01
TMEM209	33	6	8.42e-04	24913	171	2.75e-01
TSPYL1	583	4	9.25e-06	4728	33	5.29e-01

Table 13: Most Significant Genes Identified by Logistic Regression with PR+ Tumor

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
AIM1	2169	16	3.34e-05	67054	411	1.00e+00
BCL9L	68	8	2.44e-04	2868	41	6.70e-01
C12orf42	1835	5	6.50e-04	111518	989	8.73e-01
CD68	913	5	8.33e-06	2203	15	8.79e-01
IL1RAP	964	5	9.08e-04	95092	761	7.60e-01
LIF	43	2	9.03e-04	4224	30	3.82e-01
MON1B	72	4	5.30e-04	15354	80	4.41e-01
MPDU1	572	4	1.65e-05	1635	14	1.00e+00
MYH6	1643	13	7.55e-04	13928	135	4.45e-01
NPM1	1554	1	9.19e-04	22640	86	4.36e-01
PCDHA4	1750	6	7.66e-04	4486	22	3.49e-02
PDIA4	118	13	4.99e-04	14658	139	6.83e-01
QRSL1	500	6	7.17e-06	37658	242	3.68e-01
RNF214	1259	2	5.71e-04	38468	274	3.76e-01
SHBG	403	4	4.25e-04	9594	68	8.89e-01
SLC4A7	2604	9	8.35e-04	81498	538	3.72e-01
TDRD5	2514	8	9.34e-04	88788	487	1.65e-01
TGFBI	46	16	2.13e-04	32096	183	6.85e-01
UBN1	141	11	7.77e-04	32105	185	1.68e-02
UTP23	540	2	2.20e-04	3557	43	2.69e-01

Table 14: Most Significant Genes Identified by Logistic Regression with HER2+ Tumor

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
ANGPT2	14	3	1.50e-03	80784	593	9.01e-01
CERS4	720	6	1.89e-04	32624	202	1.22e-01
CPM	23	2	3.06e-04	54047	636	6.27e-01
CROCC	199	13	1.01e-03	15275	152	1.42e-01
CTSL	8	2	7.12e-04	2255	31	8.91e-01
DACT2	515	6	5.63e-04	22474	203	4.29e-01
FAM171B	10	3	1.47e-03	23841	190	3.39e-01
GALNT16	1100	4	9.30e-04	64711	501	4.73e-01
GPT	661	9	1.50e-03	2003	8	6.99e-01
HEATR6	98	8	6.17e-04	2894	69	5.23e-01
MCM7	385	8	4.03e-04	3023	28	6.73e-01
METAP1D	556	6	1.27e-03	41095	442	1.25e-01
PCYOX1	245	2	1.01e-03	6244	88	4.32e-01
PIAS3	29	5	4.16e-04	34	1	9.17e-01
PIGP	226	3	4.14e-04	6035	49	8.13e-01
RECQL4	594	13	1.74e-04	4361	16	5.90e-01
RWDD4	291	1	7.73e-04	23355	141	2.62e-01
TFAP2B	495	4	9.89e-04	11301	131	7.63e-01
VWC2	757	5	5.61e-04	57463	576	1.00e+00
ZNF620	59	4	8.75e-04	3274	27	5.39e-01

Table 15: Most Significant Genes Identified by Logistic Regression with High Tumor Grade

Gene	Minor Allele Count	Variants in Gene	p-value
C8orf37-AS1	1057	2	5.47e-04
CNTNAP3	628	1	1.09e-04
COL27A1	7713	26	1.48e-04
CTDP1	604	5	1.21e-03
GUCY2C	26	5	5.01e-04
HOXC4	1596	3	2.52e-04
NPLOC4	1863	4	7.00e-04
NSUN2	686	5	2.36e-04
PNLIPRP3	306	8	5.98e-04
PNPLA7	329	14	1.05e-03
RBM27	15	3	7.47e-04
SCG2	84	4	2.06e-05
SH2D4A	837	7	2.21e-05
SSC4D	2202	6	7.28e-04
STRA6	2276	10	1.65e-04
TAF1C	1623	15	6.48e-04
TCTE1	1314	11	9.92e-04
TMEM132A	886	9	1.17e-03
TMEM88B	172	2	4.66e-04
ZBTB43	14	4	9.55e-04

Tumor grade was not available for the TCGA participants

Table 16: Most Significant Genes Identified by Logistic Regression with High Tumor Stage

	Exome Chip			TCGA		
Gene	Minor Allele Count	Variants in Gene	p-value	Minor Allele Count	Variants in Gene	p-value
C8orf37-AS1	1057	2	5.47e-04	525417	3137	5.96e-01
CNTNAP3	628	1	1.09e-04	42217	161	8.01e-01
COL27A1	7713	26	1.48e-04	165929	986	1.46e-01
CTDP1	604	5	1.21e-03	92110	465	1.00e+00
GUCY2C	26	5	5.01e-04	14059	333	5.68e-01
HOXC4	1596	3	2.52e-04	14994	116	2.28e-01
NPLOC4	1863	4	7.00e-04	96837	387	3.65e-01
NSUN2	686	5	2.36e-04	48493	236	2.07e-01
PNLIPRP3	306	8	5.98e-04	11371	200	6.80e-01
PNPLA7	329	14	1.05e-03	20244	89	1.00e+00
RBM27	15	3	7.47e-04	33619	222	6.18e-01
SCG2	84	4	2.06e-05	346	23	4.21e-01
SH2D4A	837	7	2.21e-05	47631	522	4.89e-01
SSC4D	2202	6	7.28e-04	27812	136	3.39e-01
STRA6	2276	10	1.65e-04	20707	136	2.85e-01
TAF1C	1623	15	6.48e-04	15310	98	4.25e-01
TCTE1	1314	11	9.92e-04	10972	116	3.86e-01
TMEM132A	886	9	1.17e-03	16624	65	1.31e-01
TMEM88B	172	2	4.66e-04	1058	11	7.67e-01
ZBTB43	14	4	9.55e-04	6467	111	6.02e-02

Table 17: Association with Tumor Characteristics of Genes Previously Associated with Breast Cancer Phenotypes

first author	journal	year	chro-mo-some	gene	p-value in ER analysis	p-value in PR analysis	p-value in HER2 analysis	p-value in high grade analysis	p-value in high stage analysis
Michailidou K	Nat Genet	2013	1	DCLRE1B	5.97E-01	8.44E-01	1.01E-01	3.88E-01	4.97E-01
Michailidou K	Nat Genet	2013	1	EMBP1	7.61E-01	9.15E-01	3.36E-01	6.49E-01	5.85E-01
Thomas G	Nat Genet	2009	1	EMBP1	7.61E-01	9.15E-01	3.36E-01	6.49E-01	5.85E-01
Barnett GC	Radiother Oncol	2014	1	KCND3	9.61E-02	6.94E-01	5.81E-01	2.31E-01	4.48E-01
Garcia-Closas M	Nat Genet	2013	1	LGR6	1.00E+00	6.72E-01	3.17E-02	1.35E-01	5.95E-01
Garcia-Closas M	Nat Genet	2013	1	MDM4	9.24E-01	8.34E-01	9.06E-01	2.17E-01	3.10E-01
Michailidou K	Nat Genet	2013	1	PEX14	1.00E+00	8.33E-01	4.70E-01	2.94E-01	6.39E-01
Garcia-Closas M	Nat Genet	2013	1	PEX14	1.00E+00	8.33E-01	4.70E-01	2.94E-01	6.39E-01
Cai Q	Nat Genet	2014	1	ZC3H11A	3.42E-01	6.15E-01	2.30E-01	3.78E-01	2.41E-01
Michailidou K	Nat Genet	2013	2	DIRC3	1.88E-01	1.00E+00	5.14E-01	1.00E+00	1.00E+00
Michailidou K	Nat Genet	2013	2	DLX2	7.36E-01	7.42E-01	1.00E+00	5.79E-01	8.61E-01
Kim HC	Breast Cancer Res	2012	2	ERBB4	1.86E-02	1.98E-02	3.91E-01	1.00E+00	5.71E-01
Michailidou K	Nat Genet	2013	3	ITPR1	7.43E-01	2.42E-01	3.15E-01	6.48E-01	8.65E-01
Ahsan H	Cancer Epi Bio Prev	2014	3	NEK10	1.22E-02	6.19E-03	7.79E-01	7.59E-03	5.61E-01
Michailidou K	Nat Genet	2013	3	SLC4A7	4.84E-03	8.35E-04	5.56E-01	1.23E-01	8.12E-01
Fletcher O	J Natl Cancer Inst	2011	3	SLC4A7	4.84E-03	8.35E-04	5.56E-01	1.23E-01	8.12E-01
Michailidou K	Nat Genet	2013	3	TGFBR2	3.52E-01	3.45E-01	8.78E-01	7.13E-01	4.49E-01
Michailidou K	Nat Genet	2013	4	ADAM29	7.51E-01	6.78E-01	1.00E+00	1.00E+00	3.66E-01
Michailidou K	Nat Genet	2013	4	TET2	1.09E-01	2.37E-01	9.37E-02	5.52E-01	6.11E-01
Michailidou K	Nat Genet	2013	5	EBF1	5.47E-01	2.81E-01	1.00E+00	1.00E+00	3.95E-01
Ahsan H	Cancer Epi Bio Prev	2014	5	MAP3K1	6.60E-01	6.16E-01	7.21E-01	6.86E-01	1.65E-03
Michailidou K	Nat Genet	2013	5	PDE4D	5.95E-02	2.93E-01	5.48E-01	5.55E-02	9.01E-01
Lindstrom S	Nat Commun	2014	5	PRDM6	8.60E-01	8.28E-01	5.87E-01	7.00E-01	3.93E-01
Garcia-Closas M	Nat Genet	2013	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Michailidou K	Nat Genet	2013	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Haiman CA	Nat Genet	2011	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Michailidou K	Nat Genet	2013	6	CCDC170	1.59E-01	3.26E-02	2.08E-01	5.05E-01	2.74E-01
Gold B	Proc Natl Acad Sci U S A	2008	6	RNF146	1.80E-01	1.11E-01	9.48E-01	5.22E-01	6.51E-01
Long J	PLoS Genet	2012	6	TAB2	5.77E-01	7.20E-01	7.86E-01	6.08E-01	5.05E-02
Easton DF	Nature	2007	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Michailidou K	Nat Genet	2013	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Ahsan H	Cancer Epi Bio Prev	2014	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Fletcher O	J Natl Cancer Inst	2011	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Liu M	Mol Endocrinol	2013	8	LOC101927066	2.17E-01	1.37E-01	6.10E-02	1.39E-01	8.07E-01
Turnbull C	Nat Genet	2010	9	CDKN2B-AS1	6.02E-01	7.27E-02	1.00E+00	3.00E-01	3.75E-01
Kiyotani K	Hum Mol Genet	2011	10	C10orf11	7.48E-01	4.40E-01	3.62E-01	3.25E-01	4.20E-02
Low SK	PLoS One	2013	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Fletcher O	J Natl Cancer Inst	2011	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Michailidou K	Nat Genet	2013	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Hunter DJ	Nat Genet	2007	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Thomas G	Nat Genet	2009	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Easton DF	Nature	2007	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Li J	Breast Cancer Res Treat	2010	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Ahsan H	Cancer Epi Bio Prev	2014	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Gaudet MM	PLoS Genet	2010	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01

Table 18: Association with Tumor Characteristics of Genes Previously Associated with Breast Cancer Phenotypes (Continued)

first author	journal	year	chro-mo-some	gene	p-value in ER analysis	p-value in PR analysis	p-value in HER2 analysis	p-value in high grade analysis	p-value in high stage analysis
Michailidou K	Nat Genet	2013	1	DCLRE1B	5.97E-01	8.44E-01	1.01E-01	3.88E-01	4.97E-01
Michailidou K	Nat Genet	2013	1	EMBP1	7.61E-01	9.15E-01	3.36E-01	6.49E-01	5.85E-01
Thomas G	Nat Genet	2009	1	EMBP1	7.61E-01	9.15E-01	3.36E-01	6.49E-01	5.85E-01
Barnett GC	Radiother Oncol	2014	1	KCND3	9.61E-02	6.94E-01	5.81E-01	2.31E-01	4.48E-01
Garcia-Closas M	Nat Genet	2013	1	LGR6	1.00E+00	6.72E-01	3.17E-02	1.35E-01	5.95E-01
Garcia-Closas M	Nat Genet	2013	1	MDM4	9.24E-01	8.34E-01	9.06E-01	2.17E-01	3.10E-01
Michailidou K	Nat Genet	2013	1	PEX14	1.00E+00	8.33E-01	4.70E-01	2.94E-01	6.39E-01
Garcia-Closas M	Nat Genet	2013	1	PEX14	1.00E+00	8.33E-01	4.70E-01	2.94E-01	6.39E-01
Cai Q	Nat Genet	2014	1	ZC3H11A	3.42E-01	6.15E-01	2.30E-01	3.78E-01	2.41E-01
Michailidou K	Nat Genet	2013	2	DIRC3	1.88E-01	1.00E+00	5.14E-01	1.00E+00	1.00E+00
Michailidou K	Nat Genet	2013	2	DLX2	7.36E-01	7.42E-01	1.00E+00	5.79E-01	8.61E-01
Kim HC	Breast Cancer Res	2012	2	ERBB4	1.86E-02	1.98E-02	3.91E-01	1.00E+00	5.71E-01
Michailidou K	Nat Genet	2013	3	ITPR1	7.43E-01	2.42E-01	3.15E-01	6.48E-01	8.65E-01
Ahsan H	Cancer Epi Bio Prev	2014	3	NEK10	1.22E-02	6.19E-03	7.79E-01	7.59E-03	5.61E-01
Michailidou K	Nat Genet	2013	3	SLC4A7	4.84E-03	8.35E-04	5.56E-01	1.23E-01	8.12E-01
Fletcher O	J Natl Cancer Inst	2011	3	SLC4A7	4.84E-03	8.35E-04	5.56E-01	1.23E-01	8.12E-01
Michailidou K	Nat Genet	2013	3	TGFBR2	3.52E-01	3.45E-01	8.78E-01	7.13E-01	4.49E-01
Michailidou K	Nat Genet	2013	4	ADAM29	7.51E-01	6.78E-01	1.00E+00	1.00E+00	3.66E-01
Michailidou K	Nat Genet	2013	4	TET2	1.09E-01	2.37E-01	9.37E-02	5.52E-01	6.11E-01
Michailidou K	Nat Genet	2013	5	EBF1	5.47E-01	2.81E-01	1.00E+00	1.00E+00	3.95E-01
Ahsan H	Cancer Epi Bio Prev	2014	5	MAP3K1	6.60E-01	6.16E-01	7.21E-01	6.86E-01	1.65E-03
Michailidou K	Nat Genet	2013	5	PDE4D	5.95E-02	2.93E-01	5.48E-01	5.55E-02	9.01E-01
Lindstrom S	Nat Commun	2014	5	PRDM6	8.60E-01	8.28E-01	5.87E-01	7.00E-01	3.93E-01
Garcia-Closas M	Nat Genet	2013	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Michailidou K	Nat Genet	2013	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Haiman CA	Nat Genet	2011	5	TERT	1.75E-02	4.97E-02	1.00E+00	7.74E-01	4.62E-01
Michailidou K	Nat Genet	2013	6	CCDC170	1.59E-01	3.26E-02	2.08E-01	5.05E-01	2.74E-01
Gold B	Proc Natl Acad Sci U S A	2008	6	RNF146	1.80E-01	1.11E-01	9.48E-01	5.22E-01	6.51E-01
Long J	PLoS Genet	2012	6	TAB2	5.77E-01	7.20E-01	7.86E-01	6.08E-01	5.05E-02
Easton DF	Nature	2007	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Michailidou K	Nat Genet	2013	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Ahsan H	Cancer Epi Bio Prev	2014	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Fletcher O	J Natl Cancer Inst	2011	8	CASC8	1.05E-01	2.47E-01	1.00E+00	2.14E-01	6.42E-01
Liu M	Mol Endocrinol	2013	8	LOC101927066	2.17E-01	1.37E-01	6.10E-02	1.39E-01	8.07E-01
Turnbull C	Nat Genet	2010	9	CDKN2B-AS1	6.02E-01	7.27E-02	1.00E+00	3.00E-01	3.75E-01
Kiyotani K	Hum Mol Genet	2011	10	C10orf11	7.48E-01	4.40E-01	3.62E-01	3.25E-01	4.20E-02
Low SK	PLoS One	2013	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Fletcher O	J Natl Cancer Inst	2011	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Michailidou K	Nat Genet	2013	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Hunter DJ	Nat Genet	2007	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Thomas G	Nat Genet	2009	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Easton DF	Nature	2007	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Li J	Breast Cancer Res Treat	2010	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Ahsan H	Cancer Epi Bio Prev	2014	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01
Gaudet MM	PLoS Genet	2010	10	FGFR2	1.95E-01	1.51E-01	4.69E-01	5.03E-02	4.98E-01