A network analysis to identify mediators of germline-driven differences in breast cancer prognosis

Maria Escala-Garcia[1], Jean Abraham[2-4], Irene L. Andrulis[5, 6], Hoda Anton-Culver[7], Volker Arndt[8], Alan Ashworth[9], Paul L. Auer[10, 11], Päivi Auvinen[12-14], Matthias W. Beckmann[15], Jonathan Beesley[16], Sabine Behrens[17], Javier Benitez[18, 19], Marina Bermisheva[20], Carl Blomqvist[21, 22], William Blot[23, 24], Natalia V. Bogdanova[25-27], Stig E. Bojesen[28-30], Manjeet K. Bolla[31], Anne-Lise Børresen-Dale[32, 33], Hiltrud Brauch[34-36], Hermann Brenner[8, 36, 37], Sara Y. Brucker[38], Barbara Burwinkel[39, 40], Carlos Caldas[41, 42], Federico Canzian[43], Jenny Chang-Claude[17, 44], Stephen J. Chanock[45], Suet-Feung Chin[46], Christine L. Clarke[47], Fergus J. Couch[48], Angela Cox[49], Simon S. Cross[50], Kamila Czene[51], Mary B. Daly[52], Joe Dennis[31], Peter Devilee[53, 54], Janet A. Dunn[55], Alison M. Dunning[2], Miriam Dwek[56], Helena M. Earl[4, 57], Diana M. Eccles[58], A. Heather Eliassen[59, 60], Carolina Ellberg[61], D. Gareth Evans[62-64], Peter A. Fasching[15, 65], Jonine Figueroa[45, 66, 67], Henrik Flyger[68], Manuela Gago-Dominguez[69, 70], Susan M. Gapstur[71], Montserrat García-Closas[45, 72], José A. García-Sáenz[73], Mia M. Gaudet[71], Angela George[74, 75], Graham G. Giles[76-78], David E. Goldgar[79], Anna González-Neira[18], Mervi Grip[80], Pascal Guénel[81], Qi Guo[82], Christopher A. Haiman[83], Niclas Håkansson[84], Ute Hamann[85], Patricia A. Harrington[2], Louise Hiller[55], Maartje J. Hooning[86], John L. Hopper[77], Anthony Howell[87], Chiun-Sheng Huang[88], Guanmengqian Huang[85], David J. Hunter[60, 89, 90], Anna Jakubowska[91, 92], Esther M. John[93], Rudolf Kaaks[17], Pooja M. Kapoor[17, 94], Renske Keeman[1], Cari M. Kitahara[95], Linetta B. Koppert[96], Peter Kraft[60, 89], Vessela N. Kristensen[32, 33], Diether Lambrechts[97, 98], Loic Le Marchand[99], Flavio Lejbkowicz[100], Annika Lindblom[101, 102], Jan Lubiński[91], Arto Mannermaa[14, 103, 104], Mehdi Manoochehri[85], Siranoush Manoukian[105], Sara Margolin[106, 107], Maria Elena Martinez[70, 108], Tabea Maurer[44], Dimitrios Mavroudis[109], Alfons Meindl[110], Roger L. Milne[76, 77, 111], Anna Marie Mulligan[112, 113], Susan L. Neuhausen[114], Heli Nevanlinna[115], William G. Newman[62, 63], Andrew F. Olshan[116], Janet E. Olson[117], Håkan Olsson[61], Nick Orr[118], Paolo Peterlongo[119], Christos Petridis[120], Ross L. Prentice[10], Nadege Presneau[56], Kevin Punie[121], Dhanya Ramachandran[26], Gad Rennert[100], Atocha Romero[122], Mythily Sachchithananthan[47], Emmanouil Saloustros[123], Elinor J. Sawyer[120], Rita K. Schmutzler[124, 125], Lukas Schwentner[126], Christopher Scott[117], Jacques Simard[127], Christof Sohn[128], Melissa C. Southey[111, 129], Anthony J. Swerdlow[74, 130], Rulla M. Tamimi[59, 60, 89], William J. Tapper[131], Manuel R. Teixeira[132, 133], Mary Beth Terry[134], Heather Thorne [135, 136], Rob A.E.M. Tollenaar[137], Ian Tomlinson[138, 139], Melissa A. Troester[116], Thérèse Truong[81], Clare Turnbull[74], Celine M. Vachon[117], Lizet E. van der Kolk[140], Qin Wang[31], Robert Winqvist[141, 142], Alicja Wolk[84, 143], Xiaohong R. Yang[45], Argyrios Ziogas[7], Paul D.P. Pharoah[2, 31], Per Hall[51, 106], Lodewyk F. A. Wessels[144, 145], Georgia Chenevix-Trench[16], Gary D. Bader[6, 146], Thilo Dörk[26], Douglas F. Easton[2, 31], Sander Canisius[1, 144]* and Marjanka K. Schmidt[1, 147]*.

* Shared last and corresponding authors

[1] The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Division of Molecular Pathology, Amsterdam, The Netherlands.
[2] University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Oncology, Cambridge, UK.
[3] Cambridge Experimental Cancer Medicine Centre, Cambridge, UK.
[4] University of Cambridge NHS Foundation Hospitals, Cambridge Breast Unit and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.

[5] Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Fred A. Litwin Center for Cancer Genetics, Toronto, ON, Canada.
[6] Department of Molecular Genetics, University of Toronto, Toronto, Canada.
[7] University of California Irvine, Department of Epidemiology, Genetic Epidemiology Research Institute, Irvine, CA, USA.
[8] German Cancer Research Center (DKFZ), Division of Clinical Epidemiology and Aging Research, Heidelberg, Germany.
[9] UCSF Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA.
[10] Fred Hutchinson Cancer Research Center, Cancer Prevention Program, Seattle, WA, USA.
[11] University of Wisconsin-Milwaukee, Zilber School of Public Health, Milwaukee, WI, USA.
[12] Kuopio University Hospital, Cancer Center, Kuopio, Finland.
[13] University of Eastern Finland, Institute of Clinical Medicine, Oncology, Kuopio, Finland.
[14] University of Eastern Finland, Translational Cancer Research Area, Kuopio, Finland.
[15] University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, Erlangen, Germany.
[16] QIMR Berghofer Medical Research Institute, Department of Genetics and Computational Biology, Brisbane, Queensland, Australia.
[17] German Cancer Research Center (DKFZ), Division of Cancer Epidemiology, Heidelberg, Germany.
[18] Spanish National Cancer Research Centre (CNIO), Human Cancer Genetics Programme, Madrid, Spain.
[19] Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain.
[20] Ufa Scientific Center of Russian Academy of Sciences, Institute of Biochemistry and Genetics, Ufa, Russia.
[21] University of Helsinki, Department of Oncology, Helsinki University Hospital, Helsinki, Finland.
[22] Örebro University Hospital, Department of Oncology, Örebro, Sweden.
[23] Vanderbilt University School of Medicine, Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Nashville, TN, USA.
[24] International Epidemiology Institute, Rockville, MD, USA.
[25] Hannover Medical School, Department of Radiation Oncology, Hannover, Germany.
[26] Hannover Medical School, Gynaecology Research Unit, Hannover, Germany.
[27] N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus.
[28] Copenhagen University Hospital, Copenhagen General Population Study, Herlev and Gentofte Hospital, Herlev, Denmark.
[29] Copenhagen University Hospital, Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Herlev, Denmark.
[30] University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, Denmark.
[31] University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Cambridge, UK.
[32] Oslo University Hospital-Radiumhospitalet, Department of Cancer Genetics, Institute for Cancer Research, Oslo, Norway.
[33] University of Oslo, Institute of Clinical Medicine, Faculty of Medicine, Oslo, Norway.
[34] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany.
[35] iFIT-Cluster of Excellence, University of Tuebingen, Tuebingen, Germany.
[36] German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany.
[37] German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Division of Preventive Oncology, Heidelberg, Germany.
[38] University of Tübingen, Department of Gynecology and Obstetrics, Tübingen, Germany.
[39] German Cancer Research Center (DKFZ), Molecular Epidemiology Group, C080, Heidelberg, Germany.
[40] University of Heidelberg, Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, Heidelberg, Germany.
[41] Cancer Research UK Cambridge Institute, Department of Oncology, University of Cambridge, Li Ka Shing Centre, Cambridge, UK.
[42] Breast Cancer Programme, CRUK Cambridge Cancer Centre and NIHR Biomedical Research Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

[43] German Cancer Research Center (DKFZ), Genomic Epidemiology Group, Heidelberg, Germany.
[44] University Medical Center Hamburg-Eppendorf, Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), Hamburg, Germany.
[45] National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA.
[46] University of Cambridge, Cancer Research UK Cambridge Institute, Cambridge, UK.
[47] University of Sydney, Westmead Institute for Medical Research, Sydney, New South Wales, Australia.
[48] Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN, USA.
[49] University of Sheffield, Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, Sheffield, UK.
[50] University of Sheffield, Academic Unit of Pathology, Department of Neuroscience, Sheffield, UK.
[51] Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden.
[52] Fox Chase Cancer Center, Department of Clinical Genetics, Philadelphia, PA, USA.
[53] Leiden University Medical Center, Department of Pathology, Leiden, The Netherlands.
[54] Leiden University Medical Center, Department of Human Genetics, Leiden, The Netherlands.
[55] University of Warwick, Warwick Clinical Trials Unit, Coventry, UK.
[56] University of Westminster, Department of Biomedical Sciences, Faculty of Science and Technology, London, UK.
[57] University of Cambridge, Department of Oncology, Cambridge, UK.
[58] University of Southampton, Cancer Sciences Academic Unit, Faculty of Medicine, Southampton, UK.
[59] Harvard Medical School, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[60] Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA, USA.
[61] Lund University, Department of Cancer Epidemiology, Clinical Sciences, Lund, Sweden.
[62] University of Manchester, Manchester Academic Health Science Centre, Division of Evolution and Genomic Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester, UK.
[63] Genomic Medicine, St Mary's Hospital, Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester Centre for Genomic Medicine, Manchester, UK.
[64] Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Manchester, UK.
[65] University of California at Los Angeles, David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, Los Angeles, CA, USA.
[66] The University of Edinburgh Medical School, Usher Institute of Population Health Sciences and Informatics, Edinburgh, UK.
[67] Cancer Research UK Edinburgh Centre, Edinburgh, UK.
[68] Copenhagen University Hospital, Department of Breast Surgery, Herlev and Gentofte Hospital, Herlev, Denmark.
[69] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Genomic Medicine Group, Galician Foundation of Genomic Medicine, Santiago de Compostela, Spain.
[70] University of California San Diego, Moores Cancer Center, La Jolla, CA, USA.
[71] American Cancer Society, Epidemiology Research Program, Atlanta, GA, USA.
[72] Institute of Cancer Research, Division of Genetics and Epidemiology, London, UK.
[73] Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Medical Oncology Department, Hospital Clínico San Carlos, Madrid, Spain.
[74] The Institute of Cancer Research, Division of Genetics and Epidemiology, London, UK.
[75] The Institute of Cancer Research, Section of Cancer Genetics, London, UK.
[76] Cancer Council Victoria, Cancer Epidemiology Division, Melbourne, Victoria, Australia.
[77] The University of Melbourne, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, Melbourne, Victoria, Australia.
[78] Monash University, Department of Epidemiology and Preventive Medicine, Melbourne, Victoria, Australia.
[79] Huntsman Cancer Institute, University of Utah School of Medicine, Department of Dermatology, Salt Lake City, UT, USA.

[80] University of Oulu, Department of Surgery, Oulu University Hospital, Oulu, Finland.
[81] INSERM, University Paris-Sud, University Paris-Saclay, Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), Villejuif, France.
[82] University of Cambridge, Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, Cambridge, UK.
[83] University of Southern California, Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA.
[84] Karolinska Institutet, Institute of Environmental Medicine, Stockholm, Sweden.
[85] German Cancer Research Center (DKFZ), Molecular Genetics of Breast Cancer, Heidelberg, Germany.
[86] Erasmus MC Cancer Institute, Department of Medical Oncology, Family Cancer Clinic, Rotterdam, The Netherlands.
[87] University of Manchester, Division of Cancer Sciences, Manchester, UK.
[88] National Taiwan University Hospital and National Taiwan University College of Medicine, Department of Surgery, Taipei, Taiwan
[89] Harvard T.H. Chan School of Public Health, Program in Genetic Epidemiology and Statistical Genetics, Boston, MA, USA.
[90] University of Oxford, Nuffield Department of Population Health, Oxford, UK.
[91] Pomeranian Medical University, Department of Genetics and Pathology, Szczecin, Poland.
[92] Pomeranian Medical University, Independent Laboratory of Molecular Biology and Genetic Diagnostics, Szczecin, Poland.
[93] Stanford Cancer Institute, Stanford University School of Medicine, Department of Medicine, Division of Oncology, Stanford, CA, USA.
[94] University of Heidelberg, Faculty of Medicine, Heidelberg, Germany.
[95] National Cancer Institute, Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA.
[96] Erasmus MC Cancer Institute, Department of Surgical Oncology, Family Cancer Clinic, Rotterdam, The Netherlands.
[97] VIB, VIB Center for Cancer Biology, Leuven, Belgium.
[98] University of Leuven, Laboratory for Translational Genetics, Department of Human Genetics, Leuven, Belgium.
[99] University of Hawaii Cancer Center, Epidemiology Program, Honolulu, HI, USA.
[100] Carmel Medical Center and Technion Faculty of Medicine, Clalit National Cancer Control Center, Haifa, Israel.
[101] Karolinska Institutet, Department of Molecular Medicine and Surgery, Stockholm, Sweden.
[102] Karolinska University Hospital, Department of Clinical Genetics, Stockholm, Sweden.
[103] University of Eastern Finland, Institute of Clinical Medicine, Pathology and Forensic Medicine, Kuopio, Finland.
[104] Kuopio University Hospital, Imaging Center, Department of Clinical Pathology, Kuopio, Finland.
[105] Fondazione IRCCS Istituto Nazionale dei Tumori di Milano (INT), Unit of Medical Genetics, Department of Medical Oncology and Hematology, Milan, Italy.
[106] Sšdersjukhuset, Department of Oncology, Stockholm, Sweden.
[107] Karolinska Institutet, Department of Clinical Science and Education, Sšdersjukhuset, Stockholm, Sweden.
[108] University of California San Diego, Department of Family Medicine and Public Health, La Jolla, CA, USA.
[109] University Hospital of Heraklion, Department of Medical Oncology, Heraklion, Greece.
[110] Ludwig Maximilian University of Munich, Department of Gynecology and Obstetrics, Munich, Germany.
[111] Monash University, Precision Medicine, School of Clinical Sciences at Monash Health, Clayton, Victoria, Australia.
[112] University of Toronto, Department of Laboratory Medicine and Pathobiology, Toronto, ON, Canada.
[113] University Health Network, Laboratory Medicine Program, Toronto, ON, Canada.
[114] Beckman Research Institute of City of Hope, Department of Population Sciences, Duarte, CA, USA.
[115] University of Helsinki, Department of Obstetrics and Gynecology, Helsinki University Hospital, Helsinki, Finland.
[116] University of North Carolina at Chapel Hill, Department of Epidemiology, Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA.

[117] Mayo Clinic, Department of Health Sciences Research, Rochester, MN, USA.
[118] Queen's University Belfast, Centre for Cancer Research and Cell Biology, Belfast, Ireland, UK.
[119] IFOM - the FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, Genome Diagnostics Program, Milan, Italy.
[120] King's College London, Research Oncology, Guy's Hospital, London, UK.
[121] Leuven Cancer Institute, University Hospitals Leuven, Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven, Belgium.
[122] Hospital Universitario Puerta de Hierro, Medical Oncology Department, Madrid, Spain.
[123] University Hospital of Larisa, Department of Oncology, Larisa, Greece.
[124] University Hospital of Cologne, Center for Hereditary Breast and Ovarian Cancer, Cologne, Germany.
[125] University of Cologne, Center for Molecular Medicine Cologne (CMMC), Cologne, Germany.
[126] University Hospital Ulm, Department of Gynaecology and Obstetrics, Ulm, Germany.
[127] Centre Hospitalier Universitaire de Québec - Université Laval, Research Center, Genomics Center, Québec City, QC, Canada.
[128] University of Heidelberg, National Center for Tumor Diseases, Heidelberg, Germany.
[129] The University of Melbourne, Department of Clinical Pathology, Melbourne, Victoria, Australia.
[130] The Institute of Cancer Research, Division of Breast Cancer Research, London, UK.
[131] University of Southampton, Faculty of Medicine, Southampton, UK.
[132] Portuguese Oncology Institute, Department of Genetics, Porto, Portugal.
[133] University of Porto, Biomedical Sciences Institute (ICBAS), Porto, Portugal.
[134] Columbia University, Department of Epidemiology, Mailman School of Public Health, New York, NY, USA.
[135] Peter MacCallum Cancer Center, Melbourne, Victoria, Australia.
[136] The University of Melbourne, Sir Peter MacCallum Department of Oncology, Melbourne, Victoria, Australia.
[137] Leiden University Medical Center, Department of Surgery, Leiden, The Netherlands.
[138] University of Birmingham, Institute of Cancer and Genomic Sciences, Birmingham, UK.
[139] University of Oxford, Wellcome Trust Centre for Human Genetics and Oxford NIHR Biomedical Research Centre, Oxford, UK.
[140] The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Family Cancer Clinic, Amsterdam, The Netherlands.
[141] University of Oulu, Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, Oulu, Finland.
[142] Northern Finland Laboratory Centre Oulu, Laboratory of Cancer Genetics and Tumor Biology, Oulu, Finland.
[143] Uppsala University, Department of Surgical Sciences, Uppsala, Sweden.
[144] The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Division of Molecular Carcinogenesis, Amsterdam, The Netherlands.
[145] Faculty of EEMCS, Delft University of Technology, Delft, The Netherlands.
[146] The Donnelly Centre, University of Toronto, Toronto, Canada.
[147] The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Division of Psychosocial Research and Epidemiology, Amsterdam, The Netherlands.

Abstract

Identifying the underlying genetic drivers of the heritability of breast cancer prognosis remains elusive. We adapt a network-based approach to handle underpowered complex datasets to provide new insights into the potential function of germline variants in breast cancer prognosis. This network-based analysis studies ~7.3 million variants in 84,457 breast cancer patients in relation to breast cancer survival and confirms the results on 12,381 independent patients.

Aggregating the prognostic effects of genetic variants across multiple genes, we identify four gene modules associated with survival in estrogen receptor (ER)-negative and one in ER-positive disease. The modules show biological enrichment for cancer-related processes such as G-alpha signaling, circadian clock, angiogenesis, and Rho-GTPases in apoptosis.

Family-based studies have suggested that breast cancer survival in first-degree relatives has a hereditary component[1,2]. Nevertheless, whereas large scale genome-wide association studies (GWAS) have made considerable progress in identifying germline variants linked to breast cancer risk[3], the identification of germline variants linked to breast cancer prognosis has proven more challenging[4]. An understanding of how and which germline variants affect breast cancer prognosis could provide novel insights into the etiology of the metastatic process in breast cancer, increase knowledge on the underlying heterogeneity of the disease, and help identify new therapeutic targets or select patients most likely to benefit from existing therapies.

A major limitation of the studies to date is that the sample sizes have been insufficient to detect the small effect sizes of germline variants characteristic for breast cancer risk and survival[4–6]. Even though our previous survival GWAS included over 95,000 patients[4,5], the limiting factor was the relatively low number of events (breast cancer-specific deaths) observed. One way to overcome this limited power is to use pathway or network-based approaches[7,8]. These techniques typically use predefined gene sets, annotated pathways or protein-protein interaction (PPI) networks to detect genetic effects across multiple genes or proteins with similar or related biological functions[6,8–10]. Using such methods, a biological pathway might emerge as relevant even if none of its individual germline variants reached genome-wide significance. Moreover, assigning the variants to genes reduces dimensionality: considering several pathways as opposed to millions of individual variants leads to a substantial reduction in the number of tests performed[11]. An additional advantage of performing a pathway analysis is that it naturally suggests which biological processes mediate the genetic association with survival, making the biological interpretation easier[7,11–13].

Here we report on a network-based GWAS to identify genetic determinants of breast cancer prognosis in a dataset with a total of 84,457 breast cancer patients of European ancestry. In line with previous studies, we did not find many individual genetic variants with strong effects[14–17]. However, aggregating the survival estimates of multiple variants across genes and using a network propagation method, we identified several biological processes that may mediate a germline genetic effect on breast cancer prognosis. These include key processes in cancer biology, such as regulation of apoptosis, G-alpha signaling, and the circadian clock mechanism. In our analysis, we show that the identified polygenic effects are associated with survival not only in the discovery set, but also in an independent dataset of 12,381 patients. In addition, we studied the downstream transcriptional changes and their functional consequences due to the prognostic variants. We observed similar biological processes in the enrichment of the downstream and module-level gene analyses suggesting that both levels are perturbed by the identified genetic variants.

RESULTS

Single variant and gene analyses detect one independent hit

We performed an analysis of the association between germline genetic variants and breast cancer prognosis comprising data for 84,457 female breast cancer patients of European ancestry. To account for potential subtype-specific associations, we also performed separate analyses for ER-positive and ER-negative breast cancer. An overview of all data is given in the Methods section & **Supplementary Table 1**. As a first step in our analysis, we tested the association of ~7.3 million imputed genetic variants with breast cancer-specific survival using a Cox proportional hazard model **(Fig. 1a)**. Based on a genome-wide statistical significance P value threshold of 5×10^{-8}, we identified two variants at 8q13, in high linkage disequilibrium with each other, associated with survival in ER-positive breast cancer. The top variant was rs6990375 (chr8:70571531, $P = 6.35 \times 10^{-9}$) followed by rs13272847 (chr8:70573316, $P = 1.07 \times 10^{-8}$) . We did not find significant variants for ER-negative or all breast cancer cases.

Next, we aggregated the summary statistics of the individual variants into gene-level P values (~21,800 genes in total) using the Pascal algorithm[12] **(Fig. 1b)**. We computed the gene score based on the maximum chi-squared signal within a window size of 50-kb around the gene region (see Methods) **(Fig. 2)**. Two genes were associated with survival in ER-positive breast cancer at P < 0.05 after Bonferroni correction: *SLCO5A1* ($P = 4 \times 10^{-7}$, corrected P = 0.01) and *SULF1* ($P = 7 \times 10^{-7}$, corrected P = 0.02) **(Fig. 2c)**. These two genes are located in close proximity to each other around the significant variants at 8q13 identified in the single variant analysis. Their significance is therefore likely driven by a single causal genetic variant. The top variant rs6990375 is situated in the 3' UTR of *SULF1* where it may affect the binding of regulatory micro-RNAs. While the association of this variant with breast cancer survival has not been identified previously, it has been reported to be associated with age of onset of ovarian cancer[18]. *SULF1* has been found to be involved in cell proliferation, migration, and invasion as well as drug-induced apoptosis in cancer cell lines[19], most likely due to its regulatory role in FGF[20] and Wnt signaling[21]. Less is known about the function of *SLCO5A1*, although a role in cell proliferation has been suggested[22] . In line with the single variant analysis, we found no significant genes for all breast cancer or ER-negative breast cancer **(Fig. 2a,b)** when aggregating individual variants into genes.

Network analysis finds germline-related prognostic modules

To explore whether weaker signals of association were hidden in our data, we investigated the hypothesis that the germline genetic variants associated with breast cancer prognosis target particular biological processes, but within those processes do not uniquely target one particular gene. Different subgroups of patients might harbor variants in different genes, which ultimately affect the same biological process. Such polygenic signals, unless they have very big effects, may remain undetected if only individual variants or even individual genes are tested. We therefore applied network propagation[23], a technique that maps gene association scores onto a protein-protein interaction (PPI) network and uses the network topology to detect sub-networks, or modules, of closely interacting, high-scoring proteins **(Fig. 1c)**. In the context of this paper, we will refer to these modules also as germline-related prognostic modules (GRPMs).

For the network propagation, we used the HotNet2 method[13], which has been used previously with GWAS data[24]. We based the gene scores on the aggregate gene P values computed by the Pascal method (see Methods). The protein interaction network used by HotNet2 was obtained from iRefIndex[25].

When considering all breast cancers, the HotNet2 analysis identified no significant GRPMs (lowest P = 0.06, based on the HotNet2 permutation test). In contrast, several GRPMs were associated with prognosis in the analyses by ER subtype. For ER-positive patients, the best HotNet2 result (P value < 0.01) comprised 31 GRPMs of seven or more genes. For ER-negative patients, the best HotNet2 results (P < 0.01) included 116 GRPMs of four or more genes. A list of all significant prognostic modules is presented in Supplementary Data 1.

To help the interpretation of the identified GRPMs, we developed an extension to HotNet2 that maps the module genes to the specific genetic variants that are most strongly associated with prognosis. This was done by performing a Lasso-penalized Cox regression on the genetic variants assigned to the module genes. Using those selected variants and their effect sizes, a polygenic hazard score (PHS) was computed and used to identify a set of high-confidence GRPMs (Fig. 1d), as well as to perform a functional characterization of the downstream effects of the prognostic variants (Fig. 1e).

Prognostic modules point to underlying pathways

We restricted our scope to a subset of high-confidence GRPMs. This subset was identified by testing the association of each module's PHS with breast cancer prognosis in an independent set of 12,381 patients (with 1,120 events) (Supplementary Table 2) that were not used previously in the HotNet2 analysis or in the construction of the PHS score. GRPMs with a significant association between PHS and prognosis (P value < 0.05, based on a one-sided Wald test) in this independent set were considered high-confidence. Following this procedure, we found four high-confidence GRPMs for ER-negative breast cancer (Fig. 3a-c) and one high-confidence GRPM for ER-positive breast cancer (Fig. 3d). Hazard ratios of the association of the PHSs with breast

cancer-specific survival ranged from 1.09 to 1.28 **(Fig. 3e)**. In the remainder of this section we will discuss the high-confidence GRPMs. The term PHS P value will be used to refer to the P value of a GRPM's PHS association with survival.

To provide a functional characterization of the five high-confidence GRPMs found in the ER-negative and ER-positive subtypes, we tested each module for enriched biological processes on two levels. The first, which we call the module-level, considers the direct functions of the GRPM proteins themselves. These were identified by an enrichment analysis of the annotated biological functions of the module proteins and their direct interactors in a PPI network annotation (see Methods). For the high-confidence GRPMs in ER-negative breast cancer we identified enriched processes related to G-alpha signaling, cell growth and angiogenesis, insulin secretion and circadian clock **(Supplementary Fig. 1a-d)**. For the ER-positive high-confidence GRPM, the enriched processes included signaling by Rho GTPases and apoptosis **(Supplementary Fig. 1e)**.

The module-level enrichment provides a general summary of the biological functions of the GRPM genes. However, it is based on functional annotations that have been derived from studies in many different cell types and biological environments. To study the specific downstream effects of the identified prognostic variants in breast cancer tumors, we performed enrichment analyses on the downstream transcriptional changes due to the prognostic variants affecting the module proteins.

We estimated these downstream transcriptional effects using genetic variants and RNA expression data of female breast cancer patients from The Cancer Genome Atlas (TCGA)[26]. For each of the five GRPMs, the downstream analysis was performed on the subset of TCGA patients matching the ER subtype in which the GRPM was identified, 118 patients with ER-negative and 440 with ER-positive tumors. Using the germline genotype data of these TCGA patients, we computed the PHS for each GRPM **(Supplementary Table 3)**. Based on these PHSs, we then computed GRPM downstream transcriptional effect scores, which reflect the correlation between a module's PHS and the mRNA expression level of every gene **(Fig. 1e)** (see Methods).

Using the obtained downstream transcriptional effect scores, we performed Gene Set Enrichment Analysis (GSEA)[27] with gene sets based on Reactome[28] and the MSigDB[29] Hallmark gene sets. The enrichment results for the MSigDB Hallmark gene sets are shown in **Figure 3**, only pathways with a GSEA P value < 0.001 and FDR < 0.01 were included in the visualization. The full list of enriched processes per high-confidence GRPM can be found in **Supplementary Data 2-6** and **Supplementary Figure 2**.

The enriched pathways in the downstream analysis included biological processes such as cell cycle, DNA repair, metabolism of RNA, lipids or proteins, apoptosis, and translation of proteins. Importantly, we observed overlap of the biological processes enriched in the downstream analysis and those found for the module proteins. This observation has two important implications. First, it provides additional support for the biological role assigned to the module proteins. In addition to this, in cases where module proteins may serve several roles, it helps identify which of those roles is affected by the prognostic variants at a transcriptional level. The enriched biological processes assigned to the modules and the related downstream processes are described below.

ER-negative: G-alpha signaling events

Two high-confidence GRPMs found for patients with ER-negative tumors **(Fig. 3a)** suggested, from the module-level analysis, G-alpha signaling and G-protein activation as biological processes associated with survival. The first GRPM (PHS P = 0.0096) includes *ADCY10, GNA11, PTGIR* and *RGS3* **(Fig. 3a, right)** and the other GRPM (PHS P = 0.0082) is a larger module of 19 genes: *ADRBK2, CCL16, CNR2, CXCR5, DNAJB4, F2R, GNA15, GNAT1, GRM4, GUCA1A, GUCA1B, GUCA2B, GUCY2D, HRH4, LTB4R, OPRK1, OPRM1, RGS9* and *RGS9BP* **(Fig. 3a, left)**.

On closer inspection of the genetic variants selected for the two modules' PHSs, we observed that one genetic variant was shared by both modules. The other variants in the PHSs, two variants in total for the four-gene module and three variants for the module of 19 genes, were also located in the same genomic region on chromosome 19p13.3 **(Fig. 4a)**. These variants are

upstream of *GNA11* in the former module and *GNA15* in the latter. For the other genes in these two GRPMs, no genetic variants were selected as part of the modules' PHSs. This may be due to lack of statistical power: although the gene scores were high enough to be included in the module, none of their individual genetic variants had a strong enough association. The co-location of *GNA11* and *GNA15* provides an explanation for why the identified variants were selected for both modules. It also suggests that the genetic associations of these two genes and hence of the two modules are not independent. Indeed, the patients' PHSs for both GRPMs are highly correlated **(Fig. 4b)**, which supports a shared genetic association. This raises the question of whether the putative germline genetic effect on survival is mediated through both genes or only one of the two. In the downstream analyses of both modules, changes of *GNA15* expression were identified as one of the strongest downstream transcriptional effects, whereas this is not the case for *GNA11*. Conversely, in an independent gene expression dataset using KMplotter (kmplot.com/analysis), we found that expression of *GNA11* is significantly associated with recurrence free survival in ER-negative breast cancer **(Supplementary Fig. 3)**, while a similar effect was not seen for *GNA15*. These preliminary observations leave open the hypothesis of a role for both genes. A definitive answer will require more functional analyses.

In the module-level analysis, the GRPM formed by four genes also showed enrichment for insulin secretion. It has been shown that there is a close relationship between G-proteins and their coupled receptors (GPCR), insulin and the insulin-like growth factor I receptor (IGFIR). Altered versions of this crosstalk could play a role in cancer cells[30,31]. For example, it has been proposed that in cancer cells, insulin can increase the activity of GPCRs in cancer tissues via the mTOR (mammalian target of rapamycin) pathway[31], which was also one of the enriched processes in the downstream analysis. The highest scoring gene in the module, *GNA11*, codes for the alpha subunit of the G_{11} protein, which has been linked to insulin secretion and signaling[32,33].

For the 19-gene GRPM, we also identified thrombin signaling and platelet aggregation as two of the main module-level enriched pathways. Thrombin is a type of the above mentioned GPCRs with the capacity to upregulate genes able to induce, or contribute to oncogenesis and

angiogenesis, and is known to be able to stimulate the adhesion of tumor cells to platelets[34]. In the downstream analysis, we identified processes such as GPCR ligand binding and hemostasis which contributes to the thrombosis process and therefore is also linked to GPCRs[35] (Supplementary Fig. 2a and Supplementary Data 2). It has been reported that hemostatic elements such as platelets, coagulation and the fibrinolytic system might play an important role in breast cancer progression and metastasis[36].

ER-negative: circadian clock

Another module identified by our network analysis consists of four genes with a strong link to the circadian clock mechanism: *PER1*, *PER3*, *TIMELESS*, and *TIPIN* (PHS P = 0.030) (Fig. 3b). Having an important role in the regulation of the cell cycle[37], the circadian clock is believed to be important in the development of cancer. Disrupted sleep patterns and associated changes to the body's circadian rhythm have long been implicated in the risk of developing several cancers including breast cancer[37–39]. Although long-term night-shift work has not consistently been found to be associated with breast cancer[40], one study reported an increased risk of ER-negative breast cancer[41]. More recently, genetic variants in circadian clock genes have been reported to be associated with breast cancer risk[42,43]. In addition to risk, the circadian clock has also been suggested to be involved in breast cancer progression and prognosis[44,45].

More specifically, the circadian clock genes in this module have also individually been implicated in the biology of cancer in general and breast cancer in particular. The period genes *PER1* and *PER3* have been found to suppress cancer cell growth[46,47] and have also been observed to be deregulated in breast cancer[48]. *TIMELESS* and its interactor *TIPIN* are believed to be central players in the connection between the circadian clock and the cell cycle and apoptosis[49,50]. The importance of these genes in the regulation of cell cycle was supported by the downstream analysis, which pointed out that cell cycle-related processes are strongly enriched among the downstream transcriptional changes.

ER-negative: regulators of cell growth and angiogenesis

The last high-confidence GRPM identified for ER-negative breast cancer contains proteins that have been linked to regulation of cell growth or angiogenesis: *CHCHD4*, *PDE9A*, *SLC36A1*, and *PHYHIPL* (PHS P = 0.027) **(Fig. 3c)**. Knock down of *CHCHD4* has been found to reduce tumor growth and angiogenesis in vivo[51]. In addition, *CHCHD4* has been observed to mediate the mitochondrial translocation of p53[52] through which it may trigger apoptosis via the p53 mitochondrial pathway[53]. *PDE9A* is a regulator of cGMP signaling, a pathway that is increasingly being recognized as an important player in breast cancer biology[54]. Inhibition of *PDE9A* has been found to trigger apoptosis in both ER-positive and ER-negative breast cancer cell lines[55]. *SLC36A1*, also known as *PAT1*, has been linked to tumor cell growth through its involvement in the activation of mTORC1. *PHYHIPL* (or *PAHX-AP1*) has mostly been described in the context of neuronal cells, but no role in cancer has been described.

ER-positive: Rho GTPases in apoptosis and cell growth

For ER-positive tumors, we identified one high-confidence module (PHS P = 0.020) **(Fig. 3d)**. The module was predicted to be involved in Rho GTPases effectors, which typically function as binary switches controlling a variety of biological processes. Because of their ability to control cell motility they have been hypothesized to play a role in progression and metastatic dissemination of cancer cells[56]. This GRPM contains seven genes: *ARHGAP10*, *CCNT2*, *CDR2*, *HEXIM1*, *NEUROD2*, *PKN1* and *ZFAND6*. *ARHGAP10* (rho GTPase Activating Protein 10) was previously reported as the most significant locus ($P = 2.3 \times 10^{-7}$) in a GWAS of breast cancer survival[14]. The top scoring gene in the module, *PKN1* (protein-kinase-C-related kinase), controls processes such as regulation of the intermediate filaments of the actin cytoskeleton, tumor cell invasion and cell migration[57]. It is activated by the Rho family of small G-proteins and might mediate the Rho-dependent signaling pathway[58], which was one of the main enriched pathways in the module-level analysis. *PKN1* has also been described as an important player in other cancers: in androgen-associated prostate cancer by controlling migration and metastasis[57], or in melanomas by inhibiting Wnt/b-catenin signaling and apoptosis[58].

From the module-level analysis, another enriched main process was the pathway linked to *PTEN* (phosphatase and tensin homologue deleted on chromosome 10) regulation, which is a well characterized tumor suppressor[59]. *PTEN* is directly involved in the metabolism of phospholipids and lipoproteins[60], adaptive immune system and B-cell receptor associated events,[61] which were all hits in the downstream analysis. One of the six genes in the module, *HEXIM1* (hexamethylene bisacetamide-inducible protein 1), is a positive regulator of p53 and has been identified as a potential novel therapeutic target modulating cell death in breast cancer cells[62]. In the downstream analysis of this module we also identified processes present in the module-level analysis that highlighted key tumorigenic biological processes (Supplementary Data 6), for instance pathways related to p53 activity, WNT signaling, regulation of mRNA stability by proteins that bind AU-rich elements or apoptotic execution phase.

DISCUSSION

There is evidence that breast cancer prognosis has a heritable component[2,63,64]. Exploring the possible link between germline genetic variants and breast cancer survival may help to develop better criteria for breast cancer stratification, which might have implications for breast cancer prognostication and treatment[65]. However, identifying germline genetic variants associated with breast cancer prognosis has been challenging so far, mainly because the current sample sizes have been insufficient to detect small effect signals.

In this work, we started with a survival analysis based on individual germline variants similar to the previous GWAS we have undertaken[4]. While in the previous analyses no variants reached genome-wide significance, here, we identified two genome-wide significant variants for ER-positive tumors (rs6990375: $P < 6.35 \times 10^{-9}$ and rs13272847: $P = 1.07 \times 10^{-8}$) located in 8q13. More complete follow-up and more conservative variant filtering per dataset (only including variants with imputation $r^2 > 0.8$) may have enabled identification of these variants that remained below genome-wide significance in our previous study ($P = 3.02 \times 10^{-5}$ and $P = 1.73 \times 10^{-5}$, respectively). In the gene-level analysis, we found two significant genes (*SLCO5A1* and

555 *SULF1*, P < 0.05 after Bonferroni correction) associated with breast cancer survival. It is likely that
556 both associations were driven by the identified leading variant rs6990375.
557
558 To address the lack of power in the individual germline variant and gene-level analyses, we
559 developed a network analysis method that revealed five high-confidence GRPMs associated with
560 breast cancer prognosis. We identified four modules specific for ER-negative breast cancer and
561 one for ER-positive breast cancer. The GRPMs comprise crucial processes such as cell cycle and
562 progression, regulation of apoptosis, signaling by mTOR, immune system, G-alpha signaling, and
563 the circadian clock. These processes are already known to play a role in cancer biology in general
564 and breast cancer prognosis specifically. However, our results highlight the possible regulatory
565 impact of germline variants on these processes, which traditionally has received little attention in
566 cancer survival studies. The broad range of genes and functions seems to indicate, as already
567 hypothesized, that breast cancer survival is a complex phenotype influenced by many factors and
568 biological mechanisms.
569
570 The analysis by ER-status subtypes identified significant associations that were not present when
571 analyzing all patients together. This is in line with the breast cancer risk analyses undertaken in
572 this same dataset, where the ER-subtype analyses also identified new associations[3]. Additionally,
573 the main classification of breast cancer tumors used for prognosis and treatment selection is
574 based on immunohistochemical markers such as ER-, PR- and HER2-status, reflecting the fact
575 that each group has a different etiology and prognosis. This assumption is further supported by a
576 comparison of the gene association scores between the ER-status subtypes. The gene scores for
577 ER-positive and ER-negative breast cancer are uncorrelated **(Supplementary Fig. 4c)** (Pearson
578 correlation = -0.002), while the gene scores for all breast cancer cases seem to resemble the ER-
579 positive subtype more **(Supplementary Fig. 4a)** (Pearson correlation = 0.366) than the ER-negative
580 subtype **(Supplementary Fig. 4b)** (Pearson correlation = 0.197). In addition, we found that the
581 distribution of PHSs across patients was similar for ER-positive and ER-negative breast cancer
582 patients **(Supplementary Fig. 5)**, but importantly, each PHS was associated with prognosis only for

583 the subtype in which it was found (Supplementary Table 4). These differential associations across
584 subtypes suggest that prognosis is inherited differently for these two different disease classes.
585
586 The network-based approach and the stratification of patients by ER-status enabled a refined
587 interpretation of the GWAS results[5,66], but the findings are still limited due to the number of
588 deaths observed, limited follow-up, missing treatment information, and possibly remaining
589 heterogeneity of tumor subtype within the ER classes. Increased sensitivity and specificity of the
590 results could be achieved by including additional patients, and by adjusting for more fine-grained
591 tumor characteristics and the treatment received. Moreover, the network propagation results
592 are dependent on the completeness of the PPI network used. As a notable consequence of this,
593 we did not identify modules containing the two gene-level significant hits *SLCO5A1* and *SULF1*,
594 due to the fact that the PPI network did not contain the proteins they code for.
595
596 The modules that are identified also depend on the specificity of the PPI network to the disease-
597 relevant tissue. Many proteins have tissue-specific expression patterns and functions; hence not
598 all interactions in a generic PPI network are found in all tissues. The use of a tissue-specific PPI
599 network may prevent discovery of false positive modules. One single most relevant tissue for our
600 analysis is not easily identified though. Unlike the somatic mutations found in tumor cells, the
601 germline variants we studied are present in every cell of the body. Their effect on survival may
602 therefore be mediated by cell types or tissues other than the cancerous breast tissue. These
603 include the various cell types present in the tumor microenvironment, or distant tissues that
604 form the pre-metastatic niche. Furthermore, a PPI network specific for healthy breast tissue may
605 not accurately describe the interactions active in transformed cancer cells. In our analysis, we
606 used a generic PPI network. To prevent false positive modules, we complemented the network
607 propagation with an extra filtering step in which we select high-confidence modules based on
608 their association with survival.
609
610 Using curated protein interaction networks such as iRefIndex in propagation analyses may cause
611 a subtle type of ascertainment bias: more interactions tend to be known for better studied

proteins, which proteins involved in tumor initiation and progression often are. As a result, gene scores may correlate positively with the number of interactions in the protein interaction network. This is the case, for example, when gene scores are based on somatic mutation frequencies in cancer. HotNet2 only controls for this partially, whereas a recent extension to the HotNet2 method provides a more rigorous solution[67]. We tested whether our analysis was vulnerable to this ascertainment bias by calculating the correlation between the gene scores computed by Pascal and the number of interactions recorded by iRefIndex. For all, ER-positive, and ER-negative breast cancer, these correlations were close to zero (Pearson r2 = -0.012, r2 = -0.006, and r2 = 0.003 respectively) showing no evidence of ascertainment bias due to proteins' numbers of recorded interactions.

In summary, our network propagation analysis shows a germline genetic link to breast cancer survival and proposes a mechanism by which multiple loci with small individual effects might influence breast cancer-specific prognosis. Experimental follow-up of the high-confidence GRPMs identified is required to better understand the role of these modules. While we focused on the subset of high-confidence modules, the other modules may also yield new insights if assessed in the context of larger independent datasets. Together the results presented here may feed future hypotheses about the contribution of germline variation to breast cancer survival.

Methods

Breast cancer patient data. We used data from 12 genome-wide association studies (GWAS) that together account for 84,457 invasive breast cancer patients with 5,413 breast cancer-specific deaths within 10 years (events). These included 55,701 patients with ER-positive breast cancer (2,854 events) and 14,529 patients with ER-negative breast cancer (1,724 events), while the ER-status was unknown for the remaining 14,227 patients. All patients were females of European ancestry. A summary of the studies with the numbers of patients and events by study is given in **(Supplementary Table 1)**. The GWAS sample sets were genotyped using a variety of genotyping arrays, targeting between 200,000 and 900,000 variants across the genome, and subsequently imputed using a common reference (details given below). The majority of patients came from

641 the Breast Cancer Association Consortium (BCAC), which itself comprised 69 studies from across
642 the world that underwent a uniform data harmonization and quality control (data freeze 10).
643 Genotyping in BCAC was performed in two rounds using two different genotyping platforms:
644 iCOGS and OncoArray. In subsequent analyses, we treated these two platforms as different
645 studies. The OncoArray dataset is the largest in BCAC, with higher quality imputed genotypes
646 compared to the iCOGS data. As an independent dataset, we separated out the entire SEARCH
647 study, comprising 12,381 patients and 1,120 events, from the BCAC data. Patients in the SEARCH
648 study were recruited in the United Kingdom. Their genotypes were obtained using either iCOGS
649 or OncoArray (Supplementary Table 2). Participants of all the studies provided written informed
650 consent and studies were approved by local medical ethical committees.
651
652 **Genotype data and sample quality control.** Quality checks were performed by the original
653 studies[3,5,68]. Genotypes for all 12 datasets were imputed using a reference panel from the 1000
654 Genomes Project[69] March 2012 release. Imputation was performed by a two-stage procedure[3]
655 using SHAPEIT[70] for pre-phasing and IMPUTE2[71] for genotype imputation. The genome-wide
656 analyses were performed on ~7.3 million variants that had a minor allele frequency (MAF) > 0.05
657 and were imputed with imputation quality $r^2 > 0.8$ in at least one of the studies.
658
659 **GWAS survival analysis and summary statistics.** The survival analysis was performed for all invasive
660 breast cancer cases combined and for each of the ER-status subtypes (ER-positive and ER-
661 negative) individually. A Cox proportional hazards model was fitted to assess the association of
662 the genotype with breast cancer-specific survival. Time-to-event was calculated from the date of
663 diagnosis. Yet, because patients were recruited at different times before or after diagnosis, time
664 at risk was calculated from the recruitment date (left truncation) in order to avoid possible bias
665 produced by prevalent cases. Follow-up was right censored on the date of death if the patient
666 died from a cause other than breast cancer, the last date the patient was known to be alive if
667 death did not occur, or at 10 years after diagnosis, whatever came first. To control for cryptic
668 population substructure, we adjusted for principal components[3] (for the number of principal
669 components per study see Supplementary Table 1). Since BCAC-OncoArray and BCAC-iCOGS

comprised data from large international cohort studies, the Cox models for these datasets were stratified by country. Separate survival analyses were performed for each of the 12 main studies, after which overall results per variant were obtained by combining the results of all studies with imputation quality $r^2 > 0.8$ for that variant using a fixed-effects meta-analysis. P values were computed using a two-sided Wald test.

From variant P values to gene scores. We used the GWAS summary statistics from the survival analysis as input for computing gene scores. To obtain gene scores, we used the Pascal algorithm[12] which combines variant P values while taking into account dependence due to linkage disequilibrium (LD) structure. The Pascal method implements two gene-level statistics, corresponding to the strongest single association per gene (maximum of chi-squared statistics), or the average of all associations across the gene (sum of chi-squared statistics). After computing both statistics we tested which one had more power. To this end, we represented the set of P values into a quantile-quantile (QQ)-plot (Supplementary Fig. 6). For all breast cancer cases and for both ER-status groups, the QQ-plots suggested that the maximum statistic has more power than the sum statistic. Therefore, of the two gene statistics we chose the maximum of chi-squared statistics for the gene-level statistic.

For the LD-reference population used in the gene computation, we created an extended version that included more variants than the default library provided with Pascal. This reference population was based on 503 European genomes from the 1000 Genomes Project (1KG)[69]. For the remaining parameters, we used the default settings. First, only variants with an imputation quality $r^2 > 0.8$ and MAF > 5% in the patient data were considered. Second, the mapping of the variants to genes was based on the Pascal's default 50-kb window size from the start and end of the gene. Finally, when computing gene scores, HLA genes were excluded. After the gene score computation, we obtained 21,815 gene scores for all invasive breast cancer, 21,789 for ER-positive and 21,797 for ER-negative. The slightly different numbers of gene scores between groups are due to the distinct selection of variants, which may have different allele frequencies

across groups. The gene scores used in the HotNet2 analysis were obtained by taking the -$\log_{10}$ of the gene P values computed with Pascal.

Network propagation with HotNet2. We performed a network propagation analysis using the HotNet2 algorithm[10] and the protein-protein interaction network iRefIndex[25] applied to the -$\log_{10}$ gene scores obtained from the previous step. For edge removal on the created modules, HotNet2 automatically selects four different values which determine four different edge removal thresholds. The significance test is a two-stage statistical test based on the number and size of the identified modules compared to those found using a permutation test. We used 500 permutations and a minimum network size of two for statistical testing. Further details are provided in the original HotNet publication[72,73].

Construction of polygenic hazard scores. To summarize the total prognostic effect of the hereditary variants within the significant germline-regulated prognostic modules (GRPMs), we constructed polygenic hazard scores (PHS), using a two-step approach. First, we selected the set of variants that best represented the genetic association of breast cancer survival with each GRPM. This variant selection was performed on the BCAC-OncoArray data, since this was the largest study and had the highest imputation quality. We performed the selection using the *glmnet* R package[74], fitting a Lasso (alpha = 1) model with 10-fold cross-validation to tune the sparsity penalty and the same selection of input variants as used for the computation of the Pascal gene scores, that is, picking those variants with MAF > 5% and within a 50-kb window around the start and end of the gene. With the set of germline variants selected using the Lasso procedure **(Supplementary Table 3)**, we fitted a Cox model to estimate unpenalized coefficients, and extracted their effect size estimates to compute a PHS per GRPM, which characterized the whole set of variants for the specific module in a unique score. For a set of selected variants $\{1, \dots, n\}$, the PHS is defined as in (1):

$$\mathrm{PHS} = \sum_{i=1}^{n} X_i \beta_i \quad (1)$$

where X_i is the genotype for the i^{th} variant and β_i its associated coefficient.

Identification of high-confidence GRPMs. We obtained a selection of high-confidence GRPMs from among all modules identified using HotNet2 by testing the association of each module's PHS in two datasets. The first dataset was the BCAC-OncoArray data minus the SEARCH data component of BCAC, i.e. the same data on which the PHS was derived, which was also a subset of the data used in the HotNet2 analysis. The second dataset consisted of the SEARCH study, which was held out of the BCAC data to serve as a truly independent set. Only GRPMs that had a PHS significantly associated ($P < 0.05$) with breast cancer-specific survival in both the BCAC-OncoArray and the independent SEARCH data were considered high-confidence GRPMs and kept for further analysis. To test the association of a PHS with prognosis, we fitted a Cox model to the PHS, adjusted for the first two genetic principal components and stratified by country. We then calculated a one-sided P value for the association of the PHS covariate with survival, taking advantage of the fact that the direction of association of the PHS is predefined, i.e. lower PHS means better survival. For the BCAC OncoArray data, the P value was corrected for multiple testing using Bonferroni correction based on the number of modules tested. The independent SEARCH data comprised two subsets using either OncoArray or iCOGS data. We analyzed these two subsets separately, and then combined the results of both groups using a fixed-effect meta-analysis.

Functional enrichment analysis of GRPM members. Using Cytoscape version 3.4.0 software[75] we extended the GRPMs by adding the first direct neighboring genes in the Mentha[76] human protein-protein interaction network. With the extension of the GRPMs we obtained bigger modules placed in a functional context. We then used the Cytoscape app ClueGO[77]. ClueGO uses kappa statistics to group the elements of the network and creates organized pathway categories based on the integrated pathway annotation. We based the analysis on human Reactome[28] pathways, a Kappa Score Threshold of 0.4, and Bonferroni correction for the computed enrichment P values. For the visualization, we selected the fusion feature that groups pathways according to overlapping genes to facilitate interpretation of the results. We selected pathways with a P value < 0.05.

Downstream functional enrichment. In order to add biological and functional interpretation to the GRPMs we looked for associations between the modules' PHSs and the expression patterns of potential downstream genes **(Fig. 1e)**. From The Cancer Genome Atlas (TCGA)[26] library we extracted matched RNA-seq and genotype data of female breast cancer patients of European ancestry. This resulted in 118 patients with ER-negative breast cancer and 440 patients with ER-positive breast cancer. For each GRPM, we computed the previously obtained PHS for the subset of TCGA patients with a tumor matching the subtype for which the GRPM was found. Next, we aimed to quantify the downstream transcriptional effect of the GRPM on the expression of every individual gene. To do so, we computed the Pearson correlation between the GRPM's PHS and the RNA expression of each gene. Finally, we performed gene set enrichment analysis (GSEA)[27] to test for enrichment of biological pathways among the highly correlating genes. We used an annotation set of Reactome pathways[28] and MSigDB[29] Hallmark gene sets to perform the pre-ranked GSEA. We visualized the Reactome results with the EnrichmentMap[78] Cytoscape app. Only biological processes with P value < 0.001 and FDR < 0.05 were considered as significantly enriched.

Data availability

All 10-year breast cancer-specific survival summary estimates are available via the BCAC website (http://bcac.ccge.medschl.cam.ac.uk/bcacdata/). Individual patient data will not be made publicly available without request due to restraints imposed by the ethics committees of individual studies. Formal request can be made via the Data Access Coordination Committee (DACC) of BCAC (http://bcac.ccge.medschl.cam.ac.uk/). A subset of the data that supports the findings of this analysis is available at https://portal.gdc.cancer.gov/ (accession number phs000178).

References

1. Lindström, L. S. *et al.* Prognostic information of a previously diagnosed sister is an independent prognosticator for a newly diagnosed sister with breast cancer. *Ann. Oncol.* **25**, 1966–1972 (2014).
2. Verkooijen, H. M. *et al.* Breast cancer prognosis is inherited independently of patient, tumor and treatment characteristics. *Int. J. Cancer* **130**, 2103–2110 (2012).

3. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
4. Escala-Garcia, M. *et al.* Genome-wide association study of germline variants and breast cancer-specific mortality. *Br. J. Cancer* **120**, 647–657 (2019).
5. Guo, Q. *et al.* Identification of Novel Genetic Markers of Breast Cancer Survival. *JNCI J. Natl. Cancer Inst.* **107**, djv081–djv081 (2015).
6. Wang, K., Li, M. & Hakonarson, H. Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* **11**, 843–854 (2010).
7. Menashe, I. *et al.* Pathway analysis of breast cancer genome-wide association study highlights three pathways and one canonical signaling cascade. *Cancer Res.* **70**, 4453–9 (2010).
8. Baranzini, S. E. *et al.* Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum. Mol. Genet.* **18**, 2078–90 (2009).
9. Luo, L. *et al.* Genome-wide gene and pathway analysis. *Eur. J. Hum. Genet.* **18**, 1045–1053 (2010).
10. Leiserson, M. D. M. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–14 (2015).
11. Petersen, A., Alvarez, C., DeClaire, S. & Tintle, N. L. Assessing methods for assigning SNPs to genes in gene-based tests of association using common variants. *PLoS One* **8**, e62161 (2013).
12. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLOS Comput. Biol.* **12**, e1004714 (2016).
13. Marbach, D. *et al.* Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. *Nat. Methods* **13**, 366–370 (2016).
14. Azzato, E. M. *et al.* A Genome-Wide Association Study of Prognosis in Breast Cancer. *Cancer Epidemiol. Biomarkers Prev.* **19**, 1140–1143 (2010).
15. Kiyotani, K. *et al.* A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Hum. Mol. Genet.* **21**, 1665–72 (2012).
16. Shu, X. O. *et al.* Novel genetic markers of breast cancer survival identified by a genome-wide association study. *Cancer Res.* **72**, 1182–9 (2012).
17. Rafiq, S. *et al.* Identification of inherited genetic variations influencing prognosis in early-onset breast cancer. *Cancer Res.* **73**, 1883–91 (2013).
18. Han, C. H. *et al.* Polymorphisms in the SULF1 gene are associated with early age of onset and survival of ovarian cancer. *J. Exp. Clin. Cancer Res.* **30**, 5 (2011).
19. Lai, J.-P., Sandhu, D. S., Shire, A. M. & Roberts, L. R. The tumor suppressor function of human sulfatase 1 (SULF1) in carcinogenesis. *J. Gastrointest. Cancer* **39**, 149–58 (2008).
20. Emerson, C. P. *et al.* QSulf1, a heparan sulfate 6-O-endosulfatase, inhibits fibroblast growth factor signaling in mesoderm induction and angiogenesis. *Proc. Natl. Acad. Sci.* **101**, 4833–4838 (2004).
21. Ai, X. *et al.* QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling. *J. Cell Biol.* **162**, 341–351 (2003).
22. Sebastian, K. *et al.* Characterization of SLCO5A1/OATP5A1, a solute carrier transport protein with non-classical function. *PLoS One* **8**, e83257 (2013).
23. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: A universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
24. Nakka, P., Raphael, B. J. & Ramachandran, S. Gene and network analysis of common variants reveals novel associations in multiple complex diseases. *Genetics* **204**, 783–798 (2016).
25. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405 (2008).
26. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours.

Nature **490**, 61–70 (2012).
27. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
28. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428-32 (2005).
29. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
30. Lappano, R. & Maggiolini, M. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat. Rev. Drug Discov.* **10**, 47–60 (2011).
31. Kisfalvi, K., Rey, O., Young, S. H., Sinnett-Smith, J. & Rozengurt, E. Insulin Potentiates Ca 2+ Signaling and Phosphatidylinositol 4,5-Bisphosphate Hydrolysis Induced by G q Protein-Coupled Receptor Agonists through an mTOR-Dependent Pathway. *Endocrinology* **148**, 3246–3257 (2007).
32. Sassmann, A. *et al.* The Gq/G11-mediated signaling pathway is critical for autocrine potentiation of insulin secretion in mice. *J. Clin. Invest.* **120**, 2184–93 (2010).
33. Imamura, T. *et al.* G alpha-q/11 protein plays a key role in insulin-induced glucose transport in 3T3-L1 adipocytes. *Mol. Cell. Biol.* **19**, 6765–74 (1999).
34. Nierodzik, M. L. & Karpatkin, S. Thrombin induces tumor growth, metastasis, and angiogenesis: Evidence for a thrombin-regulated dormant tumor phenotype. *Cancer Cell* **10**, 355–62 (2006).
35. Woulfe, D. S. Platelet G protein-coupled receptors in hemostasis and thrombosis. *J. Thromb. Haemost.* **3**, 2193–200 (2005).
36. Lal, I., Dittus, K. & Holmes, C. E. Platelets, coagulation and fibrinolysis in breast cancer progression. *Breast Cancer Res.* **15**, 207 (2013).
37. Gaucher, J., Montellier, E. & Sassone-Corsi, P. Molecular Cogs: Interplay between Circadian Clock and Cell Cycle. *Trends Cell Biol.* **28**, 368–379 (2018).
38. Hansen, J. & Stevens, R. G. Case-control study of shift-work and breast cancer risk in Danish nurses: impact of shift systems. *Eur. J. Cancer* **48**, 1722–9 (2012).
39. Knutsson, A. *et al.* Breast cancer among shift workers: results of the WOLF longitudinal cohort study. *Scand. J. Work. Environ. Health* **39**, 170–7 (2013).
40. Travis, R. C. *et al.* Night shift work and breast cancer incidence: Three prospective studies and meta-analysis of published studies. *J. Natl. Cancer Inst.* **108**, djw169 (2016).
41. Rabstein, S. *et al.* Night work and breast cancer estrogen receptor status--results from the German GENICA study. *Scand. J. Work. Environ. Health* **39**, 448–55 (2013).
42. Zienolddiny, S. *et al.* Analysis of polymorphisms in the circadian-related genes and breast cancer risk in Norwegian nurses working night shifts. *Breast Cancer Res.* **15**, R53 (2013).
43. Reszka, E., Przybek, M., Muurlink, O. & Pepłonska, B. Circadian gene variants and breast cancer. *Cancer Lett.* **390**, 137–145 (2017).
44. Cadenas, C. *et al.* Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell Cycle* **13**, 3282–91 (2014).
45. Ha, N.-H., Long, J., Cai, Q., Shu, X. O. & Hunter, K. W. The Circadian Rhythm Gene Arntl2 Is a Metastasis Susceptibility Gene for Estrogen Receptor-Negative Breast Cancer. *PLoS Genet.* **12**, e1006267 (2016).
46. Yang, X. *et al.* The circadian clock gene per1 suppresses cancer cell proliferation and tumor growth at specific times of day. *Chronobiol. Int.* **26**, 1323–1339 (2009).
47. Komatsu, N. *et al.* The Circadian Gene Per1 Plays an Important Role in Cell Growth and DNA Damage Control in Human Cancer Cells. *Mol. Cell* **22**, 375–382 (2006).
48. Kuo, S.-J. *et al.* Deregulated expression of the PER1 , PER2 and PER3 genes in breast cancers. *Carcinogenesis* **26**, 1241–1246 (2005).
49. Unsal-Kacmaz, K., Mullen, T. E., Kaufmann, W. K. & Sancar, A. Coupling of Human Circadian and

Cell Cycles by the Timeless Protein. *Mol. Cell. Biol.* **25**, 3109–3116 (2005).
50. Rigaill, G. *et al.* TIPIN depletion leads to apoptosis in breast cancer cells. *Mol. Oncol.* **9**, 1580–1598 (2015).
51. Ashcroft, M. *et al.* Human CHCHD4 mitochondrial proteins regulate cellular oxygen consumption rate and metabolism and provide a critical role in hypoxia signaling and tumor progression. *J. Clin. Invest.* **122**, 600–611 (2012).
52. Zhuang, J. *et al.* Mitochondrial disulfide relay mediates translocation of p53 and partitions its subcellular activity. *Proc. Natl. Acad. Sci.* **110**, 17356–17361 (2013).
53. Sun, Y. *et al.* Haploinsufficiency in the mitochondrial protein CHCHD4 reduces brain injury in a mouse model of neonatal hypoxia-ischemia. *Cell Death Dis.* **8**, e2781 (2017).
54. Windham, P. F. & Tinsley, H. N. CGMP signaling as a target for the prevention and treatment of breast cancer. *Semin. Cancer Biol.* **31**, 106–110 (2015).
55. Saravani, R., Karami-Tehrani, F., Hashemi, M., Aghaei, M. & Edalat, R. Inhibition of phosphodiestrase 9 induces cGMP accumulation and apoptosis in human breast cancer cell lines, MCF-7 and MDA-MB-468. *Cell Prolif.* **45**, 199–206 (2012).
56. Aznar, S. & Lacal, J. C. Rho signals to cell growth and apoptosis. *Cancer Lett.* **165**, 1–10 (2001).
57. Jilg, C. A. *et al.* PRK1/PKN1 controls migration and metastasis of androgen-independent prostate cancer cells. *Oncotarget* **5**, 12646–64 (2014).
58. James, R. G. *et al.* Protein Kinase PKN1 Represses Wnt/β-Catenin Signaling in Human Melanoma Cells. *J. Biol. Chem.* **288**, 34658–34670 (2013).
59. Dillon, L. & Miller, T. Therapeutic Targeting of Cancers with Loss of PTEN Function. *Curr. Drug Targets* **15**, 65–79 (2014).
60. Garcia-Cao, I. *et al.* Systemic elevation of PTEN induces a tumor-suppressive metabolic state. *Cell* **149**, 49–62 (2012).
61. Chen, L. & Guo, D. The functions of tumor suppressor PTEN in innate and adaptive immunity. *Cell. Mol. Immunol.* **14**, 581–589 (2017).
62. Neo, S. H. *et al.* Use of a novel cytotoxic HEXIM1 peptide in the directed breast cancer therapy. *Oncotarget* **7**, 5483–94 (2016).
63. Hartman, M. *et al.* Is breast cancer prognosis inherited? *Breast Cancer Res.* **9**, R39 (2007).
64. Möller, S. *et al.* The heritability of breast cancer among women in the nordic twin study of cancer. *Cancer Epidemiol. Biomarkers Prev.* **25**, 145–150 (2016).
65. Anderson, W. F., Rosenberg, P. S., Prat, A., Perou, C. M. & Sherman, M. E. How many etiological subtypes of breast cancer: Two, three, four, or more? *J. Natl. Cancer Inst.* **106**, 1–11 (2014).
66. Kao, P. Y. P., Leung, K. H., Chan, L. W. C., Yip, S. P. & Yap, M. K. H. Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. *Biochim. Biophys. Acta* **1861**, 335–353 (2017).
67. Reyna, M. A., Leiserson, M. D. M. & Raphael, B. J. Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* **34**, i972–i980 (2018).
68. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–61, 361e1-2 (2013).
69. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73 (2010).
70. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–81 (2011).
71. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda).* **1**, 457–70 (2011).
72. Vandin, F., Clay, P., Upfal, E. & Raphael, B. J. Discovery of mutated subnetworks associated with clinical data in cancer. *Pac. Symp. Biocomput.* 55–66 (2012).

73. Vandin, F., Upfal, E. & Raphael, B. J. Algorithms for detecting significantly mutated pathways in cancer. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 6044 LNBI, 506–521 (2010).
74. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1–22 (2010).
75. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–504 (2003).
76. Calderone, A., Castagnoli, L. & Cesareni, G. mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* 10, 690–691 (2013).
77. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–3 (2009).
78. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* 5, e13984 (2010).

Supplementary Information

Acknowledgements

BCAC: We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. We acknowledge all contributors to the COGS and OncoArray study design, chip design, genotyping, and genotype analyses. **ABCFS**: Maggie Angelakos, Judi Maskiell, Gillian Dite. **ABCS:** Frans Hogervorst, Sten Cornelissen and Annegien Broeks. **ABCTB** Investigators: Rosemary Balleine, Robert Baxter, Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken, Desmond Yip. **BBCS:** Eileen Williams, Elaine Ryder-Mills, Kara Sargus. **BCINIS**: Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet. **BIGGS**: Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones. **BREOGAN**: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza, José Antúnez, Máximo Fraga; Joaquín González-Carreró and the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS. **BSUCH**: Peter Bugert, Medical Faculty Mannheim. **CCGP**: Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, Konstantinos Pompodakis. **CGPS:** Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen and the Danish Cancer Biobank. **CNIO-BCS**: Guillermo Pita, Charo Alonso, Nuria Álvarez, Pilar Zamora, and Primitiva Menendez. **CPS-II:** Centers for Disease Control and Prevention National Program of Cancer Registries. The National Cancer Institute Surveillance Epidemiology and End Results program. **CTS:** Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, and Jessica Clague DeHart. Dennis Deapen, Rich Pinder, and Eunjung Lee, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur and David Nelson,

Hoda Anton-Culver, Argyrios Ziogas, and Hannah Park and Fred Schumacher. **DIETCOMPLYF**: charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. Participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). **ESTHER**: Hartwig Ziegler, Sonja Wolf, Volker Hermann, Christa Stegmaier, Katja Butterbach. **FHRISK:** NIHR for funding. **GC-HBOC:** Stefanie Engert, Heide Hellebrand, Sandra Kröber and LIFE. Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber. **GENICA**: Dr. Margarete Fischer-Bosch [HB, Wing-Yee Lo], German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ) [HB], Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2180 - 390900677 [HB], Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, [Yon-Dschun Ko, Christian Baisch], [Hans-Peter Fischer], Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany [UH], [Thomas Brüning, Beate Pesch, Sylvia Rabstein, Anne Lotz]; [Volker Harth]. **HABCS:** Michael Bremer. **HEBCS:** Rainer Fagerholm, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä. **HUBCS:** Shamil Gantsev. **KARMA** and **SASBAC:** Swedish Medical Research Counsel. **KBCP:** Eija Myöhänen, Helena Kemiläinen. **kConFab/AOCS**: Eveline Niedermayr, Family Cancer Clinics and the Clinical Follow Up Study (received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)). **LMBC:** Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden and Kathleen Corthouts. **MARIE:** Petra Seibold, Judith Heinz, Nadia Obi, Sabine Behrens, Ursula Eilber, Muhabbet Celik and Til Olchers. **MBCSG**: Paolo Radice, Jacopo Azzollini, Bernardo Bonanni, Bernard Peissel, Roberto Villa, Giulia Cagnoli, Irene Feroce, and Cogentech Cancer Genetic Test Laboratory. **NBCS**: Kristine K. Sahlberg (PhD), Lars Ottestad (MD), Rolf Kåresen (Prof. Em.) Dr. Ellen Schlichting (MD), Marit Muri Holmen (MD), Toril Sauer (MD), Vilde Haakensen (MD), Olav Engebråten (MD), Bjørn Naume (MD), Alexander Fosså (MD), Cecile E. Kiserud (MD), Kristin V. Reinertsen (MD), Åslaug Helland (MD), Margit Riis (MD), Jürgen Geisler (MD) and OSBREAC. **NHS/NHS2:** AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. **OBCS:** Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo and Kari Mononen. **OFBCR:** Teresa Selander, Nayana Weerasooriya. **ORIGO:** E. Krol-Warmerdam, and J. Blom. **PBCS:** Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The ethical approval for the **POSH** study is MREC /00/6/69, UKCRN ID: 1137. Experimental Cancer Medicine Centre (ECMC) supported Faculty of Medicine Tissue Bank and the Faculty of Medicine DNA Banking resource. **PREFACE:** Sonja Oeser and Silke Landrith. **PROCAS**: NIHR for funding. **RBCS:** Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huijskens, Anja Kromwijk-Nieuwlaat, Annette Heemskerk, the Erasmus MC Family Cancer Clinic. **SBCS:** Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. We thank the **SEARCH** and **EPIC** teams. **SKKDKFZS: SUCCESS** Study teams in Munich, Duessldorf, Erlangen and Ulm. **SZBCS:**

Ewa Putresza. **UCIBCS:** Irene Masunaka. **UKBGS:** Breast Cancer Now and the Institute of Cancer Research and NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. **WHI:** investigators and staff for their dedication.

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

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

ABCFS was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the in the Breast Cancer Family Registry (**BCFR)**, nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The **ABCFS** was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow. The **ABCS** study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009-4363; 2015-7632]. The **ABCTB** was supported by the National Health and Medical

Research Council of Australia, The Cancer Institute NSW and the National Breast Cancer Foundation. The work of the **BBCC** was partly funded by ELAN-Fond of the University Hospital of Erlangen. The **BBCS** is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the NIHR Biomedical Research Centre, and the National Cancer Research Network (NCRN). For the **BCFR-NY, BCFR-PA, BCFR-UT** this work was supported by grant UM1 CA164920 from the National Cancer Institute. For **BIGGS**, ES is supported by NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, United Kingdom. IT is supported by the Oxford Biomedical Research Centre. The **BREOGAN** is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI17/00918/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain. The **BSUCH** study was supported by the Dietmar-Hopp Foundation, the Helmholtz Society and the German Cancer Research Center (DKFZ). **CCGP** is supported by funding from the University of Crete. The **CECILE** study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). The **CGPS** was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital. The **CNIO-BCS** was supported by the Instituto de Salud Carlos III, the Red Temática de Investigación Cooperativa en Cáncer and grants from the Asociación Española Contra el Cáncer and the Fondo de Investigación Sanitario (PI11/00923 and PI12/00070). The American Cancer Society funds the creation, maintenance, and updating of the **CPS-II** cohort. The **CTS** was initially supported by the California Breast Cancer Act of 1993 and the California Breast Cancer Research Fund (contract 97-10500) and is currently funded through the National Institutes of Health (R01 CA77398, UM1 CA164917, and U01 CA199277). Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885. The University of Westminster curates the **DietCompLyf** database funded by Against Breast Cancer Registered Charity No. 1121258 and the NCRN. The coordination of **EPIC** is financially supported by the European Commission (DG-SANCO) and the International

Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). The **ESTHER** study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). **FHRISK** is funded from NIHR grant PGfAR 0707-10031. Prof D Gareth Evans is supported by the NIHR Manchester Biomedical Research Centre (IS-BRC-1215-20007). The **GC-HBOC** is supported by the German Cancer Aid (grant no 110837, coordinator: Rita K. Schmutzler, Cologne). This work was also funded by the European Regional Development Fund and Free State of Saxony, Germany (LIFE - Leipzig Research Centre for Civilization Diseases, project numbers 713-241202, 713-241202, 14505/2470, 14575/2470). The **GENICA** was funded by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114, the Robert Bosch Foundation, Stuttgart, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, the Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, as well as the Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany. The **GESBC** was supported by the Deutsche Krebshilfe e. V. [70492] and the German Cancer Research Center (DKFZ). The **HABCS** study was supported by the Claudia von Schilling Foundation for Breast Cancer Research, by the Lower Saxonian Cancer Society, and by the Rudolf Bartling Foundation. The **HEBCS** was financially supported by the Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, and the Sigrid Juselius Foundation. The **HUBCS** was supported by a grant from the German Federal Ministry of Research and Education (RUS08/017), and by the Russian Foundation for Basic Research and the Federal Agency for Scientific Organizations for support the Bioresource collections and RFBR grants 14-04-97088, 17-29-06014 and 17-44-020498. Financial support for **KARBAC** was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County

Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation. The **KARMA** study was supported by Märit and Hans Rausings Initiative Against Breast Cancer. The **KBCP** was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland. **kConFab** is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. **LMBC** is supported by the 'Stichting tegen Kanker'. The **MARIE** study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402]. **MBCSG** is supported by grants from the Italian Association for Cancer Research (AIRC) and by funds from the Italian citizens who allocated the 5/1000 share of their tax payment in support of the Fondazione IRCCS Istituto Nazionale Tumori, according to Italian laws (INT-Institutional strategic projects “5x1000”). The **MCBCS** was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation. **MCCS** cohort recruitment was funded by VicHealth and Cancer Council Victoria. The **MCCS** was further supported by Australian NHMRC grants 209057 and 396414, and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. The **MEC** was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973. The **MISS** study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson. The **MMHS** study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. The **NBCS** has received funding from the K.G. Jebsen Centre for Breast Cancer Research; the Research Council of Norway grant 193387/V50 (to A-L Børresen-Dale and V.N. Kristensen) and grant 193387/H10 (to A-L Børresen-Dale and V.N. Kristensen), South Eastern Norway Health Authority (grant 39346 to A-L Børresen-Dale) and the Norwegian Cancer Society (to A-L Børresen-Dale and V.N. Kristensen). The **NC-BCFR** and **OFBCR** were supported by grant UM1 CA164920 from the National Cancer Institute (USA). The **NCBCS** was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund. The **NHS** was supported by NIH grants P01 CA87969, UM1 CA186107, and U19 CA148065. The **NHS2** was supported by NIH grants UM1 CA176726 and U19 CA148065. The **OBCS** was supported

by research grants from the Finnish Cancer Foundation, the Academy of Finland (grant number 250083, 122715 and Center of Excellence grant number 251314), the Finnish Cancer Foundation, the Sigrid Juselius Foundation, the University of Oulu, the University of Oulu Support Foundation and the special Governmental EVO funds for Oulu University Hospital-based research activities. The **ORIGO** study was supported by the Dutch Cancer Society (RUL 1997-1505) and the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL CP16). The **PBCS** was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. Genotyping for **PLCO** was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The **PLCO** is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health. The **POSH** study is funded by Cancer Research UK (grants C1275/A11699, C1275/C22524, C1275/A19187, C1275/A15956 and Breast Cancer Campaign 2010PR62, 2013PR044. PROCAS is funded from NIHR grant PGfAR 0707-10031. **PROCAS** is funded from NIHR grant PGfAR 0707-10031. The **RBCS** was funded by the Dutch Cancer Society (DDHK 2004-3124, DDHK 2009-4318). The **SASBAC** study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation. The **SBCS** was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank. **SEARCH** is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve. **SKKDKFZS** is supported by the DKFZ. The **SMC** is funded by the Swedish Cancer Foundation and the Swedish Research Council (SIMPLER, VR 2017-00644). The **SZBCS** was supported by Grant PBZ_KBN_122/P05/2004. The **UCIBCS** component of this research was supported by the NIH [CA58860, CA92044] and the Lon V Smith Foundation [LVS39420]. The **UKBGS** is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The **USRT** Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA. The **WHI** program is funded by the National Heart, Lung, and Blood Institute, the US National Institutes of Health and the US Department of Health and Human Services (HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C). This work was also funded by NCI U19 CA148065-01.

Author contributions: M.K.S. and S.C. conceived the study. M.E.G. performed the data analyses. M.K.S., S.C. and M.E.G. were involved in the interpretation of the data. S.C. provided statistical and computational support for the data analyses. R.K., Q.W., M.K.B. and J.D. provided database support. M.E.G., M.K.S. and S.C wrote the manuscript. All authors contributed data from their own studies, helped revise the manuscript and approved the final version.

Ethics approval: The study was performed in accordance with the Declaration of Helsinki. All individual studies, from which data was used, were approved by the appropriate medical ethical committees and/or institutional review boards. All study participants provided informed consent.

Consent for publication: All authors consented to this publication.

Competing interests: The authors declare no competing interests.

Correspondence: mk.schmidt@nki.nl and s.canisius@nki.nl

Figure Captions:

Figure 1. Network analysis pipeline (see Methods for details). **(a)** Cox models were used to estimate the association between each genetic variant and breast cancer-specific survival in 84,457 patients of the Breast Cancer Association Consortium (BCAC) dataset (discovery set). **(b)** The P values of the survival analyses for the genetic variants (blue squares) were used to compute gene scores using the Pascal algorithm. These gene scores were based on the maximum chi-squared signal within a window size of 50-kb around the gene region and accounted for linkage disequilibrium structure (depicted in a gradient blue scale). **(c)** The HotNet2 method was used to identify gene modules based on the $-\log_{10}$ P value of the computed gene scores. **(d)** The modules found by Hotnet2 were filtered to obtain a selection of high-confidence Germline-Related Prognostic Modules (GRPMs). We constructed a Polygenic Hazard Score (PHS) summarizing the prognostic effects of a set of selected genetic variants in the module. We then tested the association of this PHS with survival in both the discovery set (grey) and the independent set (orange). **(e)** We performed a functional characterization of the high-confidence GRPMs by studying the downstream transcriptional effects. For that, we used genotype and expression data from The Cancer Genome Atlas (TCGA). We computed the correlation between a GRPM's polygenic hazard score and the expression of all available genes. Based on these correlation values, a Gene Set Enrichment analysis assigned biological

processes that were enriched among the genes most correlated with the prognostic variants in the GRPM.

Figure 2. Manhattan plots of the gene-level associations with breast cancer-specific survival. Plots show the association in **(a)** all breast cancer cases (n=84,457) **(b)** Estrogen Receptor (ER)-negative (n=14,529) and **(c)** ER-positive (n=55,701). The $-\log_{10}$ gene P values from the Pascal algorithm is shown on the y axis and genomic position on the x axis. The top significant genes and the most significant gene per chromosome (if $-\log_{10}(P) > 3$) are shown in red.

Figure 3. High-confidence Germline-Related Prognostic Modules (GRPMs). The GRPM is shown at the center of the circles, surrounded by the biological processes enriched among the downstream transcriptional effects of each module. Three modules were found for Estrogen Receptor (ER)-negative breast cancer (a-c) and one module was found for ER-positive breast cancer (d). **(a)** G-alpha signaling GRPMs. **(b)** Circadian clock GRPM. **(c)** Regulators of cell growth and angiogenesis GRPM. **(d)** Rho GTPases and apoptosis GRPM. **(e)** Plots illustrating the association between each GRPM's PHS and 10-year breast cancer specific-survival in the discovery and independent sets. HR: Hazard Ratio (per standard deviation of the PHS), CI: Confidence Interval. The error bars show the 95% confidence interval. The confidence intervals shown are two-sided, whereas the significance test performed was one-sided (see Methods).

Figure 4. Genomic region 19p13.3 with the two genes *GNA11* and *GNA15*. The two G-alpha signaling high-confidence Germline-Related Prognostic Modules (GRPMs) identified in the Estrogen Receptor (ER)-negative subtype have a shared genetic signal in the same genomic region. **(a)** Top: -log10(P) for the association with survival (y axis) of all variants in the region 19p13.3 (y axis). Bottom: regression coefficients from the survival model for the genetic variants in the module's Polygenic Hazard Scores (PHSs)**. (b)** Scatter plot comparing the two modules' PHSs in the iCOGS independent validation set. PHS of the *GNA11* GRPM on the x axis and PHS of the *GNA15* GRPM on the y axis.

Supplementary Information

A network analysis to identify mediators of germline-driven differences in breast cancer prognosis

Escala-Garcia *et al.*

Supplementary Table 1. Summary of invasive breast cancer cases, events and follow-up by genotyping study and ER-status. Details about the 12 studies are described elsewhere[4].

Study	All cases		ER-positive		ER-negative		Number of principal components*
	N (breast cancer deaths)	Person-years	N (breast cancer deaths)	Person-years	N (breast cancer deaths)	Person-years	
BCAC-OncoArray *comprising 61 BCAC studies	49,843 (2,826)	280,653	3,546 (1,640)	194,729	7,826 (881)	43,008	2
BCAC-COGS *comprising 38 BCAC studies	22,708 (1,302)	121,945	15,519 (740)	85,707	3,731 (391)	19,796	9
CGEMS	1,145 (93)	7,711	--	--	--	--	0
SASBAC	787 (69)	3,739	483 (44)	2,294	108 (9)	502	0
UK2	2,763 (233)	23,112	--	--			3
Metabric	369 (86)	1,570	291 (59)	1,268	63 (25)	225	1
PG-SNPs	1,786 (204)	5,820	1,188 (116)	3,916	586 (87)	1,888	2
HEBCS	742 (285)	4,666	492 (172)	3,458	196 (101)	982	0
SUCCESS-A	3,312 (175)	13,145	2,265 (83)	9,289	1,017 (90)	3,806	0
BPC3-CPSII	293 (30)	2,544	--	--	293 (30)	2,544	0
BPC3-EPIC	476 (74)	2,226	--	--	476 (74)	2,226	0
BPC3-NHS2	233 (36)	2,732	--	--	233 (36)	2,732	0
Training set	**84,457 (5,413)**		**55,701 (2,854)**		**14,529 (1,724)**		

BCAC: Breast Cancer Association Consortium, ER: estrogen receptor

Supplementary Table 2. Summary of invasive breast cancer cases, events and follow-up by genotyping array and ER-status for the independent set. Details about the study are described elsewhere[4].

	All cases		ER-positive		ER-negative	
	N (breast cancer deaths)	**Person-years**	**N (breast cancer deaths)**	**Person-years**	**N (breast cancer deaths)**	**Person-years**
BCAC-OncoArray *SEARCH study	3,723 (110)		2,691 (55)		408 (26)	
BCAC-COGS *SEARCH study	7,539 (1,010)		5,128 (561)		1,058 (215)	
Independent set	**12,381 (1,120)**	60,025	**7,819 (616)**	36,859	**1,466 (241)**	7,088

BCAC: Breast Cancer Association Consortium, ER: estrogen receptor

Supplementary Table 3. Variants and their coefficients included in the computation of the Polygenic Hazard Score (PHS) for each Germline-Related Prognostic Module. The variant identifiers have the format "<Chromosome>_<Build19Position>_<RefAllele>_<AltAllele>". All alleles are reported on the forward strand.

GRPM	Variant	Coefficients
G-alpha signaling events (I)	19_3086486_A_G	-0.1268
	19_3089773_T_C	-0.1011
G-alpha signaling events (II)	19_3081157_T_C	-0.1430
	19_3084795_A_G	-0.0599
	19_3089773_T_C	-0.0558
Circadian clock	1_7860276_AT_ATT	0.0793
	1_7870048_T_C	0.0827
	1_7915742_CATT_C	0.0849
	1_7918598_A_C	0.0426
	1_7924023_C_T	0.1232
	1_7927086_C_T	-0.2161
	1_7946161_C_T	0.1664
	12_56849340_C_G	-0.2372
	12_56856618_C_T	0.1022
	15_66666223_T_C	-0.1605
	17_8005118_C_T	0.0091
	17_8007650_T_C	0.1323
	17_8016373_T_G	-0.1230
	17_8055999_C_A	0.1506
Regulation of cell growth and angiogenesis	21_44031933_A_G	0.1715
	21_44244882_A_G	0.1314
	3_14105089_A_G	0.0806

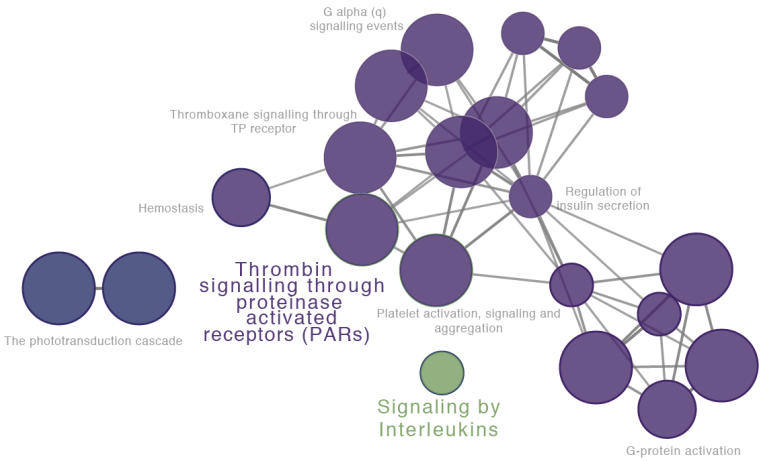
	3_14158438_C_G	-0.2045
	5_150837810_C_CAT	0.1310
Rho GTPases	15_80401077_GT_GTT	0.0926
	16_22346038_TG_T	-0.2491
	17_43185500_G_A	-0.0972
	17_43244700_A_C	0.0609
	17_43266487_G_A	-0.0965
	19_14570329_C_CA	0.1035
	2_135748039_T_G	0.1789
	4_148757466_A_C	-0.1578
	4_148946690_G_T	0.0551
	4_148949173_A_C	0.0639
	4_148970403_C_T	0.0596

Supplementary Table 4. P values obtained in the independent set for each high confidence Germline-Related Prognostic Module (GRPM)'s PHS: for the Estrogen Receptor (ER)-status group in which the GRPM was identified (in bold) versus the other ER-status group.

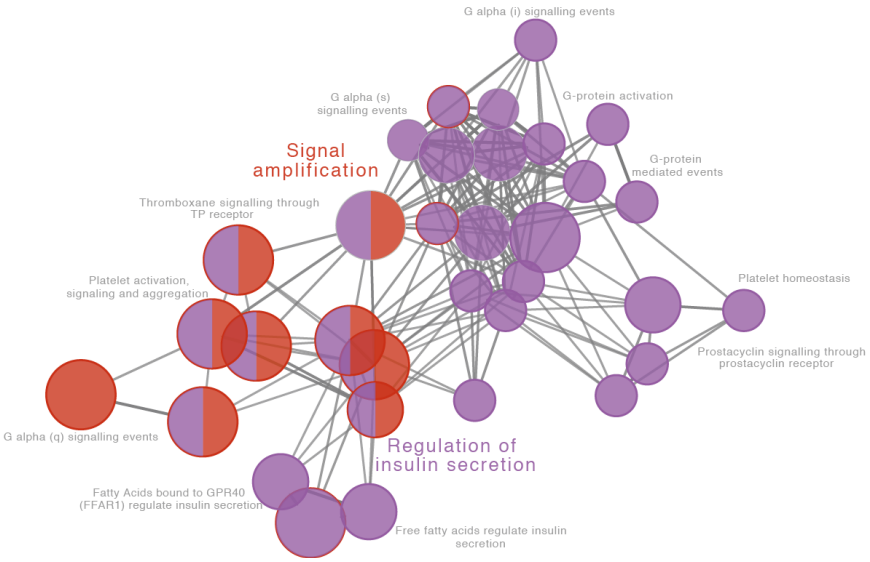
High-confidence GRPM	Independent set P value	
Identified in ER-negative tumors	ER-negative	ER-positive
G-alpha signaling events (I)	**0.008**	0.154
G-alpha signaling events (II)	**0.009**	0.171
Circadian clock	**0.030**	0.167
Regulation of cell growth and angiogenesis	**0.026**	0.145
Identified in ER-positive tumors		
Rho GTPases	0.763	**0.020**

Supplementary Figure 1. Module-level enrichment analyses for the Estrogen Receptor (ER)-negative (a-d) and ER-positive (e) high-confidence GRMPs. **(a)** G-alpha signaling (I). **(b)** G-alpha signaling (II). **(c)** Circadian clock. **(d)** Regulation of cell growth and angiogenesis. **(e)** Rho GTPases. Reactome annotations were used for the enrichment. The visualization was done using the Cytoscape app ClueGo. We selected pathways with a P value < 0.05 only. The enrichment of the nodes is represented within the node size and the functional groups are represented by the name of the most significant term in the group.

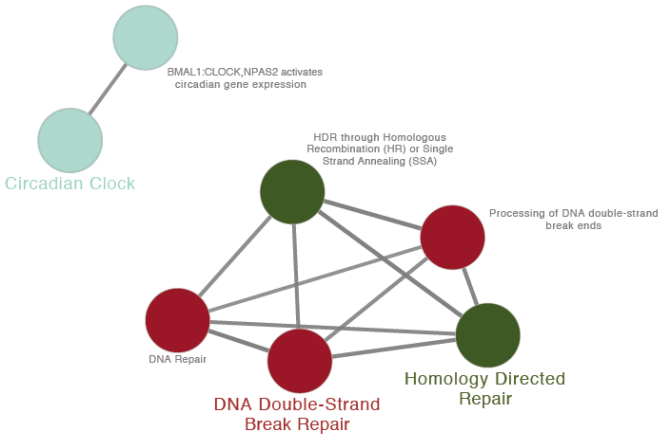
a



b



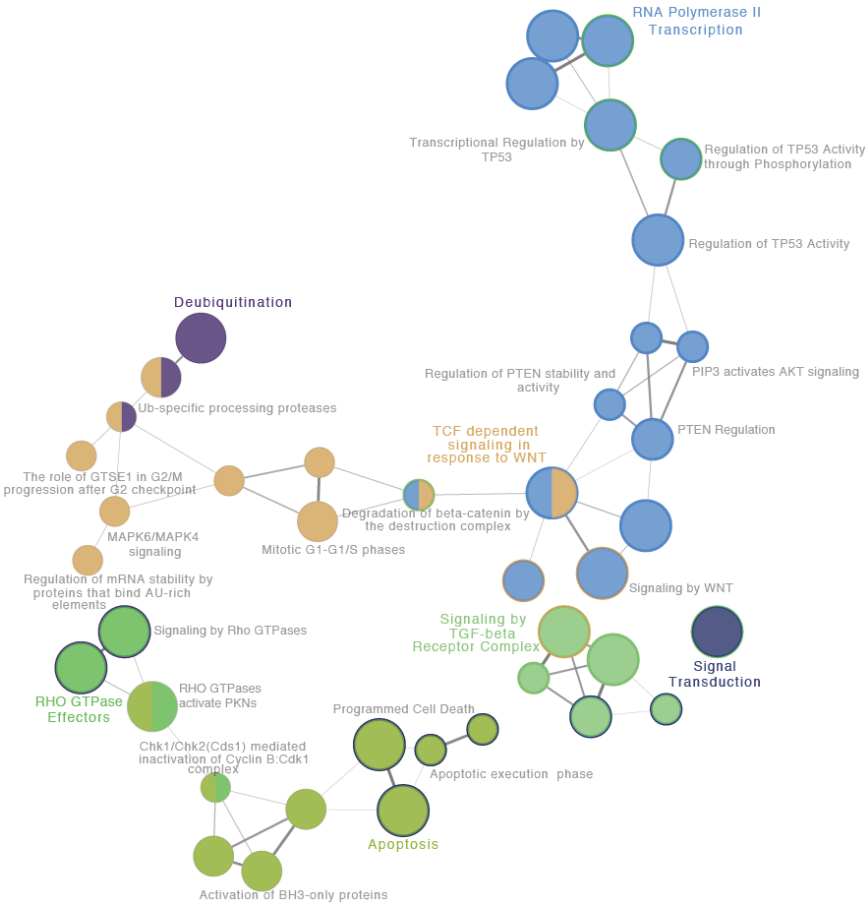
c


BMAL1:CLOCK,NPAS2 activates circadian gene expression
Circadian Clock
HDR through Homologous Recombination (HR) or Single Strand Annealing (SSA)
Processing of DNA double-strand break ends
DNA Repair
DNA Double-Strand Break Repair
Homology Directed Repair

d

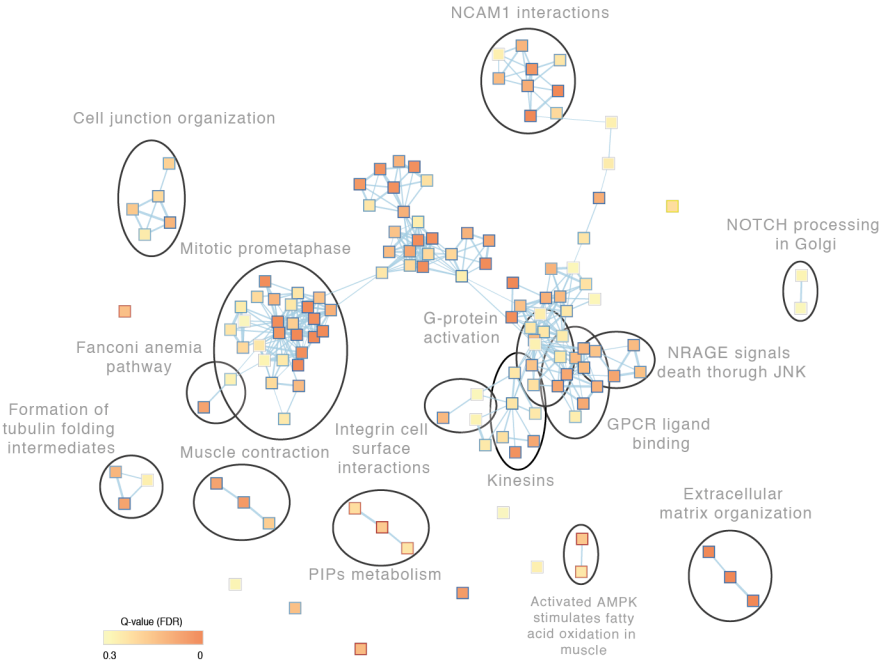

Mitochondrial protein import

e

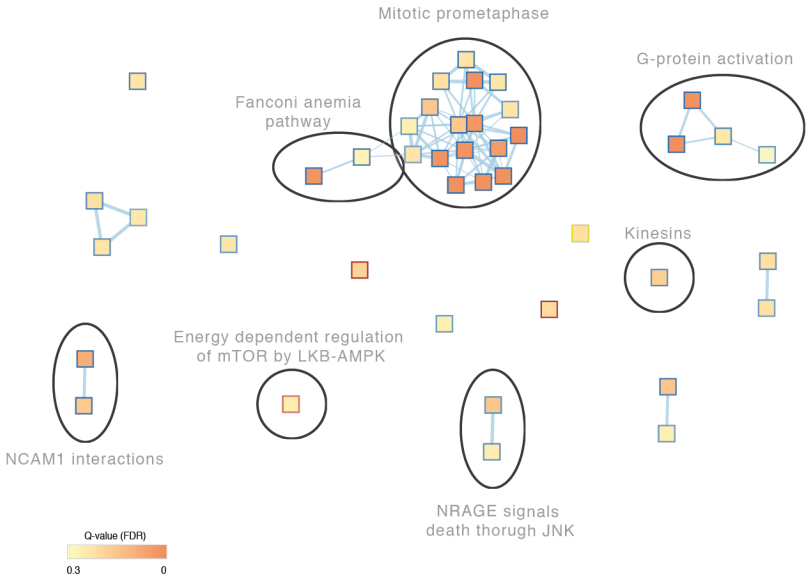

RNA Polymerase II Transcription
Transcriptional Regulation by TP53
Regulation of TP53 Activity through Phosphorylation
Regulation of TP53 Activity
Deubiquitination
PIP3 activates AKT signaling
Regulation of PTEN stability and activity
PTEN Regulation
Ub-specific processing proteases
TCF dependent signaling in response to WNT
The role of GTSE1 in G2/M progression after G2 checkpoint
Degradation of beta-catenin by the destruction complex
MAPK6/MAPK4 signaling
Mitotic G1-G1/S phases
Signaling by WNT
Regulation of mRNA stability by proteins that bind AU-rich elements
Signaling by Rho GTPases
Signaling by TGF-beta Receptor Complex
Signal Transduction
RHO GTPase Effectors
RHO GTPases activate PKNs
Programmed Cell Death
Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
Apoptotic execution phase
Apoptosis
Activation of BH3-only proteins

Supplementary Figure 2. Visualizations of the downstream enrichment analysis for the Reactome annotations for each high-confidence GRPM using the EnrichmentMap Cytoscape app. **(a)** G-alpha signaling (I). **(b)** G-alpha signaling (II). **(c)** Circadian clock. **(d)** Regulation of cell growth and angiogenesis. **(e)** Rho GTPases. Only biological processes with P value < 0.001 and False Discovery Rate (FDR) < 0.05 are shown in the representation. The colored circles represent gene sets, edges indicate overlapping genes, node size indicates the number of genes in the gene set and the color represents the associated FDR.

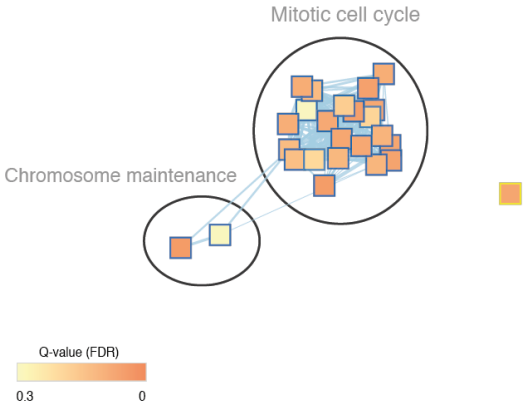
a



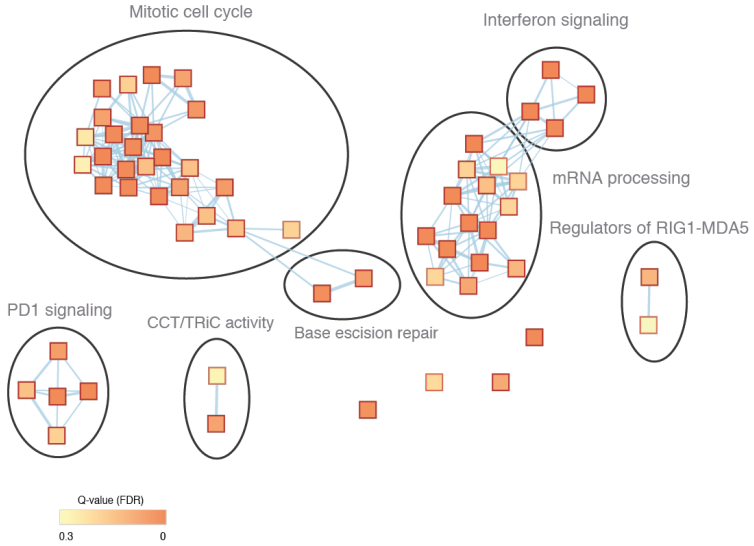
b



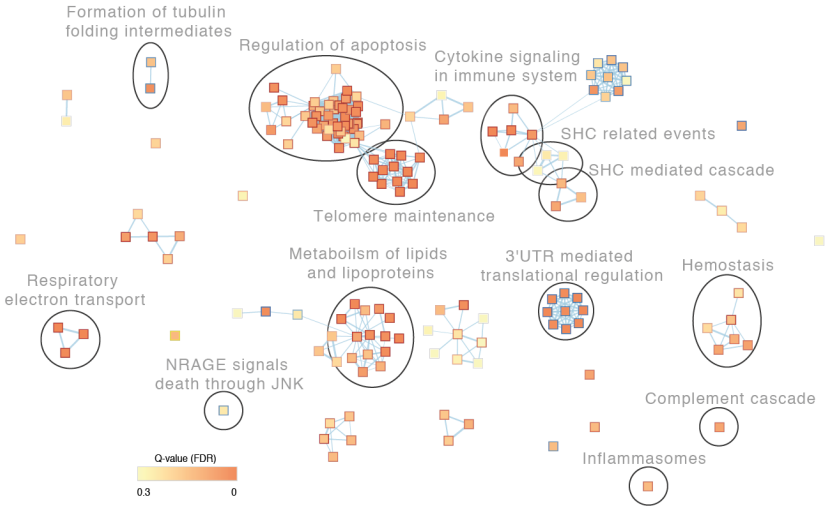
c


Mitotic cell cycle
Chromosome maintenance
Q-value (FDR)
0.3
0

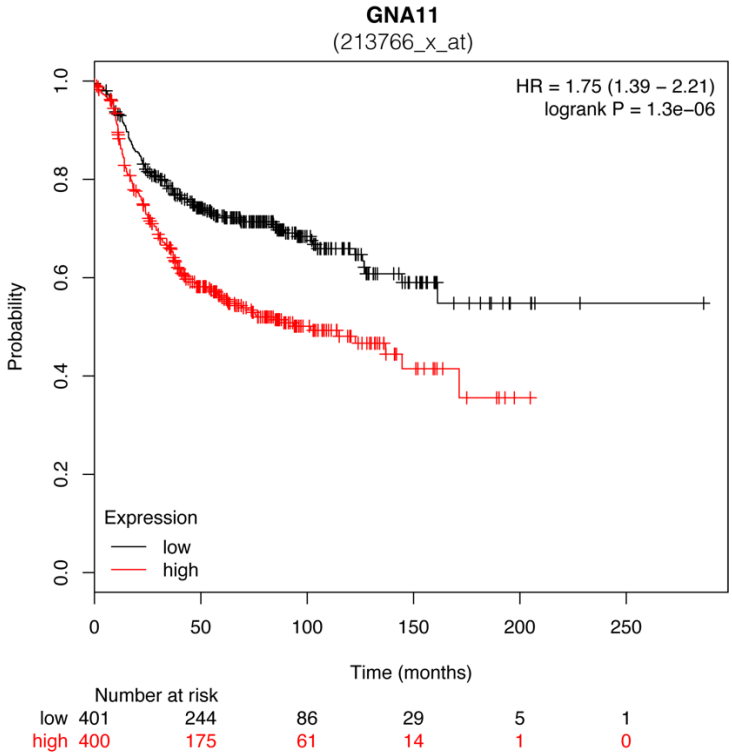
d


Mitotic cell cycle
Interferon signaling
mRNA processing
Regulators of RIG1-MDA5
PD1 signaling
CCT/TRiC activity
Base escision repair
Q-value (FDR)
0.3
0

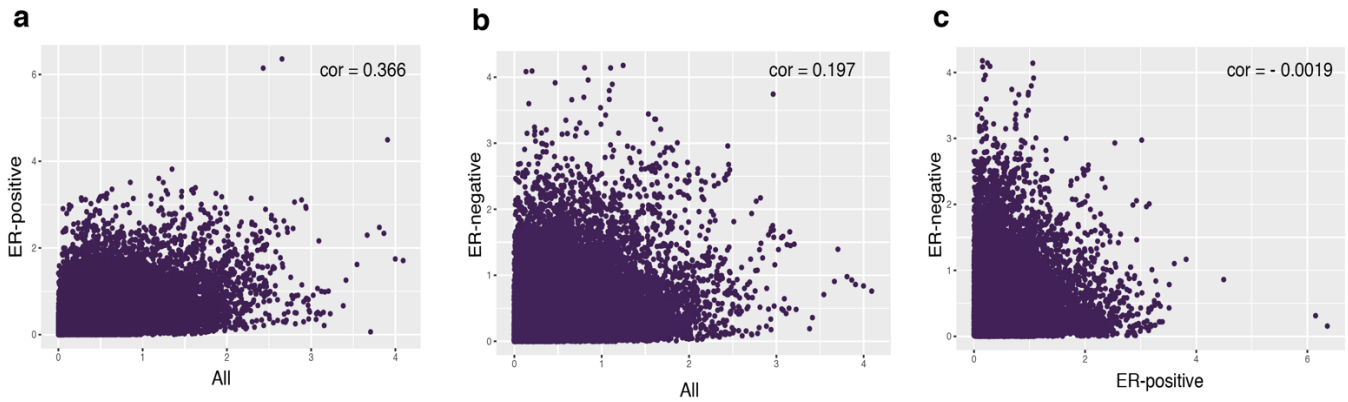
e


Formation of tubulin folding intermediates
Regulation of apoptosis
Cytokine signaling in immune system
SHC related events
SHC mediated cascade
Telomere maintenance
Respiratory electron transport
Metaboilsm of lipids and lipoproteins
3'UTR mediated translational regulation
Hemostasis
NRAGE signals death through JNK
Complement cascade
Inflammasomes
Q-value (FDR)
0.3
0

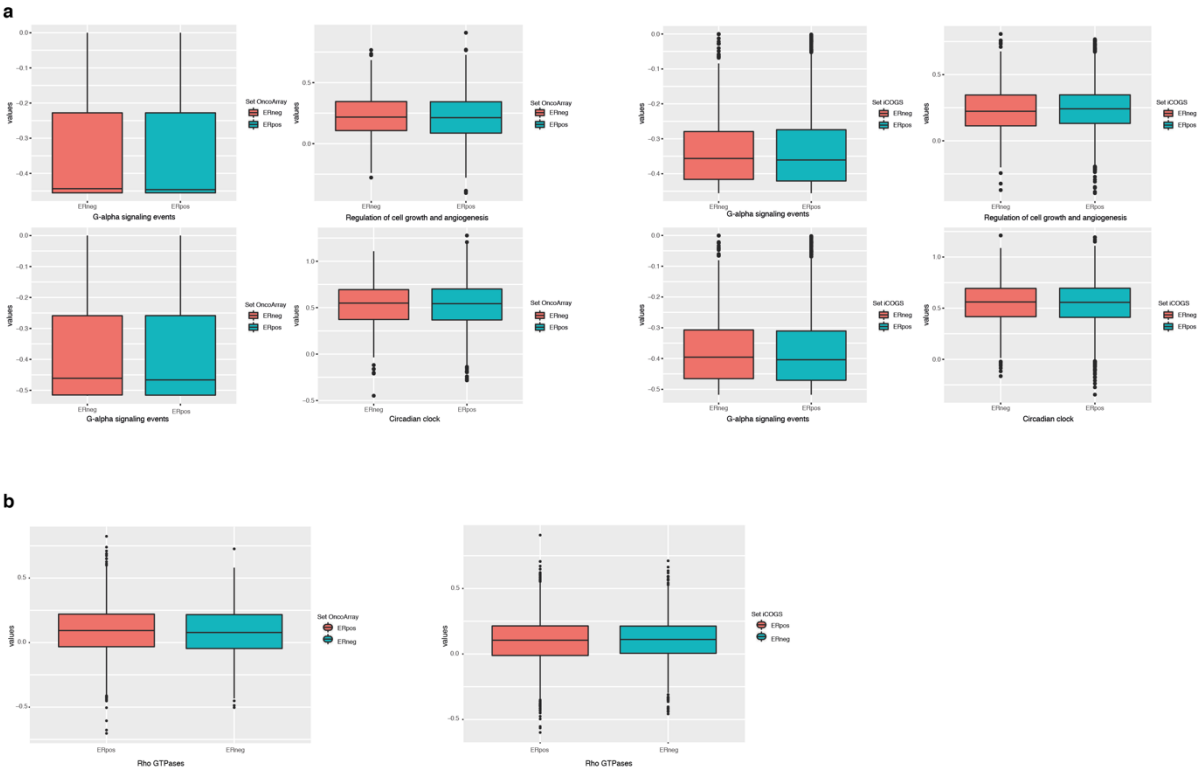
Supplementary Figure 3: Prognostic value of *GNA11* mRNA expression in Estrogen Receptor (ER)-negative breast tumors (n=1,214) with recurrence-free survival using KMplotter (kmplot.com/analysis). P value was computed using a logrank test. The Affymetrix IDs is 213766_x_at (*GNA11*).



Supplementary Figure 4: Scatter plots showing the -log10 P value of the ~21,800 gene scores computed within a 50-kb window-size around the gene region. Each dot represents a gene score. The correlations shown are Pearson correlations. **(a)** Estrogen Receptor (ER)-positive vs all breast cancers. **(b)** ER-negative vs all breast cancers. **(c)** ER-negative vs ER-positive breast cancers.



Supplementary Figure 5. Boxplots comparing the distributions of the Polygenic Hazard Scores (PHSs) for the Estrogen Receptor (ER)-status group in which the Germline-Related Prognostic Module (GRPM) was identified (red) versus the other ER-status (blue). The plot displays the median (center line), lower and upper hinges (25th and 75th percentiles respectively), two whiskers (scores outside the middle 50%) and all outlying points individually. **(a)** for the ER-negative high-confidence GRPMs. **(b)** for the ER-positive high-confidence GRPM. The Y-axes show the value of each PHS.



Supplementary Figure 6. QQ-plots of the observed and expected -log10 P values comparing Pascal's genes scores based on the maximum (left) and sum (right) statistics. **(a)** All breast cancer gene scores. **(b)** Estrogen Receptor (ER)-negative gene scores. **(c)** ER-positive gene scores.

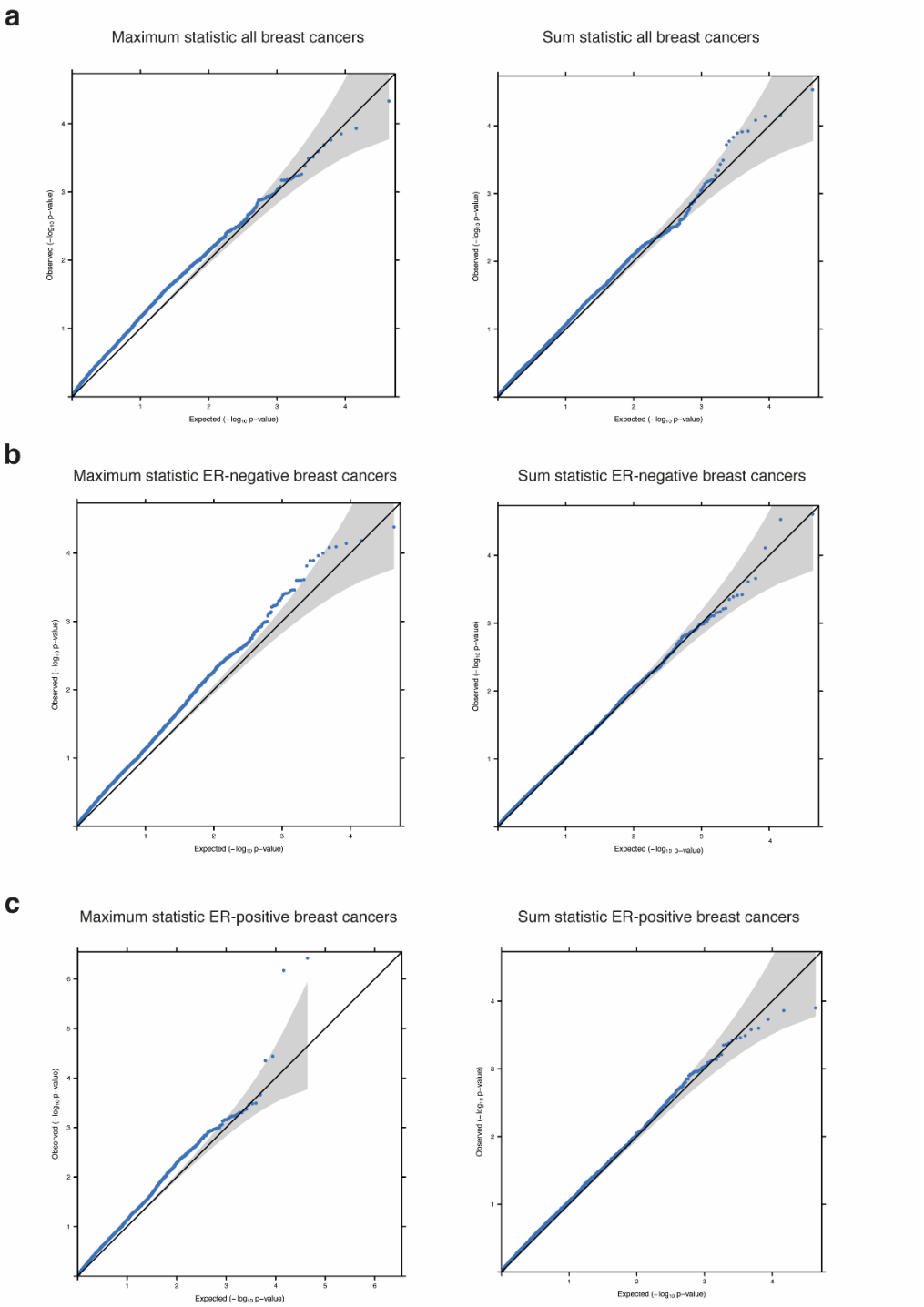


Figure 1. Network analysis pipeline (see Methods for details). **(a)** Cox models were used to estimate the association between each genetic variant and breast cancer-specific survival in 84,457 patients of the Breast Cancer Association Consortium (BCAC) dataset (discovery set). **(b)** The P values of the survival analyses for the genetic variants (blue squares) were used to compute gene scores using the Pascal algorithm. These gene scores were based on the maximum chi-squared signal within a window size of 50-kb around the gene region and accounted for linkage disequilibrium structure (depicted in a gradient blue scale). **(c)** The HotNet2 method was used to identify gene modules based on the $-\log_{10}$ P value of the computed gene scores. **(d)** The modules found by Hotnet2 were filtered to obtain a selection of high-confidence Germline-Related Prognostic Modules (GRPMs). We constructed a Polygenic Hazard Score (PHS) summarizing the prognostic effects of a set of selected genetic variants in the module. We then tested the association of this PHS with survival in both the discovery set (grey) and the independent set (orange). **(e)** We performed a functional characterization of the high-confidence GRPMs by studying the downstream transcriptional effects. For that, we used genotype and expression data from The Cancer Genome Atlas (TCGA). We computed the correlation between a GRPM's polygenic hazard score and the expression of all available genes. Based on these correlation values, a Gene Set Enrichment analysis assigned biological processes that were enriched among the genes most correlated with the prognostic variants in the GRPM.

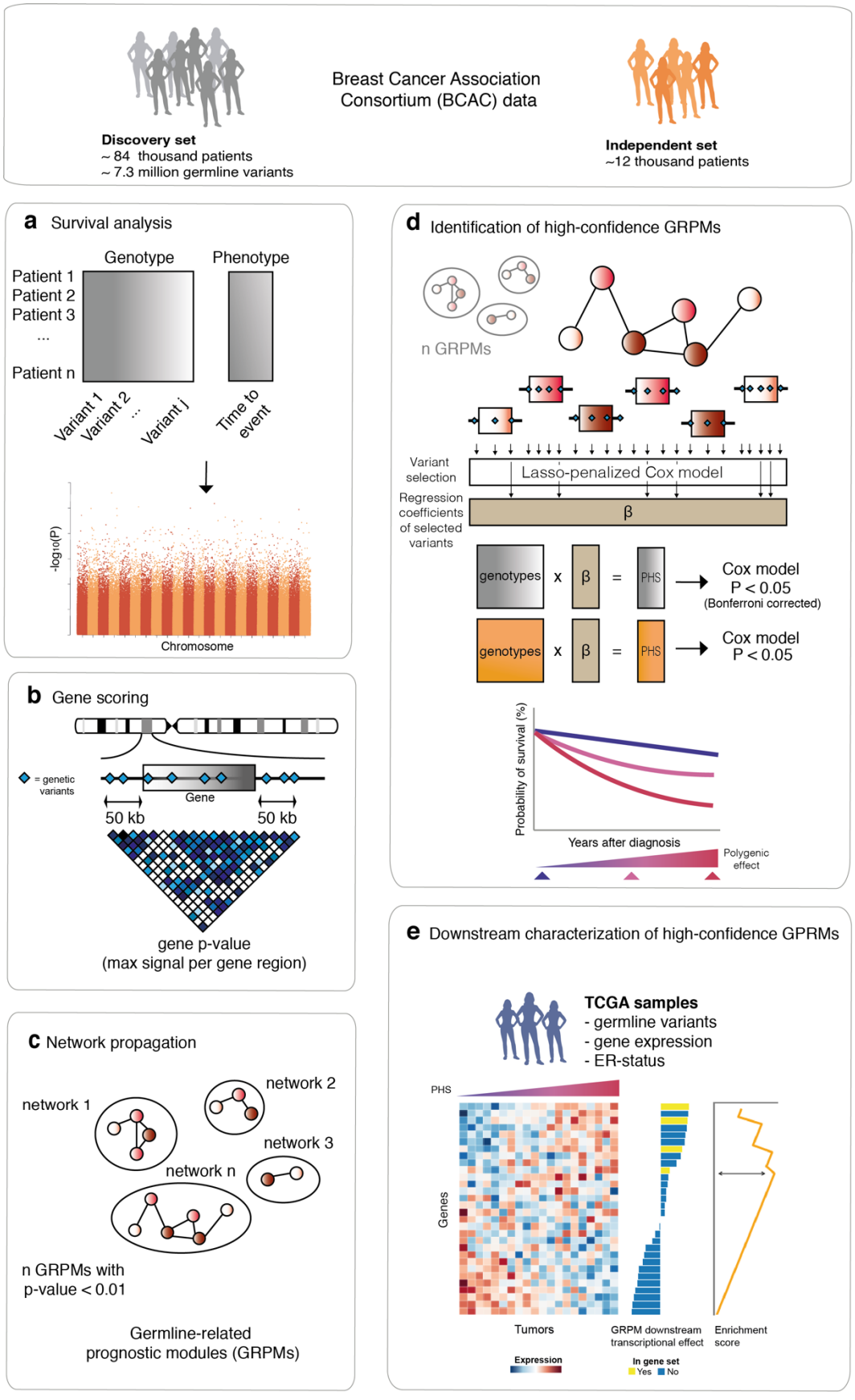

Breast Cancer Association Consortium (BCAC) data
Discovery set
~ 84 thousand patients
~ 7.3 million germline variants
Independent set
~12 thousand patients
a Survival analysis
Genotype
Phenotype
Patient 1
Patient 2
Patient 3
...
Patient n
Variant 1
Variant 2
...
Variant j
Time to event
-log10(P)
Chromosome
b Gene scoring
= genetic variants
Gene
50 kb
50 kb
gene p-value
(max signal per gene region)
c Network propagation
network 1
network 2
network 3
network n
n GRPMs with
p-value < 0.01
Germline-related
prognostic modules (GRPMs)
d Identification of high-confidence GRPMs
n GRPMs
Variant selection
Lasso-penalized Cox model
Regression coefficients of selected variants
β
genotypes
x
β
=
PHS
Cox model
P < 0.05
(Bonferroni corrected)
genotypes
x
β
=
PHS
Cox model
P < 0.05
Probability of survival (%)
Years after diagnosis
Polygenic effect
e Downstream characterization of high-confidence GPRMs
TCGA samples
- germline variants
- gene expression
- ER-status
PHS
Genes
Tumors
GRPM downstream transcriptional effect
Enrichment score
Expression
In gene set
Yes
No

Figure 2. Manhattan plots of the gene-level associations with breast cancer-specific survival. Plots show the association in **(a)** all breast cancer cases (n=84,457) **(b)** Estrogen Receptor (ER)-negative (n=14,529) and **(c)** ER-positive (n=55,701). The -$\log_{10}$ gene P values from the Pascal algorithm is shown on the y axis and genomic position on the x axis. The top significant genes and the most significant gene per chromosome (if -$\log_{10}$(P) > 3) are shown in red.

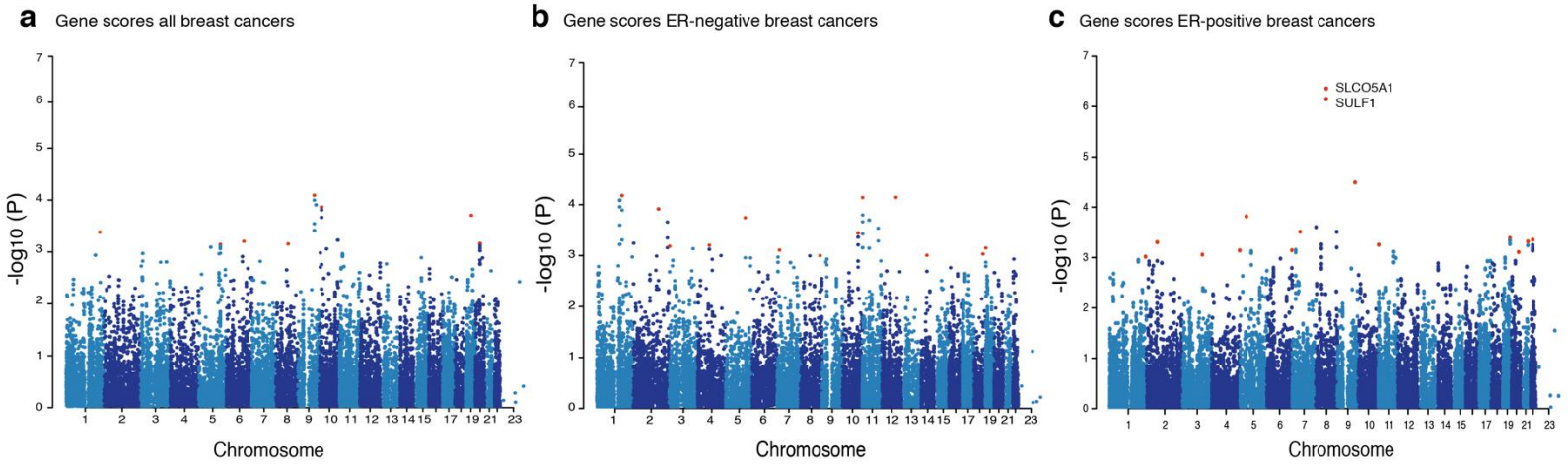


Figure 3. High-confidence Germline-Related Prognostic Modules (GRPMs). The GRPM is shown at the center of the circles, surrounded by the biological processes enriched among the downstream transcriptional effects of each module. Three modules were found for Estrogen Receptor (ER)-negative breast cancer (a-c) and one module was found for ER-positive breast cancer (d). **(a)** G-alpha signaling GRPMs. **(b)** Circadian clock GRPM. **(c)** Regulators of cell growth and angiogenesis GRPM. **(d)** Rho GTPases and apoptosis GRPM. **(e)** Plots illustrating the association between each GRPM's PHS and 10-year breast cancer specific-survival in the discovery and independent sets. HR: Hazard Ratio (per standard deviation of the PHS), CI: Confidence Interval. The error bars show the 95% confidence interval. The confidence intervals shown are two-sided, whereas the significance test performed was one-sided (see Methods).

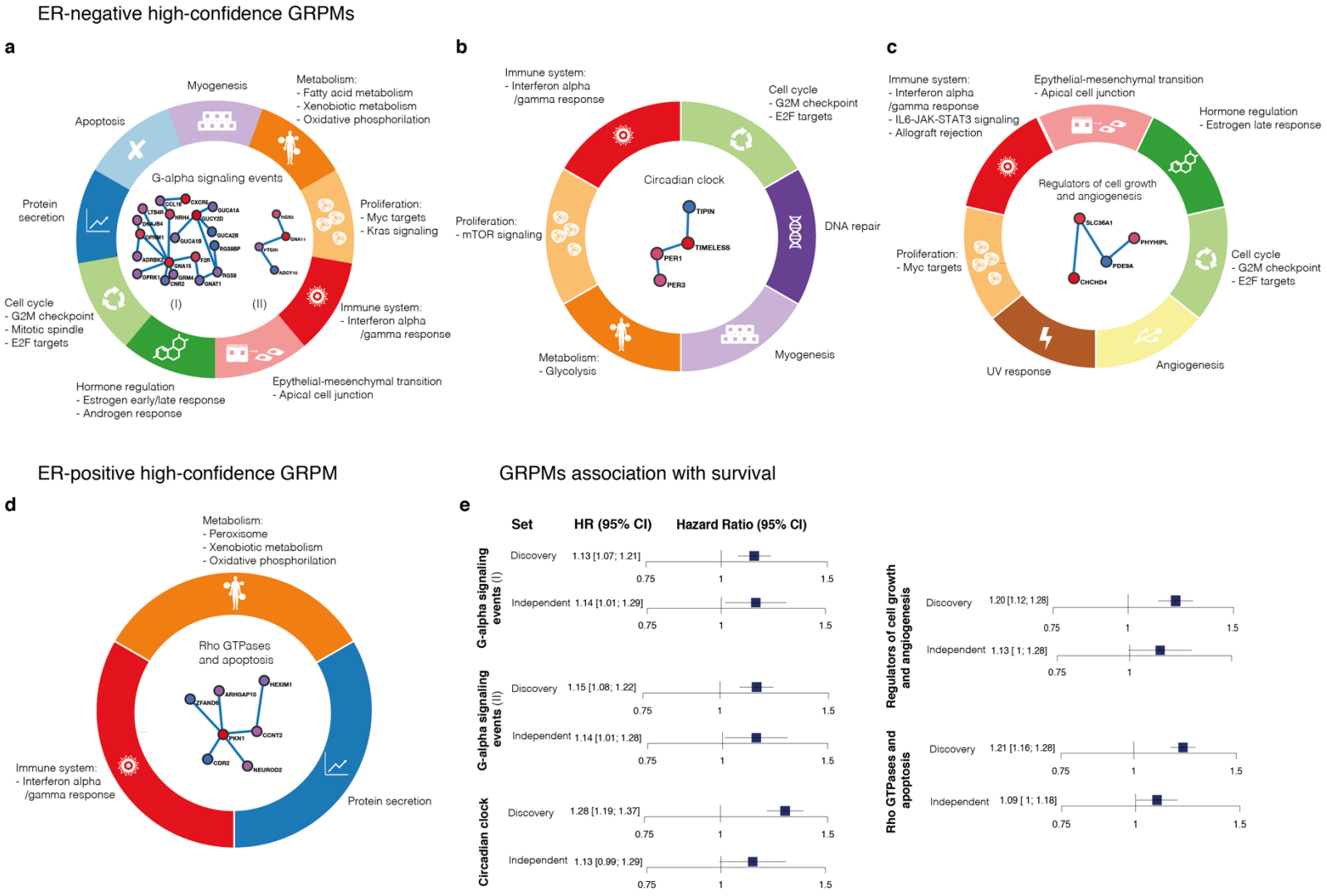


Figure 4. Genomic region 19p13.3 with the two genes *GNA11* and *GNA15*. The two G-alpha signaling high-confidence Germline-Related Prognostic Modules (GRPMs) identified in the Estrogen Receptor (ER)-negative subtype have a shared genetic signal in the same genomic region. **(a)** Top: -log10(P) for the association with survival (y axis) of all variants in the region 19p13.3 (y axis). Bottom: regression coefficients from the survival model for the genetic variants in the module's Polygenic Hazard Scores (PHSs)**. (b)** Scatter plot comparing the two modules' PHSs in the iCOGS independent validation set. PHS of the *GNA11* GRPM on the x axis and PHS of the *GNA15* GRPM on the y axis.