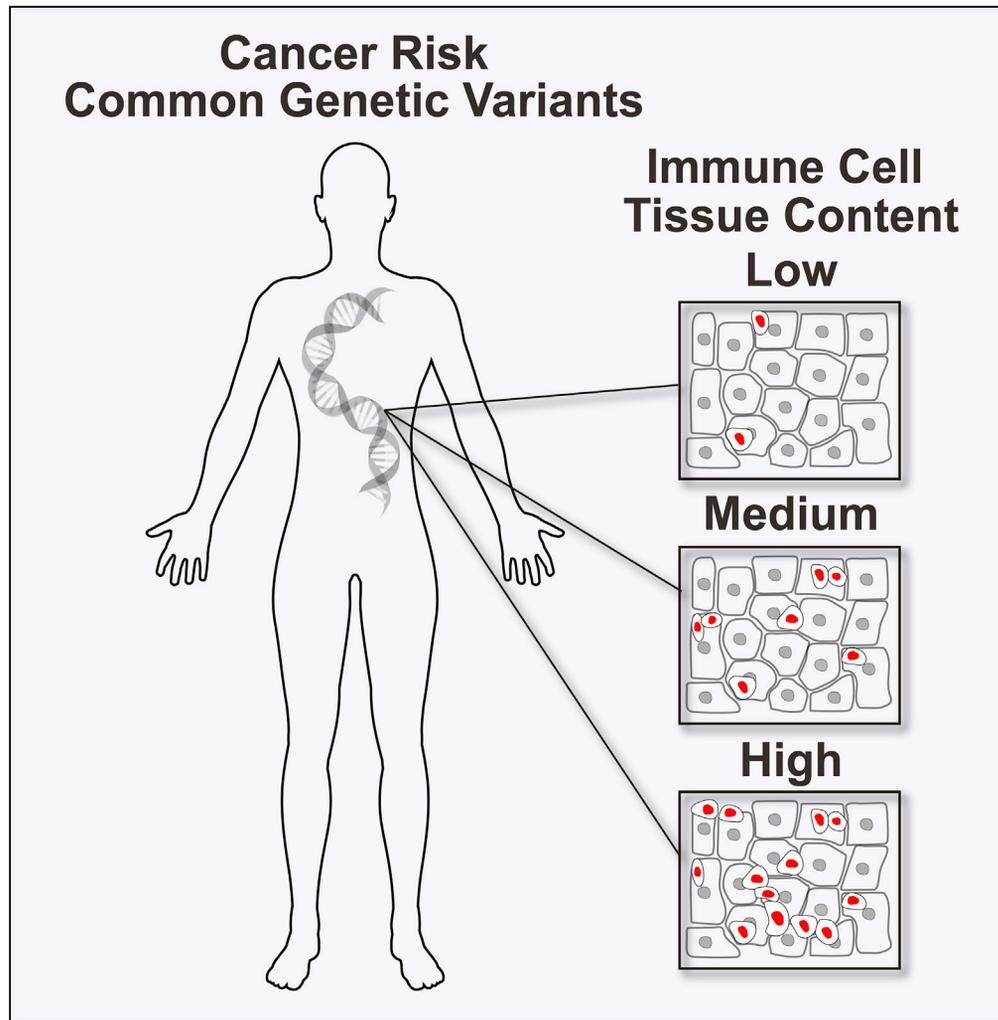


Article

Immune Cell Associations with Cancer Risk



Luis Palomero,
Ivan Galván-
Femenía, Rafael
de Cid, ..., Antonis
C. Antoniou,
Conxi Lázaro,
Miquel Angel
Pujana

aca20@medschl.cam.ac.uk
(A.C.A.)
clazaro@iconcologia.net (C.L.)
mapujana@iconcologia.net
(M.A.P.)

HIGHLIGHTS

Cancer risk genetic
variants linked to immune/
stromal cell tissue content

SH2B3 associated with
BRCA1/2 cancer risk and
immune cell counts

Peripheral immune cell
types linked to breast
cancer age at diagnosis

Palomero et al., iScience July
24, 101296
23, 2020 © 2020 The Author(s).
[https://doi.org/10.1016/
j.isci.2020.101296](https://doi.org/10.1016/j.isci.2020.101296)



Article

Immune Cell Associations with Cancer Risk

Luis Palomero,¹ Ivan Galván-Femenía,² Rafael de Cid,² Roderic Espín,¹ Daniel R. Barnes,³ CIMBA,³ Eline Blommaert,¹ Miguel Gil-Gil,⁴ Catalina Faló,⁴ Agostina Stradella,⁴ Dan Ouchi,^{5,13} Albert Roso-Llorach,^{5,13} Concepció Violan,^{5,13} María Peña-Chilet,⁶ Joaquín Dopazo,⁶ Ana Isabel Extremera,¹ Mar García-Valero,¹ Carmen Herranz,¹ Francesca Mateo,¹ Elisabetta Mereu,⁷ Jonathan Beesley,⁸ Georgia Chenevix-Trench,⁸ Cecilia Roux,⁹ Tak Mak,⁹ Joan Brunet,¹⁰ Razq Hakem,¹¹ Chiara Gorrini,⁹ Antonis C. Antoniou,^{3,*} Conxi Lázaro,^{12,*} and Miquel Angel Pujana^{1,14,*}

SUMMARY

Proper immune system function hinders cancer development, but little is known about whether genetic variants linked to cancer risk alter immune cells. Here, we report 57 cancer risk loci associated with differences in immune and/or stromal cell contents in the corresponding tissue. Predicted target genes show expression and regulatory associations with immune features. Polygenic risk scores also reveal associations with immune and/or stromal cell contents, and breast cancer scores show consistent results in normal and tumor tissue. *SH2B3* links peripheral alterations of several immune cell types to the risk of this malignancy. Pleiotropic *SH2B3* variants are associated with breast cancer risk in *BRCA1/2* mutation carriers. A retrospective case-cohort study indicates a positive association between blood counts of basophils, leukocytes, and monocytes and age at breast cancer diagnosis. These findings broaden our knowledge of the role of the immune system in cancer and highlight promising prevention strategies for individuals at high risk.

INTRODUCTION

The immune system maintains organismal integrity and function by continuously protecting itself from exogenous and endogenous assaults. The concept of “immunological surveillance of cancer” was first proposed by Burnet in 1970 (Burnet, 1970) and developed by Thomas about a decade later (Thomas, 1982). In this theory, the immune system inactivates or eliminates cancer-prone cells that are detected early in normal tissue (Ribatti, 2017). Although this idea remains a matter of debate, it is clear that some immune factors decisively influence cancer development and progression. Immunosuppression due to primary immunodeficiency or due to therapies administered to prevent organ transplant rejection and certain virus infections are associated with an increased risk of some cancers (Mortaz et al., 2016). In parallel, studies of mouse models with defined genetic alterations have demonstrated the relevance of immunosurveillance; for example, loss of *Nkg2d*, which encodes the activating receptor of natural killer (NK) and T cells, increases the risk of spontaneous neoplasms (Guerra et al., 2008).

Results from genome-wide association studies (GWASs) have identified risk-associated variants in loci coding for immune regulatory factors, such as *NKG2D* for cervical cancer risk (Chen et al., 2013). Indeed, pathway-based analyses of GWAS results have highlighted the involvement of immune-related processes in susceptibility to certain cancer types (Michailidou et al., 2017). In parallel, many germline genetic variants can influence immune cell infiltration in tumors (Lim et al., 2018). Therefore, immune-centered investigations of normal or precancerous tissue could yield fundamental evidence for improving cancer risk estimation and prevention (Spira et al., 2017). However, whether common genetic variants linked to cancer risk alter immune cell contents in the corresponding cancer target tissue, and/or at the systemic level, remains largely undetermined.

The balance between immunological surveillance and tolerance is determined from a complex interplay between different types of immune cells and other classes of stromal cells (Vinay et al., 2015; Gonzalez et al., 2018). Here, we describe an integrative analysis of genetic and transcriptome data from tissue

¹ProCURE, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona, Catalonia 08908, Spain

²GCAT-Genomes for Life, Germans Trias i Pujol Health Sciences Research Institute (IGTP), Program for Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Catalonia 08916, Spain

³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK

⁴Department of Medical Oncology, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona, Catalonia 08908, Spain

⁵Jordi Gol University Institute for Research Primary Healthcare (IDIAP Jordi Gol), Barcelona, Catalonia 08007, Spain

⁶Clinical Bioinformatics Area, Fundación Progreso y Salud (FPS), Bioinformatics in Rare Diseases (BiER), CIBERER, INB-ELIXIR-es, Hospital Virgen del Rocío, Seville 41013, Spain

⁷CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Catalonia 08003, Spain

⁸Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia

⁹Princess Margaret Cancer Centre, The Campbell Family Institute for Breast Cancer Research, Ontario Cancer

Continued



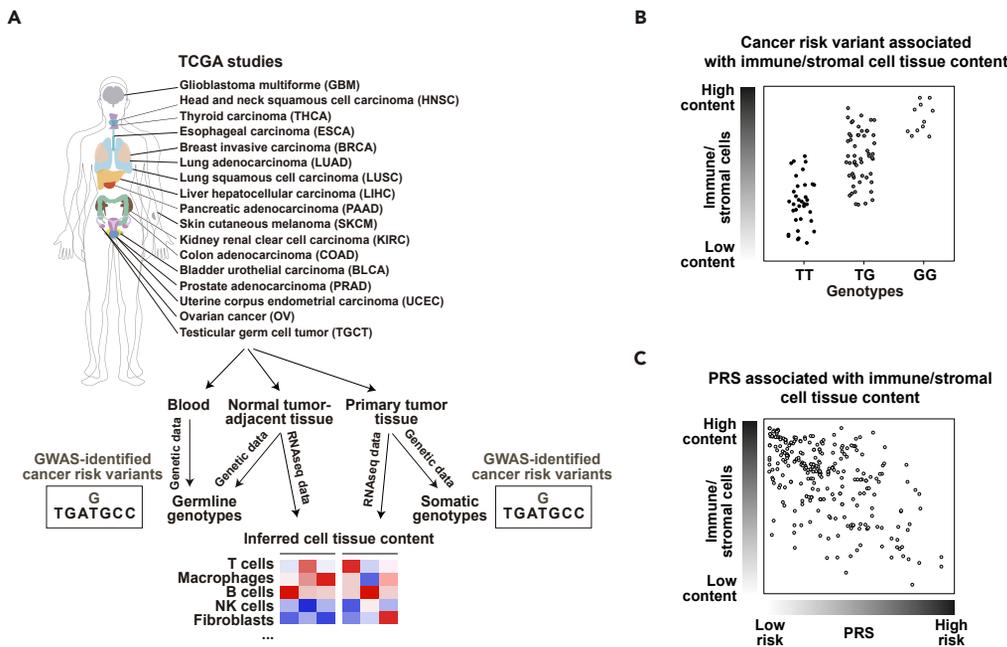


Figure 1. Strategy for Assessing the Effect of Immune/Stromal Cell Tissue Content on Cancer Risk

(A) TCGA cancer projects analyzed in this study and data analysis workflow.

(B) Association between gene expression-inferred immune/stromal cell tissue content and GWAS-identified risk variant.

(C) Association between gene expression-inferred immune/stromal cell tissue content and cancer PRS.

defined as normal and located adjacent to surgically removed tumors, and from primary tumors analyzed by The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research Network et al., 2013). This enables us to identify immune and stromal (hereafter “immune/stromal”) cell tissue content associations with the risk of several human cancer types. Beyond single variants, polygenic risk scores (PRSs) also show associations with differences in inferred immune/stromal cell tissue contents. Consistent associations among immune cell signatures, PRSs, and age at diagnosis suggest that higher immune cell infiltration reduces the risk of breast cancer. We identify the lymphocyte SH2B adaptor 3 (*LNK/SH2B3*) locus as linking immune cell counts and breast cancer risk, including that from *BRCA1/2* mutation carriers. To evaluate this connection further, we assess associations between breast cancer age at diagnosis and immune cell counts measured at diagnosis in routine clinical blood tests; the results further suggest that peripheral immune cell status influences breast cancer risk. Collectively, these findings may broaden our current knowledge of the biological basis of cancer risk and thereby suggest strategies for cancer prevention.

RESULTS

Strategy to Evaluate Immune/Stromal Cell Tissue Contents that Influence Cancer Risk

TCGA has greatly increased our knowledge of human cancer through multilayer biological analyses, which include genetic and gene expression profiling of tissue considered to be normal and situated adjacent to the cancer (hereafter referred to as “normal”) and primary tumors (Liu et al., 2018). In parallel, many successful GWASs have identified hundreds of germline genetic variants associated with the risk of common cancer types (Torkamani et al., 2018). By compiling GWAS results, we assigned cancer risk variants to 17 TCGA projects based on tissue of origin correspondences (Figure 1A lists the cancer study acronyms, and Table S1 lists the cancer risk variants). As deconvolution analyses of bulk gene expression enable robust inference of cell type content in heterogeneous samples (Avila Cobos et al., 2018), deduced cell content in normal tissue and tumors can be assessed for associations with cancer risk variants in multivariate analyses (Figure 1B). Moreover, as differences in cancer risk are more accurately defined by combinations of key variants in PRSs (Torkamani et al., 2018), it might be possible to better define the relevance of the immune/stromal cells by analyzing associations between PRSs and their corresponding signatures (Figure 1C; Table S2 summarizes the number of normal tissue and primary tumor samples available for each subsequent analysis).

Institute, University Health Network, Toronto, ON M5G 2M9, Canada

¹⁰Hereditary Cancer Program, Catalan Institute of Oncology, Biomedical Research Institute of Girona (IDIBGI), Girona, Catalonia 17190, Spain

¹¹Princess Margaret Cancer Centre, Department of Medical Biophysics, University Health Network and University of Toronto, Toronto, ON M5G 2C1, Canada

¹²Hereditary Cancer Program, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L’Hospitalet del Llobregat, Barcelona, Catalonia 08908, and Spanish Biomedical Research Network Centre in Oncology (CIBERONC), Instituto de Salud Carlos III, Madrid 28029, Spain

¹³Autonomous University of Barcelona, Bellaterra, Catalonia 08913, Spain

¹⁴Lead Contact

*Correspondence: aca20@medschl.cam.ac.uk (A.C.A.), clazaro@iconcologia.net (C.L.), mapujana@iconcologia.net (M.A.P.)

<https://doi.org/10.1016/j.isci.2020.101296>

To infer immune/stromal cell contents in normal and primary tumor samples using bulk tissue RNA sequencing (RNA-seq) data from TCGA, we applied a consensus-signature approach benchmarked against other methods (Consensus^{TME}; Jiménez-Sánchez et al., 2019). Using this approach, the computed immune/stromal estimations in the 17 identified TCGA datasets were typically found to be positively correlated with two other methods (Figure S1), as well as with estimates of leukocyte content measured by a different method (Taylor et al., 2018) (Figure S2). In turn, the estimates were generally found to be negatively correlated with aneuploidy (Figure S3), as expected (Taylor et al., 2018). In addition, the immune/stromal cell TCGA estimates showed significant differences between primary tumors with low or high levels of *CD274/PDL1* and *CD279/PDCD1* expression (Figures S4 and S5). Applying the method to RNA-seq data from whole blood samples of healthy adults also revealed positive correlations with immune cell enumerations using fluorescence-activated cell sorting (Newman et al., 2019) (Figure S6). Moreover, the estimates from this method were found to be highly correlated (Spearman's $\rho > 0.75$) with the numbers of immune/stromal cells used to generate 100 pseudo-bulk breast tumors (Figure S7 and Methods).

To further assess the validity of the inferred immune cell contents in TCGA, the deduced scores for each setting were assessed for associations with defined immune benchmark genes (Methods). In most settings, each inferred immune cell type content was found to be positively correlated with the expression of the assigned benchmark; the average Pearson's correlation coefficient values for all signature-benchmark pairs were 0.52 and 0.60 in the normal tissue and primary tumor sets, respectively (Figure S8 and Table S3). To assess further the coherence of the inferred immune cell contents, the corresponding scores were tested for association with the activity status of immune-related signaling pathways (Cubuk et al., 2018). This analysis revealed coherent clustering of immune and stromal cell types in normal tissue (Figure S9).

Identification of Cancer Risk Variants Linked to Differences in Immune/Stromal Cell Tissue Content

A total of 1,453 cancer risk variants were compiled from various sources; 214 of these were directly genotyped in TCGA, and the rest were imputed. After applying quality controls and filtering criteria (Methods), 627 and 966 variants were analyzed as potential immune/stromal quantitative trait loci (isQTLs). The isQTLs were identified using multivariate regressions including covariates of gender (when informative), age at diagnosis, tumor stage, and histology. The significance of the associations in each setting was concluded from 1,000 permutations (Methods). Tumor data were also analyzed because germline risk alleles are frequently associated with defined cancer histopathological and biomarker features (Michailidou et al., 2017). To avoid redundant tests, only cell signatures with eigenvalues >1 were examined in each setting (Table S4). Through this methodology, 22 significant isQTLs were identified. These comprised normal tissue corresponding to esophageal carcinoma (ESCA), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), and uterine corpus endometrial carcinoma (UCEC), and primary tumors of breast cancer (BRCA), head and neck squamous cell carcinoma (HNSC), LUSC, and ovarian serous cystadenocarcinoma (OV) (Figure 2A and Table S5).

Several of the identified isQTLs involved differences in endothelial and fibroblast cell content (Table S5), and these signals may also indicate links with immune cell differences: for instance, rs4072037 is associated with endothelial cell content in normal esophageal tissue, and this variant corresponds to a *cis*-expression (*cis*-e) QTL in several normal tissue types (GTEx Consortium, 2013) of genes whose products are functionally relevant in the immune system and biology of endothelial cells (Stenina-Adognravi, 2014), including *GBA*, *GBAP1*, *TSP3*/*THSB3*, and *MUC1*, which are locus-mapped genes. In addition, the cancer pleiotropy rs11168936 variant (Fehringer et al., 2016) is associated with differences in fibroblast content in normal lung tissue corresponding to LUSC, and this variant is a *cis*-eQTL for *C1QL4* in several normal tissue types (GTEx Consortium, 2013). Intriguingly, *C1q* is a regulator of dendritic cell maturation (van Kooten et al., 2008), and we found this variant also to be associated with dendritic cell content (Figure 2B). In normal lung tissue corresponding to LUAD, rs17078110 is associated with B cells, and this locus codes for *SASH1*, a regulator of TLR4 signaling and cytokine production (Dauphinee et al., 2013). Among the isQTLs identified in tumors, rs3764419 is associated with cytotoxic cell content in OV. This variant is a *cis*-eQTL for *ATAD5* (GTEx Consortium, 2013), whose product is essential for proper B cell biology and immunoglobulin production (Zanotti et al., 2015). Overall, these data suggest that some cancer risk variants are associated with immune/stromal cell tissue content, and that this link is mediated by alterations in genes of functional importance to the immune system.

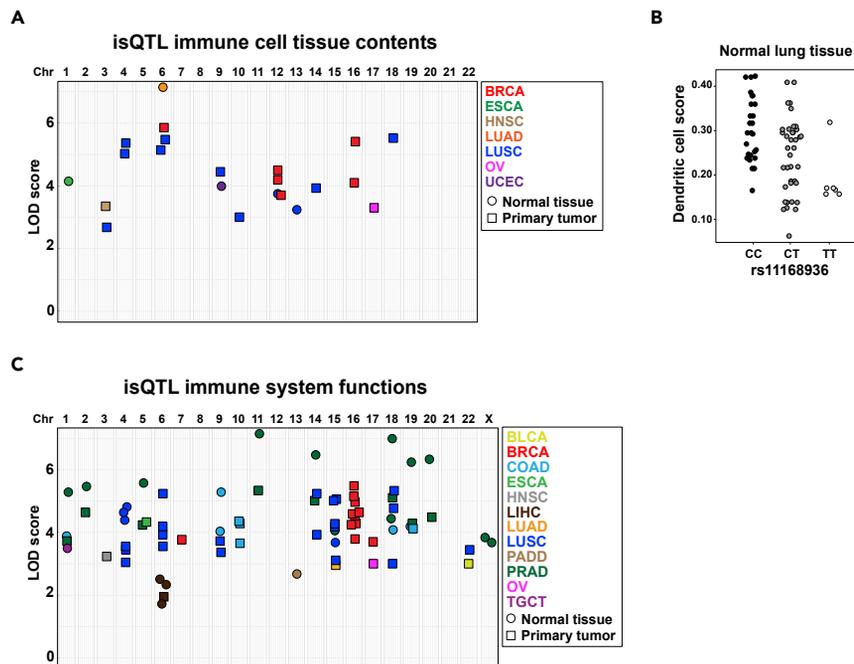


Figure 2. isQTLs Linked to Cancer Risk

(A) Graph indicating the association values (LOD scores) and relative chromosome locations of 22 isQTLs identified in normal tissue and primary tumor analyses (Table S5).

(B) isQTL rs11168936 for dendritic cell content in normal tissue corresponding to LUSC.

(C) Graph showing the LOD scores and relative chromosome locations of isQTLs identified using five TCGA-defined immune signatures (Table S6).

Identification of Cancer Risk Variants Associated with Immune System Functions in Target Tissue and Primary Tumors

The TCGA consortium examined 160 immune system-related gene expression signatures across hundreds of tumors and identified five of them as being informative for cancer classification: IFN- γ response, lymphocyte infiltration score, macrophage regulation, TGF- β response, and wound healing (Thorsson et al., 2018). Therefore, we sought to expand on the aforementioned cell-type-based associations by analyzing these five additional signatures using the same method as introduced earlier: multivariate regression with significance determined from 1,000 permutations. This study identified 75 isQTLs, of which 11 variants had been identified in the previous isQTL analyses, which represents a significant concordance (Fisher's exact test, $p < 0.0001$; Figure 2C and Table S6). Taking both analyses into account suggests that the risk of 13 cancer types may be influenced by immune/stromal cell tissue content.

Of the 57 unique variants identified from all isQTLs, five were linked to tumor suppressor genes with recognized roles in the immune system: *CDKN2A/B*, *DCC*, *MUC1*, and *SASH1*. In addition, genomic enhancers identified in T helper, regulatory, effector, memory, and mononuclear cells were significantly over-represented in this unique variant set relative to all human variants: > 2-fold enrichments, binomial test p values < 0.05 (Ward and Kellis, 2012). Consistent with this observation, 8 (14%) variants corresponded to expression (e) QTLs from 18 immune-related genes in normal human tissue (GTEx Consortium, 2013) and 13 (25%) corresponded to eQTLs identified in CD4+ and/or CD8+ T cells (Kasela et al., 2017) (Tables S5 and S6). To evaluate the relevance of these observed percentages, we examined the expected proportions when considering all cancer risk variants studied; lower percentages were identified in both analyses, with expected proportions of 11% (115/1,079) for eQTLs of immune-related genes in normal human tissue (GTEx Consortium, 2013) and 14% (151/1,079) for eQTLs in CD4/8+ T cells (Kasela et al., 2017). We then examined whether the eQTL gene targets documented within the isQTLs were functionally coherent by determining the proportion of significant gene expression-immune/stromal cell signature correlations and comparing the results with those from equivalent 1,000 random gene sets. Both isQTL sets (Tables S5 and S6) included

a higher proportion of eQTL gene targets that were positively correlated with immune/stromal cell signatures than expected by chance (Figure S10). Finally, variants correlated ($r^2 > 0.8$) with each isQTL were intersected with various functional genomic data from B cells, monocytes, and CD4+ and CD8+ T cells, and for potential effects on protein coding sequences (Methods). These analyses identified two additional candidate genes (*LIF* and *OSM*) with established functions in the immune system, being involved in cytokine signaling (Table S7). Together, these data indicate that a substantial proportion of the isQTLs identified influence genes whose expression is associated with immune system functions.

PRS Associations with Immune/Stromal Cell Tissue Content Highlight Breast Cancer Risk

The effects of individual cancer risk variants are generally small, but their combinations within PRSs can potentially identify individuals who are at substantially higher risk than average for the population (Torkamani et al., 2018). Therefore, reported PRSs were computed in the corresponding normal tissue and primary tumor TCGA settings and evaluated for associations with immune/stromal cell contents using multivariate analyses as described earlier. The study of normal tissue was limited to breast. Despite the valuable TCGA resource, the available sample size sets limited the detection of nominal significant associations to those with correlation coefficients of $r > 0.3$ in normal breast and of $r > 0.12$ in BRCA; higher correlations would be required for all other normal or tumor settings (Figure S11).

In normal breast, most immune/stromal cell contents tended to be negatively correlated with the corresponding PRSs; the PRS cell signature correlation coefficients for overall and estrogen receptor (ER)-positive breast cancer were significantly less than zero (p values < 0.001 ; Figure 3A). The ER-negative PRS could not be computed because of the relatively low number of normal samples of this subtype and with complete data. Analogous limitations were encountered when attempting to analyze triple-negative breast cancer (TNBC) and human epidermal growth factor receptor 2 (HER2)-positive breast cancers, and there were no HER2-specific PRSs to analyze whatsoever. Potentially protective cell types (i.e., those exhibiting a nominally significant negative correlation between cell content and PRS) in the aforementioned two breast cancer settings included dendritic cells, eosinophils, macrophage M2, monocytes, neutrophils, and T cell terminal differentiation (Figure 3A).

In addition to the breast cancer PRSs, eight other scores (Fritsche et al., 2018) were examined in their corresponding primary tumor TCGA settings. The distribution of the correlation coefficients between immune/stromal cell tissue content and the PRS was again found to be less than zero not only in BRCA but also in glioblastoma multiforme (GBM; with a major contribution for fibroblast content) and thyroid carcinoma (THCA; Figure 3B). Conversely, positive correlations were detected in bladder urothelial carcinoma (BLCA), OV, prostate adenocarcinoma (PRAD), skin cutaneous melanoma (SKCM), and, principally, in LUAD and LUSC (Figure 3B). Conversely, positive correlations were detected in BLCA, serous OV, PRAD, SKCM, and, principally, in LUAD and LUSC (Figure 3B). Therefore, risk stratification based on PRSs may also be linked to differences in immune/stromal cell content in normal and/or tumor tissue. LUAD and LUSC PRSs shared positive correlations ($p < 0.05$) with cytotoxic and NK cell tissue contents; however, these associations may be influenced by smoking status, because LUAD current smokers showed an opposite trend (Figure S12).

Combined analyses of normal tissue and primary tumor data further suggested common protective effects for high immune cell content in breast and colorectal tissue, and also potentially in brain and a few other settings (Figure 3C). In contrast, high immune cell content might principally increase the risk of lung, bladder, and pancreatic cancer (Figure 3C), although, as already noted, smoking may influence these associations. Then, analyses of COAD subtypes (Methods) suggested protective effects for high immune cell content in genomic stable tumors (Figure 3D, left panel), but this association might be biased due to PRS development in overall incident cases. When analyzing the COAD molecular subtypes, lower risk of CSM3 might also be associated with higher immune cell content (Figure 3D, right panel). The sample sets of these subtype analyses were relatively small to obtain robust conclusions, but, when compared with normal colorectal tissue, an opposite trend was observed (Figure 3D, left panel), which suggests that immune cell infiltration has different roles between normal tissue and tumors.

As described earlier, the normal breast and BRCA settings both showed PRS-cell signature negative correlations. To assess these observations further, the correlation estimates were compared with those from similar analyses using age at diagnosis instead of the PRSs. In normal breast tissue, the

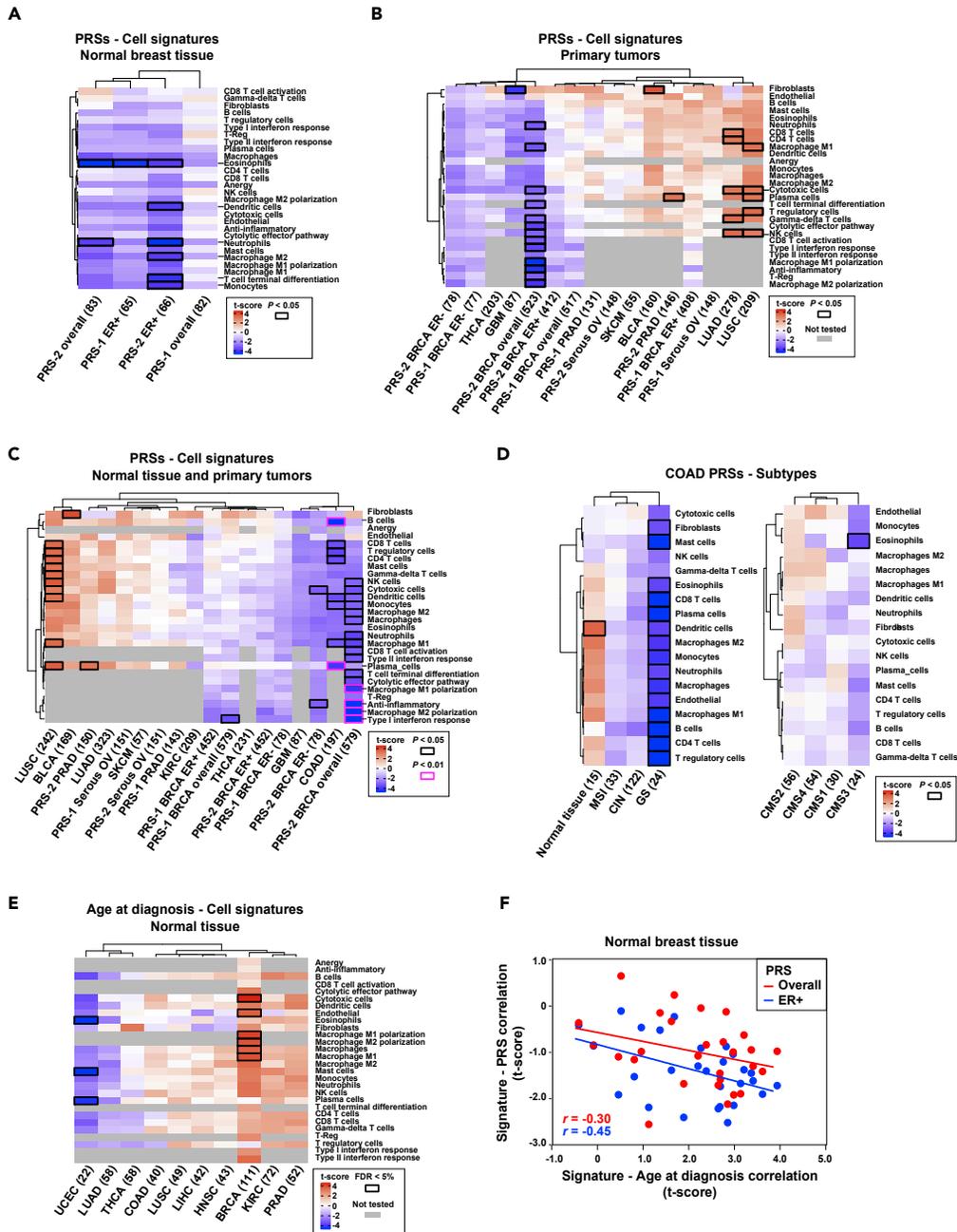


Figure 3. Associations between Immune/Stromal Cell Signatures and PRSs

(A) Unsupervised clustering of the results of the regression analysis between cell signatures and PRSs in normal breast tissue. The y axis depicts the cell type signatures, and the x axis shows the PRSs. Sources #1 and #2 of the PRSs are detailed in [Methods](#). ER+ and ER- indicate estrogen receptor-positive and estrogen receptor-negative subsets, respectively. The maximum sample size used in each analysis is shown. The color scale (t-score) is calculated as the β estimate divided by the standard error. Nominally significant associations are indicated by black-outlined rectangles.

(B) Unsupervised clustering of the coefficients of the regression of cell signature values in primary tumor TCGA studies and the corresponding PRSs. The gray-filled rectangles indicate “not tested” correlations because the corresponding cell signatures were only defined for breast cancer.

(C) Unsupervised clustering of the coefficients of the regression of cell signature values in combined normal tissue and primary tumor datasets, and the corresponding PRSs. The regression p values <0.01 are also indicated as depicted in the inset.

Figure 3. Continued

(D) Unsupervised clustering of the coefficients of the regression analysis between PRSs and cell signatures in combined normal tissue and primary tumors of the COAD study, divided by cancer subtypes.

(E) Unsupervised clustering of the results of the regression analysis of cell signatures in normal tissue and age at diagnosis across TCGA studies. Associations significant at a false discovery rate (FDR) < 5% are indicated by black-outlined rectangles.

(F) Negative correlations between the coefficients of regressions of immune/stromal cell contents and age at diagnosis or PRSs in normal breast tissue, for all cases and only ER-positive cases. The correlation coefficients are shown.

immune/stromal cell contents tended to show positive correlations with age at diagnosis ($p < 0.001$; [Figure 3E](#)). Consequently, negative correlations were detected between the estimates from the two parallel analyses, considering all cases or solely ER-positive cases ([Figure 3F](#)). Therefore, relatively higher immune/stromal cell content in normal breast might be a factor protecting against development of malignancy.

SH2B3 Connects Immune Cell Tissue Content with Breast Cancer Risk

The identified cancer risk isQTLs could be explained by peripheral alterations in immune cells. Examination of GWAS results for blood cell traits revealed that the tumor COAD isQTL rs12412391 in chromosome 10 ([Table S6](#)) is in linkage disequilibrium ($r^2 = 0.93$) with rs11190133, which is associated with differences in platelets in the UK Biobank study ([Astle et al., 2016](#)). These variants constitute an eQTL of *NKX2-3* ([GTEx Consortium, 2013](#)), and, remarkably, loss of the mouse ortholog causes developmental alterations in the spleen, colonic crypts, and lymphocyte tissue homing ([Pabst et al., 1999](#)). In addition to this locus, the tumor BRCA isQTL rs11065979 in chromosome 12 ([Table S5](#)) was associated with blood count differences in basophils, erythrocytes, eosinophils, leukocytes, monocytes, and neutrophils in the UK Biobank study ([Astle et al., 2016](#)) ([Table S8](#)). The same study also indicated an association with breast cancer risk ($p = 0.0003$; [Table S8](#)). This variant has also been linked to cancer pleiotropy ([Fehring et al., 2016](#)) and psoriasis ([Tsoi et al., 2017](#)), among other traits (GWAS Catalog). A variant in linkage disequilibrium, rs3184504 ($r^2 = 0.89$), had also been associated with breast cancer risk ([Fehring et al., 2016](#)), serum IgA levels ([Jonsen et al., 2017](#)), and various autoimmune diseases ([Webb and Hirschfield, 2016](#)), among other traits (GWAS Catalog).

To investigate further the role of the isQTL identified in chromosome 12 and linked to breast cancer risk, we analyzed association results from *BRCA1/2* mutation carriers. Both depicted variants showed nominal associations with breast cancer risk in women carriers of germline *BRCA1* or *BRCA2* mutations: *BRCA1* mutation carriers, rs11065979 hazard ratio (HR) = 0.96, 95% confidence interval (CI) 0.92–0.99, $p = 0.018$; rs3184504 HR = 0.95, 95% CI 0.92–0.99, $p = 0.006$; *BRCA2* mutation carriers, rs11065979 HR = 0.94, 95% CI 0.90–0.99, $p = 0.019$; and rs3184504 HR = 0.93, 95% CI 0.89–0.98, $p = 0.003$. Then, wider examination of this region in chromosome 12 identified several genetic associations ($p < 0.01$) with breast and/or ovarian cancer risk in these women ([Figure 4A](#) and [Table S9](#)).

The chromosome 12 locus identified here includes many eQTL signals for *SH2B3* in EBV-transformed lymphocytes and normal tissue ([Figure 4A](#), bottom panel). Next, to evaluate potential causality linked to *SH2B3*, complementary gene expression analyses were performed using the normal breast tissue TCGA data. First, the expression of *SH2B3* was found to be positively correlated with most of the immune cell/stromal cell signatures ([Figure 4B](#)); second, *SH2B3* expression was also found to be positively correlated with age at diagnosis, adjusted for tumor stage and regardless of cancer subtype ([Figure 4C](#)); third, an 84-gene signature corresponding to gene and protein functional relationships with mouse *Sh2b3* and/or human *SH2B3* ([Huan et al., 2015](#)) was also positively correlated with age at diagnosis ([Figure 4D](#)); and last, *SH2B3* expression was positively correlated with the protein measures of CD26, cell surface glycoprotein receptor important for T cell activation ([Klemann et al., 2016](#)), and TFCR, transferrin receptor required for erythropoiesis and immune system development ([Jabara et al., 2016](#)) ([Figure 4E](#)). In addition, the association between *SH2B3* expression in normal breast and age at diagnosis was replicated in an independent dataset ([Terunuma et al., 2014](#)): $n = 47$, $r = 0.30$, $p = 0.039$. Therefore, an identified isQTL may influence breast cancer risk through perturbation of *SH2B3* expression, which is expected to be fundamental for accurate systemic development and function of immune cell populations ([Li et al., 2000](#); [Velazquez et al., 2002](#); [Jabara et al., 2016](#)).

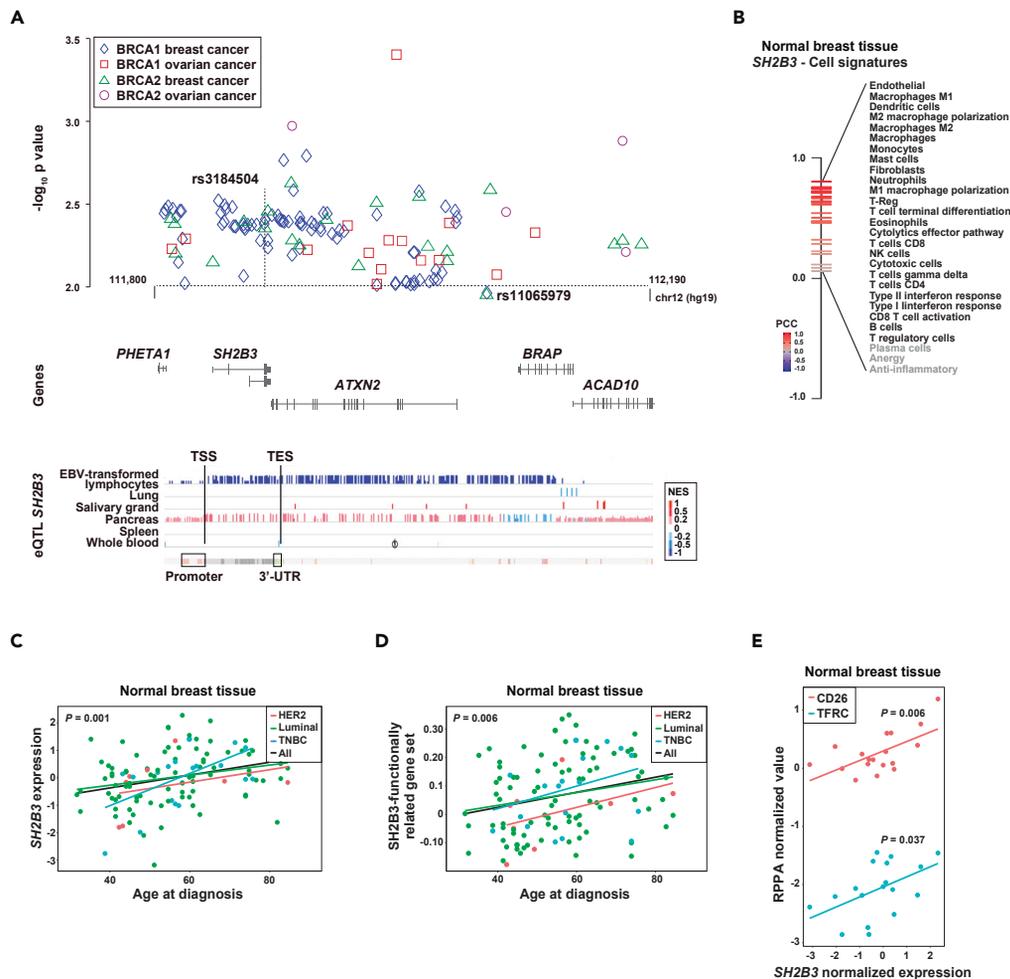


Figure 4. The *SH2B3* Locus Shows Associations with Breast Cancer Risk and Immune Cell Features

(A) Graph showing the chromosome 12 association results ($-\log_{10} p$ value, y axis) with breast and ovarian cancer risk (as depicted in the inset) in women carriers of *BRCA1/2* mutations. The rs3184504 and rs11065979 variants are indicated.

(B) Rank of expression correlations (Pearson's correlation coefficients [PCCs]) between *SH2B3* and immune/cell signatures in normal breast. All PCCs were >0 , but three of them did not reach nominal significance (marked gray).

(C) Positive correlation between *SH2B3* expression in normal breast and age of diagnosis of breast cancer. The trend lines for all cases and subtypes (luminal, HER2-positive, and triple-negative breast cancer [TNBC]) are shown. The correlation p value from the multivariate regression analysis is shown.

(D) Positive correlation between *SH2B3* functionally related gene set in normal breast and age of diagnosis of breast cancer.

(E) Positive correlation between *SH2B3* expression and CD26 and TFRC protein expression as measured by TCGA reverse-phase protein array (RPPA) assays. The correlation p value from the multivariate regression analysis is shown.

Peripheral Immune Cell Counts Are Associated with Breast Cancer Risk

To assess the proposed link between breast cancer risk and peripheral immune cell counts, which in turn might be influenced by specific genetic variants and gene candidates, a retrospective case-cohort study was performed. Data on age at diagnosis, tumor stage and subtype, and blood test results from 259 breast cancer cases were compiled in a tertiary referral hospital (Methods). The cases were randomly selected from clinical health records and showed an average age at diagnosis of 55.6 years, 95% CI 54.0–57.1 years. The blood test data were those collected on the date closest to diagnosis: 6 patients had the blood test on the same date as their diagnosis, 40 were earlier (on average 40 days before), and 182 were later (on average 45 days later): the average time between the blood test and disease diagnosis was 23.9 days, 95% CI 16.3–31.5 days. A multivariate regression analysis including tumor stage and subtype revealed three

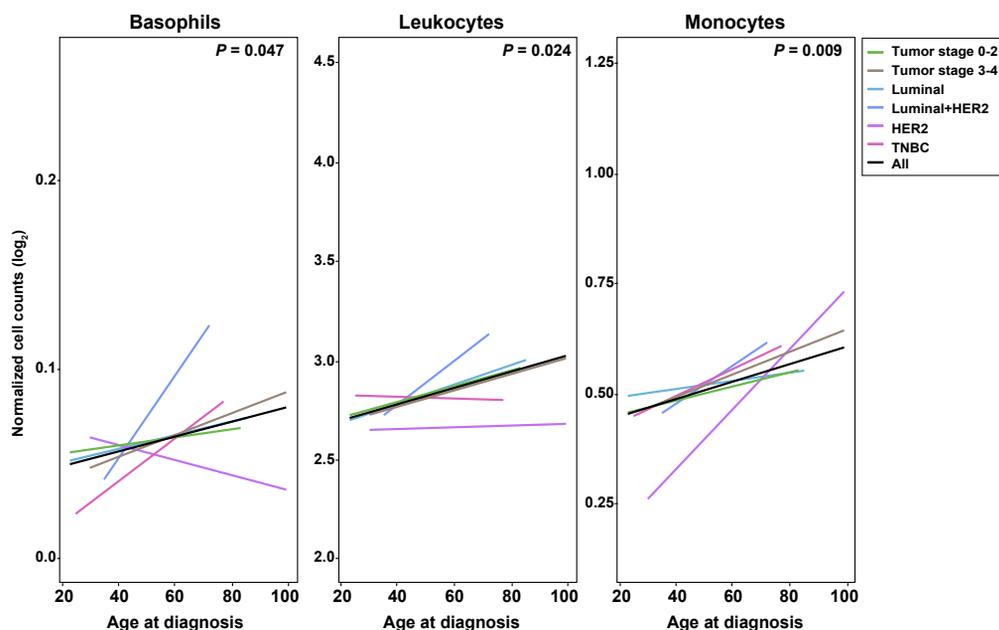


Figure 5. Positive Correlation between Peripheral Immune Cell Counts around Time of Diagnosis and Breast Cancer Age at Diagnosis

Positive correlations between basophil, leukocyte, and monocyte blood counts and age of diagnosis of breast cancer in a retrospective case-cohort hospital-based study. The value of p shown here is that associated with the coefficient calculated as part of the multivariate regression analysis. The trends for tumor subgroups are shown as depicted in the inset.

immune cell types to be significantly ($p < 0.05$) positively correlated with age at diagnosis: basophils, leukocytes, and monocytes (Figure 5). The trends were consistent for different tumor stages (0–2 and 3–4) and the major cancer subtype (i.e., luminal); the smaller patient sets of HER2-positive breast cancer ($n = 18$) and TNBC ($n = 17$) showed greater variability (Figure 5). The neutrophil to lymphocyte ratio, which is an established rate associated with breast cancer prognosis (Ethier et al., 2017), was not found to be associated with age at diagnosis in this study ($p = 0.65$).

DISCUSSION

The results of this study support the idea that the risk of certain cancers is influenced by the content of immune/stromal cells in the target tissue and/or by differences in peripheral immune cell counts. Of the 17 cancer settings analyzed, 57 risk loci comprising 13 cancer types were associated with differences in immune/stromal cell content with respect to the corresponding normal tissue and/or primary tumors. The gene candidates linked to these associations include several with key functions in the immune system, and they show significant enrichments in immune-related regulatory features and expression profiles. Detection of associations between immune/stromal cell signatures and PRSs provide further evidence that differences in these cell contents influence cancer risk. Nevertheless, the role of some cell types is multifaceted; for example, endothelial cells can regulate trafficking and activation of immune cells in a given tissue, but, critically, also determine angiogenesis (Hendry et al., 2016). Similarly, a given genetic variant may influence the expression of more than one gene target and/or indirectly alter the immune system by different mechanisms, such as by provoking oncogenic-induced signals.

Unexpectedly, there appear to be opposing cancer risk effects for immune cell contents across cancer types. These might be due to differences in tissue microenvironment conditions, such as inflammation caused by smoking or other factors (Shalapour and Karin, 2015). However, the study was limited by the relatively low numbers of normal tissue samples available for analysis, and, potentially, by gene expression alterations in normal tissue adjacent to neoplasms. This study had more power to detect significant results in normal breast tissue and BRCA and, consequently, the results prove to be more relevant and coherent in

these settings. Carrying out similar analyses in other normal and cancer tissue contexts would appear to be worthwhile. Such additional studies would benefit from complementary molecular marker and signaling analyses, which would definitively establish the functional consequences of inferred cell alterations.

At the same time as providing insight into the biological basis of cancer initiation, this study yields data that could be useful for analyses of cancer risk and prevention. Associations of PRSs with immune cell signatures could inform preventive strategies by modulating specific cell functions and/or their signaling molecules in individuals at high risk (Spira et al., 2017). This idea is particularly relevant in breast cancer. Our study shows consistent associations between immune/stromal cell signatures and breast cancer PRSs or age at diagnosis in normal tissue. A recognized risk locus connects differences in most peripheral immune cell types to breast cancer risk. This locus harbors the *SH2B3* gene, which is altered in hematological neoplasms and autoimmune diseases (Maslah et al., 2017, p. 3). Common genetic variation at this locus has been linked to cancer pleiotropy, including breast cancer susceptibility (Hung et al., 2015; Fehring et al., 2016). We extend these observations by identifying potential associations with breast and ovarian cancer risk in *BRCA1/2* mutation carriers. Our study shows consistent expression correlations of *SH2B3* or *SH2B3* functionally related genes with age at diagnosis using normal breast tissue data. Thus, pharmacological enhancement of *SH2B3* function might reduce cancer risk in individuals with high PRSs and/or carriers of *BRCA1/2* mutations. However, the functional impact on *SH2B3* remains to be established, and, therefore, prospective studies determining the expression and/or functional differences of *SH2B3* among individuals with specific alleles in the corresponding locus, and their associations with peripheral immune cell counts and cancer risk, are needed.

The effect of systemic differences of immune cell counts on breast cancer risk is further supported by unexpected associations between basophil, leukocyte, and monocyte blood counts and age at diagnosis from a retrospective case-cohort study. Relatively low monocyte counts collected over a 1-year period of disease diagnosis have recently been associated with increased breast cancer risk (Kresovich et al., 2020). However, high baseline leukocyte counts in a prospective study of postmenopausal women were found to be associated with increased breast cancer incidence (Margolis et al., 2007). In our study, we aimed to assess whether individuals' status of having relatively low peripheral immune cell counts was associated with initial cancer development, hypothetically due to reduced immunosurveillance. Our results are consistent with this explanation, and among other factors, altered *SH2B3* function might give rise to these observations. As a whole, the results of this study may be useful for improving cancer risk estimation, and for identifying preventive approaches.

Limitations of the Study

The present report identifies cancer-associated genetic variants and polygenic risk scores linked to the alteration of immune and/or stromal cell systemic and/or tissue contents. These links could explain the greater cancer risk. However, the study has several limitations that should be borne in mind. The cell content inferences were based on gene expression profiles, and therefore, molecular and cellular analyses are required to corroborate them and accurately assess their functional consequences. The observed associations could also be indirect in some instances. The study was also limited by the original sample sets, and observed associations could be confounded by other factors, such as the level of tissue inflammation, individual hormonal status, and lifestyle aspects. The genetic basis of the proposed associations between blood cell count and age at breast cancer diagnosis in the studied cohort remains unknown, and it is unclear whether similar associations exist in *BRCA1/2* mutation carriers.

Resource Availability

Lead Contact

Further information and requests for resources and data should be directed to and fulfilled by the Lead Contact Miquel Angel Pujana (mapujana@iconcologia.net).

Materials Availability

No materials were generated.

Data and Code Availability

TCGA data were obtained from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>) and from the corresponding consortium publications. Individual genetic data were obtained following

specific approval: dbGaP Data Access Committee project #11689. The R software algorithms developed by others and applied in this study are detailed in the [Transparent Methods supplemental file](#). A complete pipeline was implemented and is available at <https://github.com/pujana-lab/systematicQTL/>.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101296>.

ACKNOWLEDGMENTS

The results presented here are partly based on data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>), and we would like to express our gratitude to the TCGA consortia and their coordinators for providing the data and clinical information. This study was supported by the following patient foundations in Catalonia: DACMA, GINKGO, Sociathlon, and “Viladecans Contra el Càncer”. The work was also supported by grants from the Carlos III Institute of Health funded by FEDER funds – a way to build Europe – (ISCIII; Ministry of Science, Innovation and Universities; PI16/00563, PI18/01029, PI19/00553, and CIBERONC), the Generalitat de Catalunya (SGR 2017-449 and 2017-1282; and PERIS MedPerCan and URD-Cat), and the CERCA Program.

AUTHOR CONTRIBUTIONS

L.P. and M.A.P. designed and performed the experiments. I.G.-F. and R.d.C. performed variant imputation in TCGA. L.P., R.E., E.B. and M.G.-V. performed bioinformatic analyses of genetic variants, PRSs, gene expression, and cell signatures. D.R.B. and A.C.A. performed association studies in *BRCA1/2* mutation carriers. J. Beesley and G.C.-T. contributed to the association studies and performed candidate target gene analyses. M.G.-G., C.F., A.S., and A.I.E. contributed to the cohort study. D.O., A.R.-L., and C.V. contributed to the data normalization and multivariate regression designs. M.P.-C. and J.D. contributed to the pathway analyses. C.H., F.M., E.M., C.R., T.M., J. Brunet, R.H., and C.G. contributed data, evaluated the results, and critically revised the manuscript. C.L. contributed to the study design. M.A.P. wrote the paper.

DECLARATION OF INTERESTS

M.A.P. is recipient of an unrestricted research grant from Roche Pharma for the development of the PROCURE ICO research program. C.F. received support from Pfizer unrelated to this study.

Received: March 18, 2020

Revised: May 23, 2020

Accepted: June 15, 2020

Published: July 24, 2020

REFERENCES

- Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., et al. (2016). The allelic landscape of human blood cell Trait variation and links to common complex disease. *Cell* 167, 1415–1429.e19.
- Avila Cobos, F., Vandesompele, J., Mestdagh, P., and De Preter, K. (2018). Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* 34, 1969–1979.
- Burnet, F.M. (1970). The concept of immunological surveillance. *Prog. Exp. Tumor Res.* 13, 1–27.
- Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45, 1113–1120.
- Chen, D., Juko-Pecirep, I., Hammer, J., Ivansson, E., Enroth, S., Gustavsson, I., Feuk, L., Magnusson, P.K.E., McKay, J.D., Wilander, E., and Gyllensten, U. (2013). Genome-wide association study of susceptibility loci for cervical cancer. *J. Natl. Cancer Inst.* 105, 624–633.
- Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., Mateo, F., Herranz, C., Carbonell-Caballero, J., and Dopazo, J. (2018). Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. *Cancer Res.* 78, 6059–6072.
- Dauphinee, S.M., Clayton, A., Hussainkhel, A., Yang, C., Park, Y.-J., Fuller, M.E., Blonder, J., Veenstra, T.D., and Karsan, A. (2013). SASH1 is a scaffold molecule in endothelial TLR4 signaling. *J. Immunol.* 191, 892–901.
- Ethier, J.-L., Desautels, D., Templeton, A., Shah, P.S., and Amir, E. (2017). Prognostic role of neutrophil-to-lymphocyte ratio in breast cancer: a systematic review and meta-analysis. *Breast Cancer Res.* 19, 2.
- Fehringer, G., Kraft, P., Pharoah, P.D., Eeles, R.A., Chatterjee, N., Schumacher, F.R., Schildkraut, J.M., Lindström, S., Brennan, P., Bickeböller, H., et al. (2016). Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations. *Cancer Res.* 76, 5103–5114.
- Fritsche, L.G., Gruber, S.B., Wu, Z., Schmidt, E.M., Zawistowski, M., Moser, S.E., Blanc, V.M.,

- Brummett, C.M., Khetarpal, S., Abecasis, G.R., and Mukherjee, B. (2018). Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from the Michigan Genomics Initiative. *Am. J. Hum. Genet.* **102**, 1048–1061.
- Gonzalez, H., Hagerling, C., and Werb, Z. (2018). Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes Dev.* **32**, 1267–1284.
- GTEX Consortium (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585.
- Guerra, N., Tan, Y.X., Joncker, N.T., Choy, A., Gallardo, F., Xiong, N., Knoblauch, S., Cado, D., Greenberg, N.M., Greenberg, N.R., and Raulet, D.H. (2008). NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. *Immunity* **28**, 571–580.
- Hendry, S.A., Farnsworth, R.H., Solomon, B., Achen, M.G., Stacker, S.A., and Fox, S.B. (2016). The role of the tumor vasculature in the host immune response: implications for therapeutic strategies targeting the tumor microenvironment. *Front. Immunol.* **7**, 621.
- Huan, T., Meng, Q., Saleh, M.A., Norlander, A.E., Joehanes, R., Zhu, J., Chen, B.H., Zhang, B., Johnson, A.D., Ying, S., et al. (2015). Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol. Syst. Biol.* **11**, 799.
- Hung, R.J., Ulrich, C.M., Goode, E.L., Brhane, Y., Muir, K., Chan, A.T., Marchand, L.L., Schildkraut, J., Witte, J.S., Eeles, R., et al. (2015). Cross cancer genomic investigation of inflammation pathway for five common cancers: lung, ovary, prostate, breast, and colorectal cancer. *J. Natl. Cancer Inst.* **107**, djv246.
- Jabara, H.H., Boyden, S.E., Chou, J., Ramesh, N., Massaad, M.J., Benson, H., Bainter, W., Fraulino, D., Rahimov, F., Sieff, C., et al. (2016). A missense mutation in TFRC, encoding transferrin receptor 1, causes combined immunodeficiency. *Nat. Genet.* **48**, 74–78.
- Jiménez-Sánchez, A., Cast, O., and Miller, M.L. (2019). Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res.* **79**, 6238–6246.
- Jonsson, S., Sveinbjornsson, G., de Lapuente Portilla, A.L., Swaminathan, B., Plomp, R., Dekkers, G., Ajore, R., Ali, M., Bentlage, A.E.H., Elmér, E., et al. (2017). Identification of sequence variants influencing immunoglobulin levels. *Nat. Genet.* **49**, 1182–1191.
- Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.-J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet.* **13**, e1006643.
- Klemann, C., Wagner, L., Stephan, M., and von Hörsten, S. (2016). Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin. Exp. Immunol.* **185**, 1–21.
- van Kooten, C., Fiore, N., Trouw, L.A., Csomor, E., Xu, W., Castellano, G., Daha, M.R., and Gelderman, K.A. (2008). Complement production and regulation by dendritic cells: molecular switches between tolerance and immunity. *Mol. Immunol.* **45**, 4064–4072.
- Kresovich, J.K., O'Brien, K.M., Xu, Z., Weinberg, C.R., Sandler, D.P., and Taylor, J.A. (2020). Prediagnostic immune cell profiles and breast cancer. *JAMA Netw. Open* **3**, e1919536.
- Li, Y., He, X., Schembri-King, J., Jakes, S., and Hayashi, J. (2000). Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation. *J. Immunol.* **164**, 5199–5206.
- Lim, Y.W., Chen-Harris, H., Mayba, O., Lianoglou, S., Wuster, A., Bhangale, T., Khan, Z., Mariathasan, S., Daemen, A., Reeder, J., et al. (2018). Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proc. Natl. Acad. Sci. U S A* **115**, E11701–E11710.
- Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416.e11.
- Margolis, K.L., Rodabough, R.J., Thomson, C.A., Lopez, A.M., and McTiernan, A.; Women's Health Initiative Research Group (2007). Prospective study of leukocyte count as a predictor of incident breast, colorectal, endometrial, and lung cancer and mortality in postmenopausal women. *Arch. Intern. Med.* **167**, 1837–1844.
- Maslah, N., Cassinat, B., Verger, E., Kiladjian, J.-J., and Velazquez, L. (2017). The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. *Leukemia* **31**, 1661–1670.
- Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94.
- Mortaz, E., Tabarsi, P., Mansouri, D., Khosravi, A., Garsen, J., Velayati, A., and Adcock, I.M. (2016). Cancers related to immunodeficiencies: update and perspectives. *Front. Immunol.* **7**, 365.
- Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782.
- Pabst, O., Zweigerdt, R., and Arnold, H.H. (1999). Targeted disruption of the homeobox transcription factor Nkx2-3 in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development* **126**, 2215–2225.
- Ribatti, D. (2017). The concept of immune surveillance against tumors. The first theories. *Oncotarget* **8**, 7175–7180.
- Shaloupour, S., and Karin, M. (2015). Immunity, inflammation, and cancer: an eternal fight between good and evil. *J. Clin. Invest.* **125**, 3347–3355.
- Spira, A., Yurgelun, M.B., Alexandrov, L., Rao, A., Bejar, R., Polyak, K., Giannakis, M., Shilatifard, A., Finn, O.J., Dhodapkar, M., et al. (2017). Precancer atlas to drive precision prevention trials. *Cancer Res.* **77**, 1510–1541.
- Stenina-Adognravi, O. (2014). Invoking the power of thrombospondins: regulation of thrombospondins expression. *Matrix Biol.* **37**, 69–82.
- Taylor, A.M., Shih, J., Ha, G., Gao, G.F., Zhang, X., Berger, A.C., Schumacher, S.E., Wang, C., Hu, H., Liu, J., et al. (2018). Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e3.
- Terunuma, A., Putluri, N., Mishra, P., Mathé, E.A., Dorsey, T.H., Yi, M., Wallace, T.A., Issaq, H.J., Zhou, M., Killian, J.K., et al. (2014). MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. *J. Clin. Invest.* **124**, 398–412.
- Thomas, L. (1982). On immunosurveillance in human cancer. *Yale J. Biol. Med.* **55**, 329–333.
- Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. *Immunity* **48**, 812–830.e14.
- Torkamani, A., Wineinger, N.E., and Topol, E.J. (2018). The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590.
- Tsoi, L.C., Stuart, P.E., Tian, C., Gudjonsson, J.E., Das, S., Zawistowski, M., Ellinghaus, E., Barker, J.N., Chandran, V., Dand, N., et al. (2017). Large scale meta-analysis characterizes genetic architecture for common psoriasis associated variants. *Nat. Commun.* **8**, 15382.
- Velazquez, L., Cheng, A.M., Fleming, H.E., Furlonger, C., Vesely, S., Bernstein, A., Paige, C.J., and Pawson, T. (2002). Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. *J. Exp. Med.* **195**, 1599–1611.
- Vinay, D.S., Ryan, E.P., Pawelec, G., Talib, W.H., Stagg, J., Elkord, E., Lichtor, T., Decker, W.K., Whelan, R.L., Kumara, H.M.C.S., et al. (2015). Immune evasion in cancer: mechanistic basis and therapeutic strategies. *Semin. Cancer Biol.* **35** Suppl, S185–S198.
- Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934.
- Webb, G.J., and Hirschfield, G.M. (2016). Using GWAS to identify genetic predisposition in hepatic autoimmunity. *J. Autoimmun.* **66**, 25–39.
- Zanotti, K.J., Maul, R.W., Castiblanco, D.P., Yang, W., Choi, Y.J., Fox, J.T., Myung, K., Saribasak, H., and Gearhart, P.J. (2015). ATAD5 deficiency decreases B cell division and Igh recombination. *J. Immunol.* **194**, 35–42.

Supplemental Information

Immune Cell Associations with Cancer Risk

Luis Palomero, Ivan Galván-Femenía, Rafael de Cid, Roderic Espín, Daniel R. Barnes, CIMBA, Eline Blommaert, Miguel Gil-Gil, Catalina Falo, Agostina Stradella, Dan Ouchi, Albert Roso-Llorach, Concepció Violan, María Peña-Chilet, Joaquín Dopazo, Ana Isabel Extremera, Mar García-Valero, Carmen Herranz, Francesca Mateo, Elisabetta Mereu, Jonathan Beesley, Georgia Chenevix-Trench, Cecilia Roux, Tak Mak, Joan Brunet, Razq Hakem, Chiara Gorrini, Antonis C. Antoniou, Conxi Lázaro, and Miquel Angel Pujana

Supplementary figures legends

Fig. S1. Evaluation of immune/stromal cell tissue content estimates in relation to two other methods. Related to Figure 1.

(A) Heatmap showing the correlations (Spearman's ρ) between Consensus^{TME}-based values and analogous TIMER cell type estimates.

(B) Heatmap showing the correlations (Spearman's ρ) between Consensus^{TME}-based values and analogous MCP-counter cell type estimates.

Fig. S2. Evaluation of immune/stromal cell tissue content estimates in relation to independent leukocyte estimates. Related to Figure 1. Heatmap showing the correlations (Spearman's ρ) between Consensus^{TME}-based values and independent leukocyte estimates using the approach of Taylor et al. (2018).

Fig. S3. Evaluation of immune/stromal cell tissue content estimates in relation to aneuploidy scores. Related to Figure 1.

(A) Heatmap showing the correlations (Spearman's ρ) between Consensus^{TME}-based values and aneuploidy scores (Taylor et al., 2018) across major cancer types.

(B) Heatmap showing the correlations (Spearman's ρ) between Consensus^{TME}-based values and aneuploidy scores (Taylor et al., 2018) across in BRCA subtypes, which show positive correlations in claudin-low.

Fig. S4. Differences of inferred immune/stromal cell content between primary tumors with low and high levels of *CD274/PDL1* expression.

Related to Figure 1. The graphs show the median cell content value in each group and the significance of the difference (Wilcoxon test *P* value).

Fig. S5. Differences of inferred immune/stromal cell content between primary tumors with low and high levels of *CD279/PDCD1* expression.

Related to Figure 1. The graphs show the median cell content value in each group and the significance of the difference (Wilcoxon test *P* value).

Fig. S6. Correlations between inferred blood immune cell contents and measures from fluorescence-activated cell sorting in blood samples.

Related to Figure 1. Forest plot showing correlation estimates and 95% CIs of each inferred cell type (data from whole blood samples of healthy adults; $n = 12$, GEO GSE127813).

Fig. S7. Correlations between inferred immune/stromal cell tissue contents and single cells used to generate pseudo-bulk breast tumors.

Related to Figure 1. Each panel shows the correlation between immune cell signature scores (Y-axis) and the number of cells (X-axis) used to generate 100 pseudo-bulk breast tumors (data from Gene Expression Omnibus reference GSE75688). The trend lines, Spearman's correlations and *P* values are shown.

Fig. S8. Correlations between immune/stromal cell tissue contents and expression of immune benchmark genes. Related to Figure 1. Top panel, distribution of PCCs using data from normal TCGA tissue. Bottom panel,

distribution of PCCs using data from primary tumors of TCGA. Mean PCCs and 95% CIs are shown.

Fig. S9. Correlations between immune cell signatures and pathway signaling-inferred activities. Related to Figure 1. Unsupervised clustering of the correlation coefficients between inferred cell contents (Y-axis) and KEGG pathway activities (X-axis). Differentiated clusters in normal tissue are marked by red-outlined rectangles.

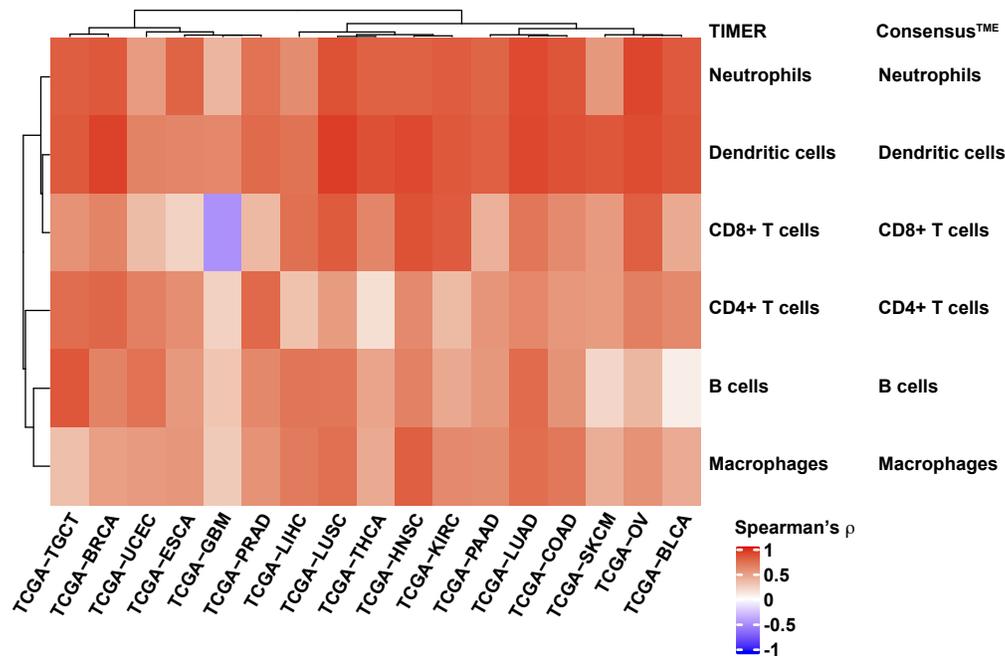
Fig. S10. Gene targets of eQTL recognized in isQTLs are frequently correlated with the corresponding immune/stromal cell signatures. Related to Figure 2. Distributions of random gene sets (same gene set size and equivalent comparisons for each signature and TCGA setting) relative to the number of significant correlations between eQTL-target and immune/stromal signatures. Left- and right-hand panels show results for the first and second isQTL sets presented in the main text, respectively. Empirical test probabilities are shown.

Fig. S11. Minimal correlation estimates to detect significant signature-PRS associations. Related to Figure 3. Left and right panels show the lowest correlations required in each normal and primary tumor setting, respectively, to detect nominal ($P < 0.05$) associations given the TCGA sample sizes.

Fig. S12. LUAD and LUSC PRS correlations with NK cell content. Related to Figure 3. Top panels, positive correlations between NK cell content in primary tumors of LUAD and LUSC, and the corresponding PRSs. The adjusted- R^2 and P values of the linear regression model are shown. Bottom panels, correlation trends of patients stratified by smoking status, as depicted in the insets. The estimate for LUAD cases classified as current smokers was found to be significantly less than zero ($r = -0.12$, $P = 0.012$).

Figure S1

A



B

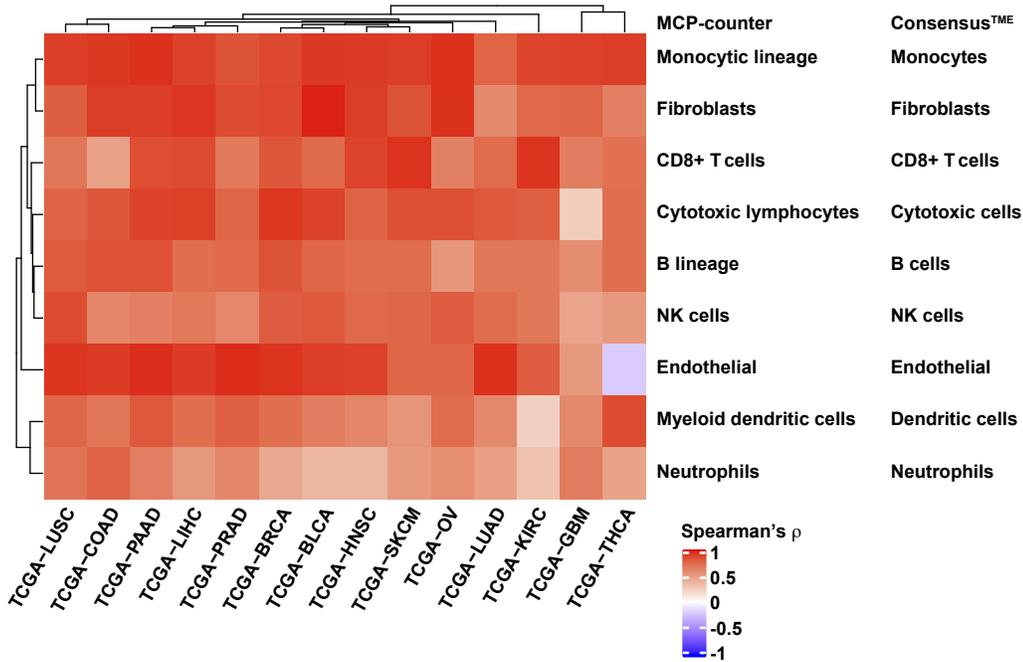


Figure S2

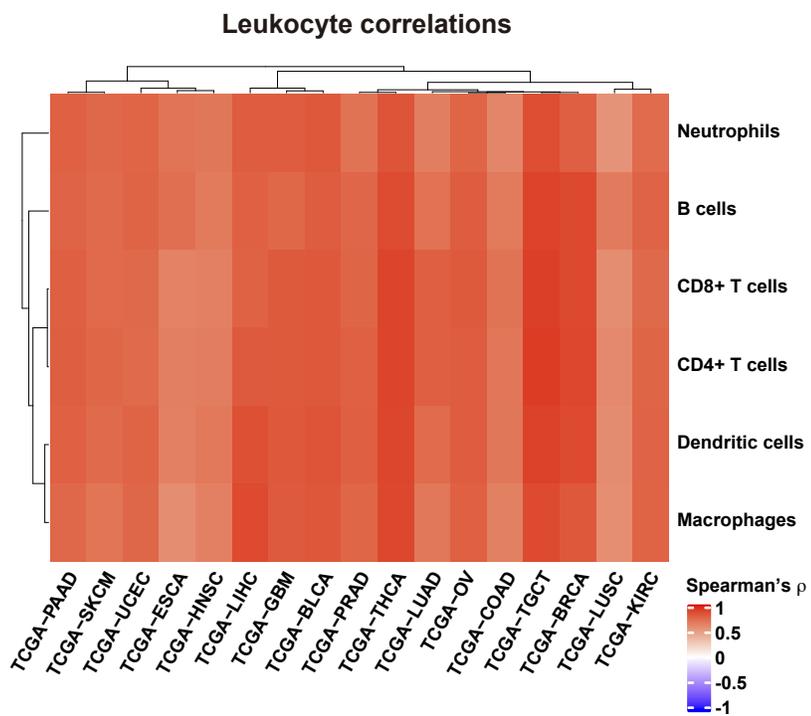
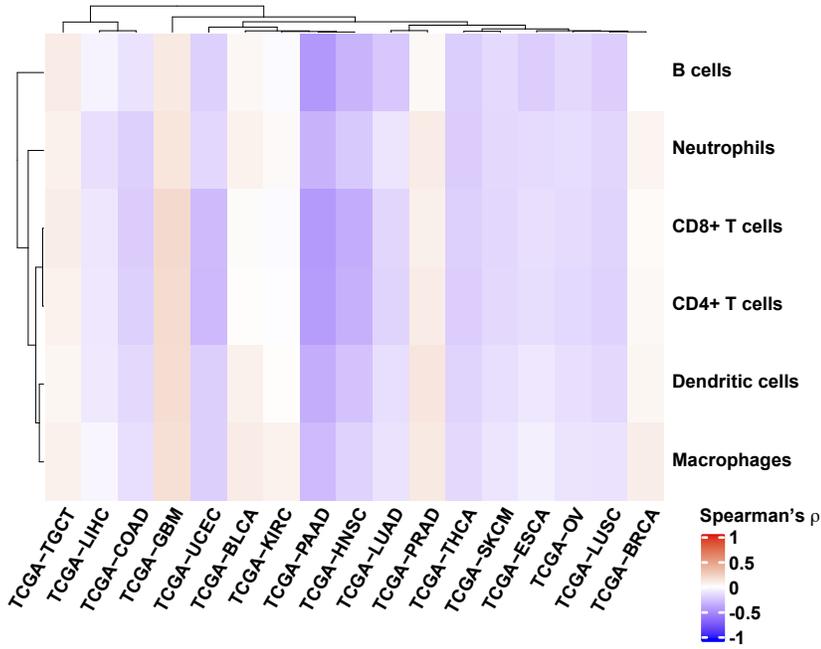


Figure S3

A

Aneuploidy correlations



B

Aneuploidy correlations

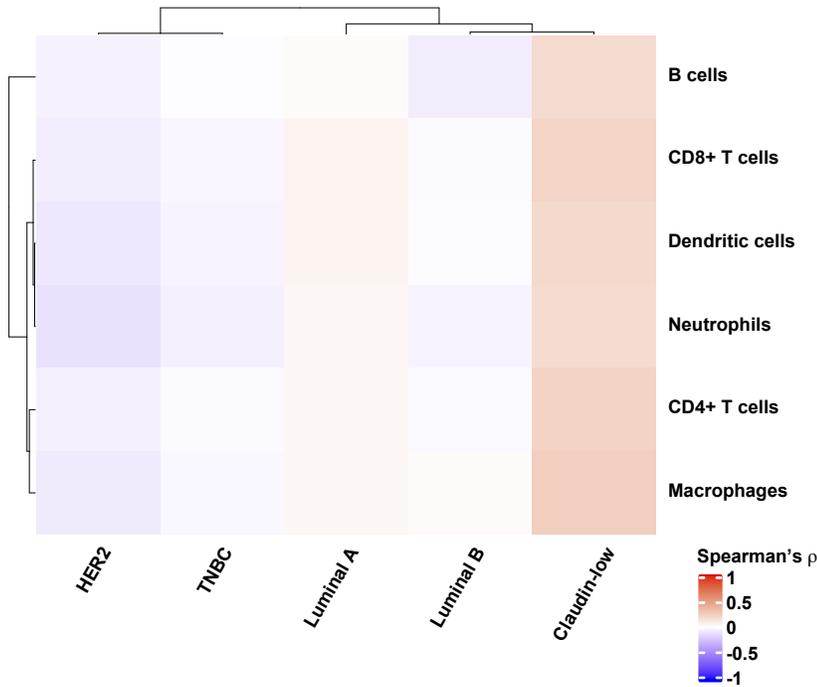


Figure S4

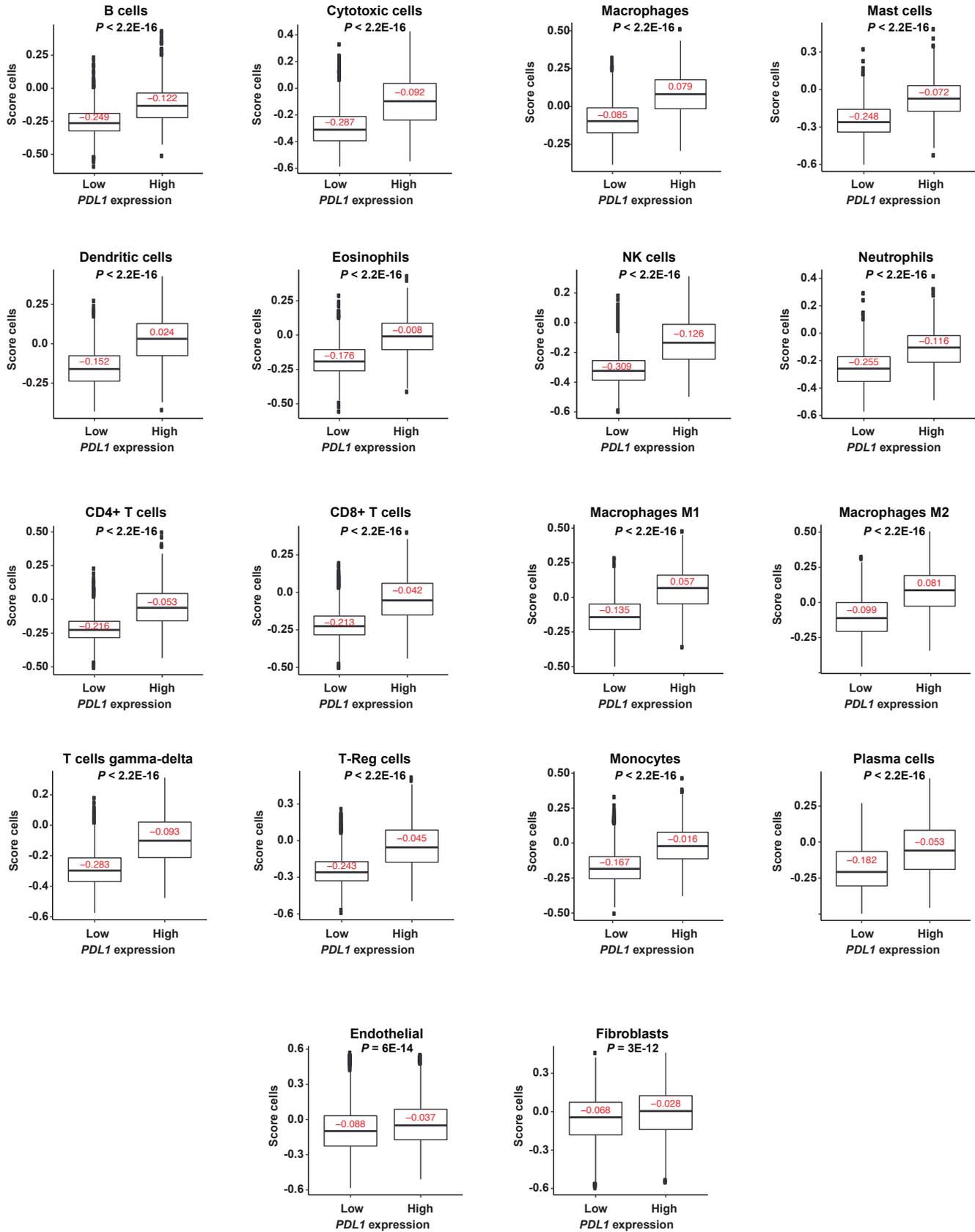


Figure S5

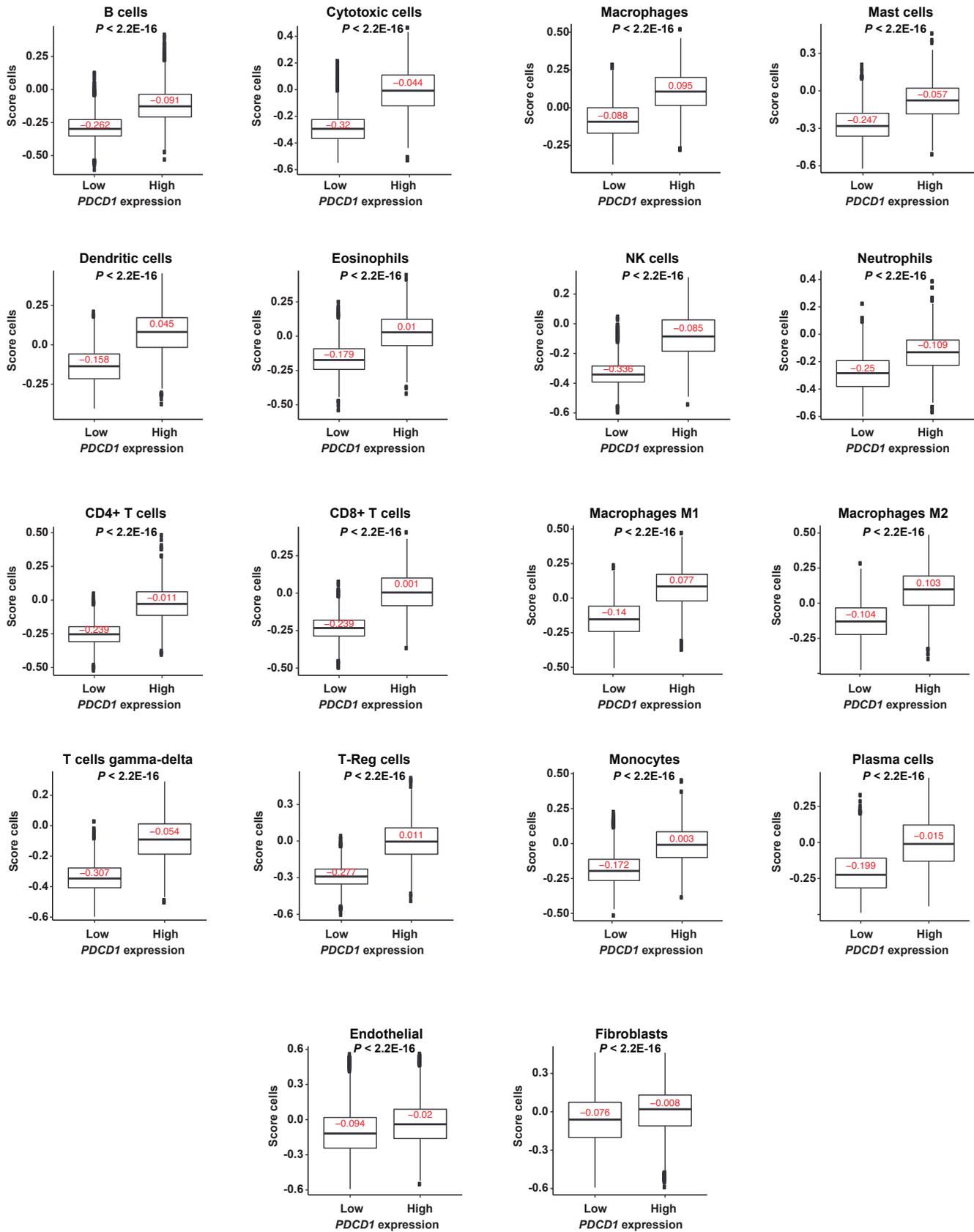


Figure S6

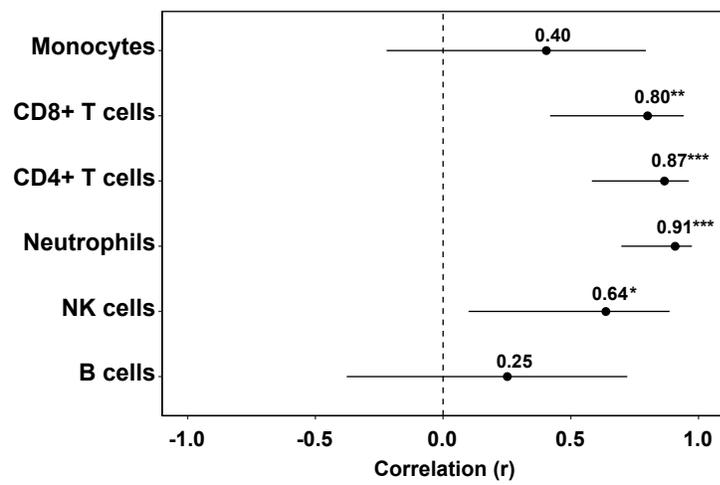


Figure S7

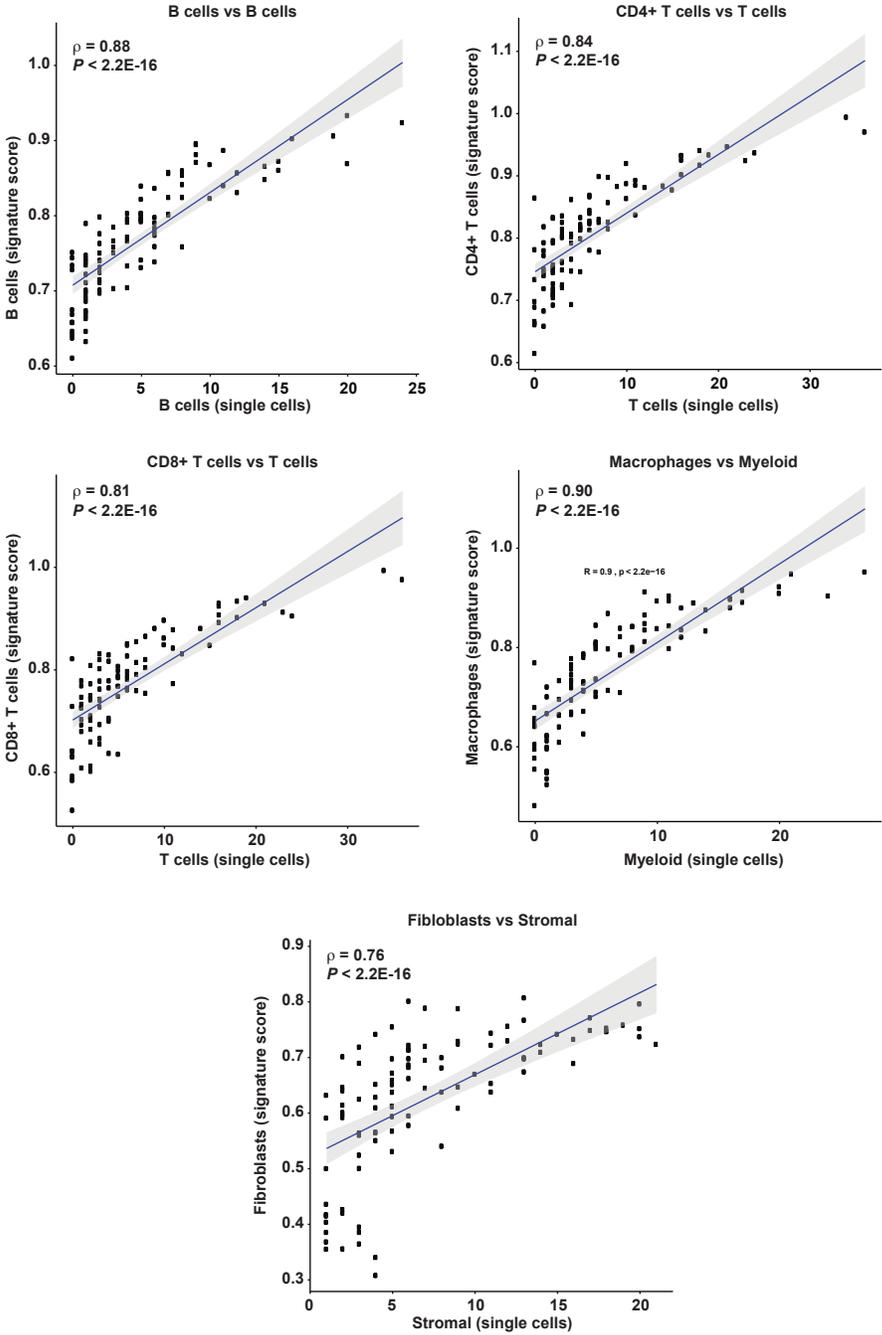


Figure S8

**Immune cell tissue content correlations
with defined immune gene benchmarks**

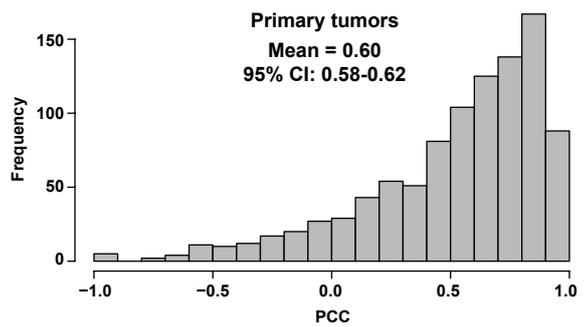
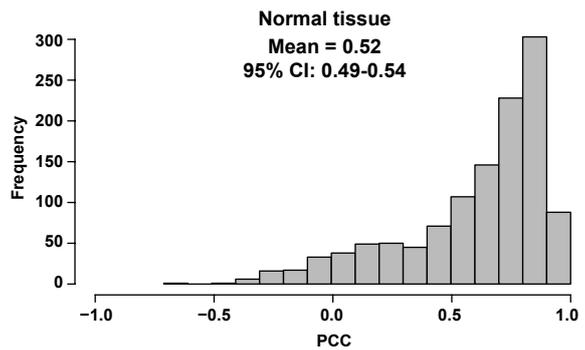


Figure S9

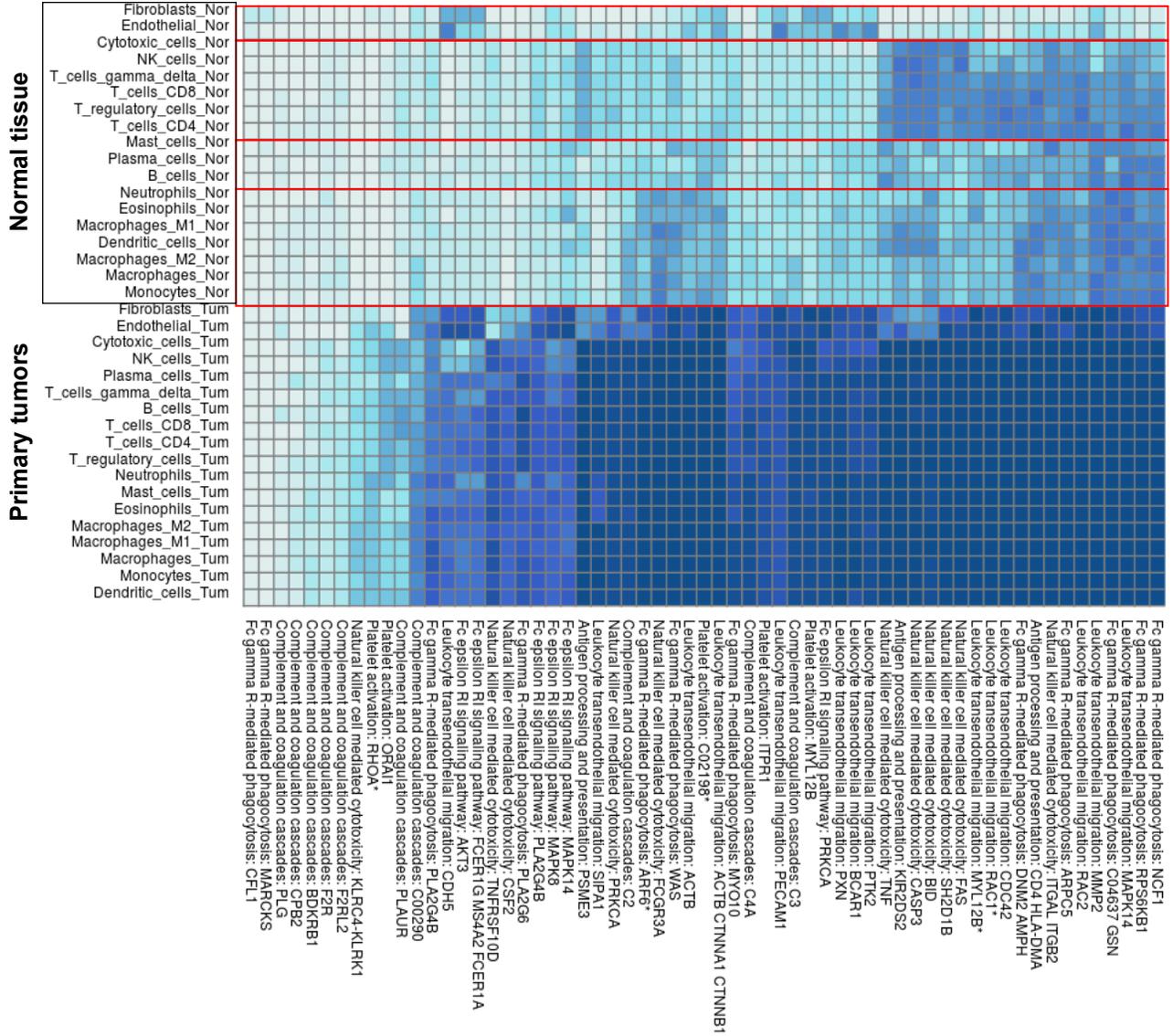


Figure S10

eQTL-gene target correlations with immune/stromal cell signatures

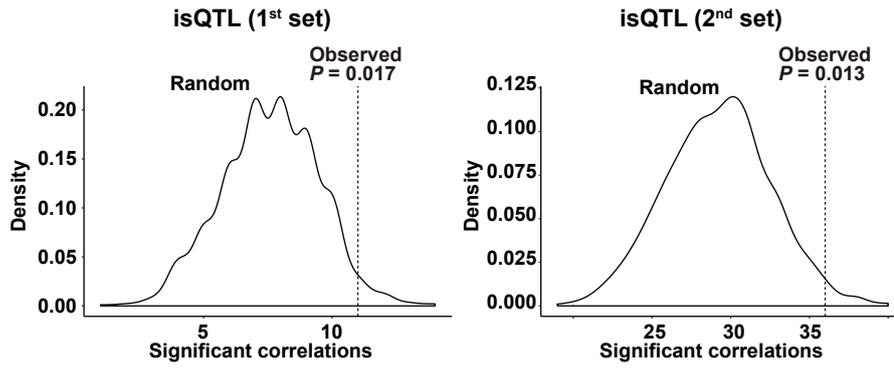


Figure S11

Minimal correlation value to detect a significant PRS-cell signature associations ($P < 0.05$)

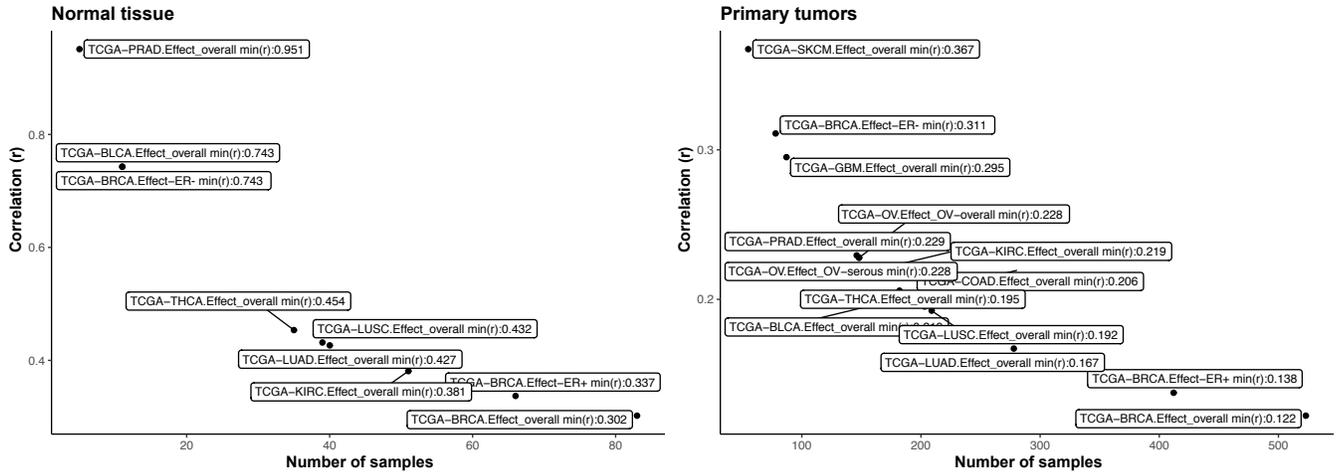
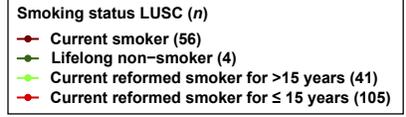
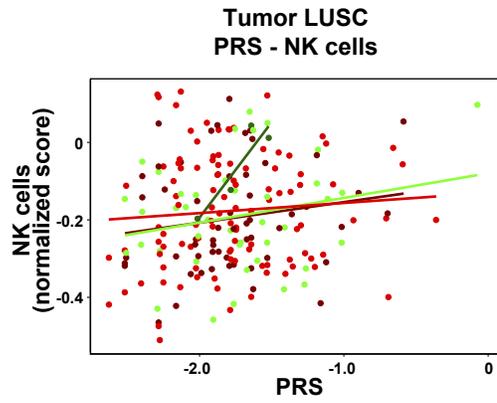
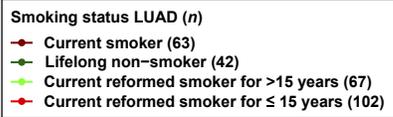
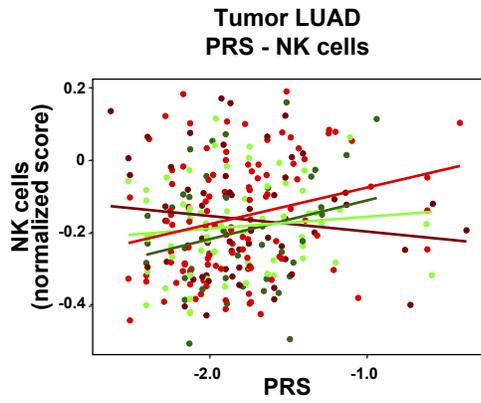
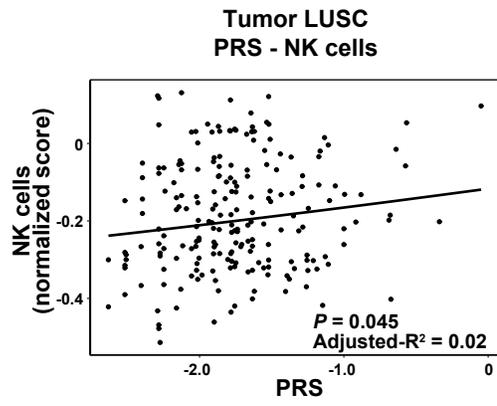
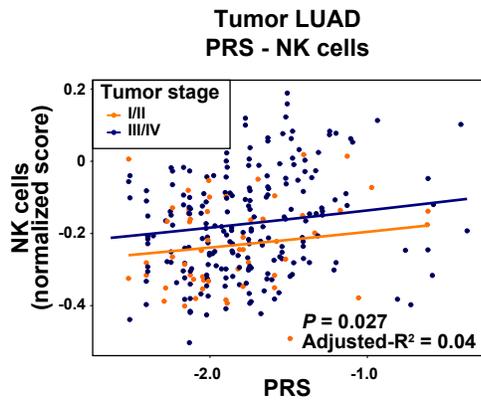


Figure S12



Transparent Methods

TCGA data

Clinical and gene expression (RNA-seq fragments per kilobase of transcript per million mapped reads (FPKM) upper quartile normalized (UQ)) data from The Cancer Genome Atlas (TCGA) projects were obtained from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>) and from the corresponding publications. Genetic data at the individual level were obtained following approval by the dbGaP Data Access Committee (project #11689). Metastases and recurrent tumors were excluded from this study, making normal tissue (blood or solid tissue) and primary tumor samples the focus of the analyses. The cancer types are named using the corresponding TCGA study abbreviations (<https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations>). For normal tissue, according to the TCGA protocols, these samples were collected > 2 cm from the tumor margin and/or did not contain tumor identified by histopathological review. The protein expression measures of CD26 and TFCR corresponded to those obtained by TCGA using reverse-phase protein arrays (RPPAs; level 4 data, <https://tcpaportal.org/tcpa/>). The COAD subtypes were defined based genomic/genetic alterations (chromosomal instability (CIN), genomic stable (GS), and microsatellite instability (MSI) tumors) and on molecular features (consensus molecular subtypes, CMS1-4) (Guinney et al., 2015).

Cancer risk variants

The variants were compiled from the GWAS Central (Beck et al., 2014) and GWAS Catalog (Buniello et al., 2019) databases, and by literature searches using the PubMed MeSH terms “GWAS”, “association”, “cancer”, and “risk”. The variants are listed in **Table S1**. The UK Biobank GWAS results were taken from the public repository at <http://www.nealelab.is/uk-biobank>.

Benchmark immune genes

These genes were compiled from The Immunological Genome Project (ImmGen) (Shay and Kang, 2013) and CellMarker (Zhang et al., 2019) databases, and by a literature search using the MeSH terms corresponding to the specific immune cell types represented by the gene expression signatures. The benchmarks and their cell type assignments are included in **Table S3**.

Genotype data and imputation

Bulk genotyping data corresponding to the Affymetrix Genome-Wide Human SNP 6.0 Array were downloaded from the TCGA legacy archive (<https://gdc-portal.nci.nih.gov/legacy-archive/>). Of the initial normal tissue and primary tumor samples (n = 16,599), those corresponding to individuals of self-reported non-white origin (n = 4,770), and those of non-European origin based on principal component analysis using variants intersected in the 1000 Genome Project phase III (n = 2,598) were excluded from subsequent analyses; these filters were applied because summary

statistics of the GWASs used in this study are strongly biased towards populations of European origin. Normal and tumor samples were then examined separately for duplicates and up to third-degree relatives (kinship cutoff = 0.05), which resulted in the exclusion of an additional 672 samples. In the joint dataset, 765 samples were also excluded because they showed a gender mismatch in an analysis of pseudoautosomal genomic regions. Considering genetic variants, 108 samples that deviated by four or more standard deviations from the mean heterozygosity rate were also excluded. For imputation, variants were excluded if they fulfilled any of the following criteria: they mapped to chromosome Y, pseudoautosomal regions or the mitochondrial genome; they had a call rate < 100%; their minor allele frequency was < 0.01; they departed from Hardy–Weinberg equilibrium ($P < 5 \times 10^{-6}$); or they mapped to AT-CG sites. Finally, 7,686 samples (4,154 normal, comprising 3,287 blood-derived and 867 solid-tissue samples; and 3,532 primary tumors, of which 94.4% were paired) and 589,101 variants were retained for subsequent analyses. Imputation was performed using the Shape-IT V2 (Delaneau et al., 2008) and IMPUTE2 (Howie et al., 2009) algorithms, and the 1000 Genome Project Phase III panel as reference. Poorly imputed variants (accuracy score < 0.7) were excluded from subsequent analyses. A standard cutoff dose was applied to calculate genotypes using a hard-call threshold of 0.1 (i.e., 0 – 0.1, 0.9 – 1.1, 1.9 – 2.0 for reference homozygote, heterozygous and alternative homozygous genotypes, respectively).

Immune/stromal cell signatures

Immune/stromal cell gene expression signatures for each TCGA cancer setting were computed using the Consensus^{TME} method (Jiménez-Sánchez et al., 2019), which was provided available as an R package (<https://github.com/cansysbio/ConsensusTME>). Ten single-cell breast cancer signatures (Azizi et al., 2018) were included in the TCGA BRCA analyses. Therefore, 18 signatures were examined in each normal tissue and primary tumor setting, except for normal breast and breast cancer tissue, for which a total of 28 signatures were analyzed. The signature scores were computed using the single-sample Gene Set Expression Analysis (ssGSEA) algorithm calculated within the Gene Set Variation Analysis (GSVA) software (Hänzelmann et al., 2013). These scores were calculated for normal tissue and primary tumors, but not for blood samples, since data from blood are limited to germline genotypes. Genes whose expression was uninformative in more than half the samples in a given setting were excluded from the signature calculations; otherwise, missing data were assigned the average value of the informative samples. Evaluation of signature scores computed by two different methods—ssGSEA and summing normalized gene expression Z-scores—revealed global coherence, whereby Pearson correlation coefficients (PCCs) were > 0.80 in 99% (571/578) of the score comparisons. To select independent signatures in each normal and cancer setting, we performed a principal component analysis using the *prcomp* function in R. Components with eigenvalues > 1 were retained to study quantitative trait loci (subsequent sections). Estimates of immune-related pathway activities were calculated using directed graphs from the Kyoto Encyclopedia of Genes and Genomes

(KEGG, <https://www.genome.jp/kegg/>). Briefly, gene expression profiles were converted into pathway module activity scores by taking into account the chain of reactions from a defined molecular input to a specific molecular output (Cubuk et al., 2018). The 84-gene signature linked to SH2B3 included the genes differentially expressed in *Sh2b3*-null cells and that participate in genetic and/or protein interactions to this gene/protein (Huan et al., 2015); *SH2B3* was excluded from this signature for subsequent analyses.

Pseudo-bulk breast tumors

To generate 100 pseudo-bulk breast tumors, we used the single-cell RNA-seq data from the Gene Expression Omnibus (GEO) reference GSE75688 (Chung et al., 2017) and aggregated read counts using the `aggregateData` function in R (<https://github.com/HelenaLC/muscat>). Each simulated sample of 100 cells was forced to include >50% tumor cells (average 75.3%, 95% CI 72.53 – 77.93%). For non-tumoral cells, 10% of them were fixed as stromal (bulk average 7.22%, 95% CI 6.22 – 8.36%), while the other 90% were a random combination of B cells (average 5.16%, 95% CI 4.28 – 6.28%), T cells (average 6.21%, 95% CI 5.05 – 7.48%), and myeloid cells (average 6.11%, 95% CI 5.05 – 7.39%). Most of the myeloid cells were originally assigned to macrophages (Chung et al., 2017).

Quantitative trait loci of immune/stromal cell tissue content

The germline genetic calls corresponded to genotype data obtained from blood or normal tissue samples. For cases with both types of sample, the variants with discordant

calls were excluded from subsequent analyses. As specified above, the somatic genetic calls corresponded to primary tumors only. The immune/stromal cell-content quantitative trait loci (isQTL) were analyzed using the *R/qtl2* package in R (Broman et al., 2019). These analyses included the covariates of gender (when informative), age at diagnosis, tumor stage and histology. The Haley–Knott regression method was used to compute the log odds (LOD) of the associations between genetic variants and immune/stromal cell scores. One thousand permutations were performed in each setting to obtain significance thresholds (Manichaikul et al., 2007) and the variant-signature associations with empirical values of $P < 0.05$ were considered significant isQTL. The gene targets were defined according to the genomic location of the identified variants. Additional targets were identified by analyzing all variants correlated ($r^2 > 0.8$, 1000 Genomes phase 3, version 5) with each isQTL and intersect them with various functional genomic data, including promoter capture Hi-C (Javierre et al., 2016), annotated enhancers (Hnisz et al., 2013, p.), and eQTL (Schmiedel et al., 2018) from B cells, monocytes, and CD4+ and CD8+ T cells. In addition, correlated variants were queried using the Ensembl Variant Effect Predictor (McLaren et al., 2016) for potential effects on protein coding sequences.

Computation of PRSs

The PRSs were compiled from the literature and computed by summing the products of the per-allele LOD ratio assigned to each risk variant, and the corresponding allele dosage, for the total number of variants initially defined for each PRS. There was no

previous evidence of significant interactions or deviations from a log-additive model in BRCA PRSs (Mavaddat et al., 2019), but it is not known for other cancers. In the analyses of BRCA, OV (no normal tissue data available), and PRAD PRSs, two sets were analyzed, both based on GWAS-identified variants: set #1 (hereafter PRSs-1), which corresponded to scores derived from large collections of GWAS cohorts and validated in independent studies (Mavaddat et al., 2019); and set #2 (hereafter PRSs-2), which corresponded to scores derived from a phenome-wide longitudinal study using electronic health records collected by the Michigan Genomics Initiative (Fritsche et al., 2018). In both sets, PRSs were developed for all BRCA patients, and separately for the estrogen receptor (ER)-positive and ER-negative subtypes. The number of initial variants in these BRCA PRSs and those included in our study, based on available genotypes and obtained imputations were 307 and 185 for PRSs-1, and 3,820 and 3,629 for PRSs-2. As expected, the PRSs from the two sets were found to be positively correlated using germline or primary tumor data: BRCA PRSs PCCs = 0.60 – 0.66, $P < 10^{-5}$; OV tumors PRSs PCC = 0.72, $P < 10^{-25}$ (serous PCC = 0.72); and PRAD PRSs PCCs = 0.23 – 0.99, $P < 0.01$. The Michigan Genomics Initiative also provided PRSs for seven other cancer types, and the number of variants originally included and analyzed in this study were, respectively: 103 and 21 for PRAD; 42 and 41 for COAD; 16 and 16 for BLCA and SKCM; 15 and 15 for OV; 9 and 9 for GBM, LUAD and LUSC; 8 and 7 for THCA; and 7 and 6 for KIRC.

Cell signature associations with PRSs

The *bestNormalize* package in R (<https://github.com/petersonR/bestNormalize>) was used to normalize the cell signature values. The transformation that produced the lowest value from the Pearson's statistic divided by the degrees of freedom was taken to indicate the best function. The error distributions of the models and Q-Q plots were examined individually. The parameters of each signature transformation are provided in **Table S10**. Outliers were identified using the interquartile range rule and excluded from subsequent analyses; these were < 5% in all settings. Normalized signature values were used as dependent variables in a linear regression analysis relative to the PRSs. Stepwise analyses including covariates of gender, tumor stage and histology were performed, and the best model was selected based on the Akaike information criterion (AIC). For normal tissue, only those studies with at least 50 informative samples were analyzed. The small number of samples in each setting meant that these analyses could only detect significant (nominal $P < 0.05$) correlation estimates > 0.27 and > 0.09 in normal breast tissue and BRCA, and stronger correlations would be required in all other settings if nominal significance were to be reached (**Fig. S11**). The significance of the associations was corrected for multiple testing using the false-discovery rate (FDR) method.

Cell signature associations with age at diagnosis

The associations between the cell signature scores (dependent variables) and age at diagnosis were evaluated by multiple linear regression, including gender and tumor

stage as covariates, the best model being determined from an AIC-based stepwise selection algorithm. The statistical significances of the associations were corrected for multiple testing separately in normal tissue and primary tumor analyses (since the expected effects were the opposite of what they proved to be) using the FDR method.

Breast cancer risk in *BRCA1/2*-mutation carriers

Analyses were performed using data from the OncoArray and Collaborative Oncological Gene-environment Study (iCOGS) consortiums with the participation of the Consortium of Investigators of *BRCA1/2* Modifiers (CIMBA). The OncoArray and iCOGS designs, quality controls, and statistical analyses have been described previously (Milne et al., 2017). Summary statistics from the retrospective likelihood method are reported.

Analysis of blood cell parameters and age at diagnosis of breast cancer

Clinical and histopathological data from breast cancer patients were compiled through manual curation of hospital records of the Catalan Institute of Oncology, L'Hospitalet del Llobregat (Barcelona, Catalonia, Spain). Patients were randomly selected from health records collected between 2009 and 2014. The compiled data included date of birth, age, gender (only women selected), date at diagnosis, tumor stage, subtype and/or ER status, and date at initial-diagnostic blood test. The blood test parameters analyzed were the normalized numbers ($\times 10^9/L$) of basophils, eosinophils, leucocytes, lymphocytes, monocytes, neutrophils, and platelets. Linear regressions of each of these parameters

on age at diagnosis, including tumor stage and subtype as covariates, were performed.

The IDIBELL's Research Ethics Committee approved this study (reference PR066/20).

Supplemental References

- Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., et al. (2018). Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* *174*, 1293-1308.e36.
- Beck, T., Hastings, R.K., Gollapudi, S., Free, R.C., Brookes, A.J. (2014). GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.* *22*, 949–952.
- Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Š., Yandell, B.S., Churchill, G.A. (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* *211*, 495–502.
- Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* *47*, D1005–D1012.
- Chung, W., Eum, H.H., Lee, H.-O., Lee, K.-M., Lee, H.-B., Kim, K.-T., Ryu, H.S., et al. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* *8*, 15081.
- Cubuk, C., Hidalgo, M.R., Amadoz, A., Pujana, M.A., Mateo, F., Herranz, C., Carbonell-Caballero, J., Dopazo, J. (2018). Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. *Cancer Res.* *78*, 6059–6072.
- Delaneau, O., Coulonges, C., Zagury, J.-F. (2008). Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics* *9*, 540.
- Fritsche, L.G., Gruber, S.B., Wu, Z., Schmidt, E.M., Zawistowski, M., Moser, S.E., Blanc, V.M., Brummett, C.M., Kheterpal, S., Abecasis, G.R., Mukherjee, B. (2018). Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from The Michigan Genomics Initiative. *Am. J. Hum. Genet.* *102*, 1048–1061.
- Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Sonesson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. *Nat. Med.* *21*, 1350–1356.
- Hänzelmann, S., Castelo, R., Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* *14*, 7.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* *155*, 934–947.
- Howie, B.N., Donnelly, P., Marchini, J. (2009). A flexible and accurate genotype

- imputation method for the next generation of genome-wide association studies. *PLoS Genet.* *5*, e1000529.
- Huan, T., Meng, Q., Saleh, M.A., Norlander, A.E., Joehanes, R., Zhu, J., Chen, B.H., Zhang, B., Johnson, A.D., Ying, S., et al. (2015). Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol. Syst. Biol.* *11*, 799.
- Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* *167*, 1369-1384.e19.
- Jiménez-Sánchez, A., Cast, O., Miller, M.L. (2019). Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res.* *79*, 6238–6246.
- Manichaikul, A., Palmer, A.A., Sen, S., Broman, K.W. (2007). Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genetics* *177*, 1963–1966.
- Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.-H., Wang, Q., Bolla, M.K., et al. (2019). Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am. J. Hum. Genet.* *104*, 21–34.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* *17*, 122.
- Milne, R.L., Kuchenbaecker, K.B., Michailidou, K., Beesley, J., Kar, S., Lindström, S., Hui, S., Lemaçon, A., Soucy, P., Dennis, J., et al. (2017). Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat. Genet.* *49*, 1767–1778
- Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. *Cell* *175*, 1701-1715.e16.
- Shay, T., Kang, J. (2013). Immunological Genome Project and systems immunology. *Trends Immunol.* *34*, 602–609.
- Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019). CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* *47*, D721–D728.