


RESEARCH

Open Access



Biological basis of extensive pleiotropy between blood traits and cancer risk

Miguel Angel Pardo-Cea^{1†}, Xavier Farré^{2†}, Anna Esteve^{3†}, Joanna Palade^{4†}, Roderic Espín^{1†}, Francesca Mateo¹, Eric Alsop⁴, Marc Alorda¹, Natalia Blay², Alexandra Baiges¹, Arzoo Shabbir¹, Francesc Comellas⁵, Antonio Gómez⁶, Montserrat Arnan⁷, Alex Teulé⁸, Monica Salinas⁸, Laura Berrocal⁹, Joan Brunet^{8,9,10}, Paula Rofes^{8,10}, Conxi Lázaro^{8,10}, Miquel Conesa¹¹, Juan Jose Rojas¹¹, Lars Velten^{12,13}, Wojciech Fendler¹⁴, Urszula Smyczynska¹⁴, Dipanjan Chowdhury^{15,16,17}, Yong Zeng¹⁸, Housheng Hansen He^{18,19}, Rong Li²⁰, Kendall Van Keuren-Jensen^{4*}, Rafael de Cid^{2*} and Miquel Angel Pujana^{1,21*} 

Abstract

Background The immune system has a central role in preventing carcinogenesis. Alteration of systemic immune cell levels may increase cancer risk. However, the extent to which common genetic variation influences blood traits and cancer risk remains largely undetermined. Here, we identify pleiotropic variants and predict their underlying molecular and cellular alterations.

Methods Multivariate Cox regression was used to evaluate associations between blood traits and cancer diagnosis in cases in the UK Biobank. Shared genetic variants were identified from the summary statistics of the genome-wide association studies of 27 blood traits and 27 cancer types and subtypes, applying the conditional/conjunctional false-discovery rate approach. Analysis of genomic positions, expression quantitative trait loci, enhancers, regulatory marks, functionally defined gene sets, and bulk- and single-cell expression profiles predicted the biological impact of pleiotropic variants. Plasma small RNAs were sequenced to assess association with cancer diagnosis.

Results The study identified 4093 common genetic variants, involving 1248 gene loci, that contributed to blood–cancer pleiotropism. Genomic hotspots of pleiotropism include chromosomal regions 5p15-*TERT* and 6p21-*HLA*. Genes whose products are involved in regulating telomere length are found to be enriched in pleiotropic variants. Pleiotropic gene candidates are frequently linked to transcriptional programs that regulate hematopoiesis and define progenitor cell states of immune system development. Perturbation of the myeloid lineage is indicated by pleiotropic associations with defined master regulators and cell alterations. Eosinophil count is inversely associated with cancer risk. A high frequency of pleiotropic associations is also centered on the regulation of small noncoding Y-RNAs. Predicted pleiotropic Y-RNAs show specific regulatory marks and are overabundant in the normal tissue and blood

[†]Miguel Angel Pardo, Xavier Farré, Anna Esteve, Joanna Palade, and Roderic Espín contributed equally.

*Correspondence:

Kendall Van Keuren-Jensen

kjensen@tgen.org

Rafael de Cid

rdecid@igtp.cat

Miquel Angel Pujana

mapujana@iconcologia.net

Full list of author information is available at the end of the article



of cancer patients. Analysis of plasma small RNAs in women who developed breast cancer indicates there is an overabundance of Y-RNA preceding neoplasm diagnosis.

Conclusions This study reveals extensive pleiotropism between blood traits and cancer risk. Pleiotropism is linked to factors and processes involved in hematopoietic development and immune system function, including components of the major histocompatibility complexes, and regulators of telomere length and myeloid lineage. Deregulation of Y-RNAs is also associated with pleiotropism. Overexpression of these elements might indicate increased cancer risk.

Keywords Blood trait, Cancer, Eosinophil, Hematopoiesis, Myeloid, Pleiotropy, Telomere, Y-RNA

Background

Cancer cells have evolved multiple mechanisms to avoid their recognition and elimination by the immune system [1]. Cancer immune evasion can be achieved by modulating antigen presentation, promoting immune tolerance, and/or recruiting immunosuppressive cell types, among several complementary strategies [2]. While these mechanisms are well-established in cancer progression, analogous tactics may endorse cancer initiation [3]. Evidence from mouse models with defined alterations of immune system factors [4–10], and epidemiological data from immunodeficient conditions [11, 12], indicate that immune surveillance substantially contributes to eliminating malignant cells at early stages. Characterization of premalignant lesions in mouse and human tissue also reveals meaningful changes in immune system factors and cell populations [13–16]. Indeed, a substantial proportion of genetic variants associated with cancer risk converges on immune system-related genes, pathways, and/or cell phenotypes [17–20]. However, we do not yet fully understand which systemic immune cell alterations markedly influence cancer risk [21, 22].

Naïve and educated immune cells circulate through the blood from one tissue to another, functioning to protect against harmful internal and external factors. However, there is substantial interindividual variation in the normality of blood traits. This variability is largely determined by inherited genetic factors [23, 24]. More than 7000 genetic loci have been associated with differences in blood traits among individuals in the general population, and several of the corresponding loci are linked to Mendelian blood disorders and the risk of a range of immune-related conditions [25]. Analysis of a subset of rare genetic variants associated with blood traits identified pleiotropic loci for the risk of breast and skin cancer [25]. However, despite the key role of the immune system in preventing carcinogenesis, the impact of common genetic variation on blood trait–cancer pleiotropism remains relatively undetermined.

To examine the basis of blood trait–cancer pleiotropism, we analyzed the results of the genome-wide association studies (GWASs) of 27 blood traits [24, 25]

and 27 cancer types, including breast cancer subtypes [26]. The results reveal extensive pleiotropy, identifying thousands of genetic variants that influence one or more blood trait, as well as one or more of the common cancer types and/or subtypes. Pleiotropism is thought to be caused by the perturbation of telomere length control, and alteration of immune system processes, in which master regulators and transcriptional programs of hematopoiesis are of particular relevance. The pleiotropic loci are also found to be enriched in the presence of functional and derived Y-RNA sequences, whose overexpression is associated with cancer status [27, 28] and that might indicate a relatively high risk of cancer.

Methods

Blood trait–cancer diagnosis association study

The UK Biobank (UKBB: <https://www.ukbiobank.ac.uk/>) is a large prospective cohort study for research into the causes of human disease. Full details of the UKBB have been described previously [29]. Briefly, it includes approximately half a million individuals, aged 40–69 years, recruited between 2006 and 2010 in the UK. Baseline sociodemographic, medical history, lifestyle exposures, and physical information, and blood samples were collected at the time of recruitment. Cancer diagnoses were obtained by linkage to electronic medical records, and national cancer and death registries. Data from 503,317 individuals were obtained following approval of project application #61744. To analyze the associations, and following the original study [24], we excluded individuals who showed (1) a discrepancy between self-reported sex and inferred genetic sex ($n = 373$); (2) heterozygosity outlier ($n = 968$); (3) chromosome aneuploidy ($n = 651$); (4) no information about genetic principal components ($n = 14,242$); (5) a cancer diagnosis before blood test ($n = 28,795$); (6) no information from the blood test ($n = 23,153$); (7) a discrepancy between the dates of the health care record and of the blood test ($n = 1$); (8) an outlier measure (> 3 times the interquartile range) for the leukocyte ($n = 1,124$) or platelet ($n = 871$) count; and (9) a C-reactive protein (CRP) value > 10 mg/L ($n = 19,475$). The outlier threshold applied to the leukocyte and platelet counts was based

on a previous study of prostate cancer risk [30] and aimed to exclude individuals with probable chronic inflammation and thrombocytosis, respectively. These pathological processes could have confounded the study conclusions as they have been associated with cancer development and progression [31–34]. Similarly, individuals with a CRP measure > 10 mg/L were excluded because this threshold constitutes clinical evidence of an acute infection or inflammatory reaction [35, 36], which could also confound the conclusions concerning cancer risk. Data from 32 individuals who withdrew from the UKBB project were also discarded. In total, 170,512 men and 198,331 women were included in the study. The cancer types were based on the International Classification of Diseases – 10th Edition (ICD-10) code for malignant cancer (ICD-10 Chapter C) [37]. Benign neoplasms (ICD-10-CM D10-D49) were not considered. The main outcome of the study was defined as a first diagnosis of cancer after the date of recruitment or a cancer-related death. Similarly, secondary outcomes of the study were considered for the most common cancer types: breast, colon, lung, and prostate. Peripheral blood samples of the UKBB participants were typically taken at the time of enrollment [29]. Values of all blood traits were \log_2 -transformed for the analysis. Multivariate Cox proportional hazards models were used to assess the association between blood traits and cancer diagnosis by considering a descriptive model-building strategy. The follow-up time was defined as being from the date of enrollment to the date of cancer diagnosis, death, loss to follow-up, or administrative censoring (March 31st 2016 for England and Wales, and November 30th 2015 for Scotland), whichever occurred first. We estimated HRs and 95% confidence intervals (CIs) associated with the risk of cancer diagnosis for a doubling of the value of each log-transformed blood trait. Models for the main outcome (all-cancer diagnosis), as well as separate lung and colon cancer diagnosis outcomes, were stratified by sex, alcohol consumption (non-drinker, drinker, unknown), the number of self-reported comorbidities (0, 1, 2, 3–5, > 5), and region of recruitment (England, Wales, Scotland), and adjusted by age at enrollment, body mass index (BMI), smoking status (non-smoker, smoker, unknown), highest level of educational qualifications (preparatory school, high school, college, other, unknown), the Townsend deprivation index (grouped into quintiles), and the top 40 genetic principal components [24]. To account for departures from the proportional hazards assumption more accurately, we used penalized splines for age at enrollment and BMI. Multicollinearity was assessed using the variance inflation factor. To consider the potential influence of an underlying cancer on blood traits levels, we conducted separate analyses for cancer diagnoses after

1 year and within 1 year following enrollment. Analyses were performed in R v 4.1.2 (R Core Team, 2020) using the *survival* and *survminer* packages.

GWAS data processing

The GWAS summary statistics of blood traits and cancer risk studies were obtained from the corresponding data sources, detailed in Additional file 1: Table S1. The study did not require individual data. For each of the variant-summary statistics, the following quality controls were applied, removing cases of single-nucleotide polymorphisms (SNPs) without a reference identifier (rs ID); duplication; poor imputation (information score < 0.9); value of minor allele frequency (MAF) \leq 0.01; strand-ambiguous alleles; and/or allele sample sizes five standard deviations or more away from the mean.

Shared genetic architecture analysis

The heritability of all phenotypes and genetic correlations were estimated by the linkage disequilibrium (LD) score regression method [38], restricted to HapMap3 SNPs. The pleiotropy-informed conditional false-discovery rate approach [39] was employed to detect shared genetic factors, using pleio-false discovery rate (pleioFDR) software (<https://github.com/precimed/pleiofdr/>) and computing conjFDR statistics. The conjFDR is given as the maximum value between the conditional FDRs (condFDR) of two given conditions. The method is not affected by the direction of the allele effects [40, 41]. To ensure the results were comparable, we analyzed a common set of 5,264,785 SNPs, from which all summary statistics were derived. Shared genetic variants were defined by conjFDR < 0.05. We performed LD clumping to define independently significant SNPs (PLINK software, $p_1 = 0.05$, LD threshold $r^2 = 0.6$, and physical distance threshold for clumping 1000 kb) and lead SNPs (PLINK software, $p_1 = 0.05$, $r^2 = 0.1$, and distance 1000 kb). Genomic risk loci were found by merging lead SNPs if they were closer than 250 kb. Candidate SNPs were mapped to independently significant SNPs using this clumping strategy. Stratified Q-Q plots were obtained using pleioFDR to visualize shared genetic architecture. In these representations, the probabilities of the primary phenotype were plotted against the null distribution. In the same plots, SNP subsets of the primary phenotype were represented as being conditioned by the significance of the association with the secondary phenotype ($p < 0.1$, 0.01, and 0.001). The genomic inflation factor (λ) for each of the thresholds was computed to establish the existence of pleiotropy in the stratified Q-Q plots.

Genetic data and functional associations

Positional information about genetic elements was obtained from ENSEMBL BioMart [42] version 2.52.0, genome build GRCh37/hg19. This resource was used to assign the identified pleiotropic variants to defined gene loci. The variants linked to the genes previously associated with leukocyte telomere length were identified using the original study annotations [43–45] and not considering other types of data. Functional annotations (GO terms and Reactome pathways) of positional protein-coding genes were analyzed using the *gost* tool of *gprofiler2* [46], with default parameters and using the FDR approach for multiple-test correction. The *cis* eQTL data from blood and immortalized lymphocytes were obtained from the GTEx project [47]. The pleiotropic variants in specific loci were examined for eQTLs of the corresponding positional gene, and the resulting pleiotropic/eQTL proportion compared with the frequency of eQTLs identified in sets of 200 randomly selected variants with defined MAF (European > 0.05), using different LD thresholds (five random sets; average $r^2 = 0.10$, 0.25, 0.50, 0.75, or 0.90) in 1000 random protein-coding gene loci. These genes were randomly selected from among those detected (defined as RNA-seq transcripts per million (TPM) > 1) in all immune major cell types [48]. The MAF information was obtained from the 1000 Genomes Project (<ftp.1000genomes.ebi.ac.uk>) [49]. The SNPs were assigned to the nearest gene locus (± 100 kb) using ENSEMBL BioMart [42] 2.52.0 (GRCh37/hg19), and LD was estimated using LDlinkR software [50]. A two-proportion Z-test was done to assess the enrichment of eQTLs in sets of pleiotropic variants of defined gene loci relative to randomly selected variants/genes. The enhancer data from immune cell types were obtained from the FANTOM Consortium [51] (predefined enhancer data; <https://enhancer.binf.ku.dk/presets/>). Fisher's exact test and the FDR approach were used to assess the proportion of pleiotropic variants identified in immune cell enhancers, relative to the proportions in adipose and brain data from the same study [51]. The list of mammalian phenotypes (MPs) and the corresponding mouse genes and human orthologs linked to immune system alterations was obtained from The Mammalian Phenotype Browser (keywords: "inflammation", "inflammatory", and "immune"; MP:0005387) [52]. Myeloid-related gene sets were also obtained from this source [52]. The hypergeometric test was applied to assess the degree of overlap of pleiotropic gene candidates (positional) among all genes annotated with the given term, and considering all protein-coding human orthologs as background. The Locus Overlap Analysis (LOLA; R version 1.28.0) [53] was applied for enrichment assessment of regulatory features (default reference database)

in defined genomic intervals centered in the TSSs of pleiotropic *RNYs* and the results compared with equivalent intervals of non-pleiotropic *RNYs*. The RNA repeat genome annotations were obtained from RepeatMasker (hg19, version 2020-02-20).

Phylogenetic analysis

Human *RNY*-related sequences were downloaded from BioMart (version 3.17), FASTA files compiled using *readDNAStringSet* in Biostrings (version 3.17), and sequences aligned using *msa* and ClustalW [54], and stored as DNABin and DNAStringSet (version 5.7) in APE [55]. The *msaplot* function in *ggtree* [56], *ggplot2* [57], and *dist.dna* in APE [55] were used to construct and visualize the phylogenetic tree. The pairwise sequence distance was computed using the K80 model [58]. The phylogenetic tree was estimated using the *nj* function implemented in APE [55].

Gene expression data

Data from The Cancer Genome Atlas (TCGA) were obtained via the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>) and gene expression information corresponded to FPKM-UQ values. The expression signature scores were computed using the single-sample GSEA (ssGSEA) algorithm calculated with GSVA software [59] (version 1.42.0). Analysis and visualization were carried out using the *ggplot2* [57] (version 3.3.5), *complexHeatmap* [60] (2.10.0), *circlize* [61] (version 0.4.13), and R *base* graphs (version 4.1.2) packages. To estimate the expression correlations empirically, 1000 sets of randomly selected ncRNA genes with the same length as the pleiotropy *RNY* set were selected, computed in ssGSEA, and analyzed to establish any association with age at diagnosis (TCGA clinical data annotation). The sRNA-seq data of plasma and the clinical and individual information from the corresponding healthy donors and cancer patients were downloaded from the exRNA Atlas [62]: Gene Expression Omnibus reference GSE71008 [28]. The difference in the levels of expression between the *RNY* signatures was examined using the Mann-Whitney test.

Cell-free plasma small-RNA library preparation, sequencing, and analysis

The genetic and clinical data of the two sample sets analyzed are detailed in Additional file 1: Tables S2 and S3. Plasma small RNAs were isolated using the Plasma/Serum Circulating and Exosomal RNA Purification Mini Kit (51000, Norgen Biotek) and washed and concentrated using the RNeasy MinElute Cleanup Kit (74204, QIAGEN). For plasma collected in heparin tubes (used in the prospective study), the RNA samples were further

purified using a heparinase-based protocol [63]. RNA concentration was measured using the Quant-it™ RiboGreen RNA Assay Kit and RiboGreen RNA Reagent (R11490, Thermo Fisher Scientific). Perkin Elmer's NEXTFLEX Small RNA-seq v3 kit (NOVA-5132-06) was used to prepare the small RNA libraries, with slight modifications to the manufacturer's protocol: up to 5 ng of total RNA was denatured at 70°C, and subjected to 3' ligation using 0.5x diluted adenylated adapter for 2 h at 25°C. NEXTFLEX Cleanup Beads were used to remove excess adapter. The adapter inactivation step was skipped, and 5' ligation was carried out with 0.5x diluted adenylated adapter. After cDNA synthesis and another bead cleanup, samples were PCR-amplified with UDI primers for 18 cycles. Finally, libraries were size-selected by gel electrophoresis. Samples were separated on 6% polyacrylamide gels, stained with SYBRgold, and bands of interest were excised, minced, and incubated in water overnight, with constant agitation. Gel-extracted libraries were treated with a DNA Clean and Concentrate kit (D4014, Zymo) following the manufacturer's instructions. Library size and concentration were determined with an Agilent 2100 Bioanalyzer, using a High Sensitivity DNA kit. Libraries were then pooled equimolarly, and the pool was quantified with KAPA SYBR FAST Universal qPCR Kit (KK4824) and loaded at 3.8 pM with 5% PhiX spike-in. Sequencing was done with Illumina's NovaSeq 6000 apparatus, using v1.5 SP 100 cycle reagents with XP workflow. Sequencing data were demultiplexed using Illumina's *bcl2fastq* software to generate *fastq* files for each sample. Samples were analyzed with the *exceRpt* small RNA pipeline [64] using the option to trim 4 bp from the 5' and 3' ends of the sequencing data, as specified by PerkinElmer.

Results

Blood traits associated with cancer diagnosis

Systemic alteration of specific immune cell types may enable cancer development [65]. We analyzed the association between blood traits and cancer diagnosis in the prospective cohort of the UKBB [24, 25]. After data filtering and quality control (Methods), the normalized blood trait measures of 364,791 individuals were examined for associations with cancer diagnosis using a Cox proportional hazard model that included individual and biological covariates. To prevent confounding effects from hidden tumors, the analysis was limited to individuals with a first cancer diagnosed >12 months after a basal blood test, and without considering benign neoplasms. As in previous studies [66], the C-reactive protein (CRP) was found to be associated with increased risk of cancer, although with a marginal effect: hazard

ratio (HR) = 1.02, 95% CI 1.00–1.04, $p = 0.035$ (Fig. 1 and Additional file 1: Table S4). Individuals with an indication of an acute inflammatory condition (CRP > 10 mg/L) were excluded from the analysis. Then, five blood traits were found to be significantly associated with increased risk of cancer: counts of lymphocytes (HR = 1.14, 95% CI 1.09–1.19, $p < 0.001$), erythrocytes (HR = 1.19, 95% CI 1.02–1.38, $p = 0.025$), and basophils (HR = 1.41, 95% CI 1.17–1.70, $p < 0.001$), and the distribution widths of erythrocytes (HR = 1.42, 95% CI 1.22–1.64, $p < 0.001$) and platelets (PDW: HR = 1.73, 95% CI 1.31–2.29, $p < 0.001$). In turn, two blood traits were found to be significantly associated with reduced risk: eosinophil count (HR = 0.66, 95% CI 0.60–0.71, $p < 0.001$) and platelet crit (PC: HR = 0.63, 95% CI 0.49–0.80, $p < 0.001$; Fig. 1 and Additional file 1: Table S4). The contrary effects of PDW and PC were consistent with a predictable negative correlation of these measures, and the association between platelet activation—inferred from the high PDW—and increased cancer risk might be akin to the role of this feature in tumor growth and invasion [67]. A subsequent sensitivity analysis of diagnoses within the first year after the basal blood test showed a greater effect of CRP (HR = 1.15, CI 1.10–1.20, $p < 0.001$), and predictable cancer associations with conditions analogous to anemia, indicated by cancer-risk associations with low erythrocyte count (HR = 0.54, 95% CI 0.36–0.81, $p = 0.003$) and low mean corpuscular hemoglobin concentration (HR = 0.38, 95% CI 0.23–0.63, $p < 0.001$; Additional file 1: Table S5; and Additional file 2: Fig. S1).

Analysis of cancer diagnosis >12 months after the blood test and stratified by sex showed similar results to those from the complete cohort, except for indications of a higher cancer risk linked to high neutrophil counts in women, and a lower cancer risk linked to low monocyte counts in men (Additional file 1: Tables S6 and S7). Stratified analyses for the most common cancer types (breast, colon, lung, and prostate; Additional file 1: Table S8) showed greater heterogeneity in the predicted effects of the blood traits, except for eosinophil counts, which were found to be significantly associated with a lower risk of the four cancer types (Additional file 1: Tables S9–S12). An inverse relationship between eosinophils and colorectal cancer incidence had been previously noted [68], and analogous trends towards a protective association were suggested for prostate and lung cancer risk [30, 69]. The data suggest that inter-individual differences in systemic immune cell levels influence cancer risk; however, the genetic factors and biological processes underlying pleiotropism are mostly unknown.

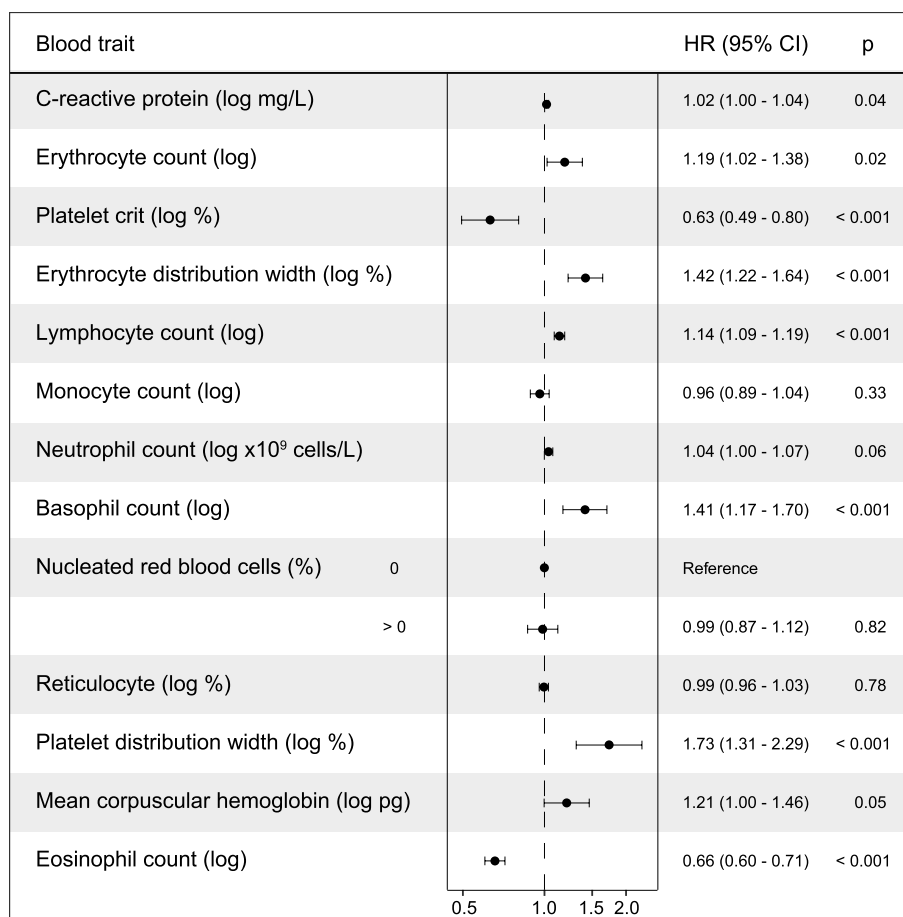


Fig. 1 Study of association of blood traits with cancer diagnosis. Forest plot showing the associations between blood traits and cancer diagnosis in the UK Biobank ($n = 364,791$). The trait units, HR, 95% CI, and significance (p) of the multivariate Cox proportional model are indicated. The dataset was filtered, blood traits \log_2 -transformed, and regression models stratified and adjusted as described in the “Methods”

Lack of global genetic correlation between blood traits and cancer risk

Host and exposome factors can alter the function of the immune system and thereby influence cancer risk [70]. Since blood traits are strongly determined by common genetic variation [24, 25], we examined the shared genetic basis of blood traits and cancer risk. We analyzed the GWAS results of 27 blood traits [24, 25] and of the risk of 27 cancer types and subtypes (subtypes of breast cancer; Additional file 1: Tables S1). After data processing and quality control analyses of the summary statistics, genetic correlations were computed using the HapMap3 [71] catalog of SNPs. Consistent with the original UKBB study [24], approximately 50% (177/351) of the pairwise comparisons of blood traits showed significant genetic correlations (FDR-adjusted $p < 0.05$; Additional file 2: Fig. S2a). By contrast, few significant genetic correlations were identified in the cancer-risk analyses, and these were only detected among the

overall and subtype-specific breast cancer studies, and for the breast-colon, breast-cervix, and colon-rectum comparisons (Additional file 2: Fig. S2b). Two GWASs were included for the analysis of breast cancer: BC#1 refers to the results from the Breast Cancer Association Consortium (BCAC) [72], including subtype analyses [26]; and BC#2 refers to the results from the UKBB [73] (Additional file 1: Tables S1). Next, analysis of the genetic correlation between blood traits and cancer risk did not reveal any significant associations (FDR-adjusted). A few nominally significant correlations were indicated, including lung cancer with white blood cell (leukocyte) counts (Additional file 1: Table S13; and Additional file 2: Fig. S2c), which was consistent with an independent observation in the UKBB [69]. Therefore, the genetics of blood traits and cancer risk are not globally correlated in the same direction when considering > 5 million variants, although pleiotropic signals might exist at specific loci.

Identification of blood trait–cancer pleiotropic variants

To identify the genetic factors shared by blood traits and cancers, we examined Q-Q plots stratified by SNP significance and conditioned for the corresponding blood trait or cancer type. Each cancer type showed evidence of deviation from expectation for an association with one or more blood traits (Additional file 2: Fig. S3). To evaluate deviation from expectation, genomic inflation scores were computed. Evidence of shared genetics ($\lambda > 1$) was obtained in 400 blood trait–cancer risk comparisons (Additional file 1: Table S14). An example of the evidence for shared genetics, the comparison between BC#1 and “lymphocyte count” (LYMPH#) at three SNP significance thresholds (LYMPH# $p < 10^{-1}$, 10^{-2} , and 10^{-3}) and for all SNPs, is shown in Fig. 2a.

Next, the condFDR/conjFDR method [39, 74] was used to leverage and identify genetic associations between blood traits and cancer risk. With a conjFDR < 0.05 , 4093 pleiotropic variants were identified, ranging from 3 to 1689, associated with gastroesophageal cancer and BC#1, respectively (Fig. 2b and Additional file 1: Table S15). Analyses of breast and prostate cancer included the data solely for females and males, respectively. The causal gene for a genetic association is often the closest gene to the specific variant [75, 76]. Next, mapping the variants to genetic elements using BioMart annotations [42] identified a range from 0 (gastroesophageal cancer) to 560 (BC#1) protein-coding genes, and relatively minor contributions from other elements (Fig. 2c). As expected, the larger cancer studies revealed more pleiotropic associations, with the exception of HER2-positive breast cancer, which yielded only 26 variants; in contrast, the melanoma and prostate studies showed comparatively more pleiotropic associations (385 and 356 variants, respectively) (Fig. 2b,d and Additional file 1: Table S15).

From the perspective of blood traits, mean corpuscular volume (MCV) and platelet count (PLT#) showed the greatest number of shared genetic variants and pleiotropic gene candidates (i.e., genes mapped to pleiotropic variants), respectively, while nucleated red blood cells

showed the weakest evidence of pleiotropy (Fig. 2e,f and Additional file 1: Table S15). Despite these profiles, all blood traits were linked to cancer risk to some extent (Fig. 2g). Subsequent grouping of blood traits by immune cell type identified specific overrepresentation and underrepresentation (FDR < 0.05) of shared variants with cancer risk. For instance, a significant enrichment of shared variants was found between reticulocytes and triple-negative breast cancer (TNBC) (Fig. 2h). Therefore, it may be concluded that broad perturbations of blood cells might influence cancer risk, although the specific processes remain to be determined.

Pleiotropism is partially linked to telomere length control

A previous study of pan-cancer pleiotropy—not considering blood traits, but including a meta-analysis of cancer GWAS UKBB results—identified 85 leading variants that influenced two or more cancer types in the same direction [73]. Our blood trait–cancer pleiotropy study identified nine variants in this set (Additional file 1: Table S16), which represents a highly significant overlap if an equivalent genome coverage is assumed: identifying nine pleiotropic variants among sets of 85 variants against a background of approximately 5 million variants has a significance of $p_{\text{hypergeometric}} = 1 \times 10^{-19}$. The nine pleiotropic variants were found to be associated with 17 blood traits and nine cancer types. The corresponding gene candidates included the telomerase RNA component (*TERC*), which had previously been shown to be associated with leukocyte telomere length [43] and the risk of diverse cancer types [77]. Following on from this observation, we identified a significant overlap of 20 genes that were linked to leukocyte telomere length [45] and that mapped to the 4093 pleiotropic variants (total pleiotropic gene candidates $n = 1228$; $p_{\text{hypergeometric}} = 0.001$). In addition to the *TERC*, the pleiotropic gene set included the telomerase reverse transcriptase (*TERT*) and the regulator of telomere elongation helicase 1 (*RTEL1*; Additional file 1: Table S17). Next, analysis of the proportion of pleiotropic variants linked to genes associated with

(See figure on next page.)

Fig. 2 Shared genetics of blood traits and cancer risk. **a** Stratified Q-Q plot for breast cancer risk (BC#1) as a function of the significance of SNP associations with LYMPH#, as indicated in the inset. The dotted line indicates no association. **b** Histogram depicting the number of variants ($n \times 10^{-3}$; conjFDR < 0.05) shared between cancer risk and blood traits. The colored bar indicates the number of individuals originally included in each cancer GWAS, as denoted in the inset. **c** Histogram depicting the distribution of classes of genetic elements (denoted in the inset) across the identified pleiotropic loci and cancer studies. **d** Plot depicting the relationship between the number (X-axis; \log_{10}) of individuals in each GWAS analyzed and the number of identified pleiotropic variants (conjFDR < 0.05 ; \log_{10}). **e** Histogram depicting the number of variants ($n \times 10^{-3}$; conjFDR < 0.05) shared by blood traits and cancer risk. **f** Histogram depicting the distribution of classes of genetic elements (denoted in the inset) across the identified pleiotropic loci and blood traits. **g** Pie charts showing the contribution of each blood trait to each cancer risk study based on the number of shared variants. Color-coded blood trait acronyms are depicted in the inset. **h** Heatmap showing the overrepresentation and underrepresentation of shared blood-trait variants for each cancer study. The significant associations (FDR-adjusted $p < 0.05$) are indicated by black-bordered squares

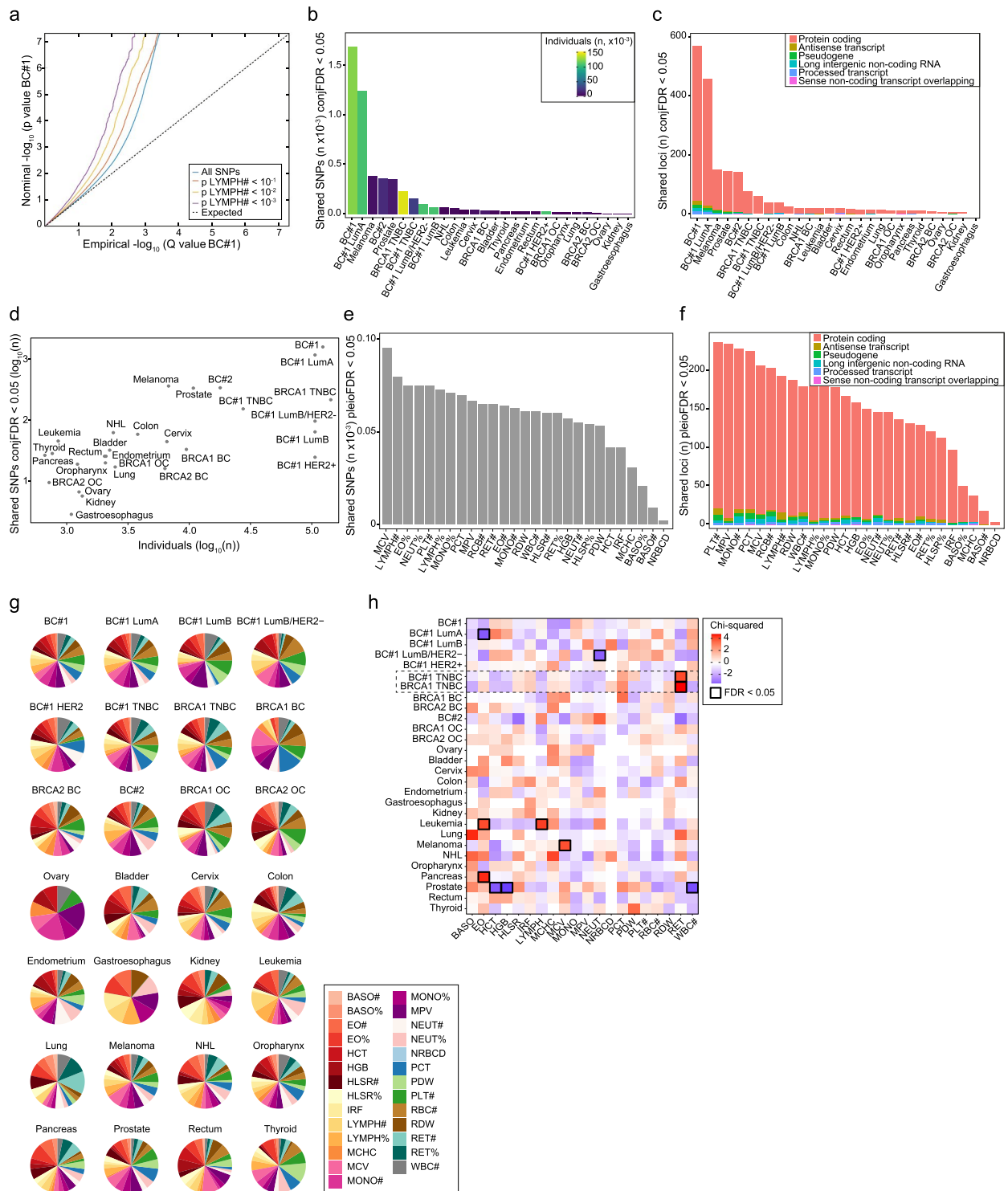


Fig. 2 (See legend on previous page.)

leukocyte telomere length revealed an enrichment in breast cancer caused by pathological variants of *BRCA1* and TNBC (32% of variants), followed by luminal A breast cancer (LumA; 16%) and melanoma (12%; Fig. 3a).

Intriguingly, luminal progenitors, the cells of origin of *BRCA1*-associated breast tumors [78], are particularly sensitive to telomere dysfunction [79]. Therefore, more than 4000 variants concurrently influence one or more

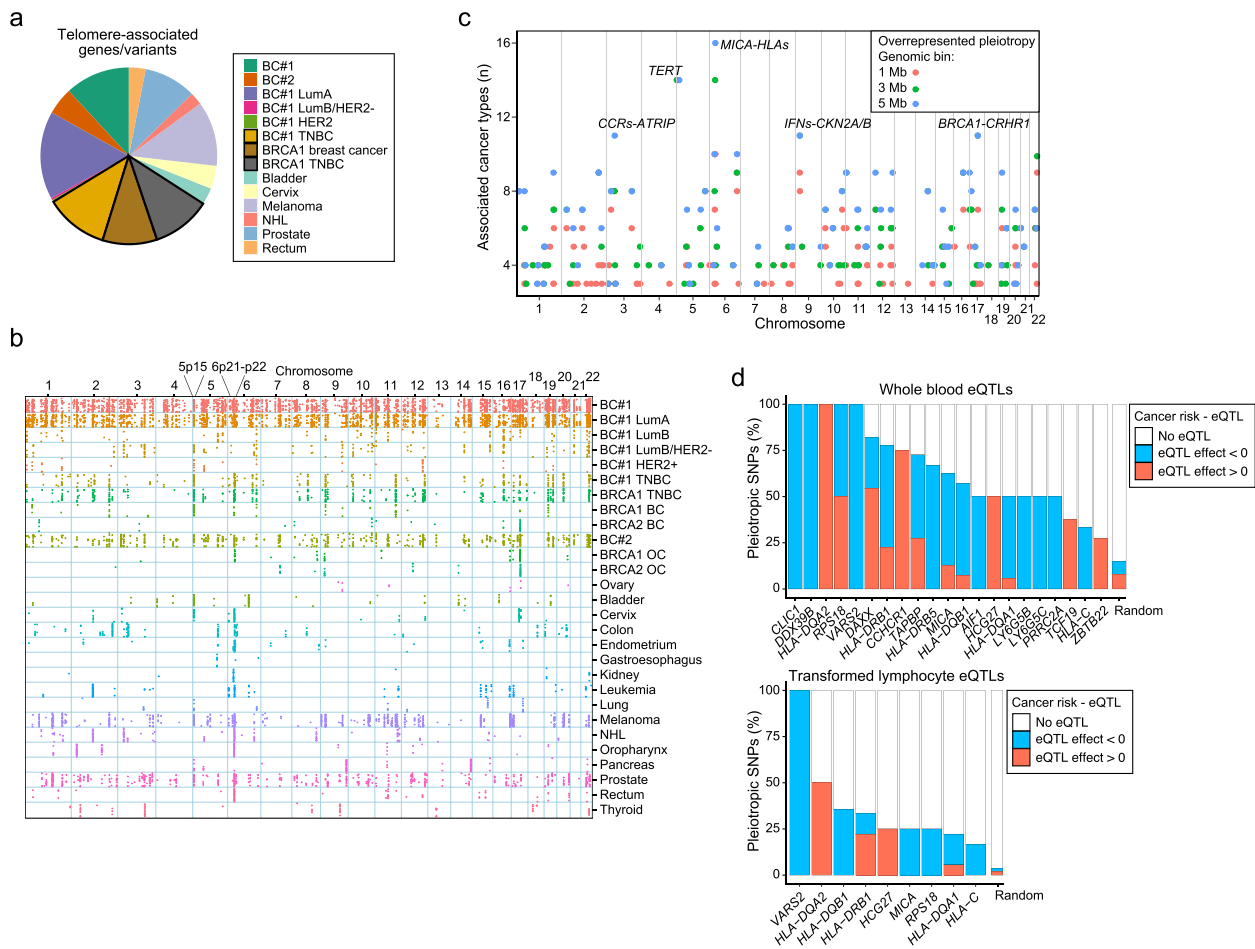


Fig. 3 Link of pleiotropism with telomere length regulation and genomic hotspots. **a** Pie charts showing the contribution of pleiotropic variants in telomere length-associated gene loci across the cancer studies. The proportion of variants associated with breast cancer caused by *BRCA1* pathological variants and/or TNBC is denoted by solid triangles, as indicated in the inset. **b** Genomic diagram showing the relative position of the pleiotropic variants (dots) across human chromosomes 1–22 (X-axis) and cancer-risk studies (Y-axis). **c** Graph showing the identified pleiotropic hotspots across human chromosomes 1–22. Results are shown for the regions including associations with > 2 cancer types and corresponding to genomic bins of 1, 3, and 5 Mb, as indicated in the inset. The hotspots including > 10 cancer trait associations are denoted by candidate gene names. **d** Histograms showing the percentage of the 6p21-p22 pleiotropic variants identified as cis-eQTL in whole blood (left panel) or immortalized lymphocytes (right panel) of the corresponding 6p21-p22 genes (X-axis). The direction of the eQTL effect is defined by the slope color (inset). The indicated genes showed significant enrichment (FDR-adjusted $p < 0.05$) of pleiotropy-eQTL correspondences relative to equivalent randomly chosen variants in 1000 gene loci expressed in all major immune cell types

blood trait and cancer risk, and regulation of telomere length in immune and/or epithelial cells might underlie this pleiotropism.

Hotspots of blood trait–cancer pleiotropism are present in the TERT and HLA regions

Examining the location of pleiotropic variants throughout the genome indicated regions with a relatively high frequency of associations (Fig. 3b). Analysis of the representation of pleiotropic associations relative to all examined variants in genomic bins of 1, 3 and 5 megabases (Mb) identified 81–159 regions with a significant

pleiotropy enrichment (chi-squared test FDR-adjusted $p < 0.05$; Fig. 3c and Additional file 1: Table S18). The genomic bins comprising associations with > 10 cancer types corresponded to the chromosomes 3p21, 5p15, 6p21-p22, 9p21, and 17q21, which, among other genes, encompass CC-motif chemokine receptors, *TERT*, human leukocyte antigens, interferons, and corticotropin-releasing hormone receptor 1, respectively (Fig. 3d).

The chromosome region with the greatest number of cancer associations ($n = 16$) corresponded to 6p21-p22 (chromosome bin from 30 to 35 Mb; Additional file 1: Table S18). To assess the regulatory impact of the

pleiotropic variants identified in this hotspot, we analyzed the correspondence with expression quantitative trait locus (eQTL) identified in whole blood and transformed lymphocytes [47], and compared the observed eQTL frequencies with those of randomly selected genetic variants (European MAF > 0.01) across different LD thresholds: $r^2 < 0.2$, 0.2–0.8, and > 0.8) from 1000 randomly chosen genes that were substantially expressed (TPM > 1) in all major immune cell types [48]. Thus, pleiotropic variants in 21 genes of chromosome 6p21-p22 were frequently found to be eQTLs in blood cells and/or lymphocytes (FDR < 0.05; Fig. 3d). Alteration of the regulation of some of these genes might therefore determine blood-cancer pleiotropism. The candidates include five *HLAs* and the major histocompatibility complex (MHC) class I polypeptide-related sequence A (*MICA*) genes.

Pleiotropic factors are frequent regulators of hematopoiesis and myeloid lineage

Telomere dysfunction alters hematopoiesis [80]. To assess the connection between pleiotropy and immune cell regulation further, we analyzed the genomic location of the pleiotropic variants in relation to enhancers identified in immune cell types and whole blood, and compared the results with those of enhancers from predicted unrelated tissue origins (adipose and brain) [81]. In six of the 12 (50%) immune cell types analyzed, the proportion of pleiotropic variants mapped to defined enhancers was significantly higher than expected, with the highest pleiotropic enrichment for enhancers in monocytes (FDR-adjusted $p < 0.05$; Fig. 4a). Next, we analyzed the occurrence of DNase I hypersensitivity and transcription factor binding sites, and epigenetic marks [53, 82, 83], in the genomic regions encompassing the positions of the identified pleiotropic variants ± 10 base pairs, and compared the observed frequency of regulatory features with that of equivalent regions in 100,000 randomly chosen variants (European MAF > 0.05). Several transcription factors were found to be overrepresented in the pleiotropic set, including some of those involved in hematopoiesis (*EGR1*, *GATA1*, and *IRF1*; Fig. 4b and Additional file 1: Table S19). The regulatory features with the greatest overrepresentation in the pleiotropic variants were the binding of RNA polymerase II (*POL2*) and the tri-methylation of the fourth lysine residue of histone H3 (*H3K4me3*), which marks transcription start sites of active genes (Fig. 4b and Additional file 1: Table S19).

We further evaluated the pleiotropic connection with master regulators of hematopoiesis. Considering the 62 curated regulators identified in the literature (Additional file 1: Table S20), 18 gene loci (29%) were found to be identified with pleiotropic variants, a significantly higher proportion than expected, given the proportion among

all protein coding genes: OR = 5.0; $P_{\text{hypergeometric}} = 9 \times 10^{-9}$. The occurrence of the candidate pleiotropic genes in the gene expression modules that portray a hematopoiesis cell hierarchy [84] was then examined. This analysis revealed a significant overlap of the pleiotropic gene set with seven modules (FDR-adjusted $P_{\text{hypergeometric}}$ values < 0.05; Fig. 4c and Additional file 1: Table S21), including a module regulated by the canonical myeloid lineage factor *SPI1*, also known as *PU.1* [85].

Next, we analyzed the profile of the pleiotropic gene set in the cell states of the hematopoietic system [86]. The signature of the pleiotropic gene set was found to be underexpressed in several progenitor cell states (Fig. 4d). Comparison of the pleiotropic signature against 100 equivalent randomly chosen gene sets (random genes among those expressing TPM > 1 in all major immune cell types [48]) confirmed significant underexpression in progenitor cell populations (Fig. 4e). The pleiotropic gene set appeared to be particularly strongly underexpressed in myeloid progenitor cell populations, including granulocyte–monocyte progenitors (GMPs), erythro-myeloid progenitors (EMP), and multipotent progenitors (MPPs) (Fig. 4e). Indeed, the pleiotropic gene set was found to have an overrepresentation of regulators of myeloid leukemia [87]: *DOT1L*, *EP300*, *FLII*, *GSE1*, and *MED24* (OR = 7.1; $P_{\text{hypergeometric}} = 4 \times 10^{-4}$). In addition, there was an overrepresentation (OR = 3.7; $P_{\text{hypergeometric}} = 5 \times 10^{-4}$) of genes that have been associated with clonal hematopoiesis through germline variation [88]. These included *ATM*, *CHEK2*, *LY75*, *PARP1*, *TERT*, *TET2*, *THADA*, *TP53*, and *ZNF318*.

Following on from the indication that perturbed hematopoiesis is linked to blood trait–cancer pleiotropism, the pleiotropic gene set was found to have an overrepresentation of mouse orthologs that cause immune system alterations when mutated or altered by allelic variants [89] (Mammalian Phenotype ontology code MP:0005387; Fig. 4f). A detailed analysis of the five ontology terms corresponding to myeloid cell alterations revealed three of them to be significantly overrepresented in the pleiotropic gene set: “decreased myeloid cell number”, “abnormal myeloid cell number”, and “abnormal myeloid cell morphology” (Fig. 4g). Therefore, the genes predicted to influence blood trait–cancer pleiotropism are frequently associated with regulating hematopoiesis and progenitor cell states, leading to potential alterations of the myeloid lineage.

High frequency of pleiotropic variants in loci containing Y-RNA-related sequences

The human genome has four functional Y-RNAs (*RNY1*, 2, 3, and 5), which are a class of small noncoding RNAs that bind and regulate Ro60 [90–92], a protein involved

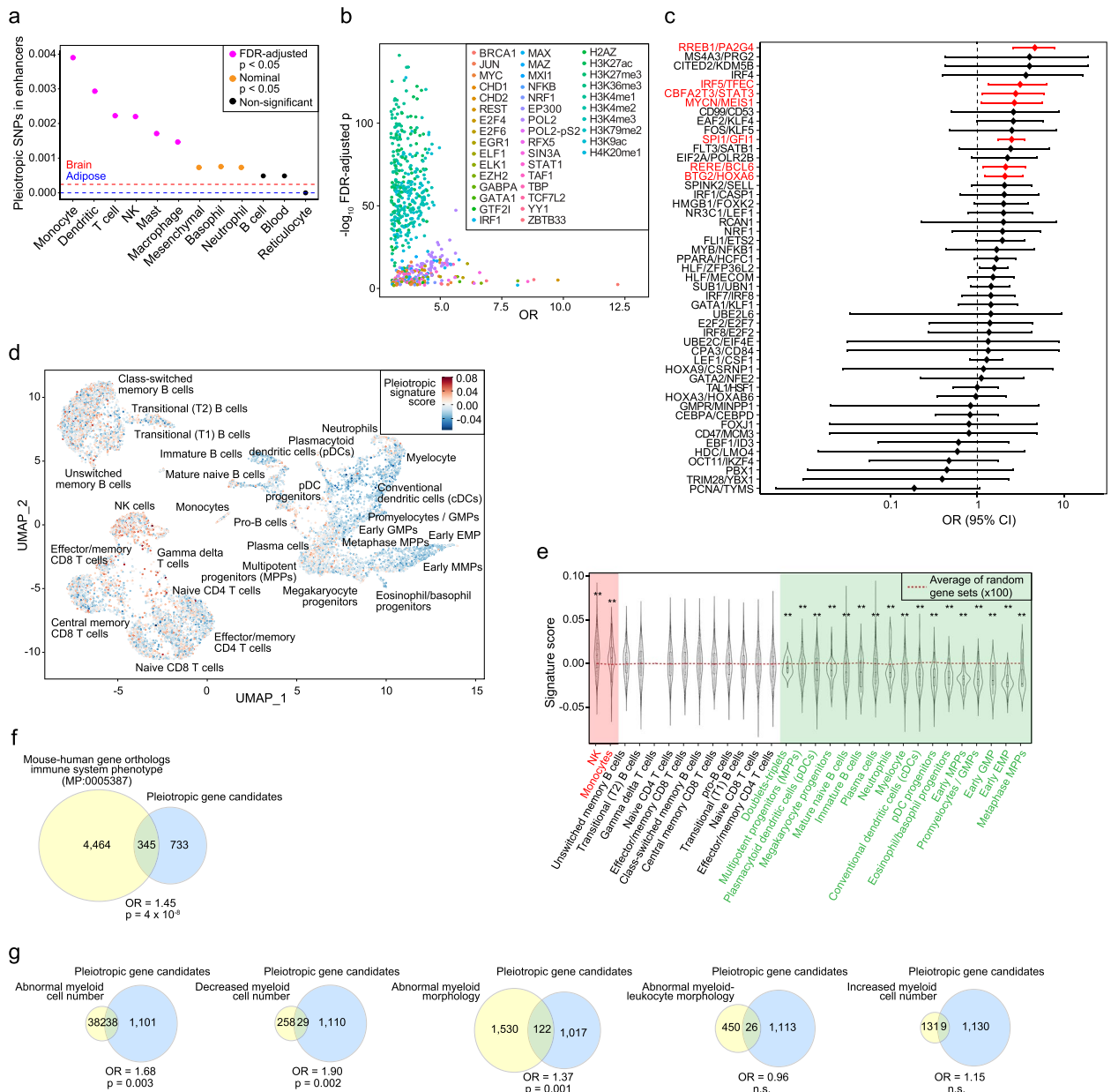


Fig. 4 Link between pleiotropic gene candidates and hematopoiesis. **a** Graph showing the proportion of pleiotropic variants (all cancers included) mapped in enhancers from immune cell types and blood (X-axis). The pink dots indicate significant overlap, as indicated in the inset. The variant-enhancer overlap proportions in brain and adipose tissue are indicated by red and blue horizontal dashed lines, respectively. **b** Graph showing the overrepresented ($-\log_{10}$ FDR-adjusted p) genomic regulatory features (binding of transcription factors and defined histone marks, denoted in the inset) in the genomic sequences centered (± 10 base pairs) on the identified pleiotropic variants ($n = 4,093$). **c** Forest plot showing the OR and 95% CI of the overlap between the pleiotropic gene set and hematopoiesis gene modules, depicted by the corresponding master regulators (Y-axis). Red bars indicate significant overlap. **d** Uniform Manifold Approximation and Projection (UMAP) of the pleiotropic gene signature expression (score indicated in inset) in the bone marrow single-cell RNA sequencing profiles. Cell clusters are annotated. **e** Violin plot showing the distribution of the pleiotropic signature expression score in each bone marrow cell type (X-axis). The horizontal line corresponds to the average score of 100 random equivalent gene sets. The asterisks indicate a significant expression difference in the pleiotropic gene signature relative to equivalent random gene sets ($**p_{\text{empirical}} < 0.01$). **f** Venn diagram showing the overlap between mouse gene orthologs that, when mutated, cause immune system alterations (MP:0005387; “immune system phenotype”) and the pleiotropic gene set (all cancers included). The OR and significance ($p_{\text{hypergeometric}}$) are indicated. **g** Venn diagrams showing the overlap between mouse gene orthologs linked to myeloid cell alterations (phenotypes are indicated) and the pleiotropic gene set (all cancers included). The OR and significance ($p_{\text{hypergeometric}}$) value are indicated; n.s., not significant

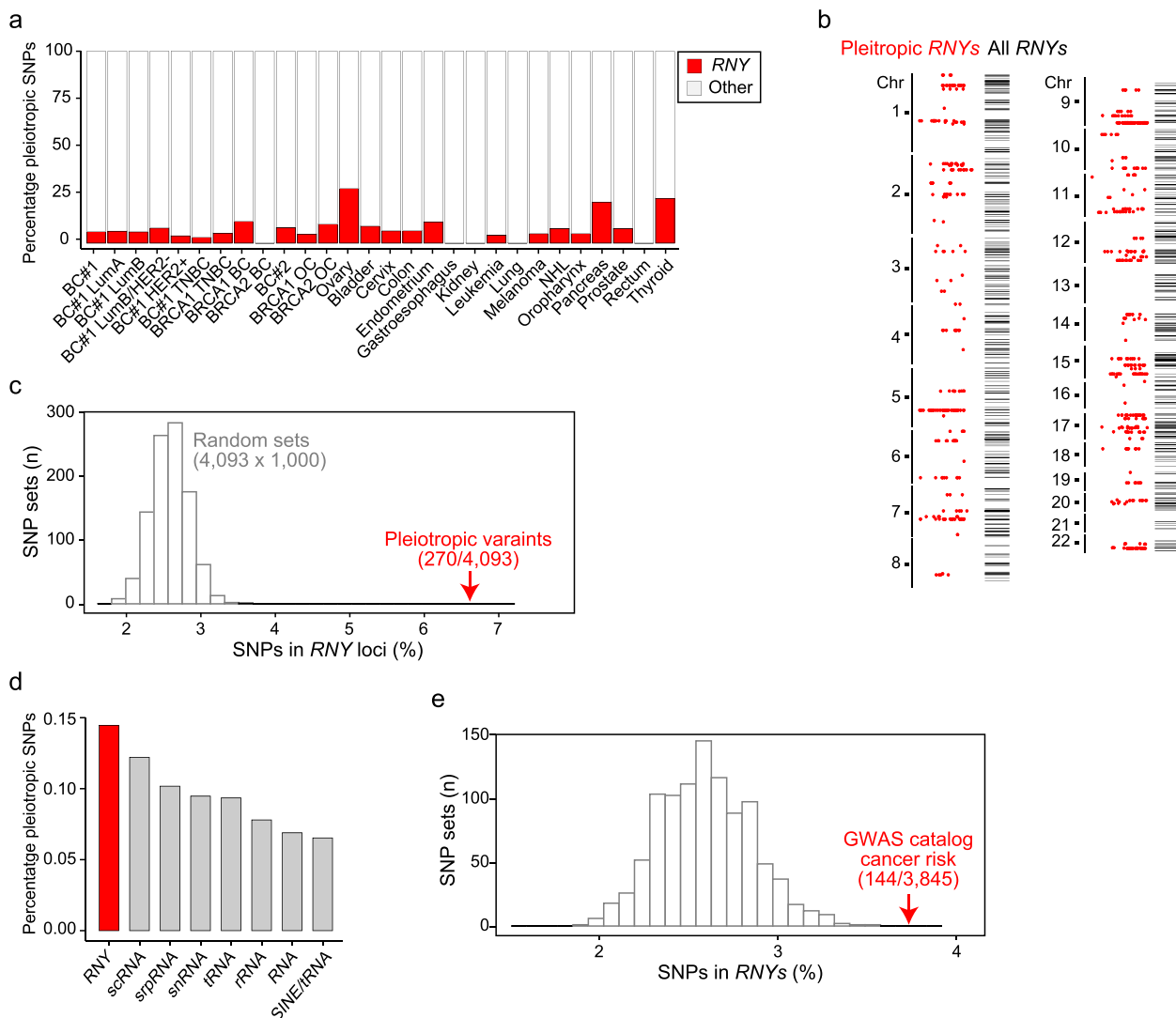


Fig. 5 High frequency of pleiotropic variants in *RNY*-containing loci. **a** Histogram showing the relative contribution of pleiotropic variants (%; Y-axis) in *RNY*-containing loci (± 50 kb centered on each variant) across cancer studies (X-axis). **b** Genomic distribution of pleiotropic variants in *RNY*-containing loci (red dots) and all *RNY*-containing loci (horizontal bars) from chromosome (chr) 1 to 22. **c** Graph showing the percentage of variants (SNPs) mapped to *RNYs* (± 50 kb) in 1000 random sets of 8155 SNPs (European MAF > 0.01 and $r^2 < 0.8$) and the observed percentage in the blood trait–cancer pleiotropy set (6.6%; 270/4,093). **d** Histogram showing the distribution of identified RNA repeat elements across the pleiotropic loci (4093 variants; ± 50 kb). The families of repeat elements are indicated (X-axis). **e** Graph showing the percentage of variants (SNPs) mapped to *RNYs* (± 50 kb) in 1000 random sets of 3847 SNPs (no filter criteria) and the observed percentage in the GWAS catalog of cancer risk variants (3.7%; 144/3,845)

in the cell’s response to stress and one identified as an autoantigen in autoimmune diseases [93]. Detailed examination of the pleiotropic loci identified numerous *RNY* genes, pseudogenes, and derived sequences (total $n = 118$) mapped in a region ± 50 kb from the pleiotropic variants across the cancer studies, with the exception of three settings: breast cancer caused by pathological variants in *BRCA2*, and gastroesophageal and kidney cancers (Fig. 5a). The *RNY*-containing loci were identified by mapping 270 pleiotropic variants (6.6% of the total 4,093

variants). They included *RNY1* and *RNY3*, four *RNY4* pseudogenes, and 112 miscellaneous Y-RNA sequences (Additional file 1: Table S22). There was no difference in the genomic distribution of the *RNY*-containing pleiotropic loci relative to all human *RNY*-derived sequences (Kolmogorov–Smirnov test $p > 0.05$; Fig. 5b). Then, the percentage of pleiotropic variants linked to *RNY* sequences was significantly higher than the expectation based on 1000 sets of 4093 randomly chosen variants—European MAF > 0.01 and $r^2 < 0.8$ in any pair— and

considering 767 *RNY* sequences annotated in the human genome, from chromosome 1 to 22, for which an average 2.8% of random variants mapped to *RNY* loci ($p_{\text{empirical}} < 0.001$; Fig. 5c). Indeed, among the established families of small noncoding RNAs, *RNY* sequences showed the closest concordance with pleiotropic loci (Fig. 5d).

Two breast cancer associations were previously predicted to target *RNY*-derived transcripts [18], and we identified these variants as being pleiotropic: rs12962334 in chromosome 18q11, which potentially targets Y-RNA ENSG00000223023; and rs1061657 in chromosome 12q24, which potentially targets Y-RNA ENSG00000199220. In addition, the study of pan-cancer pleiotropism [73] identified a potential pleiotropic *RNY* transcript in chromosome 2q14, ENSG00000201006. To assess the link between cancer risk and *RNY* sequences further, we analyzed the catalog of GWAS results [94]. Of the 3847 variants associated with cancer risk and mapped between chromosomes 1 to 22, 142 (3.7%) were found in the vicinity of an *RNY* sequence (± 50 kb; Additional file 1: Table S23). Notably, this percentage was significantly higher than expected from a consideration of 1000 sets of 3847 randomly chosen variants (dbSNP build 154; $p_{\text{empirical}} < 0.001$; Fig. 5e). We conclude that an excess of blood trait–cancer pleiotropic variants is located near *RNY* sequences, including functional *RNYs*, pseudogenes, and derived sequences.

Pleiotropic *RNYs* show specific regulatory features and relative overexpression

The pleiotropic variants identified in *RNY*-containing loci were found to be relatively highly concentrated around the corresponding transcription start sites (TSSs) and 3' regions (Fig. 6a). Only one pleiotropic variant (rs10193900) mapped within a transcribed *RNY*: the *RNYI*-derived sequence, ENSG00000201160 (Additional file 2: Fig. S4). To further determine the functionality of the pleiotropic *RNYs*, we analyzed the occurrence of DNase I hypersensitivity sites and epigenetic marks [53, 82, 83] in the regions encompassing the corresponding TSSs ± 50 kb and compared the observed frequency of regulatory elements with equivalent regions in the non-pleiotropic *RNY* loci ($n = 698$). The 5' and 3' regions of the pleiotropic *RNYs* were found to be significantly enriched in DNase I hypersensitivity sites identified in several cell lineages [82], including hematopoietic: ORs > 2 ; FDR-adjusted $p < 0.05$ (Fig. 6a and Additional file 1: Table S24). Both regions were also found to be significantly enriched in the enhancer-linked histone marks H3K4me1 and H3K27ac [83], observed in >1 assays (ORs > 3 ; FDR-adjusted $p < 0.05$) (Fig. 6a and Additional file 1: Table S24).

Consistent with marks of active transcription and enhancers, the average expression value of the pleiotropic *RNYs* in normal tissue was found to be higher than that of non-pleiotropic *RNYs*, established from the data from 15 studies included in TCGA [95] (tissue samples $n = 593$; Wilcoxon rank-sum $p = 0.014$; Fig. 6b). This difference in expression was detected despite the positive correlation between the pleiotropic and non-pleiotropic *RNY* transcript sets (hereafter “signatures”): Pearson’s correlation coefficient (PCC) = 0.82, $p < 2 \times 10^{-16}$ (Fig. 6c). Then, analysis of the *RNY* signatures in blood cell populations of neutrophils, monocytes, B, CD4 T, CD8 T, and natural killer cells [96] corroborated the overexpression of the pleiotropic set, and further indicated higher levels of this signature in myeloid relative to lymphoid cell types (2-tailed t -test $p = 0.0003$; Fig. 6d).

Analysis of the *RNY* signatures in normal tissue of TCGA showed a negative correlation with age at diagnosis for both, although it was stronger for the pleiotropic set: PCC = -0.17 vs. -0.10 ; $p = 5 \times 10^{-5}$ and 0.018, respectively (Fig. 6e). An analogous analysis using 1000 signatures of equivalent randomly selected sets of microRNAs in TCGA indicated that the negative correlation between age at diagnosis and the pleiotropic *RNY* signature was significant ($p_{\text{empirical}} = 0.035$; Fig. 6f). Multivariate logistic regression including patient sex, cancer type and subtype, and tumor stage (matched with the normal tissue analyzed) confirmed the negative correlation between the pleiotropic *RNY* signature and age at diagnosis: $\beta = -0.10$, $p = 0.025$. The analysis stratified by TCGA study was limited by the sample sizes, but reached nominal significance for the pleiotropic *RNY* signature in normal breast and esophageal tissue ($n = 112$ and 12, respectively; the non-pleiotropic *RNY* signature was also found to be significantly correlated in esophageal tissue; Additional file 2: Fig. S5). By contrast, the *RNY* association with age at diagnosis was not observed in the expression profiles of primary tumors (Fig. 6g), regardless of the high positive correlation between the two *RNY* signatures (PCC = 0.89, $p < 2 \times 10^{-16}$; Fig. 6h).

Products derived from processing *RNY* transcripts are highly abundant in body fluids and their relative overexpression has been noted in the plasma of cancer patients [27, 28, 97–100]. A large fraction of circulating *RNY* products might be derived from the *RNY4* pseudogenes [101], but phylogenetic analysis did not detect an association between *RNY4*-derived sequences and pleiotropic identification in *RNYs* (Additional file 2: Fig. S6). Subsequent examination of public plasma RNA profiles of healthy individuals and cancer patients [28] confirmed the significant overexpression of the pleiotropic *RNY* signature relative to the non-pleiotropic set (Fig. 6i). Therefore, blood trait–cancer pleiotropic variants are

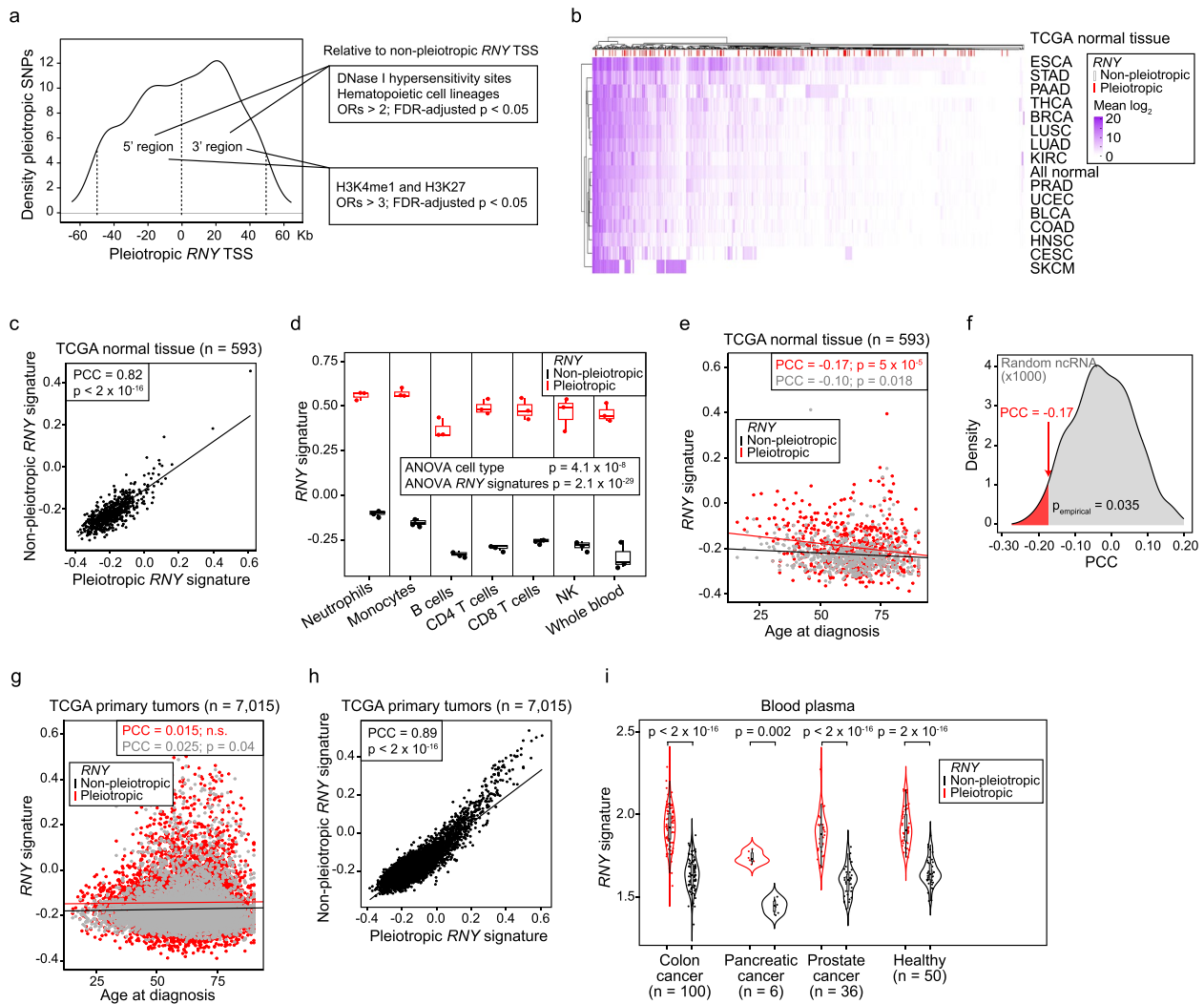


Fig. 6 Regulatory features and relative overexpression of pleiotropic *RNY*s. **a** Density distribution of the pleiotropic SNPs identified nearby (± 50 kb) *RNY* TSSs. The 5' and 3' 50-kb regions are delimited by vertical dashed lines. Genomic regulatory features found to be significantly enriched in each region are denoted in boxes. **b** Unsupervised hierarchical clustering of the average expression level of each pleiotropic and non-pleiotropic *RNY* transcript (as depicted in the inset) across normal tissue from TCGA (study acronyms are depicted on the Y-axis). **c** Scatter plot of the expression correlation between the pleiotropic and non-pleiotropic *RNY* signatures across normal tissue from TCGA. The PCC and corresponding significance (p) are indicated. **d** Box plots showing of the pleiotropic and non-pleiotropic *RNY* signature scores across primary immune cell populations isolated from whole blood. The two-way ANOVA comparisons and significance (p) are indicated. **e** Scatter plot of the correlation (PCC and p are indicated) between the pleiotropic or non-pleiotropic *RNY* expression signatures and age at diagnosis of cancer, using the corresponding normal tissue TCGA data. **f** Density distribution of the PCCs between equivalent random sets of microRNAs and age at diagnosis of cancer, using the normal tissue TCGA data ($n = 593$). The observed PCC for the pleiotropic *RNY* expression signature is indicated by an arrow, and the significant PCC tail and $p_{\text{empirical}}$ threshold are denoted. **g** Scatter plot of the correlation (PCC and p are indicated) between the pleiotropic or non-pleiotropic *RNY* signatures and age at diagnosis of cancer, using primary tumor TCGA data. **h** Scatter plot of the expression correlation between the pleiotropic and non-pleiotropic *RNY* signatures across TCGA primary tumors. The PCC and corresponding significance (p) are indicated. **i** Violin plot of the expression level of the pleiotropic and non-pleiotropic *RNY* signatures in blood plasma from cancer patients and healthy individuals, as indicated on the X-axis. Significance of the Wilcoxon rank test comparing the two signatures in each setting is shown

frequently located relatively close to *RNY* sequences, which are differentially regulated, and tend to be over-expressed in normal tissue and blood plasma of cancer patients.

Pleiotropic RNAs linked to loci influencing systemic lupus erythematosus

Ro60 controls the quality of noncoding RNAs [102, 103] and Ro60 loss causes anomalous activation of inflammatory pathways [104–106]. Ro60 binding to *RNY1*

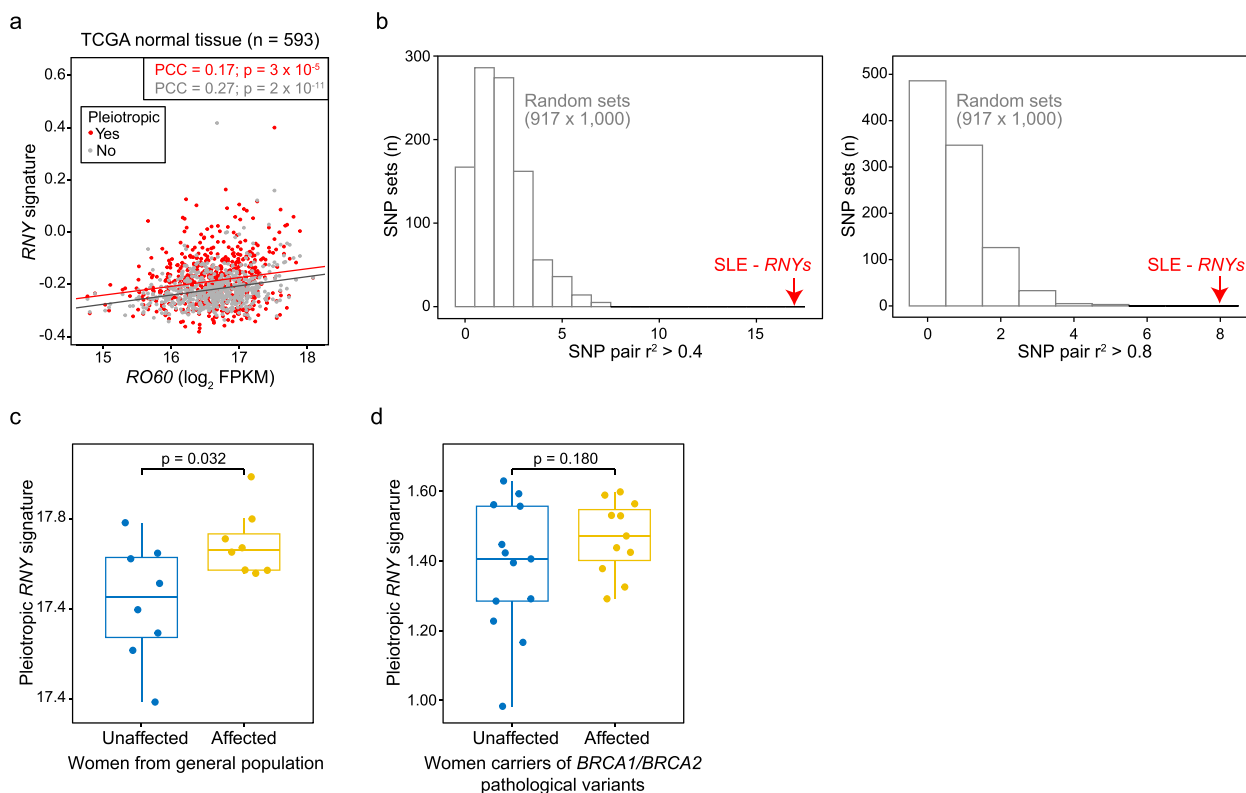


Fig. 7 Pleiotropic *RNYs* are linked to SLE risk and plasma *RNYs* are relatively abundant preceding breast cancer diagnosis. **a** Scatter plot of the correlation of the levels of expression between *RO60* and the pleiotropic or non-pleiotropic *RNY* signatures in TCGA normal tissue. The PCCs and p values are indicated. **b** Graphs showing the number of variants (SNPs) identified as pleiotropic in *RNYs* (± 50 kb) and correlated (European $r^2 > 0.4$, left panel; $r^2 > 0.8$, right panel) with SLE GWAS catalog variants, and compared with the results of equivalent 1000 random variant sets (European MAF > 0.01). **c** Box plot showing overexpression of the pleiotropic *RNY* signature in plasma of women who developed sporadic breast cancer (< 12 months after blood test) relative to matched controls who did not develop any neoplasm. The significance (p) of the Wilcoxon rank test is shown. **d** Box plot showing overexpression of the pleiotropic *RNY* signature in plasma of women carriers of pathological variants of *BRCA1* and *BRCA2* who developed breast cancer (< 12 months after blood test) relative to matched controls who did not develop any neoplasm. The significance (p) of the Wilcoxon rank test is shown

and *RNY3* is necessary to sustain a normal Ro60 level in cells, and these functional *RNYs* also influence Ro60's subcellular location and interactions [92]. In turn, Ro60 loss is correlated with reduced levels of functional *RNY* expression [104]. Similarly, we found that the expression profiles of the pleiotropic and non-pleiotropic *RNY* signatures were positively correlated with *RO60* expression in TCGA normal tissue: PCC = 0.17 and 0.27; $p = 3 \times 10^{-5}$ and 2×10^{-11} , respectively (Fig. 7a).

Ro60 was originally identified as a soluble antigen targeted by autoantibodies from patients with autoimmune rheumatic diseases; systemic lupus erythematosus (SLE) and Sjögren's syndrome [107, 108]. SLE patients have increased risk of several cancer types [109]. Next, we analyzed the GWAS catalog of SLE risk variants ($n = 917$) in search of a link to pleiotropic variants in *RNY* loci. Seventeen and eight pleiotropic variants in *RNY* TSSs ± 50 kb were found to be linked to SLE risk variants when using two thresholds (European $r^2 > 0.4$ and > 0.8 ,

respectively), and these figures of correlated genetic elements were found to be greater than expected from 1000 sets of 917 randomly selected variants (European MAF > 0.01 ; Fig. 7b and Additional file 1: Table S25). None of the pleiotropic variants was found to be linked to variants of risk for Sjögren's syndrome ($n = 48$).

Overabundance of plasma *RNY* transcripts preceding breast cancer diagnosis

Since the overexpression of *RNYs* might be associated with an increased risk of cancer, we analyzed the levels of *RNY* transcripts in plasma collected from women before they developed breast cancer and compared the results with those of matched women who remained unaffected. Using small RNA-sequencing (srRNA-seq), two independent breast cancer sets were analyzed: a set of women carriers of pathogenic variants in *BRCA1* and *BRCA2*, and diagnosed with breast cancer as a first neoplasm within 12 months of their blood test ($n = 11$), or who provided a

blood sample at a similar age and remained unaffected ($n = 13$; Additional file 1: Table S2); and a set from a long-term prospective study [110], comprising eight sporadic breast cancer cases (diagnosed within 12 months of the blood test) and eight controls matched for individual and epidemiological variables (Additional file 1: Table S3).

Unsupervised hierarchical clustering of individual *RNY* expression profiles did not distinguish women by their cancer-affected or cancer-unaffected status (Additional file 2: Fig. S7). However, computing the signature score of the pleiotropic *RNYs* showed significant overexpression in the plasma of the sporadic cases relative to unaffected women (Wilcoxon rank test $p = 0.032$; Fig. 7c). A similar, though not significant, difference was observed when comparing affected and unaffected women carriers of pathogenic variants in *BRCA1* and *BRCA2* (Fig. 7d). Consistent with the high correlation of levels of expression between *RNY* signatures (Fig. 6c,h), analysis of the non-pleiotropic *RNYs* showed similar differences in both sets (Additional file 2: Fig. S8). By contrast, the expression of four miRNAs known to be abundant in extracellular vesicles and/or lipoprotein particles of plasma (miR-16-5p, miR-21-5p, and miR-122-5p, miR-150-5p) was not significantly different in either set (Additional file 2: Fig. S9). These data suggest that overexpression of *RNY* sequences is associated with an increased risk of breast cancer.

Discussion

This study identifies 4093 pleiotropic variants influencing blood traits and cancer risk in populations of European origin. A substantial proportion of blood-cancer pleiotropism is connected to immune-related molecules and regulators of telomere length in immune and/or epithelial cells. Expanding on these observations, the predicted pleiotropic genes converge on regulatory features, gene expression profiles, and master regulators of hematopoiesis, in which factors that control myeloid lineage appear to be of greater relevance. The data provide evidence that disrupted immune surveillance increases the risk of cancer [111–113]. However, additional studies, including Mendelian randomization [114] to assess causality of the identified genetic factors, and functional assays of defined gene candidates, are required to determine the mechanisms of pleiotropism accurately.

Myeloid lineage may be of major relevance to blood trait–cancer pleiotropism, as indicated by the identification of key master regulators, their transcriptional programs and associated progenitor cell states. A recent study showed that breast tumor cells can distantly remodel the cellular cross-talks in the bone marrow niche to increase myelopoiesis [115]. Our study identifies the pleiotropic candidate *SPI1/PU.1*, which is necessary for normal myeloid and lymphoid

development [116, 117], as controlling progenitor fate, but it is specifically required for the maturation of myeloid progenitors [118]. The pleiotropic variant rs71475909 was found to be associated with breast cancer risk and eosinophil counts, and this variant is in LD with a splicing QTL of *SPI1* in blood cells [119]. In addition, *SPI1* and another proposed pleiotropic factor, *ZFPM1/FOG1* (which is linked to *BRCA1*-associated breast cancer and eosinophil counts, among other blood traits), are involved in the lineage commitment of eosinophils [120, 121]. It is of particular note that the systemic increase and tissue activation of eosinophils are associated with beneficial responses to immunotherapy in breast cancer [122], non-small cell lung cancer [123, 124], melanoma [125–127], and renal cell carcinoma [128]. In turn, high levels of circulating immunoglobulin E (IgE), and conditions of allergy and atopy may be protective of specific tumor types [129], whereas IgE immunodeficiency may increase cancer risk [130]. Thus, identified pleiotropic factors may influence cancer risk by determining myeloid lineage and the ultimate differentiation of cells, including that of eosinophils. The inferred protective effect of eosinophil counts for common cancer types in the UKBB supports this hypothesis.

Alteration of hematopoiesis and myeloid differentiation influencing blood trait–cancer pleiotropism might in turn be associated with the phenomenon of “clonal hematopoiesis”: i.e., clonal expansion of hematopoietic stem cells and their progeny due to acquired somatic mutations in driver genes, frequently linked to myeloid malignancies [131, 132]. This phenomenon causes immune dysregulation, inflammatory disease, and increased risk of hematological and solid cancers, among other consequences [133–135]. Pathological variants of genes functionally linked to the regulation of telomere length have been associated with sporadic and familial clonal hematopoiesis [88, 136]. Mendelian randomization analyses have indicated causality linking relative long telomere length to increased cancer risk [137, 138]. Further studies including clonal hematopoiesis as an additional trait are required to determine the interplay between perturbed hematopoiesis and cancer risk.

The overexpression of functional *RNYs* and of their processed fragments may induce inflammatory responses directly and/or indirectly from their interaction with Ro60 [105, 106, 139]. The plasma ratios of *RNY* subtypes are altered upon systemic inflammation [140], and *RNY*-derived sequences can activate macrophages [139]. The identification of an excess of pleiotropic signals in *RNY*-containing loci might indicate that deregulated expression of these sequences

influences cancer risk by altering the levels of immune cell types and/or inflammatory signals. According to the hypothesis, the pleiotropic variants identify *RNY* transcripts that tend to be overexpressed in normal and cancer tissue, and in plasma samples of cancer patients. Analysis of plasma *RNYs* in women prior to breast cancer development supports the link between *RNY* overexpression and increased risk, although our sample sets were of limited size. Larger studies across a range of cancer settings are needed to confirm the cancer-predictive capacity of *RNY* in body fluids. Future studies and attempts to assess applicability would also benefit from developing an informative *RNY* panel in which the corresponding transcripts are analyzed by a cost-effective method [141].

Conclusions

The study draws further attention to the relevance of the influence of systemic immune cell alterations on cancer development. The analysis reveals extensive blood–cancer pleiotropy and predicts that alteration of hematopoietic development and immune cell function principally underlies this connection. Myeloid lineage bias may be particularly relevant for blood–cancer pleiotropism. In addition, the study shows that overexpression of Y-RNAs potentially contributes to pleiotropism and might predict cancer initiation, but that larger retrospective and prospective studies across the full spectrum of settings are warranted to assess these indications. The biological factors identified here suggest opportunities for better estimating cancer risk and for developing targeted prevention approaches.

Abbreviations

BC	Breast cancer
BCAC	Breast cancer association consortium
BMI	Body mass index
CI	Confidence interval
condFDR	Conditional false discovery rate
conjFDR	Conjunctive false discovery rate
CRP	C-reactive protein
eQTL	Expression quantitative trait locus
FDR	False discovery rate
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
HR	Hazard ratio
ICD-10	International classification of diseases – 10 th edition
LD	Linkage disequilibrium
MAF	Minor allele frequency
MP	Mammalian phenotype
NHL	Non-Hodgkin's lymphoma
OR	Odds ratio
PC	Platelet crit
PCC	Pearson's correlation coefficient
PDW	Platelet distribution width
pleioFDR	Pleiotropy false discovery rate
SLE	Systemic lupus erythematosus

SNP	Single-nucleotide polymorphism
sRNA-seq	Small RNA-sequencing
ssGSEA	Single-sample gene set enrichment analysis
TCGA	The cancer genome atlas
TNBC	Triple-negative breast cancer
TPM	Transcripts per million
TSS	Transcription start site
UKBB	UK Biobank

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01294-8>.

Additional file 1: Table S1. Blood traits, cancer types and GWAS data sources. **Table S2.** Plasma samples of women carriers of pathogenic variants in *BRCA1/2*, affected or unaffected by breast cancer after blood test (< 12 months) and used for circulating sRNA-seq. **Table S3.** Plasma samples of sporadic women affected or unaffected by breast cancer after blood test (< 12 months) and used for sRNA-seq. **Table S4.** Multivariate Cox regression analysis of cancer diagnosis in UKBB (all cancers; >12 months from basal blood test). **Table S5.** Multivariate Cox regression analysis of cancer diagnosis in UKBB (all cancers; within 12 months from basal blood test). **Table S6.** Multivariate Cox regression analysis of cancer diagnosis in women of the UKBB (all cancers; >12 months from basal blood test). **Table S7.** Multivariate Cox regression analysis of cancer diagnosis in men of the UKBB (all cancers; >12 months from basal blood test). **Table S8.** Patient and incident cases included in the analyses. **Table S9.** Multivariate Cox regression analysis of breast cancer diagnosis in UKBB (>12 months from basal blood test). **Table S10.** Multivariate Cox regression analysis of colon cancer diagnosis in UKBB (>12 months from basal blood test). **Table S11.** Multivariate Cox regression analysis of lung cancer diagnosis in UKBB (>12 months from basal blood test). **Table S12.** Multivariate Cox regression analysis of prostate cancer diagnosis in UKBB (>12 months from basal blood test). **Table S13.** Heritability and genetic correlations between blood cell traits and cancer risk. **Table S14.** Genomic inflation (lambda factor) analysis for the comparisons between cancer risk and blood trait GWAS results. **Table S15.** Pleiotropy leading SNPs linking blood traits and cancer risk. **Table S16.** Pan-cancer pleiotropic SNPs (Rashkin et al., 2020) identified in the blood–cancer pleiotropy study (conjFDR < 0.05). **Table S17.** Pleiotropic gene candidates previously associated with leukocyte telomere length (Codd et al., 2021). **Table S18.** Genomic hotspots (1, 3, or 5 Mb) with significant enrichment in pleiotropic variants and linked to > 2 cancer traits. **Table S19.** Regulatory marks enriched in the blood–cancer pleiotropic variants (DNase I hypersensitivity (sheffield_dname), transcription factor binding sites (encode_tfbs), and epigenetic marks (oadmap_epigenomics) data). **Table S20.** Master regulators of hematopoiesis. **Table S21.** Pleiotropic gene candidates identified in the hematopoiesis-related gene modules (Velten et al., 2017). **Table S22.** Pleiotropic variants linked to *RNY*-containing loci. **Table S23.** GWAS-catalog cancer risk associations linked to *RNY*-containing loci (chromosomes 1–22). **Table S24.** Regulatory marks enriched in the 5' and 3'TSS regions of the pleiotropic *RNY* relative to non-pleiotropic *RNY* loci. **Table S25.** SLE risk variants (GWAS) correlated with blood–cancer pleiotropic variants in *RNY*-containing loci.

Additional file 2: Fig. S1. Blood trait associations with cancer diagnosis in the first year. **Fig. S2.** Genetic correlations among blood traits and cancer risk. **Fig. S3.** Q-Q plots for the genetic comparisons between blood traits and cancer risk. **Fig. S4.** Pleiotropic variant in a *RNY*-transcribed sequence. **Fig. S5.** *RNY* signatures and age of diagnosis of cancer types in TCGA. **Fig. S6.** Phylogenetic analysis of *RNY* sequences from the human genome. **Fig. S7.** The individual profiles of *RNYs* in plasma do not predict breast cancer. **Fig. S8.** General *RNY* overabundance in plasma is associated with breast cancer development. **Fig. S9.** Absence of association between levels of miRNAs known to be abundant in human plasma and breast cancer development.

Additional file 3.

Acknowledgements

Our results are partly based on data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>), and we are grateful to the TCGA consortia and coordinators for providing these data and the clinical information used here. We also wish to thank other consortia and investigators who provided the publicly available data used in this work, and Dr. Esther N. M. Nolte-t Hoen for guidance on Y-RNA studies. The GCAT authors would like to acknowledge all the project researchers who helped generate the corresponding data. A full list of the GCAT researchers is available from the project website (www.genomesforlife.com), and we would like to particularly thank former-researchers Anna Carreras and Betty Corté for their contribution. The GCAT authors also wish thank Joan Grifols on behalf of the Blood and Tissue Bank from Catalonia (BST) and all the volunteers who participated in the study.

Authors' contributions

MA Pujana conceived the study and wrote the manuscript. KVK-J, RC, and MA Pujana designed and supervised the study. MA Pardo, XF, AE, JP, RE, FM, EA, MA, FC, AG, MC, JJR, YZ, and HHH performed the analysis. NB, AB, AS, MA, AT, MS, LB, JB, PR, CL, LV, WF, US, DC, and RL contributed to analysis tools and data interpretation. All authors read and approved the final manuscript.

Funding

The study was partially funded by the patient foundations GINKGO Apac del Berguedà and Toca-te-les, the Instituto de Salud Carlos III (grant PI21/01306; and CIBERONC and CIBERES), co-funded by the European Regional Development Fund (ERDF), "A way to build Europe", the Generalitat de Catalunya (SGR 2017-449, 2017-1282, and 2021-184; and PERIS PFI-Salut SLT017-20-000076, Suport SLT017-20-000072, MedPerCan, and URDCat), NIH grant CA282303 (R.L), and CERCA Program of the Generalitat de Catalunya to IDIBELL and IGTP. This study makes use of data generated by the GCAT-Genomes for Life, cohort study of the Genomes of Catalonia, Fundació IGTP. GCAT was funded by the "Acción de Dinamización" of the Instituto de Salud Carlos III, Ministry of Economic Affairs and Digital Transformation (MINECO), and the Ministry of Health of the Generalitat de Catalunya (ADE 10/00026) and has additional support of the VEIS project (001-P-001647), co-funded by European Regional Development Fund (ERDF), "A way to build Europe" and the Instituto de Salud Carlos III (grant PI18/01512).

Availability of data and materials

The sRNA-seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) database [142] under accession number GSE239907 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE239907>) [143]. The individual UKBB [144] protected data were obtained upon application request and approval: project 61744 (<https://www.ukbiobank.ac.uk/enable-your-research/approved-research/study-of-white-blood-cell-counts-in-relation-to-cancer-risk>) [145]. The sources of the summary statistics of the GWASs are denoted in Additional file 1: Table S1. Validation analyses were performed using publicly deposited data: GTEx Portal [47], Open Access Datasets (https://www.gtportal.org/home/downloads/adult-gtex/bulk_tissue_expression) [146]; FANTOM5 Human Enhancers [51] (https://enhancer.binfku.dk/human_enhancers/presets) [147]; gene expression of immune cell states [86], BioStudies accession S-EPMC8642243 (<https://www.ebi.ac.uk/biostudies/europepmc/studies/S-EPMC8642243>) [148]; Mammalian Phenotype Browser [52], immune system phenotypes (https://www.informatics.jax.org/vocab/mp_ontology/MP:0005387) [149]; GWAS Catalog [94] (<https://www.ebi.ac.uk/gwas/api/search/downloads/full>) [150]; TCGA [95] data, Genomics Data Commons Portal (<https://portal.gdc.cancer.gov/>) [151]; RNA-seq data of blood immune cell populations [96], GEO [142] accession GSE60424 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60424>) [152]; and plasma extracellular RNA profiles [28], GEO [142] accession GSE71008 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71008>) [153]. All original code has been deposited at GitHub (<https://github.com/pujana-lab/PleiotropyBloodCancer>) [154] and is publicly available.

Declarations

Ethics approval and consent to participate

All research was carried out in accordance with relevant national and European guidelines and regulations. The study of UKBB individual data was approved with reference 61744. The study of plasma biomarkers was

approved by IDIBELL's Ethics Committee with reference PR217/21. The GCAT study was carried out using anonymized data provided by the Catalan Agency for Quality and Health Assessment, within the framework of the PADRIS Program. The participants provided informed written consent. The research conformed to the principles of the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ProCURE, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, 08908 Barcelona, Catalonia, Spain. ²Genomes for Life – GCAT Lab Group, Institut Germans Trias i Pujol (IGTP), Badalona, 08916 Barcelona, Catalonia, Spain. ³Badalona Applied Research Group in Oncology (B-ARGO), Catalan Institute of Oncology, Institut Germans Trias i Pujol (IGTP), Badalona, 08916 Barcelona, Catalonia, Spain. ⁴Cancer and Cell Biology, Translational Genomics Research Institute (TGen), Arizona, Phoenix, AZ 85004, USA. ⁵Department of Mathematics, Technical University of Catalonia, Castelldefels, 08860 Barcelona, Catalonia, Spain. ⁶Department of Biosciences, Faculty of Sciences and Technology (FCT), University of Vic - Central University of Catalonia (UVic-UCC), Vic, 08500 Barcelona, Catalonia, Spain. ⁷Department of Hematology, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, 08908 Barcelona, Catalonia, Spain. ⁸Hereditary Cancer Program, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, 08908 Barcelona, Catalonia, Spain. ⁹OncoGir, Catalan Institute of Oncology, Girona Biomedical Research Institute (IDIBGI), 17190 Salt, Catalonia, Spain. ¹⁰Biomedical Research Network Centre in Cancer (CIBERONC), Instituto de Salud Carlos III, 28222 Madrid, Spain. ¹¹Department of Pathology and Experimental Therapies, University of Barcelona (UB), Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, 08908 Barcelona, Catalonia, Spain. ¹²Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Spain. ¹³University Pompeu Fabra (UPF), 08002 Barcelona, Spain. ¹⁴Department of Biostatistics and Translational Medicine, Medical University of Lodz, 92-215 Lodz, Poland. ¹⁵Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ¹⁶Center for BRCA and Related Genes, Dana-Farber Cancer Institute, Boston, MA 02115, USA. ¹⁷Harvard Medical School, Boston, MA 02115, USA. ¹⁸Princess Margaret Cancer Center, University Health Network, Toronto, ON M5G 2C4, Canada. ¹⁹Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada. ²⁰Department of Biochemistry and Molecular Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, DC 20052, USA. ²¹Biomedical Research Network Centre in Respiratory Diseases (CIBERES), Instituto de Salud Carlos III, 28222 Madrid, Spain.

Received: 11 August 2023 Accepted: 22 January 2024

Published online: 02 February 2024

References

- Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell*. 2017;168:707–23.
- van Weverwijk A, de Visser KE. Mechanisms driving the immunoregulatory function of cancer cells. *Nat Rev Cancer*. 2023;23:193–215.
- Swann JB, Smyth MJ. Immune surveillance of tumors. *J Clin Invest*. 2007;117:1137–46.
- Dighe AS, Richards E, Old LJ, Schreiber RD. Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFN gamma receptors. *Immunity*. 1994;1:447–56.
- van den Broek ME, Kägi D, Ossendorp F, Toes R, Vamvakas S, Lutz WK, et al. Decreased tumor surveillance in perforin-deficient mice. *J Exp Med*. 1996;184:1781–90.
- Kaplan DH, Shankaran V, Dighe AS, Stockert E, Aguet M, Old LJ, et al. Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci U S A*. 1998;95:7556–61.

7. Smyth MJ, Thia KY, Street SE, Cretney E, Trapani JA, Taniguchi M, et al. Differential tumor surveillance by natural killer (NK) and NKT cells. *J Exp Med*. 2000;191:661–8.
8. M G, De O, Cr S, JM L, E G, R F, et al. Regulation of cutaneous malignancy by gammadelta T cells. *Science*. 2001;294:605–9.
9. Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature*. 2001;410:1107–11.
10. Street SEA, Trapani JA, MacGregor D, Smyth MJ. Suppression of lymphoma and epithelial malignancies effected by interferon gamma. *J Exp Med*. 2002;196:129–34.
11. Engels EA, Pfeiffer RM, Fraumeni JF, Kasiske BL, Israni AK, Snyder JJ, et al. Spectrum of cancer risk among US solid organ transplant recipients. *JAMA*. 2011;306:1891–901.
12. Frisch M, Biggar RJ, Engels EA, Goedert JJ. AIDS-Cancer Match Registry Study Group. Association of cancer with AIDS-related immunosuppression in adults. *JAMA*. 2001;285:1736–45.
13. Wang DJ, Ratnam NM, Byrd JC, Guttridge DC. NF- κ B functions in tumor initiation by suppressing the surveillance of both innate and adaptive immune cells. *Cell Rep*. 2014;9:90–103.
14. Ratnam NM, Peterson JM, Talbert EE, Ladner KJ, Rajasekera PV, Schmidt CR, et al. NF- κ B regulates GDF-15 to suppress macrophage surveillance during early tumor development. *J Clin Invest*. 2017;127:3796–809.
15. Bach K, Pensa S, Zarocsinceva M, Kania K, Stockis J, Pinaud S, et al. Time-resolved single-cell analysis of Brca1 associated mammary tumorigenesis reveals aberrant differentiation of luminal progenitors. *Nat Commun*. 2021;12:1502.
16. Mateo F, He Z, Mei L, de Garibay GR, Herranz C, García N, et al. Modification of BRCA1-associated breast cancer risk by HMMR overexpression. *Nat Commun*. 2022;13:1895.
17. Ferreira MA, Gamazon ER, Al-Ejeh F, Aittomäki K, Andrusis IL, Anton-Culver H, et al. Genome-wide association and transcriptome studies identify target genes and risk loci for breast cancer. *Nat Commun*. 2019;10:1741.
18. Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. *Nat Genet*. 2020;52:56–73.
19. Palomero L, Galván-Femenía I, de Cid R, Espín R, Barnes DR, et al. Immune cell associations with cancer risk. *iScience*. 2020;23:101296.
20. Lim YW, Chen-Harris H, Mayba O, Lianoglou S, Wuster A, Bhargale T, et al. Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proc Natl Acad Sci U S A*. 2018;115:E11701–10.
21. Song M, Tworoger SS. Systemic immune response and cancer risk: Filling the missing piece of immuno-oncology. *Cancer Res*. 2020;80:1801–3.
22. Srivastava S, Ghosh S, Kagan J, Mazurchuk R. The PreCancer Atlas (PCA). *Trends Cancer*. 2018;4:513–4.
23. Evans DM, Frazer IH, Martin NG. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res*. 1999;2:250–7.
24. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167:1415–1429.e19.
25. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The polygenic and monogenic basis of blood traits and diseases. *Cell*. 2020;182:1214–1231.e11.
26. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52:572–81.
27. Christov CP, Trivier E, Krude T. Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br J Cancer*. 2008;98:981–8.
28. Yuan T, Huang X, Woodcock M, Du M, Dittmar R, Wang Y, et al. Plasma extracellular RNA profiles in healthy and cancer patients. *Sci Rep*. 2016;6:19413.
29. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
30. Watts EL, Perez-Cornago A, Kothari J, Allen NE, Travis RC, Key TJ. Hematologic markers and prostate cancer risk: A prospective analysis in UK Biobank. *Cancer Epidemiol Biomark Prev*. 2020;29:1615–26.
31. Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002;420:860–7.
32. Greten FR, Grivennikov SI. Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity*. 2019;51:27–41.
33. Haemmerle M, Stone RL, Menter DG, Afshar-Kharghan V, Sood AK. The platelet lifeline to cancer: Challenges and opportunities. *Cancer Cell*. 2018;33:965–83.
34. Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W. Clinical relevance of thrombocytosis in primary care: A prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract*. 2017;67:e405–13.
35. Pepys MB, Hirschfeld GM. C-reactive protein: A critical update. *J Clin Invest*. 2003;111:1805–12.
36. Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, Criqui M, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*. 2003;107:499–511.
37. World Health Organization. ICD-10 : international statistical classification of diseases and related health problems / World Health Organization. 10th ed. Geneva: World Health Organization; 2016.
38. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5.
39. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet*. 2013;9:e1003455.
40. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet*. 2013;45:670–5.
41. Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA. New statistical approaches exploit the polygenic architecture of schizophrenia-implications for the underlying neurobiology. *Curr Opin Neurobiol*. 2016;36:89–98.
42. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart Central Portal - Unified access to biological data. *Nucleic Acids Res*. 2009;37:W23–7.
43. Codd V, Mangino M, van der Harst P, Braund PS, Kaiser M, Beveridge AJ, et al. Common variants near *TERC* are associated with mean telomere length. *Nat Genet*. 2010;42:197–9.
44. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nat Genet*. 2013;45:422–7–427e1–2.
45. Codd V, Wang Q, Allara E, Musicha C, Kaptoge S, Stoma S, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet*. 2021;53:1425–33.
46. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*. 2019;47:W191–8.
47. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348:660–5.
48. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell*. 2018;175:1701–1715.e16.
49. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res*. 2020;48:D941–7.
50. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front Genet*. 2020;11:157.
51. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455–61.

52. Smith CL, Eppig JT. The mammalian phenotype ontology: Enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med.* 2009;1:390–9.
53. Sheffield NC, Bock C. LOLA: Enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinforma.* 2016;32:587–9.
54. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinforma.* 2002. Chapter 2:Unit 2.3.
55. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinforma.* 2004;20:289–90.
56. Xu S, Dai Z, Guo P, Fu X, Liu S, Zhou L, et al. ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Mol Biol Evol.* 2021;38:4039–42.
57. Tyner S, Briatte F, Hofmann H. Network visualization with ggplot2. *R J.* 2017;9:27–59.
58. Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc Natl Acad Sci U S A.* 1981;78:454–8.
59. Hänzelmann S, Castelo R, Guinney J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
60. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinforma.* 2016;32:2847–9.
61. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinforma.* 2014;30:2811–2.
62. Murillo OD, Thistlethwaite W, Rozowsky J, Subramanian SL, Lucero R, Shah N, et al. exRNA atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. *Cell.* 2019;177:463–477.e15.
63. Kondratov K, Kurapeev D, Popov M, Sidorova M, Minasian S, Galagudza M, et al. Heparinase treatment of heparin-contaminated plasma from coronary artery bypass grafting patients enables reliable quantification of microRNAs. *Biomol Detect Quantif.* 2016;8:9–14.
64. Rozowsky J, Kitchen RR, Park JJ, Galeev TR, Diao J, Warrell J, et al. exceRpt: A comprehensive analytic platform for extracellular RNA profiling. *Cell Syst.* 2019;8:352–357.e3.
65. Gonzalez H, Hagerling C, Werb Z. Roles of the immune system in cancer: From tumor initiation to metastatic progression. *Genes Dev.* 2018;32:1267–84.
66. Zhu M, Ma Z, Zhang X, Hang D, Yin R, Feng J, et al. C-reactive protein and cancer risk: A pan-cancer study of prospective cohort and Mendelian randomization analysis. *BMC Med.* 2022;20:301.
67. Gay LJ, Felding-Habermann B. Contribution of platelets to tumor metastasis. *Nat Rev Cancer.* 2011;11:123–34.
68. Prizment AE, Anderson KE, Viswanathan K, Folsom AR. Inverse association of eosinophil count with colorectal cancer incidence: atherosclerosis risk in communities study. *Cancer Epidemiol Biomark Prev.* 2011;20:1861–4.
69. Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji B-T, et al. White blood cell count and risk of incident lung cancer in the UK Biobank. *JNCI Cancer Spectr.* 2020;4:pkz102.
70. Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA. Inflammation-induced cancer: Crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer.* 2013;13:759–71.
71. International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426:789–96.
72. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551:92–4.
73. Rashkin SR, Graff RE, Kachuri L, Thai KK, Alexeeff SE, Blatchins MA, et al. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat Commun.* 2020;11:4423.
74. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet.* 2013;92:197–209.
75. Stacey D, Fauman EB, Ziemek D, Sun BB, Harshfield EL, Wood AM, et al. ProGeM: A framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res.* 2019;47:e3.
76. Weeks EM, Ulirsch JC, Cheng NY, Trippe BL, Fine RS, Miao J, et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat Genet.* 2023;55:1267–76.
77. McNally EJ, Luncsford PJ, Armanios M. Long telomeres and cancer risk: The price of cellular immortality. *J Clin Invest.* 2019;129:3474–81.
78. Molyneux G, Geyer FC, Magnay F-A, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell.* 2010;7:403–17.
79. Kannan N, Huda N, Tu L, Droumeva R, Aubert G, Chavez E, et al. The luminal progenitor compartment of the normal human mammary gland constitutes a unique site of telomere dysfunction. *Stem Cell Rep.* 2013;1:28–37.
80. Morrison SJ, Prowse KR, Ho P, Weissman IL. Telomerase activity in hematopoietic cells is associated with self-renewal potential. *Immunity.* 1996;5:207–16.
81. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, et al. A promoter-level mammalian expression atlas. *Nature.* 2014;507:462–70.
82. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* 2013;23:777–88.
83. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518:317–30.
84. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol.* 2017;19:271–81.
85. Nerlov C, Graf T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* 1998;12:2403–12.
86. Triana S, Vonficht D, Jopp-Saile L, Raffel S, Lutz R, Leonce D, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat Immunol.* 2021;22:1577–89.
87. Wang E, Zhou H, Nadorp B, Cayanag G, Chen X, Yeaton AH, et al. Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. *Cell Stem Cell.* 2021;28:718–731.e6.
88. Kessler MD, Damask A, O'Keefe S, Banerjee N, Li D, Watanabe K, et al. Common and rare variant associations with clonal haematopoiesis phenotypes. *Nature.* 2022;612:301–9.
89. Blake JA, Baldarelli R, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Res.* 2021;49:D981–7.
90. Lerner MR, Boyle JA, Hardin JA, Steitz JA. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science.* 1981;211:400–2.
91. Hendrick JP, Wolin SL, Rinke J, Lerner MR, Steitz JA. Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: Further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol Cell Biol.* 1981;1:1138–49.
92. Leng Y, Sim S, Magidson V, Wolin SL. Noncoding Y RNAs regulate the levels, subcellular distribution and protein interactions of their Ro60 autoantigen partner. *Nucleic Acids Res.* 2020;48:6919–30.
93. Boccitto M, Wolin SL. Ro60 and Y RNAs: Structure, functions, and roles in autoimmunity. *Crit Rev Biochem Mol Biol.* 2019;54:133–52.
94. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malagone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–12.
95. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell.* 2018;173:400–416.e11.
96. Linsley PS, Speake C, Whalen E, Chaussabel D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS One.* 2014;9:e109760.
97. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep sequencing of serum small RNAs identifies patterns of 5' tRNA half and YRNA fragment expression associated with breast cancer. *Biomark Cancer.* 2014;6:37–47.
98. Victoria Martinez B, Dhahbi JM, Nunez Lopez YO, Lamperska K, Golusinski P, Luczewski L, et al. Circulating small non-coding RNA signature in head and neck squamous cell carcinoma. *Oncotarget.* 2015;6:19246–63.

99. Tolkach Y, Niehoff E-M, Stahl AF, Zhao C, Kristiansen G, Müller SC, et al. YRNA expression in prostate cancer patients: diagnostic and prognostic implications. *World J Urol.* 2018;36:1073–8.
100. Solé C, Tramonti D, Schramm M, Goicoechea I, Armesto M, Hernandez LI, et al. The circulating transcriptome as a source of biomarkers for melanoma. *Cancers.* 2019;11:E70.
101. Lovisa F, Di Battista P, Gaffo E, Damanti CC, Garbin A, Galligani I, et al. RNY4 in circulating exosomes of patients with pediatric anaplastic large cell lymphoma: An active player? *Front Oncol.* 2020;10:238.
102. Fuchs G, Stein AJ, Fu C, Reinisch KM, Wolin SL. Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. *Nat Struct Mol Biol.* 2006;13:1002–9.
103. O'Brien CA, Wolin SL. A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors. *Genes Dev.* 1994;8:2891–903.
104. Hung T, Pratt GA, Sundararaman B, Townsend MJ, Chaivorapol C, Bhargale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science.* 2015;350:455–9.
105. Reed JH, Sim S, Wolin SL, Clancy RM, Buyon JP. Ro60 requires Y3 RNA for cell surface exposure and inflammation associated with cardiac manifestations of neonatal lupus. *J Immunol.* 1950;2013(191):110–6.
106. Clancy RM, Alvarez D, Komissarova E, Barrat FJ, Swartz J, Buyon JP. Ro60-associated single-stranded RNA links inflammation with fetal cardiac fibrosis via ligation of TLRs: A novel pathway to autoimmune-associated heart block. *J Immunol.* 1950;2010(184):2148–55.
107. Clark G, Reichlin M, Tomasi TB. Characterization of a soluble cytoplasmic antigen reactive with sera from patients with systemic lupus erythematosus. *J Immunol.* 1950;1969(102):117–22.
108. Alsbaugh M, Maddison P. Resolution of the identity of certain antigen-antibody systems in systemic lupus erythematosus and Sjögren's syndrome: An interlaboratory collaboration. *Arthritis Rheum.* 1979;22:796–8.
109. Song L, Wang Y, Zhang J, Song N, Xu X, Lu Y. The risks of cancer development in systemic lupus erythematosus (SLE) patients: A systematic review and meta-analysis. *Arthritis Res Ther.* 2018;20:270.
110. Obón-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galván-Femenía I, et al. GCAT[Genomes for life: a prospective cohort study of the genomes of Catalonia. *BMJ Open.* 2018;8:e018324.
111. Dersh D, Holly J, Yewdell JW. A few good peptides: MHC class I-based cancer immunosurveillance and immune evasion. *Nat Rev Immunol.* 2021;21:116–28.
112. Lanna A, Vaz B, D'Ambra C, Valvo S, Vuotto C, Chiurchiù V, et al. An intercellular transfer of telomeres rescues T cells from senescence and promotes long-term immunological memory. *Nat Cell Biol.* 2022;24:1461–74.
113. Schratz KE, Flasch DA, Atik CC, Cosner ZL, Blackford AL, Yang W, et al. T cell immune deficiency rather than chromosome instability predisposes patients with short telomere syndromes to squamous cancer. *Cancer Cell.* 2023;41:807–817.e6.
114. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet.* 1986;1:507–8.
115. Gerber-Ferder Y, Cosgrove J, Duperray-Susini A, Missolo-Koussou Y, Dubois M, Stepaniuk K, et al. Breast cancer remotely imposes a myeloid bias on haematopoietic stem cells by reprogramming the bone marrow niche. *Nat Cell Biol.* 2023;25:1736–45.
116. McKercher SR, Torbett BE, Anderson KL, Henkel GW, Vestal DJ, Baribault H, et al. Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. *EMBO J.* 1996;15:5647–58.
117. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science.* 1994;265:1573–7.
118. Iwasaki H, Somoza C, Shigematsu H, Duprez EA, Iwasaki-Arai J, Mizuno S-I, et al. Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood.* 2005;106:1590–600.
119. Garrido-Martín D, Borsari B, Calvo M, Reverter F, Guigó R. Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat Commun.* 2021;12:727.
120. Gombart AF, Kwok SH, Anderson KL, Yamaguchi Y, Torbett BE, Koeffler HP. Regulation of neutrophil and eosinophil secondary granule gene expression by transcription factors C/EBP epsilon and PU.1. *Blood.* 2003;101:3265–73.
121. Querfurth E, Schuster M, Kulesha H, Crispino JD, Döderlein G, Orkin SH, et al. Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev.* 2000;14:2515–25.
122. Blomberg OS, Spagnuolo L, Garner H, Voorwerk L, Isaeva OI, van Dyk E, et al. IL-5-producing CD4+ T cells and eosinophils cooperate to enhance response to immune checkpoint blockade in breast cancer. *Cancer Cell.* 2023;41:106–123.e10.
123. Alves A, Dias M, Campainha S, Barroso A. Peripheral blood eosinophilia may be a prognostic biomarker in non-small cell lung cancer patients treated with immunotherapy. *J Thorac Dis.* 2021;13:2716–27.
124. Okauchi S, Shiozawa T, Miyazaki K, Nishino K, Sasatani Y, Ohara G, et al. Association between peripheral eosinophils and clinical outcomes in patients with non-small cell lung cancer treated with immune checkpoint inhibitors. *Pol Arch Intern Med.* 2021;131:152–60.
125. Simon SCS, Hu X, Panten J, Grees M, Renders S, Thomas D, et al. Eosinophil accumulation predicts response to melanoma treatment with immune checkpoint inhibitors. *Oncoimmunology.* 2020;9:1727116.
126. Delyon J, Mateus C, Lefevre D, Lanoy E, Zitvogel L, Chaput N, et al. Experience in daily practice with ipilimumab for the treatment of patients with metastatic melanoma: An early increase in lymphocyte and eosinophil counts is associated with improved survival. *Ann Oncol.* 2013;24:1697–703.
127. Wolf MT, Ganguly S, Wang TL, Anderson CW, Sadtler K, Narain R, et al. A biologic scaffold-associated type 2 immune microenvironment inhibits tumor formation and synergizes with checkpoint immunotherapy. *Sci Transl Med.* 2019;11:eaat7973.
128. Verhaar SL, Abu-Ghanem Y, Mulder SF, Oosting S, Van Der Veldt A, Osanto S, et al. Real-world data of nivolumab for patients with advanced renal cell carcinoma in the Netherlands: An analysis of toxicity, efficacy, and predictive markers. *Clin Genitourin Cancer.* 2021;19:274.e1–274.e16.
129. Turner MC, Chen Y, Krewski D, Ghadirian P. An overview of the association between allergy and cancer. *Int J Cancer.* 2006;118:3124–32.
130. Ferastraoraru D, Bax HJ, Bergmann C, Capron M, Castells M, Dombrowicz D, et al. AllergoOncology: ultra-low IgE, a potential novel biomarker in cancer—a Position Paper of the European Academy of Allergy and Clinical Immunology (EAACI). *Clin Transl Allergy.* 2020;10:32.
131. Jaiswal S, Ebert BL. Clonal hematopoiesis in human aging and disease. *Science.* 2019;366:eaan4673.
132. Belizaire R, Wong WJ, Robinette ML, Ebert BL. Clonal haematopoiesis and dysregulation of the immune system. *Nat Rev Immunol.* 2023;23.
133. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med.* 2014;371:2488–98.
134. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371:2477–87.
135. Buttigieg MM, Rauh MJ. Clonal hematopoiesis: Updates and implications at the solid tumor-immune interface. *JCO Precis Oncol.* 2023;7:e2300132.
136. Ea D, Mg T, Ke S, Sm Y, Zi C, Ej M, et al. Familial clonal hematopoiesis in a long telomere syndrome. *N Engl J Med.* 2023;388:2422–33.
137. Telomeres Mendelian Randomization Collaboration, Haycock PC, Burgess S, Nounu A, Zheng J, Okoli GN, et al. Association between telomere length and risk of cancer and non-neoplastic diseases: A Mendelian randomization study. *JAMA Oncol.* 2017;3:636–51.
138. Zhang C, Doherty JA, Burgess S, Hung RJ, Lindström S, Kraft P, et al. Genetic determinants of telomere length and risk of common cancers: A Mendelian randomization study. *Hum Mol Genet.* 2015;24:5356–66.
139. Hizir Z, Bottini S, Grandjean V, Trabucchi M, Repetto E. RNY (YRNA)-derived small RNAs regulate cell death and inflammation in monocytes/macrophages. *Cell Death Dis.* 2017;8:e2530.
140. Driedonks TAP, Mol S, de Bruin S, Peters A-L, Zhang X, Lindenberg MFS, et al. Y-RNA subtype ratios in plasma extracellular vesicles are cell type-specific and are candidate biomarkers for inflammatory diseases. *J Extracell Vesicles.* 2020;9:1764213.

141. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 2005;33:e179.
142. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* 2013;41:D991–5.
143. Palade J, Alsop E, Jensen K, Mateo F, de Cid R, Pujana MA. Analysis of plasma small RNAs prior to breast cancer diagnosis. In: GSE239907, NCBI Gene Expression Omnibus. 2023. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE239907>.
144. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–9.
145. Pujana MA. Study of white blood cell counts in relation to cancer risk. In: UK Biobank Approved Research ID: 61744. 2020. Available from: <https://www.ukbiobank.ac.uk/enable-your-research/approved-research/study-of-white-blood-cell-counts-in-relation-to-cancer-risk>. Accessed 2 Sept 2020.
146. The Genotype-Tissue Expression (GTEx) Consortium. Adult Genotype-Tissue Expression Open Access Datasets. Analysis V8. 2017. Available from: https://www.gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression. Accessed 1 Feb 2022.
147. FANTOM Consortium. FANTOM5 Human Enhancer Tracks. 2014. Available from: https://slidebase.binf.ku.dk/human_enhancers/presets. Accessed 18 May 2023.
148. Triana S, Vonficht D, Jopp-Saile L, Raffel S, Lutz R, Leonce D, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *S-EPMC8642243, BioStudies*; 2021. Available from: <https://www.ebi.ac.uk/biostudies/europepmc/studies/S-EPMC8642243>. Accessed 16 Oct 2022.
149. Smith CL, Eppig JT. Mammalian Phenotype Browser. Immune System Phenotype, MP:0005387. 2022. Available from: https://www.informatics.jax.org/vocab/mp_ontology/MP:0005387. Accessed 25 Oct 2020.
150. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI Catalog of human genome-wide association studies. All associations V1.0. 2021. Available from: <https://www.ebi.ac.uk/gwas/api/search/downloads/full>. Accessed 5 Nov 2021.
151. TCGA Consortium. Genomic Data Commons (GDC) Data Portal. Biospecimen, clinical, and RNA-seq data. 2021. Available from: <https://portal.gdc.cancer.gov/>. Accessed 7 Jan 2020.
152. Speake C, Linsley PS, Whalen E, Chaussabel D, Presnell S, Mason M. Next generation sequencing of human immune cell subsets across diseases. GSE60424, NCBI Gene Expression Omnibus. 2015. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60424>. Accessed 14 July 2023.
153. Yuan T, Huang X, Wang L. Plasma extracellular RNA profiles in healthy and cancer patients. GSE71008, NCBI Gene Expression Omnibus. 2016. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71008>. Accessed 20 Oct 2022.
154. Pardo M, Espin R, Farré X, Esteve A, Pujana MA. Code repository for "Biological basis of extensive pleiotropy between blood traits and cancer risk". GitHub. 2023. Available from: <https://github.com/pujana-lab/PleiotropyBloodCancer>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.