Title

Identification of Pleiotropic Cancer Susceptibility Variants from Genome-Wide Association Studies Reveals Functional Characteristics

Journal

Cancer Epidemiology Biomarkers & Prevention, 27(1)

ISSN

1055-9965

Authors

Wu, Yi-Hsuan
Graff, Rebecca E
Passarelli, Michael N
et al.

Publication Date

2018

DOI

10.1158/1055-9965.epi-17-0516

Peer reviewed

HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2018 January ; 27(1): 75–85. doi:10.1158/1055-9965.EPI-17-0516.

Identification of pleiotropic cancer susceptibility variants from genome-wide association studies reveals functional characteristics

Yi-Hsuan Wu[1], **Rebecca E. Graff**[1], **Michael N. Passarelli**[2], **Joshua D. Hoffman**[1], **Elad Ziv**[3,4,5], **Thomas J. Hoffmann**[1,3], and **John S. Witte**[1,3,5,6]

[1]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California

[2]Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire

[3]Institute for Human Genetics, University of California San Francisco, San Francisco, California

[4]Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, California

[5]Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California

[6]Department of Urology, University of California San Francisco, San Francisco, California

Abstract

Background—There exists compelling evidence that some genetic variants are associated with the risk of multiple cancer sites (i.e., pleiotropy). However, the biological mechanisms through which the pleiotropic variants operate are unclear.

Methods—We obtained all cancer risk associations from the National Human Genome Research Institute-European Bioinformatics Institute GWAS Catalog, and correlated cancer risk variants were clustered into groups. Pleiotropic variant groups and genes were functionally annotated. Associations of pleiotropic cancer risk variants with non-cancer traits were also obtained.

Results—We identified 1,431 associations between variants and cancer risk, comprised of 989 unique variants associated with 27 unique cancer sites. We found 20 pleiotropic variant groups (2.1%) composed of 33 variants (3.3%), including novel pleiotropic variants rs3777204 and rs56219066 located in the *ELL2* gene. Relative to single-cancer risk variants, pleiotropic variants were more likely to be in genes (89.0% versus 65.3%, $p = 2.2\times10^{-16}$), and to have somewhat larger risk allele frequencies (median RAF=0.49 versus 0.39, p=0.046). The 27 genes to which the pleiotropic variants mapped were suggestive for enrichment in response to radiation and hypoxia, alpha-linolenic acid metabolism, cell cycle, and extension of telomeres. In addition, we observed that 8 out of 33 pleiotropic cancer risk variants were associated with 16 traits other than cancer.

Corresponding Author: John S. Witte, Mailing address: 1450 3rd St, San Francisco, CA 94158, Phone: 415-502-6882, jwitte@ucsf.edu.

Conclusions—This study identified and functionally characterized genetic variants showing pleiotropy for cancer risk.

Impact—Our findings suggest biological pathways common to different cancers and other diseases, and provide a basis for the study of genetic testing for multiple cancers and repurposing cancer treatments.

Keywords

Pleiotropy; Genome-wide association Study; GWAS Catalog; Cancer susceptibility; Single nucleotide polymorphism

INTRODUCTION

An emerging focus in cancer research is the discovery and understanding of the shared genetic basis underlying the development of different cancer types. In the past 10 years, genome-wide association studies (GWAS) have identified hundreds of genetic variants associated with cancer risk (1-3), and several loci have been associated with multiple cancer sites. For example, variants at the 8q24 locus have been associated with prostate (4,5), colorectal (6–8), bladder (9), breast (10), and ovarian cancers (11), glioma (12), and chronic lymphocytic leukemia (13,14). The genes closest to this locus are *FAM84B* and *MYC*, both known cancer-related genes. As another example, the 5p15 locus containing *TERT* and *CLPTM1L* is associated with multiple cancer types, including lung (15,16), testicular (17), prostate (18-20), breast (21), colorectal (22) cancers, and glioma (12).

Pleiotropy refers to the phenomenon of a gene or genetic variant affecting more than one phenotypic trait. Identifying and characterizing pleiotropic genes and variants may have important clinical and pharmacological implications (23,24). For example, a drug used for one cancer type may be repurposed to treat another cancer type if the therapeutic target is common to both cancers. In addition, genetic tests for pleiotropic variants may provide an efficient way to identify patients at high risk of multiple cancers. Understanding the functional mechanisms by which variants exhibit pleiotropy is important toward prioritizing potential drug or genetic testing targets.

Recent studies have looked at whether genetic variants previously associated with one cancer are associated with other cancers. Cancers studied in this way include endometrial (25), colorectal (26,27), pancreatic (28), esophageal (29,30), prostate (31), lung (32), ovarian (33), gastric (34), and estrogen receptor negative (ER-) breast cancers (35), and non-Hodgkin lymphoma (36). Cross-cancer GWAS analyses for two to five cancers have also been conducted to identify pleiotropic variants (37-40). Previous work has also estimated the genetic correlation between pairs of cancers using data from GWAS for multiple cancer sites (41,42).

We build on this previous work by investigating pleiotropy across all cancer results presented in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS Catalog (1-3). The GWAS Catalog provides publicly

available, manually curated, and literature-derived single nucleotide polymorphism (SNP)-trait associations with p-values < 10^{-5} from GWAS assessing at least 100,000 SNPs.

Previous analyses of the GWAS Catalog data found substantial evidence of pleiotropy across various traits (43). However, this work did not fully investigate potential pleiotropy arising from variants in linkage disequilibrium (LD) with the associated variants, the functional implications of pleiotropic variants, or the ancestral populations in which the variants were detected. In this study, we addressed these limitations by evaluating variants in LD with the reported variants, functionally characterizing the GWAS reported variants, and incorporating ancestry information. Furthermore, we investigated the associations of the pleiotropic cancer risk variants with other diseases and traits.

MATERIALS AND METHODS

Determining associations with cancer risk in the GWAS Catalog

We accessed all associations reported in the GWAS Catalog as of September 27, 2016. These were mapped to Ensembl release version 85 and contained associations published from March 10, 2005 through October 30, 2015. For associations with any given trait, the data contained the most statistically significant variant from each independent locus for each study. To perform an initial screening for associations with cancer risk, we utilized Experimental Factor Ontology (EFO) terms (release 2016-03-15) (44,45). The curated traits in the GWAS Catalog are mapped to EFO terms to facilitate cross-study comparisons. The initial set of associations we evaluated included traits mapped to the term “neoplasm”, defined as benign or malignant tissue growth resulting from uncontrolled cell proliferation (44). “Neoplasm” and its descendant terms include both cancers and benign tumors (Supplementary Figure S1).

Our goal was to identify variants pleiotropic for cancer susceptibility, so we limited the associations included in our analysis to those specific to the risk of individual cancer types and not to other cancer outcomes. We excluded associations with curated traits containing any of the following terms: “survival”, “recurrence”, “prognosis”, “level”, “symptom”, “toxicity”, “mortality”, “treatment”, “response”, “metastasis”, “aggressiveness”, and “interaction”. We then manually reviewed each of the remaining associations and excluded those reported for gene-gene interactions, non-cancerous traits, and mixed cancer sites (i.e., combining lung, gastric and esophageal cancers). We also excluded associations reported for haplotypes and for variants in the HLA region for which rs number, chromosomal position, and allele name were unavailable.

Associations with the same cancer site but different histological subtypes were categorized as being from the same cancer site. We then calculated the number of variants associated with each cancer site, and the number of studies reporting associations for each cancer site.

Estimating linkage disequilibrium among cancer risk-associated variants

To determine the ancestry of the discovery samples within which associations were identified, we relied on data provided by the GWAS Catalog, which assigned one of 15 ancestral categories to each association. We collapsed categories by the 1000 Genomes

Project's super populations (European [EUR], East Asian [EAS], Ad Mixed American [AMR], African [AFR], and South Asian [SAS]). The number of cancer risk associations for each super population was calculated. For associations without ancestry data reported in the Catalog, we reviewed the original publications to obtain the ancestry information.

Our goal was to identify the following two types of cancer risk variants: 1) pleiotropic within ethnic group, and 2) pleiotropic across ethnic groups. To identify the first type, we estimated pairwise LD among variants discovered in the same super population using reference genotype data from the corresponding super population. To identify the second type, we estimated pairwise LD among all variants, regardless of the discovery population, using reference genotype data from each of the five 1000 Genomes Project's (46) super populations individually.

We ensured that all rs numbers were updated to build 142 of the Single Nucleotide Polymorphism Database (dbSNP; http://www.ncbi.nlm.nih.gov/projects/SNP/) (47). For variants lacking rs numbers in the original publications, we used chromosomal positions and the UCSC Genome Browser (48) to obtain rs numbers. LD was estimated with LDlink (http://analysistools.nci.nih.gov/LDlink/) (49), which uses genotype data from Phase 3 of the 1000 Genomes Project and variant rs numbers indexed based on dbSNP build 142. HaploReg v4.1 (http://www.broadinstitute.org/mammals/haploreg) (50,51) was used to evaluate LD for variants that could not be assessed by LDlink. We were unable to calculate LD for variants that were monoallelic in a given population and/or not in the 1000 Genomes data.

Identifying variants associated with the risk of multiple cancer sites

First, to identify variants pleiotropic within the same ethnic group, we grouped variants based on LD estimated in each super population. Second, to identify variants pleiotropic across ethnic groups, we grouped all variants based on LD estimated in each of the five super populations; doing so yielded five different sets of variant groupings.

We took two steps to group cancer risk variants in high LD: 1) variant pairs with $R^2 \geq 0.8$ were clustered into variant groups, and 2) variant groups sharing at least one variant were merged. Therefore, within each variant group, each variant was in LD with at least one other variant (e.g., Supplementary Figure S2). A variant group was defined as pleiotropic if it was associated with the risk of more than one cancer site ($p < 10^{-5}$). Single-cancer variant groups were associated with the risk of only one cancer site. In sensitivity analyses, we explored variant groupings based on different levels of LD (R^2 of ≥ 0.7 or ≥ 0.6).

We calculated the median and interquartile range (IQR) of the association odds ratios and risk allele frequencies (RAFs) for pleiotropic and single-cancer variants. Since these were not normally distributed, we mainly compared them using the Wilcoxon test.

Functional annotations of variants and genes pleiotropic for cancer risk

For variant-level functional annotation, we first used HaploReg v4.1 (http://compbio.mit.edu/HaploReg) (50,51) to obtain all variants in strong LD ($R^2 \geq 0.8$) with the pleiotropic risk variants (based on the 1000 Genomes Phase 1 European (EUR) population).

The locations of and consequences on protein sequences for these variants were determined using the Ensembl Variant Effect Predictor (VEP) (52). We picked consequence types for each variant using two different options in VEP. The "–most_severe" option was used to select only the most severe consequence (Supplementary Table S1). The "–pick" option was used to select one or more consequences according to an ordered set of criteria (Supplementary Table S2). We grouped the consequences into three main categories: "gene variant", "intergenic variant", and "regulatory region variant". In addition, we did variant group-level functional annotation, and the "–most_severe" and "–pick" categories were used to select one consequence type per variant group.

Functional annotation was also performed on the gene-level. Genic cancer risk variants and all other variants in strong LD ($R^2 \geq 0.8$) were mapped to RefSeq genes using HaploReg v4.1(50,51). We used the Gene ID Conversion Tool in DAVID (http://david.ncifcrf.gov/) (53,54) to convert RefSeq Accession to Entrez Gene ID. Overrepresentation of pleiotropic genes in biological processes based on Gene Oncology (GO) was tested using ConsensusPathDB (55). Overrepresentation tests comparing pleiotropic genes in Reactome (release 2016-12-07) (56,57) pathways were conducted using the PANTHER Overrepresentation Test (release 2017-04-13) (58).

Assessing associations between pleiotropic cancer risk variants and other traits

As above, we used LDlink or HaploReg to obtain all variants in strong LD ($R^2 \geq 0.8$) with the pleiotropic cancer risk variants. LD was based on the 1000 Genomes Project super population that reflected the discovery sample of the variant. These variants were searched in the GWAS Catalog to identify associations with traits other than cancer risk.

RESULTS

Summary of cancer risk associations in the GWAS Catalog

We evaluated the 28,643 associations with $p < 10^{-5}$ published in the GWAS Catalog, and identified 1,711 that mapped to one of 1,395 relevant EFO terms (i.e., "neoplasm" or its descendants) (Figure 1). Among the 1,711 associations, we excluded 171 that did not address susceptibility (e.g., gene-gene and gene-environment interactions, survival, and aggressiveness). After manually reviewing each of the remaining associations, we further excluded 85 with non-cancerous traits (e.g., cutaneous nevi, percent mammographic density), two associations with mixed cancer sites (both combining lung, gastric and esophageal cancers) (39), 20 associations with haplotypes, and two associations missing rs number, chromosomal position, and name of HLA allele. Ultimately, 1,431 cancer risk associations with $p < 10^{-5}$ (927 with $p < 5 \times 10^{-8}$) formed by 989 variants were identified from 227 studies.

The associations were grouped into 27 cancer sites (Table 1). The number of variants associated with prostate or breast cancer was almost twice the number of variants associated with all other individual cancer sites, partially reflecting the larger number of GWAS conducted for these two cancer sites. Other cancers with more than 50 associated variants were leukemia, lymphoma, and colorectal, pancreatic, skin, and lung cancers.

Cancer risk associations were discovered from 12 different populations. We observed that 993 (69.4%) associations were identified in an initial sample of Europeans, 250 (17.5%) were from East Asians, and 98 (6.8%) were from a sample containing European, East Asian, Hispanic, and African ancestries (Supplementary Table S3). The variation in these percentages reflects the differences in how many people from each of these populations have been included in GWAS.

Genetic variants showing pleiotropy for cancer risk

Among the 939 variant groups obtained using $R^2 \geq 0.8$ as a threshold (Supplementary Table S3), 20 (2.1%) exhibited pleiotropy for cancer risk within the same ethnic group (Table 2). We confirmed that all grouped variants had high LD. In particular, within the 20 pleiotropic variant groups that we identified, the lowest LD between any two variants was $R^2 = 0.735$.

Of the 20 pleiotropic variant groups, 17 variant groups were from European populations and three variant groups were from East Asian populations. The 20 pleiotropic variant groups were composed of 33 (3.3%) variants, and the remaining 956 (96.7%) cancer risk variants were classified as single-cancer variants. The pleiotropic variants are located in 27 genes such as *MDM4*, *ELL2*, *MLLT10*, *BCL2*, *BRCA2*, *BABAM1* and *ANKLE1*. We observed that the cancer risk associations were similar for pleiotropic variants (median association odds ratio=1.26; interquartile range [IQR]=1.15-1.27) and for single-cancer risk variants (median association odds ratio=1.23; IQR=1.14-1.38) (Wilcoxon test p = 0.15). However, the pleiotropic variants had slightly higher risk allele frequency (median RAF=0.49; IQR=0.30-0.54) than observed for single-cancer risk variants (median RAF=0.39; IQR=0.21-0.59) (Wilcoxon test p = 0.046) (Supplementary Table S4).

Additionally, we clustered variant groups according to different LD thresholds (Supplementary Table S3). Using a threshold of $R^2 \geq 0.7$, we identified one additional variant group (21 total groups; 41 variants) showing pleiotropy for cancer risk within the same ethnic group. Using a threshold of $R^2 \geq 0.6$, we identified yet one more additional variant group (22 total groups; 48 variants).

Variants were also grouped regardless of the discovery samples to identify those which were pleiotropic across ethnic groups. Using $R^2 \geq 0.8$ as the threshold, we identified 9 variant groups (18 variants) pleiotropic for cancer risk across ethnic groups (Supplementary Tables S5 and S6). The lower the R^2 threshold used for variant grouping, the more pleiotropic variants we identified. Overall, approximately 2-4% of variant groups (3–7% of variants) were pleiotropic (Supplementary Figure S3).

Functional characterizations of pleiotropic cancer risk variants and genes

As variants reported in the GWAS Catalog may not be causal but rather tag the true causal variants, we incorporated variants in LD when performing functional annotations to try to capture as much information as possible. We identified 518 variants in strong LD ($R^2 \geq 0.8$) with the 33 pleiotropic cancer risk variants, and 18,069 variants in strong LD with the 956 single-cancer variants. We observed that the most severe consequences were statistically different between pleiotropic and single-cancer variants (Fisher's exact test $p = 2.2 \times 10^{-16}$; Table 3). A higher percentage of pleiotropic cancer risk variants were genic (89.0%)

compared to single-cancer variants (65.3%). Within genes, most of the pleiotropic variants were in introns, 3' untranslated regions (UTRs), or they changed exon sequence in a non-coding transcript. Pleiotropic cancer risk variants were also less likely to be intergenic (11.0%) compared to single-cancer variants (31.8%) and more likely to be upstream of genes (7.4% vs. 5.3%). Interestingly, none of the pleiotropic cancer risk variants were located in regulatory regions such as transcription factor binding sites or other non-genic regions (0% vs. 2.9%). Selecting consequence according to an ordered set of criteria ("–pick" option) provided similar percentages of gene and intergenic variants (Supplementary Table S7). Likewise, annotations on variant-group level showed that pleiotropic variant groups tended to be in genes more often than in intergenic regions (Supplementary Tables S8 and S9).

For the genic variants, the 460 pleiotropic ones mapped to 27 genes, while the 11,755 single-cancer variants mapped to 612 genes. Relative to single cancer genes, pleiotropic genes had suggestive enrichment ($p < 0.007$) in the following: response to stimuli such as light, radiation, oxygen and organic cyclic compounds, cell aging, and directing movement of a protein to a specific location on a chromosome (Supplementary Table S10). Although not statistically significant after correction for multiple testing, the most overrepresented pathways for pleiotropic genes included alpha-linolenic acid (ALA) metabolism, cell cycle, and extension of telomeres (Supplementary Table S11).

Associations between pleiotropic cancer risk variants and traits other than cancer

Detecting the associations between pleiotropic cancer risk variants and non-cancer traits has the potential to suggest shared underlying biology across different traits, which may reflect common underlying mechanisms (e.g., inflammation). We found that 8 out of 33 pleiotropic cancer risk variants that we identified were associated with 16 other complex diseases or traits investigated by GWAS (Figure 2). Variants rs10936599 and rs12696304 located in *MYNN* and near *TERC* were associated with telomere length, celiac disease, and multiple sclerosis. Variant rs2736100 in *TERT* was associated with telomere length, red blood cell count, and lung diseases. The *CLPTM1L* gene variants rs401681, rs31489, rs31490, and rs4975616 were associated with serum prostate-specific antigen (PSA) levels. Variant rs2294008 located in *PSCA* gene was associated with duodenal ulcers. Pleiotropic variants rs174537 in *MYRF* and rs174549 in *FADS1* were associated with many lipid metabolism-related traits. Variants rs4430796 and rs8064454 in *HNF1B* were also associated with type 2 diabetes and PSA levels. The 672 variants in strong LD ($R^2 \geq 0.8$) with the 33 pleiotropic cancer risk variants were associated with an additional 24 traits (Supplementary Table S12).

DISCUSSION

There is considerable interest in cancer pleiotropy as it may highlight important molecular mechanisms and have implications for drug development. Our analysis across 27 cancer sites using the publicly available NHGRI-EBI GWAS Catalog detected numerous pleiotropic cancer risk variants and evaluated their functional characteristics.

We uncovered some novel patterns of pleiotropy for known cancer risk loci. Our study is the first to highlight that variants in *MLLT10* at 10p12.31 are associated with both ovarian

cancer and meningioma. *MLLT10* is known to encode a transcription factor involved in chromosomal rearrangements in leukemia (59). Nonetheless, there is currently no direct evidence for how *MLLT10* is involved in developing ovarian cancer or meningioma other than GWAS. Another novel pattern of pleiotropy that we found was that the variant rs4245739 in *MDM4* at 1q32.1 is associated with prostate cancer and ER-negative and triple-negative breast cancer. *MDM4* encodes a repressor that binds and inactivates p53 and is considered important in cancer development (60,61). Interestingly, we observed that the association of the C allele of rs4245739 is in the opposite direction for prostate (OR = 0.91, 95% CI: 0.88, 0.95) (62) and breast cancer (OR = 1.14, 95% CI: 1.10, 1.18) (63). Based on the Genotype-Tissue Expression (GTEx) Project (64), rs4245739 is not associated with expression in prostate or breast tissue, but is correlated with *PIK3C2B* expression in testis. *PIK3C2B* encodes a phosphoinositide 3-kinase that plays a role in many oncogenic pathways.

Our analysis of the GWAS Catalog also identified the novel pleiotropic gene *ELL2*. The variant rs3777204 (exm2265979 in the original publication) is associated with salivary gland carcinoma and rs56219066 is associated with multiple myeloma in European populations ($R^2 = 0.971$ between the two variants). *ELL2* encodes an elongation factor for RNA polymerase II, an important component of the super elongation complex (SEC) (65) that regulates the transcriptional elongation checkpoint control (TECC) stage of transcription. Dysregulation is related to carcinogenesis (66). We also found that variant rs56219066 is associated with a reduction of IgA and IgG levels, which could affect the pre-mRNA processing and malignant transformation in multiple myeloma (67).

We also replicated a number of previously known pleiotropic variants and loci. Variants in the *TERC-MYNN* region at 3q26.2 and *TERT-CLPTM1L* region at 5p15.33 are pleiotropic for many cancers (68,69). The telomerase reverse transcriptase (TERT) (70) and its integral RNA template (TERC) (71) are two subunits of the telomerase ribonucleoprotein complex that maintain telomere length. Variant rs2736100 in *TERT* is associated with lung and testicular cancers and glioma in European populations, but only the lung cancer association in East Asian populations was identified. To our knowledge, only one candidate gene study found an association between rs2736100 and glioma risk in East Asians ($p = 3.69 \times 10^{-4}$) (72). The association between rs2736100 and testicular cancer is in the opposite direction for lung cancer and glioma. It has been speculated that rs2736100-T mediates the recruitment of sex determining region Y transcription factor to *TERT*, which might increase the telomerase in germ cells, leading to a potential increased risk of testicular cancer (73). As another example, *BRCA2* at 13q13.1 is a well-known pleiotropic gene (74-76). Variants in the *BABAM1-ANKLE1* region at 19p13.11 are associated with breast and ovarian cancer. The *BABAM1* gene encodes a component of the *BRCA1* complex and *BRCA1* activates DNA repair in double-strand breaks (DSB) in cooperation with *BRCA2* (77); defective repair in DSB can lead to tumorigenesis (78). Since mutations in *BRCA1* affect both breast and ovarian cancer risk, the association of *BABAM1* with both of these cancers is consistent with its known interaction with *BRCA1*. Finally, the 8q24 region is associated with diverse cancers, and the *HNF1B* gene variants at 17q12 are specifically associated with hormone-related cancers, including prostate, endometrial, ovarian and testicular cancers.

Previous work evaluating the associations in the GWAS Catalog through 2011 estimated that 4.8% of SNPs associated with cancer exhibit pleiotropy (43). There are several possible reasons why we detected a lower level of pleiotropy (2.1% variant groups [3.3% variants]) within ethnic groups. First, the previous study included SNPs pleiotropic across ethnic groups. Second, we estimated LD using the populations in which the variants were discovered, while the previous study used the HapMap CEU population to calculate LD for all variants. Finally, we focused specifically on cancer risk instead of all cancer outcomes. The evaluation of pleiotropy for treatment response is certainly compelling, but it is also very complicated. One must decide how to consider treatment type, efficacy, toxicity, and time to response, among other dimensions. To maintain focus and comparability with previous research, we elected to only evaluate here cancer susceptibility. Future work exploring pleiotropy specific to the treatment of cancer can be performed.

Spurious cross-phenotype associations can occur when individuals suffering from one cancer are more likely to receive diagnostic evaluation and detection of other cancers (ascertainment bias) (23,79). In our study, this potential bias was minimized because the associations studied generally came from distinct GWAS in which few subjects had multiple cancers. Nevertheless, the frequency of pleiotropy that we observed depended upon the associations collected in the GWAS Catalog, and the traits that we selected to study were not a random sample of all traits. In addition, the frequency of pleiotropy is larger than we estimated because many common and rare variants associated with cancer risk have yet to be found. That said, utilizing the EFO terms that map the traits increased the sensitivity of detecting cancer risk associations, while manually reviewing each of the identified associations avoided including non-cancer risk outcomes.

We estimated that 89.0% of pleiotropic variants were in genes, in contrast with only 65.3% of non-pleiotropic variants. Pleiotropic variants may be more likely to be within a gene than to affect gene regulation because gene regulation is highly tissue specific. Thus, for a variant to affect risk in multiple tissues, it needs to affect the function of a gene in a way that transcends tissue specific gene regulation in regulatory elements.

The suggestion of overrepresentation of the pathways response to radiation and hypoxia, ALA metabolism, cell cycle, and extension of telomeres in pleiotropic genes implies carcinogenic mechanisms common to the development of different cancer sites. The comparison group for these functional enrichment analyses was the single-cancer variants. Misclassification of truly pleiotropic variants as observed single-cancer variants is possible if associations with other cancer sites have not yet been detected (e.g., due to limited power). Future GWAS or sequencing focused on rarer variants, cancer subtypes, or epigenetic regulations may help identify additional novel pleiotropic mechanisms.

Some of our findings of pleiotropy are in agreement with current clinical practices. For example, we observed that two highly correlated variants ($R^2 = 0.95$) located in or near the *BCL2* gene, rs17749561 and rs4987855, were associated with follicular lymphoma (FL) and chronic lymphocytic leukemia (CLL). The *BCL2* gene family encodes proteins that regulate cellular apoptosis, and serve essential roles of balancing cell survival and cell death (80,81). The drug Venetoclax (also known as ABT-199) is a BCL-2 inhibitor that binds to BCL-2

with high affinity and selectivity (82). It was approved for CLL (83) but also shows favorable efficacy and safety in patients with FL (84), reflecting the pleiotropic role of *BCL2* in these two blood cancers. Our findings also suggest potential clinical applications. For example, variants in the novel pleiotropic gene *ELL2* that we identified were associated with salivary gland carcinoma and multiple myeloma. Expression of *ELL2* was down-regulated by microRNAs miR-155 (85) and miR-299 (86). With the extensive development of microRNA therapeutics (87), *ELL2* exhibits a promising candidate to treat these two cancers. Furthermore, our discovery of variants that exhibit opposite effects on cancers (e.g., variants in *MDM4* or in *TERT*) may help identify drugs that should not be explored for repurposing across cancers.

Our approach was limited by the selection of variants for genotyping arrays and the imputation reference panels in the published GWAS. Genotyping arrays may have preferentially included variants located in genes or previously associated with diseases, giving such variants an increased chance of being pleiotropic. To try to avoid this potential bias, we compared the pleiotropic variants to single-cancer variants, which were also tagged by GWAS arrays. We were also limited in our ability to distinguish biological from spurious pleiotropy. It is possible that variants in high LD could be functional in different genes. Our study also had potential bias in that some cancers have more than one ethnicity represented in GWAS and these cancers may be more or less likely to have shared genetic causes. The absolute risk of cancer affects how many GWAS are performed and their power, so pleiotropic variants may appear to cluster for common cancers, which might also have more ethnicities involved. In addition, we were only able to evaluate pleiotropy among loci with strong enough associations to be reported in the GWAS Catalog. As all summary statistics from GWAS become more widely available, more extensive evaluations of pleiotropy can be undertaken.

Overall, identification of pleiotropic cancer risk variants and genes has important implications. The biological functions and pathways overrepresented in pleiotropic genes may inform our understanding of the underlying mechanisms shared by different cancers. Genetic tests for pleiotropic variants could be developed to efficiently identify high-risk patients. Drugs developed for one cancer might be valuable for use in treating other cancers as well. Such repurposing of anti-cancer therapies may offer promising opportunities for improving cancer treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support

This work supported by NIH grants CA127298, CA088164, CA112355, and CA201358 (REG, MNP, JDH, JSW), and the UCSF Goldberg-Benioff Program in Cancer Translational Biology (JSW).

References

1. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. 2009; 106(23):9362–7.
2. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research. 2016:gkw1133.
3. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic acids research. 2014; 42(D1):D1001–D6. [PubMed: 24316577]
4. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nature genetics. 2008; 40(3):310–5. [PubMed: 18264096]
5. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nature genetics. 2007; 39(5):645–9. [PubMed: 17401363]
6. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24. 21. Nature genetics. 2007; 39(8):984–8. [PubMed: 17618284]
7. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nature genetics. 2007; 39(8):989–94. [PubMed: 17618283]
8. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nature genetics. 2008; 40(5):631–7. [PubMed: 18372901]
9. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nature genetics. 2010; 42(11):978–84. [PubMed: 20972438]
10. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447(7148): 1087–93. [PubMed: 17529967]
11. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature genetics. 2010; 42(10):874–9. [PubMed: 20852632]
12. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, et al. Genome-wide association study identifies five susceptibility loci for glioma. Nature genetics. 2009; 41(8):899–904. [PubMed: 19578367]
13. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37. *3*, 8q24. 21, 15q21. 3 and 16q24. 1 influence chronic lymphocytic leukemia risk. Nature genetics. 2010; 42(2):132–6. [PubMed: 20062064]
14. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. Journal of the National Cancer Institute. 2008; 100(13):962–6. [PubMed: 18577746]
15. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, et al. Lung cancer susceptibility locus at 5p15. 33. Nature genetics. 2008; 40(12):1404–6. [PubMed: 18978790]
16. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15. 33 and 6p21. 33 variants influence lung cancer risk. Nature genetics. 2008; 40(12):1407–9. [PubMed: 18978787]
17. Turnbull C, Rapley EA, Seal S, Pernet D, Renwick A, Hughes D, et al. Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. Nature genetics. 2010; 42(7): 604–7. [PubMed: 20543847]

18. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Dadaev T, Jugurnauth-Little S, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. Human molecular genetics. 2013; 22(12):2520–8. [PubMed: 23535824]
19. Kote-Jarai Z, Al Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nature genetics. 2011; 43(8):785–91. [PubMed: 21743467]
20. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nature genetics. 2009; 41(2):221–7. [PubMed: 19151717]
21. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nature genetics. 2011; 43(12):1210–4. [PubMed: 22037553]
22. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Human genetics. 2012; 131(2):217–34. [PubMed: 21761138]
23. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics. 2013; 14(7):483–95.
24. Manolio TA. Bringing genome-wide association findings into clinical use. Nature Reviews Genetics. 2013; 14(8):549–58.
25. Setiawan VW, Schumacher F, Prescott J, Haessler J, Malinowski J, Wentzensen N, et al. Cross-cancer pleiotropic analysis of endometrial cancer: PAGE and E2C2 consortia. Carcinogenesis. 2014:bgu107.
26. Cheng I, Kocarnik JM, Dumitrescu L, Lindor NM, Chang-Claude J, Avery CL, et al. Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. Gut. 2013 gutjnl-2013-305189.
27. Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins S, Casey G, et al. The TERT variant rs2736100 is associated with colorectal cancer risk. British journal of cancer. 2012; 107(6):1001–8. [PubMed: 22878375]
28. Pierce BL, Ahsan H. Genome-wide “pleiotropy scan” identifies HNF1A region as a novel pancreatic cancer susceptibility locus. Cancer research. 2011; 71(13):4352–8. [PubMed: 21498636]
29. Lee E, Stram DO, Ek WE, Onstad LE, MacGregor S, Gharahkhani P, et al. Pleiotropic analysis of cancer risk loci on esophageal adenocarcinoma risk. Cancer Epidemiology and Prevention Biomarkers. 2015; 24(11):1801–3.
30. Chang J, Wei L, Miao X, Yu D, Tan W, Zhang X, et al. Two novel variants on 13q22. 1 are associated with risk of esophageal squamous cell carcinoma. Cancer Epidemiology and Prevention Biomarkers. 2015; 24(11):1774–80.
31. Panagiotou OA, Travis RC, Campa D, Berndt SI, Lindstrom S, Kraft P, et al. A genome-wide pleiotropy scan for prostate cancer risk. European urology. 2015; 67(4):649–57. [PubMed: 25277271]
32. Park SL, Fesinmeyer MD, Timofeeva M, Caberto CP, Kocarnik JM, Han Y, et al. Pleiotropic associations of risk variants identified for other cancers with lung cancer risk: the PAGE and TRICL consortia. Journal of the National Cancer Institute. 2014; 106(4):dju061. [PubMed: 24681604]
33. Li D-K, Han J, Liu J-B, Jin G-F, Qu J-W, Zhu M, et al. Genetic variants at 6p21. 1 and 7p15. 3 Identified by GWASs of multiple cancers and ovarian cancer risk: a case-control study in Han Chinese women. Asian Pac J Cancer Prev. 2014; 15:123–7. [PubMed: 24528012]
34. Du J, Xu Y, Dai J, Ren C, Zhu C, Dai N, et al. Genetic variants at 5p15 are associated with risk and early onset of gastric cancer in Chinese populations. Carcinogenesis. 2013; 34(11):2539–42. [PubMed: 23901064]
35. Campa D, Barrdahl M, Tsilidis KK, Severi G, Diver WR, Siddiq A, et al. A genome-wide “pleiotropy scan” does not identify new susceptibility loci for estrogen receptor negative breast cancer. PloS one. 2014; 9(2):e85955. [PubMed: 24523857]

36. Lim U, Kocarnik JM, Bush WS, Matise TC, Caberto C, Park SL, et al. Pleiotropy of cancer susceptibility variants on the risk of non-Hodgkin lymphoma: the PAGE consortium. PLoS one. 2014; 9(3):e89791. [PubMed: 24598796]
37. Cheng TH, Thompson D, Painter J, O'Mara T, Gorman M, Martin L, et al. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. Scientific reports. 2015; 5
38. Fehringer G, Kraft P, Pharoah PD, Eeles RA, Chatterjee N, Schumacher FR, et al. Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations. Cancer research. 2016; 76(17):5103–14. [PubMed: 27197191]
39. Jin G, Ma H, Wu C, Dai J, Zhang R, Shi Y, et al. Genetic variants at 6p21. 1 and 7p15. 3 are associated with risk of multiple cancers in Han Chinese. The American Journal of Human Genetics. 2012; 91(5):928–34. [PubMed: 23103227]
40. Law PJ, Sud A, Mitchell JS, Henrion M, Orlando G, Lenive O, et al. Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci. Scientific reports. 2017; 7:41071. [PubMed: 28112199]
41. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, et al. Analysis of heritability and shared heritability based on genome-wide association studies for 13 cancer types. Journal of the National Cancer Institute. 2015; 107(12):djv279. [PubMed: 26464424]
42. Lindström S, Finucane H, Bulik-Sullivan B, Schumacher FR, Amos CI, Hung RJ, et al. Quantifying the genetic correlation between multiple cancer types. Cancer Epidemiology and Prevention Biomarkers. 2017 cebp. 0211.2017.
43. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. The American Journal of Human Genetics. 2011; 89(5):607–18. [PubMed: 22077970]
44. Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics. 2010; 26(8):1112–8. [PubMed: 20200009]
45. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic acids research. 2011; 39(suppl 2):W541–W5. [PubMed: 21672956]
46. Consortium GP. A global reference for human genetic variation. Nature. 2015; 526(7571):68. [PubMed: 26432245]
47. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001; 29(1):308–11. [PubMed: 11125122]
48. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome research. 2002; 12(6):996–1006. [PubMed: 12045153]
49. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015:btv402.
50. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40(D1):D930–D4. [PubMed: 22064851]
51. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic acids research. 2016; 44(D1):D877–D81. [PubMed: 26657631]
52. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome biology. 2016; 17(1):122. [PubMed: 27268795]
53. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009; 4(1):44–57. [PubMed: 19131956]
54. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009; 37(1):1–13. [PubMed: 19033363]

55. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic acids research. 2013; 41(D1):D793–D800. [PubMed: 23143270]
56. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic acids research. 2014; 42(D1):D472–D7. [PubMed: 24243840]
57. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The reactome pathway knowledgebase. Nucleic acids research. 2016; 44(D1):D481–D7. [PubMed: 26656494]
58. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome research. 2003; 13(9):2129–41. [PubMed: 12952881]
59. Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J, et al. A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t (10; 11) translocation in acute leukemia. Blood. 1995; 85(6):1435–41. [PubMed: 7888665]
60. Shvarts A, Steegenga W, Riteco N, Van Laar T, Dekker P, Bazuine M, et al. MDMX: a novel p53-binding protein with some functional properties of MDM2. The EMBO journal. 1996; 15(19): 5349. [PubMed: 8895579]
61. Shvarts A, Bazuine M, Dekker P, Ramos YF, Steegenga WT, Merckx G, et al. Isolation and identification of the human homolog of a new p53-binding protein, Mdmx. Genomics. 1997; 43(1):34–42. [PubMed: 9226370]
62. Eeles RA, Al Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nature genetics. 2013; 45(4)
63. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN. Genome-wide association studies identify four ER negative–specific breast cancer risk loci. Nature genetics. 2013; 45(4):392. [PubMed: 23535733]
64. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. Nature genetics. 2013; 45(6):580–5. [PubMed: 23715323]
65. Liu M, Hsu J, Chan C, Li Z, Zhou Q. The ubiquitin ligase Siah1 controls ELL2 stability and formation of super elongation complexes to modulate gene transcription. Molecular cell. 2012; 46(3):325–34. [PubMed: 22483617]
66. Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. Nature reviews Molecular cell biology. 2012; 13(9):543–7. [PubMed: 22895430]
67. Swaminathan B, Thorleifsson G, Jöud M, Ali M, Johnsson E, Ajore R, et al. Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. Nature communications. 2015; 6
68. Campa D, Rizzato C, Stolzenberg-Solomon R, Pacetti P, Vodicka P, Cleary SP, et al. TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. International journal of cancer. 2015; 137(9):2175–83. [PubMed: 25940397]
69. Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. Nature Reviews Cancer. 2010; 10(5):353–61. [PubMed: 20414203]
70. Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, et al. Telomerase catalytic subunit homologs from fission yeast and human. Science. 1997; 277(5328):955–9. [PubMed: 9252327]
71. Feng J, Funk WD, Wang S-S, Weinrich SL. The RNA component of human telomerase. Science. 1995; 269(5228):1236. [PubMed: 7544491]
72. Chen H, Chen Y, Zhao Y, Fan W, Zhou K, Liu Y, et al. Association of sequence variants on chromosomes 20, 11, and 5 (20q13. *33*, 11q23. 3, and 5p15. 33) with glioma susceptibility in a Chinese population. American journal of epidemiology. 2011; 173(8):915–22. [PubMed: 21350045]
73. Kinnersley B, Migliorini G, Broderick P, Whiffin N, Dobbins S, Casey G, et al. The TERT variant rs2736100 is associated with colorectal cancer risk. British journal of cancer. 2012; 107(6):1001. [PubMed: 22878375]
74. Cavanagh H, Rogers KM. The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. Hereditary cancer in clinical practice. 2015; 13(1):16. [PubMed: 26236408]

75. Antoniou A, Pharoah P, Narod S, Risch HA, Eyfjord JE, Hopper J, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. The American Journal of Human Genetics. 2003; 72(5):1117–30. [PubMed: 12677558]
76. King M-C, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003; 302(5645):643–6. [PubMed: 14576434]
77. Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer science. 2004; 95(11):866–71. [PubMed: 15546503]
78. Helleday T, Lo J, van Gent DC, Engelward BP. DNA double-strand break repair: from mechanistic understanding to cancer treatment. DNA repair. 2007; 6(7):923–35. [PubMed: 17363343]
79. Smoller JW, Lunetta KL, Robins J. Implications of comorbidity and ascertainment bias for identifying disease genes. American Journal of Medical Genetics Part A. 2000; 96(6):817–22.
80. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nature reviews Molecular cell biology. 2014; 15(1):49. [PubMed: 24355989]
81. Shamas-Din A, Brahmbhatt H, Leber B, Andrews DW. BH3-only proteins: Orchestrators of apoptosis. Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research. 2011; 1813(4):508–20. [PubMed: 21146563]
82. Souers AJ, Leverson JD, Boghaert ER, Ackler SL, Catron ND, Chen J, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature medicine. 2013; 19(2):202–8.
83. Stilgenbauer S, Eichhorst B, Schetelig J, Coutre S, Seymour JF, Munir T, et al. Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. The Lancet Oncology. 2016; 17(6):768–78. [PubMed: 27178240]
84. Gerecitano JF, Roberts AW, Seymour JF, Wierda WG, Kahl BS, Pagel JM, et al. A phase 1 study of venetoclax (ABT-199/GDC-0199) monotherapy in patients with relapsed/refractory non-Hodgkin lymphoma. Am Soc Hematology. 2015
85. Bhattacharyya S, Balakathiresan NS, Dalgard C, Gutti U, Armistead D, Jozwik C, et al. Elevated miR-155 promotes inflammation in cystic fibrosis by driving hyperexpression of interleukin-8. Journal of Biological Chemistry. 2011; 286(13):11604–15. [PubMed: 21282106]
86. Huang Q, Zhang X-W, Ma Y-S, Lu G-X, Xie R-T, Yang H-Q, et al. Up-regulated microRNA-299 corrected with poor prognosis of glioblastoma multiforme patients by targeting ELL2. Japanese Journal of Clinical Oncology. 2017:1–7.
87. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nature Reviews Drug Discovery. 2017; 16(3):203–22. [PubMed: 28209991]

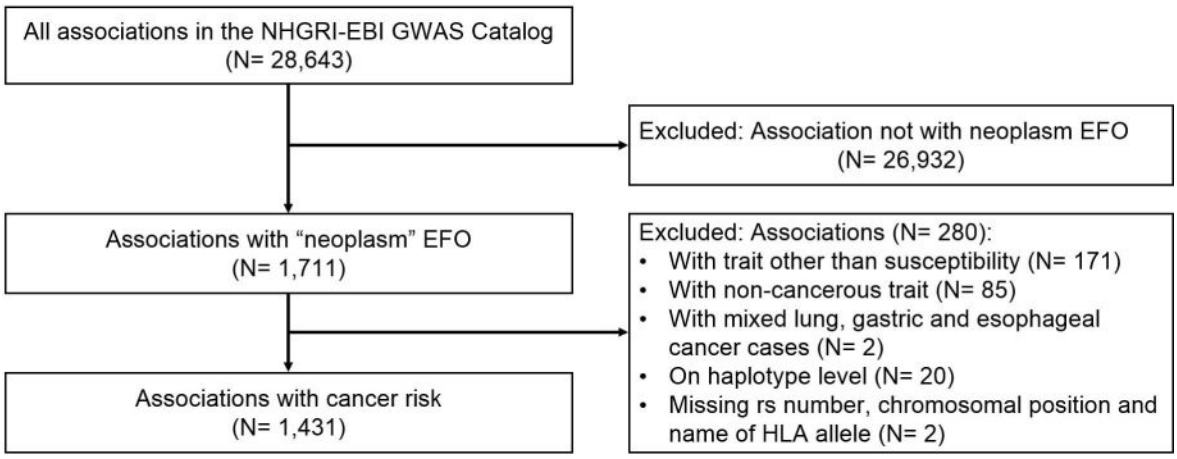

Figure 1.

Flow diagram for obtaining all cancer risk associations. Based on the 1,395 Experimental Factor Ontology (EFO) "neoplasm" terms, we identified 1,711 associations with "neoplasm". Following exclusions, we obtained 1,431 cancer risk associations.

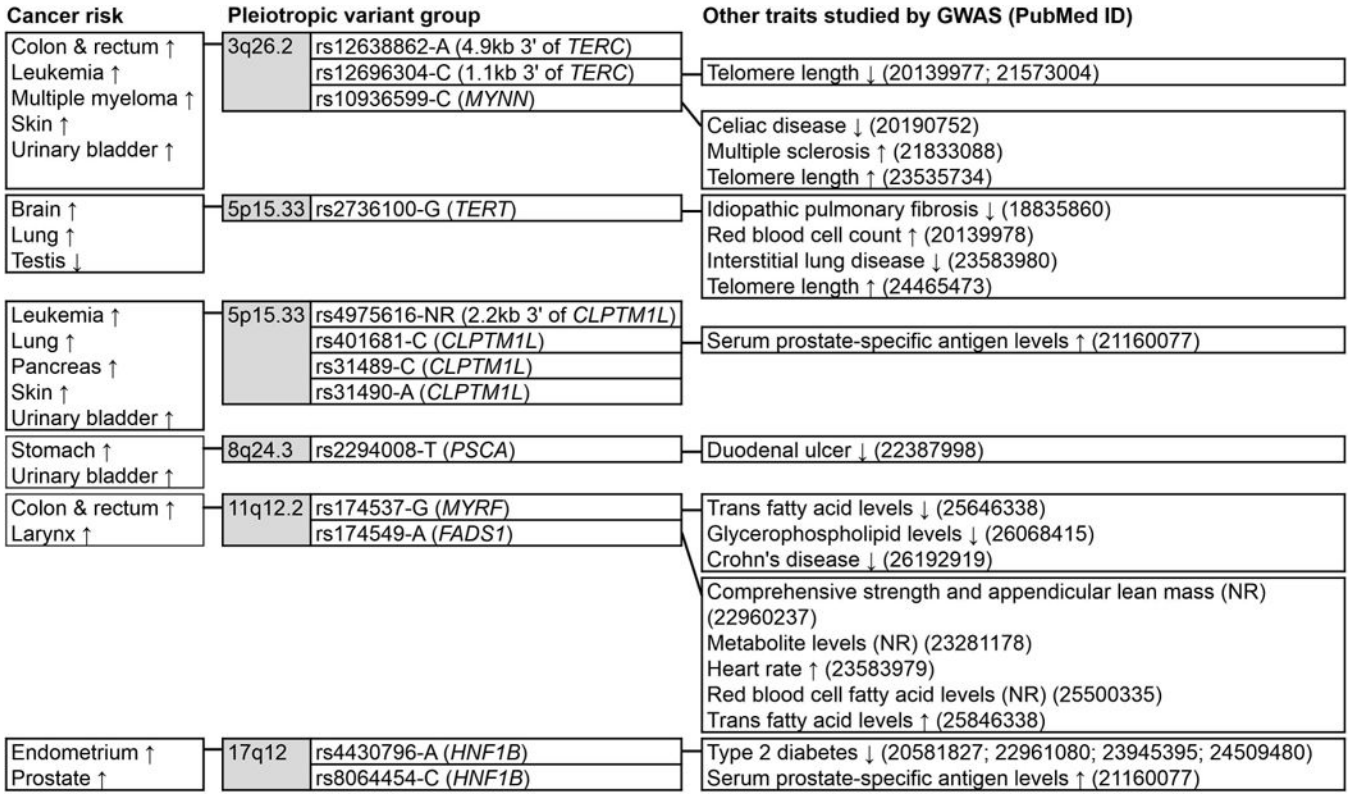

Figure 2.

Associations of pleiotropic cancer risk variants with other complex diseases and traits studied by GWAS. We identified 8 out of 33 pleiotropic variants were associated with other 16 distinct traits. The arrow indicates the direction of the association. NR: Information not reported in the GWAS Catalog.

Table 1

Number of variants and GWAS for each of the 27 cancer sites.

Cancer site	Number of variants (studies)	Cancer site	Number of variants (studies)
Prostate	166 (24)	Multiple myeloma	45 (4)
Breast	145 (29)	Stomach	17 (5)
Leukemia	95 (13)	Brain	15 (6)
Colon & rectum	81 (18)	Nasopharynx	14 (4)
Pancreas	59 (6)	Liver	10 (5)
Skin	55 (12)	Neuroblastoma	10 (4)
Lung	52 (16)	Bone	9 (2)
Lymphoma	52 (13)	Endometrium	9 (2)
Esophagus	37 (7)	Thyroid	8 (4)
Ovary	34 (8)	Larynx	6 (1)
Testis	27 (7)	Gallbladder	5 (1)
Urinary bladder	22 (6)	Salivary gland	5 (1)
Kidney	18 (5)	Upper aerodigestive tract[a]	5 (1)
Cervix	17 (3)	**All sites**	**989 (227)**

[a]Upper aerodigestive tract (UADT): oral cavity, pharynx, larynx and esophagus.

Table 2

The 20 variant groups (33 variants) pleiotropic for cancer risk within the same ethnic group.

Variant group	Region	Cancer sites	Variant rs number	LD (R^2)[a]	Cancer site associated with the variant	EA[b]	EAF[b]	P-value[c]	OR[b]	95% confidence interval of OR[b]	Ancestry of the discovery sample[d]	Mapped gene[e]	PubMed ID of the study
rs4245739	1q32.1	Breast, Prostate	rs4245739	1	Breast	C	0.26	2E-12	1.14	[1.10–1.18]	European	*MDM4*	23535733
			rs4245739	1	Prostate	C	0.25	2E-11	0.91	[0.88–0.95]	European	*MDM4*	23535732
rs10936599, rs12638862, rs12696304	3q26.2	Colon & rectum, Leukemia, Multiple myeloma, Skin, Urinary bladder	rs12638862	1 (Ref)	Multiple myeloma	A	0.74	2E-06	1.37	[1.20–1.56]	European	4.9kb 3' of *TERC*	23502783
			rs12696304	0.945	Skin	C	0.73	3E-07	1.10	[NR]	European	1.1kb 3′ of *TERC*	26237428
			rs10936599	0.928	Urinary bladder	C	0.76	5E-09	1.18	[1.11–1.23]	European	*MYNN*	24163127
			rs10936599	0.928	Colon & rectum	C	0.75	3E-08	1.08	[1.04–1.10]	European	*MYNN*	20972440
			rs10936599	0.928	Leukemia	C	0.75	2E-09	1.26	[1.17–1.35]	European	*MYNN*	24292274
			rs10936599	0.928	Multiple myeloma	C	0.80	9E-14	1.26	[1.18–1.33]	European	*MYNN*	23955597
rs10069690	5p15.33	Breast, Ovary	rs10069690	1	Breast[f]	T	0.32	5E-12	1.15	[1.11–1.20]	European	*TERT*	23535733
			rs10069690	1	Ovary	T	0.26	9E-09	1.14	[1.10–1.19]	European	*TERT*	25581431
rs2736100	5p15.33	Brain, Lung, Testis	rs2736100	1	Brain	G	0.49	2E-17	1.27	[1.19–1.37]	European	*TERT*	19578367
			rs2736100	1	Lung[f]	G	0.50	2E-10	1.12	[1.08–1.16]	European	*TERT*	19836008
			rs2736100	1	Testis	G	0.51	8E-15	0.75	[0.67–0.85]	European	*TERT*	20543847
rs31489, rs31490, rs401681, rs4975616	5p15.33	Leukemia, Lung, Pancreas, Skin, Urinary bladder	rs4975616	1 (Ref)	Lung	NR	NR	3E-09	1.15	[1.10–1.20]	European	2.2kb 3' of *CLPTM1L*	19654303
			rs401681[g]	0.849	Urinary bladder	C	0.54	4E-11	1.12	[1.08–1.16]	European	*CLPTM1L*	24163127
			rs401681	0.849	Lung	C	0.57	8E-09	1.15	[1.09–1.19]	European	*CLPTM1L*	18978787
			rs401681	0.849	Skin	C	0.55	9E-13	1.21	[NR]	European	*CLPTM1L*	25855136
			rs31489	0.735[h]	Lung	C	0.59	2E-10	1.12	[1.09–1.16]	European	*CLPTM1L*	19836008
			rs31490	0.849	Leukemia	A	0.43	2E-07	1.18	[1.11–1.26]	European	*CLPTM1L*	24292274
			rs31490	0.849	Pancreas	A	0.44	2E-11	1.20	[1.14–1.27]	European	*CLPTM1L*	25086665
rs3777204, rs56219066	5q15	Salivary gland, Multiple myeloma	rs3777204[i]	1 (Ref)	Salivary gland	C	0.29	1E-07	1.86	[1.48–2.34]	European	*ELL2*[j]	25823930
			rs56219066	0.971	Multiple myeloma	T	0.71	1E-09	1.25	[1.16–1.34]	European	*ELL2*[j]	26007630
rs2494938	6p21.1	Lung, Stomach	rs2494938	1	Lung	A	0.23	2E-06	1.15	[1.08–1.22]	East Asian	*LRFN2*	23103227
			rs2494938	1	Stomach	A	0.23	5E-09	1.18	[1.12–1.25]	East Asian	*LRFN2*	23103227

Variant group	Region	Cancer sites	Variant rs number	LD (R^2)[a]	Cancer site associated with the variant	EA[b]	EAF[b]	P-value[c]	OR[b]	95% confidence interval of OR[b]	Ancestry of the discovery sample[d]	Mapped gene[e]	PubMed ID of the study
rs2285947	7p15.3	Esophagus, Lung, Stomach	rs2285947	1	Esophagus	A	0.27	3E-06	1.14	[1.08–1.21]	East Asian	*DNAH11*	23103227
			rs2285947	1	Lung	A	0.26	2E-08	1.17	[1.11–1.24]	East Asian	*DNAH11*	23103227
			rs2285947	1	Stomach	A	0.27	1E-06	1.14	[1.08–1.21]	East Asian	*DNAH11*	23103227
rs10505477, rs6983267	8q24.21	Colon & rectum, Prostate	rs10505477	1 (Ref)	Colon & rectum	T	0.54	8E-13	1.20	[NR]	European	20kb 5' of *POU5F1B*	24737748
			rs10505477	1 (Ref)	Prostate	T	0.49	9E-09	1.39	[1.28–1.50]	European	20kb 5' of *POU5F1B*	24740154
			rs6983267	0.916	Colon & rectum[f]	G	0.49	1E-14	1.27	[1.16–1.39]	European	15kb 5' of *POU5F1B*	17618284
			rs6983267	0.916	Prostate[f]	G	0.50	4E-15	1.34	[1.25–1.43]	European	15kb 5' of *POU5F1B*	24753544
rs2294008	8q24.3	Stomach, Urinary bladder	rs2294008	1	Urinary bladder	T	0.46	3E-15	1.13	[1.10–1.16]	European	*PSCA*	24163127
			rs2294008	1	Stomach	T	0.47	2E-07	1.21	[NR]	European	*PSCA*	26098866
rs11012732, rs1243180	10p12.31	Brain, Ovary	rs11012732	1 (Ref)	Brain	A	0.32	2E-14	1.46	[1.32–1.61]	European	*MLLT10*	21804547
			rs1243180	0.858	Ovary	A	0.31	1E-09	1.10	[1.06–1.14]	European	*MLLT10*	25581431
rs174537, rs174549	11q12.2	Colon & rectum, Larynx	rs174537	1 (Ref)	Colon & rectum	G	0.59	9E-21	1.16	[1.12–1.19]	East Asian	*MYRF*	24836286
			rs174549	0.923	Larynx	A	0.59	1E-20	1.37	[1.28–1.47]	East Asian	*FADS1*	25194280
rs735665	11q24.1	Leukemia, Lymphoma	rs735665	1	Leukemia	A	0.19	4E-39	1.62	[NR]	European	35kb 5' of *GRAMD1B*	23770605
			rs735665	1	Lymphoma	A	0.21	4E-09	1.81	[1.50–2.20]	European	35kb 5' of *GRAMD1B*	20639881
rs11571833	13q13.1	Breast, Lung	rs11571833	1	Breast	T	0.01	5E-08	1.26	[1.14–1.39]	European	*BRCA2*	23535729
			rs11571833	1	Lung	T	0.01	5E-20	2.47	[2.03–3.00]	European	*BRCA2*	24880342
rs35158985, rs9929218	16q22.1	Colon & rectum, Skin	rs35158985	1 (Ref)	Skin	G	0.30	3E-07	1.10	[NR]	European	*CDH1*	26237428
			rs9929218	0.861	Colon & rectum	G	0.29	1E-08	1.10	[1.06–1.12]	European	*CDH1*	19011631
rs4430796, rs8064454	17q12	Endometrium, Prostate	rs4430796	1 (Ref)	Endometrium	A	0.52	7E-10	1.19	[1.12–1.27]	European	*HNF1B*	21499250
			rs4430796	1 (Ref)	Prostate[f]	A	0.49	1E-11	1.22	[1.15–1.30]	European	*HNF1B*	17603485
			rs8064454	0.949	Prostate	C	0.52	8E-29	1.24	[1.19–1.29]	European	*HNF1B*	25939597
rs7501939, rs757210	17q12	Prostate, Testis	rs757210	1 (Ref)	Ovary	G	0.37	8E-10	1.12	[1.08–1.17]	European	*HNF1B*	23535730
			rs7501939	0.8	Prostate[f]	NR	NR	3E-18	NR	[NR]	European	*HNF1B*	19767753
			rs7501939	0.8	Testis	C	0.62	1E-09	1.28	[1.19–1.39]	European	*HNF1B*	25877299
rs17749561, rs4987855	18q21.33	Leukemia, Lymphoma	rs17749561	1 (Ref)	Lymphoma	G	0.91	8E-10	1.34	[1.22–1.47]	European	7.4kb 3' of *BCL2*	25279986
			rs4987855	0.951	Leukemia	G	0.91	3E-12	1.47	[1.32–1.61]	European	*BCL2*	23770605

Variant group	Region	Cancer sites	Variant rs number	LD (R^2)[a]	Cancer site associated with the variant	EA[b]	EAF[b]	P-value[c]	OR[b]	95% confidence interval of OR[b]	Ancestry of the discovery sample[d]	Mapped gene[e]	PubMed ID of the study
rs8170	19p13.11	Breast, Ovary	rs8170	1	Breast	C	0.48	4E-13	1.19	[1.14–1.25]	European	*BABAM1*	23544013
			rs8170	1	Ovary	NR	0.19	3E-14	1.19	[1.14–1.25]	European	*BABAM1*	23535730
rs2363956	19p13.11	Breast, Ovary	rs2363956	1	Breast	NR	NR	2E-08	1.22	[1.14–1.30]	European	*ANKLE1*	24325915
			rs2363956	1	Ovary	T	0.43	1E-07	1.10	[1.06–1.15]	European	*ANKLE1*	20852633

[a] Linkage disequilibrium (LD).

[b] Effect allele (EA), effect allele frequency (EAF), odds ratio (OR), 95% confidence interval of OR were retrieved from the GWAS Catalog. We described the effect allele in RefSNP orientation as dbSNP did. We also manually reviewed the original publications and dbSNP to fill in the information not reported in the GWAS Catalog. NR: Information not reported in either the GWAS Catalog or the original publications of the study, and not able to be retrieved from dbSNP.

[c] For the same variant-cancer risk association repeatedly identified by multiple studies, the one with the smallest p-value and/or identified by the most recent study was reported in this table.

[d] Clustering of variants into variant groups based on pairwise LD was performed within each of the 17 ethnic groups shown in Table 2, and only the variant groups in the European and East Asian population showed pleiotropic effects for cancer risk.

[e] For variants located within a gene, the gene symbol was reported directly. For intergenic variants, distance and direction to the closest gene were reported.

[f] Statistically significant association was also observed in other populations:

rs10069690 and breast cancer in European and African American/Afro-Caribbean (PubMed ID: 22037553).

rs2736100 and lung cancer in East Asian (PubMed ID: 21725308).

rs6983267 and colorectal cancer in East Asian (PubMed ID: 24836286); rs6983267 and prostate cancer in a sample of European, African American/Afro-Caribbean, East Asian, Hispanic/Latin American (PubMed ID: 26034056).

rs4430796 and prostate cancer in East Asian (PubMed ID: 26443449).

rs7501939 and prostate cancer in East Asian (PubMed ID: 20676098), and in a sample of European, African American/Afro-Caribbean, East Asian, Hispanic/Latin American (PubMed ID: 26034056).

[g] Variant rs401681 has been associated with pancreatic cancer in a sample of European, African unspecified, Asian unspecified, Hispanic/Latin American (PubMed ID: 26098869) and another sample of European and East Asian (PubMed ID: 20101243).

[h] Variant rs31489 is in high LD with rs31490 ($R^2 = 0.874$) and rs401681 ($R^2 = 0.814$).

[i] Variant rs3777204 was shown as exm2265979 in the original publication.

[j] *ELL2* is a pleiotropic gene that is novel to this study.

Table 3

Comparison of the variant consequences between the 518 pleiotropic variants and the 18,069 single-cancer variants (p-value of Fisher's exact test = 2.2×10^{-16}) using the "–most_severe" option.

Variant consequence[a]	Impact[b]	Number of pleiotropic variants (%)[c]	Number of single-cancer variants (%)[c]
Gene variant		**Total 460 (89.0)**	**Total 11755 (65.3)**
Intron variant	Modifier	369 (71.4)	10703 (59.4)
3 prime UTR variant	Modifier	29 (5.6)	277 (1.5)
Non coding transcript exon variant	Modifier	26 (5.0)	454 (2.5)
5 prime UTR variant	Modifier	11 (2.1)	85 (0.5)
Synonymous variant	Low	11 (2.1)	94 (0.5)
Missense variant	Moderate	8 (1.5)	102 (0.6)
Splice region variant	Low	4 (0.8)	26 (0.1)
Stop gained	High	2 (0.4)	3 (0.0)
Frameshift variant	High	0 (0.0)	4 (0.0)
Splice donor variant	High	0 (0.0)	3 (0.0)
Splice acceptor variant	High	0 (0.0)	2 (0.0)
Inframe insertion	Moderate	0 (0.0)	1 (0.0)
Start lost	High	0 (0.0)	1 (0.0)
Intergenic variant		**Total 57 (11.0)**	**Total 5737 (31.8)**
Upstream gene variant	Modifier	38 (7.4)	954 (5.3)
Downstream gene variant	Modifier	11 (2.1)	731 (4.1)
(Other intergenic variant)	-	8 (1.5)	4052 (22.5)
Regulatory region variant		**Total 0 (0.0)**	**Total 523 (2.9)**
TF binding site variant	Modifier	0 (0.0)	9 (0.0)
(Other regulatory region variant)	-	0 (0.0)	514 (2.9)

[a] Variant consequence was predicted using the Ensembl Variant Effect Predictor (VEP).

[b] Impact was defined by Ensembl to classify the severity of the variant consequence, with four categories: high, moderate, low, and modifier.

[c] Variant consequence annotations for one pleiotropic variant (rs35464379) and 54 single-cancer variants were not available.