bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cross-Cancer Evaluation of Polygenic Risk Scores for 17 Cancer Types in Two Large Cohorts

Rebecca E. Graff[1,2], Taylor B. Cavazos[3], Khanh K. Thai[2], Linda Kachuri[1], Sara R. Rashkin[1], Joshua D. Hoffman[1], Stacey E. Alexeeff[2], Maruta Blatchins[2], Travis J. Meyers[1], Lancelote Leong[1], Caroline G. Tai[1], Nima C. Emami[1,3], Douglas A. Corley[2], Lawrence H. Kushi[2], Elad Ziv[4,5,6], Stephen K. Van Den Eeden[2,7], Eric Jorgenson[2], Thomas J. Hoffmann[1,2,4], Laurel A. Habel[2], John S. Witte[1,4,6,7]*, Lori C. Sakoda[2]*

* Shared senior authorship

Affiliations:

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

[2)] Division of Research, Kaiser Permanente Northern California, Oakland, CA

[3)] Program in Biological and Medical Informatics, University of California San Francisco, San Francisco, CA

[4)] Institute for Human Genetics, University of California San Francisco, San Francisco, CA

[5)] Department of Medicine, University of California San Francisco, San Francisco, CA

[6)] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA

[7)] Department of Urology, University of California San Francisco, San Francisco, CA

Corresponding Authors:

John S. Witte

Department of Epidemiology and Biostatistics

University of California, San Francisco

1450 3rd Street, Room 388, San Francisco, CA 94158

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phone: (415) 502-6882; Fax: N/A

Email: jwitte@ucsf.edu

Lori C. Sakoda

Division of Research

Kaiser Permanente Northern California

2000 Broadway, Oakland, CA 94612

Phone: (510) 891-3677; Fax: N/A

Email: lori.sakoda@kp.org

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

Background: Genetic factors that influence etiologic mechanisms shared across cancers could affect the risk of multiple cancer types. We investigated polygenic risk score (PRS)-specific pleiotropy across 17 cancers in two large population-based cohorts.

Methods: The study population included European ancestry individuals from the Genetic Epidemiology Research on Adult Health and Aging cohort (16,012 cases, 50,552 controls) and UK Biobank (48,969 cases, 359,802 controls). We selected known independent risk variants to construct a PRS for each cancer type. Within cohorts, each PRS was evaluated in multivariable logistic models with respect to the cancer for which it was developed and each other cancer type. Results were then meta-analyzed across cohorts. In the UK Biobank, each PRS was additionally evaluated relative to 20 cancer risk factors or biomarkers.

Results: All PRS replicated associations with their corresponding cancers ($p<0.05$). Eleven cross-cancer associations – ten positive and one inverse – were found after Bonferroni correction ($p<0.05/17=0.0029$). Two cancer pairs showed bidirectional associations; the melanoma PRS was positively associated with oral cavity/pharyngeal cancer and vice versa, whereas the lung cancer PRS was positively associated with oral cavity/pharyngeal cancer, and the oral cavity/pharyngeal cancer PRS was inversely associated with lung cancer. We identified 65 associations between a cancer PRS and non-cancer phenotype.

Conclusions: In this study examining cross-cancer PRS associations in two cohorts unselected for phenotype, we validated known and uncovered novel patterns of pleiotropy. Our results have the potential to inform investigations of risk prediction, shared etiology, and precision cancer prevention strategies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Neoplasms are diverse in their clinical presentation, but they share biological hallmarks acquired during the transformation of normal cells into neoplastic ones (1). Inherited genetic factors underpinning shared hallmarks could alter cancer risk in a pleiotropic manner. Indeed, genome-wide association studies (GWAS) of individual cancer types have identified loci associated with other cancer types, including 5p15 (*TERT-CLPTM1L)* (2), 6p21 (HLA complex) (3,4), and 8q24 (5). Non-GWAS approaches have yielded further pleiotropic cancer risk variants (6-26), and genetic correlation studies have identified cancer pairs with shared heritability (27-29).

Polygenic risk scores (PRS) capture a different aspect of pleiotropy. By combining variants into scores that summarize genetic susceptibility, PRS typically explain a larger proportion of disease risk than single low-penetrance variants. Relative to genetic correlations, PRS offer greater specificity by selecting a refined set of disease-specific risk variants. Therefore, PRS analyses have the potential to inform etiology and identify possible precision prevention targets shared across cancers. They are also plausibly valuable for risk prediction; there is potential clinical advantage in knowing that an individual with a high PRS for one cancer is at risk for another. While PRS have been extensively investigated for individual cancers, cross-cancer portability of PRS has remained understudied.

To comprehensively investigate pleiotropy across cancers, we leveraged results from 273 published GWAS to systematically construct PRS specific to 17 cancer types. We then evaluated associations between each PRS and the risk of each cancer type in European ancestry individuals from two large independent cohorts with genome-wide array data – the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort and UK Biobank. We also assessed associations between each genetic variant contributing to a PRS and the risk of each cancer type and characterized pleiotropy between each PRS and 20 cancer risk factors.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

METHODS

Study Populations

GERA is a prospective cohort of 102,979 adults drawn from >400,000 Kaiser Permanente Northern California (KPNC) health plan members who participated in the Research Program on Genes, Environment and Health. Participants answered a baseline survey regarding lifestyle and medical history, provided a saliva specimen between 2008 and 2011, and were successfully genotyped (30,31). Phenotyping of GERA cancer cases and controls is described in the **Supplementary Methods**. Following quality control (QC; **Supplementary Methods**), the GERA analytic population included 16,012 cases and 50,552 controls.

The UK Biobank is a population-based prospective cohort of 502,611 individuals from the United Kingdom, ages 40 to 69 at recruitment between 2006 and 2010 (32). Participants were evaluated at baseline visits during which assessment center staff introduced a touch-screen questionnaire, conducted a brief interview, gathered physical measurements, and collected biological samples. Phenotyping of UK Biobank cancer cases and controls, as well as cancer risk factors and serum biomarkers, is described in the **Supplementary Methods**. Following QC (**Supplementary Methods**), the UK Biobank analytic population included 48,969 cases and 359,802 controls.

This study was approved by the KPNC and University of California Institutional Review Boards and the UK Biobank data access committee.

Variant Selection

The process to determine variants previously associated with each cancer type is detailed in the **Supplementary Methods**. Briefly, we abstracted data from relevant cancer GWAS in the National Human Genome Research Institute-European Bioinformatics Institute Catalog of published GWAS (33,34). We identified additional studies via cited references and PubMed searches (**Supplementary Table 1**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Following a rigorous process to catalog variants available in both cohorts (**Supplementary Methods**, **Supplementary Figure 1**), we used PriorityPruner (35) and LDlink (36) to select a set of independent risk variants with linkage disequilibrium (LD) <0.3 for each cancer type. The process preferentially selected variants with the smallest *p*-values and highest imputation scores associated with the broadest phenotype (e.g., lymphocytic leukemia over pediatric lymphocytic leukemia).

Statistical Analysis

For each cancer type, we calculated the PRS based on additive dosages of the individual risk variants: ∑(# risk alleles*log(odds ratios [OR]) from the literature) for i = 1 to n risk alleles. Each PRS was then standardized based on its mean and standard deviation, and evaluated in multivariable logistic regression models with respect to the cancer for which it was developed and each of the other cancer types. ORs were estimated per standard deviation increase in the PRS. Models were adjusted for age at specimen collection, first 10 ancestry PCs, sex (except models for sex-specific cancers), reagent kit used for genotyping (Axiom v1 or v2; GERA only), and genotyping array (UK Biobank only). After conducting analyses by cohort, we combined results across cohorts using fixed effects meta-analyses. Heterogeneity was assessed based on I^2 and Cochran's Q.

For variants contributing to any of the 17 PRS, we estimated the associated risk for each cancer type using logistic regression adjusted for the aforementioned covariables. Variants were modeled individually on a log-additive scale. Results from both cohorts were meta-analyzed. We then visualized the genomic regions that were overrepresented among pleiotropic variants relative to all PRS variants.

In secondary analyses, we used the UK Biobank to explore associations between each PRS and 20 cancer risk factors or serum biomarkers (**Supplementary Methods**). Briefly, logistic (smoking status) or linear (remaining phenotypes) regression models were restricted to

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cancer-free controls and adjusted for the covariables noted above, as well as cigarette pack-years (forced expiratory volume in 1 second [FEV_1]/forced vital capacity [FVC]), assay date (serum biomarkers), and use of medications to lower cholesterol (high-density lipoprotein [HDL] and low-density lipoprotein [LDL]), control blood pressure (systolic and diastolic blood pressure), and regulate insulin (glycated hemoglobin [HbA1c]).

All statistical analyses were performed using R 3.2.2 or 3.3.3 (http://www.r-project.org/).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

We abstracted 17,868 genome-wide significant associations from 273 published GWAS (**Supplementary Table 1**). Of the selected set of 880 risk variants independent within the 17 cancer types, 808 variants were independent across all cancer types (**Supplementary Tables 2a-2q**). Endometrial cancer had the fewest independent risk variants (n = 9), and breast cancer had the most (n = 187) (**Figure 1**).

Participants were more commonly female than male (**Supplementary Table 3**). GERA participants were older than UK Biobank participants (mean age: cases, 69 versus 60; controls, 62 versus 57). Case counts ranged from 665 for pancreatic cancer to 17,901 for breast cancer (**Figure 1**). Meta-analyses of non-sex-specific cancers included 410,354 controls. Female-specific meta-analyses included 219,648 controls. Meta-analyses of prostate cancer included 190,706 male controls. For testicular cancer, analyses included 169,967 male controls (UK Biobank only).

Each PRS replicated at a nominal significance level (p<0.05; dark gray cells in **Figure 1**) for its corresponding cancer outcome. The largest effect sizes per standard deviation increase in the PRS were observed for testicular (OR=2.29; p=6.82x10^{-105}) and thyroid cancers (OR=1.55; p=6.38x10^{-33}). The smallest were observed for esophageal/stomach (OR=1.07; p=0.039) and oral cavity/pharyngeal cancers (OR=1.08; p=0.007). None of these replicative associations demonstrated significant heterogeneity across cohorts ($p_{\text{Cochran's-Q}}$<0.05). **Supplementary Tables 4a, 4b, and 4c** include summary statistics from the meta-analyses, GERA, and UK Biobank, respectively.

Eleven associations between a PRS and cross-cancer outcome were found after Bonferroni correction (p<0.05/17=0.0029; **Figure 1**). Results remained unchanged correcting for the false discovery rate at q<0.05 (**Supplementary Figure 2**). Ten pairs showed a positive association: bladder cancer PRS with cervical cancer (OR=1.04; p=9.04x10^{-4}); endometrial cancer PRS with prostate cancer (OR=1.06; p=5.34x10^{-9}); lung cancer PRS with non-Hodgkin's

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

lymphoma (NHL; OR=1.11; p=5.57x10^{-7}), colorectal cancer (OR=1.04; p=1.22x10^{-3}), and oral cavity/pharyngeal cancer (OR=1.11; p=1.06x10^{-4}); lymphocytic leukemia PRS with NHL (OR=1.08; p=1.48x10^{-4}); melanoma PRS with breast (OR=1.04; p=6.33x10^{-7}) and oral cavity/pharyngeal cancers (OR=1.10; p=7.84x10^{-4}); and oral cavity/pharyngeal cancer PRS with melanoma (OR=1.04; p=2.04x10^{-3}) and NHL (OR=1.10; p=2.67x10^{-6}). The oral cavity/pharyngeal cancer PRS was inversely associated with lung cancer (OR=0.93; p=6.25x10^{-4}). Only the melanoma PRS-breast cancer association demonstrated heterogeneity (I^2=0.79; $p_{\text{Cochran's-Q}}$=0.029). Thirty additional associations (24 positive, six inverse) were nominally significant (p<0.05).

Associations between each PRS variant and cancer type are compiled in **Supplementary Tables 5a-5q**. In total, 141 cross-cancer associations were detected at a threshold corrected for the number of effective independent tests (p<0.05/808=6.2x10^{-5}; **Supplementary Table 6**; includes 18 duplicate associations in which the same variant originated from multiple PRS). They included associations for 55 variants in LD with previously identified risk variants for the outcome cancer. Among the remaining 86 associations, 60 were novel, in that the variant (or variants with r^2>0.3 in the 1000 Genomes EUR superpopulation reference panel) had not previously been associated with the outcome cancer at p<1x10^{-6} (**Figures 2a and 2b**; includes five duplicate associations originating from multiple PRS). The cancer types with the largest number of novel risk variants were prostate (n = 15), NHL (n = 14; includes one variant originating from multiple PRS), and cervical (n = 12).

Several genomic regions were overrepresented among pleiotropic compared to all PRS variants (**Figure 2c**). Across the 141 cross-cancer associations, pleiotropic variants were most commonly found in *TERT-CLPTM1L* (16% versus 3.0%) and HLA (6p21.32: 16% versus 3.6%; 6p21.33: 13% versus 3.6%). Additional regions enriched for pleiotropy included 9q34.2 (2.1% versus 0.23%), 10q24.33 (4.2% versus 0.46%), 12q24.12 (1.1% versus 0.11%), and 17q12

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(5.3% versus 0.69%). These regions remained enriched following normalization by region size (**Supplementary Figure 3**).

At a Bonferroni-corrected significance threshold (p<0.05/20=0.0025), we identified 65 associations between a cancer PRS and cancer risk factor or biomarker (**Figure 3**, **Supplementary Table 7**). The lung cancer PRS appeared most pleiotropic, with 12 associated phenotypes. Positively associated phenotypes included cigarettes per day in smokers (p=6.06x10^{-32}), pulmonary obstruction (decreasing FEV_1/FVC; p=1.97x10^{-25}), HbA1c (p=1.59x10^{-22}), height (p=1.30x10^{-4}), and multiple metrics of adiposity (e.g., body mass index [BMI]: p=7.63x10^{-9}). The lung cancer PRS was associated with lower levels of insulin-like growth factor-1 (IGF-1; p=8.58x10^{-18}) and HDL cholesterol (p=3.94x10^{-17}). The NHL and oral cavity/pharyngeal cancer PRS were each associated with nine secondary phenotypes. Among the associations for the former were increasing levels of LDL (p=1.53x10^{-21}), IGF-1 (p=2.13x10^{-9}), and C-reactive protein (p=5.50x10^{-7}). The latter was associated with increasing alcohol intake (p=7.28x10^{-11}) and pulmonary obstruction (p=1.26x10^{-10}). PRS for breast, prostate, and ovarian cancers were not clearly associated with any secondary phenotypes. Among the secondary phenotypes, height showed the most cancer PRS associations (n=8), followed by HbA1c (n=7), and BMI, FEV_1/FVC, and LDL (n=6 each).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

In this comprehensive study of PRS-specific cancer pleiotropy, we constructed 17 PRS based on systematic review of the cancer GWAS literature. Analyses identified 11 statistically significant cross-cancer PRS associations, as well as novel cancer associations with 55 unique risk variants in known susceptibility regions. We further identified 65 cancer PRS associations with selected non-cancer phenotypes.

Of all PRS evaluated, the oral cavity/pharyngeal and lung cancer PRS were most commonly implicated in associations with cross-cancer and non-cancer phenotypes. These results support existing evidence of cancer pleiotropy, given that the PRS for both cancers included variants in two well-known pleiotropic cancer regions – *TERT-CLPTM1L* (2) and HLA (3,4). Notwithstanding shared susceptibility regions, the relationship between oral cavity/pharyngeal and lung cancers was inconsistent. In one direction, the oral cavity/pharyngeal cancer PRS was *inversely* associated with lung cancer. The negative pleiotropy could be partly attributable to the oral cavity/pharyngeal PRS variants (rs467095 and rs10462706; **Supplementary Tables 8 and 9**), both expression quantitative trait loci for *TERT* and *CLPTM1L*, that were inversely associated with lung cancer risk and in LD (r^2=0.96 and 0.66, respectively) with variants in the lung cancer PRS. The oral cavity/pharyngeal cancer PRS was also associated with increasing alcohol intake, an established risk factor for such cancers (37). The relationship between alcohol intake and lung cancer remains controversial, with the possibility of an inverse or J-shaped relationship (38,39). In the other direction, the lung cancer PRS was *positively* associated with oral cavity/pharyngeal cancer risk. The positive pleiotropy may be partially explained by the association between the lung cancer PRS and increasing cigarettes per day among smokers. PRS for both cancers were also associated with pulmonary obstruction (i.e., decreasing FEV_1/FVC), a known lung cancer risk factor (40), as well as higher HbA1c and lower IGF-1 levels, both of which indicate insulin resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Oral cavity/pharyngeal cancer also showed a bidirectional, positive relationship with melanoma, even though the two PRS share only one pair of variants in LD in *TERT-CLPTM1L* (**Supplementary Tables 8 and 9**). PRS for both cancer types were inversely associated with height, which is somewhat surprising since increasing height has been strongly associated with melanoma risk (41).

The lung cancer PRS was positively associated with colorectal cancer and NHL. The former association did not appear to be driven by variants in LD; only two out of 109 lung cancer risk variants (rs2853677 and rs1333040) are in high LD (r^2=0.62 and 0.49, respectively) with colorectal cancer risk variants (rs2735940 and rs1537372, respectively), and neither was strongly associated with colorectal cancer risk in our data. Given that the lung cancer PRS was associated with increasing BMI, body fat, and cigarettes per day, its association with colorectal cancer risk coheres with known risk factors. As five of the lung cancer variants in HLA are in LD with NHL risk variants, LD structure likely played a larger role in the latter association. Both the lung cancer and NHL PRS were associated with increasing HbA1c levels, implicating insulin resistance as a possible shared mechanism. It could be that the NHL PRS was not associated with lung cancer risk because it included only 19 SNPs (relative to 109 SNPs in the lung cancer PRS). We also identified a novel association between a lung cancer risk variant and NHL; rs652888 (6p21.33 in *EHMT2*) has been linked to several autoimmune and infectious diseases (42,43), as well as infection with Epstein-Barr virus (44), a known NHL risk factor (45).

Among the remaining Bonferroni-significant cross-cancer PRS associations, two included cancers with PRS variants that were independent at the r^2=0.3 threshold: bladder cancer PRS with cervical cancer and oral cavity/pharyngeal cancer PRS with NHL. Accordingly, cervical cancer and NHL were among the cancers with the most novel risk variants. Although none of the 15 bladder cancer variants are in LD with known genome-wide significant risk variants for cervical cancer, one *CLPTM1L* variant (rs401681-C) was associated with increased cervical cancer risk at a genome-wide significance level in our study, confirming a suggestive

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

association signal reported previously (8). Similarly, two oral cavity/pharyngeal cancer variants in HLA (rs9271378 and rs3135006), a region that has previously been implicated in NHL (46), were strongly associated with NHL risk in our analyses.

Increasing NHL risk was also associated with the lymphocytic leukemia PRS. Out of 64 lymphocytic leukemia risk variants, only one (rs4987855) is in LD (r^2=0.95) with an NHL risk variant (rs17749561). Our results align with those from Sampson, *et al*., which showed an association between a PRS for chronic lymphocytic leukemia (CLL) and the risk of diffuse large B-cell lymphoma, the most common NHL subtype (29). Both lymphocytic leukemia and NHL originate from lymphocytes, and while SEER considers CLL a leukemia (47), more recent classifications include CLL with NHL and other B-cell lymphomas (48).

The association between the endometrial cancer PRS and prostate cancer risk also validated results from Sampson, *et al*. (29) The remaining cross-cancer association from our study – between the melanoma PRS and breast cancer – was not evaluated. Their study did, however, identify two associations that our analyses did not validate: 1) lung cancer PRS and bladder cancer risk, and 2) endometrial cancer PRS and testicular cancer risk. Given differences in study design and the many additional SNPs that have been discovered since 2015, it is not especially surprising that some results were distinct.

The genomic regions overrepresented among pleiotropic variants support existing knowledge about shared mechanisms of carcinogenesis. In addition to *TERT-CLPTM1L* and HLA, 9q34.2, 10q24.33, 12q24.12, and 17q12 have been implicated in susceptibility for multiple cancer types. Variants in the breast (49) and pancreatic cancer (50) susceptibility locus 9q34.2 influence estrogen receptor signaling and insulin resistance, and were recently associated with protein biomarkers affecting carcinogenesis (51). The 10q24.33 region containing *OBFC1,* a known telomere maintenance gene, has been implicated in lymphocytic leukemia, melanoma, and kidney, ovarian, and thyroid cancers (52-57). A previous cross-cancer analysis linked 12q24.12 to both colorectal and endometrial cancer risk (58). This locus includes *SH2B3,* a

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gene involved in regulating signaling pathways related to hematopoiesis, inflammation, and cell migration. The 17q12 locus includes *HNF1B*, which has been extensively characterized with respect to hormonally driven cancers (59).

The non-cancer phenotypes that most frequently surfaced in associations with cancer PRS offer additional mechanistic insights. The lymphocytic leukemia, NHL, and kidney, lung, oral cavity/pharyngeal, and pancreatic cancer PRS were associated with at least one anthropometric trait and showed directionally consistent associations with HbA1c and IGF-1 levels. Obesity-induced chronic inflammation and oxidative stress create a milieu conducive to malignant transformation (60). Furthermore, the metabolic reprogramming necessary to meet the increased energy requirements of proliferating malignant cells is a known hallmark of cancer (1). There is also complex interplay between genetic determinants of adiposity and smoking behaviors (61). Taken together, the findings further implicate obesity-related metabolic dysregulation in cancer susceptibility for multiple sites.

Among the limitations of our study were modest numbers for some cancers. We favored their inclusion in an effort to evaluate more cancer types than previous investigations. We furthermore combined esophageal and stomach cancers and, separately, oral cavity and pharyngeal cancers into composite phenotypes. While there is precedent for doing so (62-64), we acknowledge the potential resulting phenotypic heterogeneity. We note that our analyses included prevalent and incident cases. However, results from *a posteriori* cross-cancer PRS analyses restricted to incident cases mirrored those of the primary analyses (**Supplementary Table 10**). Our findings are thus unlikely to be driven by associations with survival rather than risk. We also note that our PRS were comprised of exclusively genome-wide significant variants. While a less stringent threshold for inclusion might have yielded more signal, it would not have been based on convincing *a priori* evidence. Finally, while all PRS replicated for their target cancers, some individual risk variants did not. Nevertheless, 92% had effect estimates with consistent directionality relative to the published literature.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Among the strengths of our study was use of two large cohorts with abundant individual-level genetic and phenotypic data, independent of those from which risk variants were identified in prior cancer GWAS (except by Huyghe, *et al*) (65). We also comprehensively reviewed the contemporary literature to identify genome-wide significant risk variants for 17 cancer types. Evaluating risk variants identified for one cancer with respect to risk for others enabled discovery of novel susceptibility loci that would not otherwise meet the strict criteria for genome-wide significance. By additionally evaluating associations with cancer risk factors, we generated insights into pathways that may be influenced by genetic variants implicated in cancer.

Our work expands the repertoire of genetic susceptibility variants for multiple cancers, which should prompt future investigations of their biological and clinical relevance. Even if the precise biological mechanisms underpinning the associations remain ambiguous, the associations may still be leveraged toward a more integrated model of cancer risk prediction considering cross-phenotype effects. Even for cancers that do not share non-genetic risk factors, shared genetic profiles have the potential to aid in risk prediction. Combined with future research that investigates pleiotropy in cancer subgroups (e.g., by smoking status or histology) and clinical applications of PRS, our results may inform new strategies toward reducing the burden of cancer.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FUNDING

This work was supported by the National Institutes of Health (grant numbers K07CA188142, K24CA169004, R01CA088164, R01CA201358, R25CA112355, RC2AG036607, and U01CA127298) and the UCSF Goldberg-Benioff Program in Cancer Translational Biology. Support for participant enrollment, survey administration, and biospecimen collection for the Research Program on Genes, Environment and Health was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente national and regional community benefit programs.

ACKNOWLEDGMENTS

Data from the UK Biobank resource was obtained under application number 14105. We thank Alexander Dean for his initial contributions to the development of PRS in GERA.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.

2. Gaspar TB, Sa A, Lopes JM, Sobrinho-Simoes M, Soares P, Vinagre J. Telomere maintenance mechanisms in cancer. *Genes (Basel).* 2018;9(5).

3. Karnes JH, Bastarache L, Shaffer CM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med.* 2017;9(389).

4. Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet.* 2011;7(4):e1001378.

5. Huppi K, Pitt JJ, Wahlberg BM, Caplen NJ. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front Genet.* 2012;3:69.

6. Cheng I, Kocarnik JM, Dumitrescu L, et al. Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. *Gut.* 2014;63(5):800-807.

7. Jones CC, Bradford Y, Amos CI, et al. Cross-cancer pleiotropic associations with lung cancer risk in African Americans. *Cancer Epidemiol Biomarkers Prev.* 2019;28(4):715-723.

8. Rafnar T, Sulem P, Stacey SN, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet.* 2009;41(2):221-227.

9. Setiawan VW, Schumacher F, Prescott J, et al. Cross-cancer pleiotropic analysis of endometrial cancer: PAGE and E2C2 consortia. *Carcinogenesis.* 2014;35(9):2068-2073.

10. Chang J, Wei L, Miao X, et al. Two novel variants on 13q22.1 are associated with risk of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2015;24(11):1774-1780.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11. Du J, Xu Y, Dai J, et al. Genetic variants at 5p15 are associated with risk and early onset of gastric cancer in Chinese populations. *Carcinogenesis.* 2013;34(11):2539-2542.

12. Kinnersley B, Migliorini G, Broderick P, et al. The TERT variant rs2736100 is associated with colorectal cancer risk. *Br J Cancer.* 2012;107(6):1001-1008.

13. Lee E, Stram DO, Ek WE, et al. Pleiotropic analysis of cancer risk loci on esophageal adenocarcinoma risk. *Cancer Epidemiol Biomarkers Prev.* 2015;24(11):1801-1803.

14. Li DK, Han J, Liu JB, 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(1):123-127.

15. Lim U, Kocarnik JM, Bush WS, et al. Pleiotropy of cancer susceptibility variants on the risk of non-Hodgkin lymphoma: the PAGE consortium. *PLoS One.* 2014;9(3):e89791.

16. Panagiotou OA, Travis RC, Campa D, et al. A genome-wide pleiotropy scan for prostate cancer risk. *Eur Urol.* 2015;67(4):649-657.

17. Park SL, Fesinmeyer MD, Timofeeva M, et al. Pleiotropic associations of risk variants identified for other cancers with lung cancer risk: the PAGE and TRICL consortia. *J Natl Cancer Inst.* 2014;106(4):dju061.

18. Pierce BL, Ahsan H. Genome-wide "pleiotropy scan" identifies HNF1A region as a novel pancreatic cancer susceptibility locus. *Cancer Res.* 2011;71(13):4352-4358.

19. Fehringer G, Kraft P, Pharoah PD, et al. Cross-cancer genome-wide analysis of lung, ovary, breast, prostate and colorectal cancer reveals novel pleiotropic associations. *Cancer Res.* 2016;76(17):5103-5114.

20. Hung RJ, Ulrich CM, Goode EL, et al. Cross cancer genomic investigation of inflammation Pathway for five common cancers: lung, ovary, prostate, breast, and colorectal cancer. *J Natl Cancer Inst.* 2015;107(11).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21. Kar SP, Beesley J, Amin Al Olama A, et al. Genome-wide meta-analyses of breast, ovarian, and prostate cancer association studies identify multiple new susceptibility loci shared by at least two cancer types. *Cancer Discov.* 2016;6(9):1052-1067.

22. Qian DC, Byun J, Han Y, et al. Identification of shared and unique susceptibility pathways among cancers of the lung, breast, and prostate from genome-wide association studies and tissue-specific protein interactions. *Hum Mol Genet.* 2015;24(25):7406-7420.

23. Scarbrough PM, Weber RP, Iversen ES, et al. A cross-cancer genetic association analysis of the DNA repair and DNA damage signaling pathways for lung, ovary, prostate, breast, and colorectal cancer. *Cancer Epidemiol Biomarkers Prev.* 2016;25(1):193-200.

24. Wang Z, Zhu B, Zhang M, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet.* 2014;23(24):6616-6633.

25. Karami S, Han Y, Pande M, et al. Telomere structure and maintenance gene variants and risk of five cancer types. *Int J Cancer.* 2016;139(12):2655-2670.

26. Toth R, Scherer D, Kelemen LE, et al. Genetic variants in epigenetic pathways and risks of multiple cancers in the GAME-ON consortium. *Cancer Epidemiol Biomarkers Prev.* 2017;26(6):816-825.

27. Jiang X, Finucane HK, Schumacher FR, et al. Shared heritability and functional enrichment across six solid cancers. *Nat Commun.* 2019;10(1):431.

28. Lindstrom S, Finucane H, Bulik-Sullivan B, et al. Quantifying the genetic correlation between multiple cancer types. *Cancer Epidemiol Biomarkers Prev.* 2017;26(9):1427-1435.

29. Sampson JN, Wheeler WA, Yeager M, et al. Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. *J Natl Cancer Inst.* 2015;107(12):djv279.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

30. Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics.* 2015;200(4):1285-1295.

31. Kvale MN, Hesselson S, Hoffmann TJ, et al. Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics.* 2015;200(4):1051-1060.

32. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203-209.

33. Burdett T, Hall PN, Hastings E, et al. The NHGRI-EBI Catalog of published genome-wide association studies. http://www.ebi.ac.uk/gwas. Accessed June 30, 2017.

34. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 2017;45(D1):D896-D901.

35. Edlund CK, Anker M, Schumacher FR, Gauderman WJ, Conti DV. PriorityPruner version 0.1.4 http://prioritypruner.sourceforge.net/. Accessed June 30, 2017.

36. 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;31(21):3555-3557.

37. Kawakita D, Matsuo K. Alcohol and head and neck cancer. *Cancer Metastasis Rev.* 2017;36(3):425-434.

38. Brenner DR, Fehringer G, Zhang ZF, et al. Alcohol consumption and lung cancer risk: A pooled analysis from the International Lung Cancer Consortium and the SYNERGY study. *Cancer Epidemiol.* 2019;58:25-32.

39. Fehringer G, Brenner DR, Zhang ZF, et al. Alcohol and lung cancer risk among never smokers: A pooled analysis from the international lung cancer consortium and the SYNERGY study. *Int J Cancer.* 2017;140(9):1976-1984.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

40. Kachuri L, Johansson M, Rashkin SR, et al. Immune-mediated genetic pathways resulting in pulmonary function impairment increase lung cancer susceptibility. *Nat Commun.* 2020;11(1):27.

41. Yu D-J, Li X-J, Morice A, Wu L-J, Sun W, Zhao T-L. Height and risk of melanoma: a systematic review and meta-analysis. *Int J Clin Exp Med.* 2018;11(5):4426-4435.

42. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* 2017;18(1):76.

43. Shin JG, Cheong HS, Kim JY, et al. Identification of additional EHMT2 variant associated with the risk of chronic hepatitis B by GWAS follow-up study. *Genes Immun.* 2019;20(1):1-9.

44. Rubicz R, Yolken R, Drigalenko E, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet.* 2013;9(1):e1003147.

45. Teras LR, Rollison DE, Pawlita M, et al. Epstein-Barr virus and risk of non-Hodgkin lymphoma in the Cancer Prevention Study-II and a meta-analysis of serologic studies. *Int J Cancer.* 2015;136(1):108-116.

46. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet.* 2014;46(11):1233-1238.

47. Site Recode ICD-O-3/WHO 2008 Definition. https://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html. Accessed June 30, 2017.

48. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;127(20):2375-2390.

49. Michailidou K, Lindstrom S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-94.

50. Wolpin BM, Rizzato C, Kraft P, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet.* 2014;46(9):994-1000.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

51. Shu X, Bao J, Wu L, et al. Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk. *Int J Cancer.* In Press.

52. Robles-Espinoza CD, Velasco-Herrera Mdel C, Hayward NK, Adams DJ. Telomere-regulating genes and the telomere interactome in familial cancers. *Mol Cancer Res.* 2015;13(2):211-222.

53. Duffy DL, Zhu G, Li X, et al. Novel pleiotropic risk loci for melanoma and nevus density implicate multiple biological pathways. *Nat Commun.* 2018;9(1):4774.

54. Gudmundsson J, Thorleifsson G, Sigurdsson JK, et al. A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun.* 2017;8:14517.

55. Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet.* 2017;49(5):680-691.

56. Scelo G, Purdue MP, Brown KM, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat Commun.* 2017;8:15724.

57. Walsh KM, Whitehead TP, de Smith AJ, et al. Common genetic variants associated with telomere length confer risk for neuroblastoma and other childhood cancers. *Carcinogenesis.* 2016;37(6):576-582.

58. Cheng TH, Thompson D, Painter J, et al. Meta-analysis of genome-wide association studies identifies common susceptibility polymorphisms for colorectal and endometrial cancer near SH2B3 and TSHZ1. *Sci Rep.* 2015;5:17369.

59. Wu YH, Graff RE, Passarelli MN, et al. Identification of pleiotropic cancer susceptibility variants from genome-wide association studies reveals functional characteristics. *Cancer Epidemiol Biomarkers Prev.* 2018;27(1):75-85.

60. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body fatness and cancer--viewpoint of the IARC Working Group. *N Engl J Med.* 2016;375(8):794-798.

61. Carreras-Torres R, Johansson M, Haycock PC, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ.* 2018;361:k1767.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

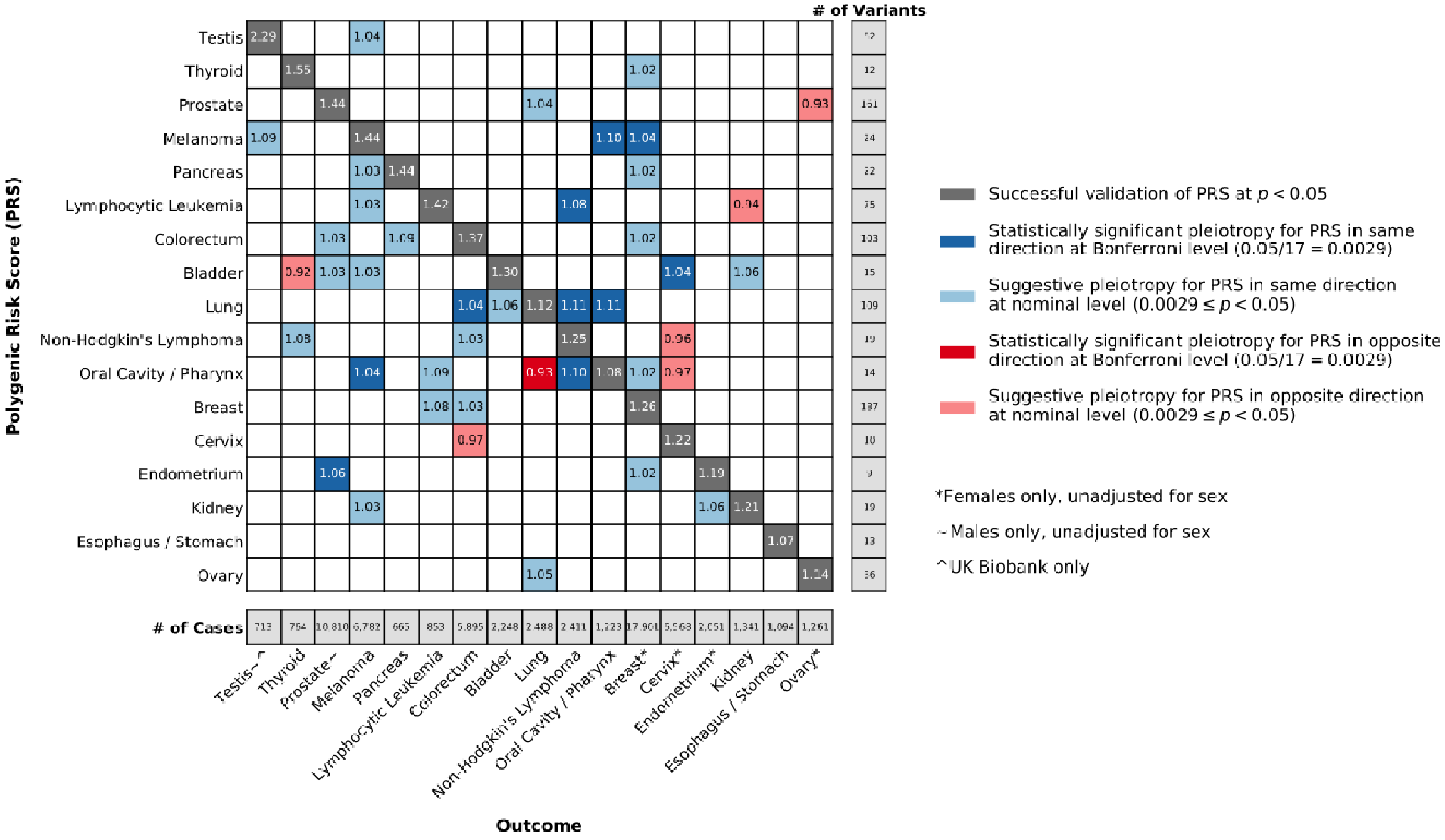
62. Okines AF, Gonzalez de Castro D, Cunningham D, et al. Biomarker analysis in oesophagogastric cancer: results from the REAL3 and TransMAGIC trials. *Eur J Cancer.* 2013;49(9):2116-2125.

63. McKay JD, Truong T, Gaborieau V, et al. A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet.* 2011;7(3):e1001333.

64. Lesseur C, Diergaarde B, Olshan AF, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat Genet.* 2016;48(12):1544-1550.

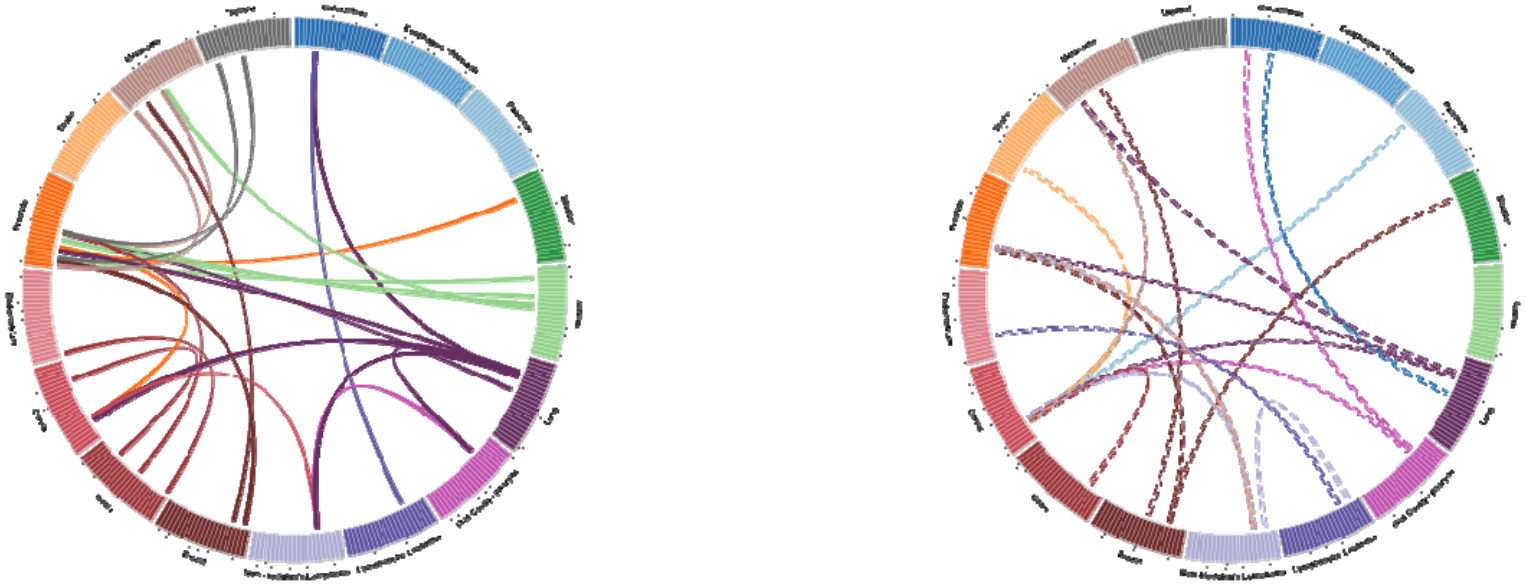
65. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet.* 2019;51(1):76-87.

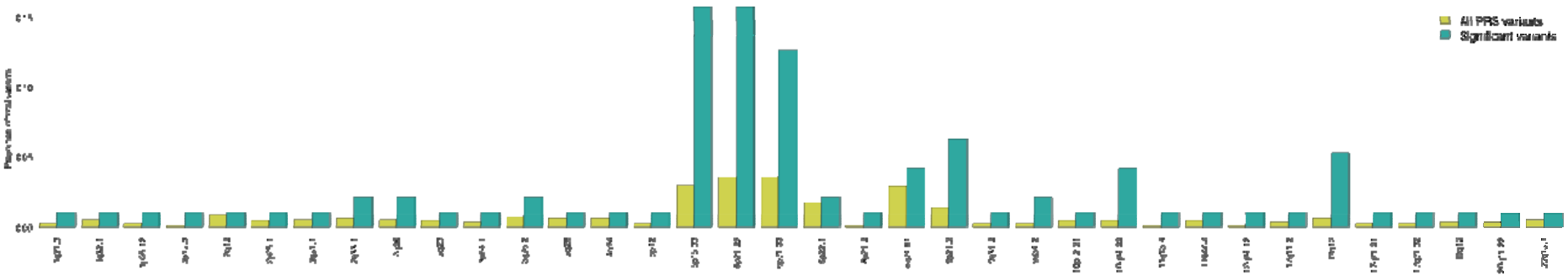
Figure 1. Odds ratios for at least nominally significant associations between cancer-specific polygenic risk scores (PRS) and cancer outcomes, based on meta-analyses of European ancestry participants from the Genetic Epidemiology Research on Aging (GERA) cohort and UK Biobank. Cancers are ordered based on hierarchical clustering of the odds ratios for each PRS across cancer outcomes.



bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

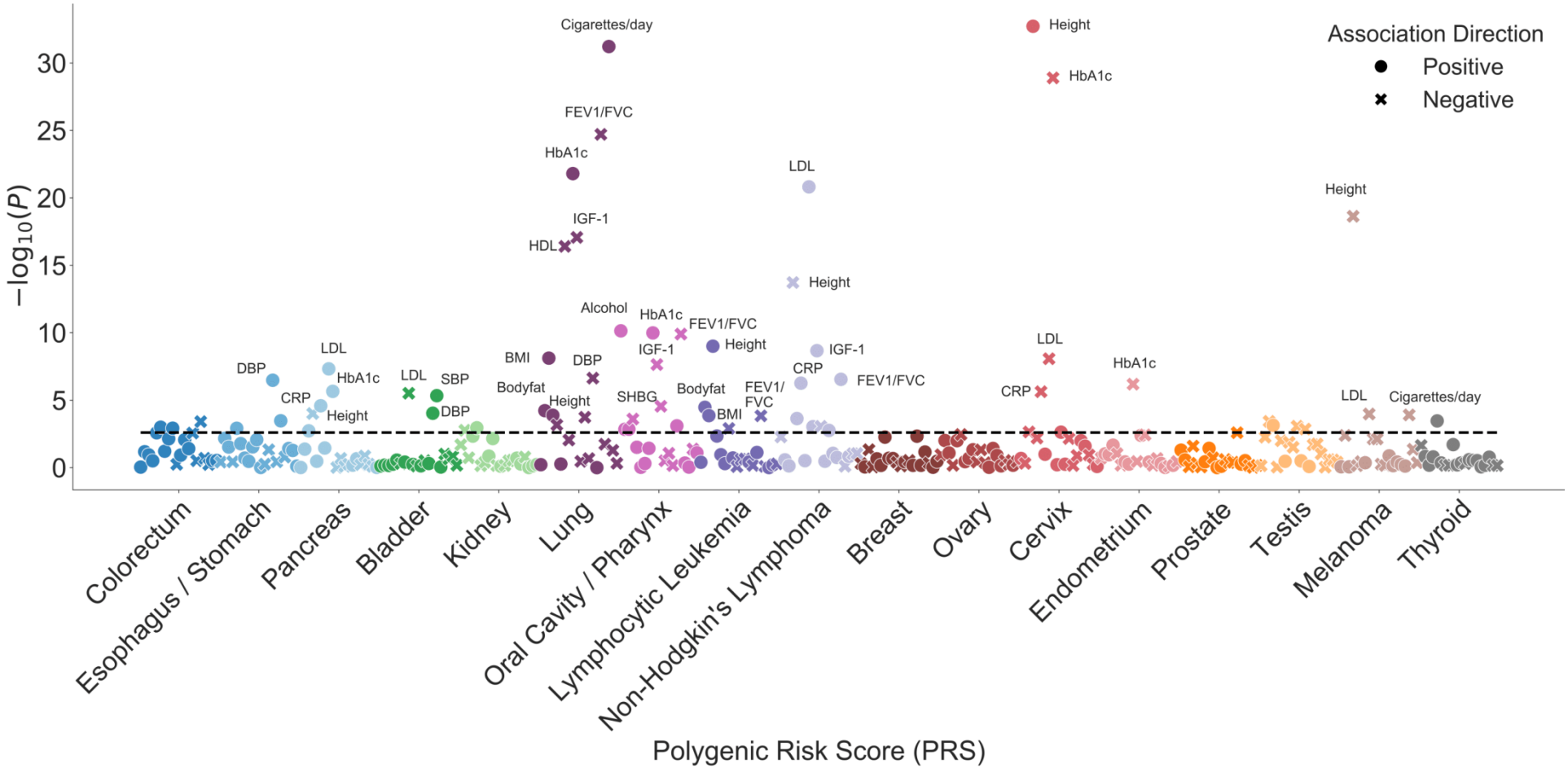
Figure 2. Pleiotropic risk variants from the 17 cancer-specific polygenic risk scores (PRS). (a) Circos plot describing each *positive* association between a known risk variant for one cancer type and a novel cancer phenotype. (b) Circos plot describing each *inverse* association between a known risk variant for one cancer type and a novel cancer phenotype. Each line in (a) and (b) represents a significant association, corrected for the number of independent tests (p<0.05/808=6.2x10^{-5}), between a risk variant for the cancer from which the line originates (denoted by line color) and the cancer type to which the line connects. Cancers are organized by organ site. (c) Region enrichment for 141 significant novel and known associations compared to all PRS variants.





bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Associations between each cancer-specific polygenic risk score (PRS) and 20 cancer risk factors and related serum biomarkers. All associations were estimated in cancer-free controls in the UK Biobank. Circles denote positive associations between the PRS and the secondary phenotype; crosses denote an inverse direction of association. The dashed line indicates the Bonferroni-corrected significance threshold (p<0.05/20 = 0.0025).



Abbreviations: BMI – Body Mass Index; CRP – C-Reactive Protein; DBP – Diastolic Blood Pressure; FEV1 – Forced Expiration Volume in the First Second; FVC – Forced Vital Capacity; HbA1c – Hemoglobin A1c; HDL – High-Density Lipoprotein; IGF-1 – Insulin Like Growth Factor-1; LDL – Low-Density Lipoprotein; SBP – Systolic Blood Pressure; SHBG – Sex Hormone Binding Globulin

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.18.911578; this version posted January 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.