Research Article

Quantifying the Genetic Correlation between Multiple Cancer Types


Abstract

Background: Many cancers share specific genetic risk factors, including both rare high-penetration mutations and common SNPs identified through genome-wide association studies (GWAS). However, little is known about the overall shared heritability across cancers. Quantifying the extent to which two distinct cancers share genetic origin will give insights to shared biological mechanisms underlying cancer and inform design for future genetic association studies.

Methods: In this study, we estimated the pair-wise genetic correlation between six cancer types (breast, colorectal, lung, ovarian, pancreatic, and prostate) using cancer-specific GWAS summary statistics data based on 66,958 case and 70,665 control subjects of European ancestry. We also estimated genetic correlations between cancers and 14 noncancer diseases and traits.

Results: After adjusting for 15 pair-wise genetic correlation tests between cancers, we found significant ($P < 0.003$) genetic correlations between pancreatic and colorectal cancer ($r_g = 0.55, P = 0.003$), lung and colorectal cancer ($r_g = 0.31, P = 0.001$). We also found suggestive genetic correlations between lung and breast cancer ($r_g = 0.27, P = 0.009$), and colorectal and breast cancer ($r_g = 0.22, P = 0.011$). In contrast, we found no evidence that prostate cancer shared an appreciable proportion of heritability with other cancers. After adjusting for 84 tests studying genetic correlations between cancer types and other traits (Bonferroni-corrected $P$ value: 0.0006), only the genetic correlation between lung cancer and smoking remained significant ($r_g = 0.41, P = 1.03 \times 10^{-10}$). We also observed nominally significant genetic correlations between body mass index and all cancers except ovarian cancer.

Conclusions: Our results highlight novel genetic correlations and lend support to previous observational studies that have observed links between cancers and risk factors.

Impact: This study demonstrates modest genetic correlations between cancers; in particular, breast, colorectal, and lung cancer share some degree of genetic basis. Cancer Epidemiol Biomarkers Prev; 26(9): 1427–35. © 2017 AACR.
Introduction

In the United States, cancer remains the second leading cause of death, with an estimated 1.69 million new cancer diagnoses and 600,000 cancer-related deaths in 2017 (1). Six cancer types—breast, colorectal, lung, ovarian, pancreatic, and prostate cancer—contribute more than 50% of annual cancer diagnoses (1). The etiologies of these cancers are complex, and associated heritability estimates (2) from twin studies range between 0.15 (colon) and 0.57 (prostate), indicating genetic components. For pancreatic cancer, data have been too sparse to estimate heritability based on twin studies. However, Muñoz and colleagues (2) did observe significant, albeit lower relative to other cancers, familial risks for pancreatic cancer. These results are supported by a meta-analysis based on more than 6,500 cases, where having a relative diagnosed with pancreatic cancer was associated with a 1.8-fold [95% confidence interval (CI) 1.5–2.1] risk increase of pancreatic cancer (3).

Various cancers share both environmental and genetic risk factors, including rare high-penetrant mutations in genes such as BRCA2 which predisposes to breast, ovarian, lung, prostate, and pancreatic cancers (4). Genome-wide association studies (GWAS) have identified more 350 distinct (reported SNPs > 500-kb apart) genomic regions that are associated with cancer (5) of which several are shared between cancer types (6). For example, the 8q24 (7–9) and 10q25 (10–13) regions have been associated with multiple cancer types (pleiotropy), although specific alleles often differ. In contrast to pleiotropy, which does not take the direction of association into account, genetic correlation describes the genome-wide correlation in allele effects and thus, considers the allele-specific direction of association between two traits. Thus, the shared genetic etiology between two traits can be either due to a shared genetic variant (or variants) with nonequal non-zero effect sizes (pleiotropy) or via a correlation between effect sizes for causal variants (genetic correlation; ref. 14). Identifying genetic regions that are associated with multiple cancer types may be useful for determining mechanisms involved in global carcinogenesis. However, the benefit of simultaneously studying multiple cancer types relies on their genetic correlation.

With the introduction of GWAS, it is now possible to quantify the phenotypic variance explained by genotyped single nucleotide polymorphisms (SNP) in single-trait (h²; refs. 15–17) and two-trait (ρg; ref. 18) settings by using variance component methods. Lu and colleagues (19) used this approach to estimate h² for 12 cancers and found significant nonzero genetic contribution for eight of them. However, their average sample size for each cancer was only 1,793 cases (range: 564–2,848) and 2,200 controls (range: 574–3,159), leading to imprecise heritability estimates. Sampson and colleagues (20) estimated the genetic correlation across 13 different cancers in an average sample size across cancers of 3,807 (range: 535–5,942) cases and 2,625 (range: 1,056–10,857) controls. Although no genetic correlation withstood multiple testing after adjusting for 91 tests, they observed the strongest genetic correlations (all P < 0.01) between kidney and testis cancer (rG = 0.73, SE = 0.28), diffuse large B-cell lymphoma and osteosarcoma (rG = 0.53, SE = 0.21), diffuse large B-cell lymphoma and chronic lymphocytic leukemia (rG = 0.51, SE = 0.18), and bladder and lung (rG = 0.35, SE = 0.14) cancer.

The sample sizes of these prior studies make interpretation of their findings difficult. Indeed, a drawback with the variance component approach is its requirement for individual-level data, which prohibits researchers from leveraging GWAS results based on meta-analyses which are often based on much larger sample sizes. The recently developed cross-trait linkage disequilibrium (LD) score regression approach overcomes this limitation by estimating the proportion of phenotypic variance explained by common SNPs (21) and the genetic correlation between two traits (22) using summary statistics only. Here, we set out to quantify pair-wise genetic correlation across breast, colorectal, lung, ovarian, pancreatic, and prostate cancer, capitalizing on summary statistics obtained from GWAS data in 66,958 case and 70,665 control subjects obtained from the GAME-ON, PanScan and GECCO consortia. In addition, we estimated the genetic correlation between each of these cancers and 14 noncancer traits which have all been suggested to be linked to cancer and for which we had access to GWAS summary statistics.

Materials and Methods

The GAME-ON network of consortia for post-GWA research, PanScan, and GECCO

We utilized three large-scale cancer genetic epidemiological consortia: GAME-ON, PanScan III, and GECCO. The Genetic Associations and Mechanisms in Oncology (GAME-ON) consortium is a network of cancer-specific post-GWA initiatives. The five GAME-ON sites are breast (DRIVE), colorectal (CORECT), lung (TRICI-ILLCO), ovarian (FOCI), and prostate (ELLIPSE; ref. 23). One of the main goals with GAME-ON was to test hypotheses across the cancer types that might illuminate common mechanisms of susceptibility. PanScan is a part of The Pancreatic Cancer Cohort Consortium with the goal of conducting GWAS to identify susceptibility markers for pancreatic cancer. For this study, we utilized genome-wide summary statistics from PanScan III (24). The Genetic Epidemiology of Colorectal Cancer Consortium (GECCO) is a large collaborative consortium evaluating genetic and environmental risk factors for colorectal cancer (25). Details of GAME-ON, PanScan, GECCO and the participating studies are available at http://epi.grants.cancer.gov/gameon/, http://epi.grants.cancer.gov/PanScan/ and https://www.fredhutch.org/en/labs/phs/projects/cancer-prevention/projects/gecco.html. Sample sizes for each cancer is listed in Table 1. These studies have been described in detail previously (23). For each cancer type, genotyping was performed using Illumina or Affymetrix arrays of varying densities described elsewhere (23, 26, 27). For all studies except GECCO, imputation was performed using Illumina or Affymetrix arrays of varying densities described elsewhere (23, 26, 27). For all studies except GECCO, imputation was performed using the 1,000 Genomes Project Phase 1 version 3 reference haplotypes, resulting in up to approximately 10 million SNPs available for the analysis for each cancer type. For GECCO, data was imputed using an in-house reference panel of 2,159 whole-genome sequenced European ancestry GECCO participants (28). As imputation quality scores were not readily accessible across all GWAS, we only included HapMap 3 SNPs as a proxy for well-imputed SNPs (29).

Noncancer traits

We also estimated the genetic correlations between the six cancer types and 14 noncancer traits (29) for which GWAS summary statistics were publicly available (Supplementary Table S1). The included traits were schizophrenia, bipolar disorder, coronary artery disease, type II diabetes, Crohn disease, ulcerative colitis, rheumatoid arthritis, ever/never smoked tobacco, height,
body mass index (BMI), fasting serum glucose, triglycerides, LDL cholesterol, and HDL cholesterol. The average sample size for the noncancer traits was 70,488 subjects.

### Statistical analysis

Associations between SNPs and cancer risk were estimated by unconditional logistic regression adjusted for age, sex (when applicable), and top principal components (ranging from two to six across contributing GWAS) to adjust for potential population stratification. We used a newly developed method for estimating heritability due to common SNPs and genetic correlations that employs only summary statistics together with linkage disequilibrium (LD) information from a reference panel (21, 22). This method, known as LD score regression, relies on the fact that SNP-specific association statistics reflect the associations of all SNPs in LD with that SNP. Thus, for a polygenic trait, SNPs in high-LD regions will on average have higher \( h^2 \) statistics than SNPs in low-LD regions and similarly, for two polygenic, genetically correlated traits with z-scores \( z_1 \) and \( z_2 \), the product \( z_1z_2 \) will on average be higher for SNPs with high LD than SNPs with low LD. Formally, the relationship between the expected \( \chi^2 \) statistic for SNP \( j \) and the LD score \( l(j) \) for SNP \( j \) can be described by

\[
E[\chi^2_j] \approx \frac{N_j}{2} l(j) + 1,
\]

where \( N_j \) is the sample size, \( h^2_j \) is the heritability due to included SNPs, \( M \) is the number of SNPs and \( l(j) := \sum r^2(j,k) \) where \( r^2(j,k) \) is the correlation between SNP \( j \) and all other SNP \( k \). As the observed SNP heritability estimates for binary traits are not directly comparable with more traditional estimates of heritability (e.g., from twin studies), we transformed the observed heritability to the liability scale which takes both ascertainment and disease prevalence into account as described previously (30). To obtain estimates of cancer-specific prevalence, we used SEER cumulative risks (31). We can extend the calculations of single-trait heritability to include two traits:

\[
E[z_1 z_2] = \frac{\sqrt{N_1 N_2}}{N_1 N_2} l_1 l_2 + \frac{N_1 N_2}{N_1 N_2}
\]

where \( r_g \) is the genetic covariance, \( N_1 \) and \( N_2 \) are sample sizes for trait 1 and 2 respectively, and \( N \) is the number of overlapping samples and \( \rho \) is the phenotypic correlation in the overlapping samples. We estimate \( r_g \) by the slope of the regression of \( z_1/z_2 \) on the LD score and tested for its difference from 0, as described previously (22). In total, we conducted 15 pair-wise genetic correlation tests between cancer types at a significance threshold of \( P = 0.05/15 = 0.003 \).

### Results

Cancer-specific GWAS summary statistics from data of 66,958 case and 70,665 control subjects across six cancer types were shared between the cancer sites. We first estimated the heritability due to common SNPs (\( h^2_g \)) on the observed and liability scale (Table 1) using cancer-specific cumulative risks (Supplementary Table S2) as reported in SEER (31). Heritability estimates on the liability scale ranged between 0.04 (pancreatic cancer) and 0.27 (prostate cancer). Using cross-trait LD score regression, we quantified the pairwise genetic correlations between breast, colorectal, lung, ovarian, pancreatic and prostate cancer. A summary of the results is displayed as a 6 \( \times \) 6 matrix (Table 2) with the estimated pair-wise genetic correlations and associated SEs in the upper right corner of the matrix, and corresponding \( P \) values for each genetic correlation can be seen in the lower left corner of the matrix. We found significant positive genetic correlations between pancreatic and colorectal cancer (\( r_g = 0.55 \), \( P = 0.003 \)), lung and colorectal cancer (\( r_g = 0.31 \), \( P = 0.001 \)), as well as suggestive positive genetic correlations between lung and breast cancer (\( r_g = 0.27 \), \( P = 0.009 \)), and colorectal and breast cancer (\( r_g = 0.22 \), \( P = 0.01 \)).

We also estimated the genetic correlation between each of these six cancer types and 14 other diseases and traits implicated to be linked to cancer [schizophrenia, bipolar disorder, coronary artery disease, type II diabetes, Crohn disease, ulcerative colitis, rheumatoid arthritis, ever/never smoked tobacco, height, BMI, fasting serum glucose, triglycerides, LDL cholesterol and HDL cholesterol (Fig. 1; Table 3; Supplementary Tables S1 and S3)]. The strongest positive genetic correlation was, as expected, between lung cancer and smoking status (never/ever; \( r_g = 0.41 \),

### Table 1. Overview of cancers analyzed

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Cases</th>
<th>Controls</th>
<th>( h^2_g ) (SE) – observed scale</th>
<th>( h^2_g ) (95% CI) – liability scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>15,748</td>
<td>18,084</td>
<td>0.12 (0.02)</td>
<td>0.14 (0.09–0.18)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>15,716</td>
<td>18,154</td>
<td>0.13 (0.02)</td>
<td>0.11 (0.07–0.14)</td>
</tr>
<tr>
<td>Lung</td>
<td>12,160</td>
<td>16,838</td>
<td>0.14 (0.03)</td>
<td>0.13 (0.08–0.19)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>4,369</td>
<td>9,123</td>
<td>0.10 (0.04)</td>
<td>0.07 (0.02–0.12)</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>5,107</td>
<td>8,845</td>
<td>0.07 (0.04)</td>
<td>0.05 (0.0–0.10)</td>
</tr>
<tr>
<td>Prostate</td>
<td>14,160</td>
<td>12,724</td>
<td>0.25 (0.03)</td>
<td>0.27 (0.21–0.33)</td>
</tr>
</tbody>
</table>

*NOTE: Number of subjects, observed heritability (standard error) and heritability (95% CI) on the liability scale explained by studied SNPs for each cancer. Cumulative risks used for calculating \( h^2_g \) on the liability scale were obtained from SEER.*

### Table 2. Genetic correlations between cancers

<table>
<thead>
<tr>
<th></th>
<th>Breast</th>
<th>Colorectal</th>
<th>Lung</th>
<th>Ovarian</th>
<th>Pancreatic</th>
<th>Prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>1</td>
<td>0.22 (0.091)</td>
<td>0.27 (0.11)</td>
<td>-0.08 (0.13)</td>
<td>0.17 (0.19)</td>
<td>0.06 (0.09)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>0.014</td>
<td>1</td>
<td>0.31 (0.097)</td>
<td>-0.17 (0.17)</td>
<td>0.55 (0.19)</td>
<td>0.09 (0.07)</td>
</tr>
<tr>
<td>Lung</td>
<td>0.009</td>
<td>0.001</td>
<td>1</td>
<td>-0.17 (0.17)</td>
<td>0.32 (0.19)</td>
<td>0.095 (0.08)</td>
</tr>
<tr>
<td>Ovarian</td>
<td>0.18</td>
<td>0.57</td>
<td>0.32</td>
<td>1</td>
<td>-0.40 (0.29)</td>
<td>0.02 (0.14)</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>0.37</td>
<td>0.003</td>
<td>0.08</td>
<td>0.17</td>
<td>1</td>
<td>-0.06 (0.14)</td>
</tr>
<tr>
<td>Prostate</td>
<td>0.52</td>
<td>0.2</td>
<td>0.25</td>
<td>0.89</td>
<td>0.68</td>
<td>1</td>
</tr>
</tbody>
</table>

*NOTE: Genetic correlations with standard errors are in the upper right part of the table; corresponding \( P \) values are in the lower left part of the table. Nominal significant (\( P < 0.05 \)) genetic correlations are highlighted in bold.*
NOTE: For each cancer-trait genetic correlation observed, current evidence from observational studies and corresponding reference are also listed.

Table 3. Nominally significant genetic correlations between cancers and noncancer traits

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Noncancer trait</th>
<th>r_g</th>
<th>SE</th>
<th>P</th>
<th>Epidemiologic observations</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>Ulcerative colitis</td>
<td>0.243</td>
<td>0.08</td>
<td>0.002</td>
<td>+</td>
<td>55</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Schizophrenia</td>
<td>0.142</td>
<td>0.049</td>
<td>0.004</td>
<td>+</td>
<td>56</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Height</td>
<td>0.143</td>
<td>0.055</td>
<td>0.03</td>
<td>+</td>
<td>57</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Triglycerides</td>
<td>-0.13</td>
<td>0.061</td>
<td>0.033</td>
<td>-</td>
<td>58</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>BMI</td>
<td>-0.112</td>
<td>0.053</td>
<td>0.035</td>
<td>- (premenopausal)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+ (postmenopausal)</td>
<td>35</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>BMI</td>
<td>0.157</td>
<td>0.046</td>
<td>6.2E–04</td>
<td>+</td>
<td>35</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Triglycerides</td>
<td>0.126</td>
<td>0.058</td>
<td>0.029</td>
<td>+</td>
<td>59</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Schizophrenia</td>
<td>-0.091</td>
<td>0.046</td>
<td>0.048</td>
<td>No association</td>
<td>56</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Ever/never smoked</td>
<td>0.412</td>
<td>0.084</td>
<td>1.0E–06</td>
<td>+</td>
<td>60</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>HDL</td>
<td>-0.151</td>
<td>0.063</td>
<td>0.017</td>
<td>-</td>
<td>59</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>BMI</td>
<td>0.116</td>
<td>0.054</td>
<td>0.032</td>
<td>- (smokers)</td>
<td>35</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Type-2 diabetes</td>
<td>0.469</td>
<td>0.191</td>
<td>0.014</td>
<td>+</td>
<td>61</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Ulcerative colitis</td>
<td>0.291</td>
<td>0.137</td>
<td>0.034</td>
<td>No association</td>
<td>55</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>HDL</td>
<td>-0.405</td>
<td>0.167</td>
<td>0.015</td>
<td>Unknown</td>
<td>N/A</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>BMI</td>
<td>0.243</td>
<td>0.077</td>
<td>0.038</td>
<td>+</td>
<td>35</td>
</tr>
</tbody>
</table>

NOTE: For each cancer-trait genetic correlation observed, current evidence from observational studies and corresponding reference are also listed. P values are not corrected for multiple testing.

Discussion

It is well known that cancer tends to cluster in families, which has been attributed to shared environmental factors and genetics (2). Many lifestyle factors associated with cancer including smoking, obesity, and alcohol intake have been shown to influence risk of multiple cancer types, implying that different cancer types share an underlying biological mechanism. Similarly, genetic variation in genes such as BRCA1/2 and TERT has been associated with risks of multiple different cancer types, providing empirical support that there are specific regions in the genome that harbor genetic variation influencing risk of multiple cancer sites. However, although many cancer types might share susceptibility loci (i.e., pleiotropy), their genetic correlation, which depends on specific risk alleles and their direction of associations, might not be strong. Here, we aimed to assess the latter among six cancers, breast, colorectal, lung, ovary, pancreatic and prostate, as well as between cancer types and seven additional disease and seven nondisease traits that have all been found to have heritable components.

We used SEER estimates (31) to obtain cancer-specific cumulative risks, recognizing that not all study subjects came from U.S. populations. For comparison, we also calculated the heritability on the liability scale using cumulative risks obtained from Mucci and colleagues (2). No qualitative difference was observed for the liability estimates using the two different sources of cumulative risks. In concordance with previous studies, the cancer-specific heritability estimates observed here were lower than what has been observed in twin studies. This is not unexpected given that here, we are only estimating the additive genetic component based on common SNPs captured by GWAS, and thus, any contribution to the heritability based on factors such as gene–gene interactions, gene–environment interactions, structural variants or rare variants, will not be captured by our analysis. Among the studied cancers, we observed the largest heritability for prostate cancer (h² = 0.27) in agreement with previous twin studies (2). We also compared our cancer-specific heritability results to previous studies estimating heritability based on GWAS data. In general, our results were comparable with previous studies. Lu and colleagues (19) estimated h² for breast cancer to be 0.13 (95% CI: 0.0–0.56) compared with our estimate of 0.14 (95% CI:...
0.09–0.18). For lung cancer, previous estimates have varied with Lu and colleagues (19) estimating \( h^2_g \) to 0.10 (95% CI: 0.0–0.24) in European populations while Sampson and colleagues (20) estimated \( h^2_g \) to be 0.21 (95% CI: 0.14–0.27), compared with our estimate of 0.13 (95% CI: 0.08–0.19). For ovarian cancer, we observed a small heritability \( h^2_g = 0.07 \), 95% CI: 0.02–0.12) compared to Lu and colleagues (ref. 19; \( h^2_g = 0.30 \), 95% CI: 0.18–0.42). It is not clear why we observe this discrepancy in results. We also observed lower \( h^2_g \) of pancreatic cancer (0.05, 95% CI: 0–0.10) than previously observed [0.18, 95% CI: 0.06–0.30 for Lu and colleagues (19) and 0.10, 95% CI: 0.04–0.16 for Sampson and colleagues (20)] but we note that the CIs are wide and overlap. For prostate cancer, we observed a heritability of 0.27 (95% CI: 0.21–0.37) compared with Lu (ref. 19; 0.81, 95% CI: 0.32–1), and Sampson (ref. 20; 0.29, 95% CI: 0.15–0.42). We note that heritability estimates reported here are all on the liability scale and were calculated using SEER rates for all three studies including ours.

We found that colorectal cancer showed significant genetic correlations with pancreatic and lung cancers, with the largest genetic correlation observed for the two gastrointestinal tract cancers: the \( r_g \) for colorectal and pancreatic cancer was 0.55 (\( P = 0.003 \)). Amundadottir and colleagues studied cancer risk for first up to fifth degree relatives in an Icelandic population and observed an increased risk for pancreatic cancer among colon cancer patients (and vice versa) for first degree relatives but not beyond (32). Colorectal cancer patients have been observed to have a higher incidence of pancreatic cancer than the general population (33). Furthermore, Lynch syndrome, the most common hereditary colorectal cancer syndrome, has also been shown to increase risk for pancreatic cancer (34). Obesity is a well-established risk factor for both colorectal and pancreatic cancer (35, 36) and we observed nominally significant \( (P < 0.05) \) genetic correlations between BMI and both colorectal and pancreatic cancer.

Colorectal cancer also showed suggestive genetic correlation with breast cancer in agreement with the study from Amundadottir and colleagues (32). A recent study found that women diagnosed with breast cancer have a 1.59-fold (95% CI: 1.53–1.65) increased risk of developing colorectal cancer compared with the general population (37).

Breast and lung cancer showed a suggestive positive genetic correlation (\( r_g = 0.27, P = 0.009 \)), supported by observational studies finding familial cosegregation of the two cancers (38, 39) as well as overlap in multiple susceptibility genes such as BRCA2, CHEK2 (40) and LSP1 (7). In contrast, Amundadottir and colleagues did not observe a significant cooccurrence among relatives (32). A recent cross-cancer GWAS based on the GAME-ON data (23) identified a pleiotropic locus at 1q22 that was associated with both breast and lung cancer.

We expect these results to generate testable hypotheses about mechanisms. There are data to support inflammation response, DNA repair and stress responses, to list just a few. For example, the genetic correlations between breast, colorectal and lung cancer might be driven in part by genetic variants in the inflammation pathway. Indeed, a recent analysis of genetic variation in the inflammation pathway from the GAME-ON consortium identified SH2B3, a key negative regulator of cytokine signaling to be associated with all three cancers (41). Furthermore, the authors found no evidence that genetic variation in inflammation-related genes was associated with prostate cancer risk. This observation corroborates our findings that prostate cancer does not share an appreciable genetic component with other cancers and supports inflammation to be a pathway in which genetic variation affects the risk of breast, colorectal, and lung cancer. Removing all SNPs with a \( p \) value test statistic >25 in the individual cancer GWAS, did not change our results: colorectal–breast cancer \( (r_g = 0.22 \) for all SNPs and \( r_g = 0.23 \) excluding significant SNPs); colorectal–lung cancer \( (r_g = 0.31 \) for all SNPs and \( r_g = 0.34 \) excluding significant SNPs); colorectal–pancreatic cancer \( (r_g = 0.55 \) for all SNPs and \( r_g = 0.58 \) excluding significant SNPs); breast–lung cancer \( (r_g = 0.27 \) for all SNPs and \( r_g = 0.33 \) excluding significant SNPs). Thus, it is likely that the genetic correlations that we observe are due to yet unidentified SNPs, and future studies should focus on simultaneously study genetically correlated cancers with the goal of identifying SNPs that are associated with multiple cancer types.

We did not observe evidence that either ovarian or prostate cancer shared an appreciable amount of heritability with other cancers, although our sample size for ovarian cancer was relatively small, leading to wide CIs. We had higher statistical power to detect genetic correlations involving prostate cancer, but no estimate was >0.1, suggesting that prostate cancer has a genetic contribution that is distinct from that of other cancer types studied here. Witte and Hoffmann used polygenic risk scores to investigate a potential shared heritability between breast and prostate cancer, and although they observed a potential common polygenic model between nonaggressive prostate cancer and breast cancer, they observed no evidence of a common model between highly aggressive prostate cancer and breast cancer (42). It is important to note that our analysis does not capture rare higher penetrance mutations, and thus, the observed increased risk of multiple cancers among relatives to prostate cancer cases (32) is likely to at least in part be attributable to rare variants such as BRCA2.

In addition, we examined the genetic correlation between these cancer types and 14 noncancer diseases and traits. The only genetic correlation between cancer and noncancer traits that withstood correction for multiple testing was smoking status and lung cancer. The strongest lung cancer susceptibility locus is in the 15q25 region which contains genes encoding the nicotinic acetylcholine receptor subunits CHRNA5, CHRNA3, and CHRNB4 and has also been associated with smoking behavior with a trait results here (42). It is important to note that our analysis does not capture rare higher penetrance mutations, and thus, the observed increased risk of multiple cancers among relatives to prostate cancer cases (32) is likely to at least in part be attributable to rare variants such as BRCA2.
possible that smoking and BMI to some extent affect lung cancer risk through the same biological mechanism.

We observed very few genetic correlations with prostate cancer compared with the other cancers of comparable sample size and no genetic correlation with noncancer traits were >0.1. We also note the lack of genetic correlation between prostate cancer and type II diabetes ($r_g = 0.03$, 95% CI $-0.10$–0.15, $P = 0.67$). The epidemiologic inverse association between prostate cancer and type II diabetes is well-documented (50–53) and a previous study showed that 10 of 36 type II diabetes SNPs were associated (2 with increased risk and 8 with decreased risk) with advanced prostate cancer (54). The different directions of significant prostate cancer associations across type II diabetes SNPs are consistent with the lack of genetic correlation (which is sensitive to direction of effects) observed in this study.

Sampson and colleagues previously used individual-level GWAS data to estimate genetic correlations between 13 cancers in 49,492 cancer cases and 34,131 controls, including estrogen receptor negative (ER−) breast cancer, lung cancer, pancreatic, and prostate cancer (20). Although, they did not observe the statistically significant genetic correlations between the cancers studied here, their SEs were in general large, making it difficult to compare the results. We note that the ER− breast cancer and pancreatic datasets they used are a subset of the data analyzed here. While we had access to GWAS summary statistics based on cancer subtypes including ER− breast cancer, squamous cell lung cancer, lung adenocarcinoma, serous, clear cell, and endometrioid ovarian cancer and aggressive prostate cancer, sample sizes for some subtypes were too small for meaningful analysis. The basis of our experience, LD score regression requires at least 10,000 cases for adequately stable estimates at these heritabilities. We note a few limitations with only having access to summary statistics data compared to individual-level data. Most importantly, the SEs associated with the estimated genetic correlations based on LD score regression are larger than what is obtained by similar methods using individual-level data. In addition, we are not able to conduct any subtype analysis on the original traits that might be of interest, for example, it might have been of interest to study the genetic correlation between BMI and breast cancer stratified by menopausal status. LD score regression leverages summary statistics rather than individual-level data and thereby overcome many of the issues associated with relying on individual-level data. Moreover, appropriate quality control steps were conducted as part of the cancer-specific GWAS meta-analysis. Furthermore, we limited our analysis to HapMap 3 SNPs to ensure well-imputed data. In summary, our results indicate that some cancers show modest genetic correlations; in particular, breast, colorectal and lung cancer share some degree of genetic basis. In contrast, prostate cancer appears to have a unique genetic architecture that is not shared with breast, lung, ovarian, and pancreatic cancer. Furthermore, a number of cancer types show genetic correlations with obesity, highlighting the involvement of adiposity-related processes in cancer. As GWAS sample sizes continue to increase and GWAS summary statistics from other traits become available, we will be able to additionally characterize the shared heritability between cancer types including histologic subtypes as well as with noncancer traits.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Disclaimer**

The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in CORECT, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government or CORECT.

**Authors’ Contributions**

**Conception and design:** S. Lindström, F.R. Schumacher, C.I. Amos, E.L. Goode, D.J. Hunter, P. Kraft

**Development of methodology:** H. Finucane, B. Bulik-Sullivan, B. Neale, A. Price, P. Kraft


**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** S. Lindström, H. Finucane, B. Bulik-Sullivan, F.R. Schumacher, R.J. Huang, S.B. Gruber, D.V. Conti, L.T. Amundadottir, H.A. Risch, L. Hsu, B. Neale, P. Kraft


**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** C.I. Amos, K.A. Rand, S.B. Gruber, R.Z. Stolzenberg-Solomon, H.A. Risch, R. Eeles, C.A. Haiman, D.J. Hunter, P. Kraft

**Study supervision:** R.Z. Stolzenberg-Solomon, J.R. Huyge, J.C. Claude, A. Chan, R. Eeles, D.F. Easton, D.J. Hunter, P. Kraft

**Acknowledgments**

**Members of CORECT**


**Members of DRIVE**

DRIVE acknowledges the following GWAS and investigators that shared genome-wide summary data as part of the breast-cancer GWAS meta-analysis: the Australian Breast Cancer Family Study (ABCFS; John L. Hopper, Melissa C. Southey, Enes Makalic, Daniel F. Schmidt), the British Breast Cancer Study (BBCS; Olivia Fletcher, Julian Peto, Lorna Gibson, Isabel dos Santos Silva), the Breast and Prostate Cancer Consortium (BPC3) (David J. Hunter, Sara Lindstrom, Peter Kraft), the Breast Cancer Family Registries (BCFR; Habib Ahsan, Alice Whittemore), the Dutch Familial Bilateral Breast Cancer Study (DFBBCS; Quentin Wansfuus, Hanne Meijers-Heiboe, Muriel Adank, Rob B. van der Luijt, Andre G. Uitterlinden, Albert Hofman), German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC; Alfons Meindl, Rita K. Schmutzler, Bertram Müller-Myskow, Peter Lichtner), the Helsinki Breast Cancer Study (HEBCS; Hei Nevalintila, Tanu A. Mura, Maija Aittomaki, Carl Eklund), the Mammary Carcinoma Risk Factor Investigation (MARIE; Jenny Chang-Chung, Rebecca Hein, Norbert Dahmen, Lars Beckman), SardiNIA (Laura Crisponi), the Singapore and Sweden Breast Cancer Study (SASBAC; Per Hall, Kamila Czene, Astrid Iwamoto, Jianjun Liu), and the UK2 (Douglas F. Easton, Clare Turnbull, Nazneen Rahman).

**Members of ELLIPSE**

ELLIPSE acknowledges the following GWAS and investigators that shared genome-wide summary data as part of the prostate cancer GWAS meta-analysis: CRUK (Rosalind Eeles, Douglas F. Easton, Zsofia Kote-Jarai, Kenneth Muir, Graham Giles, Gianluca Severi, David Neal, Jenny L. Donovan, Freddie C. Hamdy), CAPS1 and CAPS2 (FredrikWiklund, Henrik Gronberg), BPC3-MEC (Christopher Haiman, Fred Schumacher), BPC3-EPIC (Buth Travis, Elio Riboli), BPC3-Harvard (Peter Kraft, David Hunter), BPC3-ACS (Susan Gasparrini), PEGASIS (Sonja Berndt, Stephen Chanock).
The Genetic Correlation between Multiple Cancer Types

Members of TRICL

TRICL acknowledges the following investigators: Youyghun Han, Li Su, Yonguey Wei, Rayecan J. Hung, Yonathan Bihane, John McLaughlin, Paul Brennan, James D. McKay, Heike Bickelbauer, Albert Rosenberger, Richard S. Houlston, Neel Caporaso, Maria Teresa Landi, Joachim Heinrich, Angela Risch, Xifeng Wu, Yuanqing Ye, David C. Christiani, Christopher I. Amos.

PanScan: The authors wish to thank all participants and participating institutions, investigators, and staff that helped make this project possible. The authors acknowledge the contribution of the staff of the Cancer Genomics Research Laboratory (CGR) at the National Cancer Institute, National Institutes of Health, Bethesda, MD, for their help throughout the project.

ASTERISK: A Hospital Clinical Research Program (PHRC-BRD09/C) from the Regional Council of Pays de la Loire, the Groupement des Entreprises de Biologie Clinique (MEC) and the Ligue Régionale Contre le Cancer (LRCC).

HPFS, NHS, and PHS: We would like to thank the participants and staff of the Nurses’ Health Study and the Health Professionals Follow-Up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WV. The authors assume full responsibility for analyses and interpretation of these data.

PanScan: The authors wish to thank all participants and cooperating clinicians, and Ute Handue-Deub, Ute Benscheid, Muhabbet Celik and Ursula Elbier for excellent technical assistance.

Research Laboratory (CGR) at the National Cancer Institute, National Institutes of Health, Bethesda, MD, for their help throughout the project.

The authors acknowledge Dave Duggan and team members at Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent assistance in this study, including the patients and the healthy control persons, as well as all of the physicians, technicians, and students.

DACHS: We thank all participants and cooperating clinicians, and Ute Handue-Deub, Ute Benscheid, Muhabbet Celik and Ursula Elbier for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible, in particular, Peter Brossard, the Canadian Cancer Agency, Dr. Helene Deschenes, and the Canadian Cancer Society.

Additional funding toward genetic analyses of OFCCR includes the Ontario Registry for Studies of Familial Colorectal Cancer (U10 CA074783); previous CCFR section.

Alteration of a region of chromosome 5p15.2?

The authors wish to thank all the patients and the healthy control persons, as well as all of the physicians, technicians, and students.

Nurses’ Health Study and the Health Professionals Follow-Up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WV. The authors assume full responsibility for analyses and interpretation of these data.

PanScan: The authors would like to thank the participants and staff of the nurses’ health study and the health professionals follow-up study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WV. The authors assume full responsibility for analyses and interpretation of these data.

Grant Support

This work was supported by the Genetic Associations and Mechanisms in Oncology Network, GAME-ON [http://gni.grants.gov/cancer/], which includes the following consortia: CORRECT (grant number U19 CA148107); DRIVE [H119 CA148065]; ECLIPSE [grant number H119 CA148537]; PCOC [H119 CA148112]; TRICL [U119 CA148127]. This work was also supported by NIH grants CA149439, CA173785 and CA189532. R. Eeles is supported by NHIR funding to the Biomedical Research Centre at The Institute of Cancer Research and Royal Stanley NHS Foundation Trust. HRG is supported by the Fannie and John Hertz Foundation.

PanScan is supported by the Intramural Research Program of the NIH, National Cancer Institute.

GENETICS, Cancer Research conducted genotyping.

PanScan: The authors wish to thank all the patients and the healthy control persons, as well as all of the physicians, technicians, and students.

Nurses’ Health Study and the Health Professionals Follow-Up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WV. The authors assume full responsibility for analyses and interpretation of these data.
References

42. Sacone NL, Culverhouse RC, Schwantes-An Th, Cannon DS, Chen X, Cichon S, et al. Multiple independent loci at chromosome 15q25.1 affect...


Quantifying the Genetic Correlation between Multiple Cancer Types

Sara Lindström, Hilary Finucane, Brendan Bulik-Sullivan, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-17-0211

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2017/06/21/1055-9965.EPI-17-0211.DC1

Cited articles
This article cites 56 articles, 8 of which you can access for free at:
http://cebp.aacrjournals.org/content/26/9/1427.full#ref-list-1

Citing articles
This article has been cited by 20 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/26/9/1427.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/26/9/1427.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.