Large-scale cross-cancer fine-mapping of the 5p15.33 region reveals multiple independent signals

Hongjie Chen,1 Arunabha Majumdar,2,3 Lu Wang,4 Siddhartha Kar,5 Kevin M. Brown,6 Helian Feng,7 Constance Turman,8 Joe Dennis,9 Douglas Easton,9 Kyriaki Michailidou,10,11 Jacques Simard,12 Breast Cancer Association Consortium (BCAC), Timothy Bishop,13 Iona C. Cheng,14 Jeroen R. Huyghe,15 Stephanie L. Schmit,16 Colorectal Transdisciplinary Study (CORECT), Colon Cancer Family Registry Study (CCFR), Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Tracy A. O’Mara,17 Amanda B. Spurdle,17 Endometrial Cancer Association Consortium (ECAC), Puya Gharahkhani,18 Johannes Schumacher,19 Janusz Jankowski,20,21 Ines Gockel,22 Esophageal Cancer GWAS Consortium, Melissa L. Bondy,23 Richard S. Houlston,24 Robert B. Jenkins,25 Beatrice Melin,26 Glioma International Case Control Consortium (GIxCC), Corina Lesseur,27,28 Andy R. Ness,29,30 Brenda Diergaarde,31,32 Andrew F. Olshan,33,34 Head-Neck Cancer GWAS Consortium, Christopher I. Amos,35 David C. Christiani,8,36 Maria T. Landi,6

Summary

Genome-wide association studies (GWASs) have identified thousands of cancer risk loci revealing many risk regions shared across multiple cancers. Characterizing the cross-cancer shared genetic basis can increase our understanding of global mechanisms of cancer development. In this study, we collected GWAS summary statistics based on up to 375,468 cancer cases and 530,521 controls for fourteen types of cancer, including breast (overall, estrogen receptor [ER]-positive, and ER-negative), colorectal, endometrial, esophageal, glioma, head/neck, lung, melanoma, ovarian, pancreatic, prostate, and renal cancer, to characterize the shared genetic basis of cancer risk. We identified thirteen pairs of cancers with statistically significant local genetic correlations across eight distinct genomic regions. Specifically, the 5p15.33 region, harboring the TERT and CLPTML genes, showed statistically significant local genetic correlations for multiple cancer pairs. We conducted a cross-cancer fine-mapping of the 5p15.33 region based on eight cancers that showed genome-wide significant associations in this region (ER-negative breast, colorectal, glioma, lung, melanoma, ovarian, pancreatic, and prostate cancer). We used an iterative analysis pipeline implementing a subset-based meta-analysis approach based on cancer-specific conditional analyses and identified ten independent cross-cancer associations within this region. For each signal, we conducted cross-cancer fine-mapping to prioritize the most plausible causal variants. Our findings provide a more in-depth understanding of the shared inherited basis across human cancers and expand our knowledge of the 5p15.33 region in carcinogenesis.

Introduction

Cancer is a major global public health problem. More than 19.3 million new cancer cases and 10 million cancer deaths were estimated to occur worldwide in 2020.1 In the United States, approximately 1.9 million individuals are projected to be newly diagnosed with cancer, and more than 600,000 affected individuals are projected to die of cancer in 2021.2 Inherited genetic variants, along with environmental exposures, contribute substantially to the pathogenesis of...
Cancers. Cancers tend to cluster in families, and twin studies have reported cancer-specific heritability ranging from 9% (head/neck) to 58% (melanoma).

Genome-wide association studies (GWASs) of specific types of cancer have identified genetic loci significantly associated with susceptibility to malignancies. In a recent study of 18 types of cancer in European ancestry populations, the authors identified 17 genome-wide significant variants that were associated with the risk of at least two cancers with the same direction of effect. The 8q24 region has been long recognized as a pleiotropic locus, where genetic variants have been associated with the risk of breast, colorectal, endometrial, glioma, ovarian, pancreatic, and prostate cancer, among others. The 5p15.33 region has been associated with more than ten types of cancer, with multiple independent risk alleles identified. Various biological mechanisms, including inflammation, epigenetics, gene expression, and telomere structure, have been proposed to explain these identified pleiotropic associations. For example, the 5p15.33 region harbors the TERT gene, which encodes the catalytic subunit of telomerase, as well as the CLPTML1 gene, which encodes the cleft lip and palate-associated transmembrane-1 like protein.

In addition, recent efforts have been devoted toward estimating the genetic correlation between pairs of cancers. Using the restricted maximum likelihood (REML) approach implemented in the GCTA tool, one study quantified the pairwise genetic correlation among 13 types of cancers in European ancestry populations. Four pairs of cancers, including bladder-lung, testis-renal, lymphoma-osteosarcoma, and lymphoma-leukemia, demonstrated statistically significant shared heritability. We have previously applied linkage disequilibrium (LD) score regression on cancer GWAS summary statistics and observed significant genetic correlations between multiple solid tumor pairs, including colorectal-lung, colorectal-pancreatic, breast-colorectal, breast-lung, breast-ovarian, and lung-head/neck cancer. However, these studies only quantified the pairwise genetic correlation on a genome-wide scale, ignoring variations in the local genetic correlation across the genome. As shared heritability between cancers may not be uniformly distributed across the genome, such limitation may lead to missed opportunities to discover specific regions with crucial contributions to the oncogenesis of multiple cancers.

In the present study, we collected European ancestry-derived GWAS summary statistics from large-scale meta-analysis results for 14 types of cancer, based on a total number of 375,468 cancer cases and 530,521 controls. By partitioning the genome into 1,703 blocks based on the LD pattern in the 1000 Genomes (1000G) European
ancestry populations,\textsuperscript{34} we systematically estimated pairwise local genetic correlations between cancers. After adjusting for multiple comparisons, we identified thirteen pairs of cancers with statistically significant local genetic correlations across eight distinct genomic regions. Among these, a 1.2 Mb region at 5p15.33, harboring the TERT and CLPTM1L genes, showed significant local genetic correlations across six pairs of cancers, including breast (overall and estrogen receptor [ER]-negative), colorectal, glioma, lung, melanoma, pancreatic, and prostate cancer. We then utilized an iterative analysis pipeline implementing a subset-based meta-analysis approach (Association Analysis for SubSETS [ASSET])\textsuperscript{35} and a conditional analysis tool (COndition and JOint analysis tool implemented in the Genome-wide Complex Trait Analysis software, COJO-GCTA)\textsuperscript{36} and identified ten independent cross-cancer signals within the 1.2 Mb region. For each independent signal, we conducted multi-cancer fine-mapping analysis using PAINTOR\textsuperscript{37} to prioritize the variants with the highest posterior probability of being causal. Our study provides novel evidence of shared genetic susceptibility across cancer types and contributes crucial information toward understanding the common genetic mechanisms of carcinogenesis.

Material and methods

Study sample and genotype quality control

We collected the meta-analysis results from a total of 14 cancer GWASs: breast (overall, ER-positive, and ER-negative),\textsuperscript{38} colorectal,\textsuperscript{39} endometrial,\textsuperscript{16} esophageal,\textsuperscript{40} glioma,\textsuperscript{41} head/neck,\textsuperscript{42} lung,\textsuperscript{43} melanoma,\textsuperscript{44} ovarian,\textsuperscript{45} pancreatic,\textsuperscript{46} prostate,\textsuperscript{47} and renal cancer.\textsuperscript{48} Sample size for each cancer is listed in Table 1. The GWAS summary statistics for each cancer were provided by the corresponding collaborative consortia. Details on study characteristics and subjects contributing to each cancer-specific GWAS summary dataset have been described in the original cancer-specific publications. All the GWAS results used in this study were based on European ancestry populations. Genomic positions were based on Genome Reference Consortium GRCh37 (hg19).

Individual cancer GWASs were primarily imputed to the 1000G reference panels.\textsuperscript{49} Breast, ovarian, pancreatic, and prostate cancer used the 1000G phase 3 v.5 reference panel; colorectal cancer used the Haplotype Reference Consortium (HRC); head/neck cancer used HRC; renal cancer used 1000G phase 1 v.3; melanoma results for melanoma GWASs were based on studies majorly imputed with 1000G phase 1 v.3;\textsuperscript{50} lung cancer used a mix between 1000G phase 1 and hase 3; glioma used a mix between 1000G phase 3, UK10K, and HRC; esophageal cancer used 1000G phase 1; and endometrial cancer used a mix between 1000G phase 3 v.5 and UK10K. For each dataset, we conducted comprehensive quality control to clean and harmonize the GWAS summary statistics across cancers. This included: (1) removing duplicate, structural, multi-allelic, and ambiguous variants; (2) confirming that strand and alleles at each variant were consistent across cancers; (3) creating a common unique marker ID; and (4) removing analytic artifacts (e.g., common variants with reported [log odds ratio] > 3). We also removed any variants with imputation quality score < 0.3 or minor allele frequency (MAF) < 1%, and [log odds ratio] > 3.

| Cancer types | No. of cases | No. of controls | No. of SNPs after QC\textsuperscript{a} | Reference |
|--------------|--------------|-----------------|----------------------------------------|-----------|
| Breast, overall | 122,977 | 105,974 | 9,934,907 | Michailidou et al., 2017\textsuperscript{38} |
| Breast, ER-negative | 21,468 | 100,564 | 9,942,394 | Michailidou et al., 2017\textsuperscript{38} |
| Breast, ER-positive | 69,501 | 95,042 | 10,267,258 | Michailidou et al., 2017\textsuperscript{38} |
| Colorectal | 55,168 | 65,160 | 7,910,462 | Huyghe et al., 2019\textsuperscript{39} |
| Endometrial | 12,906 | 108,979 | 11,595,492 | O’Mara et al., 2018\textsuperscript{16} |
| Esophageal | 4,112 | 13,663 | 9,038,176 | Gharakhkhani et al., 2016\textsuperscript{40} |
| Glioma | 12,488 | 18,169 | 6,931,587 | Melin et al., 2017\textsuperscript{31} |
| Head/neck | 6,034 | 6,585 | 7,471,918 | Lesueur et al., 2016\textsuperscript{42} |
| Lung | 29,266 | 56,450 | 7,673,197 | McKay et al., 2017\textsuperscript{43} |
| Melanoma | 12,814 | 23,203 | 7,748,523 | Law et al., 2015\textsuperscript{44} |
| Ovarian | 22,406 | 40,951 | 9,870,154 | Phelan et al., 2017\textsuperscript{45} |
| Pancreatic | 8,638 | 12,217 | 9,568,913 | Klein et al., 2018\textsuperscript{46} |
| Prostate | 79,166 | 61,106 | 10,002,813 | Schumacher et al., 2018\textsuperscript{47} |
| Renal | 10,784 | 20,407 | 8,362,393 | Scelo et al., 2018\textsuperscript{48} |

\textsuperscript{a}Filtered out variants with imputation quality score < 0.3, minor allele frequency (MAF) < 1%, and [log odds ratio] > 3.

Genetic correlations due to sample overlap

We estimated the number of controls overlapping between pairs of cancers, as these would induce a correlation in the GWAS summary statistics between cancers. We identified participating studies and any publicly available datasets (e.g., Wellcome Trust Case Control Consortium) to calculate the maximum number of controls overlapping between any two cancers. We also employed the tetrachoric correlation between binary-transformed GWAS
We observed six pairs of cancers that had correlations where controls were shared between groups (Table S1). Pairs and these all reflected previously known documented relationships where controls were shared between groups (Table S1). Pairs with correlations > 0.05 included breast and endometrial (0.08), ER-positive breast and endometrial (0.06), breast and ovarian (0.05), esophageal and melanoma (0.08), lung and head/neck (0.07), and lung and renal cancer (0.07).

Local genetic correlation estimation
To identify regions in the genome with local genetic correlations between pairs of cancers, we used pHESS53 (Heritability Estimation using Summary Statistics), which first estimates the local SNP-heritability for each cancer within each region based on summary statistics54 and then quantifies the covariance and correlation between pairs of cancers. Based on the LD pattern in 1000G European ancestry populations,54 pHESS partitions the genome into 1,703 approximatively independent LD blocks. We took sample overlap between pairs of cancers into account as described above. Pairwise local genetic correlations were considered statistically significant if the p value < 0.05/1,703 = 2.94 × 10^{-5}.

Searching for independent signals shared across cancers
Based on the local genetic correlation results, we identified a 1.2 Mb region at 5p15.33 (hg19 coordinates: 82,252–2,132,442 bp) harboring significant local heritability for multiple cancer pairs (see Results). We selected eight types of cancers (ER-negative breast, colorectal, glioma, lung, melanoma, ovarian, pancreatic, prostate) that had genome-wide significant associations in the region and showed evidence of pairwise genetic correlation (p < 0.05) with at least one other cancer having genome-wide significant associations in this region, which includes the TERT and CLPTM1L genes. We first performed a conditional analysis using COJO-GCTA56 on each individual cancer adjusting for the variant with the smallest p value, until no variant had a conditional p < 5 × 10^{-8}. We then performed pairwise colocalization analyses using COLOC54 to assess if any cancers shared causal variants, after controlling for the independent signals identified in analysis of individual cancers. To comprehensively enumerate the independent cross-cancer signals within this locus, we then used an agnostic subset-based meta-analysis (ASSET)55 to identify variants with the strongest cross-cancer associations in this region (Figure 1). ASSET allows for opposite direction of effects across traits when assessing the association between variants and multiple traits, as implemented in the “two-sided” option in ASSET. Overlap in controls between GWASs was addressed by using the tetrachoric correlation, as described above. To determine the number of independent signals within a region, we reran all individual cancer GWASs conditioning on the top variant identified by ASSET using COJO-GCTA. The conditional analysis may be subject to the mismatch of LD between the reference panel and the population that generated the GWAS results. Consequently, we created a LD reference panel for all cancer-specific conditional analyses using European ancestry breast cancer controls (n = 40,401),55 which was the largest population with genotype data available. After generating updated cancer-specific GWAS summary statistics conditioned on the most significant variant (top variant), we reran the two-sided ASSET meta-analysis to identify any additional significant cross-cancer signals. We then added the new top variant from the ASSET analysis to the list of lead SNPs and reran all cancer GWASs conditioning on all lead variants using COJO-GCTA. We iteratively ran cancer-specific analyses conditioning on the identified top variants using COJO-GCTA and ran two-sided ASSET on the resulting cancer-specific association results. We repeated this procedure until no variant reached genome-wide significance in the two-sided ASSET meta-analysis. The lead variants that resulted from the two-sided ASSET meta-analyses based on the conditional cancer-specific results were regarded as candidate variants that independently affect the risk of multiple cancers. Using this approach, we identified a total of ten independent signals within the 5p15.33 region.

Multi-trait fine-mapping
For each of the ten cross-cancer signals in the 5p15.33 region identified by our ASSET-COJO analysis, we created new cancer-specific GWAS summary statistics adjusting for the other nine top variants and estimated variant-specific posterior probabilities of causality using PAINTOR v.3.0.37 We varied the set of cancers included in the fine-mapping analyses of each of the ten independent cross-cancer signals as we hypothesized that not all cancers would share the same causal variant for each independent signal, but, rather, different combinations of cancers contributed to each of the ten independent signals. This was also supported by the ASSET analyses, where not all cancers contributed to the top signal for each of the ten conditional meta-analyses. In particular, ASSET provides the subset of traits that contribute to the smallest variant-specific meta-analysis p value. For each variant, two subsets are reported, with the first including traits with a positive association and the
second including traits with a negative association. For each of the ten independent signals, we included a specific cancer in the PAINTOR fine-mapping analysis if: (1) it was one of the cancers selected by ASSET as a contributing phenotype in the corresponding two-sided ASSET analysis of the lead variant, or (2) the lead variant showed genome-wide significant association for that cancer in the unadjusted cancer-specific GWAS. For each independent signal, only SNPs with data for all relevant cancers were included. We ran PAINTOR under the assumption that there was only one causal variant underlying that signal. We used the same LD reference panel for the fine-mapping analysis as we did for the conditional analysis. In our primary analyses, we performed the fine-mapping with no functional annotation implemented. Since regulation of TERT and CLPTM1L expression has been linked to open chromatin conformation in previous analyses,\textsuperscript{55,56} we conducted a secondary analysis incorporating tissue-specific open chromatin annotations as functional prior. We obtained open chromatin narrow peaks identified from normal tissue or primary cell lines of the relevant organs of each signal, based on the ENCODE project.\textsuperscript{57} By overlapping variants with open chromatin peaks, we generated a binary matrix for the region, which was then implemented as the functional prior in the fine-mapping analysis.

Results

Local genetic correlation revealed specific regions in the genome with shared heritability across cancers

We first partitioned the genome into 1,703 regions and estimated the pairwise local genetic correlation between fourteen types of cancers. After adjusting for multiple comparisons (p value < 0.05/1,703 = 2.94 × 10^{-5}), we identified thirteen pairs of cancers with statistically significant local genetic correlation across eight distinct genomic regions (Table 2). Among these, seven cancer pairs had positive genetic correlation (4q24: colorectal and prostate; 5p15.33: ER-negative breast and glioma, melanoma and pancreatic; 5q11.2: overall breast and colorectal; 8q24: colorectal and prostate; 17q12: endometrial and prostate; 19p13.11: ER-negative breast and ovarian), while six others showed negative genetic correlations (1q32: ER-negative breast and prostate; 5p15.33: glioma and prostate, colorectal and glioma, ER-negative breast and prostate, lung and pancreatic; 10q26.13: ER-positive breast and prostate). The local genetic correlation results mirrored previous observations, in that genome-wide significant variants for the identified regions have been previously reported for the individual cancers. For example, colorectal and prostate cancer showed significant local genetic correlation on chromosome 8 (126,410,917–128,659,111 bp), overlapping the 8q24.21 region, which harbors susceptibility variants for more than ten types of cancers. Similarly, a region on chromosome 19 (16,374,416–18,409,862 bp) showed significant local genetic correlation between ovarian and ER-negative breast cancer, both of which have genome-wide significant susceptibility variants in this region. One region on chromosome 5 (982,252–1,132,442 bp), harboring the TERT and CLPTM1L genes, showed significant local genetic correlation across six pairs of cancers, including ER-negative breast, colorectal, glioma, lung, melanoma, pancreatic, and prostate cancer (Figure 2). Interestingly, the direction of the genetic correlations varied between cancer pairs. For example, glioma showed significant but opposite local genetic correlations with ER-negative breast \((r_g = 0.0014, p = 2.40 \times 10^{-5})\) and colorectal cancer \((r_g = -0.0015, p = 1.24 \times 10^{-5})\). Similarly, pancreatic cancer had a positive local genetic correlation with melanoma \((r_g = 0.0034, p = 4.85 \times 10^{-6})\) but a negative genetic correlation with lung cancer \((r_g = -0.0025, p = 1.39 \times 10^{-5}).\)

### Table 2. Genomic regions with statistically significant local genetic correlations between cancers

| Cancer site 1 | Cancer site 2 | Region | Region start | Region end | No. of SNPs | Direction | p value* |
|---------------|---------------|--------|--------------|------------|-------------|-----------|----------|
| ER-negative breast | prostate | 1q32   | 203334734    | 204681068  | 2,364       | negative  | 3.45E-06 |
| Colorectal     | prostate     | 4q24   | 105305294    | 107501305  | 2,986       | positive  | 1.05E-05 |
| Glioma         | prostate     | 5p15.33| 982252       | 2132442    | 2,631       | negative  | 4.03E-19 |
| Colorectal     | glioma       | 5p15.33| 982252       | 2132442    | 2,465       | negative  | 1.24E-05 |
| ER-negative breast | prostate     | 5p15.33| 982252       | 2132442    | 3,111       | negative  | 1.90E-05 |
| ER-negative breast | glioma       | 5p15.33| 982252       | 2132442    | 2,631       | positive  | 2.40E-05 |
| Melanoma       | pancreatic   | 5p15.33| 982252       | 2132442    | 2,935       | negative  | 1.39E-07 |
| Lung           | pancreatic   | 5p15.33| 982252       | 2132442    | 4,275       | positive  | 4.85E-06 |
| Overall breast | colorectal   | 5q11.2 | 55417349     | 56621102   | 2,131       | positive  | 1.97E-05 |
| Colorectal     | prostate     | 8q24   | 126410917    | 128659111  | 4,275       | positive  | 1.97E-16 |
| ER-positive breast | prostate     | 10q26.13| 123231465    | 123900545  | 1,481       | negative  | 1.22E-06 |
| Endometrial    | prostate     | 17q12  | 34469036     | 36809344   | 2,748       | positive  | 5.01E-09 |
| ER-negative breast | ovarian      | 19p13.11| 16374416     | 18409862   | 4,103       | positive  | 1.11E-07 |

Local genetic correlation between cancers across the genome (N = 1,703 regions) was estimated using HESS. *Cutoff of the statistical significance was defined as p < 0.05/1,703 = 2.94E-05, after adjusting for multiple comparison.
Distinct patterns of regional GWAS association p values for the variants at 5p15.33

Based on the local genetic correlation results, the 5p15.33 region may harbor key genetic variants related to multiple cancer types. Indeed, multiple susceptibility variants in this region have been reported for at least ten cancer types, including ER-negative breast, colorectal, glioma, lung, melanoma, ovarian, pancreatic, and prostate cancer. To obtain a more complete understanding of the association patterns in this region, we created cancer-specific regional association plots for 5p15.33 (Figure 3A). We observed three different patterns of association. Pattern A, which includes breast (overall, ER-positive, and ER-negative), colorectal, glioma, ovarian, and prostate cancer, displayed one sharp genome-wide significant signal in a narrow region (~30 kb) overlapping the \( TERT \) gene (chr5: 1,253,282–1,295,178 bp). Pattern B, which includes lung, melanoma, and pancreatic cancer, has a broader genome-wide significant signal overlapping both the \( TERT \) (chr5: 1,253,282–1,295,178 bp) and \( CLPTM1L \) genes (chr5: 1,317,869–1,345,180 bp) (Figure 3B). Pattern C, which includes endometrial, esophageal, head/neck, and renal cancer, did not have a genome-wide significant signal in this region (Figure 3C). Interestingly, the distribution of variant-specific associations for some cancers was highly similar but in the opposite direction (Figure 3D), suggesting that GWAS associations discovered in this region may underly tissue-specific regulations across cancers. The association-based classification of cancers was highly consistent with our local genetic correlation results. All cancer types showing shared significant local genetic correlation in this region were in either pattern A or B, and thus we excluded the cancers belonging to pattern C for further analyses. For breast cancer, we limited our analysis to ER-negative breast cancer, as it had the strongest association at 5p15.33. Along with colorectal, glioma, lung, melanoma, ovarian, pancreatic, and prostate cancer, a total number of eight cancer types were used in the fine-mapping cross-cancer analyses.

Ten independent signals were identified based on multi-cancer meta-analysis results

Given the important biological function of the \( TERT \) and \( CLPTM1L \) genes, previous cancer fine-mapping efforts in this region, and the appearance of multiple association peaks for some of the cancers, it is plausible to assume that multiple variants in this region affect cancer risk independently. To test this assumption, we performed a conditional analysis using COJO-GCTA for each cancer to enumerate the independent signals at the 5p15.33 region. Six of the eight cancers of interest, including ER-negative breast, colorectal, glioma, lung, pancreatic, and prostate, were identified with two or more independent variants (Table S2). A total number of thirteen variants were identified, of which four were shared by two cancer types. By using conditional analysis results of each cancer, we then assessed the probability of two cancers sharing a single causal variant using a Bayesian-based colocalization approach.\(^{54}\) Glioma and melanoma were estimated to be likely sharing a causal variant (posterior probability \( \text{[PP]} = 0.519; \text{Table S3} \)), even after controlling for the effect of identified signals of individual cancers. These results

---

**Figure 2.** Pairwise local genetic correlation between selected cancer types at chromosome 5p15.33 (982,252–2,132,442 bp)

Cancer pairs with statistically significant (\( p \text{ value} < 0.05/1,703 = 2.94 \times 10^{-5} \)) local genetic correlation are annotated with an asterisk.
Figure 3. Categorizing 14 cancer types into three tiers based on their p value distribution at 5p15.33

Pattern A cancers (A) have one single peak by the \textit{TERT} gene; pattern B cancers (B) have a broader signal at the \textit{CLPTM1L} gene as well as a signal by the \textit{TERT} gene; pattern C cancers (C) have no genome-wide significant association in this region. Genome-wide significant levels at p value $= 5 \times 10^{-8}$ are marked with red dashed line in (A)–(C). Distribution of $Z$ scores at the 5p15.33 region from the GWAS results of ER-negative breast, glioma, and prostate cancer (D). Only variants with p $< 0.05$ for both cancers are included. While the associations for ER-negative breast cancer and glioma overlap, the SNP associations with ER-negative breast cancer and prostate, as well as glioma and prostate, are in opposite directions.
suggest that multiple independent cross-cancer signals may exist in this region.

However, current state-of-the-art statistical fine-mapping tools often struggle to make inference of causality under the assumption of multiple causal variants. Further, it is likely that not all cancers share all causal variants. To get an estimate of the number of independent association signals across cancers in this region, we conducted iterative meta-analyses using individual cancer-specific association results from conditional analysis as generated by COJO-GCTA (see Material and methods). We adopted the two-sided analysis scheme in ASSET to allow for the detection of effects in opposite directions.

The strongest associated variant in the two-sided ASSET meta-analysis was rs10069690 (chr5: 1,279,790, $p = 4.05 \times 10^{-126}$; Figure 4), which was positively associated with ER-negative breast cancer and glioma while negatively associated with pancreatic and prostate cancer. We adjusted the cancer-specific GWAS results for rs10069690 using COJO-GCTA, and then reran the two-sided ASSET meta-analysis with the rs10069690-adjusted cancer-specific results. We observed the strongest association for rs465498 (chr5: 1,325,803, $p = 1.75 \times 10^{-39}$), which was positively associated with melanoma and pancreatic cancer and negatively associated with lung cancer. We added rs465498 to the set of variants to be conditioned on in the cancer-specific GWASs and iterated this process until no variant reached genome-wide significance ($p < 5 \times 10^{-8}$) in two-sided ASSET meta-analysis. In the end, we obtained ten conditionally independent significant SNPs (Table 3; Figure S1). The pairwise $r^2$ between the ten SNPs ranged between 0.001 and 0.294 as based on 1000G European ancestry data, which indicated that the pairwise correlations between the identified signals were weak (Figure 5).

For each of the ten independent signals, the number of cancer types contributing to the association as identified by ASSET ranged from two to eight. Although SNPs rs10069690 and rs7705526 were both genome-wide significant variants for ovarian cancer ($p = 1.74 \times 10^{-8}$ and $1.34 \times 10^{-9}$, respectively), ASSET did not include ovarian cancer as a contributing cancer to the meta-analysis results for either of the SNPs. To ensure that we included all relevant cancers in the fine-mapping analysis of each independent signal, we extracted the original cancer GWAS results for the ten independent SNPs and manually added any cancers to the list of contributing cancers if that cancer showed a genome-wide significant association with a specific SNP but was not included on the list of traits optimizing the ASSET meta-analysis. For each of the ten SNPs, we then applied COJO-GCTA on each included cancer GWAS dataset to obtain cancer-specific results conditioned on the other nine lead SNPs and used these adjusted summary statistics in the fine-mapping analyses.

Cross-cancer fine-mapping proposes candidate causal variants shared by cancers

To identify candidate causal SNPs within the ten independent signals identified in the conditional analyses, we
| No. of iteration | SNP with top ASSET p value | ASSET p value<sup>a</sup> | Significant cancer subset, identified by ASSET<sup>b</sup> | GWAS p values<sup>c</sup> |
|------------------|---------------------------|------------------------|-------------------------------------------------|------------------------|
|                  |                           |                       | ER-neg BrCa, glioma, prostate set 2: colorectal, prostate | ER-neg BrCa | Colorectal | Glioma | Lung | Melanoma | Ovarian | Pancreatic | Prostate |
| Initiation       | rs10069690 (5:1279790:C:T) | 4.05E−126             | set 1: ER-neg BrCa, glioma; set 2: pancreatic, prostate | 1.34E−35* | 1.13E−01 | 2.32E−66* | 9.39E−01 | 9.08E−01 | 1.74E−08* | 3.29E−03 | 1.44E−45* |
| 1                | rs465498 (5:1325803:A:G)   | 1.75E−59              | set 1: melanoma, pancreatic; set 2: lung               | 1.61E−02 | 3.89E−02 | 6.85E−05 | 2.68E−32* | 2.08E−17* | 1.49E−01 | 7.45E−17* | 9.15E−05 |
| 2                | rs2853677 (5:1287194:G:A)  | 3.24E−39              | set 1: ER-neg BrCa, colorectal, lung; set 2: glioma, melanoma, ovarian, pancreatic, prostate | 4.94E−05 | 3.65E−10* | 1.08E−28* | 2.66E−18* | 1.12E−02 | 1.49E−06 | 2.87E−08* | 1.53E−02 |
| 3                | rs11414507 (5:1291331:A:AC) | 1.29E−19              | set 1: ER-neg BrCa; set 2: prostate                  | 1.34E−16* | NA | NA | NA | NA | 1.63E−04 | 8.18E−01 | 1.61E−45* |
| 4                | rs35033501 (5:1253918:C:T) | 9.18E−15              | set 1: lung, melanoma, prostate; set 2: ER-neg BrCa, pancreatic | 9.90E−05 | 8.55E−03 | 1.24E−02 | 3.84E−01 | 3.66E−02 | 5.55E−01 | 4.48E−05 | 2.37E−15* |
| 5                | rs7705526 (5:1285974:C:A)  | 1.42E−11              | set 1: glioma, lung, melanoma; set 2: ER-neg BrCa, colorectal, pancreatic, prostate | 1.37E−04 | 3.17E−04 | 5.01E−61* | 1.01E−18* | 3.24E−03 | 1.34E−09* | 2.15E−03 | 2.78E−14* |
| 6                | rs192723047 (5:1273183:A:G) | 1.63E−11              | set 1: prostate; set 2: ER-neg BrCa                  | 4.37E−17* | NA | NA | NA | NA | 1.21E−02 | 8.96E−02 | 5.40E−24* |
| 7                | rs35226131 (5:1295373:C:T) | 2.32E−09              | set 1: pancreatic; set 2: colorectal, glioma, prostate | 8.83E−02 | 5.17E−07 | 8.66E−01 | NA | NA | 3.63E−01 | 4.30E−08* | 3.20E−06 |
| 8                | rs35334674 (5:1292299:A:G) | 1.24E−08              | set 1: ER-neg BrCa, pancreatic; set 2: colorectal, glioma, lung, prostate | 3.36E−02 | 4.40E−07 | 4.07E−02 | 2.15E−02 | NA | 5.93E−01 | 1.43E−02 | 3.23E−02 |
| 9                | rs3888705 (5:1298645:G:A)  | 2.65E−08              | set 1: ER-neg BrCa, colorectal, ovarian; set 2: pancreatic, prostate | 1.78E−02 | 1.04E−03 | 9.32E−07 | 3.62E−03 | NA | 2.75E−03 | 7.13E−01 | 1.94E−03 |
| 10               | rs148487301 (5:1318797:T:C) | 8.70E−06              | not reached genome-wide significance, iteration stopped | |

<sup>a</sup>Genome-wide significance with p value < 5 × 10<sup>−8</sup>. BrCa, breast cancer; NA, the SNP was not included in the GWAS results of corresponding cancer.

<sup>b</sup>p values from the ASSET meta-analysis allowing opposite direction of the effect (two-sided analysis).

<sup>c</sup>Cancer subsets included in the two-sided ASSET meta-analysis with best p value. Set 1/2 represents the selected cancer types with positive/negative association with the SNP.

<sup>d</sup>p values from the original GWAS results of eight cancers.
conducted a multi-cancer fine-mapping analysis using PAINTOR for each signal. We first performed fine-mapping analyses with no functional annotation data implemented. For the ten candidate signals, the size of credible sets comprising a cumulative 95% PP of causality ranged from one to fifteen variants (Table 4; Figure S2). All SNPs identified as lead SNPs in the conditional analysis were included in the 95% PP credible set of the corresponding fine-mapping analysis, with six of them having the highest PP in its set (rs35033501: PP = 0.875; rs35334674: PP = 0.987; rs192723047, rs10069690, rs7705526, rs2853677: PP > 0.999). The fine-mapping analysis based on the signal identified by SNP rs35226131 included data on colorectal, glioma, pancreatic, and prostate cancer. Although rs35226131 was identified as the SNP with the highest PP of being causal, the PP was relatively modest (PP = 0.273) and comparable to nearby SNPs (rs35161420, PP = 0.239; rs61748181, PP = 0.228). Fine-mapping analysis of the signals indexed by rs11414507 (ER-negative breast, prostate), the size of the 95% PP credible set shrank from four to two, which included the index SNP rs11414507 (PP = 0.42) as well as rs7712562 (PP = 0.58). Both rs11414507 and rs7712562 were located in open chromatin peaks in breast epithelial tissue. Similarly, after we implemented the functional annotation data, only two SNPs were included in the 95% PP credible set of the signal indexed by rs465498 (lung, melanoma, pancreatic), compared to eight SNPs in the analysis without functional information. The index SNP rs465498 had a comparable PP (0.437) as rs421629 (PP = 0.563), and both SNPs were located within open chromatin peaks in lung tissue.

Discussion

In this study, we leveraged GWAS summary statistics from 14 cancer types to estimate local genetic correlations and conduct follow-up fine-mapping of shared cancer regions in the genome. By partitioning the genome into independent blocks as defined by LD, we comprehensively estimated pairwise local genetic correlations between the included cancers. We identified 13 cancer pairs with significant local genetic correlation across eight distinct
Table 4. Statistical fine-mapping prioritized the potential causal SNP within 10 independent cross-cancer signals in 5p15.33 region, using PAINTOR v.3.0

| Index SNP     | Fine-mapped cancer types                           | Fine-mapping without functional prior | Fine-mapping with functional priora | Index SNP     |
|---------------|-----------------------------------------------------|--------------------------------------|------------------------------------|---------------|
|               |                                                     | 95% PP credible setb                 | PP, index SNP                      |               |
|               |                                                     | SNP with highest PP                 | Highest PP                        |               |
|               |                                                     |                                      | 95% PP credible setb              |               |
|               |                                                     | SNP with highest PP                 | Highest PP                        |               |
| rs35033501    | ER-neg BrCa, lung, melanoma, pancreatic, prostate  | rs35033501, rs71595003               | 0.875 rs35033501                  | rs81595003    |
| (5:1253918:C:T)|                                                     |                                      | (<0.001 rs71595003)               | (5:1292118:G:A)| 0.999       |
| rs192723047   | ER-neg BrCa, prostate                               | rs192723047                         | 1.000 rs192723047                 | rs10069690    |
| (5:1273183:A:G)|                                                     |                                      | (<1.000 rs192723047)              | (5:1279790:C:T)| 1.000       |
| rs10069690    | ER-neg BrCa, glioma, ovarian, pancreatic, prostate | rs10069690                          | 1.000 rs10069690                  | rs10069690    |
| (5:1279790:C:T)|                                                     |                                      | (<1.000 rs10069690)               | (5:1279790:C:T)| 1.000       |
| rs7705526     | ER-neg BrCa, colorectal, glioma, lung, melanoma,   | rs7705526                           | 1.000 rs7705526                   | rs7705526     |
| (5:1285974:C:A)| ovarian, pancreatic, prostate                      |                                      | (<1.000 rs7705526)                | (5:1285974:C:A)| 1.000       |
| rs2853677     | ER-neg BrCa, colorectal, glioma, lung, melanoma,   | rs2853677                           | 1.000 rs2853677                   | rs2853677     |
| (5:1287194:A:G)| ovarian, pancreatic, prostate                      |                                      | (<1.000 rs2853677)                | (5:1287194:A:G)| 1.000       |
| rs11144507    | ER-neg BrCa, prostate                               | rs7712562, rs4768246, rs7449190      | 0.265 rs7712562                   | rs7712562     |
| (5:1291331:A:AC)|                                                     |                                      | (<0.367 rs7712562)                | (5:1296072:A:G)| 0.581       |
| rs35334674    | ER-neg BrCa, colorectal, glioma, lung, pancreatic, | rs35334674                          | 0.987 rs35334674                  | rs35334674    |
| (5:1292299:G:A)| prostate                                         |                                      | (<0.988 rs35334674)               | (5:1292299:G:A)| 0.988       |
| rs35226131    | colorectal, glioma, pancreatic, prostate           | rs35226131, rs61748181, rs33958877,  | 0.273 rs35226131                  | rs35226131    |
| (5:1295373:C:T)|                                                   | rs114616103                          | (<0.273 rs35226131)               | (5:1295373:C:T)| 0.273       |
| rs3888705     | ER-neg BrCa, colorectal, ovarian, pancreatic,      | rs34156553, rs4075202, rs3888705,     | 0.092 rs34156553                  | rs34156553    |
| (5:1298645:G:A)| prostate                                         | rs77776598, rs4975539, rs6875445,     | (<0.103 rs34156553)               | (5:1243245:C:T)| 0.103       |
|               |                                                     | rs4583925, rs78244046, rs79223805,    |                                      |               |
|               |                                                     | rs4507531, rs78368589, rs44487533,    |                                      |               |
|               |                                                     | rs6554678, rs4498293, rs4332396       |                                      |               |
| rs465498      | lung, melanoma, pancreatic                          | rs380286, rs421629, rs465498, rs452932, | 0.146 rs380286                    | rs421629      |
| (5:1325803:A:G)|                                                   | rs4559961, rs455433, rs13178886,     | (<0.462 rs421629)                 | rs421629      |
|               |                                                     | rs460073                             |                                      | rs465498      |
|               |                                                     |                                      | (<0.437 rs421629)                 | (5:1320136:G:A)| 0.563       |

aUsed open chromatin narrow peaks identified from the normal tissue or primary cells of the disease-related organs as the functional prior. Open chromatin narrow peaks were obtained from the ENCODE project.

bSNPs within the credible set were ranked by the posterior probability (PP).
patterns. Our analysis complements these, as we aggregated genome-wide significant loci of multiple cancers by LD a subset-based meta-analysis approach or categorizing devoted to identifying pleiotropic variants, by using either associated with multiple cancers. Other significant TERT containing the in previous analyses. Previous efforts have been implicated in cancers (ER-negative breast, colorectal, glioma, lung, melanoma, pancreatic, and prostate cancer. By utilizing an iterative analysis, we identified ten independent cross-cancer SNP signals within this locus. We then conducted fine-mapping analyses for each independent signal and generated 95% posterior probability credible sets both without and with a priori functional information.

Our pairwise local genetic correlation results were highly consistent with the conclusions of previous GWASs and cross-cancer analyses. The pleiotropic effect of variants in the 8q24 region between multiple types of cancer, including colorectal and prostate cancer, has been previously demonstrated and replicated in studies across populations of different ancestries. The 5p15.33 region, containing the TERT and CLPTM1L genes, has also been associated with multiple cancers. Other significant genomic regions identified in our study, including 1q32.1 (ER-negative breast and prostate), 4q24 (colorectal and prostate), 5q11.2 (overall breast and colorectal), 10q26.13 (ER-positive breast and prostate), 17q12 (endometrial and prostate), and 19p13.11 (ER-negative breast and ovarian), have also been identified as pleiotropic loci in previous analyses. Previous efforts have been devoted to identifying pleiotropic variants, by using either a subset-based meta-analysis approach or categorizing genome-wide significant loci of multiple cancers by LD patterns. Our analysis complements these, as we aggregated the per-SNP effect within the loci, estimated the local heritability of each cancer, and quantified the local genetic correlation between the cancer pairs. These “shared heritability hotspots” identified in our analysis may contain genes with strong effect on multiple cancers or harbor multiple risk variants and biological mechanisms that can independently affect the risk of different cancers. Our results can thus be utilized to prioritize candidate regions for future discoveries of causal variants and functional follow-up.

As the 5p15.33 region harboring the TERT and CLPTM1L genes was the only region that displayed more than one statistically significant pairwise genetic correlation, we focused our continued efforts on this region. The TERT gene encodes the catalytic subunit of telomerase reverse transcriptase, which is a crucial enzyme for maintaining telomere length. Mendelian randomization studies have shown that genetically determined telomere length is associated with the risk of multiple cancer types, including glioma, ovarian, lung, and melanoma, but is not associated with the risk of other cancers included here, such as breast and prostate. In our study, we observed local negative genetic correlations and opposite direction of SNP effects between specific cancer types, which indicate that genetic variation in this region is likely to affect cancer risk through multiple distinct biological pathways, of which telomere length is only one implicated mechanism. Meanwhile, the CLPTM1L gene encodes the cleft lip and palate-associated transmembrane-1 like protein, which has been reported to play a role in cell apoptosis and cytokinesis and is overexpressed in lung and pancreatic cancer. Given its important biological function and significant association with a broad set of cancers, we assumed that multiple variants in this region may independently influence the risk of various types of cancers. By iteratively conducting conditional meta-analyses, we identified ten independent signals (seven in the TERT gene, one in the CLPTM1L gene, and two between TERT and CLPTM1L). Our study results are comparable to a previous study published by Wang et al., which conducted a subset-based meta-analysis across six types of cancers (bladder, glioma, lung, pancreatic, prostate, and testicular). Several signals identified in our study have either been proposed (rs10069690, rs2853677) or are correlated with the independent signals reported in that study (rs7705526 versus rs7726159, $r^2 = 0.87$; rs465498 versus rs451360, $r^2 = 0.34$). We only included cancers with genome-wide significant signals in this region into the subset-based meta-analysis and conditional analysis. Compared to the study presented by Wang et al., our study further included several common cancers (ER-negative breast, colorectal, melanoma, and ovarian), while we did not have data on bladder and testicular cancer. With an increased number of cancers and larger sample sizes, we were able to refine the cross-cancer signals in this important region. In addition, independent signal rs465498 identified in our study was in strong correlation with two previously identified susceptible loci at the CLPTM1L gene, including pancreatic cancer SNP rs31490 ($r^2 = 0.96$) and lung, melanoma, and...
Our findings imply that the association between the CLPTM1L gene and various types of cancer can be potentially attributed to one distinct signal.

When estimating the local genetic correlation across cancers, we considered subtypes for breast (ER-negative and ER-positive) and lung cancer (adenocarcinoma, small cell, and squamous cell). Despite the relatively smaller GWAS sample size (21,468 for ER-negative breast cancer compared to 122,977 for overall breast cancer), ER-negative breast cancer showed stronger associations and higher genetic correlation with other cancers in the 5p15.33 region, as compared to ER-positive and overall breast cancer. In contrast, the three subtypes of lung cancer had either no genome-wide significant hits at the 5p15.33 region (small cell) or had weaker local genetic correlation estimates (adenocarcinoma and squamous cell, data not shown) than overall lung cancer. We thus included ER-negative breast cancer and overall lung cancer in the subsequent analyses.

Multiple lead SNPs with high posterior probability have been reported to affect telomere length. SNP rs7705526 is significantly associated with telomere length in multiple populations. \(^{72-75}\) SNP rs2853677 has been associated with relative telomere length in a breast cancer case-only cohort in Han Chinese, \(^{76}\) as well as leukocyte telomere length in a European ancestry population. \(^{77}\) SNP rs35226131 is perfectly correlated with a nonsynonymous variant (rs61748181) in TERT, which results in a protein-level change from alanine to threonine and negatively influences telomere length. \(^{15,71}\) SNP rs10069690 has been found to significantly interact with recent use of non-steroidal anti-inflammatory drugs (NSAIDs) to alter telomere length in a colorectal cancer case-control study. \(^{77}\) SNP rs465498, located in the CLPTM1L gene, has been reported to be significantly associated with telomere length among Han Chinese. \(^{78}\) We could not find previous data on the role of other five lead SNPs identified by our study, and it is thus possible that other unknown mechanisms are involved.

Since previously identified cancer risk SNPs at 5p15.33 have been linked to open chromatin conformation, \(^{35,56}\) we further included regions of open chromatin for related tissues from the ENCODE project as functional prior in our fine-mapping analysis. \(^{57}\) The results for five signals (lead SNPs rs192723047, rs10069690, rs7705526, rs2853677, and rs35334674) remained unchanged, with each having a credible set containing one single SNP with a posterior probability of 1.00. After incorporating open chromatin peaks as a prior, the 95% posterior probability credible sets became smaller for three signals (lead SNPs rs35033501, rs11414507, and rs465498), as SNPs located in open chromatin peaks obtained a higher posterior probability of being causal. For the other two signals (lead SNPs rs35226131 and rs3888705), the size of each 95% credible set was relatively large in analyses with and without functional annotations. No SNPs in these regions had a predominantly high posterior probability, nor did any of them overlap with the open chromatin peaks of any related tissue. The fine-mapping results for these two signals should thus be interpreted with extra caution.

Our study has several strengths and limitations. We used cancer GWAS summary statistics published by each collaborating consortium, which maximized our sample sizes and provided large statistical power. This is also the first study to comprehensively quantify the local genetic correlation across multiple common cancers. We innovatively adopted the joint analysis pipeline of two-sided ASSET meta-analysis and COJO-GCTA. This approach enabled us to both validate the proposed pleiotropic ASSET signals in the 5p15.33 region, as well as leukocyte telomere length in a European ancestry population. \(^{75}\) SNP rs35226131 is perfectly correlated with a nonsynonymous variant (rs61748181) in TERT, which results in a protein-level change from alanine to threonine and negatively influences telomere length. \(^{15,71}\) SNP rs10069690 has been found to significantly interact with recent use of non-steroidal anti-inflammatory drugs (NSAIDs) to alter telomere length in a colorectal cancer case-control study. \(^{77}\) SNP rs465498, located in the CLPTM1L gene, has been reported to be significantly associated with telomere length among Han Chinese. \(^{78}\) We could not find previous data on the role of other five lead SNPs identified by our study, and it is thus possible that other unknown mechanisms are involved.

Since previously identified cancer risk SNPs at 5p15.33 have been linked to open chromatin conformation, \(^{35,56}\) we further included regions of open chromatin for related tissues from the ENCODE project as functional prior in our fine-mapping analysis. \(^{57}\) The results for five signals (lead SNPs rs192723047, rs10069690, rs7705526, rs2853677, and rs35334674) remained unchanged, with each having a credible set containing one single SNP with a posterior probability of 1.00. After incorporating open chromatin peaks as a prior, the 95% posterior probability credible sets became smaller for three signals (lead SNPs rs35033501, rs11414507, and rs465498), as SNPs located in open chromatin peaks obtained a higher posterior probability of being causal. For the other two signals (lead SNPs rs35226131 and rs3888705), the size of each 95% credible set was relatively large in analyses with and without functional annotations. No SNPs in these regions had a predominantly high posterior probability, nor did any of them overlap with the open chromatin peaks of any related tissue. The fine-mapping results for these two signals should thus be interpreted with extra caution.

Our study has several strengths and limitations. We used cancer GWAS summary statistics published by each collaborating consortium, which maximized our sample sizes and provided large statistical power. This is also the first study to comprehensively quantify the local genetic correlation across multiple common cancers. We innovatively adopted the joint analysis pipeline of two-sided ASSET meta-analysis and COJO-GCTA. This approach enabled us to both validate the proposed pleiotropic ASSET signals in the 5p15.33 region, as well as leukocyte telomere length in a European ancestry population. \(^{75}\) SNP rs35226131 is perfectly correlated with a nonsynonymous variant (rs61748181) in TERT, which results in a protein-level change from alanine to threonine and negatively influences telomere length. \(^{15,71}\) SNP rs10069690 has been found to significantly interact with recent use of non-steroidal anti-inflammatory drugs (NSAIDs) to alter telomere length in a colorectal cancer case-control study. \(^{77}\) SNP rs465498, located in the CLPTM1L gene, has been reported to be significantly associated with telomere length among Han Chinese. \(^{78}\) We could not find previous data on the role of other five lead SNPs identified by our study, and it is thus possible that other unknown mechanisms are involved.

Since previously identified cancer risk SNPs at 5p15.33 have been linked to open chromatin conformation, \(^{35,56}\) we further included regions of open chromatin for related tissues from the ENCODE project as functional prior in our fine-mapping analysis. \(^{57}\) The results for five signals (lead SNPs rs192723047, rs10069690, rs7705526, rs2853677, and rs35334674) remained unchanged, with each having a credible set containing one single SNP with a posterior probability of 1.00. After incorporating open chromatin peaks as a prior, the 95% posterior probability credible sets became smaller for three signals (lead SNPs rs35033501, rs11414507, and rs465498), as SNPs located in open chromatin peaks obtained a higher posterior probability of being causal. For the other two signals (lead SNPs rs35226131 and rs3888705), the size of each 95% credible set was relatively large in analyses with and without functional annotations. No SNPs in these regions had a predominantly high posterior probability, nor did any of them overlap with the open chromatin peaks of any related tissue. The fine-mapping results for these two signals should thus be interpreted with extra caution.

Our study has several strengths and limitations. We used cancer GWAS summary statistics published by each collaborating consortium, which maximized our sample sizes and provided large statistical power. This is also the first study to comprehensively quantify the local genetic correlation across multiple common cancers. We innovatively adopted the joint analysis pipeline of two-sided ASSET meta-analysis and COJO-GCTA. This approach enabled us to both validate the proposed pleiotropic ASSET signals in the 5p15.33 region, as well as leukocyte telomere length in a European ancestry population. \(^{75}\) SNP rs35226131 is perfectly correlated with a nonsynonymous variant (rs61748181) in TERT, which results in a protein-level change from alanine to threonine and negatively influences telomere length. \(^{15,71}\) SNP rs10069690 has been found to significantly interact with recent use of non-steroidal anti-inflammatory drugs (NSAIDs) to alter telomere length in a colorectal cancer case-control study. \(^{77}\) SNP rs465498, located in the CLPTM1L gene, has been reported to be significantly associated with telomere length among Han Chinese. \(^{78}\) We could not find previous data on the role of other five lead SNPs identified by our study, and it is thus possible that other unknown mechanisms are involved.
novel evidence of the shared inherited basis of human cancers and expand our understanding of the role played by the TERT-CLPTM1L region in cancer development.

Declaration of interests

B.M.W. has received research grants from Celgene and Eli Lilly and has consulting relationship with BioLineRx, Celgene, and Grail. R.A.E. has received speaker honoraria from the GU-ASCO meeting (January 2016), RMH FR meeting (November 2017, supported by Janssen), University of Chicago invited talk (May 2018), and ESMO (September 2019, supported by Bayer & Ipsen) and served as member of external expert committee at the Prostate Dx Advisory Panel (June 2020). All other authors declare no competing interests.

Data and code availability

All code is available from the corresponding author on request. Data availability varies by cancer site. Publicly available GWAS summary statistics are available at http://bcac.ccge.medschl.cam.ac.uk/bcadata/%20oncoarray/oncoarray-and-combined-summary-result/ (breast), http://practical.icr.ac.uk/blog/?page_id=8164 (prostate), http://ocac.ccge.medschl.cam.ac.uk/data-projects/results-lookup-by-region (ovarian), and https://www.ebi.ac.uk/gwas/downloads/summary-statistics (endometrial). For additional data, please contact the corresponding author.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2021.100041.

Received: April 7, 2021
Accepted: June 4, 2021

Web resources

Breast Cancer Association Consortium, http://bcac.ccge.medschl.cam.ac.uk/bcadata/%20oncoarray/oncoarray-and-combined-summary-result/
The Institute of Cancer Research PRACTICAL, http://practical.icr.ac.uk/blog/page_id=8088
Ovarian Cancer Association Consortium, http://ocac.ccge.medschl.cam.ac.uk/data-projects/results-lookup-by-region
GWAS Catalog, https://www.ebi.ac.uk/gwas/downloads/summary-statistics

References

1. International Agency for Research on Cancer; and W.H.O (2020). GLOBOCAN 2020. https://gco.iarc.fr/.
2. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2021). Cancer Statistics, 2021. CA Cancer J. Clin. 71, 7–33.
3. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med. 343, 78–85.
4. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H., et al.; Nordic Twin Study of Cancer (NorTwinCan) Collaboration (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA 315, 68–76.
5. Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T.B., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2020). Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. Nat. Commun. 11, 4423.
6. Goode, E.L., Chenevix-Trench, G., Song, H., Ramus, S.J., Notariodou, M., Lawrenson, K., Widschwendter, M., Vierkant, R.A., Larson, M.C., Kjaer, S.K., et al.; Wellcome Trust Case-Control Consortium; Australian Cancer Study (Ovarian Cancer); Australian Ovarian Cancer Study Group; Ovarian Cancer Association Consortium (OCAC); and Ovarian Cancer Association Consortium (OCAC) (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat. Genet. 42, 874–879.
7. Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirner, C., McDonald, G.J., et al. (2007). Multiple regions within 8q24 independently affect risk for prostate cancer. Nat. Genet. 39, 638–644.
8. Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., et al. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat. Genet. 39, 989–994.
9. Al Olama, A.A., Kote-Jarai, Z., Giles, G.G., Guy, M., Morrison, J., Severi, G., Leongamornlert, D.A., Tymrakiewicz, M., Jhavar, S., Saunders, E., et al.; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons’ Section of Oncology; and UK Prostate testing for cancer and Treatment study (ProtecT Study) Collaborators (2008). Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat. Genet. 41, 1058–1060.
10. Couch, F.J., Wang, X., McWilliams, R.R., Bamlet, W.R., de Andrade, M., and Petersen, G.M. (2009). Association of breast cancer susceptibility variants with risk of pancreatic cancer. Cancer Epidemiol. Biomarkers Prev. 18, 3044–3048.
11. Fletcher, O., Johnson, N., Gibson, L., Coupland, B., Fraser, A., Leonard, A., dos Santos Silva, I., Ashworth, A., Houlston, R., and Peto, J. (2008). Association of genetic variants at 8q24 with breast cancer risk. Cancer Epidemiol. Biomarkers Prev. 17, 702–705.
12. Ghoussaini, M., Song, H., Koessler, T., Al Olama, A.A., Kote-Jarai, Z., Driver, K.E., Pooley, K.A., Ramus, S.J., Kjaer, S.K., Hodgall, E., et al.; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons’ Section of Oncology; and UK ProtecT Study Collaborators (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. J. Natl. Cancer Inst. 100, 962–966.
13. Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D.H., Helgason, A., Rafnar, T., Berghorssen, J.T., Agnarsson, B.A., Baker, A., et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat. Genet. 39, 631–637.
14. Yeager, M., Chatterjee, N., Ciampa, J., Jacobs, K.B., Gonzalez-Bosquet, J., Hayes, R.B., Kraft, P., Wacholder, S., Orr, N.
38. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al.; NBCS Collaborators; ABCTB Investigators; and ConFab/AOCS Investigators (2017). Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92–94.

39. Huyghe, J.R., Bien, S.A., Harrison, T.A., Kang, H.M., Chen, S., Schmit, S.L., Conti, D.V., Qu, C., Jeon, J., Edlund, C.K., et al. (2019). Discovery of common and rare genetic risk variants for colorectal cancer. Nat. Genet. 51, 76–87.

40. Gharahkhani, P., Fitzgerald, R.C., Vaughan, T.L., Palles, C., Gockel, I., Tomlinson, I., Buas, M.F., May, A., Gerges, C., Anders, M., et al.; Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON); Esophageal Adenocarcinoma Genomics Consortium (EAGLE); and Wellcome Trust Case Control Consortium 2 (WTCCC2) (2016). Genome-wide association studies in oesophageal adenocarcinoma and Barrett’s oesophagus: a large-scale meta-analysis. Lancet Oncol. 17, 1363–1373.

41. Melin, B.S., Barnholtz-Sloan, J.S., Wrensch, M.R., Johansen, C., I’Iyasova, D., Kinnerson, B., Ostrom, Q.T., Labrecbe, K., Chen, Y., Armstrong, G., et al.; GliomaScan Consortium (2017). Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. Nat. Genet. 49, 789–794.

42. Lesueur, C., Diergaarde, B., Olshan, A.F., Wünsch-Filho, V., Ness, A.R., Liu, G., Lasko, M., Elut-Neto, J., Franceschi, S., Lagiou, P., et al. (2016). Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nat. Genet. 48, 1544–1550.

43. McKay, J.D., Hung, R.J., Han, Y., Zong, X., Carreras-Torres, R., Christiani, D.C., Caporaso, N.E., Johansson, M., Xiao, X., Li, Y., et al.; SpirroMeta Consortium (2017). Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat. Genet. 49, 1126–1132.

44. Law, M.H., Bishop, D.T., Lee, J.E., Brossard, M., Martin, N.G., Moses, E.K., Song, F., Barrett, J.H., Kumar, R., Easton, D.F., et al.; GenoMEL Consortium; Essen-Heidelberg Investigators; SDH Study Group; Q-MEGA and QTWIN Investigators; AMFS Investigators; and ATHENS Melanoma Study Group (2015). Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat. Genet. 47, 987–995.

45. Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., Lawrence, K., Winham, S.J., Dennis, J., Pirie, A., Riggin, M.J., Chornokur, G., et al.; AOCs study group; EMBRACE Study; GEMO Study Collaborators; HEBON Study; KConFab Investigators; and OPAL study group (2017). Identification of 12 new susceptibility loci for cutaneous malignant melanoma. Nat. Genet. 49, 680–691.

46. Klein, A.P., Wolpin, B.M., Risch, H.A., Stolzenberg-Solomon, R.Z., Mocci, E., Zhang, M., Czanizan, E., Childs, E.J., Hong, R.J., Han, Y., Zong, X., Carreras-Torres, R., McKay, J.D., Hung, R.J., Han, Y., Zong, X., Carreras-Torres, R., McKay, J.D., et al. (2018). Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat. Commun. 9, 556.

47. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Ciocca-Borrella, C., et al.; Profile Study; Australian Prostate Cancer BioResource (APCB); IMPACT Study; Canary PASS Investigators; Breast and Prostate Cancer Cohort Consortium (BPC3); PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium; Cancer of the Prostate in Sweden (CAPS); Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci (PEGASUS); and Genetic Associations and Mechanisms in Oncology (GAME-ON) (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat. Genet. 50, 928–936.

48. Scelo, G., Purdue, M.P., Brown, K.M., Johansson, M., Wang, Z., Eckel-Passow, J.E., Ye, Y., Hofmann, J.N., Choi, J., Foll, M., et al. (2017). Genome-wide association study identifies multiple risk loci for renal cell carcinoma. Nat. Commun. 8, 15724.

49. Anton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature 526, 68–74.

50. Province, M.A., and Borecki, I.B. (2013). A correlated meta-analysis strategy for data mining “OMIC” scans. Pac. Symp. Biocomput. 2013, 236–246.

51. Southam, L., Gilly, A., Siuvelges, D., Farma, A.E., Schwartztenbruber, J., Tachmazidou, I., Matchan, A., Rayner, N.W., Tsafanakis, E., Karaleftheri, M., et al. (2017). Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. Nat. Commun. 8, 15606.

52. Shi, H., Mancuso, N., Spendlove, S., and Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am. J. Hum. Genet. 101, 737–751.

53. Shi, H., Khichaev, G., and Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. Am. J. Hum. Genet. 99, 139–153.

54. Giambartolomei, C., Vuicveic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383.

55. Fang, J., Jia, J., Jakowski, M., Xu, M., Wang, Z., Zhang, T., Hoskins, J.W., Choi, J., Han, Y., Zhang, M., et al.; PanScan Consortium; TRICL Consortium; and GenoMEL Consortium (2017). Functional characterization of a multi-cancer risk locus on chr15p15.33 reveals regulation of TERT by ZNF148. Nat. Commun. 8, 15034.

56. Wang, Z., Zhu, B., Zhang, M., Parikh, H., Jia, J., Chung, C.C., Sampson, J.N., Hoskins, J.W., Hutchinson, A., Burdette, L., et al. (2014). Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 8p15.33. Hum. Mol. Genet. 23, 6616–6633.

57. Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., et al.; ENCODE Project Consortium (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710.

58. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 100041, July 8, 2021.
genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. Gut 63, 800–807.

60. Panagiotou, O.A., Travis, R.C., Campa, D., Berndt, S.I., Lindstrom, S., Kraft, P., Schumacher, F.R., Siddig, A., Patapheodorou, S.I., Stanford, J.L., et al.; PRACTICAL Consortium (2015). A genome-wide pleiotropy scan for prostate cancer risk. Eur. Urol. 67, 649–657.

61. Fehringler, G., Kraft, P., Pharoah, P.D., Eeles, R.A., Chatterjee, N., Schumacher, F.R., Schildkraut, J.M., Lindstrom, S., Brennan, P., Bickeboller, H., et al.; Ovarian Cancer Association Consortium (OCAC); PRACTICAL Consortium; Hereditary Breast and Ovarian Cancer Research Group Netherlands (HE-BON); Colorectal Transdisciplinary (CORECT) Study; and African American Breast Cancer Consortium (AABC) and African Ancestry Prostate Cancer Consortium (AAPC) (2016). Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. Cancer Res. 76, 5103–5114.

62. Wu, Y.H., Graff, R.E., Passarelli, M.N., Hoffman, J.D., Ziv, E., Hoffmann, T.J., and Witte, J.S. (2018). Identification of Pleiotropic Cancer Susceptibility Variants from Genome-Wide Association Studies Reveals Functional Characteristics. Cancer Epidemiol. Biomarkers Prev. 27, 75–85.

63. Cao, X., Huang, M., Zhu, M., Fang, R., Ma, Z., Jiang, T., Dai, J., Ma, H., Jin, G., Shen, H., et al. (2019). Mendelian randomization study of telomere length and lung cancer risk in East Asian population. Cancer Med. 8, 7469–7476.

64. Haycock, P.C., Burgess, S., Nounou, A., Zheng, G.N., Bowden, J., Wade, K.H., Timpson, N.J., Evans, D.M., Willeit, P., et al.; Telomeres Mendelian Randomization Collaboration (2017). Association Between Telomere Length and Risk of Cancer and Non-Neoplastic Diseases: A Mendelian Randomization Study. JAMA Oncol. 3, 636–651.

65. Zhang, C., Doherty, J.A., Burgess, S., Hung, R.J., Lindstrom, S., Kraft, P., Gong, J., Amos, C.I., Sellers, T.A., Monteiro, A.N., et al.; GECCO and GAME-ON Network: CORECT, DRIVE, ELLIPSE, FOCI, and TRICL (2015). Genetic determinants of telomere length and risk of common cancers: a Mendelian randomization study. Hum. Mol. Genet. 24, 5356–5366.

66. James, M.A., Wen, W., Wang, Y., Byers, I.A., Heymach, J.V., Coombes, K.R., Girard, L., Minna, J., and You, M. (2012). Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the Sp15.33 locus. PLoS ONE 7, e36116.

67. Jia, J., Bosley, A.D., Thompson, A., Hoskins, J.W., Cheuk, A., Collins, I., Parikh, H., Xiao, Z., Ylaya, K., Dzyadyk, M., et al. (2014). CLPTM1L promotes growth and enhances aneuploidy in cancer cells. Cancer Res. 74, 2785–2795.

68. Ni, Z., Tao, K., Chen, G., Chen, Q., Tang, J., Luo, X., Yin, P., Tang, J., and Wang, X. (2012). CLPTM1L is overexpressed in lung cancer and associated with apoptosis. PLoS ONE 7, e52598.

69. Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. 46, 994–1000.

70. Barrett, J.H., Taylor, J.C., Bright, C., Harland, M., Dunning, A.M., Akslen, L.A., Andresen, P.A., Avril, M.F., Azizi, E., Bianchi Scarrà, G., et al.; GenoMEL Consortium (2015). Fine mapping of genetic susceptibility loci for melanoma reveals a mixture of single variant and multiple variant regions. Int. J. Cancer 136, 1351–1360.

71. Kachuri, L., Amos, C.I., McKay, J.D., Johansson, M., Vineis, P., Bueno-de-Mesquita, H.B., Boutron-Ruault, M.C., Johansson, M., Quirós, J.R., Sieri, S., et al. (2016). Fine mapping of chromosome 5p15.33 based on a targeted deep sequencing and high density genotyping identifies novel lung cancer susceptibility loci. Carcinogenesis 37, 96–105.

72. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al. (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat Genet 45, 371–384, 384e1–2.

73. Delgado, D.A., Zhang, C., Chen, L.S., Gao, J., Roy, S., Shinkle, J., Sabarinathan, M., Argos, M., Tong, L., Ahmed, A., et al. (2018). Genome-wide association study of telomere length among South Asians identifies a second RTEL1 association signal. J. Med. Genet. 55, 64–71.

74. Do, S.K., Yoo, S.S., Choi, Y.Y., Choi, J.E., Jeon, H.S., Lee, W.K., Lee, S.Y., Lee, J., Cha, S.I., Kim, C.H., and Park, J.Y. (2015). Replication of the results of genome-wide and candidate gene association studies on telomere length in a Korean population. Korean J. Intern. Med. (Korean. Assoc. Intern. Med.) 30, 719–726.

75. Li, C., Stoma, S., Lotta, L.A., Warner, S., Albrecht, E., Allione, A., Arp, P.P., Broer, L., Buxton, J.L., Da Silva Couto Alves, A., et al. (2020). Genome-wide Association Analysis in Humans Links Nucleotide Metabolism to Leukocyte Telomere Length. Am. J. Hum. Genet. 106, 389–404.

76. Wang, Z., Zhang, Z., Guo, Y., Shui, H., Liu, G., Jin, T., and Wang, H. (2018). Shorter Telomere Length Is Associated with Increased Breast Cancer Risk in a Chinese Han Population: A Case-Control Analysis. J. Breast Cancer 21, 391–398.

77. Pellatt, A.J., Wallin, M., Lundgreen, A., Cawthon, R., and Slatkin, M. (2012). Genetic and lifestyle influence on telomere length of single variant and multiple variant regions. Int. J. Cancer 1351–1360.

78. Bai, Y., Fu, W., Guan, X., Wu, X., Li, G., Wei, W., Feng, Y., Meng, H., Li, H., Li, M., et al. (2020). Co-exposure to multiple metals, TERT-CLPTM1L variants, and their joint influence on leukocyte telomere length. Environ. Int. 140, 105762.

79. Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Ricketts, M., Linger, R., Nseigimana, J., Deloukas, P., et al.; UK Testicular Cancer Collaboration (2010). Variants near DMRT1, TERT and ATFF1 are associated with testicular germ cell cancer. Nat. Genet. 42, 604–607.