

Anderson Cancer Center, Houston, TX, USA, ¹⁹⁴Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, ¹⁹⁵Department of Surgery, Division of Cardiothoracic Surgery, Queen Mary Hospital, Hong Kong, China, ¹⁹⁶Department of Neurosurgery, University of Bonn Medical Center, Bonn, Germany, ¹⁹⁷Department of Pathology, The University of Melbourne, Melbourne, VIC, Australia, ¹⁹⁸University of Minnesota, Minneapolis, MN, USA, ¹⁹⁹Department of Internal Medicine, National Cheng Kung University Hospital and College of Medicine, Tainan, Taiwan, ²⁰⁰Instituto Universitario de Oncología, Universidad de Oviedo, Oviedo, Spain, ²⁰¹Sir Peter MacCallum Department of Oncology, University of Melbourne, St Andrew's Place, East Melbourne, VIC, Australia, ²⁰²Danish Cancer Society Research Center, Copenhagen, Denmark, ²⁰³Department of Pulmonary Medicine, Chang Gung Memorial Hospital, Chiayi, Taiwan, ²⁰⁴Cancer Registry Associazione Iblea Ricerca Epidemiologica, Onlus and Asp Ragusa, Ragusa Italy, ²⁰⁵Kaiser Permanente Northern California, Oakland, CA, USA, ²⁰⁶Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands, ²⁰⁷Imperial College, London, UK, ²⁰⁸Human Genetics Foundation (HuGeF), Torino Italy, ²⁰⁹National Research Centre for the Working Environment, Copenhagen, Denmark, ²¹⁰National Food Institute, Technical University of Denmark, Soborg, Denmark, ²¹¹Department of Biochemistry and Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China, ²¹²Department of Pathology and ²¹³School of Public Health, Li Ka Shing (LKS) Faculty of Medicine, The University of Hong Kong, Hong Kong, China, ²¹⁴Division of Cancer Etiology, Department of Population Sciences, City of Hope and the Beckman Research Institute, Duarte, CA, USA, ²¹⁵Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway, ²¹⁶Department of Research, Cancer Registry of Norway, Oslo, Norway, ²¹⁷Samfundet Folkhälsan, Helsinki, Finland, ²¹⁸University of California San Francisco, San Francisco, CA, USA, ²¹⁹Guangdong Lung Cancer Institute, Medical Research Center and Cancer Center of Guangdong General Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, ²²⁰Department of Pathology and Molecular Diagnostics, Aichi Cancer Center Hospital and ²²¹University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA and ²²²Department of Environmental Health, Korea National Open University, Seoul, Republic of Korea

Received April 22, 2014; Revised June 30, 2014; Accepted July 8, 2014

Genome-wide association studies (GWAS) have mapped risk alleles for at least 10 distinct cancers to a small region of 63 000 bp on chromosome 5p15.33. This region harbors the *TERT* and *CLPTM1L* genes; the former encodes the catalytic subunit of telomerase reverse transcriptase and the latter may play a role in apoptosis. To investigate further the genetic architecture of common susceptibility alleles in this region, we conducted an agnostic subset-based meta-analysis (association analysis based on subsets) across six distinct cancers in 34 248 cases and 45 036 controls. Based on sequential conditional analysis, we identified as many as six independent risk loci marked by common single-nucleotide polymorphisms: five in the *TERT* gene (Region 1: rs7726159, $P = 2.10 \times 10^{-39}$; Region 3: rs2853677, $P = 3.30 \times 10^{-36}$ and $P_{\text{Conditional}} = 2.36 \times 10^{-8}$; Region 4: rs2736098, $P = 3.87 \times 10^{-12}$ and $P_{\text{Conditional}} = 5.19 \times 10^{-6}$, Region 5: rs13172201, $P = 0.041$ and $P_{\text{Conditional}} = 2.04 \times 10^{-6}$; and Region 6: rs10069690, $P = 7.49 \times 10^{-15}$ and $P_{\text{Conditional}} = 5.35 \times 10^{-7}$) and one in the neighboring *CLPTM1L* gene (Region 2: rs451360; $P = 1.90 \times 10^{-18}$ and $P_{\text{Conditional}} = 7.06 \times 10^{-16}$). Between three and five cancers mapped to each independent locus with both risk-enhancing and protective effects. Allele-specific effects on DNA methylation were seen for a subset of risk loci, indicating that methylation and subsequent effects on gene expression may contribute to the biology of risk variants on 5p15.33. Our results provide strong support for extensive pleiotropy across this region of 5p15.33, to an extent not previously observed in other cancer susceptibility loci.

INTRODUCTION

Genome-wide association studies (GWAS) have identified independent susceptibility loci in a region on chromosome 5p15.33 that are associated with at least 10 distinct cancers. The published findings include bladder (1), estrogen-negative breast (2), glioma (3), lung (4–7), ovary (8), melanoma (9), non-melanoma skin (10,11), pancreas (12), prostate (13) and testicular germ cell cancer (14). This degree of pleiotropy for common

susceptibility alleles suggests that the region harbors an important set of elements that could influence multiple cancers. It has been observed previously that one allele may be protective for one cancer while conferring susceptibility to another (15). These independent loci map to ~63,000 bp of 5p15.33 that harbors two plausible candidate genes: *TERT*, which encodes the catalytic subunit of telomerase reverse transcriptase (16) and *CLPTM1L*, which encodes the cleft lip and palate-associated transmembrane 1 like protein (also called cisplatin resistance

related protein, CRR9). *CLPTMIL* appears to play a role in apoptosis and cytokinesis, is overexpressed in both lung and pancreatic cancer and is required for KRAS driven lung cancer (17–21). Germline mutations in *TERT* can cause dyskeratosis congenita (DC), a cancer-prone inherited bone marrow failure syndrome caused by aberrant telomere biology (22). Clinically related telomere biology disorders, including idiopathic pulmonary fibrosis and acquired aplastic anemia, can also be caused by germline *TERT* mutations (reviewed in 23).

To investigate the genetic architecture of common susceptibility alleles across this region of 5p15.33 in multiple cancer sites, we utilized a recently developed method called association analysis based on subsets (ASSET) that combines association signals for an SNP across multiple traits by exploring subsets of studies for true association signals in the same, or the opposite direction, while accounting for the multiple testing required (24). The method has been shown to be more powerful than the standard meta-analysis in the presence of heterogeneity, where the effect of a specific SNP might be restricted to only a subset of traits or/and may have different directions of associations for different traits (24).

RESULTS

In this study, we conducted a cross-cancer fine-mapping analysis of a region on chromosome 5p15.33 known to be associated with multiple cancer sites. We imputed each dataset across a 2 Mb window (chr5: 250 000–2 250 000; hg19) using the 1000 Genomes (1000G) and DCEG reference datasets (25,26) and applied a subset-based meta-analysis method (ASSET) (24) to combine results across six cancers (11 studies) (see Materials and Methods for details). This method has been shown to improve power and interpretation when compared with other traditional methods for the analysis of heterogeneous traits (24).

In the first analysis, we focused on six distinct cancer sites in which 5p15.33 had previously been reported and had a nominal *P*-value in our dataset ('Tier-I studies' scans, see Materials and Methods). We performed the analysis across all studies (77% European, 7% African American and 16% Asian ancestry, ALL scans), and, because the majority of studies and subjects were of European ancestry, we conducted parallel analyses in this group only (EUR scans). Bonferroni correction was used to assess significance, using the threshold at 1.3×10^{-5} , based on the number of single-nucleotide polymorphisms (SNPs) analyzed across the region ($n = 1924$) and the two analyses performed (ALL or EUR scans) (see Materials and Methods). In the second analysis, we examined the regions identified above in eight cancers in which 5p15.33 had not been reported in the literature (NHGRI Catalog of Published GWAS studies: <http://www.genome.gov/gwastudies/>), or did not show a nominal *P*-value in our dataset ('Tier-II studies').

Application of ASSET by sequential conditioning of associated SNPs revealed up to six independent loci on 5p15.33, each influencing risk of multiple cancers (Fig. 1, Table 1; Supplementary Material, Table S1). In the primary analysis of all subjects, we performed the ASSET meta-analysis based on unconditional association results from each of the six cancer scans (11 studies). This identified rs7726159 with the lowest *P*-value ($P = 2.10 \times 10^{-39}$), thus marking Region 1. The next

four SNPs, ranked by *P*-values, were highly correlated with the index SNP based on 1000G CEU data: rs7725218 ($P = 2.98 \times 10^{-39}$, pair-wise $r^2 = 0.90$), rs4449583 ($P = 3.37 \times 10^{-39}$, pair-wise $r^2 = 1.0$), rs7705526 ($P = 1.00 \times 10^{-36}$, pair-wise $r^2 = 0.74$) and rs4975538 ($P = 4.11 \times 10^{-32}$, pair-wise $r^2 = 0.76$). These five SNPs reside in the second and third intron of the *TERT* gene and are common, with effect allele frequencies ranging between 0.18 and 0.43 in African (AFR), 0.35–0.37 in Asian (ASN) and 0.32–0.38 in European (EUR) populations, each estimated in the 1000G project (Supplementary Material, Table S2). A search for surrogates using an r^2 threshold of 0.7 across a 1 Mb window centered on the index SNP did not identify additional highly correlated SNPs. The effect allele (A) of rs7726159 was positively associated with glioma (Glioma Scan) and lung cancer (Asian Lung) ($P = 4.38 \times 10^{-36}$, $OR_{\text{Combined}} = 1.47$; 95% CI = 1.38–1.56), but negatively associated with testicular cancer (TGCT NCI), prostate cancer (Pegasus and AdvPrCa) and pancreatic cancer (ChinaPC) ($P = 5.07 \times 10^{-6}$, $OR_{\text{Combined}} = 0.85$; 95% CI = 0.80–0.91) (Fig. 2A).

The most significant SNP after conditioning on rs7726159 was rs451360 ($P = 1.90 \times 10^{-18}$, $P_{\text{Conditional}} = 7.06 \times 10^{-16}$), residing in intron 13 of *CLPTMIL* and marking Region 2 (Fig. 1, Table 1). Six SNPs were correlated with rs451360 with an $r^2 > 0.7$, all located within 500 kb of this SNP and spanning the entire length of *CLPTMIL*: rs380145, rs13170453, rs37004, rs36115365, rs35953391 and rs7446461. This effect allele (rs451360-A) was positively associated with pancreatic cancer (PanScan) and testicular cancer (TGCT NCI) ($P = 4.38 \times 10^{-13}$, $OR_{\text{Combined}} = 1.34$; 95% CI = 1.24–1.45), but negatively associated with lung cancer (AA Lung, Asian Lung and Eur Lung) ($P = 9.50 \times 10^{-8}$, $OR_{\text{Combined}} = 0.85$; 95% CI = 0.80–0.90) (Fig. 2B). Although large differences were seen in the effect allele frequencies across the 1000G continental populations, 0.02–0.03 in AFR, 0.12 in ASN and 0.17–0.24 in EUR (Supplementary Material, Table S2), the signal was still sufficiently strong to be detected, particularly in African and Asian lung studies, suggesting its importance in lung cancer etiology.

In our sequential conditional analysis, rs2853677 (located in the first intron of *TERT*) was the most significant SNP after conditioning on both rs7726159 and rs451360, thus marking Region 3 ($P = 3.30 \times 10^{-36}$, $P_{\text{Conditional}} = 2.36 \times 10^{-8}$) (Fig. 1, Table 1). No additional SNPs with an $r^2 > 0.7$ were located within 500 kb of this SNP, which has relatively low LD with both rs7726159 ($r^2 = 0.13$) and rs451360 ($r^2 = 0.12$) in 1000G CEU data. Region 3 (rs2853677-A) was positively associated with testicular cancer (TGCT NCI) and pancreatic cancer (PanScan and ChinaPC) ($P = 1.36 \times 10^{-7}$, $OR_{\text{Combined}} = 1.22$; 95% CI = 1.13–1.31), but negatively associated with lung cancer (Asian Lung and AA Lung) and glioma (Glioma scan) ($P = 2.79 \times 10^{-31}$, $OR_{\text{Combined}} = 0.73$; 95% CI = 0.70–0.77) (Fig. 2C). The effect allele frequency for rs2853677 was consistent across the three continental 1000G populations corresponding to the studies included in this analysis: 0.60 in EUR, 0.67 in ASN and 0.71 in AFR (Supplementary Material, Table S2).

A conditional analysis based on the three SNPs above (rs7726159, rs451360 and rs2853677) yielded Region 4, marked by rs2736098 ($P = 3.87 \times 10^{-12}$; $P_{\text{Conditional}} = 5.19 \times 10^{-6}$), a synonymous variant (A305A) in the second exon of *TERT* (Fig. 1, Table 1). Three additional SNPs with an

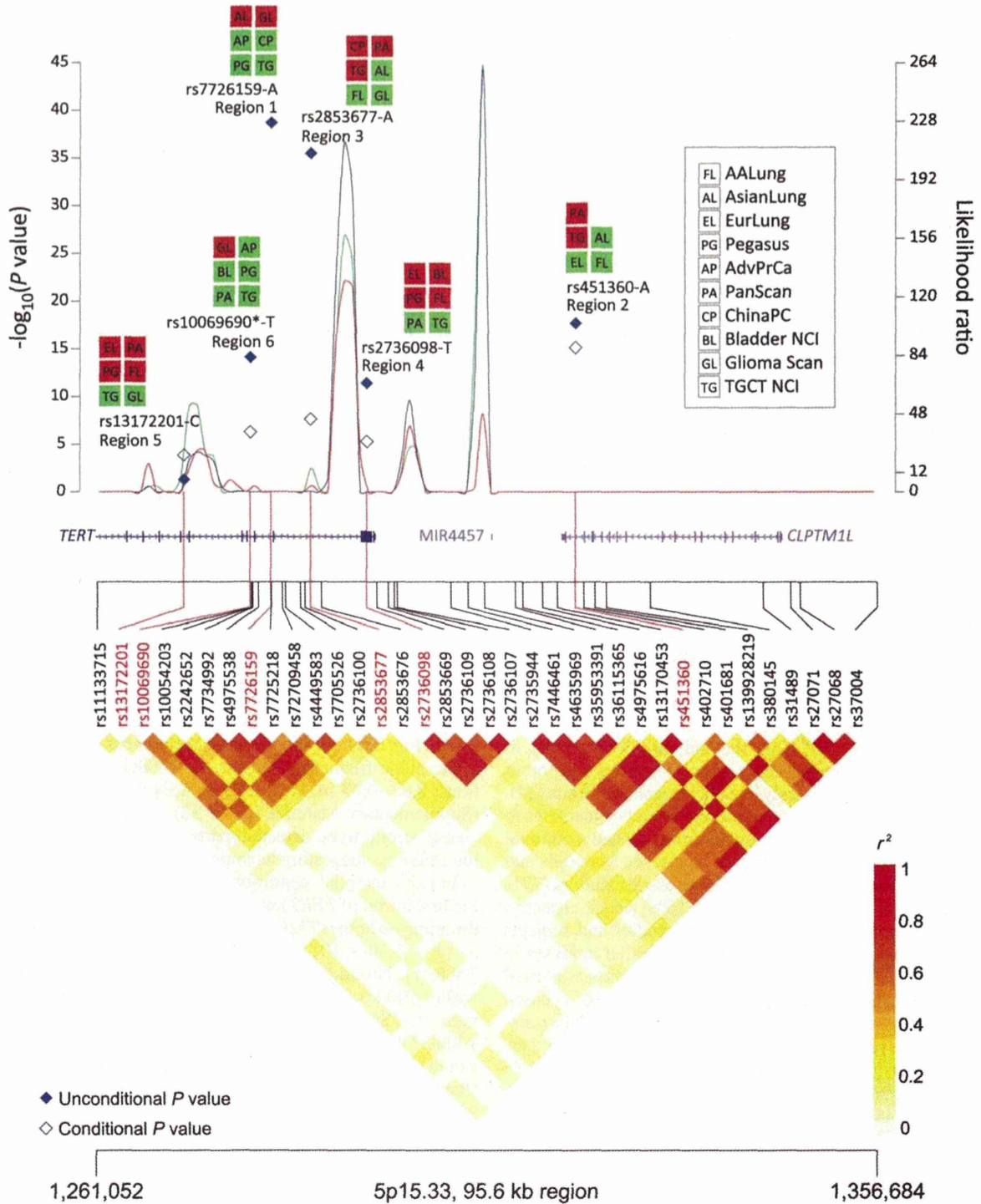


Figure 1. Sequential conditional analyses and ASSET meta-analyses identified up to six independent signals for the *TERT-CLPTM1L* region on chromosome 5p15.33. SNPs marking each region are plotted in the upper panel with two *P*-values (solid diamonds correspond to an unconditional test and open diamonds correspond to a conditional test) on a negative log scale (left *y*-axis) against genomic coordinates (*x*-axis, hg19). Cancers from different GWAS scans (acronyms detailed in box in top panel) that are associated within each region in the subset meta-analysis are listed (red, positively associated; green, negatively associated) from the unconditional ASSET meta-analysis. Effect alleles are shown next to SNP identifiers. Recombination hotspots (curved lines, top panel) were inferred from three populations from the DCEG Imputation Reference Set version 1 (red, CEU; green, ASN; blue, YRI) as the likelihood ratio statistics (right *y*-axis). Also shown are the gene structures for *TERT*, *MIR4457* and *CLPTM1L* (middle panel), and LD heatmap based on r^2 using the 1000 Genomes CEU population (lower panel). Results are shown for the ALL analysis except the region marked by rs10069690 (top panel) and labeled with a "*" that was identified in the European ancestry-only analysis (EUR).

Table 1. Association results for SNPs on chromosome 5p15.33 with the risk of cancer

SNP	Gene	Region	Position	Unconditional OR (95% CI)		Unconditional <i>P</i> -value	Significant phenotype clusters Positively associated	Negatively associated	Conditional OR (95% CI)		Conditional <i>P</i> -value
				Positively associated	Negatively associated				Positively associated	Negatively associated	
ALL											
rs7726159	<i>TERT</i>	1	1282319	1.47 (1.38–1.56)	0.85 (0.80–0.91)	2.10×10^{-39}	AsianLung, Glioma Scan	TGCT NCI, Pegasus, AdvPrCa, ChinaPC			
rs451360	<i>CLPTMIL</i>	2	1319680	1.34 (1.24–1.45)	0.85 (0.80–0.90)	1.90×10^{-18}	PanScan, TGCT NCI	EurLung, AfrAmLung, AsianLung	1.33 (1.23–1.44)	0.86 (0.81–0.92)	7.06×10^{-16}
rs2853677	<i>TERT</i>	3	1287194	1.22 (1.13–1.31)	0.73 (0.70–0.77)	3.30×10^{-26}	TGCT NCI, PanScan, ChinaPC	AsianLung ^a , Glioma Scan, AfrAmLung	1.11 (0.94–1.30)	0.80 (0.74–0.86)	2.36×10^{-8}
rs2736098	<i>TERT</i>	4	1294086	1.15 (1.10–1.21)	0.81 (0.74–0.89)	3.87×10^{-12}	AfrAmLung, Pegasus, EurLung ^a , Bladder NCI	PanScan, TGCT NCI ^a	1.18 (1.10–1.25)	0.94 (0.67–1.31)	5.19×10^{-6}
rs13172201	<i>TERT</i>	5	1271661	1.06 (0.80–1.41)	0.84 (0.73–0.96)	5.00×10^{-2}	EurLung, Pegasus ^a , PanScan, AfrAmLung ^a	TGCT NCI, Glioma Scan	1.13 (1.03–1.23)	0.81 (0.70–0.92)	1.31×10^{-4}
EUR											
rs4449583	<i>TERT</i>	1	1284135	1.50 (1.35–1.68)	0.89 (0.83–0.94)	1.02×10^{-15}	Glioma Scan	TGCT NCI, Pegasus, AdvPrCa, PanScan			
rs13170453	<i>CLPTMIL</i>	2	1317481	1.34 (1.24–1.45)	0.87 (0.80–0.95)	6.69×10^{-15}	PanScan, TGCT NCI	EurLung	1.33 (1.22–1.44)	0.86 (0.80–0.93)	6.67×10^{-14}
rs10069690	<i>TERT</i>	6	1279790	1.48 (1.31–1.67)	0.87 (0.83–0.92)	7.49×10^{-15}	Glioma Scan ^a	AdvPrCa, TGCT NCI ^a , PanScan ^a , Bladder NCI, Pegasus ^a	NA	0.77 (0.69–0.85)	5.35×10^{-7}
rs13172201	<i>TERT</i>	5	1271661	1.07 (0.88–1.29)	0.84 (0.73–0.96)	4.08×10^{-2}	EurLung, Pegasus ^a , PanScan	TGCT NCI, Glioma Scan	1.13 (1.04–1.22)	0.82 (0.75–0.90)	2.04×10^{-6}
rs2736098	<i>TERT</i>	4	1294086	1.14 (1.08–1.20)	0.81 (0.74–0.89)	5.73×10^{-10}	Pegasus, EurLung ^a , Bladder NCI ^a	PanScan, TGCT NCI	1.23 (1.11–1.35)	0.88 (0.75–1.02)	6.31×10^{-5}

The results from the imputation and subset-based ASSET meta-analysis is shown for the 'ALL' scans that include 11 GWAS scans performed in subjects of European, Asian and African American ancestry; and for the 'EUR' scans that include eight scans performed in subjects of European ancestry. Scan acronyms are detailed in Materials and Methods. Listed are SNPs that mark each of the regions identified, gene, genomic location, unconditional and conditional *P*-values and GWAS scans that were positively or negatively associated with the minor allele for each SNP/region. Note that different highly correlated SNPs may mark the same region in the 'ALL' vs. the 'EUR' analysis (Regions 1 and 2). NA indicates that no scan was associated with a particular region.

^aCancer sites that were no longer significant in the conditional analysis.

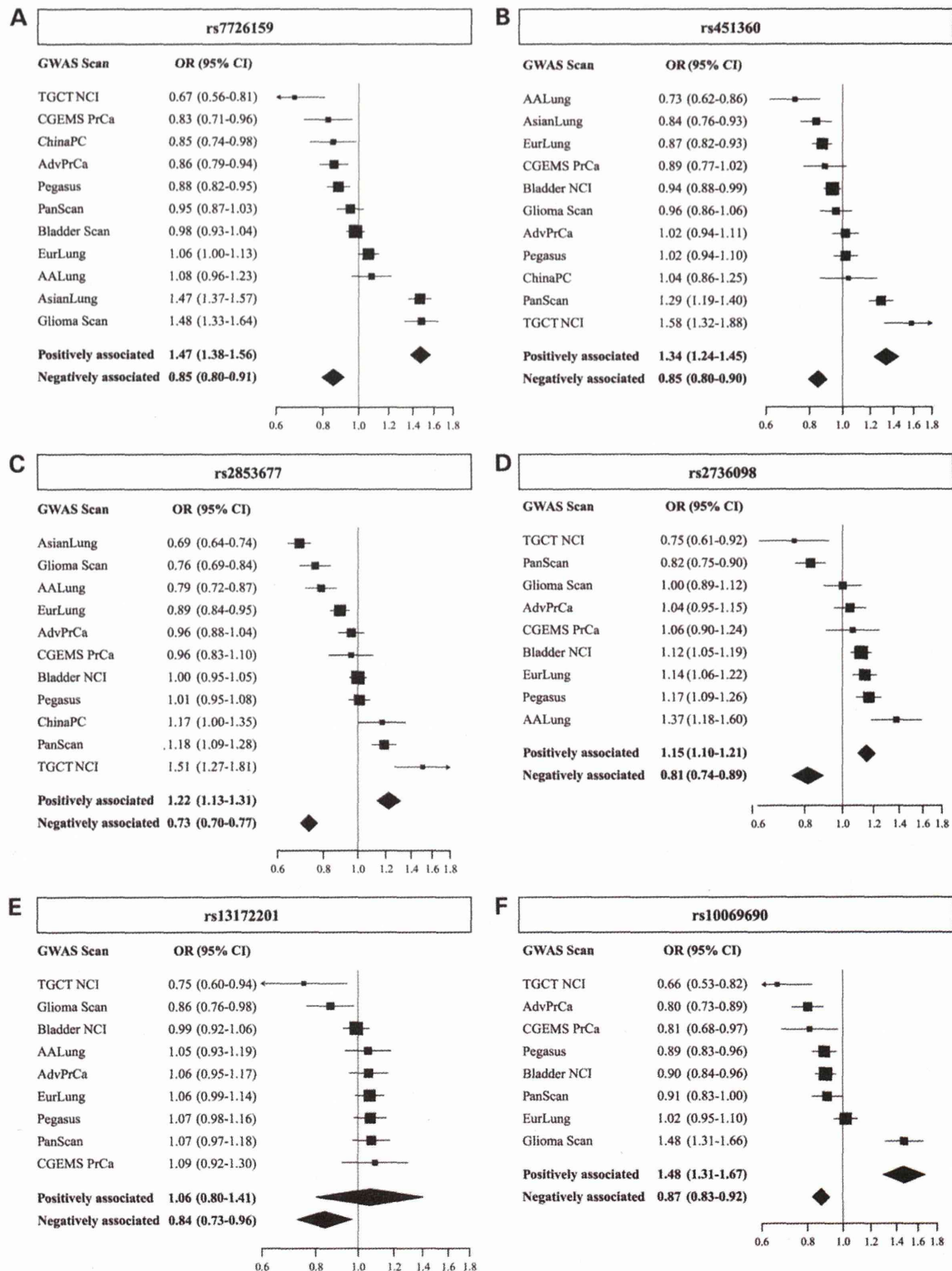


Figure 2. (A–F) Forest plots for individual risk loci on chr5p15.33 for the unconditional ASSET meta-analysis. For each cancer/GWAS scan, OR and 95% CI were listed and plotted along each line as per the unconditional association analysis. A vertical line of OR = 1 indicates the null. Two summary lines list ORs for the positively or negatively associated subsets as estimated by the ASSET program. (A) rs7726159, (B) rs451360, (C) rs2853677, (D) rs2736098, (E) rs13172201 and (F) rs10069690 in the analysis of European-ancestry studies only. Forest plots for the conditional analyses are shown in Supplementary Material, Figure S1A–E.

$r^2 > 0.7$ were located within 500 kb of this SNP: rs2853669, rs2736108 and rs2736107, all in the promoter of *TERT*, from ~200 to 2700 bp upstream of the transcriptional start site. This region (rs2736098-T) was positively associated with lung cancer (Eur Lung and AA Lung), prostate cancer (Pegasus) and bladder cancer (Bladder NCI) ($P = 2.58 \times 10^{-8}$, $OR_{\text{Combined}} = 1.15$; 95% CI = 1.10–1.21), and negatively associated with testicular cancer (TGCT NCI) and pancreatic cancer (PanScan) ($P = 4.89 \times 10^{-6}$, $OR_{\text{Combined}} = 0.81$; 95% CI = 0.74–0.89) (Fig. 2D). The effect allele frequencies displayed a wide range across the three continental populations in 1000G, interestingly with the lowest frequency in the most ancient population, 0.06–0.08 (AFR), whereas the other two populations were comparably high: 0.23–0.29 (EUR) and 0.22–0.33 (ASN) (Supplementary Material, Table S2).

An additional suggestive region (Region 5) marked by rs13172201 ($P = 0.05$; $P_{\text{Conditional}} = 1.31 \times 10^{-4}$) was determined by our sequential conditional analyses (Fig. 1, Table 1), unmasked mainly due to conditioning on rs7726159 (Region 1). The risk alleles for rs13172201 and rs7726159 were negatively correlated ($r = -0.27$, based on 1000G CEU data) and, in an exploratory analysis of rs13172201 in the Eur Lung scan, this SNP appeared to have a stronger association in rs7726159 CC carriers ($P = 7.0 \times 10^{-4}$, $OR = 1.21$ 95% CI = 1.08–1.35) when compared with rs7726159 AC/AA carriers ($P = 0.10$, $OR = 1.12$ 95% CI = 0.98–1.27).

Region 5 (rs13172201-C) was positively associated with lung cancer (Eur Lung and AA Lung), prostate cancer (Pegasus) and pancreatic cancer (PanScan) and negatively associated with testicular cancer (TGCT NCI) and glioma (Glioma scan) (Fig. 2E). The effect allele for rs13172201, the sentinel SNP in Region 5, was the minor allele in European (0.26 in EUR) and African (0.39 in AFR) populations, while it has become the major allele in Asians (0.85 in ASN).

In an analysis restricted to studies of European ancestry (EUR scans), we noted strong associations for Regions 1, 2, 4 and 5 (Table 1) but not Region 3 (marked by rs2853677). The conditional P -value for Region 5, suggestive in the analysis based on all ethnic groups, improved in this subset and surpassed the threshold of 1.3×10^{-5} (rs13172201: $P = 0.041$; $P_{\text{Conditional}} = 2.04 \times 10^{-6}$). An additional region, Region 6, marked by rs10069690 ($P = 7.49 \times 10^{-15}$; $P_{\text{Conditional}} = 5.35 \times 10^{-7}$) in intron 4 of *TERT* was identified in the European ancestry-only analysis (Fig. 1, Table 1). The significance for this region did not reach our Bonferroni-corrected P -value threshold in the analysis of all studies ($P = 5.4 \times 10^{-4}$ after conditioning on rs7726159, rs451360, rs2853677 and rs2736098). As Regions 3 and 6 were located between the same two recombination hotspots (Fig. 1), we assessed correlation in 1000G CEU subjects and noted virtually no LD (rs10069690, rs2853677, $r^2 = 0.0052$), thus supporting the notion that they are independent signals. Low LD existed for these two SNPs in the 1000G YRI ($r^2 = 0.098$) and CHB/JPT ($r^2 = 0.048$) populations (Supplementary Material, Table S3). Region 6 (rs10069690-T) was positively associated with glioma (Glioma scan) ($P = 4.07 \times 10^{-10}$, $OR_{\text{Combined}} = 1.48$; 95% CI = 1.31–1.67) and negatively associated with testicular (TGCT NCI), prostate (Pegasus and AdvPrCa), bladder (Bladder NCI) and pancreatic cancer (PanScan) ($P = 4.95 \times 10^{-7}$, $OR_{\text{Combined}} = 0.87$; 95% CI = 0.83–0.92) (Fig. 2F). Highly correlated SNPs ($r^2 > 0.7$) were

not observed within 500 kb of rs10069690. Notably, the P -value for rs10069690 in the Advanced Prostate cancer scan improved from 1.64×10^{-5} to 2.03×10^{-10} after conditioning on Region 1. The correlation between rs10069690 and rs7726159 (Region 1) is $r^2 = 0.13$ in the 1000G CEU, $r^2 = 0.30$ in YRI and $r^2 = 0.42$ in CHB/JPT populations (Supplementary Material, Table S3). SNP rs10069690 was nominally significant in the other two prostate cancer scans with unconditional P -values of 0.003 (Pegasus) and 0.02 (CGEMS PrCa) but was not significant after conditioning on the first region in these scans ($P = 0.36$ in Pegasus, $P = 0.078$ in CGEMS PrCa).

For the six signals noted, Regions 1, 3 and 6 are flanked by two recombination hotspots that separate them from Region 5 on the telomeric side and from Region 4 on the centromeric side. Recombination hotspots also separate Regions 2 and 4 (Fig. 1). The LD between SNPs in loci 1, 3 and 6 was low to moderate ($r^2 = 0.0052$, 0.131 and 0.449 in 1000G CEU, $r^2 = 0.0981$, 0.298 and 0.0765 in YRI and $r^2 = 0.0484$, 0.415 and 0.341 in CHB/JPT); however, the conditional analyses supported the presence of three signals bounded by strong recombination hotspots on either side. Region 5 is the most telomeric one and separated from the rest by a strong recombination hotspot. Supplementary Material, Table S1 shows P -values for the six regions along each step of the sequential conditional analysis to reflect the change in significance in the analysis.

We also assessed the associations for each of the regions in the ‘Tier-II studies’ comprising nine GWAS datasets across eight cancers, including 11 385 cases and 18 322 controls. None of the regions showed significant association (data not shown).

In addition to characterizing independent signals in the *TERT-CLPTMIL* region, we have fine-mapped previously reported signals. For pancreatic cancer, the reported GWAS SNP rs401681 had a P -value of 3.7×10^{-7} and an OR of 1.19 (12). After imputation, an improved P -value was seen for rs451360 (marking Region 2) ($P = 2.0 \times 10^{-10}$; $OR = 1.29$). After conditioning on rs451360, the P -value for rs401681 was no longer significant ($P = 0.1$). The LD between these two SNPs is moderate ($r^2 = 0.35$). For glioma, the GWAS SNP rs2736100 had a P -value of 8.49×10^{-9} and OR of 1.08 in the Glioma scan (27). The best imputed SNP rs449583 ($r^2 = 1$ with rs7726159, marking Region 1) showed a much improved P -value of 4.1×10^{-14} with an OR of 1.50, and the P -value of rs2736100 was no longer significant after conditioning on rs449583 ($P = 0.64$). The LD between these two SNPs was moderate ($r^2 = 0.39$).

Bioinformatic analyses using public data bases (ENCODE and TCGA) were performed to investigate the possible function of SNPs that mark each of the six regions as regulators of expression of *TERT*, or *CLPTMIL*, as well as other genes. Based on ENCODE data, the strongest evidence for putative regulatory functions was seen for SNPs in Regions 1 (rs7725218 and rs4975538), 2 (rs36115365 and rs380145), 4 (rs2736108 and rs2853669) and 5 (rs13172201) with evidence of an open chromatin conformation, regulatory histone modification marks and transcription factor binding in multiple cell types such as prostate, pancreas, breast, lung and brain (Supplementary Material, Table S2).

We next examined the TCGA datasets for expression (eQTL) and methylation (meQTL) quantitative trait loci for lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD) and

glioblastoma multiforme (GBM). We did not observe significant eQTLs ($P > 0.41$, data not shown) but noted multiple meQTLs in LUAD and PRAD tumor samples (Supplementary Material, Tables S5 and S6). Methylation at a subset of CpG probes with meQTLs correlated with expression of *TERT* and/or *CLPTMIL*, including two for Region 4 in TCGA LUAD samples (cg26209169: $\beta = -0.47$, $P = 1.18 \times 10^{-5}$; cg11624060: $\beta = -0.36$, $P = 0.001$). These CpGs are located ~1800 bp downstream of *CLPTMIL* (227 bp apart), overlap with key transcription factor binding sites (e.g. TCF3, TCF4, HNF3A, MAX, RUNX3/AML2, ATF-2 and USF1/USF2) and active histone modification marks from ENCODE, and are negatively correlated with expression of *TERT* and *CLPTMIL* (Supplementary Material, Table S5 and Fig. S2). Replication was seen in normal lung samples (cg26209169 and Region 4, $\beta = -0.650$, $P = 5.17 \times 10^{-5}$; cg11624060 and Region 4, $\beta = -0.493$, $P = 0.0027$) from EAGLE (28). The most significant meQTLs in TCGA PRAD samples were seen for Region 1 (cg03935379: $\beta = -1.06$, $P = 8.47 \times 10^{-15}$; cg06531176: $\beta = -1.18$, $P = 2.61 \times 10^{-15}$). These replicated in EAGLE ($P = 5.93 \times 10^{-8}$ and $P = 0.002$, respectively), did not correlate with expression of *TERT* or *CLPTMIL*, and were both located within exon 3 of *TERT* (Supplementary Material, Table S6).

Analysis of TCGA data also revealed increased expression of *TERT* and *CLPTMIL* in tumors compared with normal tissues for lung and prostate cancer (on average 1.29- to 2.02-fold change for paired samples). Copy number differences were more evident in lung tumors (average number of copies was 2.02 in normal and 2.54 in tumors for 51 paired samples, $P = 1.10 \times 10^{-7}$) (Supplementary Material, Fig. S3).

DISCUSSION

Chr5p15.33 harbors a unique cancer susceptibility region that contains at least two plausible candidate genes: *TERT* and *CLPTMIL*. Through a subset-based meta-analysis of GWAS data drawn from six different cancers from three continental populations, we have characterized up to six independent, common, susceptibility alleles, all with evidence of both risk-enhancing and protective effects, differing by cancer type.

TERT encodes the catalytic subunit of the telomerase reverse transcriptase, which, in combination with an RNA template (*TERC*), adds nucleotide repeats to chromosome ends (29). Although telomerase is active in germ cells and in early development, it remains repressed in most adult tissues. Telomeres shorten with each cell division and when they reach a critically short length, cellular senescence or apoptosis is triggered. Cancer cells can continue to divide despite critically short telomeres, by upregulating telomerase or by alternative lengthening of telomeres (16,30,31). While studies investigating the relationship between surrogate tissue (i.e. buccal or blood cell DNA) telomere length and cancer risk have been contradictory, larger prospective studies have not reported an association for risk but only survivorship (32–35). Heritability estimates of telomere length in twin studies suggest a significant genetic contribution, between 36 and 78% (36,37). GWAS SNPs on 5p15.33 have been associated with telomere length implying that *TERT* may indeed be the gene targeted by at least some risk variants

in this region (38–40). In addition, germline *TERT* promoter mutations have been identified in familial melanoma as well as somatic mutations in multiple cancers (41,42).

The most commonly reported SNP in the *TERT* gene, rs2736100, was first reported in several GWAS: glioma (3,43), lung cancer in European and Asians (7,44–46) and testicular cancer (14). We have fine-mapped this locus (Region 1) to a set of five correlated SNPs in the second and third intron of *TERT* (marked by rs7726159). In addition to the cancers listed above, we noted novel contributions to this locus by prostate and pancreatic cancer. Fine-mapping efforts in lung (47) and ovarian cancer (48) have reported the same SNP. Region 3 (rs2853677), located in the first intron of *TERT*, has been associated with glioma in Chinese subjects (49) and lung cancer in Japanese subjects (50), in agreement with the strong contribution to this region seen in our analysis by scans performed in individuals of Asian ancestry. In addition to lung cancer and glioma, we noted novel associations for Region 3 with pancreatic and testicular cancer. Region 4 was marked by a synonymous SNP (rs2736098) located in the second exon of *TERT*, with three additional highly correlated SNPs in the promoter region. This region has been reported via fine-mapping in lung, bladder, prostate, ovarian and breast cancer, and shown to influence *TERT* promoter activity (8). Novel contributions to Region 4 were noted for pancreatic and testicular cancer.

In our analysis, we uncovered a new susceptibility locus, Region 5 (marked by rs13172201, Fig. 1), which surpassed the Bonferroni threshold in European studies. We found evidence for a negative correlation between this SNP and rs7726159 (Region 1), indicating a possible interaction. This locus is not significant at a GWAS threshold and requires confirmation in independent samples. Region 6 (marked by rs10069690) has previously been associated with estrogen- and progesterone receptor-negative breast cancer in populations of European and African ancestry (2,51); our analysis adds five cancers to this list: glioma, prostate, testicular germ cell, pancreas and urinary bladder.

The gene adjacent to *TERT*, namely *CLPTMIL*, encodes a protein that is overexpressed in lung and pancreatic cancer, promotes growth and survival, and is required for KRAS driven lung cancer, indicating that it is a plausible candidate gene in this region (17–21). The locus in *CLPTMIL* (Region 2) has previously been associated with risk of cancer in multiple GWAS, marked by rs401681 or rs402710 in pancreatic, lung and bladder cancer as well as in melanoma (1,4,5,12,52). Our subset-based approach has fine-mapped this signal to a set of seven correlated SNPs that span the entire length of *CLPTMIL*.

Two recent papers from the Collaborative Oncology Gene-Environment Study (COGs) fine-mapped 5p15.33 in prostate, breast and ovarian cancer and identified four of the six loci noted in the current study (53,54). In prostate cancer, COGs identified three regions that corresponded to our Region 1 (COGs Region 1, rs7725218), Region 3 (COGs Region 2, rs2853676, $r^2 = 0.32$ with rs2853677) and Region 4 (COGs Region 3, rs2853669) (54). Interestingly, COGs reported protective alleles in Region 1 associated with increased *TERT* expression in benign prostate tissue samples. The fourth COGs prostate cancer locus, marked by rs13190087, was not significant in our study ($P = 0.089$), possibly due to a more specific effect for prostate cancer for this locus where our study had less power.

In breast and ovarian cancer, COGs identified three regions corresponding to our Region 1 (COGs Region 2, rs7705526, associated with risk of ovarian cancer with low malignant potential, telomere length and promoter activity), Region 4 (COGs Region 1, rs2736108, associated with risk of ER-negative and *BRCA1* mutation carrier breast cancer, telomere length and altered promoter activity) and Region 6 (COGs Region 3, rs10069690, associated with risk of ER-negative breast cancer, breast cancer in *BRCA1* carriers and invasive ovarian cancer) (53). Regions 2 (in *CLPTMIL*) and 5 (in *TERT*) were not observed in the COGs reports, perhaps due to the choice of SNPs by COGs for fine-mapping as well as the more comprehensive reference set for 1000 Genomes used to conduct our imputation, or because of cancer-specific effects for these loci.

It is becoming increasingly clear that DNA methylation is under genetic control. Regions of variable methylation exist across tissues and individuals, tend to be located in intergenic regions, overlapping known regulatory elements. Notably, these are enriched for disease-associated SNPs (28,55,56). Analysis of TCGA data, while not uncovering significant eQTLs, indicated that DNA methylation could play a role in the underlying biology at 5p15.33. Methylation in a small region downstream of *CLPTMIL*, with features supporting an active regulatory function, was consistent with lower methylation levels in carriers of risk alleles for lung cancer (Region 4) and higher expression of *TERT* and *CLPTMIL*. Increased expression of both genes is consistent with a pro-tumorigenic role in lung cancer (19,21,31). For prostate cancer, the most notable meQTLs were located within exon 3 of *TERT* with increased rates of methylation for carriers of risk alleles in Regions 1 and 6. Although gene-body methylation has been observed to positively correlate with gene expression (57), we did not see evidence to support this for this particular set of CpGs. As a large fraction of meQTLs does not overlap with eQTLs (55), they may influence molecular phenotypes other than gene expression such as alternative promoter usage, splicing and even mutations (58–60). It is intriguing that methylation QTLs observed in TCGA data differ to some degree between lung and prostate cancer, and that none were observed in glioblastoma. This indicates that the *TERT-CLPTMIL* region may harbor multiple elements that have the capacity to influence molecular phenotypes that in turn impact cancer development. However, only a subset of these elements may be active in each organ, thus leading to different mechanistic avenues for risk modulation in different tissues. It is possible that the interplay between risk variants, multiple biological mechanisms and attributed genes, in addition to environmental and lifestyle factors that differentially influence various cancers may eventually come to explain how the same alleles at this complex locus can mediate opposing cancer risk in different organs.

In summary, we report up to six independent loci on chr5p15.33, each influencing the risk of multiple cancers. We observed pleiotropy for common susceptibility alleles in this region, defined as the phenomenon wherein a single genetic locus affects multiple phenotypes (61). These alleles could influence multiple cancers distinctly, perhaps in response to environmental factors or in interactions with other genes. Our cardinal observations underscore the complexity of the alleles and suggest the importance of tissue-specific factors that contribute to cancer susceptibility. Further laboratory analysis is needed to

validate our findings using TCGA data, and investigate the optimal functional variants in each of the six independent loci in order to provide a clearer understanding of each of the loci in this multi-cancer susceptibility region.

MATERIALS AND METHODS

Study participants

Participants were drawn from a total of 20 previous GWAS scans of 13 distinct cancer types: bladder, breast, endometrial, esophageal squamous, gastric, glioma, lung, osteosarcoma, ovarian, pancreatic, prostate, renal cancer and testicular germ cell tumors. We first assessed a set of 11 GWAS representing six distinct cancers ('Tier-I studies') in which 5p15.33 had previously been implicated (NHGRI Catalog of Published GWAS studies: <http://www.genome.gov/gwastudies/>). The GWAS scans and their acronyms were: Asian lung cancer scan (AsianLung), European lung cancer scan (EurLung), African American lung (AA Lung), PanScan, China pancreatic cancer scan (ChinaPC), Testicular germ cell tumor (TGCT NCI) scan, glioma scan, Bladder NCI scan, Pegasus prostate cancer scan (Pegasus), CGEMS prostate cancer scan (CGEMS PrCa) and Advanced prostate cancer scan (Adv PrCa) (see case and control counts in Supplementary Material, Tables S4A–D). In a second analysis, we separately assessed a set of nine GWAS scans representing eight cancers ('Tier-II studies') in which 5p15.33 had not been previously reported in the literature (NHGRI Catalog of Published GWAS studies: <http://www.genome.gov/gwastudies/>). These studies were: Asian esophageal scan (Asian EsoCa), Asian gastric cancer scan (Asian GastCa), CGEMS Breast cancer scan (CGEMS Breast), Endometrial cancer scan (EndomCa), ER negative breast cancer scan (ERneg BPC3 BrCa), Ghana prostate cancer scan (Ghana PrCa), Osteosarcoma scan (OS), Ovarian cancer scan (OvCa) and Renal cancer scan (Renal US) (see case and control counts in Supplementary Material, Tables S4E–H). Studies were conducted in individuals of European background (EUR scans) but we did include studies in populations of Asian ancestry (i.e., esophageal squamous, gastric, non-smoking lung and pancreatic cancers) and African ancestry (i.e. lung and prostate cancer) (ALL scans). Study characteristics, genotyping and quality control have been previously published for all studies listed by cancer type and GWAS scan acronym: bladder cancer/Bladder NCI (1,62), breast cancer/CGEMS BrCa (63), breast cancer/ERneg BPC3 BrCa (64), endometrial cancer/EnCa (65), gastric cancer and esophageal squamous cell carcinoma/Asian UpperGI (66), glioma/Glioma scan (27), lung cancer in Europeans/EurLung (7), lung cancer in African Americans/AALung (67), lung cancer in non-smoking women from Asia/AsianLung (68,69), osteosarcoma/OS (70), ovarian cancer/OvCa (71), pancreatic cancer/PanScan (12,72), pancreatic cancer in Asians/ChinaPC (73), prostate cancer/Pegasus (unpublished data), prostate cancer/CGEMS PrCa (74), advanced prostate cancer/AdvPrCa (75), prostate cancer in Africans/GhanaPrCa (unpublished data), renal cancer/Renal US (76) and testicular germ cell tumors/TGCT NCI (77).

Each participating study obtained informed consent from study participants and approval from its Institutional Review Board (IRB) including IRB certification permitting data

sharing in accordance with the National Institutes of Health (NIH) Policy for Sharing of Data Obtained in NIH Supported or Conducted GWAS.

Genotyping

Arrays used for scanning included the Illumina HumanHap series (317 + 240S, 550, 610 K, 660 W and 1 M), as well as the Illumina Omni series (OmniExpress, Omni1M, Omni2.5 and Omni5M). The majority of the studies were genotyped at the Cancer Genomics Research Laboratory (formerly Core Genotyping Facility) of the National Cancer Institute (NCI) of the NIH. The ChinaPC GWAS (Affymetrix 6.0) was genotyped at CapitalBio in Beijing, China. This necessitated imputation before the cross-cancer subset-based meta-analysis. We used a combination of public resources, 1000 Genomes (1000G) (25) and DCEG (26) reference datasets, to impute existing GWAS datasets (78) using IMPUTE2 (79).

In addition to the standard QC procedures previously applied in the primary GWAS publications, we further filtered SNPs as follows: (i) completion rate per locus < 90%, (ii) MAF < 0.01, (iii) Hardy–Weinberg proportion P -value < 1×10^{-6} , (iv) exclusion of A/T or G/C SNPs.

Lift over the genomic coordinates to NCBI genome build 37 or hg19

Because the March 2012 release of the 1000 Genomes Project data is based on NCBI genome build 37 (hg19), we utilized the LiftOver tool (<http://hgdownload.cse.ucsc.edu/>) to convert genomic coordinates for scan data from build 36 to build 37. The tool re-maps only coordinates, but not SNP identifiers. We prepared the inference.bed file and then performed the lift over as follows:

```
~/tools/liftover/liftOver inference.bed ~/tools/liftover/hg18
ToHg19.over.chain.gz output.bed unlifted.bed
```

A small number of SNPs that failed LiftOver, mostly because they could not be unambiguously mapped to the genome by NCBI, were dropped from each imputation inference set.

Strand alignment with 1000 Genomes reference data set

Since A/T or G/C SNPs were excluded, strand alignment for the scan data required checking allele matches between the inference set and reference set locus by locus. If they did not match, alleles were complemented and checked again for matching. SNPs that failed both approaches were excluded from the inference data. Locus identifiers were normalized to those used in the 1000 Genomes data based on genomic coordinates, although the IMPUTE2 program uses only the chromosome/location to align each locus overlapping between the imputation inference and reference set.

Conversion of genotype files into WTCCC format

After LiftOver to genome build 37 and ensuring that alleles were reported on the forward strand, we converted the genotype data into IMPUTE2 format using GLU. We split the genotype file into one per chromosome and sorted SNPs in order of genomic location using the GLU *transform* module.

Imputation of a 2Mb window on chr5p15.33

We used both the 1000G data (March 2012 release) (25) and the DCEG imputation reference set (26) as reference datasets to improve overall imputation accuracy. The IMPUTE2 program (79) was used to impute a 2 Mb window on chr5p15.33 from 250 000 to 2 250 000 (hg19) with a 250 kb buffer on either side as well as other recommended default settings. For the association analysis, we focused on a smaller region from chr5: 1 250 000–1 450 000 delineated by recombination hotspots (discussed below).

Post-imputation filtering and association analysis

We excluded imputed loci with INFO < 0.5 from subsequent analyses. SNPTTEST (79) was used for the association analysis with covariate adjustment and score test of the log additive genetic effect. The same adjustments as used originally in each individual scan were used. Note that the per SNP imputation accuracy score (IMPUTE's INFO field) is calculated by both IMPUTE2 and SNPTTEST. The two INFO metrics calculated during imputation by IMPUTE2 and during association testing by SNPTTEST are strongly correlated, especially when the additive model is fitted (78). We chose the INFO metric calculated by SNPTTEST for post-imputation SNP filtering.

Subset and conditional analyses

Association outputs from SNPTTEST were reformatted and subsequently analyzed using the ASSET program, an R package (<http://www.bioconductor.org/packages/devel/bioc/html/ASSET.html>; https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/inst/doc/vignette.Rnw?root=asset) for subset-based meta-analyses (24). ASSET is a suite of statistical tools specifically designed to be powerful for pooling association signals across multiple studies when true effects may exist only in a subset of the studies and could be in opposite directions across studies. The method explores all possible subset (or a restricted set if user specifies so) of studies and evaluates fixed-effect meta-analysis-type test-statistics for each subset. The final test-statistics is obtained by maximizing the subset-specific test-statistics over all possible subsets and then evaluating its significance after efficient adjustment for multiple testing, taking into account the correlation between test-statistics across different subsets due to overlapping subjects. The method not only returns a P -value for significance for the overall evidence of association of an SNP across studies, but also outputs the 'best subset' containing the studies that contributed to the overall association signal. For detection of SNP association signals with effects in opposite directions, ASSET allows subset search separately for positively and negatively associated studies and then combines association signals from two directions using a chi-square test-statistics. The method can take into account correlation due to overlapping subject across studies (e.g. share controls). More details about these and other features of the method can be found elsewhere [22].

For our current study, the matrices of the overlapping counts for cases–controls across datasets, which are utilized by ASSET to adjust for possible correlation across studies, were constructed and passed into the ASSET program (Supplementary Tables S4A–H). We used a two-sided test P -value, which

can combine association signals in opposite directions, to assess the overall significance of whether an SNP was associated with the cancers under study. For detection of independent susceptibility SNPs, we performed sequential conditional analysis in which in each step the ASSET analysis is repeated by conditioning on SNPs that have been detected to be most significant in previous steps. The process was repeated until the P -value for the most significant SNP for a step remained $< 1.3 \times 10^{-5}$, a conservative threshold that corresponds to Bonferroni adjustment for the 1924 SNPs used in the analysis for an alpha level of 0.05 and the two analyses performed (for the ALL vs. the EUR scans).

In the primary analysis, we included all GWAS scans in which one or more susceptibility alleles on 5p15.33 had been previously noted at genome-wide significant threshold ('Tier-I studies'). We further required a nominal signal in our data ($P < 0.05$). This yielded 11 GWAS across six distinct cancer sites and includes 34 248 cases and 45 036 cancer-free controls (Supplementary Material, Tables S4A–D). In a secondary analysis, we assessed the associations for each of the six regions in scans in which 5p15.33 had not been previously reported in the literature (<http://www.genome.gov/gwastudies/>), or did not show a nominal P -value in the GWAS datasets used in the current study ('Tier-II studies'). This yielded nine GWAS datasets across eight cancers, including a total of 11 385 cases and 18 322 controls (Supplementary Material, Tables S4E–H).

Recombination hotspot estimation

Recombination hotspots were identified in the region of 5p15.33 harboring *TERT* and *CLPTMIL* (1 264 068–1 360 487) using SequenceLDhot (80), a program that uses the approximate marginal likelihood method (81) and calculates likelihood ratio statistics at a set of possible hotspots. We tested three sample sets from East Asians ($n = 88$), CEU ($n = 116$) and YRI ($n = 59$) from the DCEG Imputation Reference Set. The PHASE v2.1 program was used to calculate background recombination rates (82,83).

Validation of imputation accuracy

Imputation accuracy was assessed by direct TaqMan genotyping. TaqMan genotyping assays (ABI, Foster City, CA, USA) were optimized for six SNPs (rs7726159, rs451360, rs2853677, rs2736098, rs10069690 and rs13172201) in the independent regions. In an analysis of 2327 samples from the Glioma brain tumor study (Glioma BTS, 330 samples) (27), testicular germ cell tumor (TGCT STEED study, 865 samples) (77) and Pegasus (PLCO, 1132 samples) (unpublished data), the allelic R^2 (84) measured between imputed and assayed genotypes were 0.88, 0.98, 0.86, 0.85, 0.81 and 0.61 for the six SNPs listed in the same order as above.

Bioinformatic analysis of functional potential

HaploReg v2 (<http://www.broadinstitute.org/mammals/haploreg/haploreg.php>) was used to annotate functional and regulatory potential of highly significant and highly correlated SNPs that mark each of the regions identified (using ENCODE data) (85). RegulomeDB (<http://regulome.stanford.edu/>) was used to assess and

score regulatory potential of SNPs in each locus (86). eQTL effects were assessed using the Multiple Tissue Human Expression Resource database (<http://www.sanger.ac.uk/resources/software/genevar/>) but significant findings at a $P < 1 \times 10^{-3}$ threshold were not noted (data not shown). Predicted effects of SNPs on splicing were assessed using NetGene2 (<http://www.cbs.dtu.dk/services/NetGene2/>) (88) but no effect were seen for any of the SNPs in the six regions (data not shown).

We carried out eQTL and methylation quantitative trait locus (meQTL) analyses to assess potential functional consequences of SNPs in the six regions identified in normal and tumor derived tissue samples from TCGA: LUAD (52/403 normal/tumor samples for eQTL analysis; 26/354 normal/tumor samples for meQTL analysis), PRAD (31/133 normal/tumor for eQTL; 39/158 normal/tumor for meQTL) and GBM (109 tumor for eQTL; 83 tumor for meQTL; normal GBM samples were not available). Transcriptome (Illumina HiSeq 2000, level 3), methylation (Illumina Infinium Human DNA Methylation 450 platform, level 3), genotype data (Affymetrix Genome-Wide Human SNP Array 6.0 platform, level 2) and phenotypes were downloaded from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>). Methylation probes located on X/Y chromosomes, annotated in repetitive genomic regions (GEO GPL16304), with SNPs (Illumina dbSNP137.snpupdate.table.v2) with MAF $> 1\%$ in the respective TCGA samples, with missing rate $> 5\%$, as well as 65 quality control probes on the 450 K array. We excluded transcripts on X/Y chromosomes and those with missing rate $> 5\%$. A principle component analysis was conducted on a genome-wide level in R using gene expression and methylation data (separately in normal and tumor tissues, and after excluding transcripts with variance $< 10^{-8}$ and methylation probes with variance < 0.001). Genotype imputation was performed as described above for the 2 Mb window centered on *TERT* and *CLPTMIL*. For eQTL analysis, normalized transcript counts for *CLPTMIL* and *TERT* were normal quantile transformed and regressed against the imputed dosage of minor allele for each risk locus (six loci, 19 SNPs). The regression model included age, gender (not for PRAD), stage (only for tumor samples), copy number, top five principle components (PCs) of imputed genotype dosage and top five PCs of transcript counts to account for possible measured or unmeasured confounders and to increase detection power. The meQTL analysis was conducted in a similar manner in TCGA LUAD, PRAD GBM samples; beta-values of methylation at 169 CpG probes in the region encompassing *TERT* and *CLPTMIL* were normal quantile transformed and regressed as described above with the exception of inclusion of the top five PCs of methylation instead of expression values. We report the estimate of regression coefficient of imputed dosage, its standard error and P -values, adjusted by the Benjamini–Hochberg procedure for controlling false discovery rate (89). Spearman's rank-order correlation was calculated to assess the relationship between the methylation and gene expression for TCGA LUAD ($n = 486$), PRAD ($n = 186$) and GBM ($n = 126$) tumor samples. P -values were adjusted by the Benjamini–Hochberg procedure as described above. For the purpose of visualizing meQTLs, the most likely genotype was selected from the imputed genotype dosages.

Methylation QTLs were assessed in EAGLE normal lung tissue samples ($n = 215$) as previously described with the