Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148

Jun Fang1,*, Jinping Jia1,*, Matthew Makowski1,2, Mai Xu1, Zhaoming Wang3,4, Tongwu Zhang1, Jason W. Hoskins1, Jiyeon Choi1, Younghun Han5, Mingfeng Zhang1, Janelle Thomas1, Michael Kovacs1, Irene Collins1, Marta Dzyadyk1, Abbey Thompson1, Maura O’Neill6, Sudipto Das6, Qi Lan1, Roelof Koster1, PanScan Consortium†, TRICL Consortium‡, GenoMEL Consortiumγ, Rachael S. Stolzenberg-Solomon3, Peter Kraft7,8, Brian M. Wolpin9,10, Pascal W.T.C. Jansen2, Sara Olson11, Katherine A. McGlynn3, Peter A. Kanetsky12, Nilanjana Chatterjee3, Jennifer H. Barrett13, Alison M. Dunning14, John C. Taylor13, Julia A. Newton-Bishop13, D. Timothy Bishop13, Thorkell Andresson6, Gloria M. Petersen15, Christopher I. Amos5, Mark M. Iles13, Katherine L. Nathanson16, Maria Teresa Landi3, Michel Vermeulen2, Kevin M. Brown1,** & Laufey T. Amundadottir1,***

Genome wide association studies (GWAS) have mapped multiple independent cancer susceptibility loci to chr5p15.33. Here, we show that fine-mapping of pancreatic and testicular cancer GWAS within one of these loci (Region 2 in CLPTM1L) focuses the signal to nine highly correlated SNPs. Of these, rs36115365-C associated with increased pancreatic and testicular but decreased lung cancer and melanoma risk, and exhibited preferred protein-binding and enhanced regulatory activity. Transcriptional gene silencing of this regulatory element repressed TERT expression in an allele-specific manner. Proteomic analysis identifies allele-preferred binding of Zinc finger protein 148 (ZNF148) to rs36115365-C, further supported by binding of purified recombinant ZNF148. Knockdown of ZNF148 results in reduced TERT expression, telomerase activity and telomere length. Our results indicate that the association with chr5p15.33-Region 2 may be explained by rs36115365, a variant influencing TERT expression via ZNF148 in a manner consistent with elevated TERT in carriers of the C allele.
risk variants across a small genomic region on chromosome 5p15.33 have been reported in genome wide association studies (GWAS) for at least eleven cancer types including bladder, breast, glia, lung, melanoma, non-melanoma skin cancer, ovarian, pancreas, prostate, testicular germ cell cancer and chronic lymphocytic leukaemia14–16. Fine-mapping studies, either within a specific cancer type or across different cancers, have characterized up to seven independent loci in this region with either risk-enhancing or protective effects across a dozen cancers16–18. Notable is the fact that in nearly every locus, the effect is pleiotropic. This genomic region contains two plausible candidate genes, \( \text{TERT} \) and \( \text{CLPTM1L} \). The former encodes the catalytic subunit of the telomerase reverse transcriptase (TERT), which in combination with an RNA template (TERC) adds nucleotide repeats to chromosome ends19. Although telomerase is active in germ cells and in early development, it is 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. However, cancer cells can continue to divide despite critically short telomeres, by upregulating telomerase or by alternative lengthening of telomeres (ALT) (refs 20–22). The \( \text{CLPTM1L} \) gene encodes the cleft lip and palate associated transmembrane 1-like protein, and is overexpressed in lung and pancreatic cancer where it promotes growth and survival and is required for KRAS driven lung cancer23–25.

One of the multiple risk loci in this genomic region lies within the \( \text{CLPTM1L} \) gene and has been termed Region 2 (ref. 18), originally reported to be associated with risk of pancreatic, lung, bladder cancer, and melanoma, marked by either rs401681 or rs402710 (refs 1,4,5,11,28). By conducting fine-mapping across multiple cancers and subsequently investigating the functional consequences of the subset of genetic variants most strongly associated with cancer risk, we find that risk of pancreatic, testicular and lung cancer conferred by this locus may predominantly be explained by a single-SNP. This variant, rs36115365, exhibited preferred protein-binding and enhanced regulatory activity for the C-allele, associated with increased pancreatic and testicular but decreased lung cancer and melanoma risk.

Transcriptional gene silencing of the regulatory region encompassing this variant resulted in repression of \( \text{TERT} \) but not \( \text{CLPTM1L} \) expression in an allele-specific manner. Proteomic analysis identified allele-preferred binding of Zinc finger protein 148 (ZNF148) to rs36115365-C, a finding supported by binding of purified recombinant ZNF148 specifically to the C-allele, as well as by ChIP analysis showing allele-preferential binding of endogenous ZNF148 to rs36115365-C. Knockdown of ZNF148 resulted in reduced \( \text{TERT} \) expression, telomerase activity and telomere length. Taken together, these results indicate that the association with chr5p15.33-Region 2 may be explained by rs36115365, a variant influencing \( \text{TERT} \) via ZNF148 in a manner consistent with elevated \( \text{TERT} \) expression in carriers of the C allele.

**Results**

**Fine-mapping of the chr5p15.33 Region 2 risk locus.** We performed imputation and fine-mapping of the multi-cancer risk locus in the \( \text{CLPTM1L} \) gene (Region 2, originally marked by rs401681 and rs402710) using GWAS data for four cancers previously shown to have associations with this locus, namely pancreatic11, testicular28 and lung cancer7, and melanoma9. For pancreatic cancer, fine-mapping identified SNPs with \( P \) values significantly lower than the previously published association signal marked by rs401681, with rs451360 being the smallest (\( P = 2.0 \times 10^{-10} \) for rs451360; \( P = 3.7 \times 10^{-7} \) for rs401681; Supplementary Table 1)18. This SNP is highly correlated with eight other SNPs (\( r^2 > 0.60, 1000G \) EUR population) that collectively mark Region 2 in pancreatic cancer (Fig. 1). Fine-mapping of Region 2 for testicular germ cell tumours (TGCT) and lung cancer revealed that the strongest SNP for each was among this group of nine SNPs (rs35953391 for TGCT, \( P = 1.08 \times 10^{-9} \) and rs37004 for lung cancer, \( P = 1.18 \times 10^{-13} \); Supplementary Table 1). Conditional analysis for the most significant SNP across each cancer resulted in a substantial loss of the signal for the other eight SNPs in pancreatic \( P_{\text{Conditional}} = 0.47–0.91 \), testicular \( P_{\text{Conditional}} = 0.21–0.92 \) and lung cancer \( P_{\text{Conditional}} = 0.90–0.45 \). In contrast, for melanoma none of the nine SNPs were significantly associated with risk in an unconditional analysis. However, upon conditioning on the most significant SNP in Region 2 (rs2447853, \( P = 5.7 \times 10^{-12} \) (ref. 29), all nine SNPs became more significantly associated with melanoma risk \( P_{\text{Conditional}} = 5.77 \times 10^{-5} \) to \( 4.45 \times 10^{-3} \) consistent with the possibility that these SNPs may mark one or more risk variants independent of rs2447853.

We also noted in the 1000G Phase 3, version 1 reference dataset an insertion/deletion variant that was highly correlated with these nine SNPs (rs3030832, \( r^2 = 0.96 \) to rs451360 in EUR) that had not been included in the imputation reference based off an earlier version (1000G Phase 1, version 3). We therefore re-imputed the pancreatic cancer GWAS with the newer 1000G reference set and observed an association signal similar in strength and significance to that of the other nine variants (rs3030832, \( P = 8.25 \times 10^{-10}, OR = 1.28 \) 95% CI 1.18–1.39; Supplementary Table 2) indicating that this indel variant should likewise be considered a candidate functional risk variant. Overall, these ten variants extend across the entire length of \( \text{CLPTM1L} \) from the promoter to ~6 kb downstream of the gene (Fig. 1). Three variants, rs36115365, rs380145 and rs27071, are located within potential gene regulatory regions, annotated by the ENCODE project (Fig. 1, Supplementary Fig. 1).

**Allele-specific regulatory effects mediated by rs36115365.** We sought to assess whether any of the ten highly correlated sequence variants influence differential protein binding via electrophoretic mobility shift assays (EMSA) in the PANC-1 and/or MIA PaCa-2 pancreatic cancer cell lines (Fig. 2, Supplementary Fig. 2), only rs36115365 exhibited allele-specific binding (Fig. 2), where the pancreatic cancer risk-associated minor C-allele (MAF 0.19 in 1000G EUR) displayed selective protein binding as indicated by greater loss of C-allele-specific banding upon addition of unlabelled C-allele competitor compared to unlabelled G-allele probe. EMSA assays for rs36115365 in seven additional cancer cell lines, including pancreatic cancer (MIA PaCa-2, Supplementary Fig. 3a), testicular germ cell cancer (NTERA-2 and 2102Ep, Supplementary Fig. 3b), lung cancer (A549, Fig. 2; NCI-H460, Supplementary Fig. 3c) and melanoma (UACC903 and UACC1113; Supplementary Fig. 3d) showed a similar pattern of allele-preferential binding to the C allele of rs36115365.

This SNP is located in-between the 5’ end of \( \text{TERT} \) (~18 kb upstream) and 3’ end of \( \text{CLPTM1L} \) (~5 kb downstream), a region that overlaps active histone modification marks and multiple transcription factor binding sites according to ENCODE data (Fig. 1, Supplementary Fig. 1). The region harbouring rs36115365 demonstrated an allele-specific increase in luciferase reporter activity as compared to empty vector that was consistent across all eight cancer cell lines tested (Fig. 3, Supplementary Fig. 4), including those from pancreas (PANC-1 and MIA PaCa-2, average fold change for C versus G allele 1.38, range 1.05–2.82), testis (NTERA-2, and 2102Ep, average fold change for C/G allele 1.95, range 1.12–4.83), lung (A549 and NCI-H460,
average fold change for C/G allele 1.33, range 1.05–1.95), and melanoma (UACC903 and UACC1113, average fold change for C/G allele 1.35, range 1.30–1.42). Transcriptional activity of the genomic region surrounding rs36115365 (240 bp) was higher in the forward (plasmids FG and FC) as compared to the reverse (plasmids RG and RC) orientation. Across all cancer cell lines, the C-allele on average showed an approximately 44% higher luciferase activity than the G-allele in the forward orientation, and 23% higher activity in the reverse orientation ($P = 4.2 \times 10^{-5} - 0.031$).

Analysis of imputed GWAS data from pancreatic and testicular cancers conditioned on rs36115365 are consistent with rs36115365 accounting for the majority of the Region 2 signal ($P_{\text{Conditional}} = 0.03–0.99$ and $P_{\text{Conditional}} = 0.22–0.92$, respectively for the grouping of eight SNPs highly correlated with rs36115365; Supplementary Table 1), with the minor C allele being positively associated with risk. In lung cancer and melanoma, however, fine-mapping data suggest that the genetic architecture underlying risk in Region 2 may be more complex, but are nonetheless consistent with a functional role for rs36115365. For lung, in contrast to pancreatic and testicular cancers, the C allele of rs36115365 is negatively associated with risk. Conditioning on rs36115365 revealed a possible secondary signal for lung cancer risk within the eight highly correlated SNPs ($P_{\text{Conditional}} = 3.74 \times 10^{-5} - 0.11$; Supplementary Table 1). For melanoma, rs36115365 was not significant in single-SNP analysis ($P = 0.70$), but became more significant after conditioning on the best Region 2 SNP (rs2447853, $P_{\text{Conditional}} = 1.09 \times 10^{-4}$; Supplementary Table 1), with the C allele also being negatively associated with risk (OR = 0.86; 95% CI 0.80–0.93). After conditioning on rs36115365 for melanoma, rs2447853 also becomes more significant ($P_{\text{Conditional}} = 3.01 \times 10^{-15}$ versus $P = 5.7 \times 10^{-15}$). These data suggest rs36115365 may influence gene expression within the TERT-CLPTM1L region and may account for either some or the entire association signal in this region, depending on the cancer type.
Silencing the region harbouring rs36115365. To interrogate whether the putative gene regulatory region harbouring rs36115365 influences expression of TERT and/or CLPTM1L, siRNA mediated transcriptional gene silencing (TGS) (refs 30,31) was used to target across this region to evaluate effects on gene expression. This mechanism of gene silencing is different from the well-known siRNA-mediated post-transcriptional gene silencing (PTGS) in that it targets a genomic regulatory region that mediates gene expression rather than messenger RNA (mRNA) (refs 30,31). Eight siRNAs were designed to span the region (Fig. 4a, Supplementary Table 3) and were separately transfected into cancer cell lines from pancreas (PANC-1), lung (A549), testis (NTERA-2), and melanoma (UACC903). Three of the eight (siRNA3, siRNA5 and siRNA8; Fig. 4b, Supplementary Fig. 5) showed significant inhibition of TERT mRNA expression by RT-qPCR in all four cell lines tested compared to a scrambled siRNA control, suggesting a role for the targeted region in the regulation of TERT expression. Inhibition of TERT by the three siRNAs ranged from 24 to 74% in PANC-1, 44 to 77% in A549, 33 to 49% in NTERA-2 and 54 to 84% in UACC903. The remaining five siRNAs showed little effect on expression of TERT. In contrast, expression of CLPTM1L as well as the GAPDH and ACTB housekeeping genes were not affected by any of the eight siRNAs. In addition, four siRNAs randomly designed to target non-genic regulatory regions on chromosome 8q24.21 were used as negative controls; none altered expression of TERT, CLPTM1L, or either housekeeping gene (Supplementary Fig. 6). The three siRNAs altered TERT expression in four additional cancer cell lines from pancreas (MIA PaCa-2), testis (2102Ep), lung (NCI-H460) and melanoma (UACC1113). Both siRNA3 and siRNA8 consistently reduced expression of TERT, but not CLPTM1L or housekeeping gene expression in all four lines, while siRNA5 resulted in specific down-regulation of TERT in some but not all lines (Supplementary Fig. 7). These data suggest that the genomic region harbouring rs36115365 plays a key role in the regulation of TERT, but not CLPTM1L, expression.

Allele-specific TERT gene-regulatory activity by rs36115365. We next sought to test whether the effect of TGS by siRNA targeting this putative gene-regulatory element on TERT expression was influenced by the genotype at rs36115365 by assessing allele-specific TERT mRNA expression. The human TERT gene harbours a synonymous SNP in exon 2 (rs2736098), linked to rs36115365 (1.2 fold for A549 cells). We evaluated allele-specific levels of inhibition of TERT expression by siRNA3 (which is both closest to rs36115365 and most consistently inhibits TERT expression across the cell lines previously tested) in these three cell lines using a TagMan allelic-discrimination assay for rs2736098. Inhibition by the siRNA on the C versus the G allele of rs36115365 was 60.2 versus 49.1% in Panc 05.04 cells (P = 0.007; t-test), 70.0 versus 63.6% in A549 cells (P = 0.003; t-test) and 28.3 versus 16.4% in IMIM-PC-1 cells (P = 0.002; t-test) (Fig. 4c).

These results indicate that rs36115365 lies in a gene-regulatory element that influences TERT expression in an allele-specific manner.

Zinc-finger transcription factor 148 binds rs36115365-C. To investigate the underlying mechanism of the differential gene regulation by genotypes at rs36115365, and to identify transcription factors potentially mediating this effect, we performed pull-down with oligonucleotides corresponding to the C or the G allele of rs36115365 incubated with nuclear extracts from PANC-1 and UACC903 cell lines, followed by quantitative mass spectrometry32. While most proteins identified bind both variants equally well, we noted outliers that bound the C-allele preferentially over the G-allele, as demonstrated by their location on the two-dimensional interaction plot (lower left quadrant of each, Fig. 5a and Supplementary Fig. 9), consistent with the EMSA data and suggesting preferential protein binding to this allele. Three proteins (ZNF148, VEZFI/ZNF161 and ZNF281) were identified as binding the C variant of rs36115365 preferentially in label-swapping experiments performed across both PANC-1 and UACC903 cell lines using a poly-dAdT competitor (Fig. 5a). A fourth protein, ZNF740, was also found to preferentially bind the C variant in both cell lines using mixed poly-dAdT and poly-dIdC competitors (Supplementary Fig. 9, bottom panels). We sought to verify whether any of these four proteins differentially bound the C-allele by using antibodies against these proteins in conjunction with EMSAs for rs36115365 (Fig. 5b, Supplementary Figs 10 and 11). Only the antibody against ZNF148 consistently resulted in loss of C allele-specific banding in pancreatic (PANC-1, Fig. 5b), as well as testis (NTERA-2) and lung cancer (A549) lines (Supplementary Fig. 10). Furthermore, EMSAs using recombinant purified ZNF148 protein demonstrated specific binding of ZNF148 to the C allele of rs36115365 (Fig. 5c). Notably, the resulting band had similar mobility characteristics to both those from EMSAs of ZNF148 bound to a known binding site in the CDKN1A/p21 promoter33,34, as well as the C allele-specific band for rs36115365 using PANC-1 nuclear extracts (Fig. 5c). Consistent with these data, ZNF148, (also named ZBP-89) a zinc-finger transcription factor 148 preferentially binds a nuclear protein. Electrophoretic mobility shift assays (EMSA) with biotin-labelled oligonucleotides containing either rs36115365-C or rs36115365-G in pancreatic (PANC-1) and lung (A549) cancer cell line nuclear extracts. Two specific protein complexes bind the C allele of rs36115365 preferentially in both cell lines and are more strongly competed with unlabeled C probe as compared to unlabeled G probe. Unlabelled competitor was used at ×10 and ×100 (as indicated by gradient symbol). Arrows denote specific protein complexes bound by the C allele of rs36115365.
factor of the kruppel-like family, is predicted to bind to a consensus DNA-recognition motif created by the C-allele of rs36115365 (Fig. 5d). To further establish the binding of ZNF148 to rs36115365 and surrounding genomic region, we performed chromatin-immunoprecipitation (ChIP) for ZNF148 followed by quantitative PCR, noting an enrichment of binding at rs36115365 in pancreatic and lung cancer cell lines homozygous and heterozygous for rs36115365-C as compared to background and the surrounding area (Fig. 5e, Supplementary Fig. 12a–f). We also assessed allelic enrichment in the immunoprecipitates and noted a significant enrichment of the C allele as compared to the G allele in A549 cells (1.51 fold, \( P = 0.01; \) t-test; Supplementary Fig. 12g), with Panc 05.04 cells showing a nonsignificant trend in the same direction (1.12 fold, \( P = 0.06; \) t-test; Supplementary Fig. 12g).

ZNF148 knockdown decreases TERT mRNA and telomerase activity. To determine the effect of ZNF148 depletion on expression of TERT and CLPTM1L in pancreatic, lung, testicular, and melanoma cell lines (\( n = 8 \) total), we used siRNA-mediated PTGS. We observed that while depletion of ZNF148 resulted in little change in expression of CLPTM1L, expression of TERT was significantly decreased in most of the cell lines, with an average expression of 0.50 relative to a scrambled siRNA control (range 0.27–0.89, \( P = 2.0 \times 10^{-4} \)–0.012; t-test; Fig. 6a, Supplementary Figs 13 and 14), consistent with a role for ZNF148 in regulating TERT expression. In contrast, siRNA-mediated knockdown of VEZF1 (ZNF161), ZNF281 and ZNF740 showed no effect on expression of either TERT or CLPTM1L (Supplementary Fig. 15).

We next sought to assess if ZNF148-mediated regulation of TERT expression was accompanied by effects on telomerase activity and telomere length. Knockdown of ZNF148 via PTGS resulted in reduced telomerase activity in A549 and MIA PaCa-2 cells (Fig. 6b), as well as in NTERA-2 and UACC903 cells (Supplementary Fig. 16). This reduction was similar to that observed via siRNA-mediated depletion of TERT itself, or by transcriptional gene silencing (TGS, siRNA3) to target the gene regulatory element encompassing rs36115365. To further assess the role of ZNF148 in regulating TERT expression and activity, we performed rescue experiments after depletion of endogenous ZNF148 using an siRNA targeting the 3'UTR of ZNF148. Overexpression of exogenous ZNF148 lacking the 3'UTR indeed rescued both TERT expression and telomerase activity in A549 and MIA PaCa-2 cells (Supplementary Fig. 17). Consistent with these data, depletion of either ZNF148 or TERT, or alternatively targeting the rs36115365 regulatory region in both A549 and MIA PaCa-2 cells all resulted in similar reductions of telomere length (Fig. 6c).

Discussion

A small genomic region on chr5p15.33, that harbours the TERT and CLPTM1L genes, has been reported to influence risk of
Figure 4 | siRNA targeting the region near rs36115365 reduces expression of TERT. (a) Eight double-stranded siRNAs were designed to target the potential gene-regulatory region encompassing rs36115365. (b) Three of eight double-stranded siRNAs targeting the rs36115365 locus can inhibit the gene expression of TERT but not CLPTM1L, ACTB, and GAPDH when transfected into PANC-1 (pancreatic cancer, upper left), A549 (lung cancer, upper right), NTERA-2 (testicular cancer, lower left) and UACC903 (melanoma, lower right). Expression values were normalized to a scrambled control siRNA. Experiments were performed in triplicate and repeated at least three times; error bars represent standard error of the mean (s.e.m.). (c) Allele-specific TERT expression after siRNA targeting (by siRNA3) of the regulatory region encompassing rs36115365. TERT expression was assayed using a quantitative allelic-discrimination TaqMan assay for a synonymous mRNA coding SNP in the TERT gene (rs2736098) that is genetically linked to rs36115365 (r² = 0.14, D' = 1.0). Two pancreatic cancer cell lines (Panc 05.04, left; and IMIM-PC-1, right) and one lung cancer line (A549, center) heterozygous for both SNPs were assayed. Expression of the allele of rs2736098 that is linked to the C-allele of rs36115365 (labelled 'C allele') was reduced to a greater extent than the allele linked to rs36115365-G. The degree of knockdown of each allele was normalized to that from a scrambled siRNA control. Experiments were performed in triplicate and repeated three times. Mean measures for three independent experiments are plotted; error bars represent s.e.m.
multiple cancers and may contain up to seven or more independent susceptibility loci\textsuperscript{16–18}. The complexity of this locus is highlighted by the fact that the same alleles confer susceptibility to some cancers while they are protective for others. One of these susceptibility loci termed Region 2, initially marked by rs401681 and rs402710 in CLPTM1L, was fine-mapped in a subset-based meta-analysis across multiple cancer types\textsuperscript{18} and is the focus of the current study. The ten variants that mark Region 2 span the whole length of CLPTM1L to \(17\) kb upstream of the transcriptional start site of TERT. Here, we identify rs36115365 as a functional SNP in this region and provide a plausible biological explanation underlying risk, featuring altered TERT, but not CLPTM1L, expression. Fine-mapping of Region 2 using GWAS data from pancreatic, lung and testicular cancer confirmed significant association with this small set of tightly linked SNPs (Fig. 1). Little signal remained within Region 2 after accounting

---

**Figure 5 | ZNF148 is an allele-specific interactor of rs36115365.** (a) DNA pulldowns were performed using PANC-1 (pancreatic cancer) and UACC903 (melanoma) cell line nuclear extracts with 41 bp biotin-tagged rs36115365 bait oligos. Replicate label-swapping experiments were performed with poly-dAdT competitor. Ratios indicate enrichment for protein binding to either the C- or G-allele. Significantly enriched ratios in both a forward and reverse label-swapping experiment were called as outliers (red: C-allele; blue: G-allele. Outlier cutoff is 1.5 IQRs). Identities of C-allele preferential binding proteins identified as outliers. Boxes in boxplots represent first to third quartiles and whiskers extend to furthest data point still within 1.5 IQRs of either quartile. (b) EMSA was performed with oligonucleotides containing the C or G allele at rs36115365 using nuclear extract from PANC-1 pancreatic cancer cells. The left side (lanes 1–10) shows preferred binding of two protein complexes to the C allele of rs36115365; the right side (lanes 11–24) competition of binding to the C allele by antibodies specific for three zinc finger proteins: VEZF1/ZNF161, ZNF148/ZBP-89 and ZNF281. The gradient symbol indicates 1 or 2 \(\mu\)g antibody used for competition. An antibody to ZNF148/ZBP-89 competes efficiently for binding of a protein to the rs36115365-C oligonucleotide. (c) EMSA showing binding of recombinant human ZNF148 to the C allele of rs36115365 (lanes 1–8). A ZNF148 binding site in the CDKN1A/p21 promoter (p21 Motif, M) was used as a positive control (lanes 9–12), as well as PANC-1 nuclear extract (lanes 13–15). Arrows denote proteins bound by the C allele of rs36115365. The gradient symbol indicates addition of 90, 360 or 630 ng of recombinant ZNF148 to the reaction. (d) rs36115365 alters a predicted ZNF148/CACCC-binding protein DNA recognition motif (Transfac motif M00721). (e) Chromatin immunoprecipitation (ChIP) using an antibody against ZNF148 (black bars) in the A549 lung cancer cell line revealed enrichment of ZNF148 binding at rs36115365 over background levels using a nonspecific IgG antibody (grey bars). The location of amplicons tested with qPCR relative to rs36115365 are shown underneath the graph (drawn to scale). PCRs were conducted in triplicate and repeated for each of two immunoprecipitations; error bars represent standard deviation (s.d.) for a representative experiment.
for rs36115365 or alternatively the respective most significant SNP in pancreatic and testicular cancer, consistent with the notion that one or more of these variants (and/or an as-of-yet unidentified variant tightly linked with these SNPs) is responsible for mediating cancer risk attributable to this locus. For lung cancer, residual signal was seen after conditional analysis on rs36115365 ($P_{\text{Conditional}} = 3.74 \times 10^{-5}$ to $7.71 \times 10^{-4}$), indicating that this SNP may not explain the entire Region 2 signal for lung cancer. For melanoma, a SNP (rs2447853) highly correlated to the original GWAS SNP reported for these cancers (rs401681, $r^2 = 0.97$) represents the most significant SNP in Region 2 (ref. 29). Although rs36115365 was non-significant in single-SNP analyses, it became more significant after conditioning on rs2447853 ($P_{\text{Conditional}} = 1.09 \times 10^{-4}$). The LD structure between these SNPs and conditional analyses suggest that in melanoma both may mark independent functional variants, with the signal at rs2447853 masking the association between rs36115365 and melanoma risk in single-SNP analysis.

In contrast with the other variants, preferred protein binding was seen on the minor (C) allele of rs36115365 across multiple cell lines representing all four cancer types. Luciferase reporter assays consistently showed differential gene regulatory activity between alleles across the cancer cell lines assayed. These data suggested that rs36115365 as a strong candidate for a functional multi-cancer risk variant but did not specifically implicate which gene(s) may be influenced by this SNP.

While a suite of tools is commonly used to interrogate potential gene-regulatory GWAS loci and link regulatory variants to a specific gene or genes, their application was challenging for this locus. Expression quantitative trait locus analysis proved problematic for $TERT$ given the relatively low expression of this gene in normal tissues. Likewise, the utility of chromosome conformation capture (3C) methods to establish a physical association between risk variants and specific target genes is greatly limited by the relatively short distances between rs36115365 and the $TERT$ promoter. To establish a relationship between this element and regulation of gene expression, we targeted the intergenic risk region using siRNAs. This method has previously been used to inhibit promoter function via small RNA duplexes by a process termed TGS (refs 30,31,37–41). We applied this methodology to our study of an intergenic GWAS susceptibility variant, and established a role for the regulatory element in driving $TERT$ (but not CLPTM1L) gene expression across multiple cancer types. These data suggest that the method may be of broader utility in the functional interrogation of GWAS loci.

Our results suggest that the binding of one or more proteins to the C-allele of rs36115365 is likely to play a key role in regulating $TERT$ expression. Through quantitative mass spectrometry, we identified preferential binding of zinc finger protein 148 (ZNF148, also named ZBP-89) to the C-allele of rs36115365 in multiple cancer cell lines, and ChIP data confirmed binding of ZNF148 over rs36115365. We observed a subtle but significant preference for ZNF148 binding to the C-allele in ChIP experiments using A549 lung cancer cells, with a non-significant trend in the same direction in Panc 05.04 pancreatic cancer cells. These subtle differences in transcription factor binding are consistent with the very small effects this locus confers on cancer risk over a person’s lifetime. Consistent with a central role for ZNF148 in regulating expression of $TERT$, siRNA-mediated gene knockdown of ZNF148 consistently resulted in reduced expression of $TERT$. Furthermore, both knockdown of ZNF148 as well as TGS of the gene regulatory element in which rs36115365 resides reduced telomerase activity and telomere length, to a degree similar to knockdown of $TERT$ itself. After depletion of ZNF148, this effect was rescued by exogenous ZNF148.

ZNF148 is a transcriptional regulator of the kruppel-like family that binds GC-rich DNA sequences in a variety of promoters to either activate or repress gene expression (reviewed by Zhang et al.42). Overexpression of ZNF148 promotes growth arrest and...
apoptosis in gastrointestinal cancer cell lines in vitro and suppresses adenoma formation in the ApcMin/+ mouse model in vivo57. This may be due, at least in part, to ZNF148 binding of p53 that prevents nuclear export and results in elevated levels of nuclear p53 (ref. 44). ZNF148 has also been shown to be important in regulating CDKN1A gene expression, binding a GC-rich element in the promoter of the gene and recruiting both ataxia-telangiectasia mutated kinase and histone acetyltransferase p300 into a complex that drives histone deacetylase inhibitor (HDACi) mediated induction of this gene53,34.

Our results indicate that ZNF148 may regulate TERT expression in pancreatic, testicular, lung, and melanoma tumour cell lines via a regulatory element that is disrupted by the G allele at rs36115365. As some of these cell lines have TERT promoter mutations (UACC903, UACC1103) whereas others do not (PANC-1, MIA PaCa-2, unpublished data), our results indicate that regulation by ZNF148 is important even in the presence of these presumably activating mutations.

In summary, our work has uncovered a likely causal variant in the TERT-CLPTM1L Region 2 susceptibility locus and identified ZNF148 as a potential effector of a gene-regulatory element that mediates increased TERT expression in an allele-specific manner. Furthermore, our fine-mapping results highlight the complexity of this region and indicate that Region 2 may, in some cancers, consist of more than one underlying functional signal. Our results are remarkably consistent in eight cell lines across four different cancer types and explain, at least in part, the biological underpinnings of risk for rs36115365. Notably, our data suggest that the mechanism by which ZNF148 influences TERT is similar for cancer types in which the C-allele of rs36115365 contributes to increased risk, or alternatively to disease protection. Although TERT expression and ensuing effects on telomere length may be the crucial underlying mechanism in mediating inverse risk for different cancers, studies of surrogate tissue telomere length and cancer risk have been contradictory and shown associations with short or long telomeres, or no effect48–53. TERT could also mediate risk through its telomere-independent functions that include transcriptional regulation and mitochondrial RNA polymerase activity (for review see Martinez et al.54). Other factors may contribute to the pleiotropic effects observed for rs36115365, including environmental exposures, regulatory effects through genes beyond TERT, interaction with additional risk variants and/or somatic mutations both within Region 2 and the larger TERT/CLPTM1L locus, or tissue-specific regulation of ZNF148 and other transcription factors mediating TERT expression. Our findings represent the first steps in unravelling the complex functional consequences of carrying risk variants in Region 2 of chr5p15.33 and strongly indicate a major role for expression of TERT in influencing risk of multiple cancer types.

**Methods**

**Studies.** Subjects were drawn from GWAS studies of four cancers: pancreatic cancer: PanScan 1 and II (3,525 cases and 3,642 control subjects; dbGaP Study Accession: phs000206.v5.p3) (refs 11,18); testicular germ cell tumours: NCI (581 cases and 1055 control subjects) and PENN (477 cases) (ref. 28) and PLCO control subjects (178 control subjects)55; lung cancer from the Transdisciplinary Research (581 cases and 1055 control subjects) and PENN (477 cases) (ref. 28) and PLCO studies were reviewed and approved by institutional ethics review committees at the involved institutions. Participation of subjects in the PanScan GWAS was also reviewed and approved by institutional ethics review committees at the involved institutions. Analysis of the testicular germ cell tumour GWAS was reviewed by the NCI Special Studies Institutional Review Board and the University of Pennsylvania IRB #4.

**Fine-mapping.** Imputation across 2 Mb of chr5p15.33 (250,000 to 2,250,000 bps, hg19) was performed using phased haplotypes from the 1000G reference set (Phase 1 integrated release 3, March 2012) and IMPUTE2 for pancreatic cell tumours55 and testicular germ cell tumours34. Imputed SNPs with low MAF (<0.01) or low-quality scores (IMPUT2 information score <0.5) were removed before the association analysis. Association analysis between SNPs and case control status were performed using the score test of the log additive genetic effect with covariate adjustment using SNPRTEST as previously described25. Imputation and association analysis for melanoma was performed using 1000GC (Phase 1 integrated release 3, March 2012) as previously described29. Imputation for lung cancer7–56 was performed by using 1000GC (Phase 1 integrated release 3, March 2012) with the same quality thresholds as described, followed by association analysis and conditional analysis using summary statistics from a meta-analysis of the six studies of TERC with GCT.

Overall, Region 2 was well-imputed. Within the pancreatic cancer GWAS data, all common 1000G variants (n = 195, MAF ≥0.01) in Region 2 (defined as the genomic region between the two recombination hotspots at 1,306,281–1,367,281 in NCBI build Hg19) had imputation accuracy (INFO) scores above 0.3 (the lowest INFO score was 0.28). The imputation quality for the nine SNPs most significantly associated with pancreatic cancer risk was high in the PanScan GWAS studies, with quality scores (INFO) ranging from 0.82 to 0.96 (average 0.92). Similar imputation quality scores were observed for these SNPs in the lung cancer, TGCT, and melanoma GWAS (INFO range 0.82 to 0.98; average 0.94). In melanoma, association quality was high for all SNPs that were both statistically significant with rs36115365 in 1000 Genomes CEU data (r² ≥0.2). In PanScan, only a single such 1,000 Genomes variant had an imputation quality score (INFO) below 0.8 (rs186156459; INFO = 0.79), suggesting that poor imputation quality did not lead to the exclusion of additional strong functional candidates from consideration. Similar imputation quality was observed for these SNPs in the lung cancer, TGCT, and melanoma GWAS.

For completeness we assessed the newer 1000G (Phase 3, October 2014) reference dataset and noted an insertion/deletion variant (rs3030832) that was highly correlated to rs36115365 (r² = 0.87 in EUR). We therefore re-imputed the pancreatic cancer GWAS dataset34 with the newer 1000G reference set to re-assess the association signal across Region 2 (defined as the genomic region between the two recombination hotspots at 1,306,281–1,367,281 in NCBI build Hg19) including this variant. rs36115365 became non-significant when analysis was conditioned on rs3030832, as was rs3030832, when analysis was conditioned on rs36115365 (Supplementary Table 2), indicating that this variant is among the highly correlated variants representing Region 2 and thus represents an additional strong functional candidate. We also observed seven additional variants with similar or slightly higher ORs as compared to rs36115365 (ORmax = 1.42). To formally test if these seven variants represented potential functional variants in Region 2 we performed a series of conditional analyses. After the analysis was performed on rs36115365 we observed a large drop in risk for this variant while conditional analysis for each of the seven variants did not dramatically influence the significance or rs36115365 (Supplementary Table 2).

**Cell lines.** The human pancreatic cancer cell lines Panc-1 and Mia PaCa-2, and lung cancer cell lines A549 and NCI-H460 (purchased from ATCC) were maintained in Dulbecco’s modified Eagle’s medium (DMEM, Mediatech Inc, Herndon, VA) or RPMI 1640 (Mediatech Inc) supplemented with 10% fetal bovine serum (FBS, IFS Technologies); 10% FBS and 2.5% horse serum in RPMI 1640 supplemented with 10% FBS and 2.5% horse serum (UACC1113, UACC1113, UACC903, IMIM-PC-1 and IMIM-PC-2 did not have profiles listed). The 2012Ep profile matched a previously published profile59. The cells were routinely tested for Mycoplasma and
were negative on each occasion. None of the cell lines used are on the NCI or ICLAC lists of misidentified cells.

RNA and genomic DNA isolation. RNA was extracted using an RNeasy Plus Mini Kit (Qiagen). Quality and quantity of RNA was assessed in an Agilent 2100 Bioanalyzer (Agilent Technologies); only samples with RIN scores > 9 were used. Genomic DNA was isolated using the ZR genomic DNA kit (D3050, ZYMO Research) and assessed by Nanodrop 8000 (Thermo Scientific).

EMSA and ChIP. Nuclear extracts were purchased from Active Motif (PANC-1, MIA PaCa-2) or alternatively generated using a Nuclear Extraction Kit (A549, NCi-H460, UACC1113, UACC903, NTERA-2 and 2102Ep) (10009277, Cayman) according to the manufacturer’s instructions. Recombinant human ZNF148 protein was purchased from OriGene (TP602693, OriGene). Oligos (30-36 nt, Invitrogen, listed in Supplementary Table 3) were labelled with a biotin 3’-end-labelling kit (NEB). Forward and reverse oligos were then annealed to create double stranded 3’-end labelled or unlabelled probes. EMSAs were performed (Pierce) by incubating labelled probes for 20 min on ice with nuclear extracts (10 μg per reaction). Competition experiments were performed by adding 10–100 fold more unlabelled competitor than labelled probes. Supershift experiments were carried out by mixing 1 and 2 μg anti-ZNF148/ZBP-89 (sc-48811X, 200 μg per 0.1 ml, Santa Cruz), anti-ZNF148/ZBP-89 (sc-9878X, 200 μg per 0.1 ml, Santa Cruz), anti-ZNF281 (sc-16693X, 200 μg per 0.1 ml, Santa Cruz), anti-ZNF740 (sc-32477, 200 μg per 0.1 ml, Santa Cruz) or anti-Ilg (sc-52001, 100 μg/ml 1, Santa Cruz) antibodies with nuclear extracts for 20 min at room temperature before adding labelled probe. Recombinant human ZNF148 protein was added at 90, 360 and 630 ng to the reaction, separately. The resulting protein complexes were resolved on 4–20% TBE gels (Bio-Rad), transferred to Biodyne B membranes (VWR) and probed with ZNF148 (Stratagene) antibodies or human ZNF148 [ZBP-89, 200 μg per 0.1 ml, Invitrogen], and detected using streptavidin-HRP conjugate and a chemiluminescent substrate (20148 LighShift Kit; Pierce).

Chromatin-immunoprecipitation (ChIP) was performed with the ChIP-IT high Sensitivity Kit (active Motif) according to the manufacturer’s protocol using cells (~2 × 10^7) from each cell line of the following genotypes at rs36115365: CC (COLO 357 and IMIM-PC2, CG (Panc 05.04 and A549) and GG (Mia PaCa-2). An assembly of primers were used for RT-PCR (Life Technologies). Expression of target genes was determined on 12–24 h cultures using label-swapping, and separately conducted using poly-dAdT competitor only, giving rise to a Fisher’s Exact test P value of 0.011, indicating that the inhibition of TERT is specific.

Real-time quantitative PCR. Gene expression levels were quantified by quantitative real-time PCR using TaqMan assays for TERT (Hs09792656_m1), CLPTM1L (Hs0363947_m1), ACTB (Hs00607939_m1), ZNF148 (Hs00107570_m1), and GAPDH (Hs00468149_m1) from Life Technologies. Expression of TERT, CLPTM1L and ACTB were normalized to GAPDH, while expression of ZNF148 was normalized to ACTB. Allele-specific TERT expression was determined using an allelic discrimination TaqMan assay for rs2376989 (assay C_G2614916_20, Life Technologies), and the gene expression of each allele of TERT was also normalized to the gene expression of GAPDH. Each experiment was performed in triplicate and repeated at least three times. Significance was assessed using a Student’s two-tailed T-test (labelled significant if P < 0.01).

Luciferase cloning and expression analysis. The genomic region containing and surrounding rs36115365 (240 bp) was PCR-amplified (primers listed in Supplementary Table 5) from HapMap CEU DNA samples with the appropriate genotypes to obtain clones with each genotype, and cloned into the pGL4.23[luc2/minP] (Promega) vector via the 5’→3’ orientation. Plasmid inserts were sequenced-verified to confirm the correct inserts and genotypes. The forward (F) orientation of the insert is the same as the genomic orientation. The Firefly reporter plasmids (and a Renilla luciferase control vector) were co-transfected into pancreatic (PANC-1, MIA PaCa-2), melanoma (UACC203, UACC1113), lung (A549, NCi-H460) and testicular (NTERA-2, 2102Ep) cancer cell lines at ~70% confluence using Lipofectamine 2000 (Life Technologies). Luciferase activity was measured 36 h after transfection with the Dual Luciferase Reporter Assay System (Promega). Firefly luciferase activity was normalized to Renilla luciferase activity, and graphed as compared to the empty luciferase vector. Experiments were performed in triplicate and repeated at least three times. A T-test was used to assess significance for differences in luciferase activity.

Region targeted siRNAs to rs36115365 regulatory locus. On-target antisense enhanced siRNAs targeting the locus encompassing rs36115365 were designed by using an siRNA design tool (https://dharmacon.goldenbeardsciences.com/) and ordered from Dharmacon RNAi and Gene Expression in GE Healthcare and listed in Supplementary Table 6. No siRNAs were designable to directly overlap with rs36115365; the location of the nearest siRNA (sirNAS) was 8 bp from this variant. The siRNAs were introduced to cell lines by using RNAiMAX (Life Technologies) at a final concentration of 15 nM. RNA was extracted 48 h after transfection and reverse transcribed to cDNA by the SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies). Expression of target genes was determined on the cDNA by RT-qPCR TaqMan gene expression assays as described below. Experiments were performed in triplicate and repeated at least three times. We first tested 8 siRNAs in 4 cell lines. Three of the 8 siRNAs inhibited TERT expression in all four cell lines (PANC-1, A549, NTERA-2 and UACC903) whereas none of the 8 siRNAs inhibited CLPTM1L, ACTB or GAPDH expression. Thus, the inhibition of TERT expression by 3 of 8 out of 8 versus 0 out of the other 3 genes gives rise to a Fisher’s Exact test P value of 0.011, indicating that the inhibition of TERT is specific.

siRNA-mediated knockdown of ZNF148 and TERT mRNA. ON-TARGETplus Human SMARTpool siRNAs to ZNF148 (cat# L-012658-00-005), VEZFI (ZNF161; cat# L-019623-00-005), ZNF281 (cat# L-009585-00-005), ZNF740 (cat# L-003075-00-005), and TERT (cat# L-003547-00-005) were purchased from Dharmacon RNAi and Gene Expression in GE Healthcare. To assess possible off-target effects for the ZNF148 siRNAs we also purchased each of the four siRNAs from the SMARTpool separately and tested their effects on ZNF148 and TERT expression. All four siRNAs inhibited both ZNF148 and TERT expression indicating that off-target effects are not likely to explain our findings (Supplementary Fig. 14). Transfection, RNA purification, cDNA generation and expression analysis procedures were as described above for the region-targeted siRNA assay, except that RNA was isolated 72 h after transfection. Experiments were performed in triplicate and repeated at least three times.

Telomerase activity and telomere length. The telomeric repeat amplification protocol (TRAP) (ref. 20) was used to evaluate telomerase activity according to the manufacturer’s guidelines (Millipore, #77090), siRNAs targeting the gene regulatory region (sirNAS), ZNF148 (cat# L-012658-00-005), TERT (cat# L-003547-00-005) and a scrambled control siRNA (sequence listed in Supplementary Table 6) were introduced to MIA PaCa-2, UACC903 and NTERA-2 cells at a final concentration of 15 μM for 72 h. At that time the cells were harvested and whole cell extract prepared using CHAPS (3-[3-cholamidopropyl]-dimethylammoniomethyl)-1-propanesulfonate solution). The Bradford assay kit (Bio-Rad) was used to determine total protein concentration. Equal amounts of protein extracts were used to add telomeric repeats (CTTAGG onto 3’ end of substrate oligonucleotide (TS) at 37 °C for 30 min followed by 30 cycles of TRAP PCR and separation of PCR products on 12.5% non-denatured PAGE gels. The gels were stained with SYBR Gold Nuclear Acid Gel Stain (Life Technologies, #S-11494).

For rescue experiments, we attempted to create cell lines devoid of ZNF148 using CRISPR. Extensive screening of clones revealed none with homozygous loss of ZNF148, consistent with an essential role for ZNF148, further supported by the observed embryonic lethality of homozygous ZNF148 knock-out mice (International Mouse Phenotyping Consortium, IMPC Data Coordination Centre, MRC Hartwell Institute, Biocomputing, Harwell Campus, https://www.mousephenotype.org/data/charts?accession=MGI:1332324&allele_accession_id=MGI:56366955&bgosity=homogyzote &parameter_stable_id=IMPC_VIA_001_001&bgppelene_stable_id=BCM_001). Rescue experiments were instead performed by depletion of endogenous ZNF148 expression using an siRNA targeting the 3’→5’ UTR of ZNF148 from the 3’→5’ end targeting the ORF region (ZNF148 sense: AGUGAAGAUGUGAGCAUU; antisense: AGUGUCAACAGUGUCUC AU) and reintroduction of exogenous ZNF148 expression. Human ZNF148 ORF (Origene TrueORF RC222687) was cloned into the P-Pst6-639 (derived from pcPlUGW) lentiviral expression vector and sequence verified. For lentivirus production, lentivector co-transduced into HEK293FT cells with packaging vectors psPAX2, pMD2-G and pCA-GTR2. Virus was collected two days after transfection and concentrated by Vivaspin, before infecting MIA PaCa-2
and A549 cells. Seventy-two hours after delivery of the siRNA, ZNF148 and TERT expression, and telomerase activity, were assessed as described above.

To assess effects on telomere length, siRNA3, ZNF148 siRNA (cat# L-012658-00-0005), TERT siRNA (cat# L-003547-00-0005) and a scrambled siRNA (Supplementary Table 6) were administered to MIA PaCa-2 and A549 cells at a final concentration of 50 nM. The cells were re-transfected with siRNAs every four days and genomic DNA extracted 20 days later using the DNAeasy Blood and Tissue Kit (Qiagen, #69506). Telomere length was then determined by qPCR by comparing telomere repeat sequence copy number to a single-copy gene (RPLP0) copy number in a given sample using telomere repeat-specific primers as previously described25. The assays were performed in triplicate and repeated three times.

Data availability. The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding authors upon reasonable request. Pancreatic cancer GWAS data is available from dbGaP (phs000206.v5.p3). Testicular germ cell tumour fine-mapping data are available from the corresponding authors upon reasonable request. The lung cancer fine-mapping data that supports the findings of this study are available from the corresponding authors upon reasonable request and are currently being processed for availability through dbGaP. The melanoma fine-mapping data that support the findings of this study are available within the article and its supplementary information files or from the corresponding authors upon request.

References

1. Rothman, N. et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nat. Genet. 42, 978–984 (2010).
2. Haiman, C. A. et al. A common variant at the TERT-CLPTM1 locus is associated with estrogen receptor-negative breast cancer. Nat. Genet. 43, 1210–1214 (2011).
3. Shete, S. et al. Genome-wide association study identifies five susceptibility loci for glioma. Nat. Genet. 41, 899–904 (2009).
4. Wang, Y. et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat. Genet. 40, 1407–1409 (2008).
5. McKay, J. D. et al. Lung cancer susceptibility locus at 5p15.33. Nat. Genet. 40, 1404–1406 (2008).
6. Broderick, P. et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer. Res. 69, 6633–6641 (2009).
7. Landi, M. T. et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am. J. Hum. Genet. 85, 679–691 (2009).
8. Rafnar, T. et al. Sequence variants at the TERT-CLPTM1 locus associate with many cancer types. Nat. Genet. 41, 221–227 (2009).
9. Stacey, S. N. et al. New common variants affecting susceptibility to basal cell carcinoma. Nat. Genet. 41, 909–914 (2009).
10. Yang, X., Yang, B., Li, B. & Liu, Y. Association between TERT-CLPTM1 rs401681[C] allele and NMSC cancer risk: a meta-analysis including 45,184 cases. PLoS ONE 7, e36116 (2012).
11. Petersen, G. M. et al. Fine-mapping data are available from the corresponding authors upon reasonable request. The lung cancer fine-mapping data that supports the findings of this study are available within the article and its supplementary information files or from the corresponding authors upon reasonable request.
12. Bojesen, S. E. et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15.33 locus. PLoS ONE 7, e36116 (2012).
13. Turnbull, C. et al. Specific association of human telomerase activity with immortal cells and cancer. Science 266, 2011–2015 (1994).
14. Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. Nat. Rev. Genet. 11, 319–330 (2010).
15. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. Eur. J. Cancer. 33, 787–791 (1997).
16. Yamamoto, K., Okamoto, A., Isonishi, S., Ochiai, K. & Ohtake, Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochem. Biophys. Res. Commun. 280, 1148–1154 (2001).
17. James, M. A. et al. Functional characterization of CLPTM1L as a lung cancer risk candidate gene in the 5p15.33 locus. PLoS ONE 7, e36116 (2012).
18. Zuo, Z. et al. CLPTM1L is overexpressed in lung cancer and associated with apoptosis. Cancer. Res. 74, 2785–2795 (2014).
19. James, M. A., Vikis, H. G., Tate, E., Rynasiewicz, A. L. & You, M. CRR9/CLPTM1L regulates cell survival signaling and is required for Ras transformation and lung tumorigenesis. Cancer. Res. 74, 1116–1127 (2014).
20. Chung, C. C. et al. Meta-analysis identifies four new loci associated with testicular germ cell tumor. Nat. Genet. 45, 680–685 (2013).
21. Barrett, J. H. et al. Fine mapping of genetic susceptibility loci for melanoma reveals a mixture of single variant and multiple variant regions. Int. J. Cancer. 135, 1350–1355 (2015).
22. Carey, D. R. Regulating mammalian transcription with RNA. Trends. Biochem. Sci. 30, 655–658 (2005).
23. Malecova, B. & Morris, K. V. Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs. Curr. Opin. Mol. Ther. 12, 214–222 (2010).
24. Butter, F. et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet. 8, e1002982 (2012).
25. Bai, L. & Merchant, J. L. ATM phosphorylates ZBP-89 at Ser202 to potentiate p21wa1 induction by butyrate. Biochem. Biophys. Res. Commun. 359, 817–821 (2007).
26. Bai, L., Kao, Y., Law, D. J. & Merchant, J. L. Recruitment of ataxiatelangiectasia mutated to the p21(waf1) promoter by ZBP-89 plays a role in mucosal protection. Gastroenterology 131, 841–852 (2006).
27. Remington, M. C., Tarle, S. A., Simon, B. & Merchant, J. L. ZBP-89, a Krüppel-type zinc finger protein, inhibits cell proliferation. Biochem. Biophys. Res. Commun. 237, 230–234 (1997).
28. Freedman, M. L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat. Genet. 43, 513–518 (2011).
29. Janovská, B. A. et al. Activating gene expression in human cancer cells with promoter-targeted duplex RNAs. Nat. Chem. Biol. 3, 166–173 (2007).
30. Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. Science 305, 1289–1292 (2004).
31. Napoli, S., Pastori, C., Magistri, M., Carbone, G. M. & Catapano, C. V. Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. EMBO J. 28, 1708–1719 (2009).
32. Ting, A. H., Schuelke, K. E., Herman, J. G. & Baylin, S. B. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. Nat. Genet. 37, 906–910 (2005).
33. Xu, Z., Wei, G., Chepeln, I., Zhao, K. & Felsenfeld, G. Mapping of INS promoter interactions reveals its role in long-range regulation of SYT8 transcription. Nat. Struct. Mol. Biol. 18, 372–378 (2011).
34. Zhang, C. Z., Chen, G. G. & Lai, P. B. Transcription factor ZBP-89 plays a role in transformation and lung tumorigenesis. Cancer Res. 74, 2520–2528 (2010).
35. Yang, X., Yang, B., Li, B. & Liu, Y. Association between TERT-CLPTM1L rs401681[C] allele and NMSC cancer risk: a meta-analysis including 45,184 cases. PLoS ONE 7, e36116 (2012).
36. Weischer, M. et al. A prospective analysis of telomere length and pancreatic cancer. Cancer Epidemiol. Biomark. Prev.: Publ. Am. Assoc. Cancer Res. 20, 1238–1250 (2011).
37. Campa, D. et al. Leukocyte telomere length in relation to pancreatic cancer risk: a prospective study. Cancer Epidemiol. Biomark. Prev.: Publ. Am. Assoc. Cancer Res. 23, 2447–2454 (2014).
38. Skinner, H. G. et al. Telomere length and pancreatic cancer: a case-control study. Cancer Epidemiol. Biomark. Prev.: Publ. Am. Assoc. Cancer Res. 21, 2095–2100 (2012).
39. Lynch, S. M. et al. A prospective analysis of telomere length and pancreatic cancer risk: the alpha-toxin of beta-toxin cancer (ATBC) prevention study. Int. J. Cancer. 133, 2672–2680 (2013).
40. Nan, H. et al. Shorter telomeres associated with a reduced risk of melanoma development. Cancer. Res. 71, 6758–6763 (2011).
Acknowledgements

We acknowledge the contribution of the staff of the Cancer Genomics Research Laboratory (CGR, NCI), as well as Mr Timothy Myers and Lea Jessop (LTG, DCEG, NCI) for their invaluable help throughout the project. We are grateful to Dr Francisco X. Real, The Spanish National Cancer Institute (CNIO) in Madrid, Spain for providing reagents. We are grateful to Dr Dom Esposito, Frederick National Laboratory for Cancer Research, Frederick, MD, USA, for cloning for lentiviral production. We thank Paul D.P. Pharoah, Department of Oncology, University of Cambridge, Cambridge, UK and Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Cambridge, UK for providing GWAS data. Furthermore, we thank Dr Stephen Canchik for his advice, encouragement and critical reading of the manuscript. This work was supported by the Intramural Research Program of the US National Institutes of Health (NHI), National Cancer Institute, by contract number HHSN26120080001E, by Cancer Research UK for SEARCH (C490/A10124). Funding for M.V. was provided by the Netherlands Organization for Scientific Research (NWO Gravitation Program Cancer Genomics Netherlands). Funding for M.M. was provided by a grant from the Marie Curie Initial Training Network (ITN) DevCom (FP7, grant number 607142). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Author contributions

K.M.B and L.T.A. conceived and directed the project, with contributions from M.V. and T.A. EMSAs, suprachrom experiments and reporter assays were performed by J.J., I.C., A.T., M.D., I.W.H., J.T. and Q.L. Transcriptional gene silencing, gene knockdown, telomerase activity and telomere length assays were performed by J.F. ChIP experiments were performed by J.J. and M.X. Quantitative mass-spectrometry experiments were led by M.V. and performed by M.M., P.W.T.C.J.; additional proteomics contributions were made by T.A., M.O. and S.D. Screening and characterization of cell lines used in this work was performed by J.J., I.C., J.F., R.K., J.C. and M.K. Fine-mapping and associated analyses of were performed by L.T.A., K.M.B., M.Z., Z.W., M.M.I., M.T.L., C.I.A., G.M.P. R.S.S., P.K., B.M.W., S.O., K.M., P.A.K., N.C., J.H.B., A.M.D., J.C.T., J.A.N.B., D.T.B., K.N., as well as the PanScan, TRICE and GenoMEL Consortia. K.M.B and L.T.A wrote the manuscript, with assistance from J.F. and I.J.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Fang, I. et al. Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. Nat. Commun. 8, 15034 doi: 10.1038/ncomms15034 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
TRICL Consortium

Victoria Stevens77, Demetrios Albanes78, Neil E. Caporaso78, Paul Brennan79, James McKay79, Xifeng Wu80, Rayjean J. Hung81, John R. McLaughlin82, Heike Bickeboller83, Angela Risch84,85,86, Erich Wichmann87 & Richard S. Houlston88

GenoMEL Consortium

Graham Mann89,90, John Hopper91, Joanne Aitken92, Bruce Armstrong93, Graham Giles94, Elizabeth Holland90, Richard Keffer90,95, Anne Cust89,93, Mark Jenkins91, Helen Schmid90, Susana Puig96, Paula Aguilera97, Celia Badenas98, Alicia Barreiro96, Cristina Carrera96, Daniel Gabriel96, Pol Gimenez-Xavier96, Pablo Iglesias-Garcia96, Josep Malvehy99, Montse Mila98, Ramon Pigem99, Miriam Potrony96, Joan Anton Puig-Butillé98, Gemma Tell-Martí96, Nicholas K. Hayward100, Nicholas G. Martin100.

77American Cancer Society, Inc., Atlanta, Georgia 30303, USA; 78Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 79International Agency for Research on Cancer, World Health Organization, 69372, Lyon CEDEX 08, France; 80Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA; 81Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, University of Toronto, Toronto, Canada M5G 1X5; 82University of Toronto, Dalla Lana School of Public Health, Toronto, Ontario, Canada M5S 1A1; 83Department of Genetic Epidemiology, University Medical Center, Georg-August-University, Göttingen 37073, Germany; 84Division of Epigenomics and Cancer Risk Factors, DKFZ-German Cancer Research Center, Heidelberg 69121, Germany; 85Division of Cancer Epidemiology, DKFZ-German Cancer Research Center, Heidelberg 69121, Germany; 86Division of Cancer Research and Epigenetics, Department of Molecular Biology, University of Salzburg, Salzburg 5020, Austria; 87Helmholtz Zentrum München, Institut für Epidemiologie I, Neuherberg D-85764, Germany; 88Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; 89Division of Molecular Pathology, The Institute of Cancer Research, Surrey SM2 5NG, UK.
Sud University, Gustave Roussy, Villejuif 94801, France; 140 Département de Biologie des tumeurs–Service Génétique, Institut Curie, Paris 75248, France; 141 Dermatology Department, CHU de Poitiers, Poitiers 86021, France; 142 Laboratoire Inflammation, Tissus Epithéliaux et Cytokines, EA 4331 Université de Poitiers, Pôle Biologie Santé, Poitiers 86073, France; 143 Pathology & Laboratory Medicine, Anatomic Pathology Division, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA; 144 Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA; 145 Department of Genetics and Patomorphology, Pomeranian Medical University, Szczecin 70-115, Poland; 146 Department of Surgical Oncology, Institute of Oncology, Ljubljana 1000, Slovenia; 147 Department of Molecular Diagnostics, Institute of Oncology, Ljubljana 1000, Slovenia; 148 Department of Oncology-Pathology, Karolinska Institutet, Karolinska University Hospital, Stockholm 171 76, Sweden; 149 The Susanne Levy Gertner Oncogenetics Unit, The Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, Tel-Hashomer 52621, Israel; 150 Department of Internal Medicine and Department of Human Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Ramat Aviv 69978, Israel; 151 Department of Dermatology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Ramat Aviv 69978, Israel; 152 School of Public Health, Faculty of Welfare and Health Studies, Haifa University, Mount Carmel, Haifa 31905, Israel; 153 Department of Dermatology, Chaim Sheba Medical Center- Tel Hashomer, Ramat Gan 52621, Israel; 154 Hebrew University-Hadassah Braun School of Public Health and Community Medicine, Ein Kerem, Jerusalem 91120, Israel; 155 The Susanne Levy Gertner Oncogenetics Unit, Institute of Human Genetics, Chaim Sheba Medical Center, Tel Hashomer, Ramat Gan 52621, Israel; 156 Section of Epidemiology and Biostatistics, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds LS9 7TF, UK; 157 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.
Publisher correction: Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148

Jun Fang, Jinping Jia, Matthew Makowski, Mai Xu, Zhaoming Wang, Tongwu Zhang, Jason W. Hoskins, Jiyeon Choi, Younhun Han, Mingfeng Zhang, Janelle Thomas, Michael Kovacs, Irene Collins, Marta Dzyadyk, Abbey Thompson, Maura O’Neill, Sudipto Das, Qi Lan, Roelof Koster, PanScan Consortium, TRICL Consortium, GenoMEL Consortium, Rachael S. Stolzenberg-Solomon, Peter Kraft, Brian M. Wolpin, Pascal W.T.C. Jansen, Sara Olson, Katherine A. McGlynn, Peter A. Kanetsky, Nilanjan Chatterjee, Jennifer H. Barrett, Alison M. Dunning, John C. Taylor, Julia A. Newton-Bishop, D. Timothy Bishop, Thorkell Andresson, Gloria M. Petersen, Christopher I. Amos, Mark M. Iles, Katherine L. Nathanson, Maria Teresa Landi, Michiel Vermeulen, Kevin M. Brown & Laufey T. Amundadottir

Nature Communications 8:15034 doi: 10.1038/ncomms15034 (2017); Published 27 Apr 2017; Updated 29 Mar 2018

The original version of this Article contained an error in the spelling of two members of the GenoMEL Consortium, Joan Anton Puig-Butille and Pol Gimenez-Xavier, which were incorrectly given as Joan-Anton Puig Butille and Pol Gimenez Xavier respectively.

Additionally, the original HTML version of this Article contained errors in the spelling of the names of the GenoMEL Consortium members Nicholas K. Hayward, Nicholas G. Martin, David L. Duffy, David C. Whiteman, Stuart MacGregor, Giovanna Bianchi-Scarra, Rona Mackie, Håkan Olsson, Göran Jönsson, Åke Borg, Olivier Dereure, David E. Elder, and Veronica Höiøm, which were incorrectly given as Nick Hayward, Nicholas Martin, David Duffy, David Whiteman, Stuart Mac Gregor, Giovanna Bianchi Scarra, Rona Mackie, Hakan Olsson, Goran Jonsson, Ake Borg, Olivier Dereure, David Elder and Veronica Hoiom.

Furthermore, the original PDF version of the Article also contained an error in the publication date, which was incorrectly given as 2nd May 2017.

These errors have now been corrected in both the PDF and HTML versions of the Article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2018