nature genetics

Systematic analysis of telomere length and somatic alterations in 31 cancer types

Floris P Barthel[1–3], Wei Wei[4], Ming Tang[3], Emmanuel Martinez-Ledesma[3,5], Xin Hu[3,6], Samirkumar B Amin[3,7], Kadir C Akdemir[3], Sahil Seth[8], Xingzhi Song[8], Qianghu Wang[3,9], Tara Lichtenberg[10], Jian Hu[11], Jianhua Zhang[8], Siyuan Zheng[3,5] & Roel G W Verhaak[1,3,9]

Cancer cells survive cellular crisis through telomere maintenance mechanisms. We report telomere lengths in 18,430 samples, including tumors and non-neoplastic samples, across 31 cancer types. Telomeres were shorter in tumors than in normal tissues and longer in sarcomas and gliomas than in other cancers. Among 6,835 cancers, 73% expressed telomerase reverse transcriptase (TERT), which was associated with *TERT* point mutations, rearrangements, DNA amplifications and transcript fusions and predictive of telomerase activity. *TERT* promoter methylation provided an additional deregulatory *TERT* expression mechanism. Five percent of cases, characterized by undetectable *TERT* expression and alterations in *ATRX* or *DAXX*, demonstrated elongated telomeres and increased telomeric repeat–containing RNA (TERRA). The remaining 22% of tumors neither expressed *TERT* nor harbored alterations in *ATRX* or *DAXX*. In this group, telomere length positively correlated with *TP53* and *RB1* mutations. Our analysis integrates *TERT* abnormalities, telomerase activity and genomic alterations with telomere length in cancer.

Telomeres make up the terminal ends of each chromosome and are composed of repetitive DNA sequence $(TTAGGG)^n$ and bound proteins[1]. These complexes function by protecting the chromosome ends from being recognized as DNA double-strand breaks and preventing inadvertent activation of detrimental DNA damage response pathways[2]. Telomeres shorten with each cell division, which eventually triggers cellular senescence, resulting in growth arrest[3]. This process can be circumvented by inactivation of p53 and Rb tumor-suppressor proteins[4–7]. Further cell division leads to cellular crisis and ultimately cell death. Cells can occasionally overcome crisis through the activation of a telomere maintenance mechanism.

Senescence and crisis are potent tumor-suppressive mechanisms[8], and maintenance of telomere length is therefore an important step in oncogenesis. Telomere shortening can be counteracted by activation of telomerase[9]. The telomerase enzymatic subunit, encoded by *TERT*, is transcriptionally silent in most non-neoplastic cells, but reactivation may endow a small population of cells with the ability to survive crisis, at which point they become immortalized[10]. It has been proposed that up to 90% of human cancers reactivate telomerase[11]. Several mechanisms have since been associated with *TERT* reactivation, including *TERT* promoter mutations or rearrangements and DNA copy number amplifications[12–15]. Alternative lengthening of telomeres (ALT), a homologous recombination–based process, is frequently observed in tumors lacking telomerase activity and manifests with long but highly variable telomeres[16,17]. Deactivating mutations in *ATRX* and its binding partner *DAXX* were found to be tightly correlated with long telomeres in pancreatic neuroendocrine tumors[18] and gliomas[19]. Recent evidence suggested that loss of *ATRX* may contribute to ALT by promoting sustained sister telomere cohesion and chromatid exchange[20].

Here, we analyzed 18,430 unique samples, including tumors (n = 9,065), blood controls (n = 7,643) and solid tissue controls (n = 1,722), from 9,127 patients and 31 cancer types, to identify genomic and transcriptomic characteristics of telomere length.

RESULTS

Telomere length in human cancer and matching normal tissue

Telomere length (TL) is expected to vary among tumor types owing to varying frequencies of ALT, different age distributions and variability of telomere lengths among cells of origin from different lineages. To quantify this heterogeneity, we estimated TL for 18,430 samples across 31 cancer cohorts available through The Cancer Genome Atlas (TCGA), including samples profiled using whole-genome sequencing (WGS, n = 2,018), low-pass whole-genome sequencing

[1]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA. [2]Oncology Graduate School Amsterdam, VU University Medical Center, Amsterdam, the Netherlands. [3]Department of Genomic Medicine, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [4]Department of Biostatistics, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [5]Department of Neuro-Oncology, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [6]Program in Biostatistics, Bioinformatics, and Systems Biology, the University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas, USA. [7]Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas, USA. [8]Institute for Applied Cancer Science, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [9]Department of Bioinformatics and Computational Biology, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. [10]Biopathology Center, Nationwide Children's Hospital, Columbus, Ohio, USA. [11]Department of Cancer Biology, the University of Texas MD Anderson Cancer Center, Houston, Texas, USA. Correspondence should be addressed to S.Z. (syzhenger@gmail.com) or R.G.W.V. (roel.verhaak@jax.org).

Received 21 March 2016; accepted 4 January 2017; published online 30 January 2017; doi:10.1038/ng.3781

(LPS, $n = 1,929$); and whole-exome sequencing (WXS, $n = 14,483$)[21] (**Fig. 1a** and **Supplementary Table 1**). The full data set consisted of tumor samples, normal blood samples and solid tissue controls (**Supplementary Fig. 1a**). Matching tumor and normal (T/N) samples were available from 8,953 unique patients (**Supplementary Fig. 1b**), and using T/N TL ratios alleviated technical effects from differences in sequencing center and method (**Supplementary Fig. 2**).

To compare TL across tumors and normal tissues, we used linear mixed modeling to adjust high-confidence WGS-based TL ($n = 2,018$) for confounding effects (**Supplementary Table 2**). In addition to 734 normal blood samples, this analysis included 213 normal tissue samples of five different tissue types, including liver ($n = 21$), lung ($n = 46$) and kidney ($n = 81$). We did not detect statistically significant differences between tissue types and confirmed considerable variability among samples from the same tissue type[22] (**Supplementary Fig. 3a**,**b**) and negative correlation between TL and age (**Supplementary Fig. 3c**).

Across neoplastic samples, cervical (2.36 kb, 95% confidence interval (CI) 1.82–3.07 kb) and endometrial cancer (2.82 kb, 95% CI 2.32–3.42 kb) showed the shortest average TL, whereas glioma (5.69 kb, 95% CI 4.75–6.81 kb) and sarcoma (5.71 kb, 95% CI 4.63–7.05 kb) showed the longest (**Fig. 1b**). Tumors showed TL shortening (tumor TL < normal TL) compared to matched normal samples in 70% of our cohort, and relative TL elongation (tumor TL > normal TL) in 30% (**Supplementary Fig. 1b**). Tumor types that showed the highest rates of relatively longer TLs included testicular germ cell tumors (52%), lower-grade glioma (54%) and sarcoma (55%), possibly reflecting the high telomerase activity seen in malignant testicular germ cell tumors[23] and the high frequency of ALT in lower-grade glioma and sarcoma. Conversely, uveal melanoma (100%), kidney chromophobe (89%), kidney papillary ($n = 238/283$, 84%) and lymphoma (84%) demonstrated the highest rates of relatively shorter TLs.

Multiple modalities associated with *TERT* overexpression

To catalog somatic alterations that may lead to overexpression of *TERT* in cancer, we performed a genomic and epigenetic *TERT* survey. We curated a core sample set of the 473 T/N pairs with the most comprehensive molecular profiling and an extended set of 6,835 T/N pairs with varying numbers of cases profiled by each platform (**Fig. 1a** and Online Methods). *TERT* promoter (TERTp) mutations, predominantly chr. 5: 1,295,228 C>T and chr. 5: 1,295,250 C>T, were detected in 27% of the cases in the extended set in which TERTp status could be determined ($n = 1,581$). In agreement with previous reports[24,25], high incidence of TERTp mutations was found in bladder cancer (42/60, 70%), liver cancer (73/162, 45%), melanoma (93/129, 72%), lower-grade glioma (127/285, 45%) and glioblastoma (25/28, 89%) (**Supplementary Fig. 4a**). We found *TERT* focal amplifications in 4% of all samples, and these events were most frequently observed in ovarian cancer (6/27, 22%), lung adenocarcinoma (63/476, 13%), lung squamous cell carcinoma (23/167, 14%), esophageal carcinoma (23/168, 14%) and adrenocortical carcinoma (11/75, 15%)[14] (**Supplementary Fig. 4b**). Structural variants involving *TERT* or TERTp were detected in 15 samples (3%) and 17 samples (4%) of the core set, respectively (**Fig. 2a** and **Supplementary Table 3**). *TERT* or TERTp structural variants were most frequent in sarcoma (10/39, 26%), hepatocellular carcinoma (7/50, 14%), kidney chromophobe (5/49, 10%) and prostate cancer (2/20, 10%) (**Supplementary Fig. 4c**).

TERTp rearrangements have been proposed to lead to juxtaposition of enhancer elements in neuroblastoma[15]. To investigate this hypothesis, we overlaid the genomic positions with a database of 65,950 super-enhancers across 107 tumor and normal cell types[26]. In the majority of TERTp structural variants (65%), at least one predicted super-enhancer was found to directly overlap with the juxtaposed position (**Supplementary Table 3**). We compared signals of enhancer marks acetylation of histone H3 at K27 (H3K27ac) and monomethylation of H3 at K4 (H3K4me1)[27] for each of the breakpoints in the TERTp and the juxtaposed, distal breakpoints across 111 epigenomes available from the NIH Roadmap Epigenomics Consortium[28] and found significant enrichment for enhancer marks in the juxtaposed position compared to TERTp (**Fig. 2b** and **Supplementary Fig. 4d**). Our data suggest that TERTp rearrangements may result in repositioning of enhancer elements that activate *TERT* transcription.

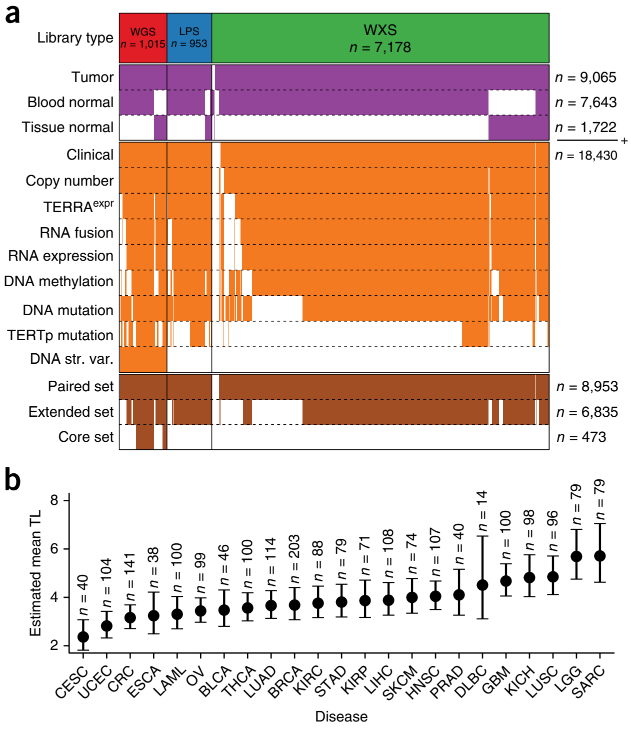

Figure 1 Telomere length in human cancer. (**a**) Heat map of data from patients in the unpaired (purple, $n = 18,430$) and fully paired sets (brown, $n = 8,953$). Each column represents a single patient. Rows in orange represent available data depending on platform. The extended ($n = 6,835$) and core sets ($n = 473$) are also shown. (**b**) Linear mixed model mean TL estimates using the high-confidence WGS set ($n = 2,018$) by sample type and for each tumor type. Error bars, 95% CI. LAML, acute myeloid leukemia; BLCA, bladder urothelial carcinoma; LGG, brain lower-grade glioma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; DLBC, lymphoid neoplasm diffuse large B cell lymphoma; OV, ovarian serous cystadenocarcinoma; PRAD, prostate adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrial carcinoma.

We next performed a supervised search for *TERT* gene fusions[29,30]. *TERT* fusion transcripts were detected in 3% of the extended set (**Supplementary Fig. 5a** and **Supplementary Table 3**). *TERT* was always the 3′ partner gene, and in 13 of 19 fusions the 5′ partner gene resided on chromosome 5. All fusions demonstrated altered exon expression flanking the fusion point, and 16/19 *TERT* fusion breakpoints

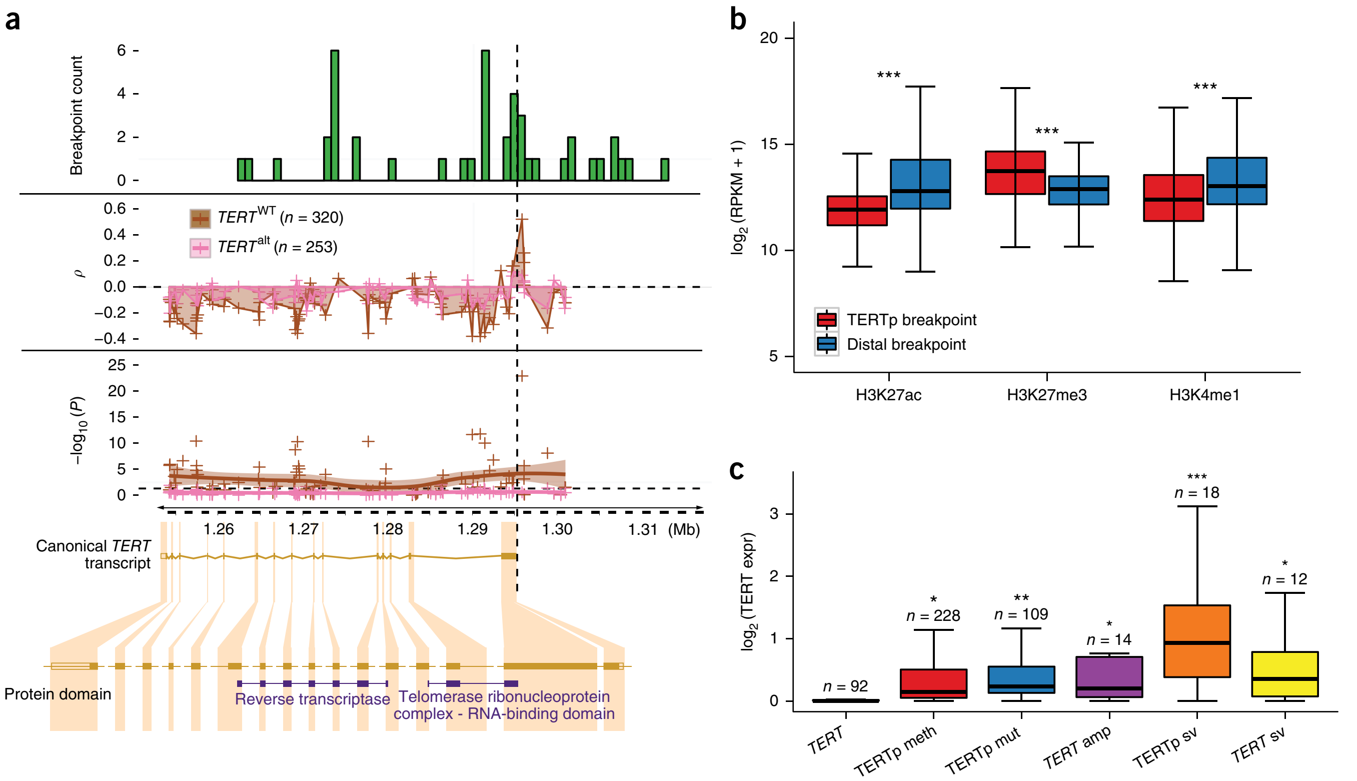

Figure 2 Multiple modalities associated with *TERT* overexpression. (**a**) Top, histogram of DNA breakpoints in *TERT* in the core set ($n = 473$). SpeedSeq detected 44 breakpoints aligning to *TERT* or TERTp in 30 samples. Middle, smoothed scatter plots of ρ and P value of probe-expression correlations for *TERT*. Each point represents an Illumina 450K probe. Vertical dashed line represents the transcription start site. The analysis was performed separately for wild-type *TERT* and *TERT*-mutant samples. Bottom, canonical *TERT* transcript visualized using Ensembl. Protein domains named according to Pfam. (**b**) H3K27ac, H3K27me3 and H3K27me1 levels from the NIH Roadmap Epigenomics data set at the locations of TERTp structural variant proximal and distal breakpoints ($n = 17$). This data set consists of 183 biological samples consolidated into 111 epigenomes (Online Methods). (**c**) *TERT* expression in *TERT*^{alt} and WT groups. ***$P < 0.0001$; **$P < 0.001$; *$P < 0.05$, two-sided *t*-test. Meth, methylation; mut, mutation; amp, amplification; sv, structural variation. Boxes, interquartile range (IQR); center lines, median; whiskers, maximum and minimum or 1.5× IQR.

fell in the second intron. Both exons 2 and 3 map to the telomerase RNA binding domain, and it is thus unlikely that the fusion product retains canonical *TERT* functionality[31,32]. We observed read coverage on exon 2 in 15 of 16 samples, suggesting that additional *TERT* transcripts were expressed.

Taken together, we found somatic *TERT* alterations, including TERTp mutations, *TERT* amplifications and *TERT* structural variants involving gene promoter or gene body, in 32% of core set samples. Somatic *TERT* alterations were associated with detectable *TERT* transcripts in 93%. Next, we evaluated whether epigenetic mechanisms could also be related to *TERT* transcriptional activation. We correlated *TERT* expression to DNA methylation probes mapping to the *TERT* gene body ($n = 72$) and promoter ($n = 3$). We observed moderate correlations between TERTp DNA methylation and *TERT* expression in samples carrying a somatic *TERT* alteration (false discovery rate (FDR) > 0.05, $|\rho| < 0.3$; **Fig. 2a**). In contrast, samples lacking somatic *TERT* alterations showed significant negative correlation between gene body methylation and expression and positive correlation between promoter methylation and expression (**Fig. 2a**). As previously described in pediatric brain tumors[33], TERTp probe cg11625005 demonstrated a strong correlation with *TERT* expression ($\rho = 0.52$, FDR < 0.0001). Further comparison of this probe using 537 paired tumor and adjacent tissue normal samples showed a general absence of methylation of this probe in normal samples (**Supplementary Fig. 5b**).

We found that 63% of tumors carrying wild-type *TERT* in the core set expressed *TERT*, of which 91% showed TERTp DNA methylation, whereas 40% of the *TERT*-nonexpressing samples showed TERTp methylation. Collectively, 95% of *TERT*-expressing samples showed TERTp mutations (31%), *TERT* amplification (3%), *TERT* structural variants (3%), TERTp structural variants (5%) or TERTp methylation (53%). Among different types of *TERT* aberrations, the TERTp structural variant group showed the highest transcription levels (two-sided *t*-test $P < 0.05$; **Fig. 2c**). TERTp methylation (two-sided *t*-test $P < 0.05$) and TERTp mutations (two-sided *t*-test $P < 0.0001$) were associated with relative TL shortening, as compared to other types of *TERT* alterations (**Supplementary Fig. 5c**).

We next performed a similar analysis of *TERC*, the RNA subunit of the telomerase complex[34]. *TERC* was amplified in 4% of samples in the extended set, most frequently in lung squamous cell carcinoma ($n = 68/167$, 41%), esophageal cancer ($n = 36/168$, 21%) and ovarian cancer ($n = 6/27$, 22%). We did not identify mutations or structural variants targeting *TERC*. Focal *TERC* amplifications were associated with increased *TERC* expression (two-sided *t*-test $P < 0.0001$) (**Supplementary Fig. 5d**) and were enriched in *TERT*-expressing samples (odds ratio (OR) 2.59, Fisher's exact $P < 0.0001$).

Inferring telomerase activity using a gene expression signature

Previous studies have shown a complex role for transcription of *TERT* and its various isoforms in determining telomerase activity[35–37], in part because only the full-length *TERT* transcript can encode enzymatically active protein[32]. The β-deletion (β⁻) isoform skips exons 7 and 8 and prematurely stops at exon 10; it is thought to be the most abundant *TERT* splice variant in cancer[38,39] (**Supplementary Table 4**). We evaluated the presence of the β⁻ transcripts in 1,201 samples with sufficient exon 6 to 9 coverage using MISO[40] (**Supplementary Fig. 6a**). Full-length transcripts were detected in all samples and were significantly more abundant than β⁻transcripts (one-sample *t*-test $P < 0.0001$, $\mu = 50\%$) (**Supplementary Fig. 6b**).

Given the limitations of *TERT* expression, as measured by RNA-seq, in predicting telomerase activity and the lack of telomerase measurements in the TCGA proteomics data sets (data not shown), we inferred telomerase activity through a gene signature (**Supplementary Table 4**). The signature showed a positive correlation with telomerase enzymatic activities determined in urothelial cancer cell lines ($n = 11$, **Supplementary Fig. 7a**) without reaching statistical significance[35] ($P = 0.07$). Telomerase activity scores were significantly higher in tumor samples than in various solid tissues (two-sided *t*-test $P < 0.01$) (**Supplementary Fig. 7b**), with kidney chromophobe carcinoma representing an exception. Both results suggested that the gene signature provided a general estimate of telomerase activity.

We found that telomerase signature scores were significantly higher in *TERT*-expressing cancers than nonexpressing cancers in both the core and extended sets (both $P < 0.0001$, Wilcoxon rank-sum test). Samples with *TERT* amplification scored highest, followed by promoter methylation, gene body and promoter structural variation and promoter mutation (**Supplementary Fig. 7c**). *TERC* amplification was additionally associated with high telomerase signature scores compared to nonamplified samples (two-sided *t*-test, $P < 0.0001$), which may in part be explained by the coexpression patterns of *TERT* and *TERC*. Pan-cancer analysis using all 31 cancer types identified a positive correlation between *TERT* expression and telomerase signature score ($\rho = 0.69$, $P < 0.0001$), with testicular germ cell tumors showing the highest average scores and pheochromocytoma and paraganglioma the lowest (**Supplementary Fig. 7d**).

Telomere elongation in *ATRX*-altered tumors

To identify *TERT*-independent mechanisms involved in TL regulation, we associated somatic alterations to TL ratio in the extended set. We reduced the search space by selecting genes significantly mutated, genes focally deleted or gained and genes included in a manually compiled list of telomere-associated genes ($n = 196$; **Supplementary Table 5**). Alterations of *ATRX* and *IDH1* were the most significantly associated with relative TL elongation (both FDR < 0.0001; **Fig. 3a**). Because *IDH1* and *ATRX* mutations frequently co-occur in glioma, we tested a model with both tumor type and *IDH1* as covariates, and found *IDH1* no longer associated with TL ratio (two-sided *t*-test $P = 0.15$). Other hits (FDR < 0.25) associated with relative TL elongation included *TP53* (TL ratio 1.15, 95% CI 1.1–1.21, FDR < 0.0001), *BCOR* (TL ratio 1.24, 95% CI 1.08–1.42, FDR = 0.04), *RB1* (TL ratio 1.14, 95% CI 1.04–1.25, FDR = 0.05), *CCNE1* (TL ratio 1.19, 95% CI 1.05–1.35, FDR = 0.07) and *TERC* (TL ratio 1.14, 95% CI 1.02–1.27, FDR = 0.16). Alterations of *VHL* were found to be associated with relative TL shortening (TL ratio 0.7, 95% CI 0.61–0.8, FDR < 0.0001). We repeated the analysis in each cancer type and found *TP53* associated with relative TL elongation in six tumor types (**Supplementary Fig. 8a**). Given the role of p53 in apoptosis regulation and senescence bypass, direct involvement of *TP53* in telomere maintenance should be carefully tested in well-controlled conditions. A linear regression model showed that in addition to older age, positive TERRA expression, *TP53* deletion, *TP53* mutations, *ATRX* deletion, *ATRX*

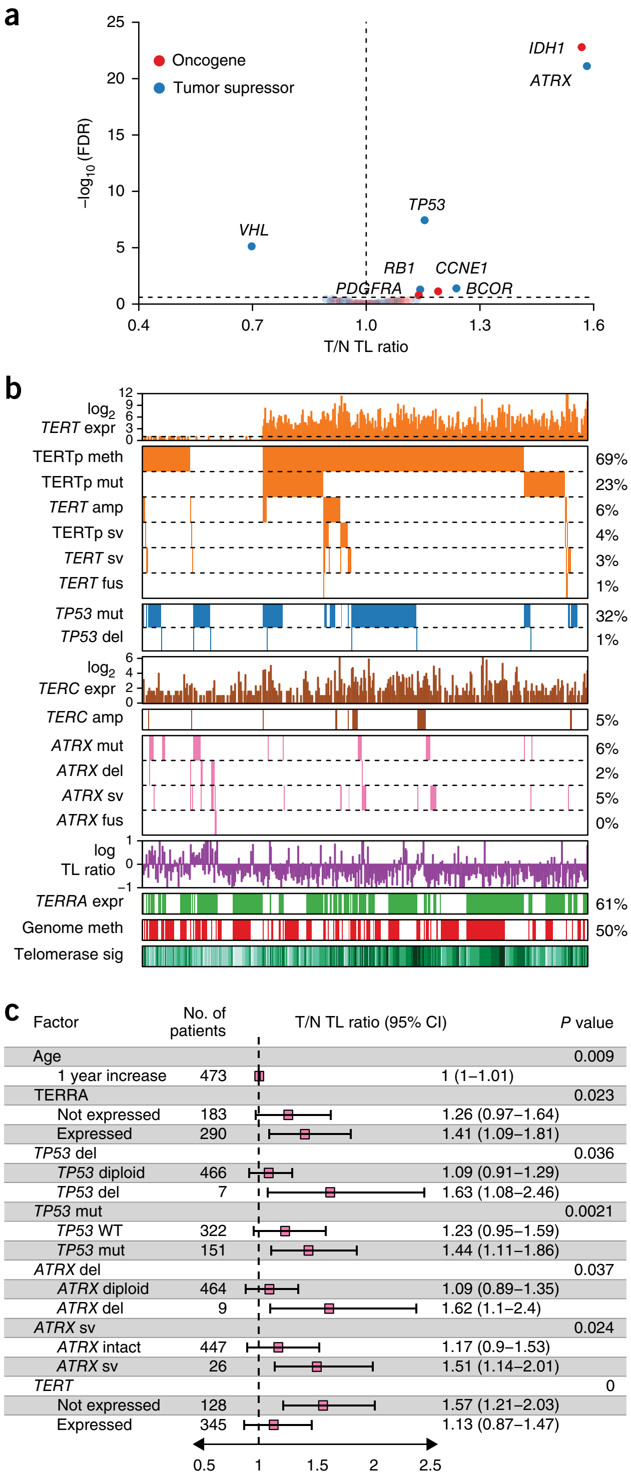

Figure 3 Multivariable genomic determinants of TL. (**a**) Gene to TL ratio associations using the extended set ($n = 6,835$). *P* values were calculated using a two-sided *t*-test and adjusted for multiple testing using FDR. (**b**) Heat map of *TERT*, *TERC*, *ATRX* and *TP53* expression and somatic alterations in the core set ($n = 473$). TL ratio, TERRA expression and telomerase signature (sig) score are also shown. Each column represents a sample. (**c**) Linear regression analysis of TL ratio. Variables shown are independent predictors of TL in the core set ($n = 473$). Variables from **b** were selected using backwards elimination to derive the final model. $R^2 = 0.16$. Meth, methylation; mut, mutation; amp, amplification; sv, structural variation; fus, fusion; del, deletion.

structural variants and absent or undetectable *TERT* expression were all independently associated with relative TL elongation (**Fig. 3b**,**c**). Although *DAXX* has been linked to telomere length and ALT[19,41], *DAXX* mutations (n = 51/6,835) and deletions (n = 5/6,835) did not associate with TL.

Because of the strong correlation between *ATRX* and TL, we performed a thorough investigation of alterations affecting this gene. In addition to nonsynonymous mutations and deletions, we detected *ATRX* structural variants from WGS data in 5% of samples from the core set (n = 26/473). To detect *ATRX* structural variants across the extended set, we used DNA copy number calls to detect breakpoints in *ATRX* (**Supplementary Fig. 8b**). Despite the high detection threshold of this method (34% sensitivity, 100% specificity), *ATRX* structural variants were predicted in 70 of 6,835 samples (1%). *ATRX* fusion transcripts were found in seven extended set samples (**Supplementary Fig. 8c**). *ATRX* was the 5′ gene partner in five fusion genes, and the fusions were predicted to result in retention of <10% of the ATRX protein sequence. *ATRX* was the 3′ partner in two fusion transcripts, and functional protein domains were retained in both.

We observed a significant decrease in *ATRX* expression in samples showing *ATRX* mutations, deletions, fusions and structural variants compared to cases with wild-type *ATRX* (**Fig. 4a**). We found that all of types of *ATRX* alteration associated with significantly longer TLs than wild-type *ATRX*, consistent with the previously established association between *ATRX* deactivation and ALT (**Fig. 4b**).

Recent studies found that *ATRX* knockdown results in higher levels of TERRA[42]. We estimated TERRA levels using RNA sequencing in the extended set and in 566 non-neoplastic samples. Our results showed a significantly higher fraction of TERRA-expressing samples in all groups of *ATRX*-altered samples (**Fig. 4c**; Fisher's exact test $P < 0.05$) than in *ATRX* wild-type samples. TERRA expression was associated with relative TL elongation (two-sided t-test TL ratio 1.13, 95% CI 1.07–1.017, $P < 0.0001$). Significantly more samples expressed TERRA in tumors lacking *TERT* expression than in tumors expressing *TERT* (Fisher's exact OR 0.73, 95% CI 0.65–0.82, $P < 0.0001$). In contrast, comparison of TERRA levels in tumor samples and their matched normal controls (n = 596, Fisher's exact test $P > 0.05$) showed no significant differences.

Absence of *TERT* expression and *ATRX* deactivation.

Owing to the association between somatic *ATRX* and *TERT* alterations with TL, we grouped tumors in the extended set as *TERT*-expressing (*TERT*$^{\text{expr}}$, n = 5,001/6,835, 73%) and *ATRX*- or *DAXX*-altered (*ATRX*/*DAXX*$^{\text{alt}}$, n = 309/6,835, 5%). *ATRX* or *DAXX* mutations were found in 210 *TERT*$^{\text{expr}}$ samples, representing 3% of the cohort. These events were mostly nontruncating, while *ATRX* and *DAXX* mutations in *TERT*-negative cases were mostly truncating (**Supplementary Note** and **Supplementary Fig. 9a**). *ATRX* and *DAXX* mutants expressing *TERT* showed higher telomerase signature scores than did *ATRX*/*DAXX*$^{\text{alt}}$ samples lacking *TERT* expression (**Supplementary Fig. 9b**). On the basis of these observations, we included these samples in the *TERT*$^{\text{expr}}$ category. The remaining 22% of samples had neither detectable *TERT* expression nor somatic alterations in *ATRX* or *DAXX* (WT/WT, n = 1,525/6,835; **Fig. 5a**). Both *TERT*$^{\text{expr}}$ and WT/WT groups showed significantly higher telomerase signature scores than the *ATRX*/*DAXX*$^{\text{alt}}$ group (**Fig. 5b**; two-sided t-test $P < 0.0001$). The intermediate telomerase activity levels and variable TLs suggested that the WT/WT group comprises a heterogeneous set of tumors, some of which may have undergone ALT through mechanisms independent of *ATRX* and *DAXX*. The WT/WT samples were most frequent among pheochromocytoma (n = 141/160, 88%), kidney papillary

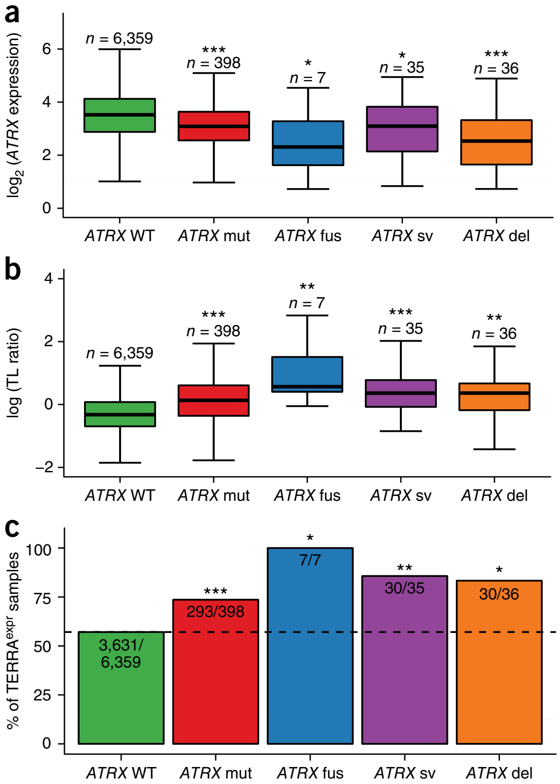

Figure 4 *ATRX*-altered tumors show profound telomere elongation. (**a**,**b**) *ATRX* expression (**a**) and TL ratios (**b**) in *ATRX*-altered and WT groups. (**c**) Percentage of TERRA-expressing (TERRA$^{\text{expr}}$) tumors per group of *ATRX* alterations. Each group was compared to the ATRX WT group; ***$P < 0.0001$; **$P < 0.001$; *$P < 0.05$, two-sided t-test (**a**,**b**) or two-sided Fisher's exact test (**c**). Boxes, interquartile range (IQR); center lines, median; whiskers, maximum and minimum or 1.5× IQR. Mut, mutation; sv, structural variation; fus, fusion; del, deletion.

(n = 113/161, 70%), kidney chromophobe (n = 52/65, 80%) and thyroid cancer (n = 343/449, 79%). Seven of the 101 WT/WT samples in the core set showed a somatic *TERT* alteration in absence of *TERT* expression, including TERTp mutation (n = 1), *TERT* amplification (n = 2) and TERTp or *TERT* structural variant (n = 4).

TERT$^{\text{expr}}$ samples showed relative TL attrition in most cancer types, *ATRX*-altered samples tended to show relative TL elongation, and WT/WT samples demonstrated cancer-type-dependent patterns (**Fig. 5c**). For example, WT/WT sarcoma (TL ratio 1.27, 95% CI 1.1–1.47) and glioma (TL ratio 1.33, 95% CI 1.01–1.75) samples showed relative TL elongation, whereas WT/WT thyroid (TL ratio 0.69, 95% CI 0.62–0.77), kidney chromophobe (TL ratio 0.74, 95% CI 0.67–0.81) and kidney clear cell (TL ratio 0.57, 95% CI 0.48–0.67) cancer samples showed relative TL attrition. Patterns of relative TL change from the WT/WT group indicate that some tumors may be detected before having acquired immortalized cells[43] or that other mechanisms to develop ALT exist.

Because this analysis suggests that the prevalence of ALT may be underestimated by *ATRX* or *DAXX* inactivation alone, we sought to compare the prevalence of *ATRX* and *DAXX* alteration to published prevalence of ALT across 26 cancer types, including 40 histological subtypes[44] (**Supplementary Table 6**). The prevalence of ALT exceeded that of *ATRX* and *DAXX* alterations in seven histological subtypes and was reduced in two, reinforcing the notion of cancer-type-specific and *ATRX*- and *DAXX*-independent ALT mechanisms.

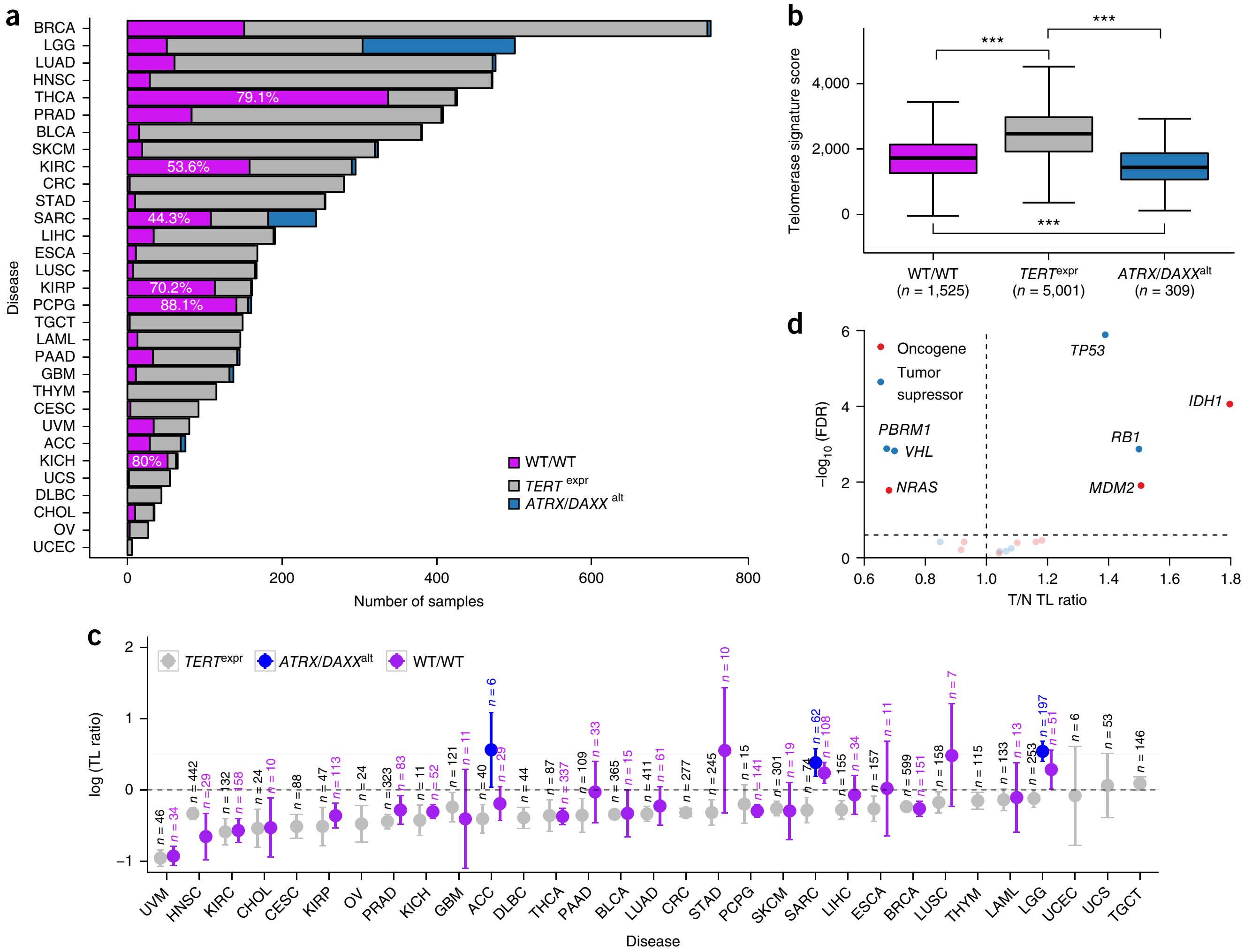

Figure 5 A substantial fraction of cancer samples lacks detectable *TERT* expression and mechanisms of *ATRX* deactivation. (**a**) Frequency of *TERT*^{expr} and *ATRX/DAXX*^{alt} groups across cancer types in the extended set. ACC, adrenocortical carcinoma; CHOL, cholangiocarcinoma; COAD, colon adenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; TGCT, testicular germ cell tumors; THYM, thymoma; UCS, uterine carcinosarcoma; UVM, uveal melanoma. Other abbreviations are as in **Figure 1**. (**b**) Telomerase signature score by *TERT* and *ATRX/DAXX* status. (**c**) T/N TL ratio by cancer type. Groups with <6 samples were omitted. Error bars, 95% CI. (**d**) Gene-to-T/N TL ratio associations within the WT/WT group (n = 1,525). ***P < 0.0001; **P < 0.001; *P < 0.01, two-sided t-test. FDR adjustment for multiple testing was used in **d**. Boxes, interquartile range (IQR); center lines, median; whiskers, maximum and minimum or 1.5× IQR.

To better understand alternative mechanisms that may contribute to telomere maintenance, we conducted an unsupervised search for TL ratio–associated somatic alterations within the WT/WT group in the extended set. This analysis recovered strong associations between relatively long TL, *IDH1* (TL ratio 1.8, 95% CI 1.39–2.33, FDR = 0.0001) and *TP53* (TL ratio 1.39, 95% CI 1.23–1.57, FDR < 0.0001), both of which were also identified in the analysis of TL length in the entire data set. The finding of *IDH1* may reflect a glioma-specific effect. In addition to *IDH1* and *TP53*, *RB1* (TL ratio 1.5, 95% CI 1.2–1.87, FDR = 0.001) and *MDM2* (TL ratio 1.51, 95% CI 1.14–2, FDR = 0.01) were revealed to associate with relatively long TLs. The finding of *RB1* is consistent with experimental data demonstrating marked elongation of telomeres in Rb1-deficient mice independent of telomerase[45,46]. In the opposite direction, somatic alterations in *PBRM1* (TL ratio 0.67, 95% CI 0.55–0.83, FDR = 0.001), *NRAS* (TL ratio 0.68, 95% CI 0.52–0.9, FDR = 0.02) and *VHL* (TL ratio 0.7, 95% CI 0.57–0.85, FDR = 0.002) were associated with relative TL shortening (**Fig. 5d**).

We compared global genomic characteristics of the three groups and found that WT/WT samples contained fewer copy number segments (two-sided t-test P < 0.0001; **Supplementary Fig. 9c**) and had a lower mutation rate (two-sided t-test P < 0.0001; **Supplementary Fig. 9d**) than *TERT*^{expr} and *ATRX/DAXX*^{alt} samples. Survival analyses within cancer types showed better survival in WT/WT than in *TERT*^{expr} sarcoma and thyroid carcinoma (**Supplementary Fig. 9e**). Telomerase activity scores in WT/WT samples in these two cancer types were significantly lower (two-sided t-test $P = 7.0 \times 10^{-7}$ and $P = 7.8 \times 10^{-3}$, respectively), suggesting that telomerase activity might represent a relevant prognostic factor in these cancers.

Telomere position effect

After the quantification of TLs, we sought to relate TL to gene expression changes using the extended set and 566 adjacent tissue normal samples (**Supplementary Table 7**). We first investigated the telomere position effect (TPE), a phenomenon that describes TL-mediated transcriptional regulation[47]. Expression of genes close to telomeres was negatively correlated with TL, and this effect was attenuated as distance to telomeres increased (**Fig. 6a,b**). Because tumors had shorter telomeres on average than matched normal samples, genes

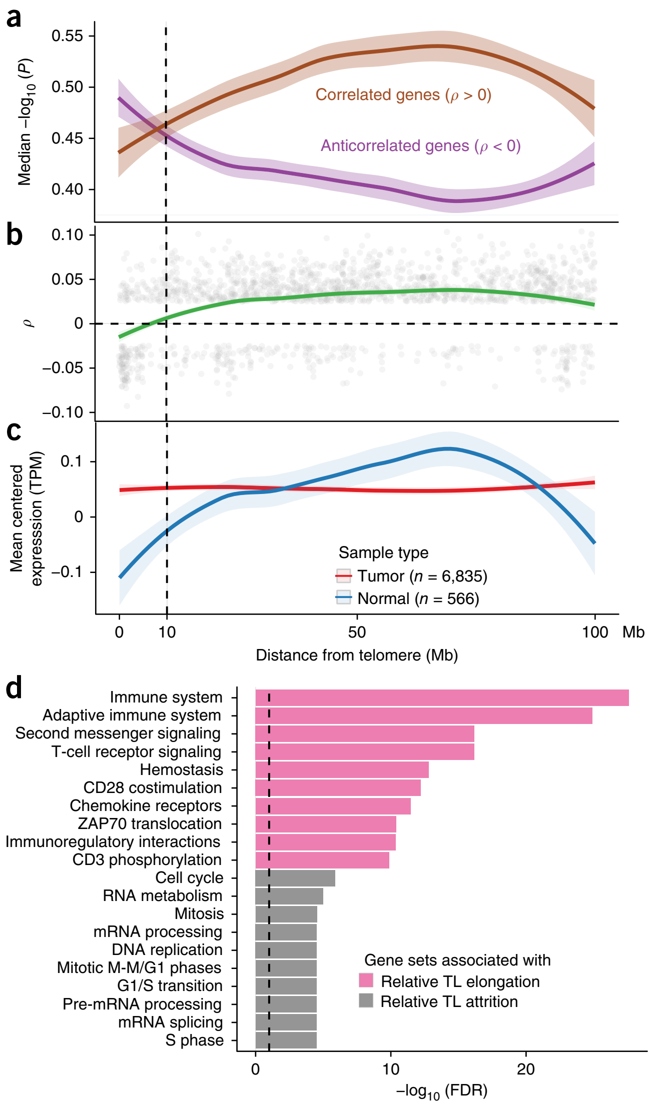

Figure 6 TPE. (**a**) $-\log_{10}(P)$ for genes showing a correlation or an anticorrelation to tumor TL relative to the distance to the respective telomere. Spearman correlation tests were conducted individually within cancer types. P values for n = 3,477 genes were averaged across cancer types. (**b**) Spearman ρ across all cancer types for n = 2,016 genes significantly associated to TL (FDR < 0.25) relative to the distance to the respective telomere. (**c**) Average gene expression (in transcripts per million (TPM)) for tumor and normal samples relative to distance to the respective telomere. Mean expression was calculated for n = 3,477 genes. (**d**) Reactome gene expression pathway analysis on the extended set (n = 6,835 samples) of the top 500 genes most associated with relative TL elongation and shortening, respectively. Top 10 pathways for each set of 500 genes are shown.

close to telomeres showed higher expression in tumor than normal tissue (**Fig. 6c**). The TPE was not detectable beyond the distance of 10 million bp from telomeres (**Fig. 6a**,**b**).

To identify gene sets associated with TL attrition and elongation in an unsupervised fashion, we associated gene expression to TL attrition and elongation. Pathway analysis of the top 500 correlated genes indicated that relative TL attrition and elongation were associated with immunoreactive and proliferative signatures, respectively (**Fig. 6d**). These results may reflect differences in telomere maintenance rather than telomere attrition and elongation and describe frequently observed patterns of expression across cancers, potentially related to aggressiveness of tumors[48].

DISCUSSION

This study represents, to our knowledge, the largest in-depth integrative analysis of TL and related somatic alterations performed to date. As expected, tumor TLs were shorter than normal tissue TLs, and TL was inversely correlated with age in both tumor and non-neoplastic tissues. Among *TERT*$^{\text{expr}}$ tumors, 32% carried at least one of three *TERT* abnormalities: promoter mutation, amplification or chromosomal rearrangement. Of *TERT*$^{\text{expr}}$ and wild-type tumors, 91% were TERTp methylated, compared to 40% of tumors lacking *TERT* transcription. This paradoxical association between TERTp methylation and increased *TERT* expression may result from loss of binding of CTCF, a transcriptional repressor reported to bind to the unmethylated TERTp[49]. Structural *TERT* variants have been documented[15,50,51], and we detected these across several cancer types previously unknown to show such abnormalities, including sarcoma, prostate and liver carcinoma. Hepatitis B virus and adeno-associated virus type 2 integration in *TERT* were found in about 5% of hepatocellular carcinomas[52]. The absence of similar observations in other virus-related cancers (including head and neck, bladder and gastric cancers) suggests that this mechanism is specific to liver cancer; we thus did not consider it in this study.

Whether *TERT* expression translates directly to active telomerase activity is unclear. Only the full-length transcript (as opposed to known isoforms) has been found to activate telomerase[32,39]. *TERC* expression is additionally needed for telomerase activity[53]. An estimated 5–10% of *TERC* and TERRA transcripts carry the poly(A) tails required for oligo(dT)-primer-based RNA sequencing quantification[54–56]. We detected *TERC* and TERRA expression across our cohort, suggesting that if expression is sufficiently high, transcripts can be detected using conventional RNA sequencing approaches. Lacking telomerase activity data, we sought to infer telomerase activity using a gene-signature-driven approach. Our results suggested a positive correlation between *TERT* expression and telomerase activity, corroborating recent findings in bladder cancer[35]. We also observed a positive correlation between *TERC* expression and telomerase activity, as well as increased telomerase activity in *TERT*-expressing tumors compared to tumors with *ATRX* or *DAXX* alterations and WT/WT tumors. This gene signature may serve as a useful proxy for estimating telomerase activity using transcriptional profiles.

Deactivation of *ATRX* and/or *DAXX* has been related to ALT[19,20,57] and was observed in 5% of cases in our core set. A detailed review of *ATRX* somatic changes showed a large spectrum of potentially protein truncating changes, including inactivating mutations, deletions and structural variants. *DAXX* alterations were much less frequent. Our analysis reinforced the association of inactivated *ATRX* or *DAXX* and ALT, demonstrating relative TL elongation in samples affected by somatic alterations in one of these two genes and a higher frequency of TERRA expression in tumors with these alterations.

The 22% of cases that lacked detectable *TERT* expression and known ALT-related abnormalities provide a notable result from our analysis. It is plausible that *TERT* transcription below the currently applied detection threshold may be sufficient for telomere maintenance, that not all tumors harbor immortalized cells with a telomere maintenance mechanism[43] or that additional telomere maintenance mechanisms exist. Such mechanisms may involve some *RB1* and *TP53* alterations, as somatic changes in these genes were associated with telomere elongation within this group. Notably, WT/WT tumors were mostly pheochromocytoma and paraganglioma, kidney chromophobe and papillary thyroid tumors. Pheochromocytoma is generally (>90%) labeled benign. Chromophobe renal cell carcinoma and papillary thyroid cancer are malignant tumors but are generally well differentiated, infrequently metastasize and demonstrate a more favorable outcome than other cancer types[58–61]. Future studies are needed to elucidate the telomere maintenance mechanisms, or lack thereof, in these tumor types.

In summary, our analysis has broadened the scope of potential *TERT*-activating changes, completed the spectrum of *ATRX*-truncating alterations and provided new insights into the telomere biology of tumors lacking these classic alterations. These findings extend current understanding of telomere biology and open avenues for functional studies of how to target this crucial pathway in oncogenesis.

URLs. PRADA fusion portal, http://www.tumorfusions.org.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all patients who consented to have their samples analyzed by The Cancer Genome Atlas Network. We thank D. Jiao and D. Jackson at the High Performance Computing facility at MD Anderson Cancer Center for assistance with data downloading and storage. We thank I.R. Watson (McGill University) and T. Wu (MD Anderson Cancer Center) for insights on TCGA melanoma project, and D. Wheeler (Baylor College of Medicine) for insights on TCGA liver cancer project. We thank S. Borah and T.R. Cech (University of Colorado) for sharing telomerase enzymatic activity data of the 23 urothelial cancer cell lines. We thank D. Broccoli (Mercer University) for sharing the 14 gene signature from GSE20559. We thank X. Wang (Baylor College of Medicine) for help with CPTAC data and A. Barbo (MD Anderson Cancer Center) for statistical advice. We thank B. Murray (Broad Institute) for suggestions regarding the analysis of copy number data. This project is supported by grants from the National Institutes of Health (R01CA190121 and P01CA085878 to R.G.W.V., P50CA127001 to Z.S. and R.G.W.V. and U24CA143883) and the Cancer Prevention and Research Institute of Texas (CPRIT) (R140606) to R.G.W.V. This work was supported by Cancer Center Support Grants P30CA16672 and P30CA034196.

AUTHOR CONTRIBUTIONS

F.P.B. was involved in all aspects of data analysis. W.W. performed linear mixed modeling. M.T. and S.B.A. analyzed whole-genome sequencing data. E.M.-L. analyzed gene expression data. X.H. and Q.W. performed gene fusion analysis. K.C.A. was involved in methylation and epigenetics analysis. S.S., X.S. and J.Z. collected and analyzed low-pass sequencing data. T.L. collected clinical data for the solid tissue samples. J.H. provided critical insights into ALT and telomerase biology. S.Z. designed the telomerase signature score. S.Z., F.P.B. and R.G.W.V. conceived the study and wrote the paper. S.Z. and R.G.W.V. supervised the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. O'Sullivan, R.J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11**, 171–181 (2010).
2. de Lange, T. How telomeres solve the end-protection problem. *Science* **326**, 948–952 (2009).
3. Olovnikov, A.M. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.* **41**, 181–190 (1973).
4. Shay, J.W., Pereira-Smith, O.M. & Wright, W.E. A role for both RB and p53 in the regulation of human cellular senescence. *Exp. Cell Res.* **196**, 33–39 (1991).
5. Stewart, S.A. & Weinberg, R.A. Telomeres: cancer to human aging. *Annu. Rev. Cell Dev. Biol.* **22**, 531–557 (2006).
6. Maser, R.S. & DePinho, R.A. Connecting chromosomes, crisis, and cancer. *Science* **297**, 565–569 (2002).
7. Sahin, E. & DePinho, R.A. Axis of ageing: telomeres, p53 and mitochondria. *Nat. Rev. Mol. Cell Biol.* **13**, 397–404 (2012).
8. Hackett, J.A. & Greider, C.W. Balancing instability: dual roles for telomerase and telomere dysfunction in tumorigenesis. *Oncogene* **21**, 619–626 (2002).
9. Greider, C.W. & Blackburn, E.H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405–413 (1985).
10. Morales, C.P. *et al.* Absence of cancer-associated changes in human fibroblasts immortalized with telomerase. *Nat. Genet.* **21**, 115–118 (1999).
11. Shay, J.W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33**, 787–791 (1997).
12. Huang, F.W. *et al.* Highly recurrent *TERT* promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
13. Horn, S. *et al. TERT* promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
14. Zhang, A. *et al.* Frequent amplification of the telomerase reverse transcriptase gene in human tumors. *Cancer Res.* **60**, 6230–6235 (2000).
15. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700–704 (2015).
16. Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A. & Reddel, R.R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* **3**, 1271–1274 (1997).
17. Dilley, R.L. & Greenberg, R.A. ALTernative telomere maintenance and cancer. *Trends Cancer* **1**, 145–156 (2015).
18. Jiao, Y. *et al.* DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**, 1199–1203 (2011).
19. Heaphy, C.M. *et al.* Altered telomeres in tumors with *ATRX* and *DAXX* mutations. *Science* **333**, 425 (2011).
20. Ramamoorthy, M. & Smith, S. Loss of ATRX suppresses resolution of telomere cohesion to control recombination in ALT cancer cells. *Cancer Cell* **28**, 357–369 (2015).
21. Ding, Z., Mangino, M., Aviv, A., Spector, T. & Durbin, R. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res.* **42**, e75 (2014).
22. Dlouha, D., Maluskova, J., Kralova Lesna, I., Lanska, V. & Hubacek, J.A. Comparison of the relative telomere length measured in leukocytes and eleven different human tissues. *Physiol. Res.* **63** (Suppl. 3), S343–S350 (2014).
23. Albanell, J. *et al.* Telomerase activity in germ cell cancers and mature teratomas. *J. Natl. Cancer Inst.* **91**, 1321–1326 (1999).
24. Killela, P.J. *et al. TERT* promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. USA* **110**, 6021–6026 (2013).
25. Vinagre, J. *et al.* Frequency of *TERT* promoter mutations in human cancers. *Nat. Commun.* **4**, 2185 (2013).
26. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.* **44**, D164–D171 (2016).
27. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
28. Kundaje, A. *et al.*; Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
29. Torres-García, W. *et al.* PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics* **30**, 2224–2226 (2014).
30. Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845–4854 (2015).
31. Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
32. Hrdlicková, R., Nehyba, J. & Bose, H.R. Jr. Alternatively spliced telomerase reverse transcriptase variants lacking telomerase activity stimulate cell proliferation. *Mol. Cell. Biol.* **32**, 4283–4296 (2012).
33. Castelo-Branco, P. *et al.* Methylation of the *TERT* promoter and risk stratification of childhood brain tumours: an integrative genomic and molecular study. *Lancet Oncol.* **14**, 534–542 (2013).
34. Blasco, M.A. Telomeres and human disease: ageing, cancer and beyond. *Nat. Rev. Genet.* **6**, 611–622 (2005).
35. Borah, S. *et al. TERT* promoter mutations and telomerase reactivation in urothelial cancer. *Science* **347**, 1006–1010 (2015).
36. Counter, C.M. *et al.* Telomerase activity is restored in human cells by ectopic expression of hTERT (hEST2), the catalytic subunit of telomerase. *Oncogene* **16**, 1217–1222 (1998).
37. Rohde, V. *et al.* Expression of the human telomerase reverse transcriptase is not related to telomerase activity in normal and malignant renal tissue. *Clin. Cancer Res.* **6**, 4803–4809 (2000).
38. Kilian, A. *et al.* Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types. *Hum. Mol. Genet.* **6**, 2011–2019 (1997).
39. Wong, M.S., Wright, W.E. & Shay, J.W. Alternative splicing regulation of telomerase: a new paradigm? *Trends Genet.* **30**, 430–438 (2014).
40. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
41. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
42. Flynn, R.L. *et al.* Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science* **347**, 273–277 (2015).

43. Reddel, R.R. The role of senescence and immortalization in carcinogenesis. *Carcinogenesis* **21**, 477–484 (2000).
44. Heaphy, C.M. *et al.* Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am. J. Pathol.* **179**, 1608–1615 (2011).
45. Gonzalo, S. *et al.* Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat. Cell Biol.* **7**, 420–428 (2005).
46. García-Cao, M., Gonzalo, S., Dean, D. & Blasco, M.A. A role for the Rb family of proteins in controlling telomere length. *Nat. Genet.* **32**, 415–419 (2002).
47. Robin, J.D. *et al.* Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* **28**, 2464–2476 (2014).
48. Martínez, E. *et al.* Comparison of gene expression patterns across 12 tumor types identifies a cancer supercluster characterized by *TP53* mutations and cell cycle defects. *Oncogene* **34**, 2732–2740 (2015).
49. Renaud, S. *et al.* Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene. *Nucleic Acids Res.* **35**, 1245–1256 (2007).
50. Valentijn, L.J. *et al.* *TERT* rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat. Genet.* **47**, 1411–1414 (2015).
51. Davis, C.F. *et al.* The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
52. Khoury, J.D. *et al.* Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq. *J. Virol.* **87**, 8916–8926 (2013).
53. Xi, L. & Cech, T.R. Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Res.* **42**, 8565–8577 (2014).
54. Chapon, C., Cech, T.R. & Zaug, A.J. Polyadenylation of telomerase RNA in budding yeast. *RNA* **3**, 1337–1351 (1997).
55. Porro, A., Feuerhahn, S., Reichenbach, P. & Lingner, J. Molecular dissection of telomeric repeat-containing RNA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol. Cell. Biol.* **30**, 4808–4817 (2010).
56. Feuerhahn, S., Iglesias, N., Panza, A., Porro, A. & Lingner, J. TERRA biogenesis, turnover and implications for function. *FEBS Lett.* **584**, 3812–3818 (2010).
57. Clynes, D. *et al.* Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat. Commun.* **6**, 7538 (2015).
58. Przybycin, C.G. *et al.* Chromophobe renal cell carcinoma: a clinicopathologic study of 203 tumors in 200 patients with primary resection at a single institution. *Am. J. Surg. Pathol.* **35**, 962–970 (2011).
59. Guo, Z. & Lloyd, R.V. Pheochromocytomas and paragangliomas: an update on recent molecular genetic advances and criteria for malignancy. *Adv. Anat. Pathol.* **22**, 283–293 (2015).
60. Davies, L. & Welch, H.G. Increasing incidence of thyroid cancer in the United States, 1973-2002. *J. Am. Med. Assoc.* **295**, 2164–2167 (2006).
61. Lenders, J.W.M., Eisenhofer, G., Mannelli, M. & Pacak, K. Phaeochromocytoma. *Lancet* **366**, 665–675 (2005).

ONLINE METHODS

Sample selection. The 31 cancer types included in this study were: acute myeloid leukemia (LAML); adrenocortical carcinoma (ACC); bladder urothelial carcinoma (BLCA); brain lower-grade glioma (LGG); breast invasive carcinoma (BRCA); cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC); cholangiocarcinoma (CHOL); colon adenocarcinoma (COAD); esophageal carcinoma (ESCA); glioblastoma multiforme (GBM); head and neck squamous cell carcinoma (HNSC); kidney chromophobe (KICH); kidney renal clear cell carcinoma (KIRC); kidney renal papillary cell carcinoma (KIRP); liver hepatocellular carcinoma (LIHC); lung adenocarcinoma (LUAD); lung squamous cell carcinoma (LUSC); lymphoid neoplasm diffuse large B cell lymphoma (DLBC); ovarian serous cystadenocarcinoma (OV); pancreatic adenocarcinoma (PAAD); pheochromocytoma and paraganglioma (PCPG); prostate adenocarcinoma (PRAD); rectum adenocarcinoma (READ); sarcoma (SARC); skin cutaneous melanoma (SKCM); stomach adenocarcinoma (STAD); testicular germ cell tumors (TGCT); thymoma (THYM); thyroid carcinoma (THCA); uterine carcinosarcoma (UCS); uterine corpus endometrial carcinoma (UCEC); uveal melanoma (UVM).

We removed cases that were annotated as having bad DNA quality or failing a quality-control (QC) step. We also removed technical replicates and samples from patients who had prior systematic treatment or revised pathological diagnosis. Tumor and control samples from each patient were paired, selecting a blood-derived control if both blood and solid tissue control were available. A comprehensive sample selection procedure can be found in the **Supplementary Note**, and a schematic is shown in **Supplementary Figure 1a**.

Data generation. Raw RNA and DNA sequencing data (n = 35,978 samples) were downloaded from CGHub[62] and processed through a flowr[63] pipeline consisting of various components depending on the data type. Processed clinical, gene expression, DNA methylation, copy number segmentation and mutation data were downloaded and compiled from previously published TCGA papers, the Firehose data portal (Broad Institute) and the TCGA data portal (NIH). GISTIC[64] was employed on the processed segmentation files to infer somatic copy number changes. A more detailed description of the data collection and generation process is described in the **Supplementary Note**.

Telomere length quantification. Quantification of telomere length (TL) was performed using TelSeq[21]. Briefly, this tool counts the number of reads containing any (range k to ∞) amount of telomeric repeats (n_k), or TTAGGG$_{[k,\infty]}$, and the GC-adjusted coverage s. The quotient of the number of telomeric repeats and the GC-adjusted coverage is then multiplied by the average chromosome length c (c = 332,720,800/46,000, in kb), resulting in the estimated telomere length (l) in kb.

$$\ell = c \times \frac{n_k}{s}$$

We used a k of 7, as per the author's recommendations. Using the default settings from TelSeq, this calculation is done individually for each read group within a sample. In order to calculate the average TL for each sample, the weighted average length was used, supplying the total number of reads in each read group as weight.

RNA-seq BAM files were also processed using TelSeq (k = 7) to obtain an estimate of TERRA expression. Because TERRA expression demonstrated a bimodal distribution, and over 40% of samples lacked any detectable TERRA expression, TERRA expression was dichotomized using a cutoff of 0.

TERTp mutation detection. We used GATK[65] pileup to determine bases mapping to each position 200 bp upstream of the *TERT* transcription start site (chr. 5:1295162–1295404, hg19) in 1,771 tumor samples and matched normal controls from 20 cancer types where whole-genome or low-pass whole-genome sequencing data were available.

We first performed an unsupervised screen, testing each of 200 sites in every tumor and matched normal. For each site, we required a minimum coverage of 10 reads, minimum variant allele fraction of 25% in tumor and maximum variant allele fraction of 2% in normal. Fisher's test was used to determine significance, and a threshold of 0.05 was used. Samples with *TERT* expression below a threshold of two reads were filtered from the analysis. Four sites demonstrated significance in at least one sample: C228, C250, C242/243 and C169. We then determined the nucleotides called in each of the affected sites, and found that all affected sites underwent a C>T transition.

Next, we performed a supervised analysis using the four sites above in all tumor samples to determine TERTp status. For each tumor sample, we required a minimum coverage of 6 bp across all aforementioned sites in order to call TERTp status. Because all GBM samples lacked any coverage at the C169 site, this site was excluded in this tumor type. In order to call the mutation, we required at least 15% variant reads.

There was sufficient coverage for TERTp mutation calling in 903 samples, and we detected TERTp mutations in n = 183, including chr. 5: 1,295,228 C>T (n = 128), 1,295,250 C>T (n = 49), 1,295,242/1,295,243 C>T (n = 5) and 1,295,169 C>T (n = 1). Next, we combined these calls with n = 1077 TERTp calls obtained by targeted sequencing from adrenocortical carcinoma (n = 91), lower-grade glioma (n = 287), hepatocellular carcinoma (n = 196), melanoma (n = 119) and papillary thyroid carcinoma (n = 384) for a grand total of n = 1,807 TERTp mutation calls. In cases where WGS and targeted sequencing based calls disagreed, we selected the mutated variant.

Structural variant detection. We used SpeedSeq[66] to call structural variants in WGS-based BAM files. The SpeedSeq pipeline was built using our in-house pipeline building tool Flowr[63], and ran on a high-performance computing cluster at the University of Texas MD Anderson Cancer Center. BAM files were realigned using BWA-MEM[67]. The resulting calls in VCF format were filtered against matching normal. Only somatic events with at least three supporting reads were retained.

Because we were interested only in variants involving *TERT* (chr. 5:1253287–1315162, including 20 kb upstream of the TSS) and *ATRX* (chr. X:76760356–77041719), variants not overlapping one of these regions were excluded. An overview of all included variants can be found in **Supplementary Table 2**.

Fusion transcript detection. All fusion transcripts were detected using the pipeline of RNA-seq Data Analysis (PRADA) as previously described[29,30]. Briefly, chimeric fusion transcripts were detected from the realigned BAM files mapped to a combined genome and transcriptome reference on the basis of the evidence of both discordant read pairs and fusion-spanning reads, where discordant read pairs represent the paired read-ends that map uniquely to two protein coding genes, fusion-spanning reads mapped to the exon–exon junctions between two coding genes. Then the confidence of the detected fusion were evaluated on the basis of the number of junction-spanning reads and discordant read pairs, gene partner uniqueness, gene homology, open reading frame preservation, transcript allele fraction and presence of DNA breakpoints in adjacent distal regions. All fusions harboring *TERT* or *ATRX* transcripts were considered bona fide fusion calls on the basis of the criteria described previously and therefore included in this study. A summary of included fusions can be found in **Supplementary Table 2**.

TERT isoform detection. PRADA-aligned RNA-seq BAM files were available for n = 6,625 (97% of the extended set) samples, missing data on ovarian carcinoma (OV, n = 27) and prostate cancer (PRAD, n = 175).

We applied the mixture of isoforms (MISO) model[40] to each BAM to infer the relative abundance of full-length and β$^-$ *TERT* transcripts in each sample. Because the detection of isoforms is limited by the high GC-content of *TERT* combined with a 3′ bias by regular poly(A) enrichment RNA-seq protocol and unevenly spread expression across exons, we limited the analysis to exons 6–9 only and applied a coverage threshold of at least 12 reads within this region. These filters reduced the sample cohort to 24% (n = 1,201) of 5,001 tumors classified as *TERT*-expressing in the extended set. The exon model used to infer isoform abundance has been included in **Supplementary Table 4**. Briefly, the β$^-$ isoform skips exons 7 and 8, and samples demonstrating predominantly β$^-$ transcripts should therefore demonstrate low coverage in exons 7 and 8 relative to 6 and 9 and show junction-spanning reads between exons 6 and 9.

Structural variant breakpoint super-enhancer and ChIP-seq analysis. Super-enhancer coordinates were downloaded from dbSuper[26], accessed 15 March 2016. At this time, the database consisted of 65,950 super-enhancers across 107 tissue and/or cell types. After excluding low-complexity regions and

doi:10.1038/ng.3781

telomeres and centromeres, regions in this database spanned approximately 18% of the genome. We used the UCSC genome browser[68] to browse the database and inspected the position of each of the juxtaposed genomic coordinates. This analysis found overlapping super-enhancers in 11/17 TERTp structural variants, and this was significantly more than expected by chance (chi-square test $P = 0.001$).

Publicly available ChIP-seq data were downloaded from the NIH Roadmap Epigenomics Mapping Consortium[28]. This data set consists of 183 biological samples from multiple individuals, sequencing centers, tissue and/or cell types and was further consolidated into 111 unique epigenomes. We identified $n = 44$ tissue and/or cell types with H3K27ac (enhancer) sequencing, $n = 53$ tissue and/or cell types with H3K4me1 (enhancer) sequencing and $n = 55$ tissue and/or cell types with H3K27me3 sequencing (inactive promoter). Altogether, we collected data from 71 distinct tissue and/or cell types across these three marks. Collected BED files were converted to BAM format and a total number of reads was computed for each sample and mark.

ChIP-seq data were downloaded as previously described. For each structural variant proximal (adjacent to the TERTp) and distal breakpoint (juxtaposed to the TERTp), we calculated the number of reads mapping to each of three histone marks (H3K27ac, H3K27me3 and H3Kme1) individually within each tissue and cell/type. Breakpoint coordinates were flanked on either side to form a 2-kb bin. Read counts were normalized to RPKM on the basis of the total number of reads for each mark and sample. We then compared the RPKM between distal and proximal breakpoints for each histone mark for each set of breakpoint coordinates individually (**Supplementary Fig. 4d**), and subsequently with all breakpoints pooled together (**Fig. 2b**). *P* values were calculated using a two-sided *t*-test for each mark individually.

Telomerase activity signature inference. Microarray gene expression data from eight dedifferentiated liposarcoma samples were downloaded from the Gene Expression Omnibus (GEO GSE20559). We performed a differential expression analysis comparing four telomerase positive and four telomerase negative (ALT) samples and identified 1,302 genes associated with telomerase positive tumors (fold-change ≥ 1.5). Intersecting with 420 genes associated with embryonic stem cells[69], which are known to be telomerase positive, further refined this gene set and resulted in a list of 43 genes (**Supplementary Table 4**). Validation of the resulting gene signature using matched telomerase activity and RNA sequencing data from eleven urothelial cell carcinoma cell lines provided some evidence that this gene signature may be able to predict telomerase activity ($\rho = 0.58$, $P = 0.07$; **Supplementary Fig. 7a**). A detailed description of the strategy used to infer a telomerase activity signature may be found in the **Supplementary Note**.

Candidate gene selection and telomere length association. To narrow down the list of candidate genes and increase the FDR-adjusted significance threshold, we took the union of the MutSig2CV[70] gene list (FDR < 0.05, downloaded and compiled from publicly available Firehose analyses) from all cancer types, and combined this list with genes within significant GISTIC 2.0 peaks (FDR < 0.10) after filtering peaks larger than 1 Mb.

Each gene was annotated as putative tumor suppressor or putative oncogene depending on the mutational patterns (**Supplementary Note**) or whether it was found within an amplification or deletion peak. In those cases where the mutation-based classification did not agree with the copy number–based classification (e.g., a gene with highly frequent hotspot mutations was found in a deletion peak), we selected to prefer the mutation-based classification. Genes lacking classification and genes where the mutation-based classification demonstrated contradicting evidence (e.g., genes that could be classified as either tumor suppressors or oncogene) were dropped, unless the gene was present on a list of telomere-related genes, in which case it was annotated as a significantly mutated gene. The list of candidate genes was further reduced by removing multiple genes from the same peak, additionally leaving only those genes closely tied to telomere function when multiple genes present in the peak (e.g., shelterin complex genes) and genes that were also found in the MutSig2CV gene list. The final list of genes can be found in **Supplementary Table 5**.

For each candidate gene and sample, we determined whether it was altered depending on the classification. For genes classified as tumor suppressors, we classified samples as altered when they showed a somatic mutation, a focal deletion or both. For genes classified as oncogenes, we classified the sample as altered when they showed a somatic mutation, focal amplification or both. For genes classified as significantly mutated gene only somatic mutations were counted. Each candidate gene was then correlated to the TL ratio in all samples and within diseases individually. This analysis was repeated specifically within samples classified as double wild type.

Statistical analysis. Summary statistics of telomere length are provided in mean, s.d. and range by tumor type and tissue type. Telomere length was transformed to the logarithmic scale owing to skewness.

A linear mixed model was used to estimate mean telomere length of tumor and normal samples for each tumor type. Age, center and sex were included as covariates in all models. Interactions between disease type and center, gender and age were assessed and none were significant. The 'patient' variable was modeled as a random effect in the mixed model to account for correlation between samples from the same patient.

Linear regression was used to correlate TL ratio to various biomarkers. Backwards elimination was then used to select the best model with only significant factors. All tests were two-sided, and *P* values ≤ 0.05 were considered statistically significant.

Spearman correlation was used to associate *TERT* expression and methylation and to correlate gene expression and TL. A two-sided *t*-test was used to correlate somatic alterations and TL. *P* values were adjusted for multiple testing using the Benjamini–Hochberg FDR. An FDR of 0.25 or lower was considered statistically significant.

Survival curves were estimated and plotted using the Kaplan–Meier method. Log-rank tests were used to compare curves between groups. Univariate and multivariate Cox modeling were used to compute hazard ratios and confidence intervals.

All statistical analyses were carried out using SAS version 9 (SAS Institute) and R (R Foundation for Statistical Computing).

Data availability. A complete overview of all gene fusions can be found on our fusion web portal (http://tumorfusions.org/). Raw RNA and DNA sequencing data ($n = 35,978$ samples) were downloaded from CGHub[62]. Processed clinical, gene expression, DNA methylation, copy number segmentation and mutation data were downloaded and compiled from previously published TCGA papers, the Firehose data portal (Broad Institute, http://gdac.broadinstitute.org/) and the TCGA data portal (NIH). All TCGA data are publicly available through NCI Genomic Data Commons (https://gdc.cancer.gov/). Cases used in this study and their telomere length estimates are listed in **Supplementary Table 1**.

62. Wilks, C. *et al.* The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database (Oxford)* https://dx.doi.org/10.1093/database/bau093 (2014).
63. Seth, S. *et al.* Flowr: robust and efficient pipelines using a simple language-agnostic approach. Preprint at http://biorxiv.org/content/early/2015/10/22/029710 (2015).
64. Mermel, C.H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
65. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
66. Chiang, C. *et al.* SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
67. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint https://arxiv.org/abs/1303.3997 (2013).
68. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
69. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* **40**, 499–507 (2008).
70. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).