bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The Telomere Length Landscape of Prostate Cancer

Julie Livingstone[1,2,3,4], Yu-Jia Shiah[5], Takafumi N. Yamaguchi[1,2,3,4], Lawrence E. Heisler[5], Vincent Huang[5], Robert Lesurf[5], Tsumugi Gebo[1,2,3,4], Benjamin Carlin[1,2,3,4], Stefan Eng[1,2,3,4], Erik Drysdale[5], Jeffrey Green[5], Theodorus van der Kwast[6,7], Robert G. Bristow[6,8,9], Michael Fraser[6], Paul C. Boutros[1,2,3,4,8,10]

[1] Department of Human Genetics, University of California, Los Angeles, CA 90095, USA

[2] Department of Urology, University of California, Los Angeles, CA 90024, USA

[3] Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, CA 90024, USA

[4] Institute for Precision Health, University of California, Los Angeles, CA 90024, USA

[5] Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada

[6] Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2M9, Canada

[7] Department of Pathology, Laboratory Medicine Program, University Health Network, Toronto, ON M5G 2C4, Canada

[8] Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada

[9] Manchester Cancer Research Centre, Manchester, United Kingdom

[10] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

Corresponding Author: Dr. Paul C. Boutros, 12-109 CHS; 10833 Le Conte Avenue; Los Angeles, CA 90095. Phone: 310-794-7160; Email: pboutros@mednet.ucla.edu

Keywords: Telomere length; prostate cancer; genomic instability; integration

Word Count of the text: 6,955 (excluding references)

Word Count of the abstract: 198

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Replicative immortality is a hallmark of cancer, and can be achieved through telomere lengthening and maintenance. We report telomere lengths (TLs) of 392 localized prostate cancer tumours and characterize their relationship to genomic, transcriptomic and proteomic features. Shorter tumour TLs were associated with elevated genomic instability, including single-nucleotide variants, indels and structural variants. Genes involved in cell proliferation and signaling were correlated with tumour TL at all levels of the central dogma. TL was also associated with multiple clinical features of a tumour. Longer TLs in non-tumour samples were associated with a lower rate of biochemical relapse after definitive local therapy. Our analysis integrates multi-omics data to illuminate the relationship of specific genomic alterations in a tumour and TL in prostate cancer. Although the role of telomere length in cancer has been well studied, its association to genomic features is less well known. We describe the multi-level integration of telomere length, genomics, transcriptomics and proteomics in localized prostate cancer. **Patient Summary** We examined the association between telomere length and multiple omics-level data in prostate cancer. We observed that traditional telomere mutations are rare in prostate cancer and that telomere length is associated with multiple measure of genomic instability.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Telomeres, which make up the ends of chromosomes, consist of a repeat TTAGGG sequence[1] along with bound proteins known as shelterin[2]. Telomeres protect chromosomal ends from degradation by the DNA double-strand break (DSB) response pathway. Due to the linearity of chromosomes and chromosomal replication, telomeres are shortened by approximately 50 bp during mitosis[3]. When telomeres become substantially shortened, cell cycle progression halts and cells enter replicative senescence; further replication leads to cellular crisis and eventually cell death[4]. Telomere maintenance and lengthening is essential for cancer cell proliferation and enables replicative immortality: a fundamental hallmark of cancer[5]. Telomere regulation occurs through two known mechanisms: activation of telomerase or alternative lengthening of telomeres (ALT) which relies on homology-directed DNA replication[6].

Despite the pan-cancer studies analyzing the telomere length from various tumour types[7,8], the role of telomere maintenance in individual tumour types is poorly understand. Moreover, the relationship between telomere length and biologically-relevant genomic indices, such as percentage of the genome altered (PGA; [9,10], and other measures of mutational density has not been assessed, nor has the association between telomere length and clinical outcome in prostate cancer.

We and others have described the genomic, transcriptomic and proteomic landscape of localized, non-indolent prostate cancer[11–18]: the most frequently diagnosed non-skin malignancy in North American men (~250,000 new cases per year). Localized prostate cancer is a C-class tumour[19], characterized by a paucity of driver single nucleotide variants (SNVs) and a relatively large number of structural variants (SVs), including copy number aberrations (CNAs) and genomic rearrangements (GRs). Several of these aberrations, including mutations in *ATM* and amplifications of *MYC* – which drive DSB repair and cell proliferation, respectively – are associated with significantly reduced time to biochemical and metastatic relapse after local therapy[20]. Intriguingly, both of these mutations have also been associated with telomere maintenance[21,22] and telomere shortening – relative to adjacent epithelium[23]. Similarly an interaction between hypoxia, dysregulated *PTEN*, *TERT* abundance and telomere shortening was recently illustrated[15]. Despite this, no well-powered study exists evaluating the association between telomere length, somatic features and clinical outcome in prostate cancer.

To fill this gap, we quantify the telomere length and somatic mutational landscapes of 392 localized prostate tumours. We explore associations between telomere length and the tumour methylome, transcriptome and proteome. Using rich clinical annotation, we further assessed the relationship between telomere length and outcome. Taken together, these data establish the role and regulation of telomere length in localized prostate cancer, and establish clear links between telomere maintenance and drivers of prostate cancer development and clinical aggression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Association of telomere length with somatic nuclear driver events

To investigate the impact of telomere length (TL) on the clinico-genomics of prostate tumours, we exploited whole genome sequencing (WGS) of 392 published tumour-normal pairs[11–14,24]. We estimated both tumour and non-tumour (blood or adjacent histologically normal tissue) TLs for each sample using TelSeq v0.0.1[25] and TelomereHunter (v1.0.4)[26]. After quality control, 381 samples were retained for further analysis (see **Methods**). All tumours were treatment-naive, and detailed clinical information was collected and is available in **Supplementary Table 1**. The cohort consisted of 11% ISUP Grade Group (GG) 1, 52% GG2, 33.5 % GG3, 6.8% GG4 and 3.4% GG5. For the majority of samples, the tumour was confined to the prostate (6.5% T1, 53.0% T2, 40.0% T3, 0.5% T4). The mean tumour coverage was 73.1x ± 20.6x; the mean non-tumour coverage was 44.1x ± 13.4x. Median clinical follow-up time was 7.46 years. TLs for each sample, along with clinical and genomic summary data are in **Supplementary Table 1**. Non-tumour TLs varied dramatically across individuals, ranging from 2.10 kbp to 15.0 kbp, with a median of 4.52 ± 1.35 kbp. Blood TLs (n = 341) were shorter than those in adjacent normal (n = 40; Mann Whitney U test; P = 2.80 x 10^{-10}; **Supplementary Fig. 1A**). By contrast, tumour TLs varied less but were significantly shorter, ranging from 1.03 to 6.45 kbp with a median of 3.36 ± 0.87 kbp. Non-tumour TLs were not associated with sequencing coverage (**Supplementary Fig. 1B**). Tumour TLs were independent of tumour purity but there was a weak negative correlation between coverage and TL driven by some samples sequenced with over 100x coverage (**Fig. 1D**; **Supplementary Figs. 1B-C**). Tumour and non-tumour TLs estimates from TelSeq and TelomereHunter were highly correlated (**Supplementary Fig. 1D**) so we decided to use TelSeq estimates throughout. To account for batch effects and the differences in blood and normal adjacent tissue, a linear model was fit and TLs were adjusted (**Supplementary Figs. 1E-F)**. TL ratios (tumour TL / non-tumour TL) were calculated to further reduce any effects caused by co-founding of sequencing method. Tumour and non-tumour TLs were positively correlated with one another (ρ = 0.37, P = 7.30 x 10^{-14}, **Fig. 1A**). As expected, TL ratio was positively correlated with tumour TL (ρ = 0.63, P < 2.2 x 10^{-16}; **Fig. 1B**) but negatively correlated with non-tumour TLs (ρ = -0.40, P < 2.2 x 10^{-16}; **Fig. 1C**). There was no difference in TL ratio between localized and metastatic samples (n = 101; P = 0.95; Mann Whitney U test; **Supplementary Fig. 1G**).

To assess whether tumour TL was related to any specific genomic property of a tumour, we evaluated a set of driver mutations previously identified in prostate cancer[14]. The relationship of each of these features with tumour TL is shown in **Fig. 1D**. While tumour TL was not associated with any known prostate cancer-related genomic rearrangement (GR) or single nucleotide variant (SNV) at current statistical power, samples with *CHD1*,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

RB1 or *NKX3-1* deletions had shorter tumour TL (**Fig. 1D; Fig. 5A**). By contrast, TL was closely associated with multiple measures of genomic instability. Tumours with shorter TLs had an elevated number of SNVs (ρ = -0.27, $P = 5.78 \times 10^{-8}$; **Fig. 2A**), indels (ρ = -0.32, $P = 2.83 \times 10^{-10}$; **Fig. 2C**) and GRs (ρ = -0.12, $P = 1.6 \times 10^{-2}$; **Fig. 2E**), as well as higher PGA (ρ = -0.21, $P = 3.95 \times 10^{-5}$; **Fig. 2G**), suggesting tumours with shorter telomeres accrue more mutations of all types without strong selective pressures for specific ones.

To determine whether these associations with somatic features were also related to an individual's non-tumour cells, we related each somatic feature against the TL ratio (tumour TL / non-tumour TL). Similar to tumour TL, the TL ratio did not significantly differ between samples with any of the recurrent prostate cancer-related GRs or CNAs but samples with a somatic SNV in the gene *SPOP* had smaller TL ratios (**Supplementary Fig. 2)**. We identified significant correlations between somatic genomic instability measures and TL ratio. Tumours with an elevated number of SNVs (ρ = -0.15, $P = 4.20 \times 10^{-3}$; **Fig. 2B**), indels (ρ = -0.18, $P = 2.97 \times 10^{-4}$; **Fig. 2D**), GRs (ρ = -0.22, $P = 1.08 \times 10^{-5}$; **Fig. 2F**) and PGA (ρ = -0.13, $P = 1.69 \times 10^{-2}$; **Fig. 2H**) had smaller TL ratios.

We also assessed the association of telomere length with chromothripsis using published ShatterProof[27] scores from a subset of samples in this cohort (n = 170)[14]. There was no correlation between scores representing chromothripsis events in either tumour TL (ρ = 0.06, $P = 0.43$) or TL ratio (ρ = 0.02, $P = 0.80$).

Fusion events are associated with telomere length

When telomeres shorten beyond a certain length, double strand break repair is activated and cell cycle progression is arrested *via* the *TP53* pathway[28]. Failure to block cell growth can lead to telomere crisis and subsequent translocations, chromothripsis or chromosome fusions[29]. We explored the association of TL and the number of gene fusions present in a tumour. There was a negative correlation between the number of gene fusions and tumour TL (ρ = -0.26; $P = 2.18 \times 10^{-3}$) but no correlation with TL ratio (**Figs. 2I-J**). In a previous study, 47 recurrent gene fusions were discovered from matched RNA-Sequencing data[18]. Differences in tumour TL and TL ratio between samples with a gene fusion and those without were investigated for each of these recurrent fusions. No gene fusions were associated with TL ratio, but the PCAT1:CASC21 gene fusion was significantly associated with tumour TL (Mann Whitney U test; $Q = 2.07 \times 10^{-4}$; **Supplementary Fig. 3** and **Supplementary Table 2**). Tumours with this fusion had shorter tumour telomeres (mean = 3.3 kbp) than those without (mean = 3.8 kbp). These data suggest that the number of fusions and specifically the long non-coding RNA *PCAT1*, which promotes cell proliferation, is related to tumour TL.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Proliferation rate is not associated to telomere length

The rapid reproduction or proliferation of a cell should reduce the telomere length in dividing tumour cells. To test this, we investigated the correlation of TL with MKI67 abundance levels and a previously published proliferation score[30]. Surprisingly, there was no association between either tumour TL (ρ = -0.14; *P* = 0.11) or TL ratio (ρ = -0.09; *P* = 0.30) and MIK67 RNA abundance (**Supplementary Fig. 3D-E**). Similarly, there was no association between proliferation scores and tumour TL (ρ = 0.01; *P* = 0.91) or TL ratio (ρ = -0.05; *P* = 0.54; **Supplementary Fig. 3F-G**). This suggests that there is a more complex relationship between proliferation and TL at play.

The role of *TERT* in prostate cancer

A pan-cancer study reported that *TERT* alterations including promoter mutations, amplifications and structural variants were seen in approximately 30% of all cancers[7]. In our cohort, 10% of samples had *TERT* amplifications, 11% had *TERC* amplifications, ~1% had *TERT* structural variants and no samples had *TERT* SNVs or gene fusions. *TERT* mutations were seen less frequently in other localized prostate cancer datasets, 1.7% (17/1,013; [31] and 0.6% (2/333; [13]), and in a metastatic dataset 3% (5/150; [20], likely reflecting the early-stage of our cohort. Mutations in *ATRX* and *DAXX*, which have been correlated with longer telomeres[32], were rare in our cohort: only two samples harboured a CNA in *DAXX*, and only four samples had an alteration in *ATRX*.

Tumour *TERT* RNA abundance was not correlated with tumour TL or TL ratio (**Fig. 3A**). Samples with higher *TERT* RNA abundance had fewer GRs (ρ = -0.17; *P* = 4.79 x 10^{-2}; **Fig. 3B**), but there was no correlation between *TERT* abundance and SNV count (ρ = -0.04, *P* = 0.67; **Fig. 3C**), indel count (ρ = -0.04, *P* = 0.132; **Fig. 3D**) or PGA (ρ = -0.13, *P* = 0.679; **Fig. 3E**). The abundance of *TERC*, the telomerase RNA component, was negatively correlated with tumour TL (ρ = -0.24; *P* = 4.55 x 10^{-3}; **Supplementary Fig. 4A**) but there was no correlation with TL ratio or GR count (ρ = 0.12; *P* = 0.145; **Supplementary Fig. 4B**). *TERC* abundance was positively correlated with SNV count (ρ = 0.23; *P* = 7.34 x 10^{-3}; **Supplementary Fig. 4C**), indel count (ρ = 0.34; *P* = 4.88 x 10^{-5}; **Supplementary Fig. 4D**) and PGA (ρ = 0.26; *P* = 1.90 x 10^{-3}; **Supplementary Fig. 4E**). *TERT* and *TERC* abundances were not correlated (ρ = 0.02; *P* = 0.794). These data suggest that TERT signaling is not significantly abrogated in localized prostate cancer either by somatic aberrations or through gene expression changes.

To explore the relationship of *TERT* RNA abundance and tumour TL further, we considered known activating transcription factors. Transcription of *TERT* can be activated by *MYC* and *SP1* and repressed by *AR*[33]. *MYC* amplifications occur in 14.5% of our samples (51/351; **Fig. 1D**), while *SP1* CNAs are rare (3/351). *TERT* and *MYC* mRNA abundance was positively correlated (ρ = 0.27; *P* = 1.46 x 10^{-3}) but *MYC* abundance was unrelated to tumour TL (**Supplementary Fig. 5A)**. Contrastingly, there was a positive correlation between tumour TL length and *SP1* abundance (ρ = 0.23; *P* =

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

6.84 x 10^{-3}) but no significant correlation between *SP1* and *TERT* abundance (**Supplementary Fig. 5B**). We did not observe any statistically significant correlations between *AR* and *TERT* abundance, or tumour TL (**Supplementary Fig. 5C**). The direct relationship of these transcription factors on *TERT* is hard to elucidate because of the low measured abundance of *TERT*. Nonetheless, the abundance of *SP1* and *AR* appear to positively and negatively affect tumour TL, respectively.

To determine whether *TERT* was being regulated epigenetically, we first investigated the correlation between its methylation status and its RNA abundance using 91 annotated sites. We identified one CpG site with a significant negative correlation and two with significant positive correlations (Spearman's correlation; $Q < 0.05$; $|\rho| > 0.2$; **Fig. 3F)**. Further, 31% (28/91) of *TERT* CpGs sites were significantly correlated to telomere length: 7 positively and 21 negatively (Spearman's correlation; $Q < 0.05$; $|\rho| > 0.2$; **Fig. 3F**). This strongly suggests that methylation of *TERT* may impact *TERT* abundance and tumour TL.

Candidate regulators of prostate tumour telomere length

Evidence of correlation between methylation and tumour TL in *TERT* led us to investigate the role of methylation on TL genome-wide. For each gene, we considered the CpG site most associated to its mRNA abundance (see **Methods**) and related that to tumour TL (n = 241). Methylation of almost half of all genes (46%; 7,088/15,492) was significantly correlated with tumour TL (Spearman's correlation; $Q < 0.05$; **Supplementary Table 3)**. Similarly, almost a third of genes showed transcriptional profiles associated with tumour TL (32%; 4,520/13,956). No proteins were significantly associated with tumour TL after FDR adjustment although 9.3% proteins showed correlation to tumour TL before adjustment (n = 548/5,881; Spearman's correlation; $P < 0.05$). There were 112 genes with methylation, transcription and proteome correlations to telomere length. Remarkably, these showed no functional enrichment. Several genes showed methylation positively correlated with tumour TL but negatively correlated with RNA and protein abundance (**Fig. 4A**), suggesting suppression of tumour TL elongation. One such gene is the oncogene *AKT1*, which regulates processes including cell proliferation, survival and growth[34]. High *AKT1* abundance may indicate an elevated proliferation and therefore shorter telomeres.

We also identified genes whose methylation was negatively correlated with tumour TL but positively correlated with RNA and protein abundance suggesting promotion of telomere elongation (**Fig. 4B**). These included *SLC14A1*, a membrane transporter that mediates urea transport, and *ITGA3*, an integrin that functions as a cell surface adhesion molecule. We used gprofiler2[35] to identify pathways enriched in genes with methylation or transcriptomic profiles that are correlated with tumour TL using KEGG pathways[36]. We identify 16 pathways enriched in genes with methylation profiles and 16 pathways that were enriched in genes with transcriptomic profiles that were correlated

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with tumour TL (**Supplementary Fig. 6A)**. To reduce false positives and account for crosstalk between pathways, we applied a crosstalk correction method[37,38]. The crosstalk matrix (**Supplementary Fig. 6B**) identified overlap between the cancer related pathways, and after crosstalk adjustment only one pathway remained enriched in genes with transcriptomic profiles that were correlated to tumour TL: hsa04519 (Focal adhesion; **Supplementary Fig. 6C**).

We similarly investigated whether TL ratio was associated with methylation and found that the methylation levels of 33.7% (5,218/15,492) of genes were significantly correlated with TL ratio (Spearman's correlation; $Q < 0.05$; **Supplementary Table 4)**. Surprisingly, fewer than 1% (n = 53/13,958) of genes with overlapping data also had a significant correlation between RNA abundance and TL ratio and none between protein abundance and TL ratio (Spearman's correlation; unadjusted $P < 0.05$). These results suggest that tumour TL, not TL ratio, is associated with tumour gene expression.

Association of telomere length and specific copy number aberrations

Since prostate tumour gene-expression and clinical behaviour is predominantly driven by CNAs[14,19] we next investigated their role in TL. As noted above (**Fig. 1D**), driver CNAs were largely unassociated with tumour TL (**Fig. 5A;** white background) or TL ratio (**Fig. 5B**; white background). We therefore considered copy number changes genome-wide for associations with TL. We identified 24 loci encompassing 35 genes in which there was a significant difference in tumour TL in samples with a copy number change compared to those without (Mann-Whitney U test; $Q < 0.05$; **Supplementary Table 5** and **Fig. 5A**). We also identified 128 loci encompassing 319 genes in which there was an association between copy number status and TL ratio (Mann-Whitney U test, $Q < 0.05$; **Supplementary Table 7**). For example, tumours with deletions in DNA methyltransferase 1, *DNMT1*, had smaller TL ratios ($Q = 0.028$, effect size = 0.11, **Fig. 5B**). An opposing trend was seen in the chromatin organization gene, *PRDM16* ($Q = 0.027$, effect size = 0.15) and the membrane metallo-endopeptidase gene, *MMEL1* ($Q = 0.027$, effect size = 0.14; **Fig. 5B**), where amplifications resulted in smaller TL ratios. This analysis highlights that copy number aberrations are more associated with TL ratio (change in length from non-tumour TL to tumour TL) than absolute tumour TL.

We also explored CNAs in genes comprising the telomere complex (*TERF1*, *TERF2*, *TERF2IP*, and *POT1*), shelterin interacting proteins (*PINX1* and *RTEL1),* and the components of telomerase (*TERT* and *TERC*). There were no differences in the tumour TL (**Supplementary Fig. 7A**) or TL ratio (**Supplementary Fig. 7B**) between samples with and without a CNA in these genes.

Next, we compared TL across previously identified CNA subtypes. There was no difference in tumour TL ($P = 0.53$; one-way ANOVA) or TL ratio ($P = 0.78$; one-way ANOVA) in the four CNA subtypes identified from aCGH arrays and associated with prognosis[9] (**Supplementary Fig. 8A-B**). There was an association between TL ratio

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and the six CNA subtypes (P = 2.12 x 10^{-2}; one-way ANOVA) identified from 284 OncoScan SNP arrays[14] but not with tumour TL (**Supplementary Fig. 8C-D**). Samples in subtype C5, which was defined by amplifications in genes near the end of chromosomes had smaller TL ratios than C3 (defined by an 8p deletion and an 8q amplification) and C4 (defined as having a quiet CNA profile). A smaller TL ratio in the samples from subtype C5 indicates that the non-tumour TL length was longer than in the tumour TL (**Supplementary Fig. 8E**): the consequences of this remain to be elucidated.

Clinical correlates of telomere length

The clinical features of a tumour can have prognostic value, and have been associated with the genomic features of tumours[14]. Higher serum abundance of prostate specific antigen (PSA), higher ISUP Grading and tumour size and extent are all associated with worse outcome. Therefore, we considered whether there was interplay between TL and the clinical features of a tumour. Tumour TL was not significantly correlated to age, (ρ = -0.10, P = 5.8 x 10^{-2}; **Fig. 6A**) but there was a significant positive correlation between age at diagnosis and TL ratio (ρ = 0.11, P = 2.53 x 10^{-2}; **Fig. 6B**). Tumour TL was shorter than non-tumour TL in younger patients. This could be related to the aggressiveness of early onset prostate cancers, which is characteristic of tumours in younger men[24]. There was a negative correlation between pre-treatment PSA levels between both tumour TL (ρ = -0.16, P = 2.23 x 10^{-3}) and TL ratio (ρ = -0.19, P = 1.70 x 10^{-4}; **Figs. 6C-D**). Neither tumour TL nor TL ratio was associated with ISUP Grade (**Figs. 6E-F**). Surprisingly, tumour TL was shorter in smaller tumours (T1) than larger tumours (T2 or T3; one-way ANOVA, P = 2.2 x 10^{-2}; **Fig. 6G**) but this can be explained by the higher average age of patients with T1 tumours (mean = 71.3) compared to other T categories (mean = 62.0). Accordingly, there was no association between TL ratio, which controls for patient age, and T category (P = 0.29; **Fig. 6H**).

Telomerase activity and TL has been proposed to have clinical utility at three different stages; diagnosis, prognosis and treatment[33]. TL from biopsies has been correlated with progression to metastasis and disease specific death[39]. As well, TL from leukocytes has been associated with poor survival[40,41]. We explored if tumour TL, non-tumour TL or TL ratio were associated with biochemical relapse (BCR), an early surrogate endpoint in intermediate-risk prostate cancer. Cox proportional hazards (Cox PH) models were fit, splitting patients (n = 290) into two groups based on their TL with increasing cutoff thresholds (50 bp each time; **Supplementary Figs. 9A-C**). From this outcome-oriented optimal cut-point analysis we discovered that samples with non-tumour TL less than 3.9 kbp had a higher rate of BCR than samples with longer TLs (HR = 2.02, P = 1.6 x 10^{-3}; **Fig. 6I**). Non-tumour TL is associated with survival independent of PGA (Cox PH model, P = 0.02). There was no association between tumour TL and BCR (**Fig. 6J**), but there was an association between TL ratio and BCR , where samples with a TL ratio greater than 0.65 had a lower rate of BCR (HR = 0.42, P = 2.6 x 10^{-3}; **Fig. 6K**). We also

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

considered TL as a continuous measurement and fit Cox PH models using tumour TL, non-tumour TL and TL ratio. Again, there was no association between continuous tumour TL and BCR but there was an association between non-tumour TL (HR = 0.768, P = 0.014) and TL ratio (HR = 1.71, P = 0.031; **Supplementary Fig. 9D**). These results suggest that non-tumour TL and TL ratio are weakly prognostic, and thus may reflect host factors that may influence patient risk categorization.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

These data emphasize the relationship of genomic instability and TL. Genomic instability has previously been linked with poor outcome in prostate cancer[9,14] and TL shortening could be the cause of some of this instability. Telomere shortening has been implicated as an early event in prostate cancer due to evidence of shortened telomeres observed in a precursor histopathology, high-grade prostatic intraepithelial neoplasia[42,43]. Since cellular proliferation in prostate cancer is increased by seven fold compared to normal prostatic epithelial cells[33], telomeres in these dividing cells will shorten with each cell division. There is no evidence that primary prostate cancer exhibits ALT lengthening[23] therefore the vast majority, if not all tumours, activate telomerase for telomere maintenance. We did not observe any *TERT* promoter mutations in our cohort but there are strong negative correlations between methylation probes in the promoter of *TERT* and tumour TL. This may be a proxy for telomerase activity since DNA methylation impedes transcription.

We see an unexpected divergence between somatic molecular features associated with TL ratio and those with tumour TL. Specifically, measures of genomic instability are linked to TL ratio (which represents the ratio between non-tumour TL and tumour TL) while specific CNAs, GRs, and SNVs are not (**Fig. 1** and **Supplementary Fig. 2**). This suggests that during the progression of cells from normal to cancerous, non-tumour TL may influence tumour genomics, where tumours with shorter TL experience more genomic instability. Alternatively, a common factor may be influencing during this epoch of the tumour's evolution. Once tumours are formed, it is the specific mutations within the cell that are more associated with tumour TL. This may be due to mutations in cell division and growth regulating genes such as *ATK1* and *SPOP*, which increases the number of divisions in the tumour and thereby shortens tumour telomeres. Further evidence of this hypothesis is seen in tumours with *PCAT1* fusions, where tumours with this fusion had shorter tumour TL than samples without it[44].

One limitation in the estimation of TL using short-read whole genome sequencing is the difficulty in estimating chromosome specific telomere lengths. Junction spanning reads from paired-end experiments, in which one read maps within the first or last band of the chromosome and the other read maps within the telomere region, are scarce. Further studies should be performed using long read sequences, in which these regions may have more coverage and can be used to determine chromosome specific shortening and its association to specific genomic events or biochemical relapse.

These data highlight the complicated relationship between telomere length in both tumour and non-tumour cells, and molecular and clinical tumour phenotypes. They highlight the need for increased study of telomere length across cancer types, and for long-read sequencing to introduce chromosome-specific analyses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

The authors thank all members of the Boutros lab for insightful commentary and technical support.

Funding

This study was conducted with the support of the Ontario Institute for Cancer Research to PCB through funding provided by the Government of Ontario. This work was supported by Prostate Cancer Canada and is proudly funded by the Movember Foundation - Grant #RS2014-01. Dr. Boutros was supported by a Terry Fox Research Institute New Investigator Award and a CIHR New Investigator Award. This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-125). This work was supported by the NIH/NCI under award number P30CA016042, by an operating grant from the National Cancer Institute Early Detection Research Network (U01CA214194) and by support from the ITCR (U24CA248265).

Author contributions

Formal Analysis: JL. Methodology: JL, SE, ED. Data curation: JL, YSY, TNY, LEH, VH, RL, TG, BC. Visualization: JL, JG. Supervision: MF., TvdK, RGB, PCB. Conceptualization, Supervision: PCB, MF, RGB. Pathology Reviews: TvdK. Writing - original draft: JL, PCB. Writing – review & editing: JL, TNY, VH, RL, MF, PCB. Approved the Manuscript: All Authors

Declaration of Interest and Financial Disclosures

All authors declare that they have no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Figure 1. Tumour telomere length (TL) is associated with genomic features. A, Correlation between tumour TL and non-tumour TL. **B,** Correlation between tumour TL and TL ratio (tumour TL / non-tumour TL). **C,** Correlation between non-tumour TL and TL ratio. **D,** Tumour TL is ranked in descending order of length (kbp; top bar plot). The association of tumour TL and measures of mutational burden, TMPRSS2:ERG (T2E) fusion status, as well as known prostate cancer genes with recurrent CNAs, coding SNVs, and GRs are shown. Bar plots to the right indicate the statistical significance of each association (see **Methods)**.

Figure 2. Mutational landscape differs with telomere length. A-B, Correlation between the number of SNVs and **A,** tumour TL and **B,** TL ratio. **C-D,** Correlation between the number of indels and **C,** tumour TL and **D,** TL ratio. **E-F,** Correlation between the number of GRs and **E,** and tumour TL and **F,** TL ratio. **G-H,** Correlation of percentage of the genome altered (PGA) and **G,** tumour TL and **H,** TL ratio. **I-J,** Correlation between the number of fusions and **I,** tumour TL and **J,** TL ratio. Orange dots indicate tumour TL while green dots indicate TL ratio. Spearman's ρ and *P* values are displayed.

Figure 3. The genomic correlates of *TERT* abundance. A, Correlation of *TERT* RNA abundance with tumour TL and TL ratio. Orange dots indicate tumour TL while green dots indicate TL ratio. Spearman's ρ and *P* values are displayed. **B-E,** Correlation of *TERT* abundance and **B,** the number of GRs, **C,** number of SNVs, **D,** number of indels, and **E,** PGA. Spearman's ρ and *P* values are displayed. **F**, Spearman's correlation of significantly associated methylation probes with RNA abundance and tumour TL. Blue dots indicate a positive correlation while orange dots indicate a negative correlation. Probes within the promoter are labeled in red while the rest are located in the gene body. Dot size indicated the magnitude of correlation. Background colour indicates unadjusted *P* values. Methylation probes are ordered by their correlation between *TERT* RNA abundance from negative to positive.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4. Association of methylation, RNA abundance, protein abundance and telomere length. A, Positive correlation of methylation and tumour TL, but negative correlation of RNA and protein abundance. Top panels in light blue represent methylation, middle panels in blue-grey represent RNA abundance and the bottom panels in purple represent protein abundance. Darker purple dots represent undetected, imputed abundance measures. Spearman's ρ and *P* values are displayed. **B,** Negative correlation of methylation and tumour TL, but positive correlation of RNA and protein abundance. Top panels in light blue represent methylation, middle panels in blue-grey represent RNA abundance and the bottom panels in purple represent protein abundance. Darker purple dots represent undetected, imputed protein abundance measures. Spearman's ρ and *P* values are displayed.

Figure 5. Telomere length differs by copy number status. A, Difference in tumour TL between samples with a copy number aberration and those without in prostate cancer related genes and associated genes. **B,** Difference in TL ratio between samples with a copy number aberration and those without in prostate cancer related and associated genes. *Q* values are from a Mann-Whitney U test and are bolded when significant (FDR < 0.05). Colour of the points indicate copy-number status of the gene: amplification (red), deletion (blue), or neutral (black).

Figure 6. Telomere length is associated with clinical features and biochemical relapse. A-B, Correlation of age at diagnosis with **A,** tumour TL and **B,** TL ratio. Spearman's ρ and *P* values are displayed. **C-D,** Correlation of pre-treatment PSA with **C,** tumour TL and **D,** TL ratio. Spearman's ρ and *P* values are displayed. **E-F,** Association of ISUP grade with **E,** tumour TL and **F,** TL ratio. *P* value is from an one-way ANOVA. **G-H,** Association of T category with **G,** tumour TL and **H,** TL ratio. *P* value is from an one-way ANOVA. On all plots, green indicates TL ratio, while orange indicates tumour TL. **I-K,** Cox proportional hazard models were created for **I,** non-tumour TL**, J,** tumour TL and **K,** TL ratio with BCR as the endpoint. Samples were split into two groups based on the optimal cut point analysis (**see Methods**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Patient cohort

Published whole-genome sequences of tumour and matched non-tumour samples were downloaded from public repositories (phs000447.v1.p1[11], phs000330.v1.p1[12], EGAS00001000900[14], phs000178.v11.p8[13], EGAS00001000400[24], phs001648.v2.p1[45]). For RadP patients, BCR was defined as two consecutive post-RadP PSA measurements of more than 0.2 ng/ml (backdated to the date of the first increase). If a patient has successful salvage radiation therapy, this is not BCR. If PSA continues to rise after radiation therapy, BCR is backdated to first PSA > 0.2. If patient gets other salvage treatment (such as hormones or chemotherapy), this is considered BCR. Tumour cellularity and Gleason grades were evaluated independently by two genitourinary pathologists on scanned haematoxylin- and eosin-stained slides. Serum PSA concentrations (ng/mL) are reported according to the reading at the time of diagnosis. Cellularity was also determined *in silico* from OncoScan SNP arrays via qpure (v1.1)[46].

Whole-genome sequencing data analysis

Raw sequencing reads were aligned to the human reference genome, GRCh37, using BWA-mem (version > 0.7.12; [47] at the lane level. Picard (v1.92; http://broadinstitute.github.io/picard/) was used to merge the lane-level BAMs from the same library and mark duplicates. Library level BAMs from each sample were also merged without marking duplicates using Picard. Local realignment and base quality recalibration was carried out on tumour/non-tumour pairs together using the Genome Analysis Toolkit (GATK; > version 3.4.0; [48]. Tumour and non-tumour sample level BAMs were extracted, headers were corrected using SAMtools (v0.1.9; [49], and files were indexed with Picard (v1.92).

Computational telomere length estimation

Tumour and non-tumour telomere lengths were estimated using TelSeq (v0.0.1; [25] and TelomereHunter (v1.0.4)[26] on BAM files generated using bwa-mem (version > 0.712; [47] and GATK (version > 3.4.0; [48]. TelSeq estimates telomere length with the following formula $l = t_k sc$, where t_k is the abundance of telomeric reads (reads that contain k or more TTAGGG repeats; $k = 7$), c is a constant for the genome length divided by the number of chromosome ends and s is the fraction of all reads with GC composition between 48-52%. TelomereHunter also identifies telomeric reads based on the number to repeat sequences with a read and normalizes by the number of reads with GC content between 48-52%. This value is multiplied by 10^6 to calculate TRPM (telomeric reads per GC content-matched million reads) values. As a quality measure, TelSeq estimates for each sample were generated per sequencing lane. Reads from lanes that contained too few reads to calculate an estimate (marked as UNKNOWN), and outlier lanes as identified by grub's test, were removed from input BAMs using BAMQL v1.6)[50].

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

After outliers were removed, TelSeq was run again ignoring read groups with the -u parameter. Samples with telomere estimates less than one were removed from further analysis. To account for differences in TL due to sequencing center, a linear model was fit with TL as the response variable and sequencing center as the predictor variable. A separate model was fit for tumour and non-tumour length.

Somatic variant calling

Single nucleotide variants (SNVs) and genomic rearrangements (GRs) were called using pipelines that have been described in detail elsewhere[14]. Briefly, SomaticSniper (v1.0.5; was used to call SNVs on bases with at least 17x coverage in tumours and 10x in non-tumours. Coding versus non-coding SNVs were determined using Annovar[52]. Genomic rearrangements were identified using Delly (version 0.7.8; [53]). Gene fusion events involving *ERG* or *ETV* were collectively referred to as ETS events. Genomic rearrangement calls were examined to determine if breakpoints led to a TMPRSS2:ERG fusion or if breakpoints were found in both 1 Mbp bins surrounding the following gene pairs: *ERG*:*SLC45A3*, *ERG*:*NDRG1*, *ETV1*:*TMPRSS2*, *ETV4*:*TMPRSS2*, *ETV1*:*SLC45A3*, *ETV4*:*SLC45A3*, *ETV1*:*NDRG1*, and *ETV4*:*NDRG1*. *ERG* immunohistochemistry and deletion calls between *TMPRSS2* and *ERG* loci in OncoScan SNP array data provided further support for these fusions.

mRNA abundance data generation and analysis

Generation and analysis of RNA abundance data has been previously described in detail[18]. Briefly, 200 ng of total RNA was used to construct a TruSeq strand specific library with the Ribo-Zero protocol (Illumina), and all samples were sequenced on a HiSeq2000v3 to a minimal target of 180 million paired-end reads. Reads were mapped using the STAR aligner (v2.5.3a; [54]) to GRCh37 with GENCODE v24lift37[55]. RSEM (v1.2.29) was used to quantify gene abundance[56].

Methylation microarray data generation

Illumina Infinium HumanMethylation 450k BeadChip kits were used to assess global methylation, using 500 ng of input genomic DNA, at McGill University and the Genome Quebec Innovation Centre (Montreal, QC). All samples used in this study were processed from fresh-frozen prostate cancer tissue. The IDAT files were loaded and converted to raw intensity values with the use of wateRmelon package (v1.15.1; [57]. Quality control was conducted using the minfi package (v1.22.1; [58]; no outlier samples were detected). Raw methylation intensity levels were then pre-processed using Dasen. Probe filtering was conducted after normalization, as previously described[14]. Annotation to chromosome location, probe position, and gene symbol was conducted using the IlluminaHumanMethylation450kanno.ilmn12.hg19 package (v0.6.0).

Association of telomere length with fusions

The association between gene fusion status and tumour TL and TL ratio was tested using a Mann-Whitney U-test in 47 previously identified recurrent gene fusions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Association of telomere length and proliferation

A proliferation score per sample was generated using a previously published signature[30] where tumours with a RNA abundance value greater than the mean for each gene in the signature were given a score of +1, and tumours with a RNA abundance value less than the mean for that gene were given a score of –1. All values were summed to generate a proliferation score. Spearman's correlations between TL, TL ratio and the proliferation score was calculated. The correlation between TL, TL ratio and MKI67 abundance was also calculated.

Association of telomere length with chromothripsis

Chromothripsis scores were previously generated using ShatterProof (v0.14; [14,27] with default settings. Spearman's correlation between the maximum ShatterProof score per sample and telomere length was calculated using samples with both available metrics (n = 170).

Association of telomere length with clinical and genomic features

Telomere length estimates were associated with genomic and clinical features. Clinical features, including ISUP Grade, pre-treatment PSA, T category and age at diagnosis, were categorized and tested for association using an one-way ANOVA. Pathological T category was used for surgery samples and diagnostic T category was used for radiotherapy samples. Binary features including the presence of specific GRs, CNAs and SNVs were tested for association using a Mann-Whitney U test. Summary features including PGA, GR count, SNV count and indel count were correlated to TL using Spearman's correlation.

Association of telomere length with methylation

The correlation matrix of methylation and mRNA abundance levels from TCGA was downloaded from https://gdac.broadinstitute.org/. For each gene, the probe showing the highest Spearman's correlation with mRNA abundance levels was used in our correlation analysis.

Association of telomere length with transcriptome and proteome abundance

Spearman's correlations between TL and RNA (n = 139; [14]) and protein abundance (n = 70; [17]) and TL were calculated.

Over-representation analysis pathway analysis

Pathway analysis was performed with the gprofiler2[35] R package using genes in which there was a significant association between TL and methylation or RNA separately using the KEGG collection of pathways[36].

Crosstalk effects in pathway analysis

To account for crosstalk effects caused by gene overlap in pathway analysis, we implemented the principle component analysis method proposed by [38]. Briefly, for genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that are overlap among pathways, each gene is only allowed membership in one of the pathways. This membership is determined by the highest correlation between the gene and the PC1 of the other genes in the pathway. A fisher's exact test was then used to determine enrichment of TL correlated genes in the reduced pathway membership.

Association of telomere length with copy number aberrations

SNP microarray data generation and analysis has been previously described in detail[14]. Briefly, SNP microarrays were performed with 200 ng of DNA on Affymetrix OncoScan FFPE Express 2.0 and 3.0 arrays. Analysis of the probe assays was performed using .OSCHP files generated by OncoScan Console (v1.1) using a custom reference. BioDiscovery's Nexus Express™ for OncoScan 3 Software was used to call copy number aberrations using the SNP-FASST2 algorithm. Gene level copy number aberrations for each patient were identified by overlapping copy number segments from OncoScan SNP 3.0 data, with RefGene (2014-07-15) annotation using BEDTools (v2.17.0; [59]. Genes with the same copy number profile across patients were then collapsed into contiguous regions. Contiguous gene segments with aberrations in less that 5% of patients were removed from the analysis. To find associations between TL and copy number segments, a Mann-Whitney U test was used to compare the mean TL between samples with a copy number aberration and those without. The copy number aberration state (either amplified or deleted) was determined as the status with the largest proportion of samples. Samples with aberrations in the other class was merged into the without group. For example, three samples have an amplification in *CHD1*, while 49 samples have a deletion. The three samples would be grouped with copy number neutral samples and the Mann-Whitney test performed comparing the two groups. *P* values were FDR adjusted to account for multiple testing.

Association with biochemical relapse

Cox proportional hazards models were fit with the R package survival (v3.2-7) using TL as a continuous variable. Age at diagnosis was controlled for in the model. Kaplan Meier plots were generated by dichotomizing samples based on the optimal cut point analysis, in which samples were dichotomized using increasing thresholds of 50 bp.

Statistical analyses and data visualization

All statistical analyses were performed within the R statistical environment (v3.3.1). Visualization in R was performed through the BoutrosLab Plotting General package (v5.6.1; [60]). *P* values from Spearman's correlations were calculated using the AS-89 algorithm[61].

Data availability

OncoScan SNP array data and whole genome DNA sequencing can be found on EGA under the accession EGAS00001000900. Processed variant calls are available through the ICGC Data Portal under the project PRAD-CA (https://dcc.icgc.org/projects/PRAD-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CA). mRNA data is available in the Gene Expression Omnibus under the accession GSE84043. Methylation data is available under the accession GSE107298.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Table Legends

Supplementary Table 1 | Clinical and genomic features of tumours

Clinical data for 382 samples used in analysis after applying quality control metrics.

Supplementary Table 2 | Association between Tumour TL and recurrent gene fusions

Statistical summary of 47 recurrent fusions pairs tested for association with TL using a Wilcoxon signed-rank test.

Supplementary Table 3 | Genomic and transcriptomic correlations with Tumour TL

Results from Spearman's correlation between tumour TL and methylation beta values, RNA abundance and protein abundance. *Q* values are FDR adjusted *P* values. NAs indicate missing values where tests could not be performed.

Supplementary Table 4 | Genomic and transcriptomic correlations with TL ratio

Results from Spearman's correlation between tumour TL and methylation beta values, RNA abundance and protein abundance. *Q* values are FDR adjusted *P* values. NAs indicate missing values where tests could not be performed.

Supplementary Table 5 | Associations between CNAs and Tumour TL

Associations between CNAs and Tumour TL, ordered by FDR adjusted *P* values. Each row represents collapsed segments containing multiple genes. Contiguous gene segments with aberrations in less that 5% of patients were removed.

Supplementary Table 6 | Associations between CNAs and TL ratio

Statistically significant associations between CNAs and TL ratio, ordered by FDR adjusted *P* values. Each row represents collapsed segments containing multiple genes. Contiguous gene segments with aberrations in less that 5% of patients were removed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Moyzis RK, Buckingham JM, Cram LS, et al. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci USA*. 1988;85(18):6622-6626.

2. Arnoult N, Karlseder J. Complex interactions between the DNA-damage response and mammalian telomeres. *Nat Struct Mol Biol*. 2015;22(11):859-866. doi:10.1038/nsmb.3092

3. Samassekou O, Gadji M, Drouin R, Yan J. Sizing the ends: normal length of human telomeres. *Ann Anat*. 2010;192(5):284-291. doi:10.1016/j.aanat.2010.07.005

4. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. *Exp Cell Res*. 1961;25:585-621.

5. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.

6. Dunham MA, Neumann AA, Fasching CL, Reddel RR. Telomere maintenance by recombination in human cells. *Nat Genet*. 2000;26(4):447-450. doi:10.1038/82586

7. Barthel FP, Wei W, Tang M, et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet*. 2017;49(3):349-357. doi:10.1038/ng.3781

8. Sieverling L, Hong C, Koser SD, et al. Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nature Communications*. 2020;11(1):733. doi:10.1038/s41467-019-13824-9

9. Lalonde E, Ishkanian AS, Sykes J, et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. *Lancet Oncol*. 2014;15(13):1521-1532. doi:10.1016/S1470-2045(14)71021-6

10. Lalonde E, Alkallas R, Chua MLK, et al. Translating a Prognostic DNA Genomic Classifier into the Clinic: Retrospective Validation in 563 Localized Prostate Tumors. *European Urology*. 2017;72(1):22-31. doi:10.1016/j.eururo.2016.10.013

11. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153(3):666-677. doi:10.1016/j.cell.2013.03.021

12. Berger MF, Lawrence MS, Demichelis F, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214-220. doi:10.1038/nature09744

13. The Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163(4):1011-1025. doi:10.1016/j.cell.2015.10.025

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

14. Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*. 2017;541(7637):359-364. doi:10.1038/nature20788

15. Bhandari V, Hoey C, Liu LY, et al. Molecular landmarks of tumor hypoxia across cancer types. *Nature Genetics*. Published online January 14, 2019:1. doi:10.1038/s41588-018-0318-2

16. Espiritu SMG, Liu LY, Rubanova Y, et al. The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. *Cell*. 2018;173(4):1003-1013.e15. doi:10.1016/j.cell.2018.03.029

17. Sinha A, Huang V, Livingstone J, et al. The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell*. 2019;35(3):414-427.e6. doi:10.1016/j.ccell.2019.02.005

18. Chen S, Huang V, Xu X, et al. Widespread and Functional RNA Circularization in Localized Prostate Cancer. *Cell*. 2019;176(4):831-843.e22. doi:10.1016/j.cell.2019.01.025

19. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*. 2013;45(10):1127-1133. doi:10.1038/ng.2762

20. Robinson D, Van Allen EM, Wu Y-M, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215-1228. doi:10.1016/j.cell.2015.05.001

21. Kim H, Chen J. c-Myc interacts with TRF1/PIN2 and regulates telomere length. *Biochem Biophys Res Commun*. 2007;362(4):842-847. doi:10.1016/j.bbrc.2007.08.064

22. Lee SS, Bohrson C, Pike AM, Wheelan SJ, Greider CW. ATM Kinase Is Required for Telomere Elongation in Mouse and Human Cells. *Cell Rep*. 2015;13(8):1623-1632. doi:10.1016/j.celrep.2015.10.035

23. Sommerfeld HJ, Meeker AK, Piatyszek MA, Bova GS, Shay JW, Coffey DS. Telomerase activity: a prevalent marker of malignant human prostate tissue. *Cancer Res*. 1996;56(1):218-222.

24. Weischenfeldt J, Simon R, Feuerbach L, et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell*. 2013;23(2):159-170. doi:10.1016/j.ccr.2013.01.002

25. Ding Z, Mangino M, Aviv A, Spector T, Durbin R, UK10K Consortium. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res*. 2014;42(9):e75. doi:10.1093/nar/gku181

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

26. Feuerbach L, Sieverling L, Deeg KI, et al. TelomereHunter – in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics*. 2019;20(1):272. doi:10.1186/s12859-019-2851-0

27. Govind SK, Zia A, Hennings-Yeomans PH, et al. ShatterProof: operational detection and quantification of chromothripsis. *BMC Bioinformatics*. 2014;15:78. doi:10.1186/1471-2105-15-78

28. Chin L, Artandi SE, Shen Q, et al. p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell*. 1999;97(4):527-538.

29. Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell*. 2015;163(7):1641-1654. doi:10.1016/j.cell.2015.11.054

30. Starmans MHW, Lieuwes NG, Span PN, et al. Independent and functional validation of a multi-tumour-type proliferation signature. *Br J Cancer*. 2012;107(3):508-515. doi:10.1038/bjc.2012.269

31. Armenia J, Wankowicz SAM, Liu D, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet*. 2018;50(5):645-651. doi:10.1038/s41588-018-0078-z

32. Heaphy CM, de Wilde RF, Jiao Y, et al. Altered telomeres in tumors with ATRX and DAXX mutations. *Science*. 2011;333(6041):425. doi:10.1126/science.1207313

33. Graham MK, Meeker A. Telomeres and telomerase in prostate cancer development and therapy. *Nat Rev Urol*. 2017;14(10):607-619. doi:10.1038/nrurol.2017.104

34. Majumder PK, Sellers WR. Akt-regulated pathways in prostate cancer. *Oncogene*. 2005;24(50):7465-7474. doi:10.1038/sj.onc.1209096

35. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res*. 2020;9. doi:10.12688/f1000research.24956.2

36. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

37. Donato M, Xu Z, Tomoiaga A, et al. Analysis and correction of crosstalk effects in pathway analysis. *Genome Res*. 2013;23(11):1885-1893. doi:10.1101/gr.153551.112

38. Zhou Y, Gao Y, Xu C, Shen H, Tian Q, Deng H-W. A novel approach for correction of crosstalk effects in pathway analysis and its application in osteoporosis research. *Sci Rep*. 2018;8(1):668. doi:10.1038/s41598-018-19196-2

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

39. Heaphy CM, Yoon GS, Peskoe SB, et al. Prostate cancer cell telomere length variability and stromal cell telomere length as prognostic markers for metastasis and death. *Cancer Discov*. 2013;3(10):1130-1141. doi:10.1158/2159-8290.CD-13-0135

40. Svenson U, Roos G, Wikström P. Long leukocyte telomere length in prostate cancer patients at diagnosis is associated with poor metastasis-free and cancer-specific survival. *Tumour Biol*. 2017;39(2):1010428317692236. doi:10.1177/1010428317692236

41. Renner W, Krenn-Pilko S, Gruber H-J, Herrmann M, Langsenlehner T. Relative telomere length and prostate cancer mortality. *Prostate Cancer Prostatic Dis*. 2018;21(4):579-583. doi:10.1038/s41391-018-0068-3

42. Koeneman KS, Pan CX, Jin JK, et al. Telomerase activity, telomere length, and DNA ploidy in prostatic intraepithelial neoplasia (PIN). *J Urol*. 1998;160(4):1533-1539.

43. Zhang W, Kapusta LR, Slingerland JM, Klotz LH. Telomerase activity in prostate cancer, prostatic intraepithelial neoplasia, and benign prostatic epithelium. *Cancer Res*. 1998;58(4):619-621.

44. Prensner JR, Chen W, Iyer MK, et al. PCAT-1, a long noncoding RNA, regulates BRCA2 and controls homologous recombination in cancer. *Cancer Res*. 2014;74(6):1651-1660. doi:10.1158/0008-5472.CAN-13-3159

45. Quigley DA, Dang HX, Zhao SG, et al. Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell*. 2018;174(3):758-769.e9. doi:10.1016/j.cell.2018.06.039

46. Song S, Nones K, Miller D, et al. qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE*. 2012;7(9):e45835. doi:10.1371/journal.pone.0045835

47. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324

48. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110

49. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

50. Masella AP, Lalansingh CM, Sivasundaram P, Fraser M, Bristow RG, Boutros PC. BAMQL: a query language for extracting reads from BAM files. *BMC Bioinformatics*. 2016;17:305. doi:10.1186/s12859-016-1162-y

51. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012;28(3):311-317. doi:10.1093/bioinformatics/btr665

52. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603

53. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):i333-i339. doi:10.1093/bioinformatics/bts378

54. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

55. Frankish A, Diekhans M, Ferreira A-M, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766-D773. doi:10.1093/nar/gky955

56. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323. doi:10.1186/1471-2105-12-323

57. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:293. doi:10.1186/1471-2164-14-293

58. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049

59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033

60. P'ng C, Green J, Chong LC, et al. BPG: Seamless, automated and interactive visualization of scientific data. *BMC Bioinformatics*. 2019;20(1):42. doi:10.1186/s12859-019-2610-2

61. Best DJ, Roberts DE. Algorithm AS 89: The Upper Tail Probabilities of Spearman's Rho. *Journal of the Royal Statistical Society Series C (Applied Statistics)*. 1975;24(3):377-379. doi:10.2307/2347111

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

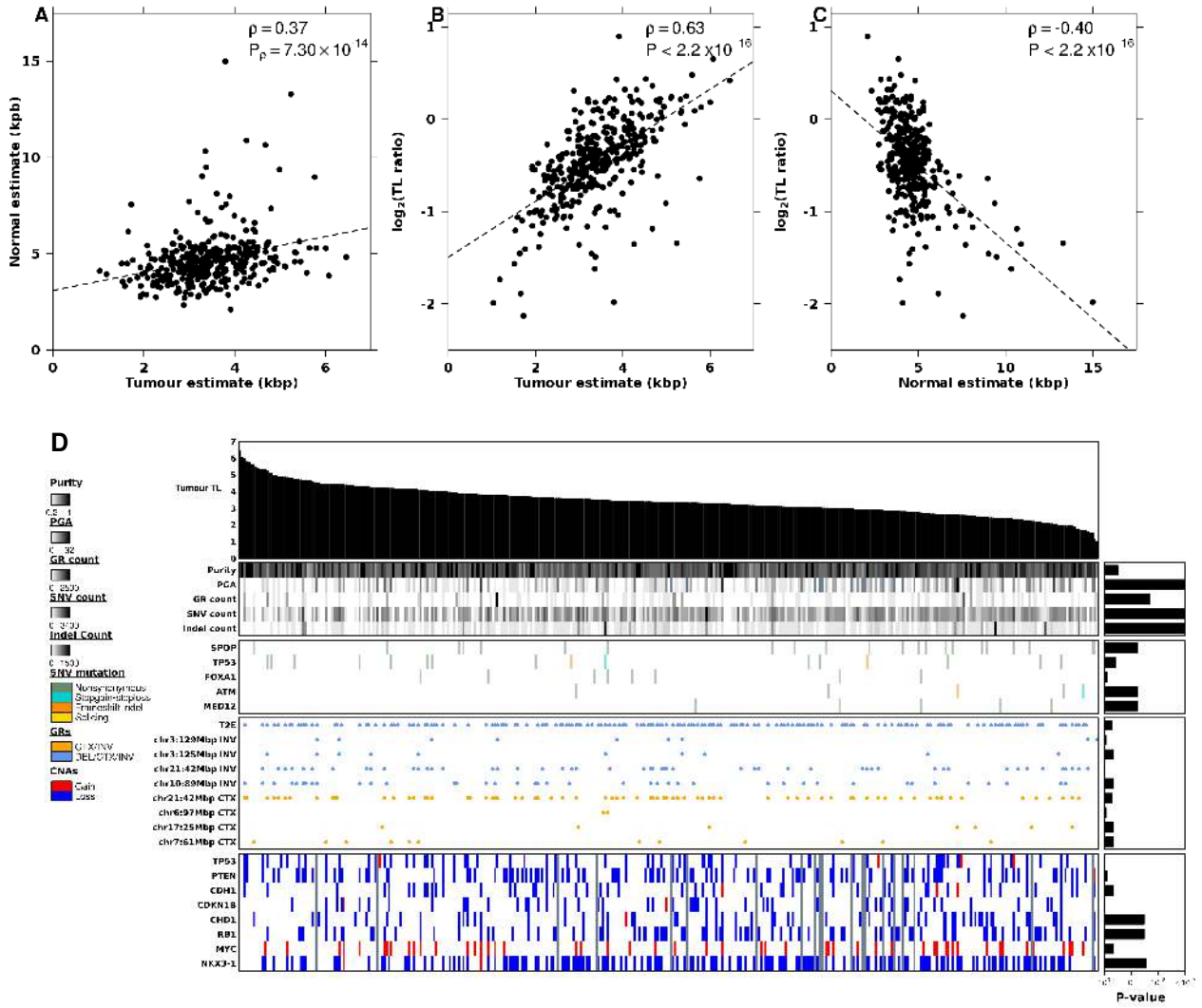

Figure 1 — Tumour telomere length (TL) is associated with genomic features

A-B, Correlation between tumour TL and **A** non-tumour (blood) TL and **B**, TL ratio (tumour TL / non-tumour (blood) TL). **C**, Correlation between non-tumour (blood) TL and TL ratio. **D**, Tumour TL is ranked in descending order of length (kbp; top bar plot). The association of tumour TL and measures of mutational burden, TMPRSS2:ERG (T2E) fusion status, as well as known prostate cancer genes with recurrent CNAs, coding SNVs, and GRs are shown. Bar plots to the right indicate the statistical significance of each association **(see Methods)**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

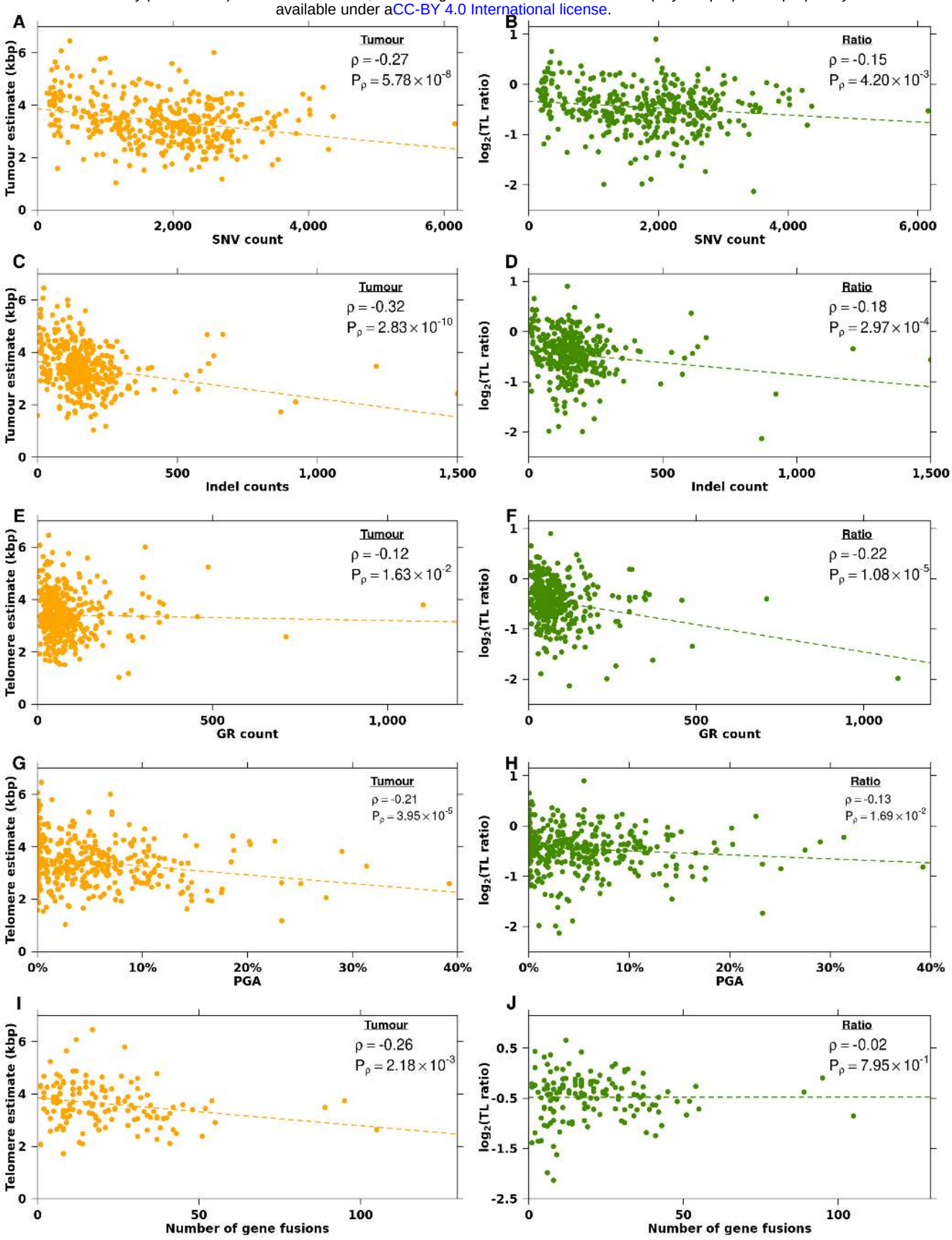

Figure 2 — Mutational landscape differs with telomere length

A-B, Correlation between the number of SNVs and **A**, tumour TL and **B**, TL ratio. **C-D**, Correlation between the number of indels and **C**, tumour TL and **D**, TL ratio. **E-F**, Correlation between the number of GRs and **E**, and tumour TL and **F**, TL ratio. **G-H**, Correlation of percentage of the genome altered (PGA) and **G**, tumour TL and **H**, TL ratio. **I-J**, Correlation between the number of fusions and **I**, tumour TL and **J**, TL ratio. Orange dots indicate tumour TL while green dots indicate TL ratio. Spearman's ρ and P-values are displayed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

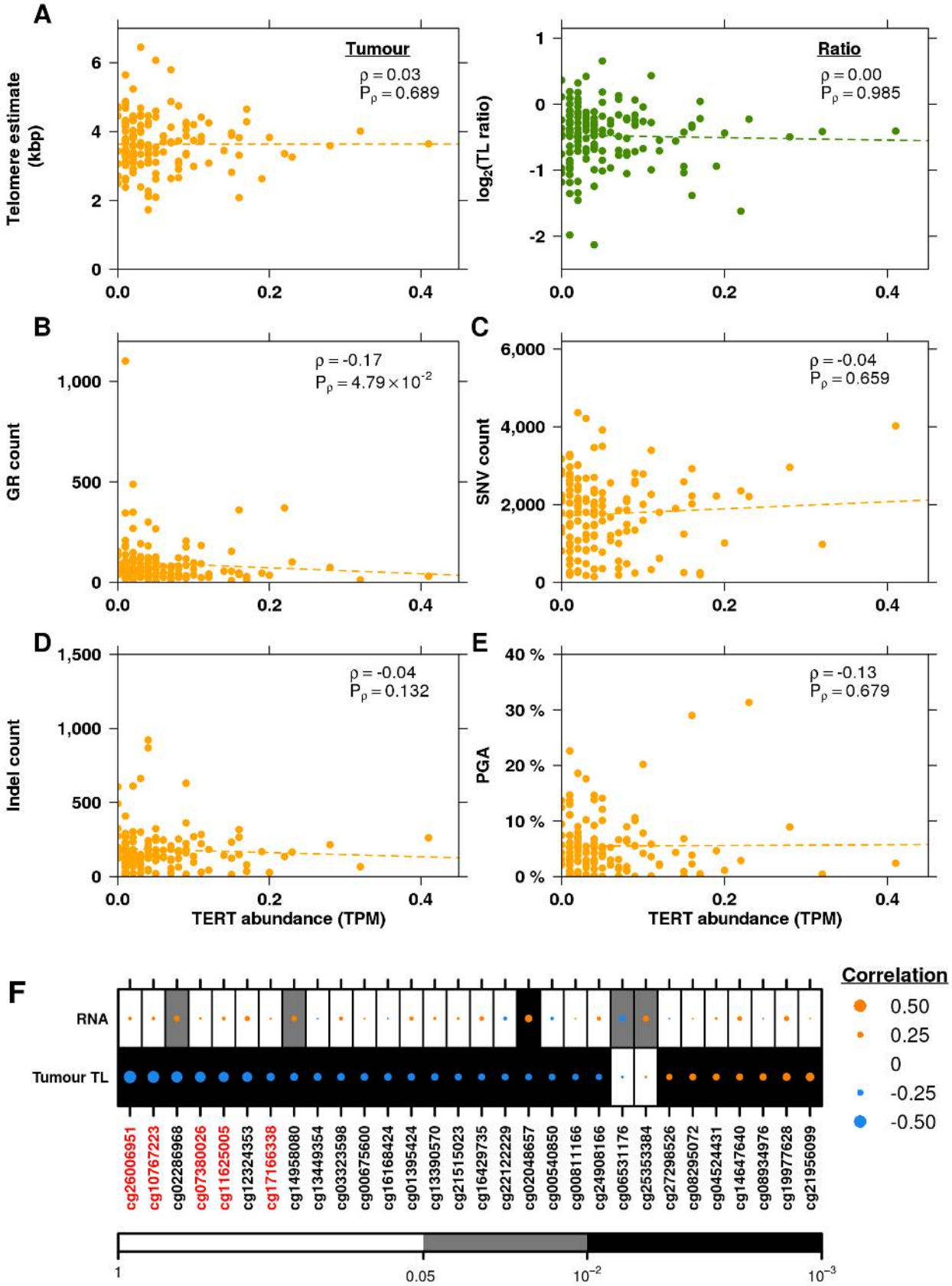

Figure 3 — The genomic correlates of TERT abundance

A, Correlation of *TERT* RNA abundance with tumour TL and TL ratio. Orange dots indicate tumour TL while green dots indicate TL ratio. Spearman ρ and P-values are displayed. **B-E**, Correlation of *TERT* abundance and **B**, the number of GRs, **C**, number of SNVs, **D**, number of indels, and **E**, PGA. Spearman ρ and P-values are displayed. **F**, Spearman's correlation of significantly associated methylation probes with RNA abundance and tumour TL. Orange dots indicate a positive correlation while blue dots indicate a negative correlation. Probes within the promoter are labeled in red while the rest are located in the gene body. Dot size indicated the magnitude of correlation. Background colour indicates unadjusted P-values. Methylation probes are ordered by their correlation between *TERT* RNA abundance from negative to positive.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

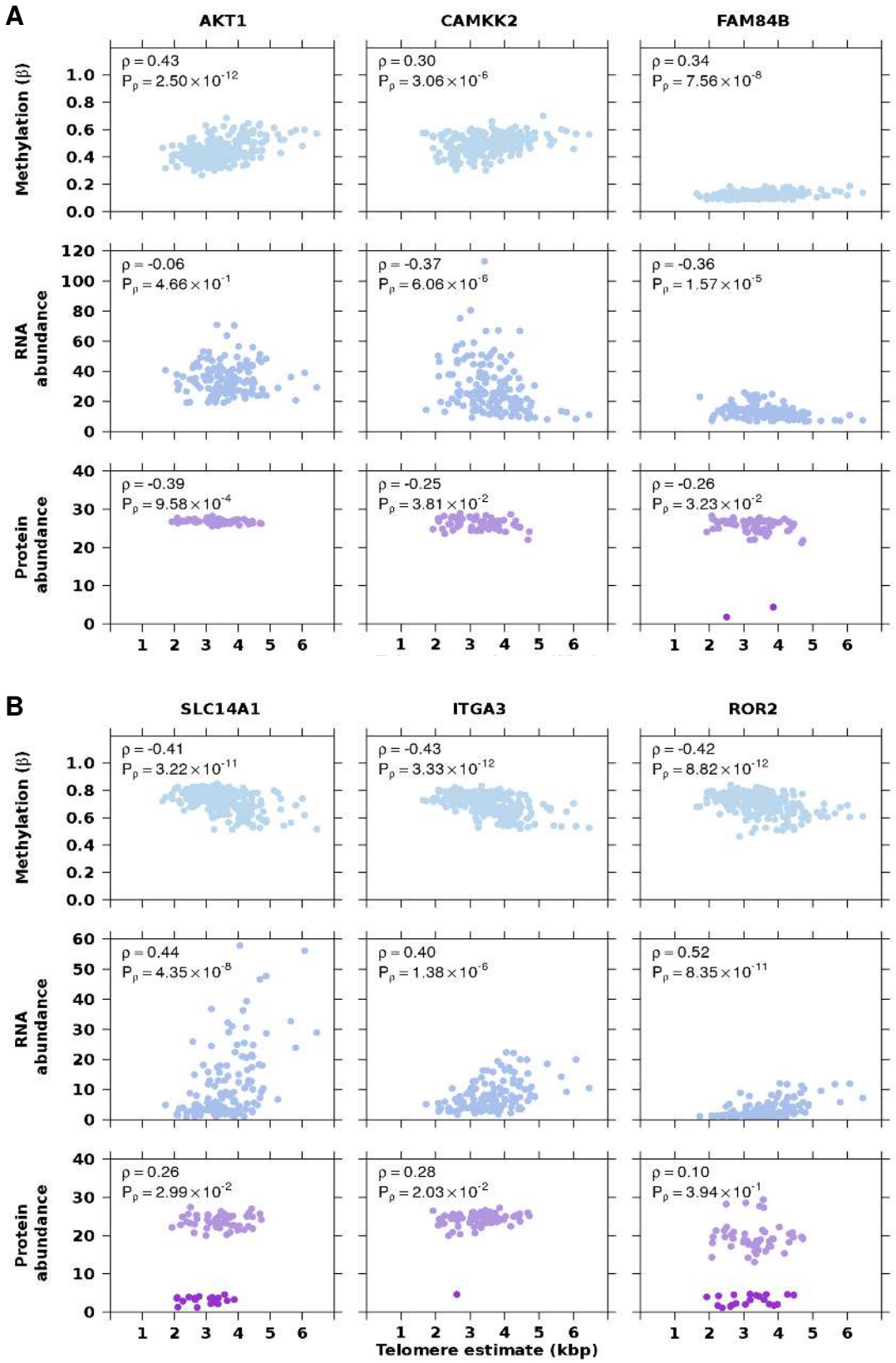

Figure 4 — Association of methylation, RNA abundance, protein abundance and telomere length

A, Positive correlation of methylation and tumour TL, but negative correlation of RNA and protein abundance. **B**, Negative correlation of methylation and tumour TL, but positive correlation of RNA and protein abundance. Top panels in light blue represent methylation beta values, middle panels in blue-grey represent RNA abundance and the bottom panels in purple represent protein abundance. Darker purple dots represent undetected, imputed protein abundance measures. Spearman ρ and P-values are displayed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

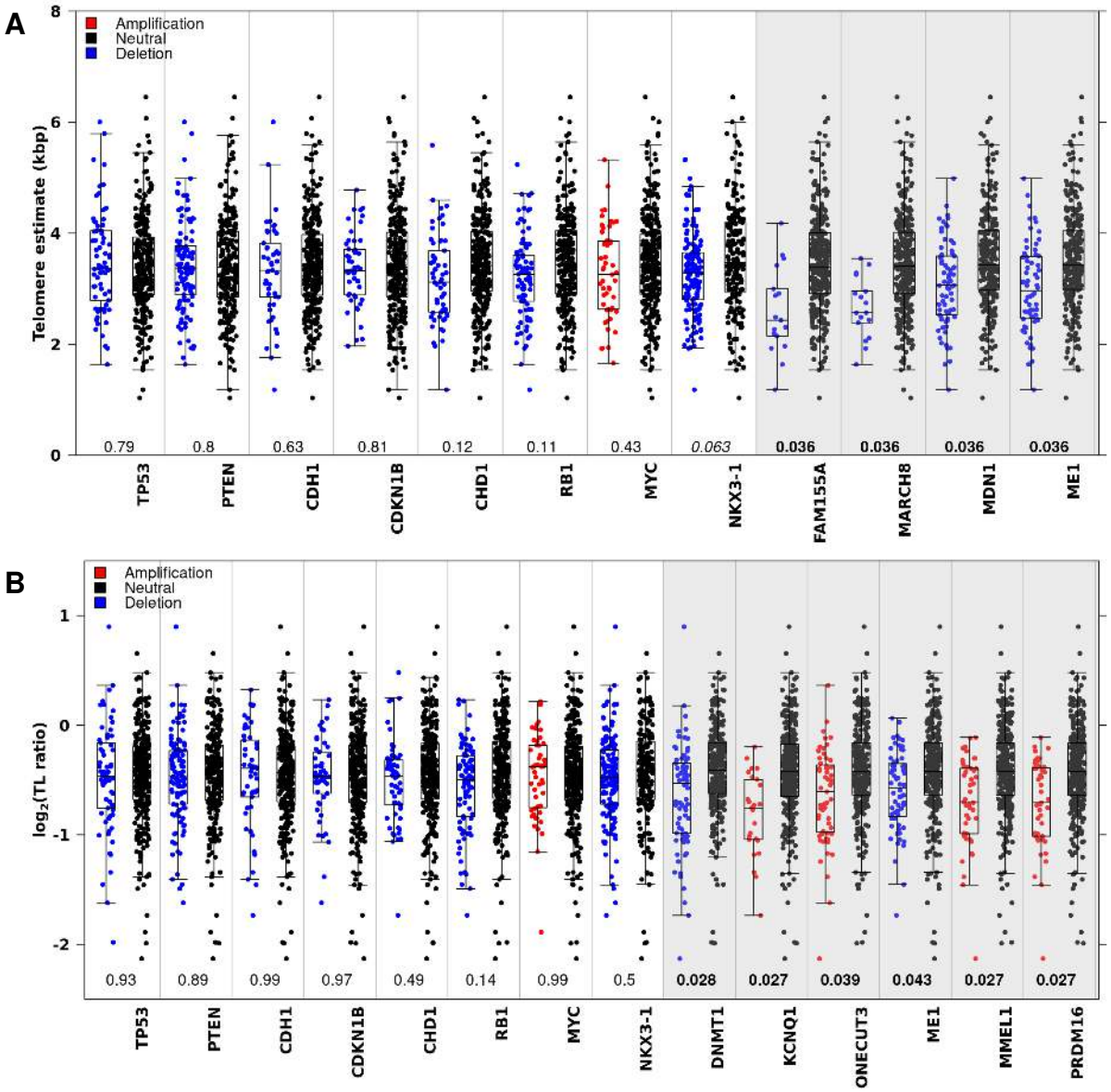

Figure 5 — Telomere length differs by copy number status

A, Difference in tumour TL between samples with a copy number aberration and those without in prostate cancer related genes and associated genes. **B**, Difference in TL ratio between samples with a copy number aberration and those without in prostate cancer related and associated genes. Q-values are from a Mann-Whitney U test and are bolded when significant (< 0.05). Colour of the dots indicate copy number status of the gene: amplification (red), deletion (blue), or neutral (black). Boxes with a white background are known prostate cancer genes, while boxes with a gray background were identified by a genome wide search.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.22.445239; this version posted September 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

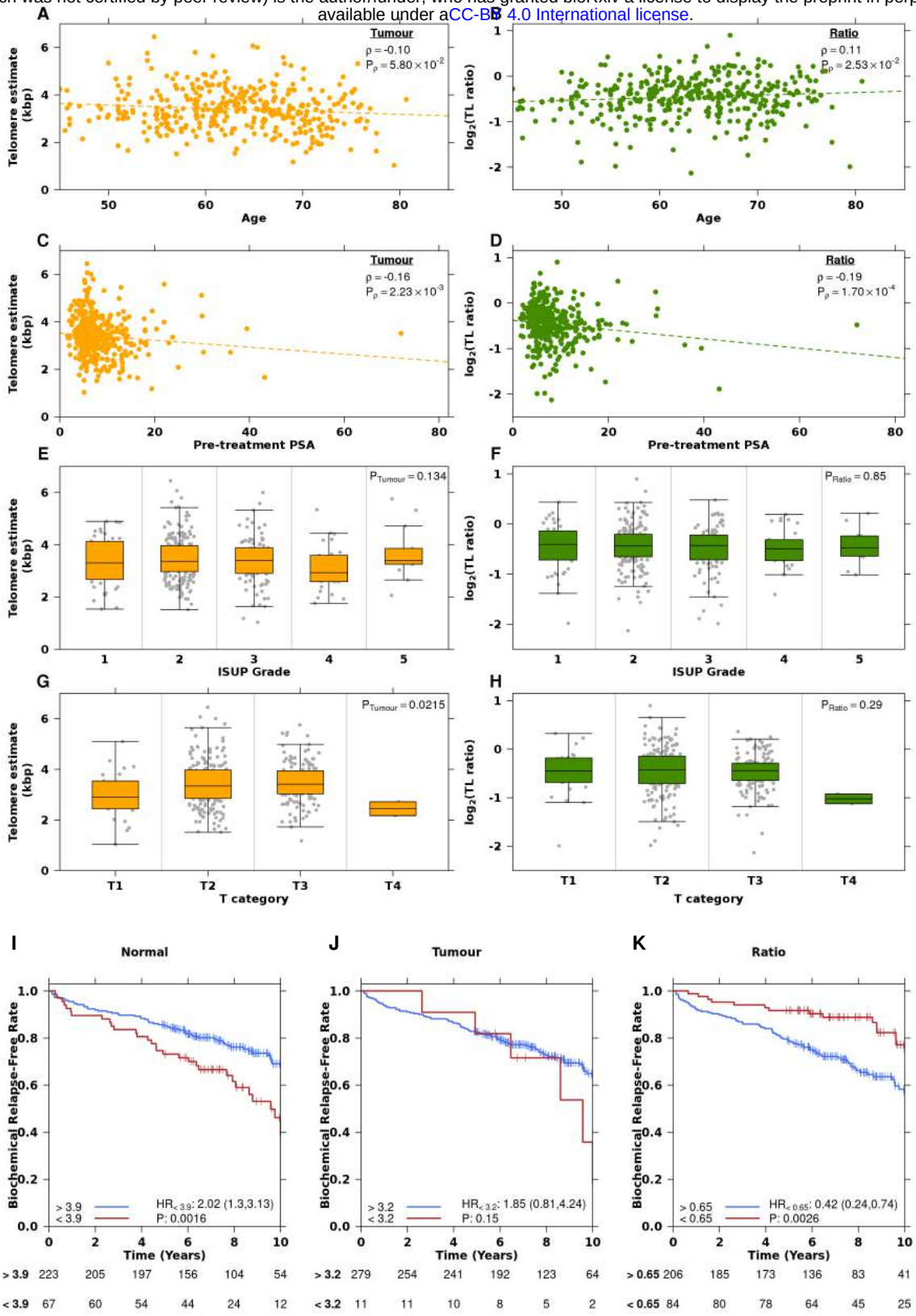

Figure 6 — Telomere length is associated with clinical features and biochemical relapse

A-B, Correlation of age at treatment with **A**, tumour TL and **B**, TL ratio. Spearman ρ and P-values are displayed. **C-D**, Correlation of pre-treatment PSA with **C**, tumour TL and **D**, TL ratio. Spearman ρ and P-values are displayed. **E-F**, Association of ISUP grade with **E**, tumour TL and **F**, TL ratio. P-value is from an one-way ANOVA. **G-H**, Association of T category with **G**, tumour TL and **H**, TL ratio. P-value is from an one-way ANOVA. On all plots, green indicates TL ratio, while orange indicates tumour TL. **I-K**, Cox proportional hazard models were created for **I**, non-tumour (blood) TL, **J**, tumour TL and **K**, TL ratio with BCR as the endpoint. Samples were split into two groups based on the optimal cut point analysis **(see Methods)**.