Inferring Telomerase Enzymatic Activity from Expression Data

Nighat Noureen[1], Shaofang Wu[3], Yingli Lyu[1], Juechen Yang[1], WK Alfred Yung[3], Jonathan Gelfond[2], Xiaojing Wang[1,2], Dimpy Koul[3], Andrew Ludlow[4], Siyuan Zheng[1,2]†

[1]Greehey Children's Cancer Research Institute, UT Health San Antonio, San Antonio, Texas, USA. [2]Department of Population Health Sciences, UT Health San Antonio, San Antonio, Texas, USA. [3]Department of Neuro-Oncology, MD Anderson Cancer Center, Houston, Texas, USA. [4]Department of Movement Science, University of Michigan, Ann Arbor, Michigan, USA

†Corresponding author. Email: zhengs3@uthscsa.edu

Abstract:

Active telomerase is essential for stem cells and most cancers to maintain telomeres. The enzymatic activity of telomerase is related but not equivalent to the expression of TERT, the catalytic subunit of the complex. Here we show that telomerase enzymatic activity can be robustly estimated from the expression of a 13-gene signature. We demonstrate the validity of the expression-based approach, named EXTEND, using cell lines, cancer samples, and non-neoplastic samples. When applied to over 9,000 tumors and single cells, we find a strong correlation between telomerase activity and cancer stemness. This correlation is largely driven by a small proliferating cancer cell population that exhibits both high telomerase activity and cancer stemness. This study establishes a novel computational framework for quantifying telomerase enzymatic activity and provides new insights into the relationships among telomerase, cancer proliferation, and stemness.

INTRODUCTION

Telomerase is the ribonucleoprotein complex that adds telomeric repeats to telomeres at chromosome ends. In the absence of telomerase, telomeres progressively shorten due to incomplete replication of chromosome ends[1]. Persistent telomere shortening leads to senescence and crisis, thus continuously dividing cells including stem cells and most cancer cells require active telomerase to maintain telomere lengths[2]. Loss of telomerase activity results in degenerative defects and premature ageing[3,4]. In contrast, reactivation of telomerase enables malignant transformation and cancer cell immortality[5,6]. These observations emphasize the pivotal role of telomerase in many human health concerns and highlight the needs for close monitoring of its activity.

The telomerase complex is composed of the reverse transcriptase subunit TERT, the template containing non-coding RNA TERC, and accessory proteins such as dyskerin (DKC1) and telomerase Cajal body protein 1 (TCAB1). The core subunits are the catalytic subunit TERT and the RNA template TERC. In vivo, processive extension of telomeres, i.e. telomerase processivity, requires binding of telomerase to telomeres through a six-protein complex, shelterin, specifically its component protein TPP1[7,8]. Another shelterin protein POT1 bridges TPP1 and chromosome 3' overhang, the DNA substrate of telomerase.

While *TERC* is thought to be abundant and ubiquitously expressed[9,10], *TERT* is transcriptionally repressed in most somatic cells[11]. Some cancer lineages (e.g. brain, liver, skin and bladder) frequently acquire recurrent mutations in the *TERT* promoter region, predominantly at -124 and -146 loci upstream from *TERT* transcription start site[12-15]. These C>T mutations create consensus binding sites for GABP transcription factors, alter chromatin states, and enhance the transcriptional output of *TERT* [16-20]. In bladder cancer, the promoter mutations correlate with increased *TERT* expression and telomerase enzymatic activity[21].

Enzymatic activity is a fundamental metric of telomerase. Several protocols have been established to measure telomerase enzymatic activity, including the PCR-based TRAP assay and direct enzymatic assays[22-24]. These assays allowed for investigations of associations between telomerase activity and clinical and histopathologic variables in cancer and other diseases[25,26], and regulation of telomerase activity by each component of the telomerase complex. Ectopic expression of *TERT* and *TERC*, in many cases *TERT* alone, increases telomerase enzymatic activity[27-29]. These data led to the view that *TERT* is the limiting component for telomerase activity.

However, new emerging data challenge the use of *TERT* expression as a surrogate for telomerase enzymatic activity. First, the *TERT* gene can transcribe more than 20 splicing isoforms[30], but only the full-length transcript bearing all 16 exons can produce the catalytic subunit[31-33]. Secondly, single-cell imaging studies showed most *TERT* mRNAs localize in the nucleus but not the cytoplasm, and thus are not translated[34].

Thirdly, endogenous TERT protein and *TERC* are far more abundant than the assembled telomerase complex in cancer cell lines[23]. Finally, *TERC* and accessory proteins can also impact telomerase activity. For example, telomerase activity in human T cells has been reported to relate to *TERC* levels rather than TERT[35]. In addition, mutations in *DKC1* and *TERC* both cause dyskeratosis congenita, a rare genetic syndrome related to impaired telomerase[10,36].

Here we report a systematic analysis of telomerase activity in cancer. This analysis was enabled by a new telomerase activity prediction algorithm, EXpression-based Telomerase ENzymatic activity Detection (EXTEND).

RESULTS

Rationale and overview of EXTEND, a tool for predicting telomerase activity

An overview of the EXTEND algorithm is shown in **Supplementary Fig. 1**. We posited that comparing expression of telomerase positive and negative tumors could yield a telomerase activity signature. Most epithelial tumors express *TERT* and use telomerase to maintain telomeres. Without experimental evidence, however, *TERT* expression cannot reliably indicate positive telomerase for the reasons noted above. We instead used *TERT* promoter mutation to stratify such tumors, reasoning that the presence of this genetic change likely reflects evolutionary selection for telomerase as the predominant telomere maintenance mechanism (TMM). We further assumed that tumors with alternative lengthening of telomeres (ALT) phenotype, an alternate TMM mechanism, were negative controls. The ALT phenotype is usually determined through ALT-associated promyelocytic leukemia nuclear bodies, extrachromosomal telomeric DNA C-Circles, or ALT-associated telomere foci[37,38]. Large cohorts of experimentally confirmed ALT tumors are not yet publicly available, largely due to the rarity of the phenotype. However, mutations in *ATRX* and its interacting partner *DAXX* are nearly perfectly correlated with ALT [39]. We thus searched the TCGA dataset for cancer types with both high frequencies of mutations in the *TERT* promoter and *ATRX* and *DAXX*. This search identified lower-grade glioma (LGG) (**Supplementary Fig. 2**), a cancer type demonstrating strong mutual exclusivity of the two TMMs[40].

We identified 108 *TERT* co-expressing genes in the LGG dataset. These genes were further intersected with genes upregulated in the *TERT* promoter mutant tumors. The resulting 12 genes were complemented with *TERC*, the RNA subunit of the telomerase complex, giving rise to a 13-gene signature. Seven of the 13 genes were highly expressed in testis but low in other tissues (**Supplementary Table 1**). Mutations in *HELLS*, a gene encoding a lymphoid-specific helicase, cause the centromeric instability and facial anomalies (ICF) syndrome, a genetic disorder associated with short telomeres[41]. Functional enrichment analysis suggested an overrepresentation of the signature genes in the cell cycle pathway (FDR=1.95e-4), particularly S phase

(FDR=0.01, **Supplementary Table 2**), a narrow time window when telomerase is active in extending telomeres[42].

To score this signature, we designed an iterative rank-sum method (**Supplementary Fig. 1**). This method first divides the signature into a constituent component (*TERT* and *TERC*) and a marker component (the other 11 genes). The constituent component is scored by the maximum ranking of *TERT* and *TERC*, whereas the marker component is scored by the rank sum of the signature genes. Because the size of the constituent component is much smaller than the marker component, we adjusted its contribution to the final score by a factor determined based on the correlation between the original constituent component score and the aggregated score. Using cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE), we estimated that the constituent component generally contributes less than 20% to the final scores (**Supplementary Fig. 3**).

Since *TERC* lacks a long poly(A) tail, its measurement is less robust with a poly(A) enriched mRNA sequencing protocol. Thus, we tested the stability of EXTEND across sequencing protocols. Using samples sequenced by both ribosomal RNA depletion and poly(A) enrichment protocols[43], we found *TERC* was indeed less concordant than *TERT* between the two protocols (Rho=0.71 vs 0.98), but EXTEND scores nevertheless agreed well (Rho=0.96, P=1.9e-6; **Supplementary Fig. 4**).

Validation and comparison with *TERT* expression in cancer

We first validated EXTEND with cancer cell lines. We performed semi-quantitative TRAP assays on 28 patient-derived glioma sphere-forming cells (GSCs) (Methods). EXTEND scores were significantly correlated with experimental readouts (Rho=0.48, P=0.01; **Fig. 1A**). Although this correlation was comparable to *TERT* expression (Rho=0.5, P=0.01), EXTEND demonstrated superior performance in differentiating two recently determined ALT lines, GSC5-22 and GSC8-18[37] (**Supplementary Fig. 5**). Using 11 bladder cancer lines with available RNAseq data, we compared predictions obtained using EXTEND with more accurate direct enzymatic assays[21]. EXTEND scores and cell line telomerase activity were significantly correlated (Rho=0.72, P=0.01), whereas the correlation for *TERT* did not reach statistical significance (Rho=0.55, P=0.08) (**Fig. 1B, Supplementary Fig. 6**). We then measured telomerase activity of 15 lung cancer cell lines using digital droplet TRAP assays[24]. EXTEND outperformed *TERT* expression in predicting telomerase on these cell lines (Rho=0.65, P=0.01 vs Rho=0.48, P=0.07) (**Fig. 1C**, **Supplementary Fig. 7**).

We next tested EXTEND in two cancer types enriched with ALTs[44], liposarcoma and neuroblastoma. In liposarcoma[45], telomerase-positive tumors had significantly higher EXTEND scores than ALT tumors (P<0.01, Student t test; **Fig. 1D**). Neuroblastomas were previously divided into five groups based on telomere maintenance mechanisms: *TERT* promoter rearrangement, *MYCN* amplification, *TERT* expression high, ALT, and no TMM[38]. Enzymatic assays suggested that the first three groups were telomerase-

positive, whereas the latter two had very low telomerase activity[38]. Consistent with these observations, EXTEND estimated higher telomerase activity in the three telomerase-positive groups (**Fig. 1E**). The *TERT* high group was estimated to have relatively lower scores than the other two telomerase-positive groups, a pattern also consistent with the experimental data[38].

Telomerase activity in non-neoplastic and embryonic samples

We then applied EXTEND to tissue samples from the GTEx dataset (**Supplementary Table 3**). These samples were collected post-mortem from healthy donors, except for a few transformed cell lines. Among the 52 sub-tissue types, we observed the highest EXTEND scores in Epstein-Barr virus-transformed lymphocytes, testis, transformed skin fibroblasts, and esophageal mucosa, a tissue with a high self-renewal rate (**Fig. 2A, Supplementary Fig. 8**). The testis was recently shown to have the longest average telomere lengths among human tissues[46]. Highly differentiated tissue skeletal muscle samples had the lowest average scores. Skin-transformed fibroblasts had almost negligible *TERT* expression despite a relatively high EXTEND score. In contrast, brain tissues, including substantia nigra (brain_SN), putamen, nucleus accumbens (brain_NA) and caudate, had low EXTEND scores but expressed *TERT* at a detectable level. To explain this disparity, we examined *TERT* alternative splicing. Compared with testis, most expressed *TERT* were short-spliced forms in the brain thus could not encode catalytically active proteins (**Supplementary Fig. 9**).

Next, we analyzed human tissues during embryonic development. We focused on heart and liver, two organs with distinct self-renewal behaviors. Cardiomyocytes were thought to lose their proliferative ability after birth[47], whereas telomerase-expressing hepatocytes robustly repopulate the liver in homeostasis throughout adulthood[48,49]. Using a recently published dataset[50], we analyzed both organs through embryonic weeks 4-20 and after birth. For heart, EXTEND scores dropped considerably between the 12th and 13th embryonic weeks but remained at modest levels in infants and toddlers (**Fig. 2B**). A further drop was observed when entering adulthood. These observations agreed with previous studies that reported decreased telomerase activity after the 12th embryonic week in the heart tissue[51], and that infant hearts have higher telomerase activity than those of adults[52].

We did not observe significant decreases in EXTEND scores for liver during late embryonic weeks, despite the declining *TERT* expression after around 10 weeks. EXTEND scores largely remained stable across the lifespan, although their levels were much lower than in fetal samples (**Fig. 2C**, P=1.1e-9).

Finally, we applied EXTEND to samples from a patient with dyskeratosis congenita[53]. This patient carried loss-of-function mutations in Poly(A)-specific ribonuclease (*PARN*), a gene required for *TERC* maturation. Inhibition of poly(A) polymerase PAP-associated domain–containing 5 (*PAPD5*) counteracts *PARN* mutations to increase *TERC* levels

and telomerase activity [53]. EXTEND predicted higher telomerase activity in *PARN* mutant cells compared with wild-type controls upon *PAPD5* knockdown, confirming the earlier finding (**Fig. 2D**).

Landscape of telomerase activity in cancer

We next analyzed more than 9,000 tumor samples and 700 normal samples from TCGA (**Supplementary Table 4**). We first evaluated if EXTEND scores were also reflective of telomerase processivity in cancer. POT1 is a shelterin component that regulates telomerase processivity through interactions with chromosome 3' overhang and TPP1, the telomerase anchor to telomeres[54]. TPP1-POT1 heterodimer has been reported to enhance telomerase processivity[55-57]. EXTEND scores were positively correlated with *POT1* expression across pan-cancer (R=0.22, P<2.2e-16). No correlation was observed between *TERT* and *POT1* (R=-0.004, p=0.71) (**Supplementary Fig. 10**). We did not find significant associations between telomere length and EXTEND scores in TCGA cohorts after adjusting for multiple hypothesis testing (**Supplementary Fig. 11**).

As expected, tumors had significantly higher EXTEND scores than matched normal samples (**Fig. 3A**; P < 2.2e-16, t test). Similar to the GTEx sample analyses, scores varied across normal tissues. Gastrointestinal organs (esophageal – ESCA, stomach – STAD, colorectal - CRC), mammary gland (breast - BRCA), and reproductive organs (uterus endometrial - UCEC) had overall higher scores. EXTEND scores varied across cancer types, with kidney (KIRP, KIRC), thyroid (THCA), prostate (PRAD), and pancreatic (PAAD) demonstrating the lowest scores (**Fig. 3A**). These cancer types also appeared to have the smallest differences between normal and cancer samples. The small difference for pancreatic cancer may reflect this cancer type's high impurity.

Using seven cancer types where each of the four stages had at least 10 samples, we found in four of seven tested cancer types (THCA, KIRC, KIRP, LUAD), EXTEND scores increased in high-stage tumors, suggesting higher telomerase activity in advanced stage tumors consistent with previous reports[58-61]. In contrast, stomach adenocarcinoma (STAD) and colorectal adenocarcinoma (CRC) showed the highest scores in stage I tumors (**Fig. 3B**). In melanoma, metastatic cancers exhibited higher scores than primary and regionally invasive tumors (**Fig. 3C**).

Given the correlation between EXTEND score and tumor stage, we further tested its association with tumor molecular subtypes. EXTEND scores were significantly different in at least one subtype classification in all tested cancer types (**Supplementary Fig. 12**), suggesting telomerase activity may be an important factor underlying the molecular heterogeneity of cancer.

We also examined associations between telomerase activity and patient prognosis. We categorized EXTEND scores into low and high tumor groups based on the median and performed univariate and multivariate survival analyses. Five cancer types (ACC, KICH,

KIRP, LUAD and SARC) exhibited worse overall survival in the high score group, whereas stomach adenocarcinoma (STAD) and thymoma (THYM) exhibited the opposite pattern (**Fig. 3D**). However, only ACC and STAD remained significant in the multivariate analysis when controlling for tumor stage and patient age at diagnosis, likely due to the association between telomerase activity and tumor stage.

Next, we correlated EXTEND scores with the ten oncogenic signaling pathways recently curated by PanCan Atlas to identify potential regulators of telomerase[62]. Cell cycle, p53, Myc, and receptor tyrosine kinase (RTK) pathways were largely positively correlated with EXTEND scores, whereas the TGF beta and Wnt pathways were negatively correlated (**Fig. 3E**). The positive correlation between cell cycle genes and EXTEND corroborates the observation that telomere extension by telomerase occurs during cell cycle[42,63]. Myc, Wnt, and TGF beta pathways, specifically c-myc, beta-catenin, and *TGFBR2*, directly regulate telomerase[64-68]. Expression of *PDGFRA*, a marker of mesenchymal cells, was negatively correlated with EXTEND scores in 21 of 31 cancer types. This was in stark contrast to other genes of the RTK pathway, suggesting possible suppression of telomerase activity in cells of mesenchymal origins.

Telomerase activity and cancer stemness are positively correlated

Increasing evidence suggests cancers exhibit stem-cell like characteristics, though it is still controversial if cancer stemness reflects the presence of cancer stem cells or stem cell-associated programs[69]. Since active telomerase is a feature of stem cells, we examined the association between cancer stemness and telomerase activity.

We obtained stemness measurements of TCGA tumors from a recent pan-cancer analysis[70]. This cancer stemness index was calculated from an expression signature of 12,945 genes independently derived by comparing embryonic stem cells and differentiated progenitor cells. We found cancer stemness and EXTEND scores were highly correlated at the cancer type level (Rho=0.85, P=2e-9) (**Fig. 4A**). This significant correlation remained within each cancer type (FDR<0.05), suggesting it was cancer lineage independent. Nine of the 13 EXTEND signature genes overlapped with the stemness signature. However, removing these nine genes from the stemness signature virtually had no impact on the resulting stemness scores as they only constituted less than 0.1% of the stemness signature. To further validate the correlation between stemness scores and EXTEND, we performed a permutation analysis by randomly shuffling gene labels of the expression data. We then calculated empirical p values by comparing observed correlations with those generated from the random permutation; 27 of 31 cancer types remained highly significant (empirical p values < 0.05) (**Supplementary Table 5**).

We next asked if these tumor and tissue level correlations reflected tumor cell behavior. We calculated cancer stemness using single-cell RNAseq data from glioblastoma[71], head and neck cancer[72], and medulloblastoma samples[73]. Though none of the datasets

had the sensitivity to detect *TERT* or *TERC* expression due to the low input materials from single cells, EXTEND successfully estimated telomerase activity scores for each cell. We again observed strong positive correlations between single cell EXTEND scores and cancer stemness ($P < 2.2e-16$) (**Figs. 4B, 4C** and **Supplementary Fig. 13A**), suggesting the associations between the two concepts are inherent to cancer cells.

To characterize these high stemness, high telomerase cells, we identified genes and their corresponding pathways differentially expressed in these cells (**Supplementary Table 6, Supplementary Fig. 14**). Cell cycle, DNA replication, and repair pathways were significantly enriched among all three cancer types (**Fig. 4D**), suggesting these were cycling cells. To further verify this finding, we divided cells into G1_S, G2_M, and non-cycling based on recently published markers[71]. Cells in G1_S phase exhibited significantly higher EXTEND scores than G2_M cells and non-cycling cells (**Figs. 4E, 4F and Supplementary Fig 13B**). Taken together, these data support a model that a group of high stemness, high telomerase cells drive tumor growth.

This model also predicts that telomerase activity is not only a marker for tumor stemness but also a marker for tumor proliferation. Although previously documented[74,75], the relationship between telomerase activity and tumor proliferation has not been quantitatively measured. Using KI67, a cell proliferation marker, we observed a linear positive correlation between tumor proliferation and EXTEND score on cancer type level (Rho=0.9, P=2.5e-11) and in most tumor cohorts (FDR<0.05). Exceptions were two tumor types that originate from reproductive organs, uterine carcinosarcoma (UCS) and testicular germ cell tumors (TGCT) (**Fig. 4G**).

Discussion

TERT has long been recognized as a pivotal determinant of telomerase enzymatic activity. However, it has been increasingly clear that *TERT* is involved in telomere independent functions and its expression has limitations in predicting telomerase activity[23,32-34]. Furthermore, *TERT* has a high GC content (58% vs genome-wide average 41% in human), making it hard to capture in sequencing even for bulk samples. This issue is exacerbated by the low expression of *TERT*, which was estimated to be between 1-40 copies per cell in cancer cells[34]. Indeed, only 73% of TCGA tumors detected *TERT* expression[76], a fraction significantly lower than the TRAP assay-based estimate of telomerase positive tumors (~90%)[22]. In single cell RNA sequencing studies, it is typical that none of *TERT* reads are detected in any cells. Thus, *TERT* expression is inadequate, both biologically and technically, for guiding a systematic analysis of telomerase activity.

We here developed EXTEND to predict telomerase activity based on gene expression data. Using cancer cell lines of various tissues of origin, we show that EXTEND outperformed *TERT* expression in predicting telomerase activity. Analyses of GTEx and

embryonic samples further supported this conclusion. Interestingly, brain tissues exhibited detectable *TERT* expression in the GTEx data but low EXTEND scores, likely due to *TERT* alternative splicing leading to isoforms that do not encode catalytically active proteins. The brain was reported to have an extreme transcriptome diversity due to alternative splicing compared with other tissues[77], thus the excessive alternative splicing of *TERT* is not surprising and unlikely specific to *TERT*. Moreover, EXTEND accurately demonstrated the point where telomerase activity is diminished during fetal heart development. These data substantiate EXTEND's validity and distinguish it from *TERT*. An important reason for EXTEND's robust performance is the eleven marker genes that have no reported functional associations with the telomerase. These genes not only reduce the impact of *TERT* expression on the final scores, but also provide an avenue for estimating telomerase activity even when *TERT* or *TERC* are beyond detection sensitivity in bulk or single cell samples.

Applying EXTEND to >9,000 tumors and 700 normal samples confirmed many previously known associations but also revealed important new insights. Telomerase activity, as quantified by EXTEND, is continuous rather than dichotomous. Variations in telomerase activity across cancer types can be partially explained by *TERT* expression (R=0.54, P=0.003). Strikingly, cancer stemness, a measure reflecting the overall similarity in self-renewal between cancer cells and embryonic stem cells, correlates significantly better than *TERT* expression with telomerase activity (Rho=0.85, P=2e-9). This correlation coefficient indicates 72% of the variation in telomerase activity across cancer types can be explained by their differences in cancer stemness. Single cell analysis confirmed the tight correlation between active telomerase and cancer stemness program, and further revealed the high stemness high telomerase cells were cycling cells. Future studies will be needed to elucidate the identifies of these cells, and whether inhibiting their telomerase activity, or targeting the stemness program, or both can effectively eliminate these cells.

The lack of correlation between telomerase activity and telomere lengths was surprising but could be explained by several reasons. First of all, telomere lengths are determined by the counteracting effects of attrition during cell division and extension by telomerase. Second, telomere lengths used in our analysis were estimated based on abundances of telomeric repeats from sequencing data[76]. At best, short read sequencing data can only estimate the average telomere length of a sample. However, telomerase preferentially acts on the shortest telomeres[5]. Furthermore, a recent study suggests telomeric repeats can frequently insert in non-telomeric regions of the genome [78]. Thus, overall telomeric repeat counts may not precisely measure a tumor's telomere lengths.

In summary, our study demonstrates the feasibility of digitalizing telomerase enzymatic activity, a pathway fundamental to cancer cell and stem cell survival. Our analysis

establishes quantitative associations between telomerase activity and cancer stemness and proliferation.

Acknowledgement

This work was supported by CPRIT (RR170055 to S.Z.) and MD Anderson Cancer Center Brain Cancer SPORE Program Career Enhancement Award (S.Z.). We thank Dr. Zhu Wang (UT Health San Antonio) for discussions on statistical issues and Karen Klein for editorial assistance. The results shown here are in whole or part based upon data generated by the TCGA Research Network. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal.

References

1 Maciejowski, J. & de Lange, T. Telomeres in cancer: tumour suppression and genome instability. *Nat Rev Mol Cell Biol* **18**, 175-186, doi:10.1038/nrm.2016.171 (2017).

2 Stewart, S. A. & Weinberg, R. A. Telomeres: cancer to human aging. *Annu Rev Cell Dev Biol* **22**, 531-557, doi:10.1146/annurev.cellbio.22.010305.104518 (2006).

3 Shay, J. W. & Wright, W. E. Hallmarks of telomeres in ageing research. *J Pathol* **211**, 114-123, doi:10.1002/path.2090 (2007).

4 Aviv, A. & Shay, J. W. Reflections on telomere dynamics and ageing-related diseases in humans. *Philos Trans R Soc Lond B Biol Sci* **373**, doi:10.1098/rstb.2016.0436 (2018).

5 Chiba, K. *et al.* Mutations in the promoter of the telomerase gene TERT contribute to tumorigenesis by a two-step mechanism. *Science* **357**, 1416-1420, doi:10.1126/science.aao0535 (2017).

6 Artandi, S. E. & DePinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9-18, doi:10.1093/carcin/bgp268 (2010).

7 Greider, C. W. Telomerase is processive. *Mol Cell Biol* **11**, 4572-4580, doi:10.1128/mcb.11.9.4572 (1991).

8 de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* **19**, 2100-2110, doi:10.1101/gad.1346005 (2005).

9 Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236-1241, doi:10.1126/science.7544491 (1995).

10 Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-555, doi:10.1038/990141 (1999).

11 Avilion, A. A. *et al.* Human telomerase RNA and telomerase activity in immortal cell lines and tumor tissues. *Cancer Res* **56**, 645-650 (1996).

12 Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959, doi:10.1126/science.1229259 (2013).

13 Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961, doi:10.1126/science.1230062 (2013).
14 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 U S A* **110**, 6021-6026, doi:10.1073/pnas.1303607110 (2013).
15 Vinagre, J. *et al.* Frequency of TERT promoter mutations in human cancers. *Nat Commun* **4**, 2185, doi:10.1038/ncomms3185 (2013).
16 Bell, R. J. *et al.* Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **348**, 1036-1039, doi:10.1126/science.aab0015 (2015).
17 Stern, J. L., Theodorescu, D., Vogelstein, B., Papadopoulos, N. & Cech, T. R. Mutation of the TERT promoter, switch to active chromatin, and monoallelic TERT expression in multiple cancers. *Genes Dev* **29**, 2219-2224, doi:10.1101/gad.269498.115 (2015).
18 Huang, F. W. *et al.* TERT promoter mutations and monoallelic activation of TERT in cancer. *Oncogenesis* **4**, e176, doi:10.1038/oncsis.2015.39 (2015).
19 Akincilar, S. C. *et al.* Long-Range Chromatin Interactions Drive Mutant TERT Promoter Activation. *Cancer Discov* **6**, 1276-1291, doi:10.1158/2159-8290.CD-16-0177 (2016).
20 Stern, J. L. *et al.* Allele-Specific DNA Methylation and Its Interplay with Repressive Histone Marks at Promoter-Mutant TERT Genes. *Cell Rep* **21**, 3700-3707, doi:10.1016/j.celrep.2017.12.001 (2017).
21 Borah, S. *et al.* Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. *Science* **347**, 1006-1010, doi:10.1126/science.1260200 (2015).
22 Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015 (1994).
23 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, doi:10.1093/nar/gku560 (2014).
24 Ludlow, A. T. *et al.* Quantitative telomerase enzyme activity determination using droplet digital PCR with single cell resolution. *Nucleic Acids Res* **42**, e104, doi:10.1093/nar/gku439 (2014).
25 Fernandez-Marcelo, T. *et al.* Clinical Relevance of Telomere Status and Telomerase Activity in Colorectal Cancer. *PLoS One* **11**, e0149626, doi:10.1371/journal.pone.0149626 (2016).
26 Werner, C. M. *et al.* Differential effects of endurance, interval, and resistance training on telomerase activity and telomere length in a randomized, controlled study. *Eur Heart J* **40**, 34-46, doi:10.1093/eurheartj/ehy585 (2019).
27 Bodnar, A. G. *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349-352, doi:10.1126/science.279.5349.349 (1998).
28 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, doi:10.1038/sj.onc.1201882 (1998).

29 Garrels, W. *et al.* Ectopic expression of human telomerase RNA component results in increased telomerase activity and elongated telomeres in bovine blastocysts. *Biol Reprod* **87**, 95, doi:10.1095/biolreprod.112.100198 (2012).
30 Hrdlickova, 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, doi:10.1128/MCB.00550-12 (2012).
31 Ludlow, A. T. *et al.* NOVA1 regulates hTERT splicing and cell growth in non-small cell lung cancer. *Nat Commun* **9**, 3112, doi:10.1038/s41467-018-05582-x (2018).
32 Ludlow, A. T., Slusher, A. L. & Sayed, M. E. Insights into Telomerase/hTERT Alternative Splicing Regulation Using Bioinformatics and Network Analysis in Cancer. *Cancers (Basel)* **11**, doi:10.3390/cancers11050666 (2019).
33 Ulaner, G. A., Hu, J. F., Vu, T. H., Giudice, L. C. & Hoffman, A. R. Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer Res* **58**, 4168-4172 (1998).
34 Rowland, T. J., Dumbovic, G., Hass, E. P., Rinn, J. L. & Cech, T. R. Single-cell imaging reveals unexpected heterogeneity of telomerase reverse transcriptase expression across human cancer cell lines. *Proc Natl Acad Sci U S A* **116**, 18488-18497, doi:10.1073/pnas.1908275116 (2019).
35 Liu, K., Hodes, R. J. & Weng, N. Cutting edge: telomerase activation in human T lymphocytes does not require increase in telomerase reverse transcriptase (hTERT) protein but is associated with hTERT phosphorylation and nuclear translocation. *J Immunol* **166**, 4826-4830, doi:10.4049/jimmunol.166.8.4826 (2001).
36 Vulliamy, T. *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432-435, doi:10.1038/35096585 (2001).
37 Farooqi, A. *et al.* Identification of patient-derived glioblastoma stem cell (GSC) lines with the alternative lengthening of telomeres phenotype. *Acta Neuropathol Commun* **7**, 76, doi:10.1186/s40478-019-0732-4 (2019).
38 Ackermann, S. *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* **362**, 1165-1170, doi:10.1126/science.aat6768 (2018).
39 Heaphy, C. M. *et al.* Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**, 425, doi:10.1126/science.1207313 (2011).
40 Cancer Genome Atlas Research, N. *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med* **372**, 2481-2498, doi:10.1056/NEJMoa1402121 (2015).
41 Thijssen, P. E. *et al.* Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat Commun* **6**, 7870, doi:10.1038/ncomms8870 (2015).
42 Vogan, J. M. & Collins, K. Dynamics of Human Telomerase Holoenzyme Assembly and Subunit Exchange across the Cell Cycle. *J Biol Chem* **290**, 21320-21335, doi:10.1074/jbc.M115.659359 (2015).

43 Zhao, W. *et al.* Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419, doi:10.1186/1471-2164-15-419 (2014).
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, doi:10.1016/j.ajpath.2011.06.018 (2011).
45 Lafferty-Whyte, K. *et al.* A gene expression signature classifying telomerase and ALT immortalization reveals an hTERT regulatory network and suggests a mesenchymal stem cell origin for ALT. *Oncogene* **28**, 3765-3774, doi:10.1038/onc.2009.238 (2009).
46 Demanelis, K. *et al.* Determinants of telomere length across human tissues. *bioRxiv*, 793406, doi:10.1101/793406 (2019).
47 Leri, A., Rota, M., Pasqualini, F. S., Goichberg, P. & Anversa, P. Origin of cardiomyocytes in the adult heart. *Circ Res* **116**, 150-166, doi:10.1161/CIRCRESAHA.116.303595 (2015).
48 Cordero-Espinoza, L. & Huch, M. The balancing act of the liver: tissue regeneration versus fibrosis. *J Clin Invest* **128**, 85-96, doi:10.1172/JCI93562 (2018).
49 Lin, S. *et al.* Distributed hepatocytes expressing telomerase repopulate the liver in homeostasis and injury. *Nature* **556**, 244-248, doi:10.1038/s41586-018-0004-7 (2018).
50 Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505-509, doi:10.1038/s41586-019-1338-5 (2019).
51 Ulaner, G. A. & Giudice, L. C. Developmental regulation of telomerase activity in human fetal tissues during gestation. *Mol Hum Reprod* **3**, 769-773, doi:10.1093/molehr/3.9.769 (1997).
52 Pavo, I. J. & Michel-Behnke, I. Clinical cardiac regenerative studies in children. *World J Cardiol* **9**, 147-153, doi:10.4330/wjc.v9.i2.147 (2017).
53 Boyraz, B. *et al.* Posttranscriptional manipulation of TERC reverses molecular hallmarks of telomere disease. *J Clin Invest* **126**, 3377-3382, doi:10.1172/JCI87547 (2016).
54 Schmutz, I. & de Lange, T. Shelterin. *Curr Biol* **26**, R397-399, doi:10.1016/j.cub.2016.01.056 (2016).
55 Xu, M., Kiselar, J., Whited, T. L., Hernandez-Sanchez, W. & Taylor, D. J. POT1-TPP1 differentially regulates telomerase via POT1 His266 and as a function of single-stranded telomere DNA length. *Proc Natl Acad Sci U S A* **116**, 23527-23533, doi:10.1073/pnas.1905381116 (2019).
56 Latrick, C. M. & Cech, T. R. POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J* **29**, 924-933, doi:10.1038/emboj.2009.409 (2010).
57 Wang, F. *et al.* The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506-510, doi:10.1038/nature05454 (2007).
58 Albanell, J. *et al.* High telomerase activity in primary lung cancers: association with increased cell proliferation rates and advanced pathologic stage. *J Natl Cancer Inst* **89**, 1609-1615, doi:10.1093/jnci/89.21.1609 (1997).

59 Yoshida, K., Sakamoto, S., Sumi, S., Higashi, Y. & Kitahara, S. Telomerase activity in renal cell carcinoma. *Cancer* **83**, 760-766, doi:10.1002/(sici)1097-0142(19980815)83:4<760::aid-cncr19>3.0.co;2-q (1998).
60 Marchetti, A. *et al.* Telomerase activity as a prognostic indicator in stage I non-small cell lung cancer. *Clin Cancer Res* **5**, 2077-2081 (1999).
61 Cheng, A. J., Lin, J. D., Chang, T. & Wang, T. C. Telomerase activity in benign and malignant human thyroid tissues. *Br J Cancer* **77**, 2177-2180, doi:10.1038/bjc.1998.363 (1998).
62 Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337 e310, doi:10.1016/j.cell.2018.03.035 (2018).
63 Zhu, X. *et al.* Cell cycle-dependent modulation of telomerase activity in tumor cells. *Proc Natl Acad Sci U S A* **93**, 6091-6095, doi:10.1073/pnas.93.12.6091 (1996).
64 Wang, J., Xie, L. Y., Allan, S., Beach, D. & Hannon, G. J. Myc activates telomerase. *Genes Dev* **12**, 1769-1774, doi:10.1101/gad.12.12.1769 (1998).
65 Greenberg, R. A. *et al.* Telomerase reverse transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. *Oncogene* **18**, 1219-1226, doi:10.1038/sj.onc.1202669 (1999).
66 Hoffmeyer, K. *et al.* Wnt/beta-catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549-1554, doi:10.1126/science.1218370 (2012).
67 Li, H., Xu, D., Li, J., Berndt, M. C. & Liu, J. P. Transforming growth factor beta suppresses human telomerase reverse transcriptase (hTERT) by Smad3 interactions with c-Myc and the hTERT gene. *J Biol Chem* **281**, 25588-25600, doi:10.1074/jbc.M602381200 (2006).
68 Yang, H., Kyo, S., Takatura, M. & Sun, L. Autocrine transforming growth factor beta suppresses telomerase activity and transcription of human telomerase reverse transcriptase in human cancer cells. *Cell Growth Differ* **12**, 119-127 (2001).
69 Miranda, A. *et al.* Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc Natl Acad Sci U S A* **116**, 9020-9029, doi:10.1073/pnas.1818210116 (2019).
70 Malta, T. M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354 e315, doi:10.1016/j.cell.2018.03.034 (2018).
71 Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e821, doi:10.1016/j.cell.2019.06.024 (2019).
72 Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624, doi:10.1016/j.cell.2017.10.044 (2017).
73 Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74-79, doi:10.1038/s41586-019-1434-6 (2019).
74 Belair, C. D., Yeager, T. R., Lopez, P. M. & Reznikoff, C. A. Telomerase activity: a biomarker of cell proliferation, not malignant transformation. *Proc Natl Acad Sci U S A* **94**, 13677-13682, doi:10.1073/pnas.94.25.13677 (1997).

75 Greider, C. W. Telomerase activity, cell proliferation, and cancer. *Proc Natl Acad Sci U S A* **95**, 90-92, doi:10.1073/pnas.95.1.90 (1998).
76 Barthel, F. P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet* **49**, 349-357, doi:10.1038/ng.3781 (2017).
77 Zaghlool, A., Ameur, A., Cavelier, L. & Feuk, L. Splicing in the human brain. *Int Rev Neurobiol* **116**, 95-125, doi:10.1016/B978-0-12-801105-8.00005-9 (2014).
78 Sieverling, L. *et al.* Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat Commun* **11**, 733, doi:10.1038/s41467-019-13824-9 (2020).

Figure Legends

Fig. 1. EXTEND Validation. (a) Correlation between EXTEND score and TRAP assay readouts in 28 glioma sphere forming cell lines. Two ALT cell lines are labeled in blue. Spearman correlation was used to calculate p value and Rho for (A-C). **(b)** Correlation between EXTEND score and direct enzymatic assay results in 11 bladder cancer cell lines. **(c)** Correlation between EXTEND score and digital droplet TRAP assay in 15 lung cancer cell lines. **(d)** EXTEND scores across ALT and telomerase positive tumors in Liposarcomas (GSE14533). Both telomerase positive tumors and cell lines show significantly higher EXTEND scores than ALT samples. **(e)** EXTEND scores across the five TMM groups of Neuroblastoma (GSE120572). Telomerase positive groups (MYCN amplification, TERT high and TERT rearrangement) show significantly higher scores than ALT and no-TMM groups.

Fig. 2. EXTEND scores across normal tissues and during embryonic development. (a) EXTEND scores and TERT expression across 52 sub tissues across Genotype-Tissue Expression (GTEx) data. The left y-axis indicates TERT expression while the right y-axis represents EXTEND scores. **(b and c)** TERT expression (blue) and EXTEND scores (red) across tissue development phases in (**b**) heart tissue and (**c**) liver tissue. Red and blue lines represent the mean of EXTEND scores and TERT expression of samples from each age group. **(d)** EXTEND predicts higher telomerase activity in PARN mutant, PARD5 knockout sample (left two bars), whereas no such effect is observed in PARN wildtype iPS cells (right two bars). Data is downloaded from GSE81507.

Fig. 3. EXTEND scores in cancer. (a) EXTEND scores for TERT-expressing tumors and normal samples across 16 TCGA cancer cohorts. **(b)** EXTEND scores across tumor stages. Only tumor types with a minimum of 10 cases for each stage were used in this analysis. P-values were calculated using two-sided t-test. *, P<0.05; **, P< 0.01. **(c)** EXTEND scores of cutaneous melanoma (TCGA SKCM). **(d)** Hazard ratio plot based on univariate cox regression model for 31 cancer types. P-values are highlighted for significant cases only. **(e)** Correlations between EXTEND scores and gene expression of ten oncogenic signaling pathways across 31 cancer types. Significant correlations (FDR < 0.05) are shown in either red (positive correlation) or blue circles (negative correlation). Frequency of each gene's positive and negative correlation patterns across different cancer types is summarized on the right.

Fig. 4. Association between EXTEND score, cancer stemness and proliferation. (**a**) Correlation between telomerase activity and cancer stemness across TCGA cohorts. Node size is proportional to correlation coefficients. In all cancer types, tumor stemness and telomerase activity is significantly correlated (FDR<0.05). (**b-c**) Correlation between EXTEND score and cancer stemness at single cell level in (**b**) glioblastoma and (**c**) head and neck cancer. Each dot represents one cell. (**d**) Top ten pathways enriched in the high stemness, high telomerase cells from head and neck cancer. (**e-f**) EXTEND scores of cycling cells (G1-S phase and G2-M phase) and non-cycling cells in (**e**) glioblastoma (**f**) head and neck cancer. (**g**) Correlation between EXTEND scores and proliferation marker MKI67 expression across 31 cancer types. This correlation is significant in all but 2 cancer types (FDR<0.05).

Supplementary Figure Legends

Supplementary Fig. 1. Schematic of EXpression-based Telomerase Enzymatic activity Detection (EXTEND) tool workflow.

Supplementary Fig. 2. (**a**) Frequencies of ATRX/DAXX alteration and TERT promoter mutation across TCGA cancer cohorts. X-axis indicates sample size. Each bar represents one cancer type from TCGA. (**b**) TERT promoter mutation and ATRX mutations are strongly mutually exclusive in LGG.

Supplementary Fig. 3. Contribution of signature genes towards EXTEND score evaluated using Cancer Cell Line Encyclopedia (CCLE) data. Each dot represents one cell line. TERT/TERC contributes more than other genes in the signature but overall the contribution is less than 20%.

Supplementary Fig. 4. Spearman correlation between PolyA enrichment protocol-based prediction and ribosomal depletion protocol-based prediction. Data from GSE51783 was used for this analysis. X-axis represents PolyA enrichment protocol-based scores, whereas y-axis represents ribosomal depletion protocol-based scores.

Supplementary Fig. 5. Spearman correlation of telomerase activity (measured by TRAP assay) with **(a)** EXTEND scores and **(b)** TERT expression across 28 glioma cell lines. Each dot represents a cell line. X-axis in both panels denotes telomerase activity from TRAP assays. Y-axis is EXTEND score and TERT expression, respectively. GSC5-22 and GSC818 are two ALT lines.

Supplementary Fig. 6. Spearman correlation of telomerase activity (obtained through enzymatic assay) with **(a)** EXTEND scores and **(b)** TERT expression across 11 bladder cancer cell lines. Correlation coefficient (Rho = 0.72, P-value = 0.013) of EXTEND scores versus telomerase activity significantly outperforms TERT expression (Rho= 0.55, P-value = 0.083).

Supplementary Fig. 7. Spearman correlation of telomerase activity (obtained through digital droplet-based TRAP assay) with **(a)** EXTEND scores and **(b)** TERT expression across 15 lung cancer cell lines. Correlation coefficient (Rho = 0.65, P-value = 0.01) of EXTEND scores versus telomerase activity significantly outperforms TERT expression (Rho= 0.48, P-value = 0.07).

Supplementary Fig. 8. Distribution of EXTEND scores across human tissues. Each group represents one tissue type from GTEx. TERT expression is denoted by the left y axis, whereas EXTEND score is denoted by the right y axis.

Supplementary Fig. 9. A screenshot from UCSC Xena shows the percentages of TERT isoforms in brain (purple) versus testis (cyan) across GTEx samples. Most of TERT expressed in the brain are short spliced forms (second last), whereas the proportion of full-length isoform is close to zero. In contrast, about 25-30% of TERT expressed in the testis are full-length isoforms.

Supplementary Fig. 10. Correlation between EXTEND and POT1 across pan-cancer (left). No correlation was observed between POT1 and TERT expression (right). For this

analysis, EXTEND scores and gene expression were z-score transformed in each cancer cohort to remove tissue effect.

Supplementary Fig. 11. Correlation between telomere lengths (obtained from Barthel et al.,2017) and EXTEND scores across 31 cancer types (TCGA) for WGS and WXS platforms. Numbers in each cell (of heatmap) represent spearman correlation coefficient while the color represents adjusted p value (multiple testing via Bonferroni correction).

Supplementary Fig. 12. EXTEND scores across various subtypes in cancer cohorts (TCGA data). P-values were calculated using two-sided t-test.

Supplementary Fig. 13. Correlation between telomerase activity and cancer stemness in medulloblastoma. **(a)** Scatter plot representing correlation (spearman rank method) of EXTEND scores with stemness index for single cell data from Medulloblastoma (Hovestadt et al., 2019). **(b)** Violin plot representing distribution of EXTEND scores across cycling cells (in G1-S and G2-M phase) and noncycling cells in Medulloblastoma data. P-values were calculated using two-sided t-test.

Supplementary Fig. 14. Distribution of EXTEND scores and cancer stemness scores. Their distributions consistently show multi-modes. Based on these distributions, we set EXTEND score to 0.5 as the cutoff to select for high stemness, high telomerase activity cells in the three single cell datasets (accounting for 10% of glioblastoma cells, 17% of head and neck cells, and 14% of medullo cells). Linear regression gives a slope of 0.67 for medulloblastoma, and 0.6 for both glioblastoma and head and neck cancer when considering EXTEND score as the dependent variable. Intercept is close to 0 for all three in linear regression.

Supplementary Table Legends

Supplementary Table 1: Telomerase Signature Genes information.

Supplementary Table 2: Pathway enrichment of telomerase signature genes retrieved through MSigDb.

Supplementary Table 3: EXTEND Scores for samples cross Genotype-Tissue Expression (GTEx) data.

Supplementary Table 4: EXTEND Scores for 31 cancer types in TCGA Pancan.

Supplementary Table 5: EXTEND Scores and Stemness index Correlation based on Permutation Test across TCGA cohort

Supplementary Table 6: Pathways enrichment (MSigDb) for Up and downregulated genes based on High and Low EXTEND Scores for 3 single cell data sets (Glioblastoma, Medulloblastoma and Head and Neck cancers)

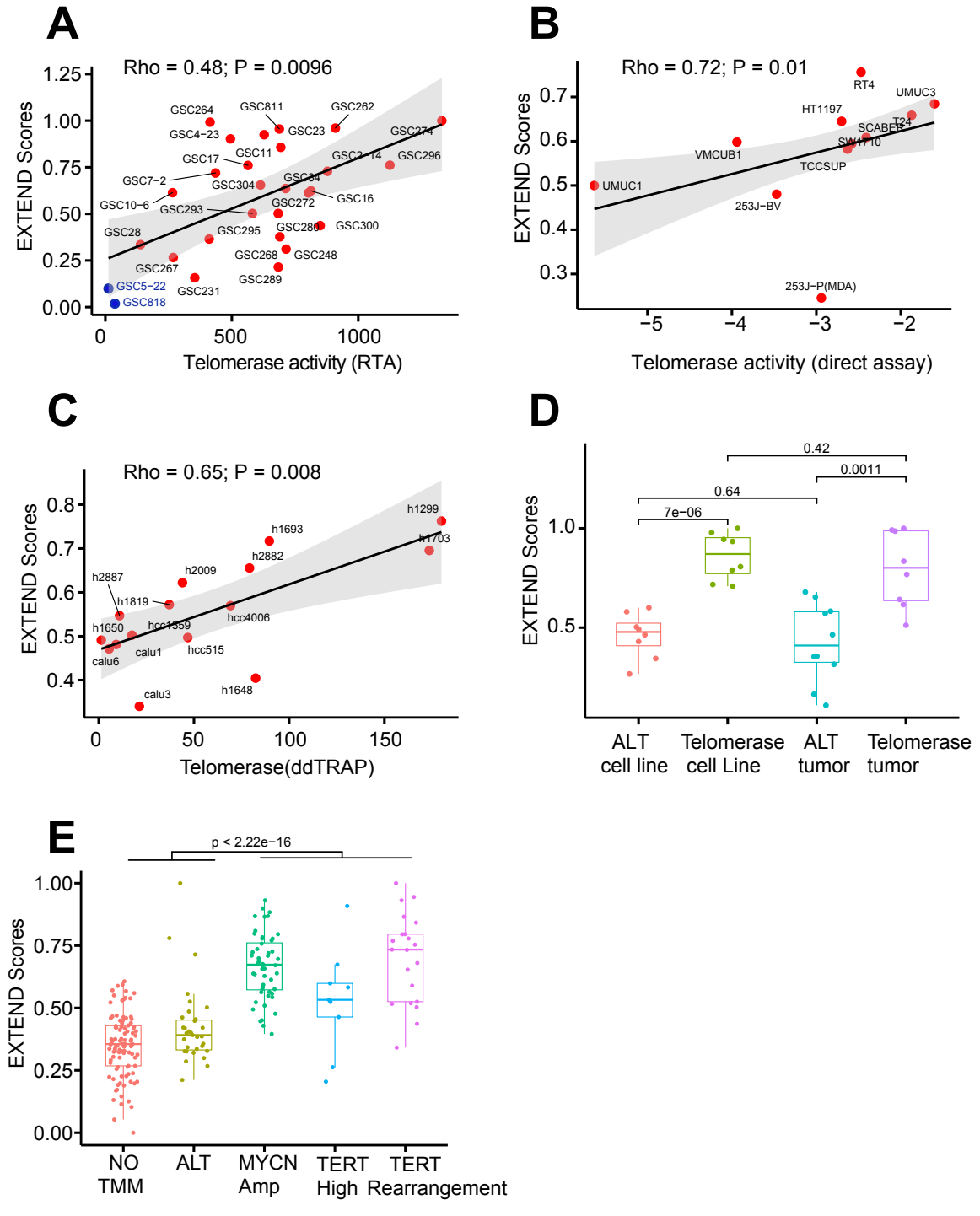

Fig. 1. EXTEND Validation. (**A**) Correlation between EXTEND score and TRAP assay readouts in 28 glioma sphere forming cell lines. Two ALT cell lines are labeled in blue. Spearman correlation was used to calculate p value and Rho for (A-C). (**B**) Correlation between EXTEND score and direct enzymatic assay results in 11 bladder cancer cell lines. (**C**) Correlation between EXTEND score and digital droplet TRAP assay in 15 lung cancer cell lines. (**D**) EXTEND scores across ALT and telomerase positive tumors in Liposarcomas (GSE14533). Both telomerase positive tumors and cell lines show significantly higher EXTEND scores than ALT samples. (**E**) EXTEND scores across the five TMM groups of Neuroblastoma (GSE120572). Telomerase positive groups (MYCN amplification, TERT high and TERT rearrangement) show significantly higher scores than ALT and no-TMM groups.

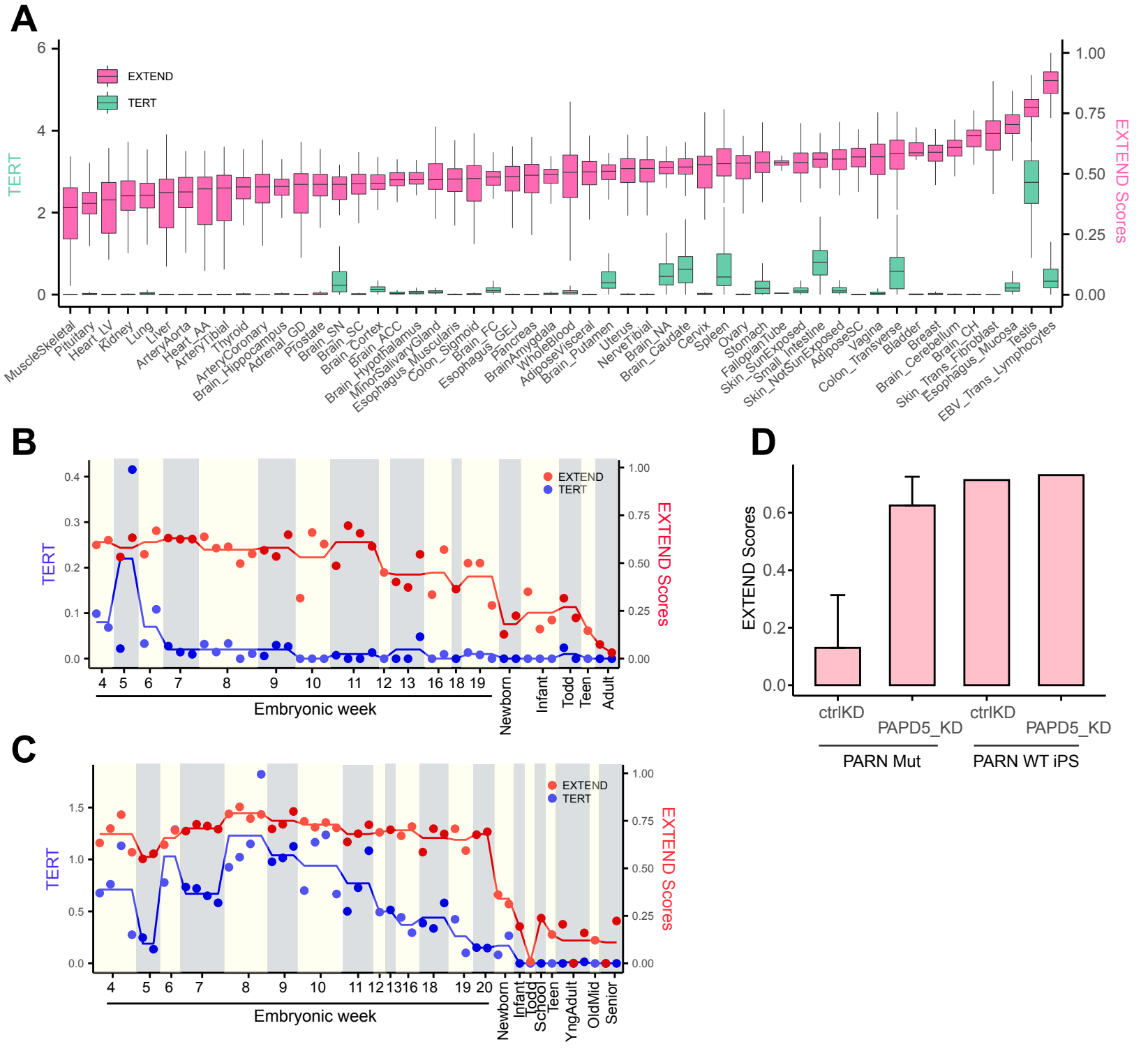

Fig. 2. EXTEND scores across normal tissues and during embryonic development. (**A**) EXTEND scores and TERT expression across 52 sub tissues across Genotype-Tissue Expression (GTEx) data. The left y-axis indicates TERT expression while the right y-axis represents EXTEND scores. (**B and C**) TERT expression (blue) and EXTEND scores (red) across tissue development phases in (B) heart tissue and (C) liver tissue. Red and blue lines represent the mean of EXTEND scores and TERT expression of samples from each age group. (**D**) EXTEND predicts higher telomerase activity in PARN mutant, PARD5 knockout sample (left two bars), whereas no such effect is observed in PARN wildtype iPS cells (right two bars). Data is downloaded from GSE81507.

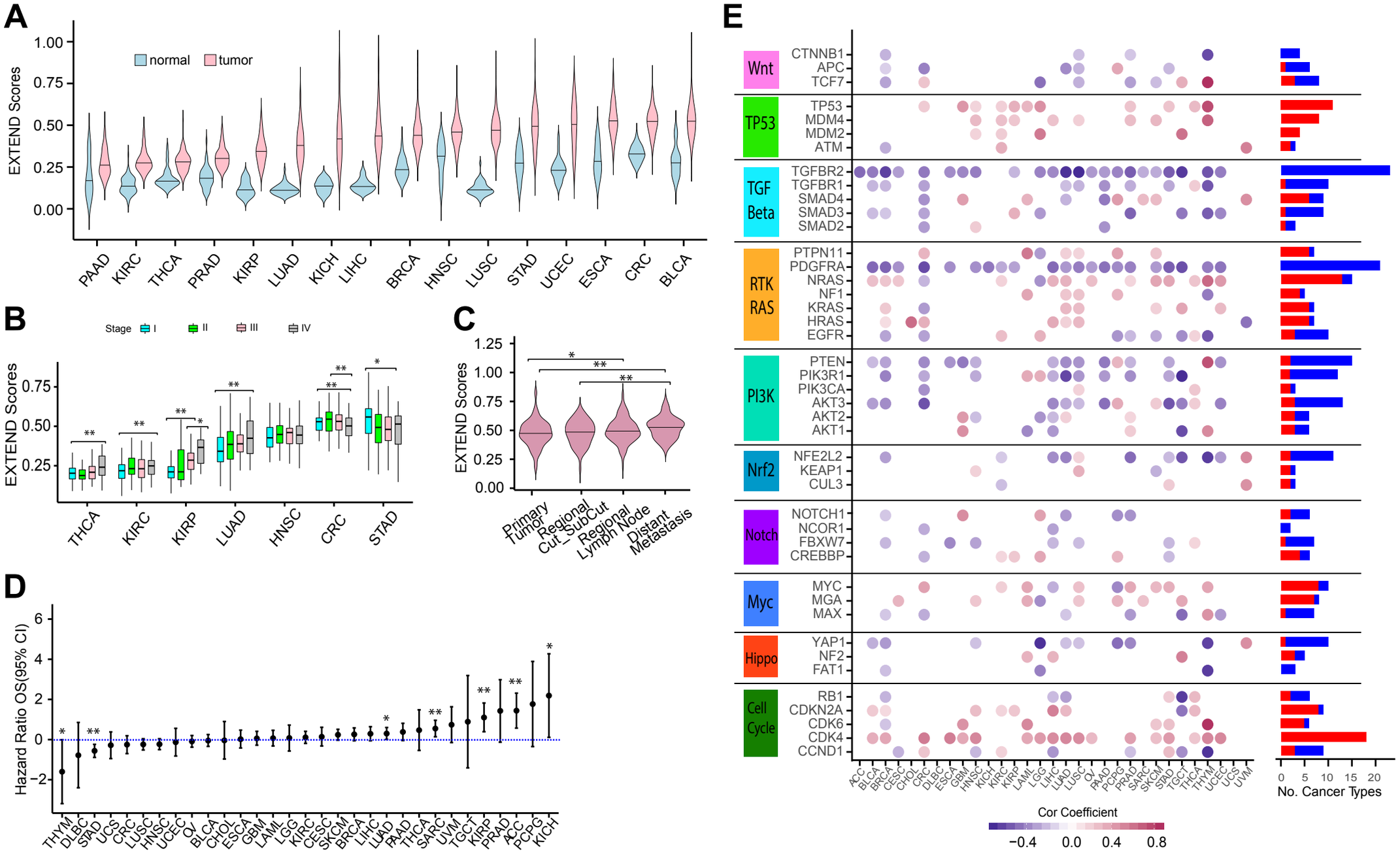

Fig. 3. EXTEND scores in cancer. (**A**) EXTEND scores for TERT-expressing tumors and normal samples across 16 TCGA cancer cohorts. (**B**) EXTEND scores across tumor stages. Only tumor types with a minimum of 10 cases for each stage were used in this analysis. P-values were calculated using two-sided t-test. *, P<0.05; **, P< 0.01. (**C**) EXTEND scores of cutaneous melanoma (TCGA SKCM). (**D**) Hazard ratio plot based on univariate cox regression model for 31 cancer types. P-values are highlighted for significant cases only. (**E**) Correlations between EXTEND scores and gene expression of ten oncogenic signaling pathways across 31 cancer types. Significant correlations (FDR < 0.05) are shown in either red (positive correlation) or blue circles (negative correlation). Frequency of each gene's positive and negative correlation patterns across different cancer types is summarized on the right.

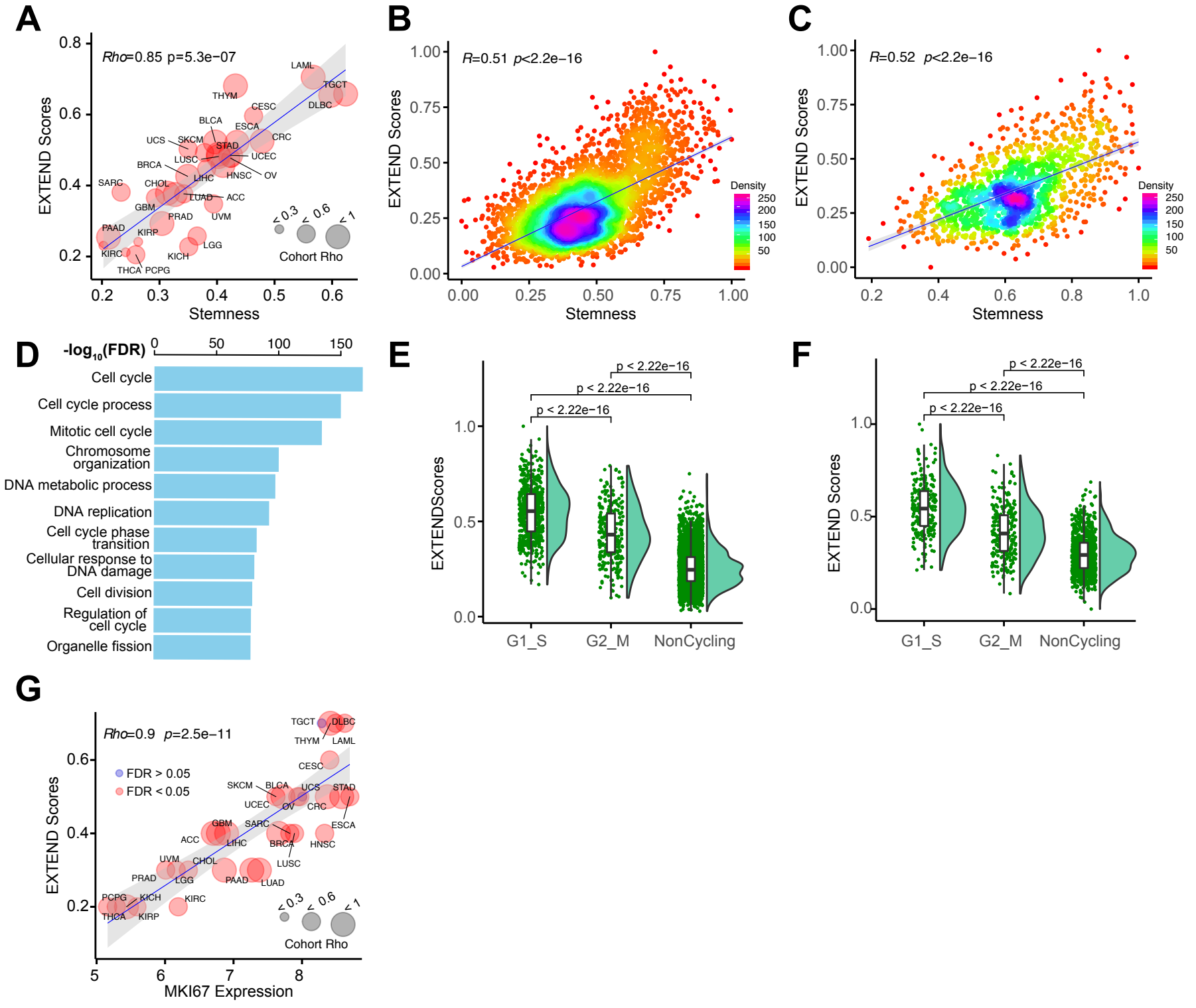

Fig. 4. Association between EXTEND score, cancer stemness and proliferation. (**A**) Correlation between telomerase activity and cancer stemness across TCGA cohorts. Node size is proportional to correlation coefficients. In all cancer types, tumor stemness and telomerase activity is significantly correlated (FDR<0.05). (**B-C**) Correlation between EXTEND score and cancer stemness at single cell level in (B) glioblastoma and (C) head and neck cancer. Each dot represents one cell. (**D**) Top ten pathways enriched in the high stemness, high telomerase cells from head and neck cancer. (E-F) EXTEND scores of cycling cells (G1-S phase and G2-M phase) and non-cycling cells in (**E**) glioblastoma (**F**) head and neck cancer. (**G**) Correlation between EXTEND scores and proliferation marker MKI67 expression across 31 cancer types. This correlation is significant in all but 2 cancer types (FDR<0.05).