TELOMERE LENGTH ALTERATIONS IN MICROSATELLITE STABLE COLORECTAL CANCER AND ASSOCIATION WITH THE IMMUNE RESPONSE

Adriana Lopez-Doriga[1], Rebeca Sanz-Pamplona[1,*], Laura Valle[2], M. Henar Alonso[1], Susanna Aussó[1], Adrià Closa[1], Xavier Sanjuan[3], David Barquero[4], Francisco Rodríguez-Moranta[5], Victor Moreno[1,6,*].

1) Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), Oncobell Program, Bellvitge Biomedical Research Institute (IDIBELL) and CIBERESP, L'Hospitalet de Llobregat, Barcelona, Spain.

2) Hereditary Cancer Program, Catalan Institute of Oncology (ICO), Bellvitge Biomedical Research Institute (IDIBELL) and CIBERONC, L'Hospitalet de Llobregat, Barcelona, Spain.

3) Department of Pathology, University Hospital Bellvitge (HUB-IDIBELL), L´Hospitalet de Llobregat, Barcelona, Spain.

4) Digestive service, Hospital de Sant Joan Despí Moisès Broggi, Sant Joan Despí, Barcelona, Spain.

5) Department of Gastroenterology, University Hospital Bellvitge (HUB-IDIBELL), L´Hospitalet de Llobregat, Barcelona, Spain.

6) Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain.

***Corresponding authors:**

Víctor Moreno and Rebeca Sanz-Pamplona

Av. Gran Vía 199-203. 08908, L'Hospitalet de Llobregat, Barcelona, Spain.

Telephone number: +34 93 260 71 86

E-mail: v.moreno@iconcologia.net

SHORT TITLE: Telomere length and immune response in colon cancer

ARTICLE CATEGORY: Research article

ABBREVIATIONS:

CRC: Colorectal Cancer

TTLR: Tumor Telomere Length Ratio

CN: Copy Number

GSEA: Gene Set Enrichment Analysis

TL: Telomeric length

KEY WORDS: Telomere length; Exome sequencing; Immune response; Colon Cancer

BRIEF DESCRIPTION: A relative measure of telomere length is extracted from the whole-exome sequencing data of colorectal tumor samples. We found compelling associations between the telomere length estimation and genes related to immune response and inflammation.

COMPETING INTERESTS

No potential competing interests were disclosed.

AUTHORS' CONTRIBUTIONS

AL-D, RS-P and MHA carried out the analysis (bioinformatics and statistical) and interpretation of data, AL-D and RS-P and LV carried out the writing, review and/or revision of the manuscript, SA, AC and LV participated in complementary analysis. XS, DB and FRM contributed with material support. VM carried out the conception and supervision of the study.

ACKNOWLEDGEMENTS

We would like to thank Carmen Atencia, Pilar Medina, and Isabel Padrol for their help with the clinical annotation of the samples.

Funding

This work was supported by the Catalan Institute of Oncology, the Instituto de Salud Carlos III and the Spanish Ministry of Economy and Competitivity, co-funded by FEDER funds –a way to build Europe– (grants PI08-1635, PS09-1037, PI11-1439, PI14-613 and SAF2016-80888-R), CIBERESP (grant CB07/02/2005), CIBERONC, and the Catalan Agency for Management of University and Research Grants (AGAUR, grant 2017SGR723). This work was supported by COST Action BM1206.

ABSTRACT

Telomeres are repetitive sequences (TTAGGG) located at the end of chromosomes. Telomeres progressively shorten with each cell replication cycle, ultimately leading to chromosomal instability and loss of cell viability. Telomere length anomaly appears to be one of the earliest and most prevalent genetic alterations in malignant transformation. Here we aim to estimate telomere length from whole-exome sequencing data in colon tumors and normal colonic mucosa, and to analyze the potential association of telomere length with clinical factors and gene expression in colon cancer.

Reads containing at least five repetitions of the telomere sequence (TTAGGG) were extracted from the raw sequences of 42 adjacent normal-tumor paired samples. The number of reads from the tumor sample was normalized to build the Tumor Telomere Length Ratio (TTLR), considered an estimation of telomere length change in the tumor compared to the paired normal tissue. We evaluated the associations between TTLR and clinical factors, gene expression and copy number (CN) aberrations measured in the same tumor samples.

Colon tumors showed significantly shorter telomeres than their paired normal samples. No significant association was observed between TTLR and gender, age, tumor location, prognosis, stromal infiltration or molecular subtypes. The functional gene set enrichment analysis showed pathways related to immune response significantly associated with TLLR.

By extracting a relative measure of telomere length from whole-exome sequencing data, we have assessed that colon tumor cells predominantly shorten telomeres, and this alteration is associated with expression changes in genes related to immune response and inflammation in tumor cells.

INTRODUCTION

Telomeres are sequences located at the end of the chromosomes composed by large tandem repeats of six nucleotides (TTAGGG). Telomeres protect the chromosome ends from nucleolytic degradation, end-to-end fusions and irregular recombination, thus being critical for genome stability and integrity. Telomere length is variable, typically 10-15-kb in humans [1], and decreases with each replication cycle. When telomeres reach a critically short length, the affected cells undergo apoptosis, senescence or acquire chromosomal structural abnormalities [2]. Telomerase catalyzes the de novo addition of telomere repeat sequences onto chromosome ends and it is usually inactive in normal somatic cells [3].

Telomere abnormal maintenance has been related to cancer [1]. However, telomeres and telomerase have a dual function and play important roles in both suppressing and facilitating malignant transformation. On the one hand, increased telomere shortening causes genomic instability, which may favor tumor development and progression [4]. On the other hand, cancer cells may show telomerase activity, which prevents further telomere shortening, thus maintaining tumor cell viability [2]. Some tumors that lack telomerase activity are able to maintain or elongate their telomeres through a phenomenon known as ALT (alternative lengthening of telomeres) [5]. In colorectal cancer (CRC), previous studies have demonstrated the presence of shortened telomeres [6-9]. Indeed, most studies have reported this telomere shortening as a main point in the process of colorectal carcinogenesis [10, 11].

Also, telomere length has been proposed as a prognostic indicator in solid tumors like breast, prostate, colorectal, hepatocellular, lung, esophagus, head and neck, renal, neuroblastoma, glioblastoma or sarcoma, as poor survival is observed in patients with longer telomere length in tumor respect to non-tumor tissue [12]. However, this is controversial in CRC, since the meta-analysis by Wang *et al.* did not find an association between long telomere length and overall survival or disease free survival [13]. However, the analysis by subgroups did show an association between long telomere length and poor overall survival for studies restricted to Europe and for patients older than 60 years of age [13].

Interestingly, recent data have shown an association between chronic inflammation and telomere length. Oxidative stress, caused by inflammation, was associated with accelerated telomere shortening [14]. Inflammation is an important component in the initiation and progression of some cancers [15]. In particular, molecular pathobiology of CRC implicates pro-inflammatory conditions to promote the tumor malignant progression, invasion, and metastasis [16]. Thus, in the context of CRC, chronic inflammation may contribute to telomere shortening.

Several analytical techniques have been used for telomere length measurement or estimation: Southern blot, quantitative PCR, flow cytometry with fluorescence in situ hybridization (FISH), single or universal single telomere length analysis (STELA), among others. [17]. Nowadays, next generation sequencing approaches have revolutionized cancer genomics research by providing fast and accurate information about individual tumors [18]. In this line, telomere length can be estimated from whole-genome sequencing data by counting the number of reads of the telomere repeat sequence [19, 20].

Here, we estimate the telomere length from whole-exome sequencing data and evaluate a potential association of tumor telomere length with the clinical features and the gene expression in colon cancer. In an attempt to find a molecular mechanism explaining telomere shortening in CRC, we also provide functional information about the genes most correlated with telomere length.

METHODS

1. Patients and samples

A subset of 42 paired adjacent normal and tumor tissues (84 samples), from a previously described set of 100 patients with stage II microsatellite stable colon cancer were included in the study [21] (hereafter "CLX subset", *http://www.colonomics.org*; NCBI BioProject PRJNA188510; Table 1). All patients were recruited at the Bellvitge University Hospital (Barcelona, Spain). Tumor samples were collected at the anatomopathological department of the hospital by expert pathologists an included in the tumor biobank. In addition, samples for this study were analyzed by a pathologist (XS) to confirm at least 80% of purity of the tumors and no pathology of the paired normal mucosa. Patients were not treated with adjuvant chemotherapy and all had a minimum of three years of follow-up after the date of diagnosis. Written informed consent was obtained from all patients and the IDIBELL Ethics Committee approved the protocol (PR178/11).

In addition, a whole-exome sequenced validation dataset was downloaded from The Cancer Genome Atlas repository (TCGA, *http://cancergenome.nih.gov/*) comprising 29 paired adjacent normal CRC tissues, 10 of them were discarded due to heterogeneous sequencing parameters as coverage or sequencing quality, so the final validation subset was composed by 19 paired

samples (hereafter "TCGA subset"; Table 1).

2. Exome sequencing pipeline and telomere length estimation

Genomic DNA from the 42 adjacent-normal and tumor paired samples was sequenced at the National Center of Genomic Analysis (CNAG, Barcelona, Spain) on an Illumina HiSeq2000 platform. Exome capture was performed with the Sure Select XT Human All Exon 50MB kit (Agilent). Tumors' exomes were sequenced at 60X coverage and exomes from adjacent normal tissues were sequenced at 40X (2x75 bp reads). The sequencing analysis pipeline has been previously described in detail [22]. From the raw sequences of each sample (*fastq* files), reads containing at least five repetitions of the telomere sequence (TTAGGG) were extracted. This number of repeats was selected to increase specificity, since some genes contain up to three repeats of the TTAGGG sequence. With this extraction criterion, none of the reads selected contained other sequence pattern than the telomere repeat. To avoid artifacts due to varying inter-sample sequencing depths, the number of extracted telomere reads was divided by the corresponding total number of reads in each sample. The "*normalized number of reads in tumor sample*" is considered as an estimation of telomere length in tumors [23, 24]. This measure divided by the "*normalized number of reads in the paired normal sample*" was therefore denominated as tumor telomere length ratio (TTLR). As Tumor and Normal tissue samples were derived from the same patient, TTLRs were consequently adjusted for any putative age and gender effects when different groups of samples were compared.

3. Copy Number Analysis

Data for copy number (CN) alterations for these samples was extracted from the Affymetrix Genome-Wide Human SNP Array 6.0, which contains 906,600 SNP probes and 946,000 non-polymorphic probes across the entire genome. Data on hybridization intensity for each probe were used to identify CN regions and to segment the genome. Segmentation calculations were analyzed by VEGA's method [25]. This algorithm splits the set of ordered data into regions of adjacent elements with similar values and assigns a unique value representing the average ratio of each region respect to the expected in normal diploid tissue DNA. Ratios that exceeded ± 0.3 were considered gains or losses of CN [26]. Two measures were performed, first a global chromosomal aberration index, that included focal and broad CN changes, including aneuploidy. This index was calculated as the proportion genome with gains or losses for each sample. Second, specific CN alterations at the ends of each chromosome were also calculated, using the average of the CN values for the more distal segments of each chromosome arm.

4. Gene expression

Gene expression data was available for the same samples, both tumor and paired normal mucosa (data available on GEO repository as dataset GSE44076). Gene expression had been generated with the Affymetrix Human Genome U219 Array, that contains over 36000 probes corresponding to more than 20000 genes. However, only 15317 genes were used for analysis after applying a filter to remove those with very low variability among samples (sd<0.2).

5. *In silico* functional analysis

Gene Set Enrichment Analysis (GSEA) was used on a pre-ranked gene list sorted by the absolute value of the partial correlation coefficient between TTLR and expression of each gene [27]. For GSEA, the databases *KEGG*, *Biocarta*, *Reactome*, *Hallmark*, *Positional*, and *Immunologic* were interrogated.

6. Proportion of stroma estimation, immune cells infiltration and molecular subtyping

An estimation of the proportion of stroma and the level of infiltrating immune cells in each tumor sample were calculated using the ESTIMATE R package [28]. Tumors were classified according to the molecular consensus subtyping [29] into four groups: CMS1, associated to microsatellite instability phenotype, comprises tumors with better prognosis, and shows activation of immune pathways; CMS2, characterized by a high chromosomal instability (CIN) and strong WNT/MYC pathways activation; CMS3, that show low CIN, but are generally KRAS mutant and have activated pathways related to energy metabolism; and CMS4, that are more mesenchymal and show up-regulation of TGF-β signaling. The CMSclassifier R package was used to obtain this classification using a Random Forest approach.

7. Statistics

Wilcoxon non-parametric test was used to analyze the differences of estimated telomere lengths between tumor and normal. Spearman's rho was calculated to assess the correlation between telomere length in normal tissue and patient's age. The associations between TTLR and clinical factors such as gender, tumor location and prognosis were analyzed by Mann-Whitney tests. Likewise, molecular CRC subtypes, number of somatic mutations or *APC, KRAS*, and *TP53* status, extracted from exome sequencing results, were analyzed with non-parametric tests. The association between TTLR and the gene expression was assessed with partial correlation, using a rank transformation of the ratio of gene expression in the tumor over the normal mucosa. The analyses were adjusted for the CN, estimated at the region where the corresponding gene was located.

RESULTS

Telomere length in normal and tumor samples and association with clinical and molecular characteristics

Colon tumors had shorter telomeres than paired normal samples as shown in Figure 1A, with a median estimated telomere length of 6.9 and 10.9, respectively (difference 3.6; 95%CI 2.59 – 4.57; P-value ≤ 0.01). All except two samples exhibited shorter estimated telomere length in tumor than their paired normal tissue (Suppl. Table 1). Neither of these two samples presented microsatellite instability or were *POLE*-mutated, which might have suggested specific tumor characteristics. When the TCGA paired subset data were analyzed, a similar pattern was observed. The median of telomere length estimation in tumor was also lower than in normal samples, with medians of 0.87 and 1.94, respectively, (difference 0.63, 95%CI 0.16 – 1.15, P-value =0.032) (Suppl. Table 1). Figure 1B shows the differences between tumor and normal samples in the TCGA paired subset, with a profile comparable to CLX samples. Regarding tumor telomere length ratio (TTLR), the mean in CLX data was 0.71 (95%CI 0.13 – 1.29), whereas in the TCGA subset, the mean TTLR was 0.70 (95%CI -0.19 – 1.61).

No significant association was observed between TTLR and gender, age, tumor location or prognosis (Table 2). Global CN aberrations, including aneuploidy, were related to TTLR, but the association was restricted to gains in CN, which correlated with lower TTLR (r=-0.48; P-value=0.0014). We explored if CN aberrations at the ends of the chromosomes could be a confounding factor for other analysis, but no correlation was found between the sum of CN alterations estimated at the ends of chromosomes and telomere length estimation in tumors (r=0.019; P-value=0.90).

No association between TTLR and the mutation status of *APC*, *KRAS* or *TP53* was observed (Table 2). However, a negative correlation between TTLR and the total number of somatic mutations was identified (r=-0.42; P-value=0.005), implying that tumors with higher number of somatic mutations also have shorter telomeres.

Stromal infiltration in tumors was first considered as a potential confounding factor, but no significant correlation with TTLR was identified. However, the level of infiltrating immune cells in tumors was slightly correlated with TTLR (r=0.26, P-value=0.095). Regarding molecular subtypes, 5% (n=2) of CLX tumors were classified as CMS1, 36% (n=15) as CMS2, 24% (n=10) as CMS3, 24% (n=10) as CMS4, and 15% (n=5) with no clear classification. No differences in TTLR were observed among CMS subgroups (Table 2, ANOVA P-value=0.84).

Though the analysis of telomere length changes In tumor respect to normal mucosa is probably the best measure because it intrinsically adjust for patient characteristics like age, sex, genetic susceptibility and others, we also performed the same analyses based on the telomere length estimated in tumors only. The same results were obtained as can be seen in Table 2.

Telomere length and differences in gene expression between tumor and normal samples

Gene expression changes between tumor and normal tissue were calculated as the logarithm of the ratio of the gene expressions between tissues: "log(T/N)". Table 3 shows the most significant genes whose expression change was associated with TTLR after multiple comparison correction (FDR Q-value <0.01). These associations were evaluated with a partial Spearman correlation test adjusted by CN variation, because in an initial unadjusted analysis most identified genes were located in chromosome ends, where gains or deletions were most frequently detected. More precisely, the CN of genes located at the beginning of chromosome 1 showed significant positive correlation with TTLR, and the CN of genes located at chromosome 8q22.1 and 8q22.2 showed significant negative correlation with TTLR. These associations might bias the correlation between TTLR and gene expression if not considered as a covariate in partial correlations.

After applying the analysis described above, six genes remained significant with FDR <0.05 (Table 3). Of those, five genes (*TCTN3, RNF7, ATP6V0E1, SNX3* and *UBAP1*) showed a positive correlation between T/N expression change and TTLR, i.e., higher differences in T/N expression were associated with higher TTLR. *TCTN3* showed the highest positive correlation with TTLR (r=0.62). Only *DDX54* gene showed a negative correlation (r=-0.56), thus indicating that higher *DDX54* expression in tumors was associated with lower TTLR (Table 3).

Of note, the T/N expression change of most genes responsible for telomere maintenance, *TERT, RAP1, DKC1, TERF1, TERF2, POT1, TERF2IP* and *TPP1*, did not show a statistically significant correlation with TTLR. Neither the analysis of telomere length assessed only in tumor showed an association with the expression of these genes. The only exception was the shelterin component *TINF2* (TERF1-interacting nuclear factor 2) that showed significant positive correlation (r=0.50; P-value=0.0003), although when corrected for multiple comparisons, this correlation was no longer significant (adjusted P-value=0.15).

***In silico* functional analysis**

Interestingly, a gene set enrichment analysis (GSEA) revealed functions and pathways related to the immune response as highly correlated with telomere length (Bonferroni corrected P-

value<0.05). Some of these were: "*antigen processing and presentation"* and "*intestinal immune network for IGA production"* (KEGG database); "*interferon gamma signaling"*, "*cytokine signaling in immune system"* and "*adaptive immune system"* (Reactome); or "*IL10 pathway"* (Biocarta), among others (Table 4). In addition, "*interferon gamma response* and *inflammatory response"* emerged as significant when the Hallmarks database was interrogated. As an example, Figure 2 shows the enrichment plots for five of the most significant pathways. Most of the significant functions (88%) and all immune-related functions and pathways presented a positive enrichment score, meaning that as TTLR increases, also increases expression differences (T/N).

Based on this result, an enrichment analysis was performed considering gene sets from the Immunological database. In all, 381 out of 4874 interrogated pathways resulted statistically significant (98% of them showed a positive enrichment score), thus reinforcing the finding that telomere length is related to the immune response in colorectal cancer.

Although CN dosage had been taken into account in the analysis, several positional gene sets were also associated with telomere length. The most significant one was the chromosomal region 6p22 (P-value = 0.001). Interestingly, most genes located in this chromosomal region codify for histones suggesting a relation between regulation of expression of such histones and telomere length.

DISCUSSION

By using whole-exome sequencing data for telomere length estimation, we observed telomere shortening in colon tumor samples compared to their corresponding adjacent normal tissues, suggesting that colon tumor cells lose telomere repeats in the replicative process. This finding is in agreement with previous reports [30, 31] and shows that exome-captured short read sequencing data can be used to estimate overall telomere length, as had already been shown for whole-genome sequencing [19, 20].

Telomere shortening has been considered a relevant mechanism in the colorectal carcinogenic process [32-34]. Specifically, telomere dysfunction may be considered a major driving force in generating chromosomal instability [10, 35, 36], a common feature in our samples. Indeed, we observed that CN aberrations, specifically genomic gains, were associated with shortened telomeres. In agreement with that, in a previous analysis of the CN of these colon tumors, we observed that focal CN gains were frequently tumor-exclusive and associated with a higher number of mutations whereas CN loses were recurrent in some chromosomal locations [26].

These results point to a more chromosomal instable phenotype in tumors losing telomeres that mainly generates CN gains. A similar result has been reported for rectal cancer, in which shorter telomeres were observed in chromosomally instable tumors [37]. Moreover, in our analysis, tumors with higher number of somatic mutations also have shorter telomeres, reinforcing the theory that shorter telomeres go along with both chromosomal and mutation instability, as it has been reported in some cancers like myeloid leukemia, breast or esophageal cancer [38, 39].

No clear associations were identified between telomere length and clinical parameters probably due to the small sample size of our study. Previous studies have reported the potential utility of telomere length as a prognostic biomarker in colon cancer when different stages were included, showing shorter telomeres among early-stage tumors. This might indicate that telomere shortening is an early carcinogenic event [30, 31]. The inclusion of only stage II colon tumors in our study prevented us from identifying similar differences. Also, the expected association between age and telomere length measured in normal tissue was not significant in our data. This has also been observed by Valls-Bautista *et al.*, who found no correlation between telomere length and age in their study that included CRC patients with ages between 48 and 82 years old [33]. However, age-related telomere length was reported in a study including patients with a wider range of ages (29-97 years) [40]. Thus, probably the association between telomere length and age is not observed in our data because most of the patients were diagnosed within a relatively short age range (60-80 years) and this association can only be observed when the age range is large.

Since our study measures telomere length in DNA extracted from bulk tumors, one potential bias for this analysis is that normal stromal cells with longer telomeres infiltrating the tumor could alter the TTLR measurement. Telomere length was independent of the proportion of stroma in tumors, but we observed a non-significant positive correlation with an estimation of the level of immune cells infiltration in tumors. Telomere shortening in peripheral blood T-cells has previously been reported to be associated with lung, colon and breast cancer [41, 42]. Thus, we cannot discard that a fraction of the observed correlation may be explained by shared telomere shortening in tumor and T-cells of the microenvironment. In relation to this, CRC has been recently classified according to four different molecular subtypes characterized by distinct biological and clinical features [43]. Although CMS4 subtype is enriched in stromal infiltration [44], no differences in telomere length were identified among molecular subtypes.

The most interesting and novel finding of our study is the association in CRC of telomere length

with the differential expression of several genes related with inflammation and immune response, and with autoimmune pathologies, such as thyroid autoimmune disease, systemic lupus erythematosus, asthma or type I diabetes; among others (Table 4). Other studies also have associated the telomere complex with autoimmune or inflammatory pathways by different mechanisms [45, 46]. In cancer epidemiology, it has been reported that prolonged stress on the immune system may shorten telomere length and result in increased risk of cancer [47, 48]. Telomeres are particularly susceptible to shortening in response to DNA damage, which might be caused by persistent inflammation. Even low-grade but prolonged inflammation is sufficient to induce telomere erosion [49, 50]. TERT can regulate pathways including Wnt/β-catenin (a central one in CRC) and can regulate inflammatory signaling through binding to NF-kB promoter that leads to transcription of genes such as IL-6 or TNF-α [51]. NF-kB is the major driver of inflammatory signaling and it has been reported that this pathway modulate in turns the activity of the telomerase [52]. Moreover, telomeric repeats (TTAGGG) can inhibit CpG binding to TLR9 to impair innate immune activation [53]. In our study, we also found interleukin (IL) signaling pathways associated with telomere length. In agreement, the expression of a number of immunological cytokines has been previously reported to be correlated with the telomere length in renal cell carcinoma, including IL-7, IL-8 and IL-10, which upregulate the activity of telomerase, suggesting that the accumulation of these cytokines in the tumor microenvironment may have effects on the telomeres of tumor cells [54]. Indeed, the IL-10 pathway is one of the most strongly associated with telomere-length in our data. IL-10 is expressed by several immune cells including Tregs, inducing immune suppression by inhibiting cytotoxic T lymphocyte activation [55]. In the same immune context, interferon pathways had been found to be associated with telomere length in our data. In agreement, it has been reported that DNA damage induces interferon production leading to the inhibition of stem cell function in response to telomere shortening [56].

On the other hand, additional functional studies are required to help us understand the identified strong correlation of telomere length with the expression of *TCTN3, RNF7, ATP6V0E1, SNX3, UBAP1* and *DDX54*; genes a priori not related to telomere maintenance or function or to immune response or inflammation. In this regard, a mechanism named TPE-OLD (for “telomere position effect over long distances”) has been described in which telomere length regulates gene expression [57]. One might hypothesize that this chromatin conformation mechanism or a related one could drive gene expression changes in these genes that are not related with telomere length or telomere maintenance.

Despite the interesting results obtained in our study, it may have some limitations. The method used to estimated telomere length from whole-exome sequencing data, was originally developed for whole-genome sequencing data [24]. The telomere sequence is not specifically captured by the Sure Select XT Human All Exon kit, and the telomere reads identified are the result of unspecific DNA fragments captured due to the abundance of the telomere sequence. It was expected that a random representation of the telomere sequences was present in the final library sequenced, and there is not *a priori* reason to believe that the number of reads sequenced may be biased by any third factor other than sampling variation. We assume that the total number of telomere reads is proportional to the average telomere length of the analyzed cells. The method, however, does not allow the estimation of telomere length of individual chromosomes. We had no means to validate telomere length estimation in our samples with a laboratory-based measurement, but we have found good correlation (r=0.26, P-value=0.016) of our measure with other published computational methods that have performed such validation [19].

Observational studies are prone to confounding. We have considered here two important ones: the proportion of stroma in the tumor and CN aberrations affecting the ends of chromosomes. We have discarded the association of telomere length with these factors. Though we found a moderate association with global CN gains, indicating that telomeres are related to chromosomal instability, there was no association with gains or losses in the ends of chromosomes, which might be correlated with telomere length. Finally, our validation dataset is limited in size. In order to avoid variation related to technical issues, only TCGA samples sequenced in an Illumina platform were used for validation purposes. The individual matching of tumors with their corresponding paired normal samples was the main resource to avoid putative biases, which was also complemented by the normalization by the total number of reads. An attempt was made to analyze a larger size of TCGA of tumor samples and compare them to a common pool of normal mucosae in an unpaired analysis, but the results were not satisfactory, because the variation among tumors was very high. The reduced analysis of paired samples, however, showed very consistent results.

CONCLUSION

Our study has shown, using a relative measure of telomere length extracted from whole-exome sequencing data, that telomere shortening occurs in colorectal tumors, and that this alteration is associated with expression changes in genes related to immune response and inflammation, deserving further study.

REFERENCES

[1] M.A. Blasco, Telomeres and human disease: ageing, cancer and beyond, Nature reviews, 6 (2005) 611-622.

[2] K. Collins, J.R. Mitchell, Telomerase in the human organism, Oncogene, 21 (2002) 564-579.

[3] T.A. Morrish, D. Bekbolysnov, D. Velliquette, M. Morgan, B. Ross, Y. Wang, B. Chaney, J. McQuigg, N. Fager, I.P. Maine, Multiple Mechanisms Contribute To Telomere Maintenance, Journal of cancer biology & research, 1 (2013).

[4] W.C. Hahn, Role of telomeres and telomerase in the pathogenesis of human cancer, Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 21 (2003) 2034-2043.

[5] J.D. Henson, A.A. Neumann, T.R. Yeager, R.R. Reddel, Alternative lengthening of telomeres in mammalian cells, Oncogene, 21 (2002) 598-610.

[6] M. Engelhardt, P. Drullinsky, J. Guillem, M.A. Moore, Telomerase and telomere length in the development and progression of premalignant lesions to colorectal cancer, Clinical cancer research : an official journal of the American Association for Cancer Research, 3 (1997) 1931-1941.

[7] R.R. Plentz, S.U. Wiemann, P. Flemming, P.N. Meier, S. Kubicka, H. Kreipe, M.P. Manns, K.L. Rudolph, Telomere shortening of epithelial cells characterises the adenoma-carcinoma transition of human colorectal cancer, Gut, 52 (2003) 1304-1307.

[8] E. Rampazzo, R. Bertorelle, L. Serra, L. Terrin, C. Candiotto, S. Pucciarelli, P. Del Bianco, D. Nitti, A. De Rossi, Relationship between telomere shortening, genetic instability, and site of tumour origin in colorectal cancers, British journal of cancer, 102 (2010) 1300-1305.

[9] F. Maxwell, L.M. McGlynn, H.C. Muir, D. Talwar, M. Benzeval, T. Robertson, C.S. Roxburgh, D.C. McMillan, P.G. Horgan, P.G. Shiels, Telomere attrition and decreased fetuin-A levels indicate accelerated biological aging and are implicated in the pathogenesis of colorectal cancer, Clinical cancer research : an official journal of the American Association for Cancer Research, 17 (2011) 5573-5581.

[10] R. Bertorelle, E. Rampazzo, S. Pucciarelli, D. Nitti, A. De Rossi, Telomeres, telomerase and colorectal cancer, World journal of gastroenterology, 20 (2014) 1940-1950.

[11] C. Pinol-Felis, T. Fernandez-Marcelo, J. Vinas-Salas, C. Valls-Bautista, Telomeres and telomerase in the clinical management of colorectal cancer, Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico, (2016).

[12] U. Svenson, G. Roos, Telomere length as a biological marker in malignancy, Biochimica et biophysica acta, 1792 (2009) 317-323.

[13] W. Wang, L. Zheng, N. Zhou, N. Li, G. Bulibu, C. Xu, Y. Zhang, Y. Tang, Meta-analysis of associations between telomere length and colorectal cancer survival from observational studies, Oncotarget, 8 (2017) 62500-62507.

[14] R.P. Barnes, E. Fouquerel, P.L. Opresko, The impact of oxidative DNA damage and stress on telomere homeostasis, Mechanisms of ageing and development, (2018).

[15] A. Mantovani, P. Allavena, A. Sica, F. Balkwill, Cancer-related inflammation, Nature, 454 (2008) 436-444.

[16] T. Francuz, P. Czajka-Francuz, S. Cison-Jurek, J. Wojnar, [The role of inflammation in colon cancer pathogenesis], Postepy higieny i medycyny doswiadczalnej, 70 (2016) 360-366.

[17] N. Basu, H.G. Skinner, K. Litzelman, R. Vanderboom, E. Baichoo, L.A. Boardman, Telomeres and telomere dynamics: relevance to cancers of the GI tract, Expert review of gastroenterology & hepatology, 7 733-748.

[18] J. Shendure, R.D. Mitra, C. Varma, G.M. Church, Advanced sequencing technologies: methods and goals, Nature reviews, 5 (2004) 335-344.

[19] Z. Ding, M. Mangino, A. Aviv, T. Spector, R. Durbin, U.K. Consortium, Estimating telomere length from whole genome sequence data, Nucleic Acids Res, 42 (2014) e75.

[20] M. Parker, X. Chen, A. Bahrami, J. Dalton, M. Rusch, G. Wu, J. Easton, N.K. Cheung, M. Dyer, E.R. Mardis, R.K. Wilson, C. Mullighan, R. Gilbertson, S.J. Baker, G. Zambetti, D.W. Ellison, J.R. Downing, J. Zhang, P. Pediatric Cancer Genome, Assessing telomeric DNA content in pediatric cancers using whole-genome sequencing data, Genome biology, 13 (2012) R113.
[21] R. Sanz-Pamplona, A. Berenguer, D. Cordero, D.G. Mollevi, M. Crous-Bou, X. Sole, L. Pare-Brunet, E. Guino, R. Salazar, C. Santos, J. de Oca, X. Sanjuan, F. Rodriguez-Moranta, V. Moreno, Aberrant gene expression in mucosa adjacent to tumor reveals a molecular crosstalk in colon cancer, Mol Cancer, 13 (2014) 46.
[22] R. Sanz-Pamplona, A. Lopez-Doriga, L. Pare-Brunet, K. Lazaro, F. Bellido, M.H. Alonso, S. Ausso, E. Guino, S. Beltran, F. Castro-Giner, M. Gut, X. Sanjuan, A. Closa, D. Cordero, F.D. Moron-Duran, A. Soriano, R. Salazar, L. Valle, V. Moreno, Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer, Clinical cancer research : an official journal of the American Association for Cancer Research, 21 (2015) 4709-4718.
[23] L. Nersisyan, A. Arakelyan, Computel: computation of mean telomere length from whole-genome next-generation sequencing data, PloS one, 10 (2015) e0125201.
[24] M. Lee, C.E. Napier, S.F. Yang, J.W. Arthur, R.R. Reddel, H.A. Pickett, Comparative analysis of whole genome sequencing-based telomere length measurement techniques, Methods, 114 (2017) 4-15.
[25] S. Morganella, L. Cerulo, G. Viglietto, M. Ceccarelli, VEGA: variational segmentation for copy number detection, Bioinformatics, 26 (2010) 3020-3027.
[26] M.H. Alonso, S. Ausso, A. Lopez-Doriga, D. Cordero, E. Guino, X. Sole, M. Barenys, J. de Oca, G. Capella, R. Salazar, R. Sanz-Pamplona, V. Moreno, Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component, British journal of cancer, 117 (2017) 421-431.
[27] A. Subramanian, H. Kuehn, J. Gould, P. Tamayo, J.P. Mesirov, GSEA-P: a desktop application for Gene Set Enrichment Analysis, Bioinformatics, 23 (2007) 3251-3253.
[28] K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Treviño, H. Shen, P.W. Laird, D.A. Levine, S.L. Carter, G. Getz, K. Stemke-Hale, G.B. Mills, R.G.W. Verhaak, Inferring tumour purity and stromal and immune cell admixture from expression data, Nature Communications, 4 (2013).
[29] J. Guinney, R. Dienstmann, X. Wang, A. de Reyniès, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster, E. Fessler, F. De Sousa E Melo, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar, The consensus molecular subtypes of colorectal cancer, Nature Medicine, 21 (2015) 1350-1356.
[30] T. Fernandez-Marcelo, A. Sanchez-Pernaute, I. Pascua, C. De Juan, J. Head, A.J. Torres-Garcia, P. Iniesta, Clinical Relevance of Telomere Status and Telomerase Activity in Colorectal Cancer, PloS one, 11 (2016) e0149626.
[31] N. Suraweera, D. Mouradov, S. Li, R.N. Jorissen, D. Hampson, A. Ghosh, N. Sengupta, M. Thaha, S. Ahmed, M. Kirwan, F. Aleva, D. Propper, R.M. Feakins, T. Vulliamy, N.J. Elwood, P. Tian, R.L. Ward, N.J. Hawkins, Z.Z. Xu, P.L. Molloy, I.T. Jones, M. Croxford, P. Gibbs, A. Silver, O.M. Sieber, Relative telomere lengths in tumor and normal mucosa are related to disease progression and chromosome instability profiles in colorectal cancer, Oncotarget, 7 (2016) 36474-36488.
[32] B.R. Druliner, X. Ruan, R. Johnson, D. Grill, D. O'Brien, T.P. Lai, S. Rashtak, D. Felmlee-Devine, J. Washechek-Aletto, A. Malykh, T. Smyrk, A. Oberg, H. Liu, J.W. Shay, D.A. Ahlquist, L.A. Boardman, Time Lapse to Colorectal Cancer: Telomere Dynamics Define the Malignant Potential of Polyps, Clinical and translational gastroenterology, 7 (2016) e188.
[33] C. Valls-Bautista, C. Pinol-Felis, J.M. Rene-Espinet, J. Buenestado-Garcia, J. Vinas-Salas, In colon cancer, normal colon tissue and blood cells have altered telomere lengths, Journal of

surgical oncology, 111 (2015) 899-904.
[34] R.A. Risques, L.A. Lai, C. Himmetoglu, A. Ebaee, L. Li, Z. Feng, M.P. Bronner, B. Al-Lahham, K.V. Kowdley, K.D. Lindor, P.S. Rabinovitch, T.A. Brentnall, Ulcerative colitis-associated colorectal cancer arises in a field of short telomeres, senescence, and inflammation, Cancer research, 71 (2011) 1669-1679.
[35] A.K. Meeker, J.L. Hicks, E.A. Platz, G.E. March, C.J. Bennett, M.J. Delannoy, A.M. De Marzo, Telomere shortening is an early somatic DNA alteration in human prostate tumorigenesis, Cancer research, 62 (2002) 6405-6409.
[36] K. Muraki, K. Nyhan, L. Han, J.P. Murnane, Mechanisms of telomere loss and their consequences for chromosome instability, Frontiers in oncology, 2 (2012) 135.
[37] L.A. Boardman, R.A. Johnson, K.B. Viker, K.A. Hafner, R.B. Jenkins, D.L. Riegert-Johnson, T.C. Smyrk, K. Litzelman, S. Seo, R.E. Gangnon, C.D. Engelman, D.N. Rider, R.J. Vanderboom, S.N. Thibodeau, G.M. Petersen, H.G. Skinner, Correlation of chromosomal instability, telomere length and telomere maintenance in microsatellite stable rectal cancer: a molecular subclass of rectal cancer, PloS one, 8 (2013) e80015.
[38] J.M. Watts, B. Dumitriu, P. Hilden, A. Kishtagari, F. Rapaport, C. Chen, J. Ahn, S.M. Devlin, E.M. Stein, R. Rampal, R.L. Levine, N. Young, M.S. Tallman, Telomere length and associations with somatic mutations and clinical outcomes in acute myeloid leukemia, Leukemia research, 49 (2016) 62-65.
[39] X.D. Hao, Y. Yang, X. Song, X.K. Zhao, L.D. Wang, J.D. He, Q.P. Kong, N.L. Tang, Y.P. Zhang, Correlation of telomere length shortening with TP53 somatic mutations, polymorphisms and allelic loss in breast tumors and esophageal cancer, Oncology reports, 29 (2013) 226-236.
[40] K. Nakamura, E. Furugori, Y. Esaki, T. Arai, M. Sawabe, I. Okayasu, M. Fujiwara, M. Kammori, K. Mafune, M. Kato, M. Oshimura, K. Sasajima, K. Takubo, Correlation of telomere lengths in normal and cancers tissue in the large bowel, Cancer letters, 158 (2000) 179-184.
[41] C. Falci, K. Gianesin, G. Sergi, S. Giunco, I. De Ronch, S. Valpione, C. Solda, P. Fiduccia, S. Lonardi, M. Zanchetta, S. Keppel, A. Brunello, V. Zafferri, E. Manzato, A. De Rossi, V. Zagonel, Immune senescence and cancer in elderly patients: results from an exploratory study, Experimental gerontology, 48 (2013) 1436-1442.
[42] Y. Qian, T. Ding, L. Wei, S. Cao, L. Yang, Shorter telomere length of T-cells in peripheral blood of patients with lung cancer, OncoTargets and therapy, 9 (2016) 2675-2682.
[43] J. Guinney, R. Dienstmann, X. Wang, A. de Reynies, A. Schlicker, C. Soneson, L. Marisa, P. Roepman, G. Nyamundanda, P. Angelino, B.M. Bot, J.S. Morris, I.M. Simon, S. Gerster, E. Fessler, E.M.F. De Sousa, E. Missiaglia, H. Ramay, D. Barras, K. Homicsko, D. Maru, G.C. Manyam, B. Broom, V. Boige, B. Perez-Villamil, T. Laderas, R. Salazar, J.W. Gray, D. Hanahan, J. Tabernero, R. Bernards, S.H. Friend, P. Laurent-Puig, J.P. Medema, A. Sadanandam, L. Wessels, M. Delorenzi, S. Kopetz, L. Vermeulen, S. Tejpar, The consensus molecular subtypes of colorectal cancer, Nat Med, 21 (2015) 1350-1356.
[44] C. Isella, A. Terrasi, S.E. Bellomo, C. Petti, G. Galatola, A. Muratore, A. Mellano, R. Senetta, A. Cassenti, C. Sonetto, G. Inghirami, L. Trusolino, Z. Fekete, M. De Ridder, P. Cassoni, G. Storme, A. Bertotti, E. Medico, Stromal contribution to the colorectal cancer transcriptome, Nature genetics, 47 (2015) 312-319.
[45] S.S. Jose, K. Bendickova, T. Kepak, Z. Krenova, J. Fric, Chronic Inflammation in Immune Aging: Role of Pattern Recognition Receptor Crosstalk with the Telomere Complex?, Frontiers in immunology, 8 (2017) 1078.
[46] S. Georgin-Lavialle, A. Aouba, L. Mouthon, J.A. Londono-Vallejo, Y. Lepelletier, A.S. Gabet, O. Hermine, The telomere/telomerase system in autoimmune and systemic immune-mediated diseases, Autoimmun Rev, 9 (2010) 646-651.
[47] R.B. Effros, Telomere/telomerase dynamics within the human immune system: effect of chronic infection and stress, Experimental gerontology, 46 (2011) 135-140.
[48] N.P. Weng, Telomeres and immune competency, Current opinion in immunology, 24 (2012) 470-475.

[49] T.B. Kryston, A.B. Georgiev, P. Pissis, A.G. Georgakilas, Role of oxidative stress and DNA damage in human carcinogenesis, Mutation research, 711 (2011) 193-201.
[50] M.K. Graham, A. Meeker, Telomeres and telomerase in prostate cancer development and therapy, Nature reviews. Urology, 14 (2017) 607-619.
[51] A. Ghosh, G. Saginc, S.C. Leow, E. Khattar, E.M. Shin, T.D. Yan, M. Wong, Z. Zhang, G. Li, W.K. Sung, J. Zhou, W.J. Chng, S. Li, E. Liu, V. Tergaonkar, Telomerase directly regulates NF-kappaB-dependent transcription, Nature cell biology, 14 (2012) 1270-1281.
[52] D. Jurk, C. Wilson, J.F. Passos, F. Oakley, C. Correia-Melo, L. Greaves, G. Saretzki, C. Fox, C. Lawless, R. Anderson, G. Hewitt, S.L. Pender, N. Fullard, G. Nelson, J. Mann, B. van de Sluis, D.A. Mann, T. von Zglinicki, Chronic inflammation induces telomere dysfunction and accelerates ageing in mice, Nat Commun, 2 (2014) 4172.
[53] I. Gursel, M. Gursel, H. Yamada, K.J. Ishii, F. Takeshita, D.M. Klinman, Repetitive elements in mammalian telomeres suppress bacterial DNA-induced immune activation, Journal of immunology, 171 (2003) 1393-1400.
[54] U. Svenson, E. Gronlund, I. Soderstrom, R.T. Sitaram, B. Ljungberg, G. Roos, Telomere length in relation to immunological parameters in patients with renal cell carcinoma, PloS one, 8 (2013) e55543.
[55] A. Ben-Baruch, Inflammation-associated immune suppression in cancer: the roles played by cytokines, chemokines and additional mediators, Seminars in cancer biology, 16 (2006) 38-52.
[56] Q. Yu, Y.V. Katlinskaya, C.J. Carbone, B. Zhao, K.V. Katlinski, H. Zheng, M. Guha, N. Li, Q. Chen, T. Yang, C.J. Lengner, R.A. Greenberg, F.B. Johnson, S.Y. Fuchs, DNA-damage-induced type I interferon promotes senescence and inhibits stem cell function, Cell reports, 11 (2015) 785-797.
[57] J.D. Robin, A.T. Ludlow, K. Batten, F. Magdinier, G. Stadler, K.R. Wagner, J.W. Shay, W.E. Wright, Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances, Genes & development, 28 (2014) 2464-2476.

Table 1) Clinical factors for CLX and TCGA subsets.

	CLX subset patients (n=42)	**TCGA subset patients (n=19)**
Gender		
Male	31 (73.8 %)	12 (63 %)
Female	11 (26.2 %)	7 (37 %)
Median age (range, years)	70 (43 ; 84)	73 (51 ; 90)
Site		
Right	12 (28.6 %)	10 (53 %)
Left	30 (71.4 %)	9 (47 %)
Recurrence-free median time *(range, months)*	60.7 (6.75 – 127.4)	
Stage		
I		2 (10.5 %)
II	42 (100 %)	7 (36.8 %)
III		7 (36.8 %)
IV		3 (15.7 %)
Microsatellite instability		
MSS	42 (100%)	17 (89.5%)
MSI	0	2 (10.5 %)

Table 2) Associations between TTLR and clinical and mutation

		TTLR			Tumor TL		
	N (%)	**Mean (sd)**	**Mean dif. (95%CI)**	**P-value**	**Mean (sd)**	**Mean dif. (95%CI)**	**P-value**
Gender							
Male	31 (73.8 %)	0.65 (0.26)			11.33 (4.91)		
Female	11 (26.2 %)	0.82 (0.38)	0.18 (-0.08 ; 0.44)	0.12	15.1 (6.05)	3.77 (-0.58; 8.06)	0.04
Tumor site							
Right	12 (28.6 %)	0.64 (0.20)			11.43 (3.9)		
Left	30 (71.4 %)	0.71 (0.34)	0.07(-0.10 ; 0.64)	0.40	12.6 (5.94)	1.23 (-1.96 ; 4.42)	0.69
Prognosis							
No Relapse	21 (50%)	0.69 (0.30)			12.24 (5.45)		
Relapse	21 (50%)	0.70 (0.31)	0.01(-0.18 ; 0.20)	0.94	12.38 (5.51)	0.15 (-3.26 ; 3.27)	0.88
APC							
WT	12 (28,6%)	0.74 (0.22)			12.98 (4.74)		
Mut	30 (71,4%)	0.67 (0.33)	-0.07(-0.25; 0.11)	0.43	12.04 (5.72)	-0.94 (-2.62 ; 4.48)	0.44
KRAS							
WT	28 (66,7%)	0.73 (0.29)			12.8 (4.79)		
Mut	14 (33,3%)	0.61 (0.32)	-0.12(-0.33 ; 0.09)	0.26	11.3 (6.58)	1.47 (-2.65 ; 5.59)	0.10
TP53							
WT	24 (57.1%)	0.68 (0.21)			12.49 (3.99)		
Mut	18 (42.9%)	0.71 (0.40)	0.03(-0.18 ; 0.25)	0.73	12.18 (7.01)	-0.31 (-4.1 ; 3.48)	0.55
Molecular subtype				0.84			0.77
CMS1	2 (5%)	0.55 (0.25)			10.39 (3.84)		
CMS2	15 (36%)	0.73 (0.32)			13.53 (7.25)		
CMS3	10 (24%)	0.74 (0.39)			13.15 (5.03)		
CMS4	5 (15%)	0.66 (0.23)			11.5 (3.69)		
			Spearman rho (95%CI)	**P-value**		**Spearman rho (95%CI)**	**P-value**
Age			-0.049 (-0.36 ; 0.26)	0.75		0.116 (-0.19 ; 0.43)	0.46
Global CN gains*			-0.31 (-0.62 ; 0.00)	0.048		-0.42 (-0.73 ; -0.11)	0.05
Global CN loses*			0.09 (-0.40 ; 0.22)	0.54		-0.098 (-0.05 ; 0.21)	0.53

Stromal infiltration			0.09 (-0.40 ; 0.22)	0.55		-0.08 (-0.39 ; 0.23)	0.59
Immune cells infiltration			0.26 (-0.16 ; 0.43)	0.095		0.14 (-0.17 ; 0.45)	0.38
Number of somatic mutations			-0.42 (-0.73 ; -0.10)	0.005		-0.41 (-0.72 ; -0.1)	0.005

*Estimated as the proportion of altered genome

Table 3) Correlations between TTLR and gene expression change (T/N) in CLX subset

Gene	r*	P-value	FDR**	Description	Function
TCTN3	0.62	7.1E-07	0.0099	Tectonic family member 3	May be involved in apoptosis regulation. Necessary for signal transduction through the sonic hedgehog (Shh) signaling pathway
RNF7	0.61	1.3E-06	0.0099	RING-box protein 2	Mediates the ubiquitination and subsequent proteasomal degradation of target proteins involved in cell cycle progression, signal transduction and transcription
ATP6V0E1	0.59	3.9E-06	0.020	ATPase, H+ transporting. lysosomal 9kDa, V0 subunit e1	Responsible for acidifying a variety of intracellular compartments in eukaryotic cells
SNX3	0.57	1.8E-05	0.046	Sorting nexin 3	Plays a role in protein transport between cellular compartments
DDX54	-0.57	1.7E-05	0.046	DEAD (Asp-Glu-Ala-Asp) box polypeptide 54	Represses the transcriptional activity of nuclear receptors
UBAP1	0.57	1.2E-05	0.046	Ubiquitin associated protein 1	Regulator of vesicular trafficking process
PHF11	0.55	3.2E-05	0.062	PHD finger protein 11	Positive regulator of Th1-type cytokine gene expression
LOC284889	-0.56	2.9E-05	0.062		
ELMO2	0.54	4.3E-05	0.074	Engulfment and cell motility 2	Involved in cytoskeletal rearrangements required for phagocytosis of apoptotic cells and cell motility
CHD8	-0.54	5.2E-05	0.080	Chromodomain helicase DNA binding protein 8	DNA helicase that acts as a chromatin remodeling factor and regulates transcription. Is a negative regulator of Wnt signaling pathway by regulating beta-catenin (CTNNB1) activity
PHAX	0.54	6.6E-05	0.089	Phosphorylated adaptor for RNA export	snRNA export from the nucleus
OR2T2	-0.53	7.0E-05	0.089	Olfactory receptor, family 2. subfamily T. member 2	G-protein coupled receptor activity

*Partial Spearman correlation coefficient, adjusted for CN

** Only significant correlations with FDR< 0.1 are shown

Table 4) Significant enriched gene sets with high telomere length correlated genes.

Database	Gene set*	Gene set Size	Enrichment Score	Nominal P-value	FDR Q-value	FWER P-value	Correlation +/-
REACTOME	INTERFERON GAMMA SIGNALING	52	0.58	<0.0001	<0.0001	<0.0001	positive
	INTERFERON SIGNALING	133	0.49	<0.0001	<0.0001	<0.0001	positive
	TCR SIGNALING	48	0.57	<0.0001	0.0015	0.004	positive
	DOWNSTREAM TCR SIGNALING	32	0.62	<0.0001	0.0014	0.005	positive
	INTERFERON ALPHA BETA SIGNALING	49	0.54	<0.0001	0.0015	0.007	positive
	CYTOKINE SIGNALING IN IMMUNE SYSTEM	233	0.44	<0.0001	0.0019	0.01	positive
	ADAPTIVE IMMUNE SYSTEM	477	0.41	<0.0001	0.0019	0.012	positive
	REGULATION OF BETA CELL DEVELOPMENT	21	-0.64	<0.0001	0.0099	0.012	negative
	MHC CLASS II ANTIGEN PRESENTATION	81	0.49	<0.0001	0.0022	0.016	positive
	RNA POL I PROMOTER OPENING	45	0.55	<0.0001	0.0023	0.019	positive
	ANTIGEN PRESENTATION FOLDING ASSEMBLY AND PEPTIDE LOADING OF CLASS I MHC	20	0.66	<0.0001	0.0035	0.032	positive
BIOCARTA	EDG1 PATHWAY	25	0.62	<0.0001	0.024	0.016	positive
	IL10 PATHWAY	16	0.69	<0.0001	0.013	0.018	positive
	SPPA PATHWAY	20	0.65	<0.0001	0.0092	0.019	positive
	EIF4 PATHWAY	24	0.62	<0.0001	0.0096	0.026	positive
KEGG	LEISHMANIA INFECTION	64	0.61	<0.0001	<0.0001	<0.0001	positive
	ANTIGEN PROCESSING AND PRESENTATION	56	0.60	<0.0001	<0.0001	<0.0001	positive
	SYSTEMIC LUPUS ERYTHEMATOSUS	100	0.51	<0.0001	0.0011	0.003	positive
	ASTHMA	18	0.74	<0.0001	0.0088	0.003	positive
	AUTOIMMUNE THYROID DISEASE	29	0.64	<0.0001	0.0063	0.003	positive
	VIRAL MYOCARDITIS	61	0.55	<0.0001	0.0052	0.003	positive
	ALLOGRAFT REJECTION	29	0.63	<0.0001	0.045	0.003	positive
	GRAFT VERSUS HOST DISEASE	32	0.60	0.0015	0.0052	0.004	positive
	LYSOSOME	113	0.46	<0.0001	0.0057	0.005	positive
	LEUKOCYTE TRANSENDOTHELIAL MIGRATION	101	0.47	<0.0001	0.0052	0.005	positive
	TYPE I DIABETES MELLITUS	34	0.56	<0.0001	0.0015	0.016	positive
	INTESTINAL IMMUNE NETWORK FOR IGA PRODUCTION	38	0.55	<0.0001	0.0014	0.016	positive
	HEMATOPOIETIC CELL LINEAGE	70	0.48	<0.0001	0.0013	0.017	positive
	RETINOL METABOLISM	48	-0.45	<0.0001	0.016	0.017	negative
	MATURITY ONSET DIABETES OF THE YOUNG	15	-0.63	0.0026	0.014	0.032	negative
	CELL ADHESION MOLECULES CAMS	117	0.43	<0.0001	0.0031	0.041	positive
HALLMARK	INTERFERON GAMMA RESPONSE	193	0.58	<0.0001	<0.0001	<0.0001	positive
	INTERFERON ALPHA RESPONSE	93	0.60	<0.0001	<0.0001	<0.0001	positive

	INFLAMMATORY RESPONSE	186	0.45	<0.0001	<0.0001	<0.0001	positive
	ALLOGRAFT REJECTION	181	0.44	<0.0001	<0.0001	<0.0001	positive
	COMPLEMENT	182	0.43	<0.0001	<0.0001	<0.0001	positive
	PROTEIN SECRETION	95	0.46	<0.0001	0.050	0.003	positive
	IL6 JAK STAT3 SIGNALING	81	0.47	<0.0001	0.057	0.004	positive
	MYC TARGETS V2	57	-0.39	<0.0001	0.0082	0.009	negative

*Gene Set Enrichment Analysis (GSEA) summary results with the most significant gene sets classified by their origin: Pathway databases (REACTOME, Biocarta, KEGG) and Hallmarks of cancer. A gene set is a group of genes involved in a certain biological process or molecular function that has been annotated in those databases.

FDR: False Discovery Rate; FWER: Family-Wise Error Rate (multiple comparison adjusted P-value)

Figure legends

Figure 1)

Differences between Normal and Tumor tissues in CLX and TCGA subset

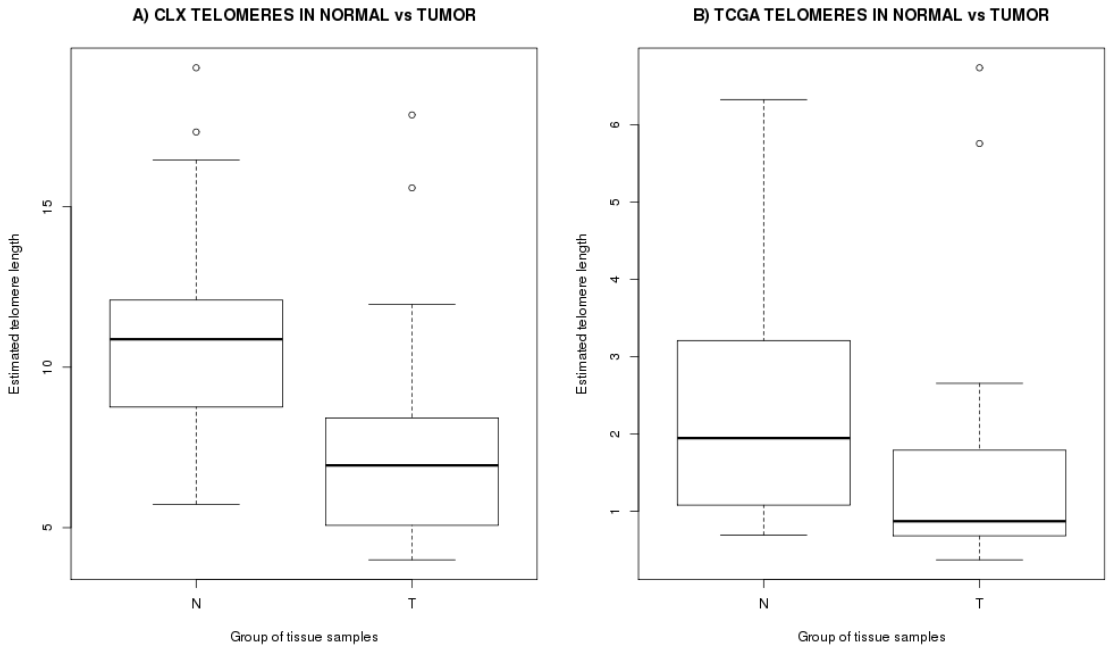
Boxplot representing the normalized number of telomere reads in Normal (green) and Tumor (red) paired samples. A) Samples from CLX subset B) Samples from TCGA subset.

Figure 2)

Enrichment statistically significant gene sets related with immune system

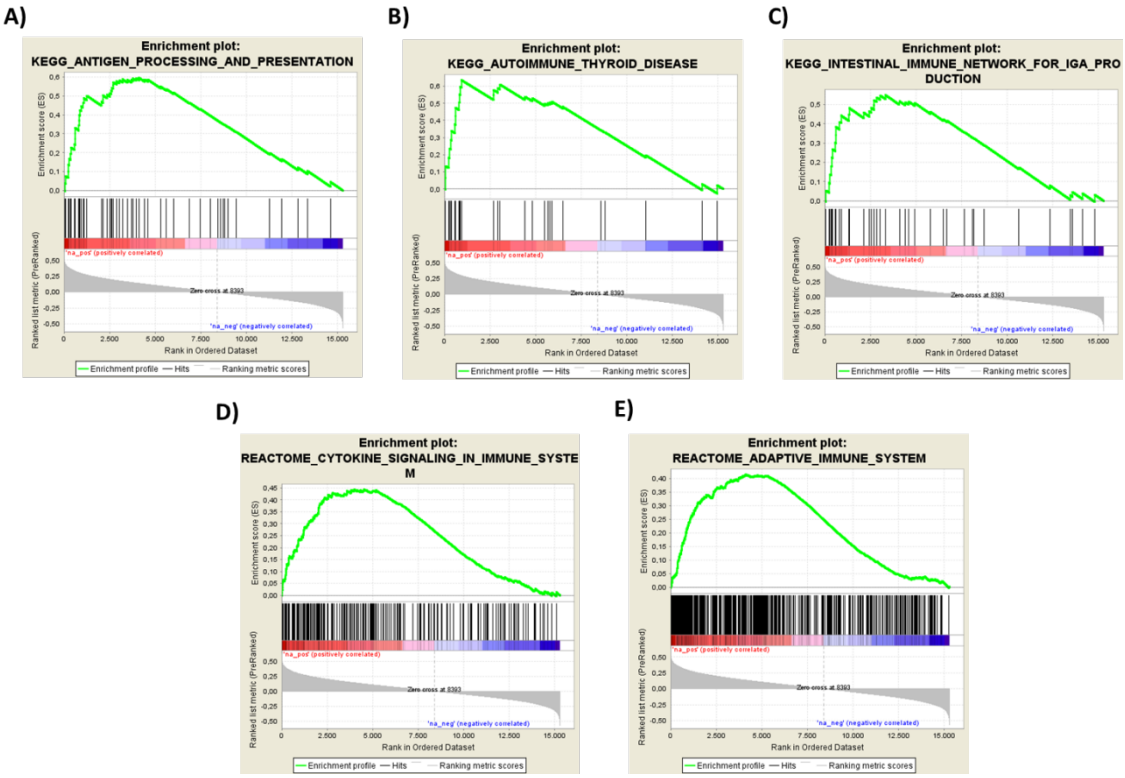
GSEA enrichment plots showing five of the most statistically significant gene sets, A) Kegg antigen processing and presentation, B) Autoimmune thyroid disease, C) Intestinal immune network for IGA production, D) Cytokine signaling in immune system and E) Adaptive immune system. In all plots, each black vertical line represents a gene in the specific gene set, the left-to-right position of each line indicates the relative position of the gene within the rank of the overall genes. The expression of genes near the middle are not correlated with the TTRL. The green line indicates the enrichment score. The section below the genes shows the rank in ordered list.

Figure 1) Differences between Normal and Tumor tissues in CLX and TCGA subset



Boxplot representing the normalized number of telomere reads in Normal (green) and Tumor (red) paired samples. A) Samples from CLX subset B) Samples from TCGA subset.

Figure 2) Enrichment statistically significant gene sets related with immune system



GSEA enrichment plots showing five of the most statistically significant gene sets, A) Kegg antigen processing and presentation, B) Autoimmune thyroid disease, C) Intestinal immune network for IGA production, D) Cytokine signaling in immune system and E) Adaptive immune system. In all plots, each black vertical line represents a gene in the specific gene set, the left-to-right position of each line indicates the relative position of the gene within the rank of the overall genes. The expression of genes near the middle are not correlated with the TTRL. The green line indicates the enrichment score. The section below the genes shows the rank in ordered list.

SUPPLEMENTAL MATERIAL

Suppl. Table 1) Number of telomeric and total reads in CLX and TCGA subsets.

	ID sample	N. telomeric reads in Tumor exome sequencing	Total N. reads in Tumor exome sequencing	N. telomeric reads in Normal mucosae exome sequencing	Total N. reads in Normal mucosae exome sequencing
CLX subset	1	432	64865000	796	43974000
	2	544	60644000	790	37618000
	3	475	61879000	1130	55826000
	4	1142	78868000	812	48313000
	5	821	64138000	617	40431000
	6	872	73334000	807	66566000
	7	832	59345000	609	44503000
	8	546	62008000	924	49030000
	9	502	57125000	547	47193000
	10	806	61459000	944	52515000
	11	1316	72766000	446	42756000
	12	595	54676000	584	41076000
	13	526	68842000	618	41073000
	14	897	71813000	728	50882000
	15	674	63031000	1515	44031000
	16	610	61752000	850	53146000
	17	584	52368000	552	34862000
	18	858	64672000	1188	53173000
	19	605	66758000	829	44618000
	20	916	62945000	808	36207000
	21	452	55740000	1841	71279000
	22	364	57771000	650	54280000
	23	986	76952000	1010	48815000
	24	651	81601000	754	48453000
	25	487	48831000	806	38520000
	26	381	54252000	860	41434000
	27	349	44804000	1035	52542000
	28	586	52342000	738	53154000
	29	358	53056000	591	33856000
	30	780	54991000	1827	109111000
	31	1179	75846000	1289	47626000
	32	1205	61263000	1308	44792000
	33	528	66201000	764	44651000
	34	1085	59911000	1739	94974000
	35	899	67916000	695	37824000
	36	1490	53509000	706	25279000

	37	1122	55471000	788	38387000
	38	581	61296000	627	34756000
	39	804	65894000	514	36532000
	40	549	74281000	621	50392000
	41	763	46147000	840	41626000
	42	1820	56891000	797	39894000
TCGA subset	1	2216	1109757040	2209	943135464
	2	1945	1121695152	2512	848429704
	3	1223	1105067448	2659	944964432
	4	849	973951640	5368	848429704
	5	2462	427531904	1365	359608600
	6	768	468376016	652	578663776
	7	1315	495137080	1133	373331672
	8	253	680319304	468	676085792
	9	1454	786691416	3080	910365632
	10	734	495137080	1837	501449272
	11	2384	353787936	2129	609203536
	12	342	747771064	388	353103072
	13	612	734277248	1145	646813376
	14	608	852112696	752	386150832
	15	558	811937072	234	293332536
	16	456	616992336	308	321714976
	17	428	630676664	362	341277224
	18	326	853395112	704	872850280
	19	298	737625176	1042	945557576