Quantifying immune cell telomere content at single-cell resolution in context of PD-1 checkpoint immunotherapy

Niklas L. Engel[1,2], Lea Herzel[1], Julie Surmely[1], Hanna Frieß[1], Malte Simon[1], Benedikt Brors[1,3,4], Charles Imbusch[1,5,6], Lars Feuerbach[1,*]

[1]Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[2]Department of Biomedical Informatics, Harvard Medical School, Boston, 02115, MA, USA
[3]National Center for Tumor Diseases (NCT), Heidelberg, Germany
[4] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[5] Institute of Immunology, University Medical Center Mainz , Mainz, Germany
[6] Research Center for Immunotherapy, University Medical Center Mainz , Mainz, Germany
*Corresponding author l.feuerbach@dkfz-heidelberg.de

Abstract

Introduction: Biological processes such as aging, carcinogenesis, and immune response rely on the ability to maintain or rapidly expand cell populations. The fitness of the involved cells is constrained by their replicative potential, which is reflected in the cellular telomere content.

Method: We apply TelomereHunter to scATAC-seq data to determine telomere content on single-cell level, in a hematopoietic dataset consisting of 35,139 cells from samples of basal cell carcinoma patients receiving programmed cell death protein 1 (PD1) blockade treatment. Integrating information from open-chromatin-based signatures to assess cell identity, we characterize the heterogeneity of telomere length for individual cell populations pre- and post-immunotherapy.

Results: The extracted telomeric reads reflect the expected telomereome-to-genome fraction. Telomere content distributions differ significantly between cell populations, and the median telomere content in intermediate and terminal exhausted CD8+ T-cells pre-treatment is significantly correlated to response to PD-1 checkpoint blockade. Likewise, telomere content correlates with post-treatment cell proliferation in terminally exhausted and T follicular helper cells from responding patients.

Conclusion: Telomere content measurement from scATAC-seq data has a sufficiently high signal-to-noise ratio to detect significant differences between cell types. Furthermore, the telomere content of CD8+ exhausted T-cells pre-treatment is a putative biomarker for successful PD-1-based immunotherapy.

Introduction

The ends of human chromosomes are protected by telomeres from degradation and fusion with other DNA molecules[1]. Constituted primarily of the hexameric repeat TTAGGG[2,3], the telomeres form a loop structure that is stabilized by the proteins of the shelterin complex[4]. Telomeres have a heterogenous length, which is anti-correlated to age, and subjected to inherited predisposition[5,6]. During cell division telomeres shorten (end-replication problem)[7], which eventually triggers cellular senescence upon reaching a critical telomere length (M1 checkpoint / Hayflick limit)[8]. During embryogenesis, the enzyme telomerase is able to elongate the telomeres by using the Telomerase RNA component (TERC) RNA as a template to add new nucleotides to the telomeric reads[9]. The telomerase components are expressed during early embryonic development, but the gene of its core component *TERT* is silenced between the 12th and 18th week of gestation[10,11]. While most somatic cells display no telomerase activity, an exception from this rule are cells of the germline, stem cells, and components of the adaptive immune system[10,12,13]. The telomere maintenance in immune cells is incomplete and processes such as aging are attributed to telomere attrition. More specifically, median telomere length in the lymphocytes of the adaptive immune system has been observed to decrease from 10 kbp at birth to 4.5 kbp for centenarians[14].

The adaptive immune system consists of a variety of cell types. Among them are T-cells that can be subdivided into CD4+ and CD8+ T-cells. The CD4+ compartment is made of T-helper (Th) and regulatory T-cells (Tregs). Among T-helper (Th) cells are T follicular helper cells (Tfh). The CD8+ compartment can be categorized into effector and memory T-cells. CD4+ T-cells primarily function as helpers, while CD8+ T-cells have a cytotoxic role.
Naïve T-cells await activation by their cognate antigen that matches a cell-specific pattern encoded in their T-cell receptor (TCR). T-cell activation leads to differentiation and eventually rapid clonal expansion of the corresponding immune cells. To maintain replication, T-cells that express the coreceptor CD28 upregulate telomerase via NF-kappaB guided signaling pathway by receptor costimulation upon activation[15]. Permanent stimulation via TCR and other factors can lead to dysfunctional T-cells which are limited in their capacity to clear pathogens and tumor cells. These dysfunctional T-cells are also termed exhausted T-cells (TEx)[16]. TEx can be divided into at least three different stages of exhaustion and can be categorized as Early TEx, Intermediate TEx, and terminal TEx[17–19].
Failure of effective immune response in elderly patients may be linked to a limited replicative potential of lymphocytes due to short telomeres. This process is also observed in accelerated form during chronic viral infection[20]. To effectively study this process at single-cell resolution novel approaches are required.

We test if computational analysis of single-cell Assay for Transposase-Accessible Chromatin with sequencing (scATAC-seq) data yields sufficient information on the telomere content of the analyzed cells. Although ATAC-seq data is primarily used to identify regions of open chromatin, it has been observed that telomeric reads enrich in these datasets in cells with intact shelterin, as well as, in shelterin-deficient model systems[21]. Our study therefore aims to clarify if unlike nucleosome-occupied DNA, shelterin-bound DNA allows unbiased Tn5 transposase activity, or if telomere sequence is as underrepresented as heterochromatin in the resulting datasets.

To this end, we use the TelomereHunter software[22], which has been previously applied to characterize the telomere compartment, or telomereome, of sequencing datasets from deep and shallow whole-genome sequencing (WGS), whole-exome sequencing (WES), and Chromatin immunoprecipitation DNA-sequencing (ChIP-seq)[22–25]. In this context, telomere content generally designates the fraction of telomeric reads over all genomic reads in a NGS dataset, which is then corrected for confounders such as sequencing depth, GC-bias, age and hereditary differences in telomere length to enable a comparison between different samples.

We test our approach on a scATAC-seq dataset derived from tumor biopsies of patients with basal cell carcinoma (BCC) that were treated with immunotherapy in form of anti-programmed cell death protein 1 (PD-1) checkpoint blockade[17], and characterize telomere content differences between individual cell types. In this dataset, telomere content of intermediate and terminal exhausted T-cells prior to therapy onset is predictive for the success of PD-1 checkpoint blockade. This highlights the potential of telomere content of T-cell populations as biomarker for therapy selection in cancer patients, as well as, the relevance of the single-cell telomere quantification workflow we showcased were.

Results

The initial scATAC-seq dataset utilized in our experiments consisted of biopsies obtained from patients with BCC before and after receiving PD-1 immunotherapy[17]. Samples from seven patients, ranging in age from 50 to 75 years, represent time points before and at various time intervals after therapy onset. Patient responses to the treatment were available from the original publication (Supplementary Table 1). Cells were grouped in multiple fractions. The *Total* cell fraction was unsorted, while the other fractions were identified by FACS using specific cell surface markers: the *T-cell* fraction (CD45+ and CD3+), the non-T *Immune* cell fraction (CD45+ and CD3-), and the stromal and *Tumor* cell fraction (CD45-). Within the *T-cell* fraction cell types were assigned based on clustering analysis of the ATAC-seq profiles. To estimate the telomere content of each cell computationally, we employed TelomereHunter[22], and derived a telomere content estimate for each individual cell, represented by their cellular barcode.

Application of a coverage-based cutoff for TelomereHunter single-cell sequencing optimization

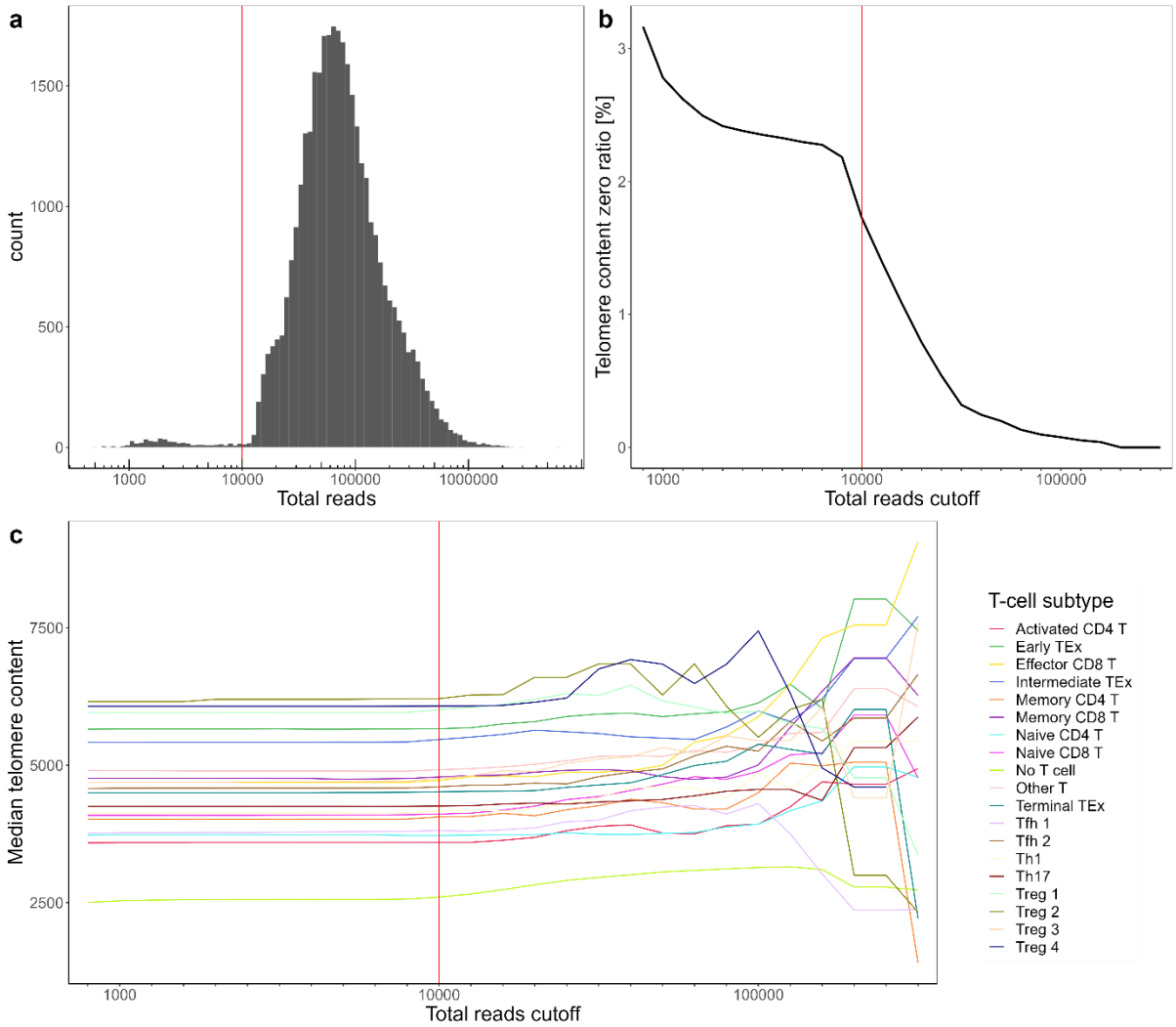

Figure 1: Analysis of progressive removal of low coverage cells from analysis.
a Histogram of the log10-total reads per cell in the Satpathy dataset. **b** Percentage of uninformative cells having a telomere content of zero (y-axis) for increasing total reads cutoffs (x-axis). **c** Trajectories of the median telomere content for all T-cell subtypes in the Satpathy dataset for all barcodes with more than the total reads cutoff (x-axis). A cutoff of 10,000 includes only cells with at least 10,000 reads. The x-axis has log10 increments but is continuously scaled. The cutoff of 10,000 or more reads is highlighted in red in all three plots.

Following this step, we obtained the GC-corrected telomere content estimates for a total of 35,139 cells. Cells with low read count may bias the telomere content estimation. Therefore, we generated four diagnostic diagrams to define a minimal read count threshold for inclusion (t_{in}). The histogram of read counts per cell, showed a bimodal distribution with a small mode representing low coverage cells, while the majority of cells in the larger mode displays read counts above 10,000 reads per cell (Figure 1, a). The second diagram shows the number of cells without any telomeric reads (T_0-cells) (Figure 1, b). The third diagram resolves the cells by cell type and compares different t_{in} values against the resulting median telomere content per immune cell type. A long stable plateau across all groups past a t_{in} of 10,000 reads per cell is observable. Beyond this point, the median telomere content begins to fluctuate in an increasing number of cell types (Figure 1, c). The last diagram indicates that this is caused by decreasing read numbers (Supplementary Figure 1, a). In case of low coverage T_0

observations may not reflect cells with low telomere content, but rather statistical artifacts due to the discrete nature of read counts. A t_{in} of 10,000 reads provides a good tradeoff between the reduction of the T_0-cells and the inclusion of a high cell count.
As a consequence, 4.1% of the cells are removed from the dataset. Beyond the 10,000 mark the stability of the groupwise telomere content estimations become unstable, as reflected in the median and standard deviation values (Supplementary Figure 1, b and c). This observation is consistent across responding and non-responding patients, pre- and post-treatment biopsies, different patients, cell types, and T-cell subtypes (Figure 1, c; Supplementary Figure 2, a-d; Supplementary Figure 3 a-e; Supplementary Table 2). We therefore selected a t_{in} of 10,000 reads per cell and removed all barcodes not meeting this threshold from the remaining analysis resulting in 33,202 cells. The telomere content, including the number of cells, median values, and standard deviations, varies among the seven patients (Supplementary Table 3; Supplementary Figure 4). Variations in telomere content can also be observed in the biological granularity (Supplementary Table 4-7).

Robustness of the ATAC-seq protocol for telomere content estimation

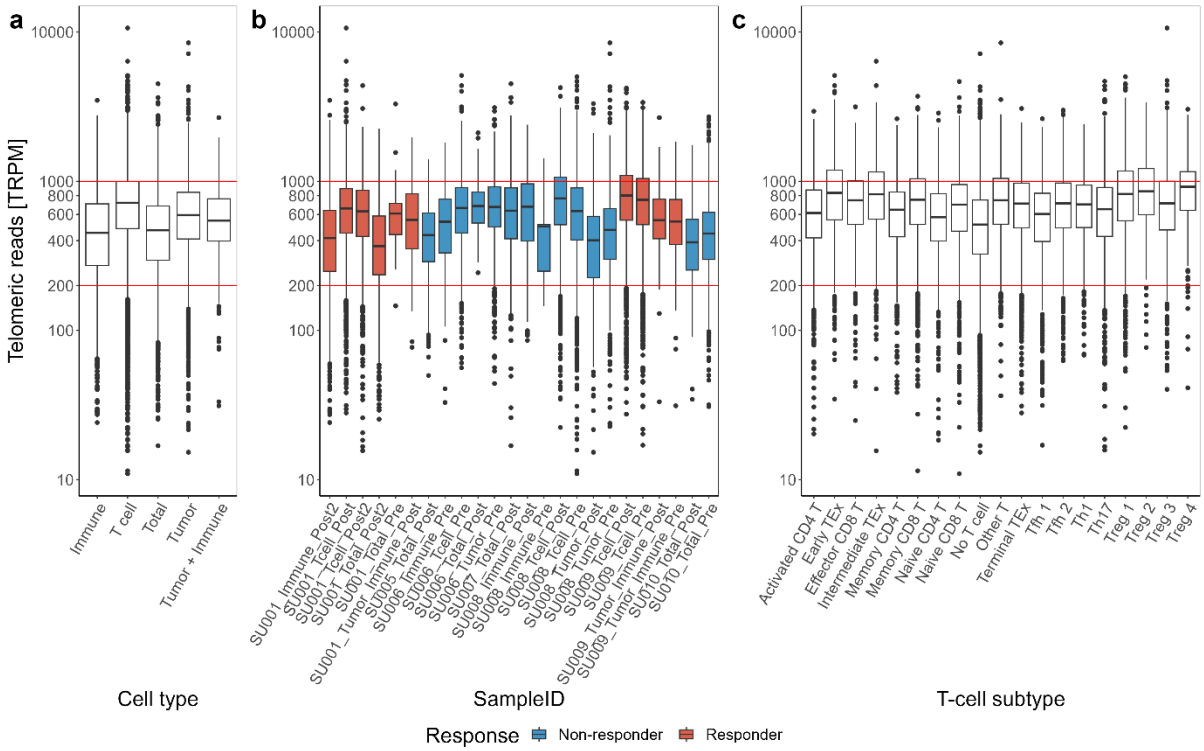

Figure 2: Comparison of TRPM estimates from TelomereHunter to *in vivo* studies.
Raw Telomere content of the cells used in the Satpathy dataset. *Immune* + *Tumor* designate samples in which *Tumor* and *Immune* cells were pooled. **a** TRPM of cells in the cell types identified by Satpathy and colleagues (*Immune*, *T-cell,* and *Tumor* cells, as well as a shared *Tumor* + *Immune* and a *Total* cell cluster). **b** TRPM of cells in the 24 scATAC-seq pseudo-bulks generated by Satpathy and colleagues. **c** TRPM of cells in the 19 T-cell subtypes identified by Satpathy and colleagues. In all three TRPM boxplots, a TRPM of 200 and 1000 is highlighted in red. In all boxplots, a total reads cutoff of 10,000 reads is applied.

Next, we tested if the scATAC-seq protocol derived observations reflect the expected range of telomere content. In particular, we tested if the telomerome is underrepresented by this assay. The bulk dataset exhibits a raw telomere content of 732 Telomeric reads per million

sequenced reads (TRPM). To compute a lower bound for the expected raw telomere content in healthy cells of young individuals, we assume an average telomere length of 10 kilobase pairs (kbp) per chromosome end. Consequently, 960 kbp out of the 6,32 gigabase pairs (Gbp) of genomic DNA in a diploid human genome (average of female genome 6.37 Gbp and male genome 6.27 Gbp) originate from telomeric regions, accounting for approximately 0.0152 % of the genome or 152 TRPM. Recent *in vivo* studies from cell lines suggest a slightly higher interval in human cells between 200 and 1000 TRPM, respectively 0.02-0.1% of the genomic DNA[26]. Then, we grouped the single-cell telomere content estimates derived from TelomereHunter data according to the available technical and biological annotations, such as FACS sorting fraction (Figure 2, a), their pseudo-bulk labels (Figure 2, b), and according to ATAC-seq-profiling-derived celltype (Figure 2, c). Across all 33,202 cells, 72.8% fall in the interval indicated by the cell line experiments (red lines) and only 3.5% have a TRPM smaller than 152. For the unfiltered data these numbers are comparable (Supplementary Figure 5, a-c). This indicates that the scATAC-seq protocol adequately represents the telomeric fraction of the genome with a tendency towards enrichment.

Telomere content associated T-cell count change

As already reported by Satpathy *et al.*, the cell counts of several immune cells are quite distinct between response status and time points. Based on the scATAC-seq profiles of the *T-cell* fraction, we conducted a UMAP dimensionality reduction. The resulting UMAP is then decomposed into a 2x2 grid representing the time points and response types, thus conserving the location of cells with similar ATAC-cell profiles even if underrepresented in their layer (Figure 3, a and b; Supplementary Figure 6, a-c). We then discretized the UMAP into squares which each represent the mean GC-corrected telomere content of all cells overlapping with it (Figure 3, a). To quantify the visible effect of telomere content associated T-cell count change in the telomere and cell-annotation UMAPs, we computed the relative differences in cell counts and median telomere content between pre- and post-treatment fractions to characterize these changes for the cell types of interest (Supplementary Figure 7, a-b and 8). In responding patient SU009, expanding cell fractions can be observed both while increasing and decreasing their median telomere content. Interestingly, in non-responding patient SU006, only T-cells that strongly decrease their cell counts post-therapy can be observed. (Figure 3, c). These observations pose the question, to what degree the changes in telomere content can be explained through telomere shortening by active cell division, activity levels of telomerase-based telomere maintenance or differentiation/exhaustion processes that change cell identity during PD-1 treatment.

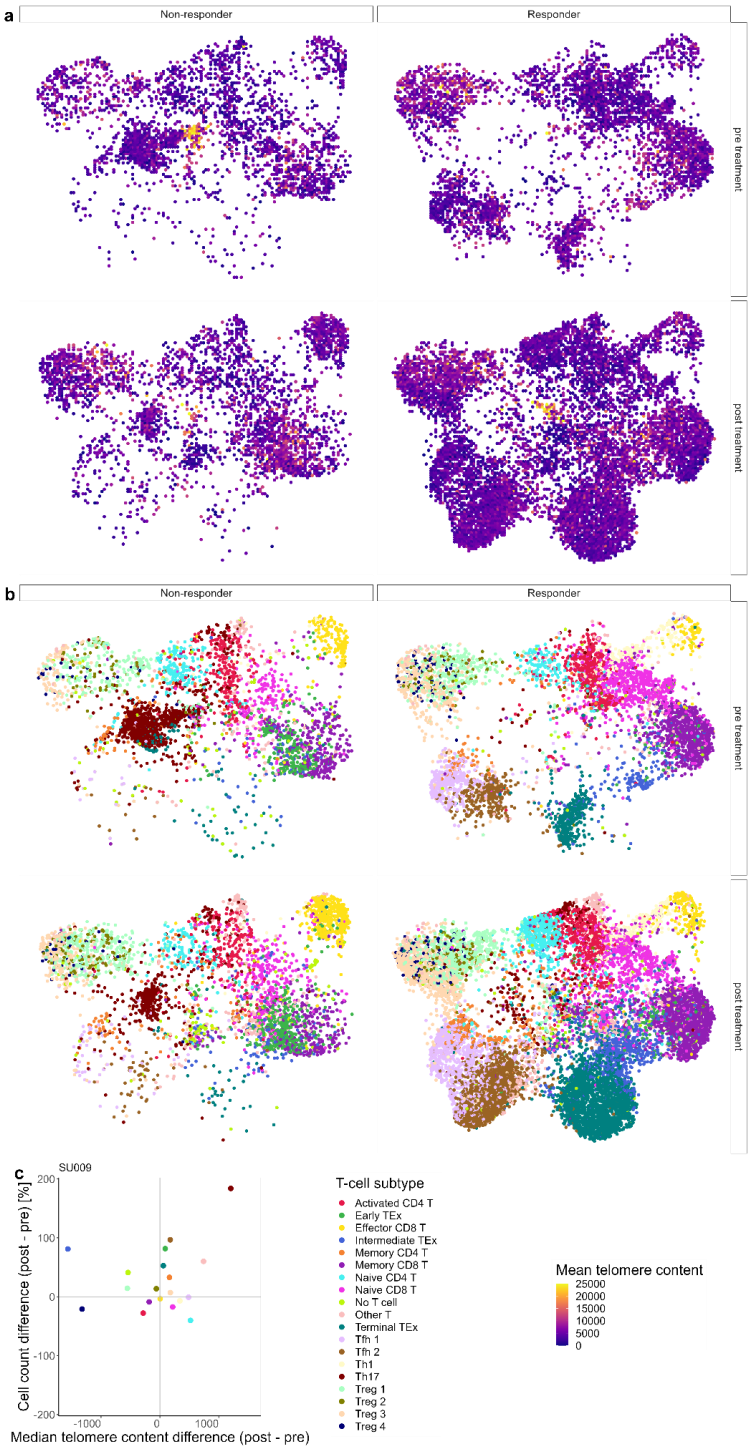

Figure 3: Projection of telomere content on a single-cell level.
a UMAP projection colored by telomere content and split by pre- and post and response to therapy status. To better resolve the single-cell telomere content, UMAP coordinates were averaged rounded to one decimal place. Thus, the mean GC-corrected telomere content of all cells falling into the cluster of the same UMAP coordinates is displayed. **b** UMAP projection of the T-cell subtypes pre- and post-PD-1 checkpoint therapy treatment and separated by response to therapy. **c** Scatter plot of the median telomere content difference between pre- and post-treatment fractions against the relative cell counts difference of all T-cell subtypes for SU009. The four significant T-cell subtypes *Early, Intermediate*, and *Terminal TEx* and *Tfh2* cells are highlighted by triangle dots. To all plots, a total reads cutoff of 10,000 reads is applied.

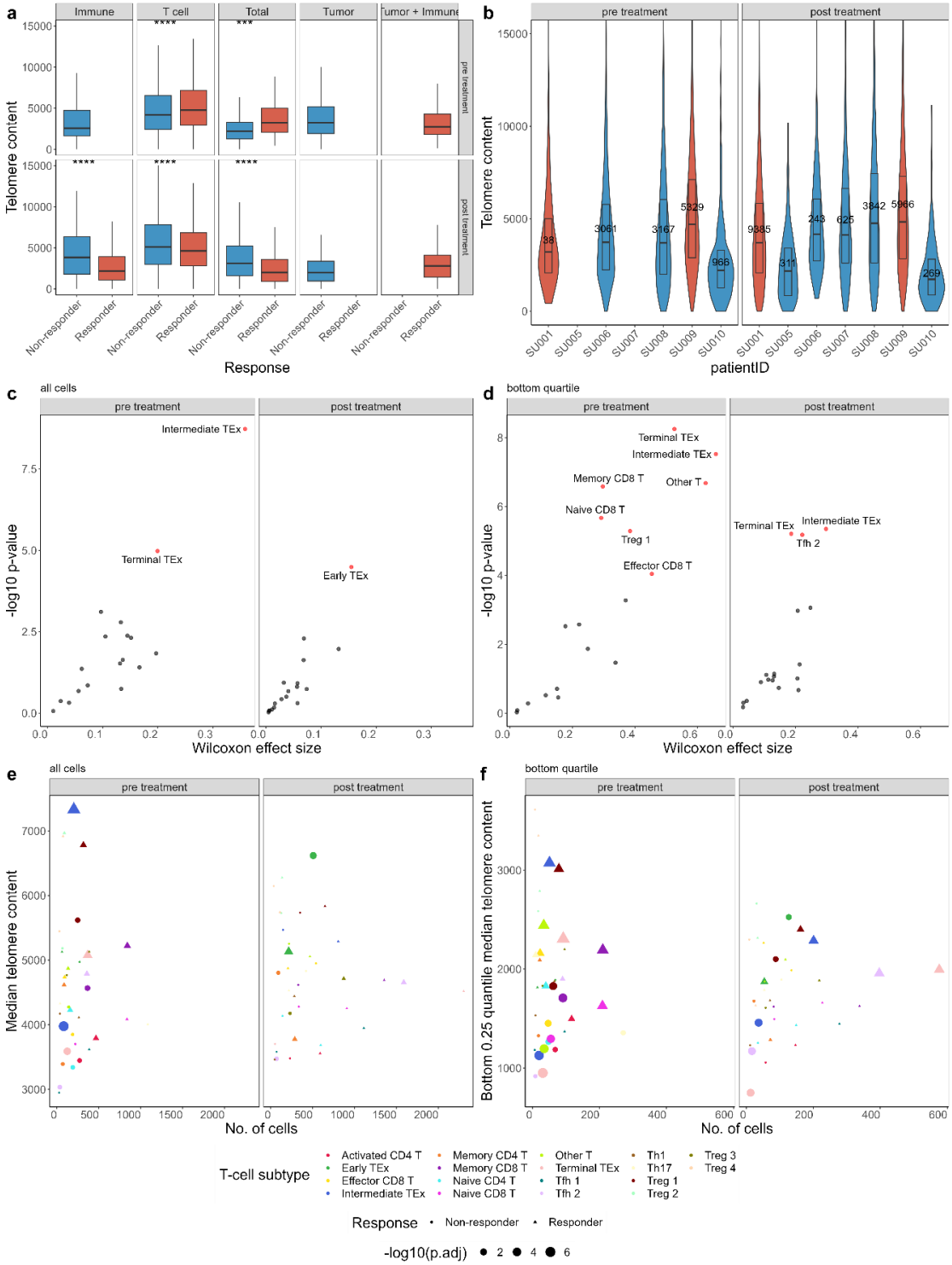

Figure 4: Association of GC-corrected telomere content to therapy response
a Wilcoxon test of the differences in telomere content at single-cell resolution between the PD-1 checkpoint therapy *Responder* and *Non-responder* patients (outliers not included in this plot). Data is grouped by the timepoint-of-biopsy (rows), FACS-sorting fractions (columns) and response status (color). **b** Telomere content at single-cell resolution grouped by patients, timepoint-of-biopsy and response status. The y-axis is limited to 15,000 for better visibility (full plot in Supplementary Figure 9). **c-d** Differences of **c** telomere content and **d** the bottom 25% quartile telomere content between *Responder* and *Non-responders* grouped by cell types. Volcano plot shows log10-p-values (y-axis) and

the Wilcoxon effect size for T-cell subtypes (x-axis). Values greater than the q-value (Benjamini-Hochberg) cutoff of 0.01 are highlighted in red. The Volcano plots are partitioned into the two biopsy timepoints pre- and post-treatment. Non T-cells are excluded from the test. **e-f** Scatter plot of Number of cells (x-axis) vs. median telomere content per T-cell type (color) and response type (shape). Data is shown for all cells (**e**) and only of the bottom quartile (**f**). To all plots and the statistical tests, a total reads cutoff of 10,000 reads is applied.

Telomere content analysis reveals a distinct PD-1 checkpoint therapy response signature in exhausted T-cells

To follow-up on the differences observed in the UMAP, we performed statistical testing of the telomere content between responding and non-responding patients. Responding patients exhibit a significantly larger median telomere content compared to non-responding patients (Wilcoxon p-value = 3.02e-36, Wilcoxon effect size = 0.07). Furthermore, the pooled pre-treatment samples show a significantly larger median telomere content compared to the pooled post-treatment samples (Wilcoxon p-value = 0.0041, Wilcoxon effect size = 0.02). Among different fractions, *T-cell*s have a significantly larger median telomere content than other cell types after multiple testing correction. The median telomere content significantly differs between the cells of 7 out of 10 fraction pairings (Supplementary Figure 10, a). Patient SU009 has a significantly larger median telomere content than all other patients. Similarly, median telomere content levels differ significantly between 19 out of 21 patient pairs (Supplementary Figure 10, b). Among T-cell subtypes, *Treg 2* cells have the largest median telomere content. This difference is significant to all other cell types, except in comparison to *Treg 1*, *Treg 4*, and *Early TEx* cells. For these cases, the effect sizes vary from very small to medium (Supplementary Figure 10, c).

Remarkably, telomere content can differentiate responding from non-responding patients based on the pre-treatment biopsies alone. *Responders* have a significantly larger median telomere content (Wilcoxon p-value = 4.66e-101, Wilcoxon effect size = 0.19) (Supplementary Table 8). At FACS fraction resolution, namely in *T-cell* and *Total* cells, the median telomere content is significantly different between the two groups, as well (Figure 4, a). This signal in the *T-cells* fraction is predominantly driven by the 5329 cells of one responding patient (SU009), while the second responder (SU001) only provides observations for 38 cells (Figure 4, b). Therefore, we resolved this comparison on the level of T-cell subtypes. After multiple testing correction, a statistically significant difference in median telomere content is observed for *Terminal* (q=3.68e-4, Wilcoxon effect size=0.2) and *Intermediate TEx* (q=6.73e-8, Wilcoxon effect size=0.36) cells pre-treatment between *Responders* and *Non-responders* (Figure 4, c). Post-treatment, this is observed for *Early TEx* cells (q=0.00111, Wilcoxon effect size=0.155) (Supplementary Table 10). This is also reflected in the density distributions of cell-specific telomere length estimates (Supplementary Figure 11). In contrast to the higher telomere content in the pre-treatment condition, *Early TEx* cells post-treatment exhibit lower telomere content in *Responders*. For *Intermediate TEx* cells, observations are exclusively derived from the 201 cells of responder SU009 (Supplementary Figure 12, a; Supplementary Table 9). When pooling the pre- and post-treatment fractions together, the population *Tfh 2* also reaches significance (q=0.00128, effect size=0.09). Since looking at the shortest telomeres of a cell population has in the past also shown to be clinically significant, we then focused on the first quartile Q1, respectively, the 25% of cells with the lowest telomere content values, for each T-cell fraction. With this approach, 7 T-cell fractions differ significantly between *Responders* and *Non-Responders* pre-treatment, including *Intermediate* and *Terminal TEx* cells. These cell types are also significant post-treatment, together with *Tfh2* cells (Figure 4, d). Focusing on the directionality of changes, almost all the differences in pre-treatment are because of responding patients' cells having higher telomere contents (Figure 4, e). This is even more pronounced among the Q1 subset (Figure 4, f).

Discussion

In this project, we implemented a workflow for the *in silico* analysis of telomere content at single-cell resolution based on scATAC-seq data. First, we established a quality control procedure for the exclusion of low-coverage cells to prevent statistical bias. For this dataset, the exclusion of cell barcodes with less than 10,000 reads was necessary due to the discrete nature of count statistics.

We then confirmed that with an observed 732 TRPM over theoretically expected 152 TRPM for telomeres of 10 kbp length the ATAC-seq protocol leads to an at least 4.8 fold enrichment of telomeric DNA rather than a depletion. At the same time, the observed telomere content is within the bounds of observations made from cell lines. This implies that the enrichment is most likely caused by the expected underrepresentation of heterochromatic genome regions and not by an artifactual amplification of the telomerome.

Moving our attention from the technical characterization of the method to a clinically relevant question, we investigate the relationship between the telomere content and the response to PD-1 checkpoint immunotherapy. The TelomereHunter-derived telomere content measurements revealed significant variations, both within and between patients. The results indicated that patients who responded to PD-1 checkpoint immunotherapy had a significantly higher median telomere content in the T-cell compartment compared to non-responders. This finding suggests a potential link between observed telomere content and the efficacy of immunotherapy. More specifically, these differences were observed in pre-treatment biopsies, indicating their predictive potential, and were mainly driven by *Intermediate* and *Terminal TEx* cells. Further, the rapid expansion of these two cell types in the post-therapy samples of responding patients makes the availability of an increased replicative potential at treatment onset a plausible contribution to therapy success. This hypothesis is supported by the comparison of the quartile of T-cells with the lowest telomere content between Responders and Non-Responders, in which the difference between the response types is even more pronounced. In other words, in Non-Responders more cells and T-cell types are closer to the Hayflick limit, which triggers cellular senescence via the M1 checkpoint taking them out of the replication cycle. Indeed, in line with this model a recent study has established an association of an expression based cellular senescence signature to decreased response to anti-PD-1 therapy[27]. As this observation was primarily driven by a single responding patient, caution is required regarding the generalization of this result, but a follow-up on this hypothesis on a larger cohort appears promising.

Additionally, *Tfh 2* cells showed a significant increase in telomere content post-treatment in responding patients, which is in line with other results that highlight the relevance of T follicular helper cells for response to immunotherapy[28–30]. Whether this increase indicates the activation of telomere maintenance to enable rapid expansion or is caused by other biological processes may be relevant for a deeper understanding of therapy success.

It was recently proposed that telomere content derived from ATAC-seq experiments may not primarily reflect telomere length, but also altered chromatin compaction in the non-telomeric genome, for instance, changes during the cell cycle[31]. Delineation of the impact of these two components on the telomere content in the future may show if both signal sources contribute to the predictive power of this observable, or if one of them is dominating. For instance, a validation experiment on a larger cohort that ideally includes orthogonal methods for single-cell telomere length measurement, such as telomere Fluorescence in situ hybridization (FISH), would help to answer this question. As distinct cell cycle activity levels, as well as, longer telomeres both are relevant for cell proliferation, we can for now summarize these processes by the term replicative potential.

The observed impact of *Intermediate* and *Terminal TEx* cells in predicting response to PD-1 checkpoint immunotherapy aligns with their nature as exhausted T-cells. The aim of PD-1 immunotherapy is to reactivate exhausted TEx cells by interfering with the repressive signaling cascade.

PD-1 therapy success during chronic viral infection of mice depends on costimulation of the CD28+ pathway and has been linked to the rescue of TEx[32]. Additionally, the transition of resting T-cells into activated T-cells is flanked by the activation of telomerase in the CD28+ subset of these cell populations, which has been recently assessed on single-cell level by droplet digital PCR based Telomerase Repeated Amplification Protocol (ddTRAP) assays[15,33]. These observations indicate the relevance of telomere maintenance for TEx rescue in context of PD-1 therapy. In context of this prior work, it would be of interest to assess if the increased telomere content in responders is associated as well with telomerase activity and an enrichment of CD28+ cell surface marker. *Terminal TEx* cells, while being more dysfunctional, could benefit more strongly from improved telomere maintenance which leads to better overall immune response[34,35].

In summary, the increased telomere content in the *Intermediate* and *Terminal TEx* cells of *Responder* patients might signify a preserved replicative potential and thus, together with a higher cell count before treatment, the ability to leverage a more intense response to PD-1 immunotherapy. T-cell exhaustion is a complex process and influenced by multiple factors. The interplay between telomere content and these factors likely contributes to the observed impact on T-cell response, underscoring the necessity of further studies. If these relationships are confirmed, a comprehensive assay based on FACS sorting and telomere length measurement by qPCR or telomere restriction fragment analysis (TRF) would be sufficient to identify patient with higher response rates to PD-1 therapy.

We have introduced a workflow for *in silico* analysis of telomere content in scATAC-seq data together with a downstream analysis of the impact of telomere content on response to immunotherapy. This workflow will aid future analyses of the telomereome in single-cell and ATAC-seq datasets that also examine differences in telomere content between different cell types and conditions.

Methods

R

R is a programming language used for statistical computing[36]. All data visualizations, data wrangling, and statistical analyses within this project were performed using R Version 4.1.3. The packages BiocompR version 0.0.219, cowplot version 1.1.1, dplyr version 1.1.2, ggExtra version 0.10.1, ggplot2 version 3.4.3, ggpointdensity 0.1.0 version, ggpmisc 0.5.4-1 version, ggpubr version 0.6.0, ggrepel version 0.9.3, ggtext version 0.1.2, gplots version 3.1.3, gridExtra version 2.3, readr version 2.1.4, reshape2 version 1.4.4, RColorBrewer version 1.1-3, rstatix version 0.7.2, Seurat version 4.3.0, tidyr version 1.3.0, and viridis version 0.6.5 were used in this project.

Statistical Testing

An unpaired Wilcoxon rank-sum test was used to determine the significance of differences between two independent groups. This non-parametric test was chosen because it can handle data that are not normally distributed and because it may be used to compare groups with small sample numbers. P-values obtained from the statistical tests were adjusted using the Benjamini-Hochberg correction to reduce the likelihood of false positives due to multiple testing. A α-level of 0.01 was used as a p- or q-value cutoff.

TelomereHunter

TelomereHunter version 1.1.0. was used to calculate the telomere content of the data in this project[22]. All files were fed into TelomereHunter with the tumor option -ibt, no healthy control samples were used and additionally, the option - -plotNone was set. TelomereHunter was used together with gcc version 7.2.0 and R version 3.4.0. TelomereHunter combines Python and R for i*n silico* estimation of telomere content and composition from cancer genomes TelomereHunter searches for the four telomeric repeats (TTAGGG, TCAGGG, TGAGGG, TTGGGG) in sequencing reads using tumor BAM files as input. Telomeric reads are characterized as intratelomeric, subtelomeric, intrachromosomal, or junction-spanning according to their mapping position. The raw telomere content is computed by dividing the telomeric reads * 10^6 by all sequenced reads and has the unit Telomeric Reads Per Million sequence Reads (TRPM). The GC-corrected telomere content is then computed by dividing the number of intratelomeric reads * 10^6 by the number of reads with comparable GC content to telomeres (48%-52% GC-content), and enables the comparison of measurements from independent sequencing runs.

Satpathy Dataset

The dataset used in this project was created by Satpathy and colleagues[17]. This dataset is based on a single-cell ATAC-seq (scATAC-seq) study of seven patients with locally progressed or metastatic basal cell carcinoma. Patient selection criteria were ensuring that no immune checkpoint blockade or other immunosuppressive therapies had been delivered during the previous four weeks. The dataset includes tumor samples from seven different patients, both before and after PD-1 blocking therapy. This results in a collection of 24 composite samples, so-called pseudo bulks. This dataset is organized in four main cellular categories: immunological, malignant, stromal, and intratumoral T-cells. Fluorescence-activated cell sorting (FACS) is used to separate these T-cells into these groups. T-cells can be further subdivided into regulatory CD4+ T (Treg) cells, exhausted CD8+ T (TEx) cells, CD4+ T follicular helper (Tfh) cells, T helper (Th) cells, memory CD4+ and CD8+ T-cells, and naïve CD4+ and CD8+ T-cells. The data includes whether the cells are from individuals who reacted to the PD-1 blockade or not. The samples were obtained from the GEO database. Following TelomereHunter analysis, the data from these samples were merged for future

examination. The 10X Genomics Chromium platform was used for the scATAC-seq method. CellRanger, a software tool created by 10X Genomics, aided in further data processing. Before beginning the studies for this thesis, the CellRanger-processed data was treated using TelomereHunter, a tool used to get insights into the telomere compositions in the data.

Data Download and Processing

Bam files were downloaded from SRA (SRP192525) and converted to fastq files with "bam2fastq" (https://github.com/jts/bam2fastq). Then, "cellranger-atac count" from the ''Cell Ranger ATAC'' (v. 1.1.0,10X Genomics) software was used for read filtering, alignment (reference genome hg19, refdata-cellranger-atac-hg19-1.1.0), peak calling and count matrix generation for each sample. A custom script obtained from 10xGenomics was run for detection of gel bead and barcode multiplets. Barcodes were kept if they passed the following filters: at least 5,000 read-pairs passed read filters; less than 20% read-pairs with low mapping quality "< 0.2 fraction of unmapped + low mapping quality (mapq < 30) + chimeric read-pairs"; less than 90% read pair duplicates; less than 10% read pairs from mitochondrial DNA; no annotation as gel bead or barcode multiplets. Subsequently, sample aggregation was performed using 'cellranger-atac aggr' without normalization for library size "–normalize = none". Barcodes with more than 50,000 fragments were subsampled to 50,000 fragments to mitigate the influence of cell count depth on the downstream analysis.

Seurat Pipeline

Fragments were loaded into R (v. 3.6.0) using the package Seurat (v. 3.1.2)[37]. Transcription start site (TSS) scores were calculated as previously described[17] and barcodes with a TSS score below 8 were excluded. For normalization and dimensionality reduction, Signac (v. 0.1.3, https://github.com/timoast/signac)[37] was used. Briefly, the peak-barcode matrix was binarized and normalized with "RunTFIDF(method = 1)". Then, singular value decomposition was run (RunSVD) on the upper quartile of accessible peaks "FindTopFeatures(min.cutoff = 'q75')". Batch correction for donors was done with Harmony[38] using "HarmonyMatrix(sigma=1)". The first 20 components from the Harmony reduction were used to calculate a UMAP representation "RunUMAP(metric = 'euclidean')" and graph-based clusters were determined "FindNeighbors(dims = 1:20)", "FindClusters(resolution =0.5)"

To extract the T-cell subset from this dataset, the authors' annotation from Satpathy et al. was matched with the barcodes in the dataset. Then, the subset of clusters with high gene activity for T-cell markers (CD4, CD8A, CD3E) and high fraction of cells previously annotated as T-cell was selected. For this subset, dimensionality reduction and harmony batch correction was repeated as described above to obtain the final dataset of T-cells.

Code Availability

The code for reproducing the statistical analysis has been deposited on GitHub: https://github.com/niklas-engel/scATAC_TelomereHunter_2024/.

Acknowledgements

This Projekt was funded by the DFG in context of the Forschergruppe “FOR2674: Alters-assoziierte epigenetische Veränderungen als therapeutischer Ansatzpunkt in der Behandlung der akuten myeloischen Leukämie”. We thank Cindy Körner for her comments regarding the manuscript. We acknowledge all patients that consented to have their samples analyzed in the Satpathy et al. study.

Author information

Authors and Affiliations

Niklas L. Engel[1,2], Lea Herzel[1], Julie Surmely[1], Hanna Frieß[1], Malte Simon[1], Benedikt Brors[1,3,4], Charles Imbusch[1,5,6], Lars Feuerbach[1,*]

[1]Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[2]Department of Biomedical Informatics, Harvard Medical School, Boston, 02115, MA, USA
[3]National Center for Tumor Diseases (NCT), Heidelberg, Germany
[4] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[5] Institute of Immunology, University Medical Center Mainz , Mainz, Germany
[6] Research Center for Immunotherapy, University Medical Center Mainz , Mainz, Germany
*Corresponding author l.feuerbach@dkfz-heidelberg.de

Contributions

N.L.E. and L.F. designed the study and wrote the manuscript. C.I. and L.F. supervised the study. N.L.E, L.H, J.S, H.F, M.S conducted the data analysis and visualization. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Counter, C. M. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
2. Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6622–6626 (1988).
3. Meyne, J., Ratliff, R. L. & Moyzis, R. K. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7049–7053 (1989).
4. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
5. Aubert, G. & Lansdorp, P. M. Telomeres and Aging. *Physiol. Rev.* **88**, 557–579 (2008).
6. Pearce, E. E. *et al.* Telomere length and epigenetic clocks as markers of cellular aging: a comparative study. *GeroScience* **44**, 1861–1869 (2022).
7. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
8. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621 (1961).
9. Blackburn, E. H. Switching and Signaling at the Telomere. *Cell* **106**, 661–673 (2001).
10. Wright, W. E., Piatyszek, M. A., Rainey, W. E., Byrd, W. & Shay, J. W. Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.* **18**, 173–179 (1996).
11. Ulaner, G. A., Hu, J.-F., Vu, T. H., Giudice, L. C. & Hoffman, A. R. Tissue-specific alternate splicing of human telomerase reverse transcriptase (hTERT) influences telomere lengths during human development. *Int. J. Cancer* **91**, 644–649 (2001).
12. Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).

13. Shay, J. W. & Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33**, 787–791 (1997).
14. Aubert, G., Baerlocher, G. M., Vulto, I., Poon, S. S. & Lansdorp, P. M. Collapse of Telomere Homeostasis in Hematopoietic Cells Caused by Heterozygous Mutations in Telomerase Genes. *PLOS Genet.* **8**, e1002696 (2012).
15. Huang, E. (Elijah) *et al.* The Maintenance of Telomere Length in CD28+ T Cells During T Lymphocyte Stimulation. *Sci. Rep.* **7**, 6785 (2017).
16. Wherry, E. J. T cell exhaustion. *Nat. Immunol.* **12**, 492–499 (2011).
17. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
18. Beltra, J.-C. *et al.* Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity* **52**, 825-841.e8 (2020).
19. Riegel, D. *et al.* Integrated single-cell profiling dissects cell-state-specific enhancer landscapes of human tumor-infiltrating CD8+ T cells. *Mol. Cell* **83**, 622-636.e10 (2023).
20. Bellon, M. & Nicot, C. Telomere Dynamics in Immune Senescence and Exhaustion Triggered by Chronic Viral Infection. *Viruses* **9**, 289 (2017).
21. Timashev, L. A., Babcock, H., Zhuang, X. & de Lange, T. The DDR at telomeres lacking intact shelterin does not require substantial chromatin decompaction. *Genes Dev.* **31**, 578–589 (2017).
22. Feuerbach, L. *et al.* TelomereHunter – in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics* **20**, 272 (2019).
23. Sieverling, L. *et al.* Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat. Commun.* **11**, 1–13 (2020).
24. Hartlieb, S. A. *et al.* Alternative lengthening of telomeres in childhood neuroblastoma from genome to proteome. *Nat. Commun.* **12**, 1269 (2021).

536 25. Stephens, Z., Ferrer, A., Boardman, L., Iyer, R. K. & Kocher, J.-P. A. Telogator: a
537 method for reporting chromosome-specific telomere lengths from long reads. *Bioinforma.*
538 *Oxf. Engl.* **38**, 1788–1793 (2022).
539 26. Mazzucco, G., Huda, A., Galli, M., Zanella, E. & Doksani, Y. Purification of mammalian
540 telomeric DNA for single-molecule analysis. *Nat. Protoc.* **17**, 1444–1467 (2022).
541 27. Wei, Q. *et al.* Multi-omics and single cell characterization of cancer immunosenescence
542 landscape. *Sci. Data* **11**, 739 (2024).
543 28. Bindea, G. *et al.* Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the
544 Immune Landscape in Human Cancer. *Immunity* **39**, 782–795 (2013).
545 29. Gu-Trantien, C. *et al.* $CD4^+$ follicular helper T cell infiltration predicts breast cancer
546 survival. *J. Clin. Invest.* **123**, 2873–2892 (2013).
547 30. Zappasodi, R. *et al.* Non-conventional Inhibitory CD4+Foxp3−PD-1hi T Cells as a
548 Biomarker of Immune Checkpoint Blockade Activity. *Cancer Cell* **33**, 1017-1032.e7
549 (2018).
550 31. Yakovenko, I. *et al.* Telomemore enables single-cell analysis of cell cycle and chromatin
551 condensation. 2023.03.19.533267 Preprint at https://doi.org/10.1101/2023.03.19.533267
552 (2024).
553 32. Kamphorst, A. O. *et al.* Rescue of exhausted CD8 T cells by PD-1 targeted therapies is
554 CD28-dependent. *Science* **355**, 1423–1427 (2017).
555 33. Ludlow, A. T. *et al.* Quantitative telomerase enzyme activity determination using droplet
556 digital PCR with single cell resolution. *Nucleic Acids Res.* **42**, e104 (2014).
557 34. Budimir, N., Thomas, G. D., Dolina, J. S. & Salek-Ardakani, S. Reversing T-cell
558 Exhaustion in Cancer: Lessons Learned from PD-1/PD-L1 Immune Checkpoint
559 Blockade. *Cancer Immunol. Res.* **10**, 146–153 (2022).
560 35. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control
561 and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
562 36. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *J. Comput.*
563 *Graph. Stat.* **5**, 299–314 (1996).

564 37. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21
565 (2019).
566 38. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with
567 Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

568