Telomemore enables single-cell analysis of cell cycle and chromatin condensation

Authors:
Ionut Sebastian Mihai[1,2,3], William Rosenbaum[2], Andy Dernstedt[4], Remigius Gröning[4], Johan Trygg[5,6], Mattias Forsell[4], Johan Henriksson[1,2,7*]

Affiliations:
1. Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå, Sweden
2. Umeå Centre for Microbial Research, Department of Molecular Biology, Umeå University, Umeå, Sweden
3. Industrial Doctoral School, Umeå University
4. Department of Clinical Microbiology, Umeå University, Umeå, Sweden
5. Department of Chemistry, Umeå University, Umeå, Sweden
6. Sartorius Corporate Research, Umeå, Sweden
7. Integrated Science Lab, Umeå, Sweden

* Corresponding author

Abstract

It is now easy to perform multiome single-cell analysis, including both RNA and ATAC readouts from the same cell. This enables a closer linkage between the two types of modalities, but it remains an open question what more information can be extracted from this type of data. ATAC-seq is normally only used to assay transcription factor binding to open regions. By reanalyzing several large datasets, and generating an atlas of B cells, we show that telomere accessibility can better pinpoint processes related to cell cycle and chromatin condensation. We provide Telomemore, a tool that can extract telomeric reads, and give examples of new findings it enables. Our new findings will aid in the annotation and analysis of single-cell ATAC or multiome datasets.

Introduction

ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) is a widely used method for finding open chromatin regions[1]. With this readout also becoming available in the commercial 10x Genomics “multiome” single-cell platform, ATAC-seq can be expected to become even more popular. It is thus important to understand how such data should be interpreted. ATAC-seq is based on the enzyme Tn5 fragmenting and tagging (“tagmenting”) chromatin, which is then sequenced (Figure 1a). The common analysis pipeline results in a dimensional reduction and clustering of similar cells. The accessibility of regions can then be compared, and in particular, transcription factor (TF) activity can be quantified based on motif presence in accessible regions[2]. Thus upstream regulators of different cell types and fates can be discovered.

There are however several open questions about the interpretation of ATAC-seq data. For example, fragments mapping outside peak regions usually remain uninvestigated, and ChIP-seq peaks also exist in places without ATAC-seq peaks[3]. This suggests that some TF-bound regions are inaccessible (Figure 1b). We thus set out to explore what other information can be extracted from ATAC-seq data. We show that telomeres are tagmented, despite being protected by the shelterin complex[4] (Figure 1c). Furthermore, we demonstrate that the telomere fragment abundance is reflected by chromatin condensation, and interpret this in terms of cell cycle progression. Because the cell cycle is linked to many biological processes, and confounds single-cell analysis, telomere accessibility is of wide utility in interpreting single cell data. We provide a new tool to compute it, Telomemore, and precomputed telomere abundances for several large public datasets. Finally, we provide a new single-cell atlas of human tonsil B cells and show that telomeres help contextualize genes related to somatic hypermutation.

Results

Telomeric but not centromeric regions are tagmentable in cells

With the new complete genome sequence from the telomere-2-telomere consortium[5], we could assay the origin of ATAC-seq reads after 10x chromium library preparation. Centromeres are

largely inaccessible (Figure 1d). We however noted the presence of telomeric reads. To annotate telomeres at high precision, we implemented a telomere counter (Telomemore) for single-cell ATAC. We define normalized telomere accessibility (nTA) as telomere counts divided by cell total ATAC read counts. A comparison of nTA for different datasets is shown in Figure 1e. We find a large spread in telomere coverage but it is the highest in samples prepared using the 10x chromium RNA+ATAC multiome kit v1. All single-cell methods capture more telomeres than one large bulk ATAC PBMC dataset[6].

Because individual cells have few telomere fragments, and they are not pure repeats of $[TTAGGG]_n$, it was possible to deduplicate the reads based on the sequence. In PBMC multiome data, with 3,800 UMI/cell and 26,000 peaks/cell (after filtering), we find approximately 40% unique telomere reads and a near linear correlation with unduplicated telomere count (Figure 1f). Thus in applications where the telomeric reads are of high importance, deeper sequencing may be beneficial. We also find global differences between PBMC datasets that may need to be batch corrected. We therefore use the within library-rank percentile to correct nTA. The telomere counts, for each cell of the Figure 1g datasets, are provided as Supplementary data. For brevity, only a number of these datasets are described in this study.

Telomere accessibility is linked to chromatin condensation, but not to telomere length

We expected telomere read abundance to be driven by telomere length, as longer telomeres can theoretically be divided into more fragments. To test this hypothesis, we used the common assumption that telomere length decreases with age. By extracting nTA from bulk ATAC-seq of PBMCs from 54 men and 66 women[6], we could however conclude that nTA is not significantly correlated with age, and thus not with length (Figure 2a). This was true also after filtering out datasets that, according to the mapped read length distribution, had differences in library size selection.

To clarify this lack of correlation, we extracted telomeres and interpreted them from a large number of bulk ATAC-seq datasets. Since telomeres are typically shortened during cell replication, we found a dataset where cells that had been sorted by cell cycle stage using flow cytometry of particular interest[7]. Our analysis shows that nTA was reduced in the S-phase (Figure 2b). We therefore hypothesize a model in which the telomere accessibility is constant, but overall chromosome accessibility is not. In such a model, nTA should increase during chromatin condensation. This hypothesis is further supported by time course ATAC-seq data of mouse and human CD4 T cells undergoing *in vitro* activation (Figure 2c). Naive T cells, until activated, remain in a condensed dormant condensed state which can even be seen using transmission electron microscopy (TEM)[8]. During activation, and entry into the cell cycle, nTA decreases, as predicted by the model.

Collectively, our analysis demonstrates that chromatin condensation is a driver for nTA, but not telomere length in any significant manner. This is contrary to our initial expectation, as the length varies approximately 5-15 kb between humans and is also cell type dependent[9] - thus providing up to a factor 3x difference. A potential explanation would be if only a fixed length part of the telomere is accessible to tagmentation, such as near the T- or D-loops (Figure 1c), or in a

specific part of the subtelomeric region (some subtelomeric patches are accessible to m6A-MTase labeling[10]). An analysis of the telomere fragment base frequency shows that the reads consist of well-aligned $[CCCTAA]_n$ repeats (Figure 2d). The strong pattern shows that Tn5 has a strong preference to tagment before the CCC site. Since Tn5 normally prefers a G at the first base[11], also for naked DNA, we cannot tell whether the shelterin proteins TRF1/TRF2, which target TTAGGG, affect the Tn5 sequence preference. The precise action of Tn5 on the telomere thus cannot be concluded given current data, and we thus simply accept the empirical trend of higher nTA in more condensed chromatinTelomere accessibility is linked to stemness and differentiation

The cell cycle is a natural driver of chromatin condensation, but not the only one. To get an overview of telomere accessibility, we particularly reanalyzed publicly available multiome datasets of PBMCs (25,000 cells). We focused on T cells (Figure 3a) and monocytes (Figure 3b), annotated based on previous atlases (see Methods). To find drivers for nTA, we calculated the correlations between nTA, gene expressions and motif activities (Supplemental File 1).

CD4 and CD8 T cells form distinct clusters but have the same most nTA-anticorrelated genes: LEF1 (CD8: -45%, CD4: -19%) and BACH2 (CD8: -50%, CD4: -18%). Similarly, the activity of the TCF1/LEF1 motifs are also the most anticorrelated (CD8:-49%, CD4:-19%). LEF1 binds to the minor groove of DNA, bending it, and is known to regulate large scale chromatin organization in CD8 T cells[12]. CD8-Cre Tcf1/Lef1 KO has shown the need of LEF1 in maintaining CD8 identity, otherwise leading to expression of CD4-related genes and aberrant cytotoxicity[13]. Similar to LEF1, BACH2 also drives the program for stem-like CD8 T cells[14]. The lower nTA suggests that stem-like CD8 T cells have an open chromatin. On the contrary, CCL5 and several granzymes (GZMA, GZMK, GZMH) are among the most positively nTA-correlating genes. NK cells, which are also cytotoxic, and separate from CD4 and CD8, have the highest nTA. The chromatin of NK[15] cells can be compared to T cells[8] using TEM; resting NK cells have a highly compact chromatin, similar to naive T cells.

All monocytes are generally characterized by bilobed horseshoe shaped nuclei, but there are several subsets with different functions[16]. In this dataset, non-classical $CD14^{lo}CD16+$ monocytes have a much higher nTA. The second most positively nTA-correlated gene is CDKN1C (24%), which inhibits proliferation during the G_1 phase[17]. This is at odds with the RNA-based annotation that shows a reduction in G_1 phase. Non-classical monocytes tend to patrol capillaries[16] and we speculate that the increased chromatin compactification may help their navigation; this would be similar to neutrophils, which also have multilobed chromatin that let them enter small gaps in the endothelium and extracellular matrix.

To summarize, nTA pinpoints several interesting cases of chromatin compactification. For the immune system, it can map to an axis from stemness to a resting state, but also predicts nuclei morphological differences that may be linked to the function of the respective cell type.

Telomere accessibility across tissues and cell types pinpoint cell cycle related transcription factors

To obtain a global overview of TA in healthy adult humans, we investigated a recent atlas of the

chromatin accessibility in the human genome, using sci-ATAC-seq[18]. The nTA of the publicly available subset (500k cells) is shown cells *vs* tissues in Figure 4a. Naive T cells have a high nTA reflecting the previous observations on their condensed state. To find common drivers of nTA, we set up two linear models *vs* TF motif activity. The results are similar, both with and without regressing cell type specific effects (Figure 4b-c, values in Supplemental File 2). At the global level, the activity of motifs TP63, GRHL1 and FOSL2 positively correlate with nTA and under chromatin condensation. Among the TP members, having similar motifs, the tumor suppressor and cell cycle inhibitor p53 is the most famous. Knock-out of GRHL1 also leads to cell cycle arrest in G_2/M phase[19], this most condensed phase. Conversely, AR, MEF2D and NR3C1 drive chromatin decondensation. The androgen receptor (AR) is important for breast and prostate cancer proliferation[20]. The pleiotropic MEFs are activated at the G_0/G_1 transition, and promote S-phase entry[21]. Overall the motifs point can be related to cell cycle regulation and are consistent with the interpretation of nTA^{hi} as a marker for chromatin condensation. However, we expect that more information can be extracted, e.g., GRHL1 has also been linked to aging, although in an insulin related manner[22]. Per-cell nTAs are available on Github for further exploration.

Telomere accessibility elucidates B cell somatic hypermutation

We reanalyzed a previous RNA-only single-cell B cell dataset[23] and noticed a strong influence of the cell cycle. Inspired by this, we generated a new 10x multiome dataset of human tonsillar B cells from 5 individuals, aiming to better understand somatic hypermutation (SHM) and class switch recombination (CSR) in the germinal center (GC). About 7,200 B cells were kept after annotation and *in silico* filtration, covering follicular and germinal center B cells (Figure 5a-b). Overall we find good agreement between clustering based on ATAC-seq *vs* RNA-seq, with some exceptions; B cells with interferon response (IFI44L+, IRF-family motifs also enriched) are scattered in the ATAC-seq UMAP. The naive B cells are loosely split into $DNMT3A^{hi/lo}$ cells in RNA and ATAC, but the differences are small. A large group of $MLLT3^{hi}AKAP6^{hi}$ naive B cells are further away from GC. These possibly reflect a house keeping state: MLLT3 is part of the polymerase super elongation complex, while AKAP6 has been linked to non-centrosomal microtubule-organizing centers (MTOCs) during differentiation[24]. These might thus be the most naive B cells in our dataset. Naive cells, which are all $Ccnd3^{hi}$, funnel through a $Ccnd2^{hi}$ state ("Follicular to GC") which form a bridge toward the GC. This bridge is up in activity for the NFKB2 and JUNB families. Coming back from the GC, there are 3 memory cell clusters in order of proximity to the GC: $FCRL4^{hi}$, $TOX^{hi}FOXP1^{lo}$, and $TOX^{lo}FOXP1^{hi}$. Despite no difference in cell cycle composition, these clusters are all nTA^{hi}, with $FCRL4^{hi}$ being the most condensed; this may possibly be linked to the fact that FCRL4+ B cells are considered to be exhausted, or have a hyporeactive phenotype[25].

We separated the GC to obtain better resolution (Figure 5c-d). Leiden clustering splits it primarily by the cell cycle phase, and at higher resolution by Ig class. The LZ-DZ gradient aligns with CD83-CXCR4. However, based on marker genes, the several orders of B cell cycling causes Leiden to possibly not recapitulate the GC maturation process[26] well. Most cycling cells are in the DZ, but MAML3/DNER (Figure 6a) delineate a large cycling Notch-signalling population (DNER+ possible late-stage), previously associated with the LZ[27]. These are

enriched in class-switched IGHA* and IGHG*. Complementary to MAML3+ is a population of cycling unswitched IGHM+, frequently also expressing CD44. These are all CUX1^{lo}, contrary to the idea of CUX1 expression being cell-cycle dependent[28]. However, CUX1 also speeds up proliferation[28], which can be related to the known acceleration of B cell proliferation[29].

The cluster of hypermutating AICDA+ cells is CD22^{lo}FOXP1^{lo}BACH2^{hi}KLF12^{lo}. However, nTA is higher in AICDA+ G2M cells, continuing into a separate cluster (from “GZ DZ G2M” to “GC DZ AICDA*”, Figure 5c). Microscopy has shown that the mutagenase AICDA enters the B cell nucleus during mitosis, and that it acts during early G_1 phase[30]. This suggests that AICDA expression does not need to overlap with its activity, and that nTA might be a more relevant measure of activity - in particular as chromatin condensation may shield other genes from mutagenesis, not just precisely during mitosis. AICDA level is also known to increase with every cell division, known to be correlated with BCR affinity, here possibly represented as an extended nTA/AICDA gradient toward the LZ clusters[31]. A pre-apoptotic subcluster is also present, being BMP7^{hi} (which is pro-apoptotic[32]) and ZNF385B^{hi} (predicted to be involved in p53 binding).

The condensation is correlated with several genes relevant to proliferation and cell cycle (Figure 6a); The top most correlating gene with nTA is however the actin polymerization regulator KANK1 (22%), for which KO can lead to cell proliferation[33]. nTA also correlates with CDK13 which has been shown to increase Pol II processivity[34]. GWAS has associated CDK13 with the amount of IgD+CD38^{dim}, and IgD+CD24- B cells[35], or lymphocyte count in general[36,37]. It is therefore possible that CDK13 upregulates DNA damage response genes, as previously shown for CDK12[38], and thereby help control the dose of SHM.

Because SHM and CSR depend on base excision repair (BER) and mismatch DNA repair (MMR), we specifically looked at the genes involved. These do not generally overlap with AICDA, with several genes limited to, or near, the S-phase (BER: POLQ, UNG, MSH2. MMR: LIG1, MSH2, MCM8, POLD3, PCNA). This suggests that BER/MMR happens separately from AICDA mutagenesis, after decondensation. AICDA targets also the telomere regions (WRCY motif), and these are lost without repair by UNG[39]. However, as the telomeres get spliced until UNG removes the uracil bases, which happens in the S-phase, we can rule out that free telomere fragments contribute to nTA during the mutagenesis. Overall our analysis shows that telomere accessibility can be given an interpretation that helps pinpoint AICDA activity during condensation of the B-cell chromatin.Discussion

In this study we have shown that telomeric reads can be useful to assay chromatin condensation. This process is in turn intimately linked to the cell cycle and of broad relevance to single-cell data analysis. There is thus value in combined single-cell RNA and ATAC-seq beyond just transcription factor activity analysis. We have presented a tool, Telomemore, that enables easy extraction of this data from 10x datasets. Furthermore, we provide this data from several published dataset, covering over a million cells. As single-cell ATAC-seq data alone can be difficult to interpret, we have also generated new multiome data where the cell cycle is a dominant component. This data will likely prove useful for further methods development in the single-cell domain.

An alternative to measuring telomere accessibility may have been to extract RNA-seq reads of TERRA[40] (Telomeric repeat-containing RNA). However, k-mer based searches in several 10x single-cell RNA-seq datasets typically yielded fewer than 5 potential reads/dataset. Thus the RT-capture efficiency of the current 10x RNAseq chemistry is insufficient, and ATAC-seq is necessary to capture telomeric information.

The production of this manuscript highlighted the need for the FAIR principles (Findability, Accessibility, Interoperability, and Reuse). Much single-cell data has only been made available as processed count tables with the implicit assumption that it contains all interesting information. However, others have made use of mtDNA to perform lineage tracing[41], and we provide further uses of the raw data. We thus urge others to always release their sequencing data as raw FASTQ, and ideally also deposit it in the Human Cell Atlas to aid reprocessing. Finally, much effort of this study went into data access agreements, and this study would not have happened unless we had easy access to primary data. Thus we call for a discussion on what data should be considered “sensitive”, as needlessly hiding raw data slows down research and is against the interest of those benefiting from drugs derived from analysis of human single-cell data.

Online Methods

Analysis of bulk ATAC-seq

CCCTAA-motif scanning was performed using a custom python script on the first reads in the file. For the PBMC *vs* aging dataset[6], we further investigated if the quality of the library impacted the telomere readout. Fragment size distributions were inspected using ATACFragQC v 0.4.5, showing differences probably caused by different size selection procedures. We removed outlier samples based on nucleosome signal, that is, the ratio of fragments between 147 bp and 294 bp (mononucleosome) to fragments < 147 bp (nucleosome-free). No qualitative difference was noticed after clean-up, nor was there a correlation between nTA and nucleosome signal.

The pattern analysis was done using several libraries from the bulk PBMC *vs* aging dataset[6]. Since the libraries are paired end, R1 and R2 were selectively swapped based on G *vs* C-frequency to effectively align them. A custom R script was used to estimate frequencies at each position.

Obtaining of existing single-cell data

Suitable datasets for reanalysis were found from the 10x genomics publication database. Datasets on NCBI SRA were inspected using the online preview to check availability of reads. In some cases the cell barcodes were missing and these datasets are thus impossible to use. In some cases the originally submitted data can be obtained through cloud retrieval but we prioritized other datasets whenever possible. In SRA, cell barcode reads are inconsistently denoted as “biological” or “technical”. To ensure that we obtained the cell barcodes, we retrieved the data using the command “fasterq-dump SRRxxx -e 10 -v --include-technical --split-files”.

Furthermore we processed and broadly analyzed the following datasets: Delacher2021[42]

(GSE156112), Jain2021_pbmc[43] (E-MTAB-11225, E-MTAB-11226), Jain2021_thymus[43] (E-MTAB-9828, E-MTAB-9840), Lyu2021[44] (GSE183684), Morabito2021[45] (GSE174367), Sarropoulos2021[46] (E-MTAB-9765), Satpathy2019[47] (GSE129785), Taavitsainen2021[48] (GSE168667), Wimmers2021[49] (GSE165904), Ziffra2021[50] (GSE163018) and Zhang2021[18] (GSE184462).

The Satpathy2019 GEO dataset GSE129785 was not compatible with ArchR. Also, as cell barcodes were missing in the SRA upload, SRA-to-AWS cloud delivery was used. The BAM files were unaligned to FASTQ using https://github.com/10XGenomics/bamtofastq/, and then realigned using CellRanger. Similarly, we solved issues with Ziffra2021 and Delacher2021 using the cloud delivery service.

Mapping of ATAC-seq fragments to T2T genome

ATAC-seq reads from our multiome B-cell dataset were mapped to the telomere-2-telomere (T2T) consortium reference[5] using STAR 2.7.10a. The genome was built using --limitOutSAMoneReadBytes 10000000 --outFilterMultimapNmax 1000 to aid with centromeric reads which we expect to be multimapping. We also qualitatively checked for centromeric reads by searching for sequences that they may contain.

Common analysis of telomeres and single cell data

Unless otherwise stated, all 10x single-cell ATAC-seq data was aligned using cellranger-atac-2.0.0. Multiome RNA+ATAC-seq datasets were aligned with cellranger-arc 2.0.0. Cell types were annotated according to markers from each respective source article, or annotation files as described for each dataset below.

Telomemore then scanned the aligned BAM-files for reads having more than 3 telomere CCCTAA-motifs. This number was empirically chosen based on a histogram of motif counts across reads, but can also be motivated mathematically. The probability of detecting N telomere motifs in a random M bp sequence is approximately:

$$P[\text{N telomere motifs in one M bp read}] \approx P[\text{one motif}]^{N} * \text{NumberOfPossibleMotifLocations} * P[\text{no motifs in remaining part of M bp read}] \approx 0.25^{N6} * \Pi_{i=1..N} (M-4-i)$$

Thus if more than 3 telomere motifs are required for a 50 bp read to be classified as telomeric, this results in a false positive rate of less than $2*10^{-8}$; or about 1 in 50 million reads.

A study has shown that the nucleosome signal around CTCF sites specifically correlates with mitosis[51]. We attempted to generate such a measure by extracting peaks having CTCF motifs in them, and then calculating the nucleosome signal. We however failed to find a consistent correlation with the cell cycle using this approach. We also correlated the “whole-cell” nucleosome signal to nTA, but obtained both positive, neutral and negative correlations depending on the dataset. We thus conclude that nTA might be linked to the nucleosome signal, but in a complex manner that we do not understand. In particular, there is no noticeable nucleosome signal variation across the GC B cells.

Specific analysis of 10x multiome PBMC data

We obtained the 10x PBMC multiome PBMC demo dataset from www.10xgenomics.com ("Fresh Frozen Lymph Node with B-cell Lymphoma (14k sorted nuclei)", "10k Human PBMCs, Multiome v1.0, Chromium Controller", "10k Human PBMCs, Multiome v1.0, Chromium X"). Cell types were predicted using SingleR[52], based on the DICE[53], HPCA[54] and Monaco[55] reference datasets. Leiden clustering was applied and clusters named according to reference datasets. The nTA-gene correlations are based on the clusters referenced in the text.

Specific analysis of Zhang2021 chromatin accessibility atlas, and chromvar analysis

This dataset further depends on reuse of a pancreas dataset, GSE160472[56], which we included. Privacy protected samples (human heart samples on dbGaP: phs001961, human islet samples on dbGaP: phs002204) were not included. Because the sciATACseq cell barcodes reside in the name of the sequencing read, and NCBI SRA strips the read names in their upload, we had to use GEO cloud delivery. Due to the size of the data, we did not perform alignment and *de novo* cluster assignment. Instead, existing cell type annotations were downloaded from https://data.mendeley.com/datasets/yv4fzv6cnm/1 (1B_Cell_metadata.tsv.gz).

The correlation between gene expression and telomere content was obtained from supplementary the file https://cdn.elifesciences.org/articles/66198/elife-66198-supp2-v2.xlsx[57].

To find motifs linked to telomere accessibility, a linear model was set up using Limma[58] with models as shown in the main figure. Motif scores were calculated using ChromVar[2] through ArchR. CISBP was used as the motif database (http://cisbp.ccbr.utoronto.ca/)[59].

Generation of tonsillar B cell multiome data

The research was carried out according to The Code of Ethics of the World Medical Association (Declaration of Helsinki). Ethical permit was obtained from the Swedish Ethical review authority (No: 2016/53-31) and all samples were collected after receiving informed consent from patient or patient's guardian. Briefly, tonsillar cell suspensions were prepared by tissue homogenizing in RPMI-1640 medium and passed through a 70 µm cell strainer. Red blood cells were lysed using BD PharmLyse lysis buffer according to manufacturer's instructions. All cell suspensions were frozen in fetal bovine serum (FBS) (Gibco) with 10% DMSO and stored in liquid N_2.

B cells from tonsils of 5 individuals were enriched using negative selection magnetic beads (EasySep Human B Cell Isolation kit, Stemcell technologies, #17954). They were then pooled to avoid batch effects, with an average viability per donor of 92% (S.D. 1.5%).

Enriched B cells were washed in ice-cold ATAC-seq resuspension buffer (RSB, 10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2), spun down, and resuspended in 100 mL ATAC-seq lysis buffer (RSB plus 0.1% NP-40 and 0.1% Tween-20 (Thermo Fisher). Lysis was allowed to proceed on ice for 5 min, then 900 mL RSB was added before spinning down again and resuspending in 50 mL 1X Nuclei Resuspension Buffer (10x Genomics). To assess nuclei purity and integrity after lysis, nuclei were stained with Trypan Blue (15250061, Thermo Fisher Scientific), and DAPI

(D1306, Thermo Fisher Scientific), according to manufacturer recommendation. If necessary, cell concentrations were adjusted to equal ratios (per donor) prior to starting single-cell GEM emulsion droplet generation with the ATAC-seq NextGEM kit (10x Genomics). Briefly, nuclei were incubated in a transposition mix. Transposed nuclei were then loaded into a Chromium Next GEM Chip J. 9.000 nuclei were loaded per lane, with a target recovery of 5.500 (doublet rate 4% - 4.8%). After GEM emulsion generation, reverse transcription, cDNA amplification and multiome sc library generation occurred, according to manufacturer specifications (*Chromium Next GEM Single Cell Multiome ATAC + Gene Expression*, CG000338 Rev A, 10x Genomics, September 2020).

Analysis of tonsillar B cell multiome data

Library reads were aligned and aggregated using CellRanger ARC 2.0.0. The classification of TCRs was done using TRUST4[60]. The SNPs of the cells were extracted using cellSNP[61], and assigned to donors using Vireo[62]. No batch effects between donors were noticed. Single-cell RNA-seq analysis was done using Seurat[43], and single-cell ATAC-seq analysis using Signac[63]. MACS2 was used for peak calling[64]. The Vireo doublet score was used to filter out droplets. Cell types were predicted using SingleR[52], based on the HPCA[54] and Monaco[55] reference datasets. Furthermore, labels were transferred using Seurat from a previous single-cell RNA-seq-only tonsillar B cell dataset[65], and compared. Major axes (GC/non-GC and naive vs switched) aligned but precise clusters did not align satisfactorily. The dataset clusters however had similar topology if annotated based on the same marker genes. We did not find discrete clusters corresponding to “activated” nor “FCRL3^{hi}”. Unwanted cells (T cells, dendritic cells) were removed and ignored in this study. Dimensional reductions were done using UMAP[66]. Especially the maker CXCR4^{hi} has been used as a marker for the dark zone (DZ), where proliferation happens. However, CXCR4 expression increases during the cell cycle and the dark-light zone axis is a result of cell migration where they compete for their ligand CXCL12[67].

All highly varying genes and motif activities were correlated to nTA using spearman's method.

Because it has been reported that GC B cells, compared to naive and memory, are higher in TERT and have on average longer telomeres[68], we also investigated this alternative explanation for AICDA^{hi} cells being nTA^{hi}. TERT is however upregulated in rather separate S-phase cells without effect on nTA. Telomeres have also been found to be 1.4kb longer in naive T cells over memory T cells[69]; however, assuming an analogy for B cells, we cannot see such a trend on nTA in our data.

We speculated if we could see differences in accessibility over the Ig region, across different GC B cell clusters - especially due to free-floating spliced Ig gDNA. Pileups over the region did however not show any significant differences.

Supplemental files

Supplemental File #1. Correlations of genes and motifs in multiome PBMC dataset to nTA

Supplemental File #2. Correlations of motifs to Zhang2021 atlas dataset to nTA

Data availability

The tonsillar multiome B cell dataset sequencing data has been uploaded to ArrayExpress (E-MTAB-12632). The Telomemore tool is available at Github, https://github.com/henriksson-lab/telomemore.java. Furthermore, scripts to reanalyze the various datasets, and precomputed abundances, are available at https://github.com/henriksson-lab/telomemore_supplemental. The human tonsil B cell data can be viewed interactively at http://data.henlab.org/.

Author contributions

W.R. conceived the name and implemented the first version of Telomemore. A.D. and R.G. contributed material and helped with flow cytometry. I.S.M. performed the single-cell data generation. J.T., M.F. and J.H. supervised. J.H. conceived the study, the algorithm, and wrote the final implementation. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

All authors have read and agreed to the published version of the manuscript. We thank Matthew Weirauch for excellent help with the CIS-BP database; Magnus Hultdin, Sofie Degerman and Pär Larsson (Umeå university) for discussions about telomeres; and Xi Chen (SUSTech) for discussions about ATAC-seq. The analysis was inspired by a collaboration with Oliver Billker and Ágnes Regős on malaria ATAC-seq. Filippe Seiz De Filippi did a first attempt at annotating the B cells.

Funding

The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2018-05973, under Project SNIC 2021/22-697, SNIC 2021/6-328 and SNIC 2022-5-18. J.H.is supported by Vetenskapsrådet grant number #2021-06602.

Conflicts of Interest

J.T. is employed at Sartorius. I.S.M is employed at Umeå university but partially funded by Sartorius. Other authors declare no conflict of interest.

References

1. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

2. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
3. Henriksson, J. *et al.* Genome-wide CRISPR Screens in T Helper Cells Reveal Pervasive Crosstalk between Activation and Differentiation. *Cell* **176**, 882–896.e18 (2019).
4. Lim, C. J. & Cech, T. R. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat. Rev. Mol. Cell Biol.* **22**, 283–298 (2021).
5. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44–53 (2022).
6. Márquez, E. J. *et al.* Sexual-dimorphism in human immune system aging. *Nat. Commun.* **11**, 751 (2020).
7. Chen, X. *et al.* ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat. Methods* **13**, 1013–1020 (2016).
8. Rawlings, J. S., Gatzka, M., Thomas, P. G. & Ihle, J. N. Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence. *EMBO J.* **30**, 263–276 (2011).
9. Pickett, H. A., Cesare, A. J., Johnston, R. L., Neumann, A. A. & Reddel, R. R. Control of telomere length by a trimming mechanism that involves generation of t-circles. *EMBO J.* **28**, 799–809 (2009).
10. Dubocanin, D. *et al.* Single-molecule architecture and heterogeneity of human telomeric DNA and chromatin. *bioRxiv* 2022.05.09.491186 (2022) doi:10.1101/2022.05.09.491186.
11. Green, B., Bouchier, C., Fairhead, C., Craig, N. L. & Cormack, B. P. Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mob. DNA* **3**, 3 (2012).
12. Shan, Q. *et al.* Tcf1 and Lef1 provide constant supervision to mature CD8+ T cell identity and function by organizing genomic architecture. *Nat. Commun.* **12**, 5863 (2021).
13. Xing, S. *et al.* Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. *Nat. Immunol.* **17**, 695–703 (2016).

14. Yao, C. *et al.* BACH2 enforces the transcriptional and epigenetic programs of stem-like CD8+ T cells. *Nat. Immunol.* **22**, 370–380 (2021).
15. Zarcone, D. *et al.* Ultrastructural analysis of human natural killer cell activation. *Blood* **69**, 1725–1736 (1987).
16. Cros, J. *et al.* Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity* **33**, 375–386 (2010).
17. Eggermann, T. *et al.* CDKN1C mutations: two sides of the same coin. *Trends Mol. Med.* **20**, 614–622 (2014).
18. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e19 (2021).
19. He, Y. *et al.* EGFR-ERK induced activation of GRHL1 promotes cell cycle progression by up-regulating cell cycle related genes in lung cancer. *Cell Death Dis.* **12**, 430 (2021).
20. Michmerhuizen, A. R., Spratt, D. E., Pierce, L. J. & Speers, C. W. ARe we there yet? Understanding androgen receptor signaling in breast cancer. *NPJ Breast Cancer* **6**, 47 (2020).
21. Di Giorgio, E., Gagliostro, E., Clocchiatti, A. & Brancolini, C. The control operated by the cell cycle machinery on MEF2 stability contributes to the downregulation of CDKN1A and entry into S phase. *Mol. Cell. Biol.* **35**, 1633–1647 (2015).
22. Grigolon, G. *et al.* Grainyhead 1 acts as a drug-inducible conserved transcriptional regulator linked to insulin signaling and lifespan. *Nat. Commun.* **13**, 107 (2022).
23. Holmes, A. B. *et al.* Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J. Exp. Med.* **217**, (2020).
24. Vergarajauregui, S. *et al.* AKAP6 orchestrates the nuclear envelope microtubule-organizing center by linking golgi and nucleus via AKAP9. *Elife* **9**, (2020).
25. Jourdan, M. *et al.* Characterization of human FCRL4-positive B cells. *PLoS One* **12**, e0179793 (2017).

26. Song, S. & Matthias, P. D. The Transcriptional Regulation of Germinal Center Formation. *Front. Immunol.* **9**, 2026 (2018).
27. Garis, M. & Garrett-Sinha, L. A. Notch Signaling in B Cell Immune Responses. *Front. Immunol.* **11**, 609324 (2020).
28. Liu, N. *et al.* CUX1, A Controversial Player in Tumor Development. *Front. Oncol.* **10**, 738 (2020).
29. Gitlin, A. D. *et al.* HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells. *Science* **349**, 643–646 (2015).
30. Wang, Q. *et al.* The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med.* **214**, 49–58 (2017).
31. Rush, J. S., Liu, M., Odegard, V. H., Unniraman, S. & Schatz, D. G. Expression of activation-induced cytidine deaminase is regulated by cell division, providing a mechanistic basis for division-linked class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13242–13247 (2005).
32. Bollum, L. K. *et al.* BMP-7 induces apoptosis in human germinal center B cells and is influenced by TGF-β receptor type I ALK5. *PLoS One* **12**, e0177188 (2017).
33. Nguyen, M. T. & Lee, W. Kank1 Is Essential for Myogenic Differentiation by Regulating Actin Remodeling and Cell Proliferation in C2C12 Progenitor Cells. *Cells* **11**, (2022).
34. Greifenberg, A. K. *et al.* Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Rep.* **14**, 320–331 (2016).
35. Orrù, V. *et al.* Complex genetic signatures in immune cells underlie autoimmunity and inform therapy. *Nat. Genet.* **52**, 1036–1045 (2020).
36. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198–1213.e14 (2020).
37. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214–1231.e11 (2020).

38. Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **25**, 2158–2172 (2011).
39. Cortizas, E. M. *et al.* UNG protects B cells from AID-induced telomere loss. *J. Exp. Med.* **213**, 2459–2472 (2016).
40. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38**, 5797–5806 (2010).
41. Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325–1339.e22 (2019).
42. Delacher, M. *et al.* Single-cell chromatin accessibility landscape identifies tissue repair program in human regulatory T cells. *Immunity* **54**, 702–720.e17 (2021).
43. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
44. Lyu, P. *et al.* Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. *Cell Rep.* **37**, 109994 (2021).
45. Morabito, S. *et al.* Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
46. Sarropoulos, I. *et al.* Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science* **373**, (2021).
47. 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).
48. Taavitsainen, S. *et al.* Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse. *Nat. Commun.* **12**, 5307 (2021).
49. Wimmers, F. *et al.* The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. *Cell* **184**, 3915–3935.e21 (2021).
50. Ziffra, R. S. *et al.* Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021).

51. Oomen, M. E., Hansen, A. S., Liu, Y., Darzacq, X. & Dekker, J. CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. *Genome Res.* **29**, 236–249 (2019).
52. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
53. Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* **175**, 1701–1715.e16 (2018).
54. Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C. & Hume, D. A. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics* **14**, 632 (2013).
55. Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* **26**, 1627–1640.e7 (2019).
56. Chiou, J. *et al.* Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat. Genet.* **53**, 455–466 (2021).
57. Hu, K., Ghandi, M. & Huang, F. W. Integrated evaluation of telomerase activation and telomere maintenance across cancer cell lines. *Elife* **10**, (2021).
58. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
59. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
60. Song, L. *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
61. Huang, X. & Huang, Y. Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab358.
62. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).

63. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

64. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

65. King, H. W. *et al.* Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol* **6**, (2021).

66. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

67. Weber, T. S. Cell Cycle-Associated CXCR4 Expression in Germinal Center B Cells and Its Implications on Affinity Maturation. *Front. Immunol.* **9**, 1313 (2018).

68. Weng, N. P., Granger, L. & Hodes, R. J. Telomere lengthening and telomerase activation during human B cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10827–10832 (1997).

69. Weng, N. P., Levine, B. L., June, C. H. & Hodes, R. J. Human naive and memory T lymphocytes differ in telomeric length and replicative potential. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11091–11094 (1995).

Figure legends

Figure 1. Assessment of centromeric and telomeric accessibility
(a) *ATAC-seq is a modern method in which accessible chromatin is "tagmented", i.e. fragmented by the enzyme Tn5 that also adds adapters for sequencing. These fragments are normally used to analyze enhancers.* ***(b)*** *Not all ChIP-seq peaks are present in ATAC-seq, and the reads outside the peaks are normally ignored.* ***(c)*** *The telomere is tagmented, despite being protected by the shelterin complex[4]. Some potential sites of tagmentation have been highlighted.* ***(d)*** *Reads mapped to the T2T reference genome, showing a tightly closed centromeric region.* ***(e)*** *Abundance of telomeres in different ATAC-seq-containing datasets. The B cell multiome dataset is included in this study; fibroblast and T cell datasets are under preparation.* ***(f)*** *Most telomeric reads contain sufficient sequence differences that deduplication is possible. The red line represents no duplicates. Each library has a different offset, suggesting the need for batch correction.*

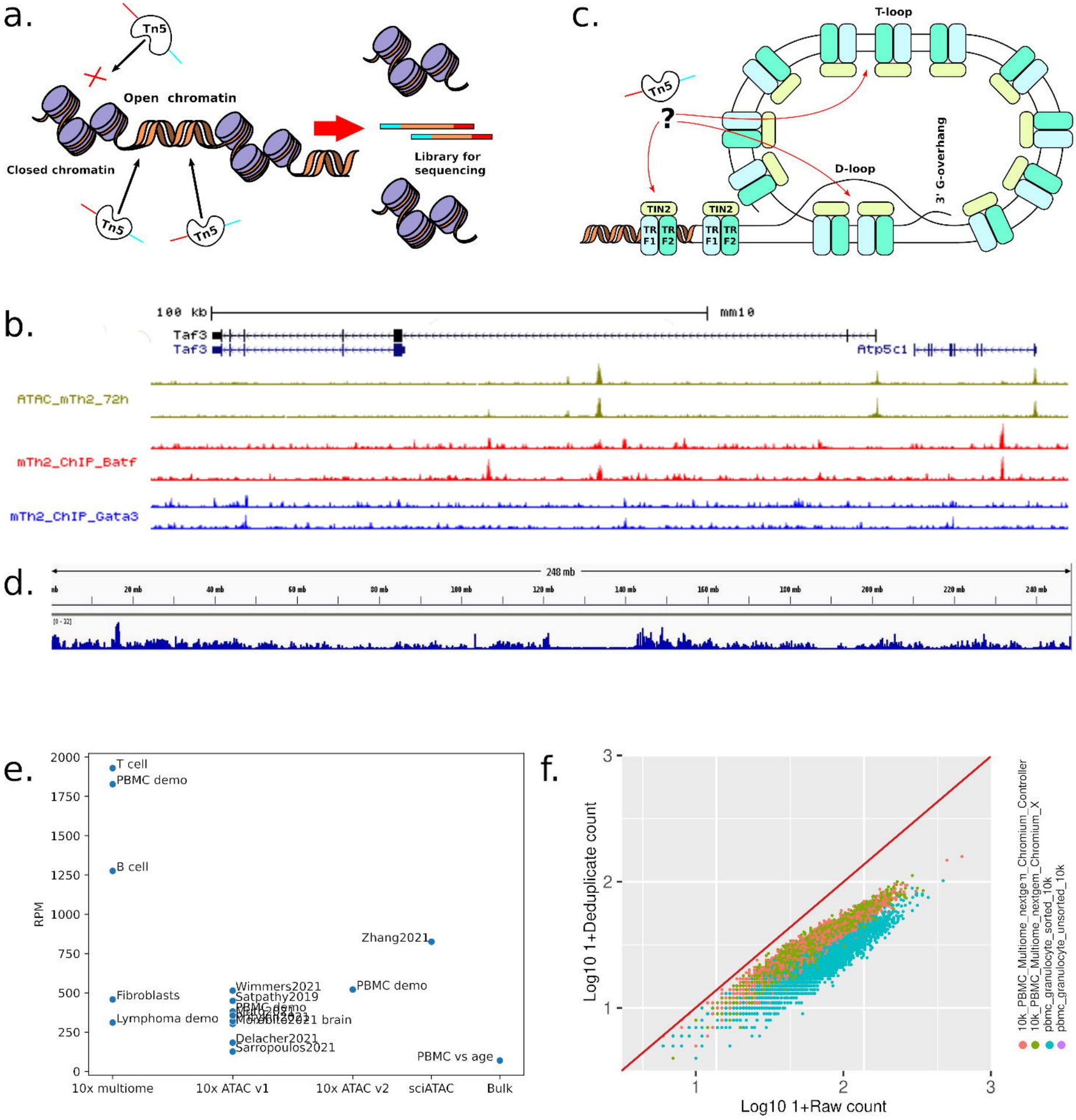
Figure 2. Telomere accessibility is linked to chromatin condensation, but not to telomere length. (a) *Bulk ATAC-seq nTA across PBMCs from donors of different age; no significant correlation is seen, and if any, it is not in the expected direction if driven by telomere length.* ***(b)*** *The bulk ATAC-seq nTA from cells sorted by cell cycle stage[7].* ***(c)*** *Bulk ATAC-seq nTA in human and mouse CD4 T helper type 2 cells during the first 72h of activation[3]. The nTA drops as expected if it is driven by other enhancers opening as part of decondensation (green arrows are conceptual only). Transmission electron microscopy images showing the rapid chromatin decondensation have been generated previously[8].* ***(d)*** *Analysis of sequencing data shows clear Tn5 sequence preference. Reads have been aligned based on GC-richness. Telomeres are not perfect repeats, degenerating the pattern over distance.*

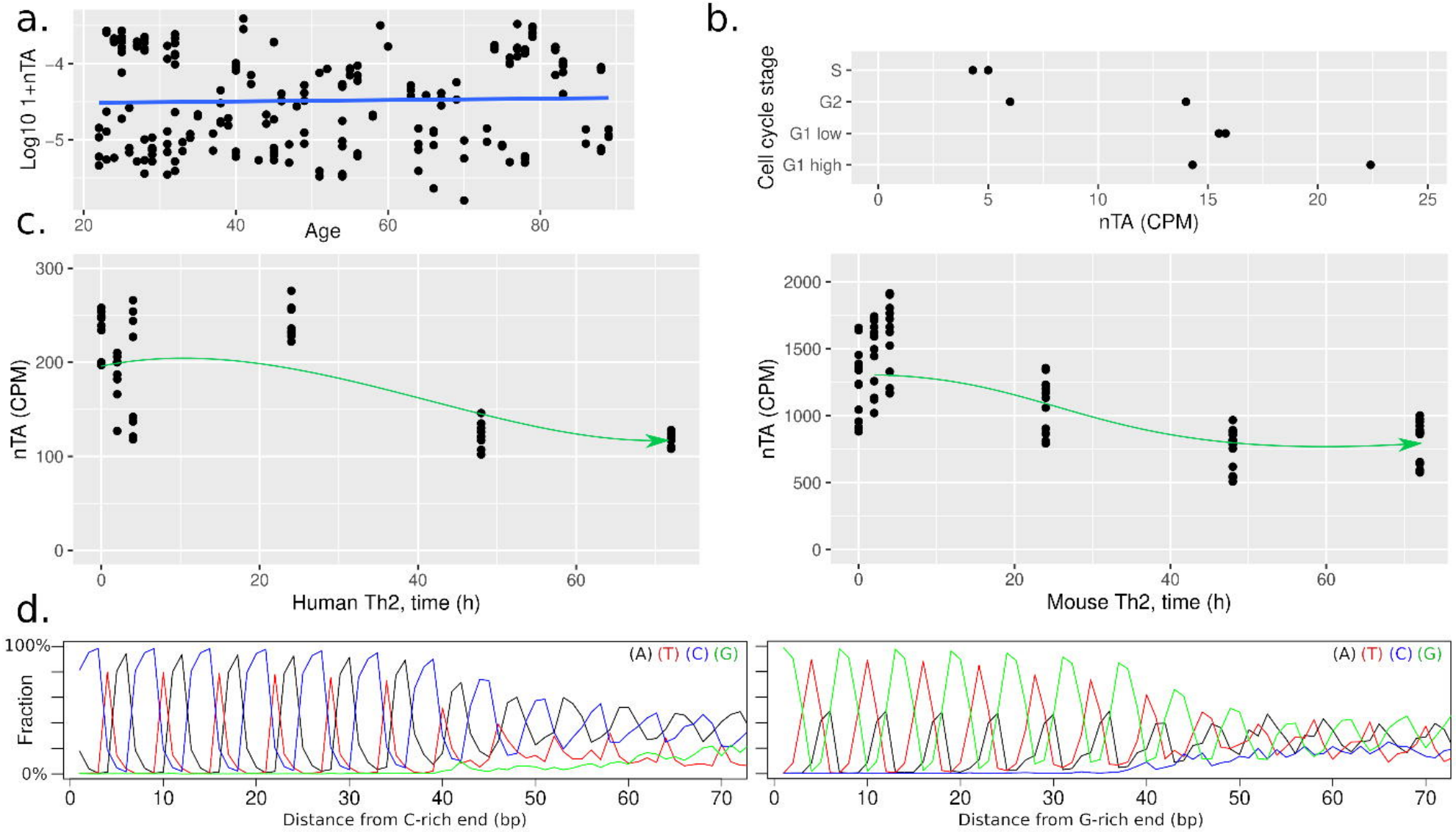
Figure 3. Telomere accessibility is linked to stemness and differentiation. (a) *UMAP of NK and T cells from multiome PBMC data. Effectorness/cytotoxicity axis along which nTA is correlated.* ***(b)*** *UMAP of monocytes, where nTA is up in CD14^{lo}CD16+ non-classical monocytes (CD14=FCGR3A). Cell cycle inhibitor CDKN1C is up in this subset which is not reflected in the standard RNA-seq based cell cycle inference method.*

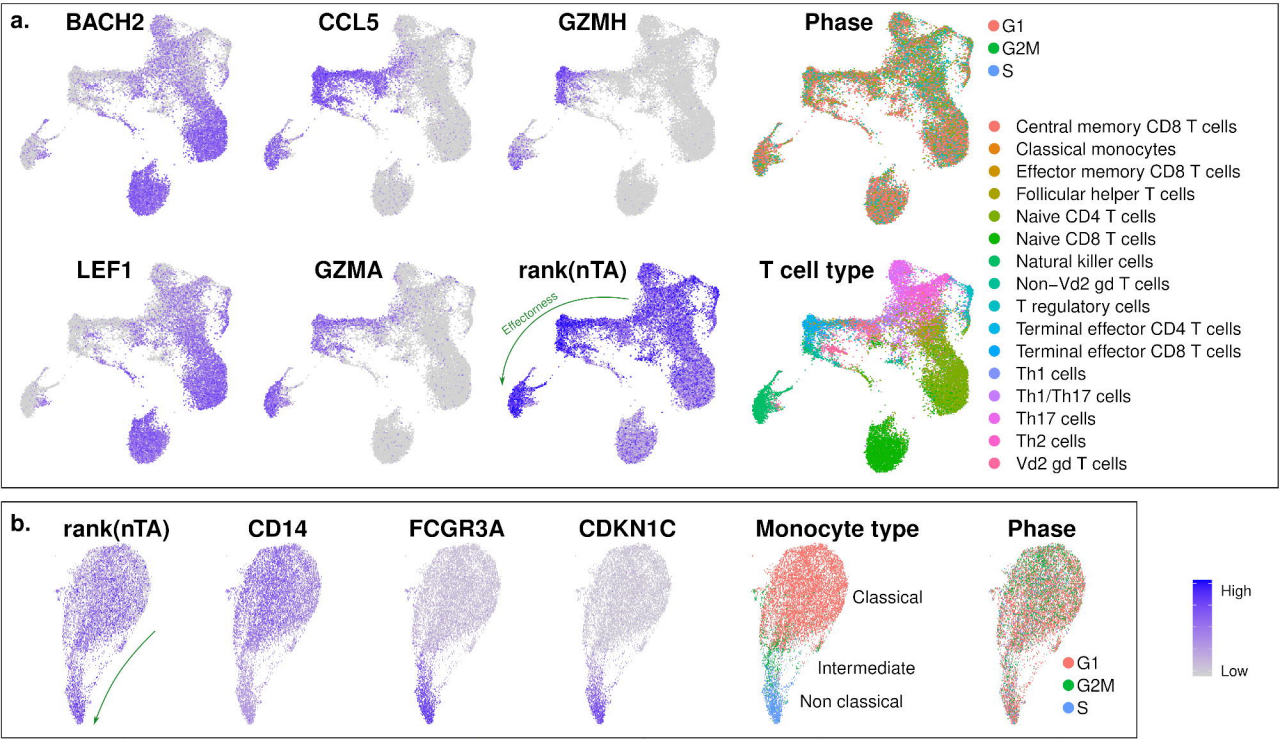
Figure 4. A single-cell atlas of telomere accessibility in the human genome. (a) *Normalized telomere accessibility across human cells and tissues[18].* ***(b)*** *Linear model showing motifs of enhancers that are correlated with nTA.* ***(c)*** *Linear model also removing cell type specific differences.* ***(d)*** *Motif sequences of some of the top drivers of nTA.*

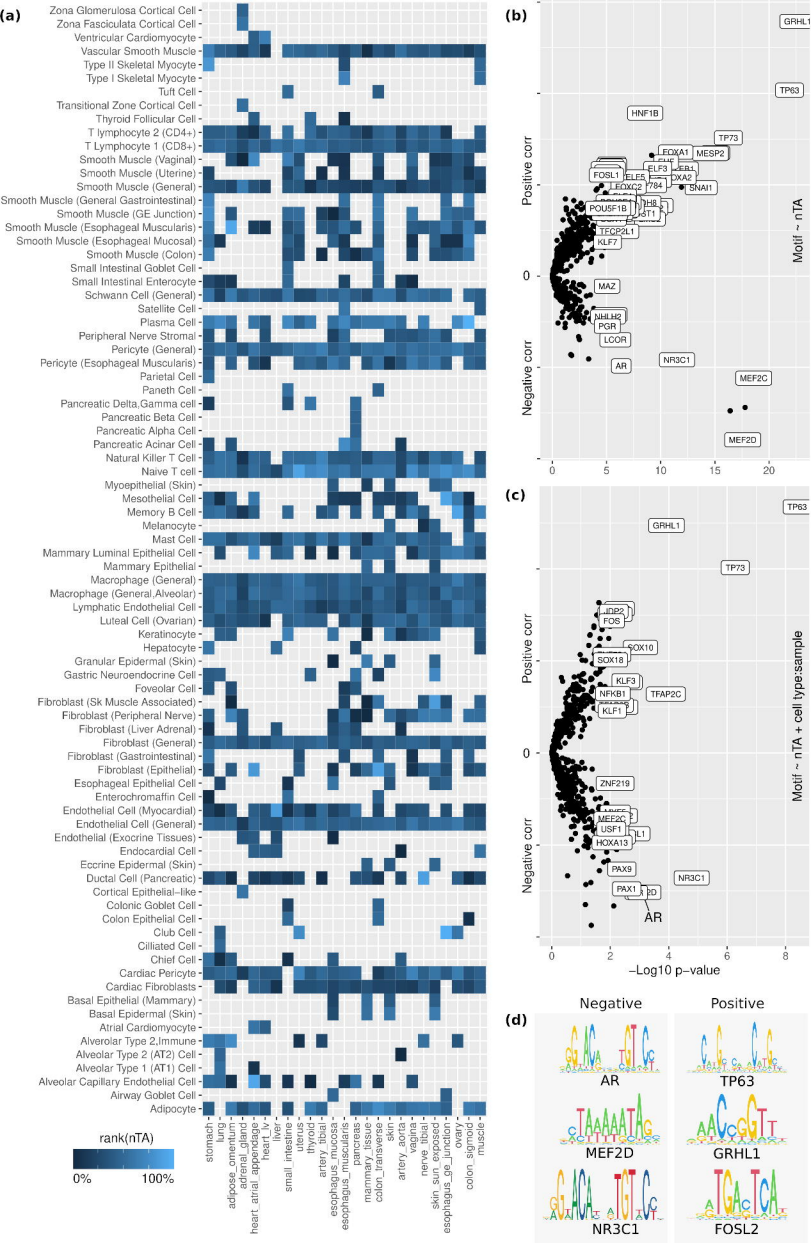
Figure 5. Telomere accessibility pinpoints B cell somatic hypermutation. (a) *UMAP of the different B cell subtypes.* ***(b)*** *Expression levels of some marker genes for the different clusters/cell types.* ***(c)*** *UMAP of the GC, with AICDA expression and telomere accessibility.* ***(d)*** *Marker genes of the GC.*

Figure 6. TF activity, and marker genes for the germinal center. (a) *Volcano plot of correlations between TF gene expression and corresponding motif activity.* ***(b)*** *Motif activities of TFs across clusters. (c) Gene expression patterns, not always well captured by per-cluster summaries.*

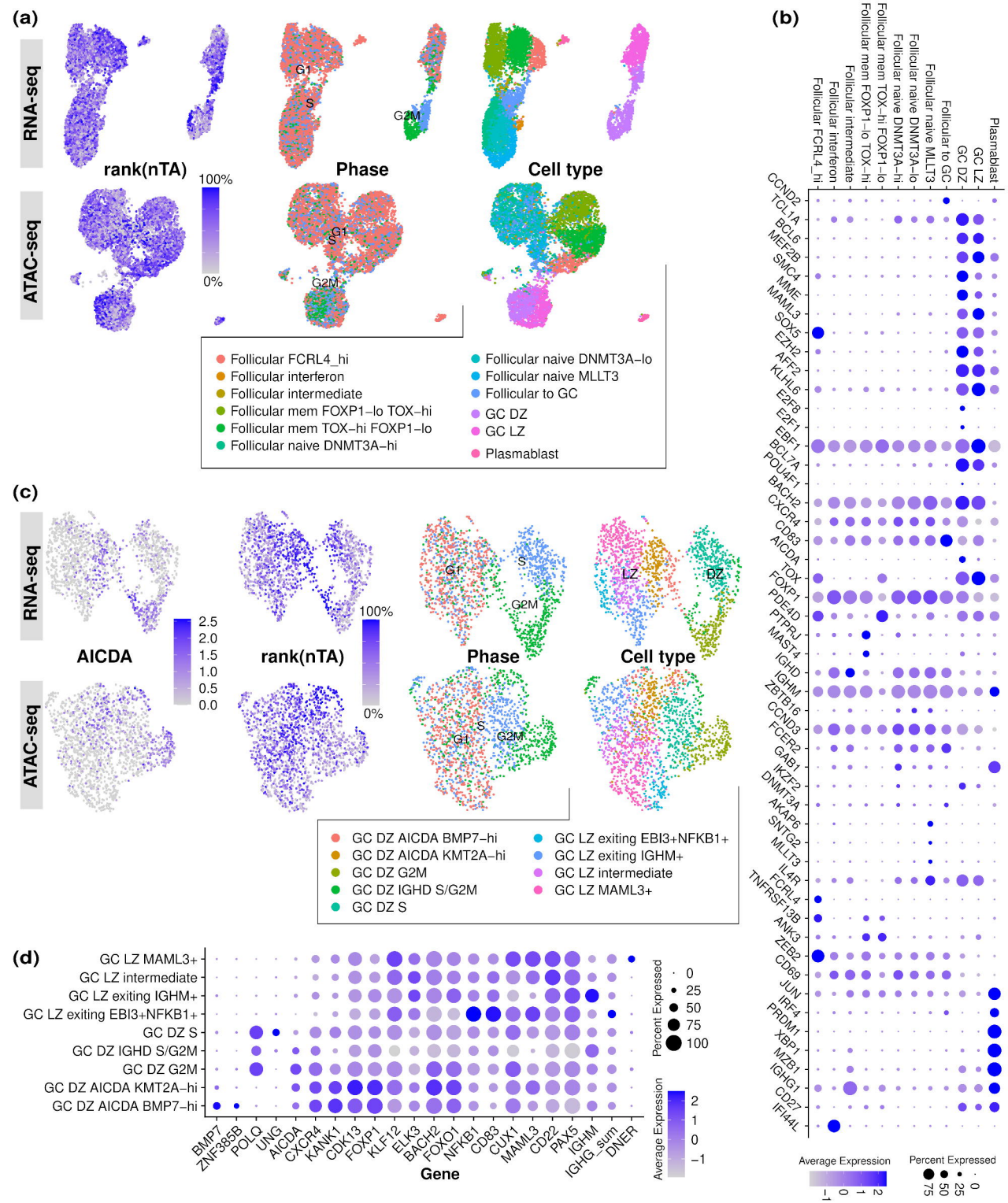

a.
Tn5
Open chromatin
Closed chromatin
Tn5
Tn5
Library for sequencing
c.
T-loop
Tn5
?
D-loop
3' G-overhang
TIN2
TIN2
TR F1
TR F2
TR F1
TR F2
b.
100 kb
mm10
Taf3
Taf3
Atp5c1
ATAC_mTh2_72h
mTh2_ChIP_Batf
mTh2_ChIP_Gata3
d.
248 mb
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
140 mb
160 mb
180 mb
200 mb
220 mb
240 mb
e.
RPM
T cell
PBMC demo
B cell
Zhang2021
Fibroblasts
Lymphoma demo
Wimmers2021
Satpathy2019
PBMC demo
Mohabit2021 brain
Delacher2021
Sarropoulos2021
PBMC demo
PBMC vs age
10x multiome
10x ATAC v1
10x ATAC v2
sciATAC
Bulk
f.
Log10 1+Deduplicate count
Log10 1+Raw count
10k_PBMC_Multiome_nextgem_Chromium_Controller
10k_PBMC_Multiome_nextgem_Chromium_X
pbmc_granulocyte_sorted_10k
pbmc_granulocyte_unsorted_10k

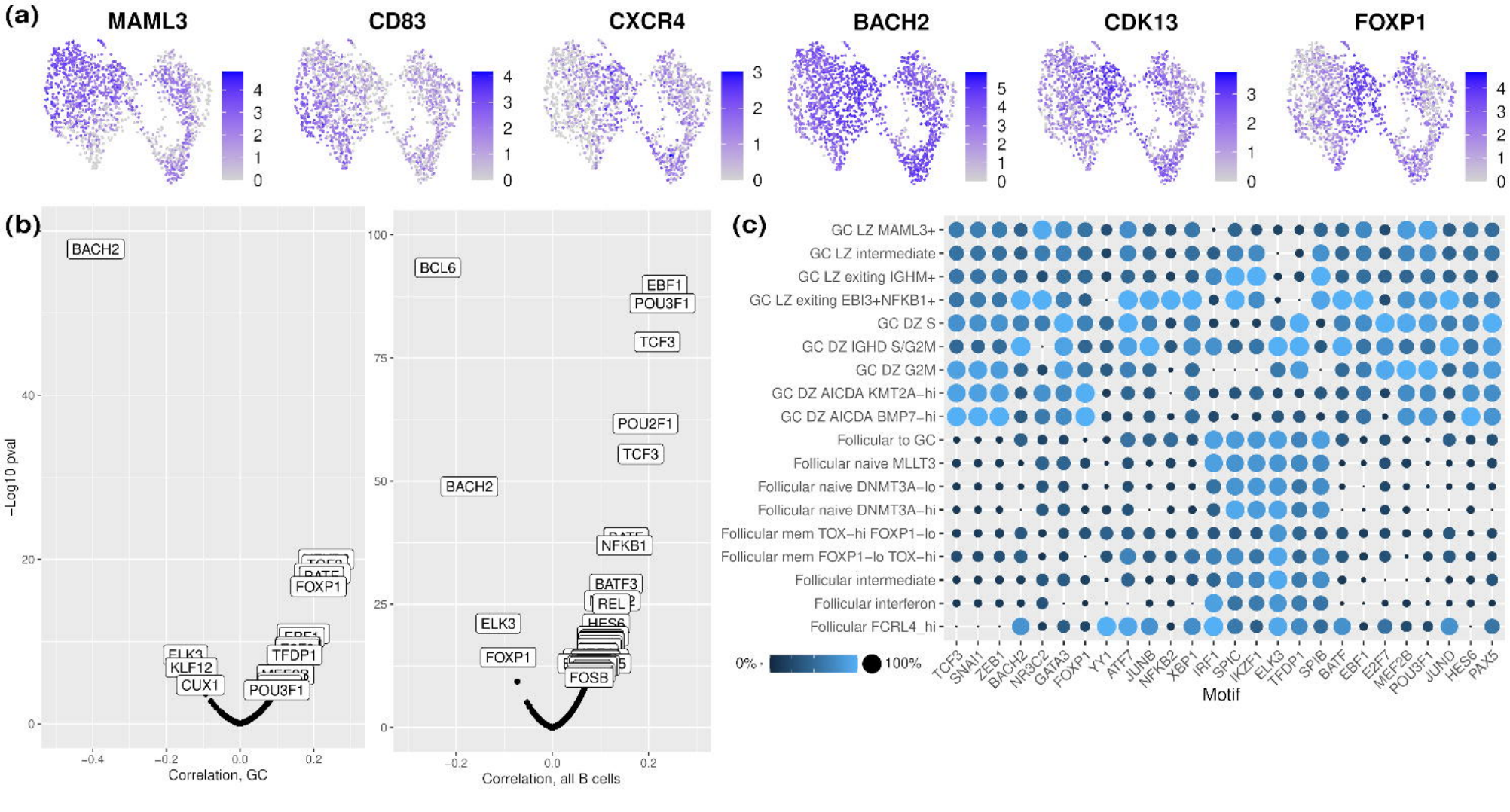

a.
Log10 1+nTA
−4
−5
20
40
60
80
Age
b.
Cell cycle stage
S
G2
G1 low
G1 high
0
5
10
15
20
25
nTA (CPM)
c.
nTA (CPM)
300
200
100
0
0
20
40
60
Human Th2, time (h)
nTA (CPM)
2000
1500
1000
500
0
0
20
40
60
Mouse Th2, time (h)
d.
100%
Fraction
0%
(A) (T) (C) (G)
0
10
20
30
40
50
60
70
Distance from C-rich end (bp)
(A) (T) (C) (G)
0
10
20
30
40
50
60
70
Distance from G-rich end (bp)


a.
BACH2
CCL5
GZMH
Phase
G1
G2M
S
LEF1
GZMA
rank(nTA)
Effectorness
T cell type
Central memory CD8 T cells
Classical monocytes
Effector memory CD8 T cells
Follicular helper T cells
Naive CD4 T cells
Naive CD8 T cells
Natural killer cells
Non−Vd2 gd T cells
T regulatory cells
Terminal effector CD4 T cells
Terminal effector CD8 T cells
Th1 cells
Th1/Th17 cells
Th17 cells
Th2 cells
Vd2 gd T cells
b.
rank(nTA)
CD14
FCGR3A
CDKN1C
Monocyte type
Classical
Intermediate
Non classical
Phase
G1
G2M
S
High
Low



(a)
Zona Glomerulosa Cortical Cell
Zona Fasciculata Cortical Cell
Ventricular Cardiomyocyte
Vascular Smooth Muscle
Type II Skeletal Myocyte
Type I Skeletal Myocyte
Tuft Cell
Transitional Zone Cortical Cell
Thyroid Follicular Cell
T lymphocyte 2 (CD4+)
T Lymphocyte 1 (CD8+)
Smooth Muscle (Vaginal)
Smooth Muscle (Uterine)
Smooth Muscle (General)
Smooth Muscle (General Gastrointestinal)
Smooth Muscle (GE Junction)
Smooth Muscle (Esophageal Muscularis)
Smooth Muscle (Esophageal Mucosal)
Smooth Muscle (Colon)
Small Intestinal Goblet Cell
Small Intestinal Enterocyte
Schwann Cell (General)
Satellite Cell
Plasma Cell
Peripheral Nerve Stromal
Pericyte (General)
Pericyte (Esophageal Muscularis)
Parietal Cell
Paneth Cell
Pancreatic Delta,Gamma cell
Pancreatic Beta Cell
Pancreatic Alpha Cell
Pancreatic Acinar Cell
Natural Killer T Cell
Naive T cell
Myoepithelial (Skin)
Mesothelial Cell
Memory B Cell
Melanocyte
Mast Cell
Mammary Luminal Epithelial Cell
Mammary Epithelial
Macrophage (General)
Macrophage (General,Alveolar)
Lymphatic Endothelial Cell
Luteal Cell (Ovarian)
Keratinocyte
Hepatocyte
Granular Epidermal (Skin)
Gastric Neuroendocrine Cell
Foveolar Cell
Fibroblast (Sk Muscle Associated)
Fibroblast (Peripheral Nerve)
Fibroblast (Liver Adrenal)
Fibroblast (General)
Fibroblast (Gastrointestinal)
Fibroblast (Epithelial)
Esophageal Epithelial Cell
Enterochromaffin Cell
Endothelial Cell (Myocardial)
Endothelial Cell (General)
Endothelial (Exocrine Tissues)
Endocardial Cell
Eccrine Epidermal (Skin)
Ductal Cell (Pancreatic)
Cortical Epithelial-like
Colonic Goblet Cell
Colon Epithelial Cell
Club Cell
Ciliated Cell
Chief Cell
Cardiac Pericyte
Cardiac Fibroblasts
Basal Epithelial (Mammary)
Basal Epidermal (Skin)
Atrial Cardiomyocyte
Alverolar Type 2,Immune
Alveolar Type 2 (AT2) Cell
Alveolar Type 1 (AT1) Cell
Alveolar Capillary Endothelial Cell
Airway Goblet Cell
Adipocyte
stomach
lung
adipose_omentum
adrenal_gland
heart_atrial_appendage
heart_lv
liver
small_intestine
uterus
thyroid
artery_tibial
esophagus_mucosa
esophagus_muscularis
pancreas
mammary_tissue
colon_transverse
skin
artery_aorta
vagina
nerve_tibial
skin_sun_exposed
esophagus_gej_junction
ovary
colon_sigmoid
muscle
rank(nTA)
0%
100%
(b)
GRHL1
TP63
HNF1B
TP73
FOXA1
MESP2
FOSL1
ELF5
ELF3
DXA2
SNAI1
FOXC2
POU5F1B
TFCP2L1
KLF7
MAZ
NHLH2
PGR
LCOR
AR
NR3C1
MEF2C
MEF2D
Positive corr
Negative corr
Motif ~ nTA
0
5
10
15
20
(c)
TP63
GRHL1
TP73
JDP2
FOS
SOX10
SOX18
KLF3
NFKB1
TFAP2C
KLF1
ZNF219
MEF2C
USF1
HOXA13
PAX9
NR3C1
PAX1
AR
Positive corr
Negative corr
Motif ~ nTA + cell type:sample
0
2
4
6
8
−Log10 p−value
(d)
Negative
Positive
AR
TP63
MEF2D
GRHL1
NR3C1
FOSL2


(a)
RNA-seq
ATAC-seq
rank(nTA)
Phase
Cell type
100%
0%
G1
S
G2M
Follicular FCRL4_hi
Follicular interferon
Follicular intermediate
Follicular mem FOXP1−lo TOX−hi
Follicular mem TOX−hi FOXP1−lo
Follicular naive DNMT3A−hi
Follicular naive DNMT3A−lo
Follicular naive MLLT3
Follicular to GC
GC DZ
GC LZ
Plasmablast
(b)
CCND2
TCL1A
BCL6
MEF2B
SMC4
MME
MAML3
SOX5
EZH2
AFF2
KLHL6
E2F8
E2F1
EBF1
BCL7A
POU4F1
BACH2
CXCR4
CD83
AICDA
TOX
FOXP1
PDE4D
PTPRJ
MAST4
IGHD
IGHM
ZBTB16
CCND3
FCER2
GAB1
IKZF2
DNMT3A
AKAP6
SNTG2
MLLT3
IL4R
FCRL4
TNFRSF13B
ANK3
ZEB2
CD69
JUN
IRF4
PRDM1
XBP1
MZB1
IGHG1
CD27
IFI44L
Average Expression
Percent Expressed
(c)
AICDA
2.5
2.0
1.5
1.0
0.5
0.0
LZ
DZ
GC DZ AICDA BMP7−hi
GC DZ AICDA KMT2A−hi
GC DZ G2M
GC DZ IGHD S/G2M
GC DZ S
GC LZ exiting EBI3+NFKB1+
GC LZ exiting IGHM+
GC LZ intermediate
GC LZ MAML3+
(d)
GC LZ MAML3+
GC LZ intermediate
GC LZ exiting IGHM+
GC LZ exiting EBI3+NFKB1+
GC DZ S
GC DZ IGHD S/G2M
GC DZ G2M
GC DZ AICDA KMT2A−hi
GC DZ AICDA BMP7−hi
BMP7
ZNF385B
POLQ
UNG
AICDA
CXCR4
KANK1
CDK13
FOXP1
KLF12
ELK3
BACH2
FOXO1
NFKB1
CD83
CUX1
MAML3
CD22
PAX5
IGHM
IGHG_sum
DNER
Gene
Percent Expressed
Average Expression


(a)
MAML3
CD83
CXCR4
BACH2
CDK13
FOXP1
(b)
BACH2
KLF12
CUX1
TFDP1
POU3F1
FOXP1
−Log10 pval
Correlation, GC
BCL6
EBF1
POU3F1
TCF3
POU2F1
TCF3
BACH2
NFKB1
BATF3
REL
ELK3
FOXP1
FOSB
Correlation, all B cells
(c)
GC LZ MAML3+
GC LZ intermediate
GC LZ exiting IGHM+
GC LZ exiting EBI3+NFKB1+
GC DZ S
GC DZ IGHD S/G2M
GC DZ G2M
GC DZ AICDA KMT2A−hi
GC DZ AICDA BMP7−hi
Follicular to GC
Follicular naive MLLT3
Follicular naive DNMT3A−lo
Follicular naive DNMT3A−hi
Follicular mem TOX−hi FOXP1−lo
Follicular mem FOXP1−lo TOX−hi
Follicular intermediate
Follicular interferon
Follicular FCRL4_hi
0%
100%
TCF3
SNAI1
ZEB1
BACH2
NR3C2
GATA3
FOXP1
YY1
ATF7
JUNB
NFKB2
XBP1
IRF1
SPIC
IKZF1
ELK3
TFDP1
SPIB
BATF
EBF1
E2F7
MEF2B
POU3F1
JUND
HES6
PAX5
Motif