ABSTRACT

The Chromatin Accessibility Signature of Aging in Human Blood Leukocytes Stems from CD8+ T cells

Cheng-Han Chung, Ph.D.

Mentor: Jacques Banchereau, Ph.D.

Human aging is linked to changes in immune function that contribute to decreased responses to pathogens and increased systemic inflammation. Human aging is also associated with profound epigenetic changes across cell types and tissues. How these changes affect the aging –associated decline of the immune system is unknown. The Assay for Transposase Accessible Chromatin with sequencing technology (ATAC-seq) allowed us to study, at a system biology level, the open chromatin landscapes of human peripheral blood mononuclear cells (PBMCs), monocytes, purified B and T cell subsets from healthy young and healthy elderly individuals. We captured aging-associated epigenomic remodeling in PBMCs consisting of (1) systematic chromatin closing at promoters and enhancers targeting the T cell signaling and development and (2) chromatin opening, mostly at quiescent and repressed sites associated with cytotoxicity. Transcriptome profiling of the same individuals revealed gene expression changes concordant with epigenomic changes. Analysis of naïve and memory CD4+ and CD8+ T

cell subsets demonstrated that the epigenomic signature of aging in PBMCs arises mostly from memory CD8+ T cells, indicating that aging differentially affects T cell epigenomes in a subset-specific manner.

This study provides the first systems-level description of chromatin accessibility changes associated with immune aging in human PBMCs and T cell subsets. It revealed in PBMCs significant chromatin closing at promoters and enhancers, including at the *IL7R* locus and the IL-7 signaling pathway. Our study revealed individual-level variability in aging-associated chromatin remodeling and provided a systematic and modular tool for assessing deviations from chronological age. The open chromatin profiling of sorted T cell subsets, concluded that the chromatin "aging signature" captured in PBMCs, mostly stems from memory CD8+ T cells. The combined ATAC-seq/RNA-seq analyses uncovered epigenetic changes poised for expression changes and active noncoding elements (e.g., enhancers), both of which will be essential for understanding the regulatory mechanisms underlying immunosenescence.

Nevertheless, ATAC-seq based open chromatin profiling is a straightforward approach to identify functional genomic regulatory regions, master regulators, and gene regulatory networks controlling complex in vivo processes. In our lab, ATAC-seq is utilized to understand the epigenetics differences in different immune cells and diseases.

The Chromatin Accessibility Signature of Aging in Human Blood Leukocytes Stems from CD8+ T Cells

by

Cheng-Han Chung, B.S., M.S.

A Dissertation

Approved by the Institute of Biomedical Studies

_____

Robert R. Kane, Ph.D., Director

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

_____

Jacques Banchereau, Ph.D., Chairperson

_____

Karolina Palucka, M.D., Ph.D.

_____

Christopher M. Kearney, Ph.D.

_____

Robert R. Kane, Ph.D.

_____

Duygu Ucar, Ph.D.

_____

Elyssia Gallagher, Ph.D.

Accepted by the Graduate School

August 2017

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

Copyright © 2017 by Cheng-Han Chung

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATAC-seq ..... Assay for transposase-accessible chromatin using sequencing

CD ..... Cell differentiation

CMV ..... Cytomegalovirus

ChIP ..... Chromatin Immunoprecipitation

EBV ..... Epstein-Barr virus

ELISA. ..... Enzyme-linked immunosorbent assay

FACS ..... Fluorescence-activated cell sorting

FDR ..... False Discovery Rate

FITC ..... Fluorescein

FRAIRE-seq ..... Formaldehyde-assisted isolation of regulatory elements with sequencing

GLM ..... Generalized linear model

GO ..... Gene Ontology

HIF-1α ..... Hypoxia-inducible factor-1α

IFN ..... Interferon

IL ..... Interleukin

IL-7 ..... Interleukin-7

CHAPTER ONE

Introduction

Epigenetics of Aging

Aging is the process of growing old, and is a complex phenomenon characterized by progressive functional decline at the molecular, cellular, tissue, and organ level (*1*) and results in physical, psychological, and social changes. During aging, the mechanisms that normally maintain health and stress resistance strikingly decline, resulting in decrepitude, frailty, and ultimately death. Many factors can potentially affect the reaction time of aging, such as nutrition, environment, metabolism, exercise, disease, etc. Approximately 150,000 people die each day across the globe, and about two thirds die from age-related causes (http://www.who.int/ageing/publications/global_health.pdf). The major leading causes of death in aging people are heart disease, cancer, stroke, chronic lower respiratory diseases, influenza and pneumonia, Alzheimer disease, diabetes, nephritis, accidents, and septicemia. Unsurprisingly, the majority of these diseases are attributed to genomic instability such as epigenetic modification and DNA damage, causing biological systems to fail. Epigenetics has recently emerged as another possible determinant of aging. Carlos Lopez-Otın, et al. identified nine cellular and molecular hallmarks of aging (Figure 1)(*2*). Undoubtedly, immune responses in the cellular senescence play an important role in aging.

Epigenomic changes associated with aging encompass alterations in transcription factor binding, histone marks, DNA methylation, nucleosome position, and non-coding RNA.

Figure 1. The hallmarks of aging from Lopez-Otin et al., *Cell*, 2013 (*2*). Nine hallmarks described in this figure: genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication, Lopez-Otin et al., *Cell*, 2013 (*2*).

These epigenetic modifications may be used to track donor age forensic analysis or to estimate biological age, C. I. Weidner et al. 2014 (37). Epigenetics refers to changes in phenotype or gene expression caused by mechanisms other than changes in the underlying DNA sequence. The gene regulatory landscape includes many interconnected players that promote or repress gene expression, L. N. Booth et al. (Figure 2).

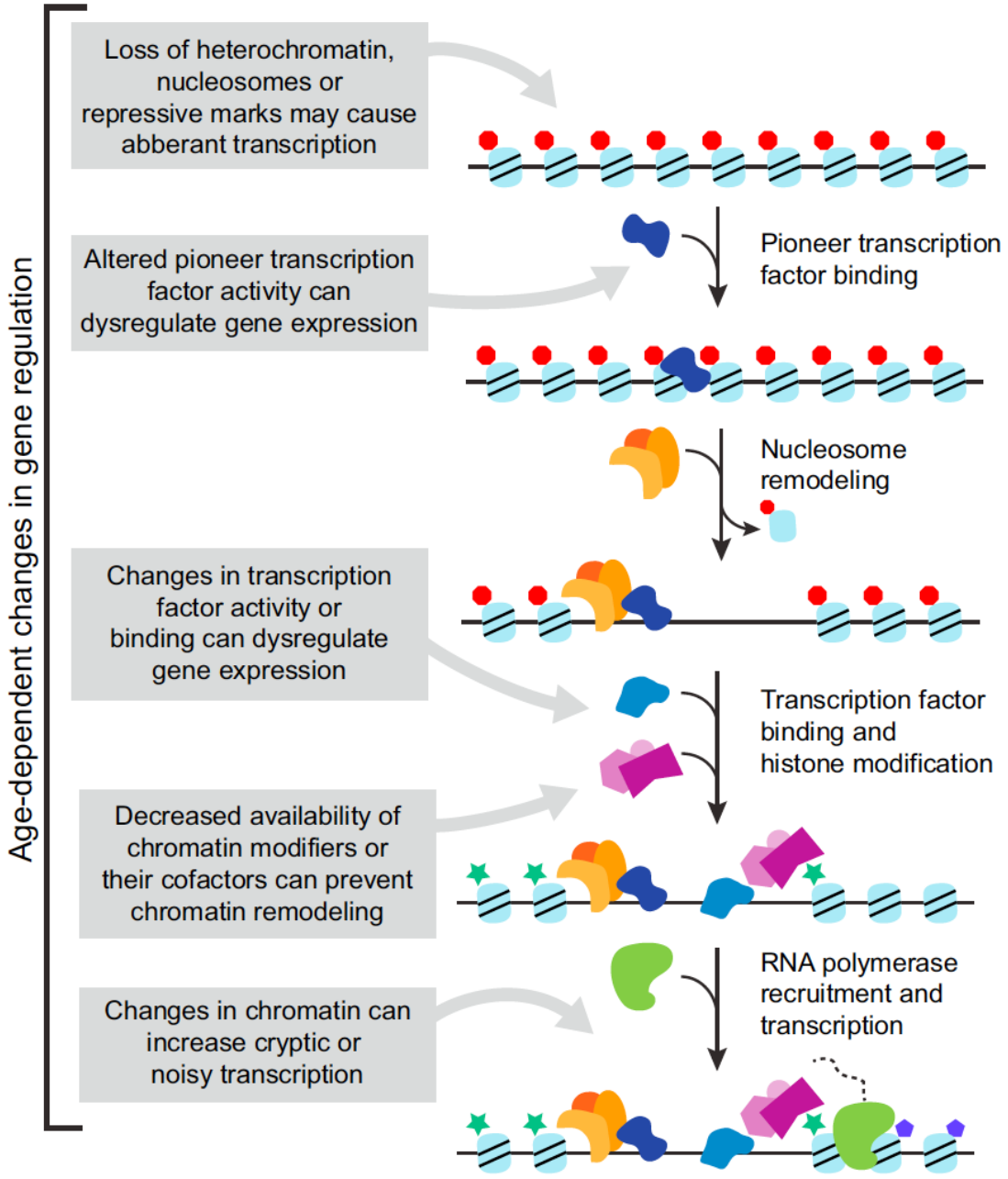


Figure 2. Transcription factors and chromatin modifiers act together to regulate gene expression from L. N. Booth et al *Molecular cell*, 2016 (*1*). The remodeling and transcription activation of a gene occurs in a stepwise fashion that begins with the binding of a sequence-specific DNA-binding protein called a pioneer transcription factor. Following recruitment by a pioneer transcription factor, nucleosome remodeling complexes displace or move nucleosomes and open chromatin. This increases DNA accessibility and allows binding of additional transcription factors, recruitment of chromatin modifiers that remove repressive marks (red hexagons) and add activating marks (green stars), and finally, transcription of the gene. Age-associated changes have been observed at every step of this process, and changes in one step (for example, the activity of pioneer transcription factor) can have down-stream consequences for gene regulation(*1*).

Chromatin and regulation of transcription play key roles in the age-dependent manifestation of these hallmarks of aging and in the response to their emergence. Epigenetic changes, such as histone markers and DNA methylation, take time to develop and are long-lived (*62, 63, 64, 65, 66*). The epigenetic responses to the emergence of a hallmark of aging could create a "memory" of the event, causing it to be perpetuated over time and through cellular divisions. There are a series of factors that induce genome dysregulation (Figure 3)(*1*). First, with age, genomic instability can be attributed to a high frequency of mutations within the genome of a cellular lineage; including changes in nucleic acid sequences, chromosomal rearrangements or aneuploidy, telomere attrition and transposon activation (*67, 68*). Decreased telomerase activity weakens DNA protection against external caused damage, which results in errors in repair and mutations. Another source of genome instability is epigenetic or mutational reductions in expression of DNA repair genes. Secondly, during aging, mitochondria turnover declines and mitochondrial dysfunction increases, C. Lopez-Otin et al (*2*). The activity of mitochondria with age can result in epigenomic changes that alter lifespan. Thirdly, appropriate epigenomic regulation is key to proteostasis, and age-associated epigenomic changes can modify the ability of a cell to respond to proteostatic stress (*69*). Last but not least, an increase in the number of senescent cells and a decline in tissue regeneration due to the loss of stem cell proliferation are hallmarks of aging (*2, 70, 71*). In recent years, it has been appreciated that senescent cells manifest dramatic alteration in their secretome, which is particularly enriched in proinflammatory cytokines (*2*). This proinflammatory secretome, mostly driven by immune cells, may contribute to aging (*72*, *73*).

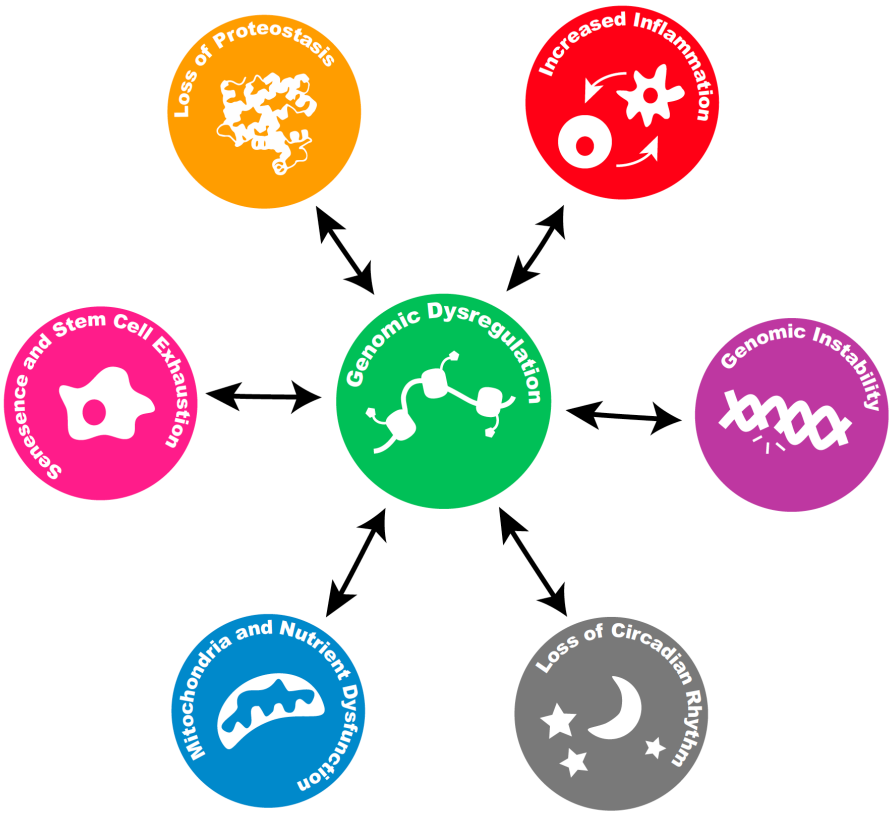


Figure 3. Epigenomic changes are a hub in the progression of aging from L. N. Booth et al. *Molecular cell*, 2016 (*1*). Epigenomic changes—changes in transcription factors, histone marks, nucleosome position, and DNA methylation— are connected with the other hallmarks of aging (*2-4*). Epigenomic changes can trigger the emergence of other hallmarks of aging and can also be affected by them in L. N. Booth et al. *Molecular cell*, 2016(*1*).

Immunosenescence

Aging accompanies gradually diminished immunity. Immunosenescence refers to the decline of the immune system related to aging, contributing to high risk of infection, cancer, and autoimmune diseases in the elderly, G. A. Poland et al (*5, 74*). The progress of immunosenescence is full of complexities and affects on both innate and adaptive immunity limiting the response to pathogens and to different therapies. Some of the age-dependent biological changes that contribute to the onset of immunosenescence are listed in Figure 4(*5*). Hematopoietic stem cells (HSCs) diminish in their self-renewal capacity.

It is due to the accumulation of oxidative damage to DNA by aging and cellular metabolic activity and shortening of telomeric terminals of chromosomes. The total number of phagocytes is decreased in aged hosts. The cytotoxicity of natural killer (NK) cells and the antigen-presenting function of dendritic cells are decreased with old age (*76*). Thus, it is the inability for effector T-lymphocytes to modulate an adaptive immune response (*75*). The humoral immunity is decline because of a reduction in the population of antibody producing B-cells along with a smaller immunoglobulin diversity and affinity.

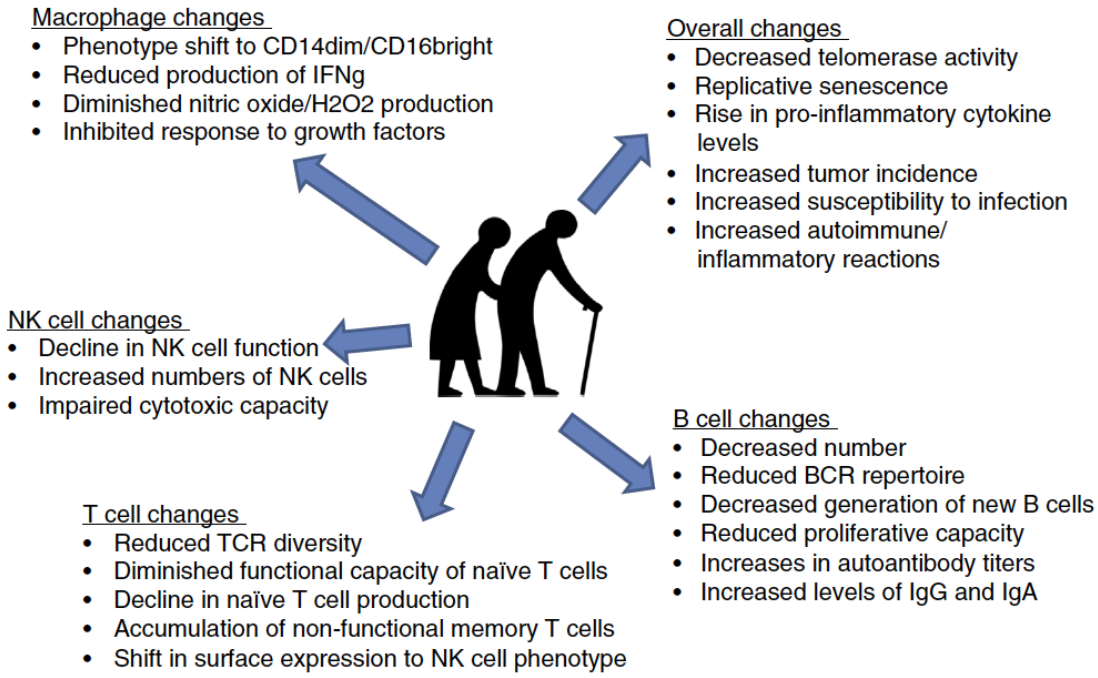


Figure 4. Immunological changes due to immunosenescence from G. A. Poland et al *Current opinion in immunology,* 2014 (*5*).

The changes in the immune response with aging are often paralleled by inflamm-aging or inflammaging (*6, 77*), a state that is associated with gradually increased pro-

inflammatory mediators, which develops due to continuous antigenic stimulation and cellular deterioration in aged subjects. Inflammaging is a highly significant risk factor for both morbidity and mortality in the elderly people (*78*). Generally, stress of immunity can be induced by pathogens, such as cytomegalovirus (CMV), Epstein-Barr virus (EBV) or by other cellular and molecular debris generated by the damage of reactive oxygen species (ROS) (*79, 80*). These factors can induce chronic inflammation and exhaust the adaptive immune response, accelerating unrelated age-associated pathologies. Immune cells responsible for innate immunity (neutrophils and nature killer cells (NK)) or for adaptive immunity (T and B lymphocytes) are continuously stimulated and secrete pro-inflammatory cytokines, such as tumor necrosis factor-α (TNF-α) and IL-6. This is referred to as a low-grade inflammatory response in aging. Pro-inflammatory cytokines and acute-phase proteins (IL-6, IL-8, TNF-α, C-reactive protein) are two characteristics of inflamm-aging that negatively modulate T-cell receptor (TCR)-mediated signaling pathways (*81, 82, 83, 84*).

Alterations in innate immunity may affect the priming of adaptive immunity. For example, the malfunction of aged CD4+ T cells fails to sustain the responses of innate and adaptive immune cells against pathogens or cancer cells. The generation of novel naïve T cells is entirely dependent on thymic function (*7, 85,*). The loss of the thymus prompted a natural supposition that thymic involution is responsible for the age-associated failure of the adaptive immune system (Figure 5) (*8*). Moreover, peripheral expansion of naïve T cells is driven by tonic T cell receptor (TCR) signals and homeostatic cytokines, in particular by interleukin-7 (IL-7). IL-7 drives T cell proliferation and its concentration does not decline with age; therefore, it does not

become limiting (*86, 87*). However, increased IL-7-driven proliferation, possibly due to a lower threshold for IL-7 receptor signaling, leads to abrupt declines in T cell compartment size several years later, D. B. Palmer et al (*8, 87, 88*).

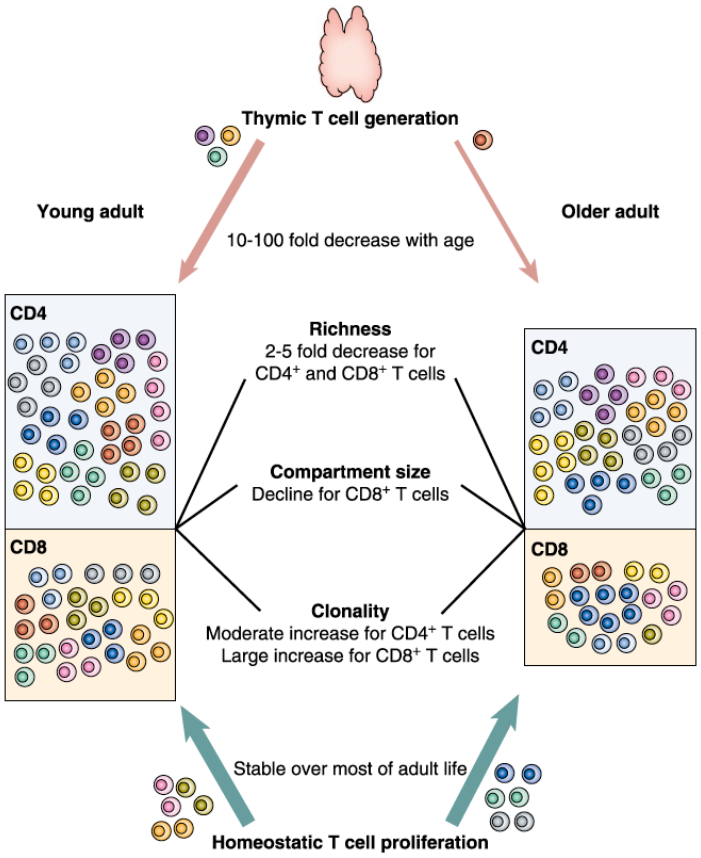


Figure 5. Naive T cell homeostasis and age from D. B. Palmer et al. *Frontiers in immunology,* 2013. Thymic T cell regeneration is quantitatively irrelevant throughout adult life, and homeostatic proliferation is responsible for maintaining the size of the naive T cell compartment. Although only thymic T cell generation can add novel naive T cells and enrich diversity, homeostatic T cell proliferation can sustain the richness of the TCR repertoire (i.e., the total number of T cells with different TCR sequences), whereas peripheral selection during homeostatic proliferation may result in increasing unevenness, that is, increasing inequalities in clonal sizes and clonal expansions of selected few clones. Age-associated changes in these metrics between CD4 and CD8 T cells of young (35 y) and older (65–80 y old) healthy adults are illustrated, D. B. Palmer et al. *Frontiers in immunology,* 2013(*8*).

There is no evidence that T cell turnover decreases with age; possibly, it increases in the very old. Transcriptional and epigenetic profiling could address whether naïve T cells differentiate with age and have transcriptional signatures reminiscent of memory genes or

an epigenetic landscape in which T effector genes are poised. For these reasons, understanding how to overcome the reduced immunity in elderly individuals is in an important research field in aging.

These changes may result in serious clinical consequences such as increased infections, cancers, and autoimmune disease (*9*). Aging is associated with alterations in immune responses, such as altered signaling pathways in immune cells. For example, the mammalian target of rapamycin, mTOR, is a critical kinase in a pathway that regulates many processes, but has been linked mainly to glucose metabolism and longevity (*89*). Rapamycin, an anti-aging drug, has been researched and used to extend the life span of yeasts and humans, and it has been published that rapamycin counteracts certain aging-related alterations in both young and old mice (*89, 90*). The mTOR-signaling pathway is under control of TCR/CD28 stimulation, S. Sauer *et al* (*38*).

Human aging is an ubiquitous complex phenomenon that results from environmental, stochastic, genetic, and epigenetic events in different cell and tissue types and their interactions throughout life(*10*). Human aging is characterized by a chronic, low-grade inflammation, which is involved in innate immunity, crosstalk between innate and adaptive immunity, cell signaling, homeostasis, etc. Adaptive immunity, antigen presenting cells (APCs), T lymphocytes and B lymphocytes, decline with age, whereas innate immunity, NK cells and neutrophils, undergo more subtle changes that could result in mild hyperactivity (*91*).

Cytokines are central to immune cell communication. Therefore, age-related changes in cytokine profiles contribute to many changes in the immune system. Chronic inflammation in adaptive immunity contains several cytokines, molecular pathways,

effector cells, and tissue responses that appear to be shared across multiple age-related diseases. IL-6 is a commonly used marker of inflammatory status, and a hallmark of chronic morbidity. Other inflammatory mediators that increase across multiple age-related diseases include IL-1β and TNF-α. In the cell signaling cascades and transcriptional pathways, IL-1β and TNF-α are regulated by similar upstream molecules, NF-κB, STAT, etc. M. Maggio (39). How the immune responses interact to trigger complicated signaling pathways inducing different cytokines to affect aging is profound.

Next Generation Sequencing Technologies

There are different sequencing technologies to explore immunosenescence at the molecular level. Next Generation Sequencing (NGS), also known as high-throughput sequencing, comprise a number of different modern sequencing technologies including Illumina sequencing and long read (PacBio) sequencing (*92*). These recent technologies, which allow us to sequence DNA and RNA much faster and less expensively than the previously used Sanger sequencing, have revolutionized the study of genomics and molecule biology. Epigenetic processes control gene expression by altering chromatin structure. Epigenomics can be measured by DNA methylation, analysis of DNA/protein interactions, chromatin accessibility and conformation assays. Among these, chromatin accessibility assays include DNase sequencing (DNA-seq), Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq) (*94, 95, 96*), and Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq). DNase-seq and FAIRE-seq, require millions of cells as starting material and take longer to generate sequencing libraries (*93*). In these assays, cells must be expanded ex vivo to gain sufficient material, and the culture conditions will perturb the in vivo context and change the epigenetic state

in unknown ways, and input requirements often prevent application of these assays to clinical samples, thereby, precluding generation of personal epigenomic on diagnostic timescales (*11*). In contrast, ATAC-seq technology has a series of advantages: (1) sequence can be generated from small number of cells even from single cells, J. D. Buenrostro *et* al. (*40*) (2) It has a higher signal-to-noise ratio with low background signal and sharper peaks J. D. Buenrostro et al. (*11*) and (3) It has the ability to determine nucleosome positioning when using paired-end sequencing; and (4) ATAC-seq protocol is straightforward and libraries are easier to generate. ATAC-seq uses a bacterial (Tn5) transposase, an enzyme that inserts a short fragment of DNA (a transposon) into another molecule of DNA, or in this case, inserts two short fragments separate from each other. This recent method can also provide better readouts. Indeed, the lab of William J. Greenleaf at Stanford University concluded that ATAC-seq is compatible with clinical timescales and standard blood draws by generating the open-chromatin landscape from small cell numbers (*11*). Therefore, ATAC-seq technology provides a unique opportunity to profile and study chromatin accessibility profiles of blood-derived human immune cells in the context of aging.

CHAPTER TWO

Objectives

Rationale

Aging is a very complicated process accompanied by significant decline in immune defense that results in increased susceptibility to infections or malignancy compared to young adults. Therefore, trying to find a way to reverse immunity from elderly to young adult, or to slow down the decline of immunity is very important. Nevertheless, trying to identify a set of reliable biomarkers of aging is important not only to monitor the effect of pharmacological interventions and predict the timing of pathologies associated with aging but also to understand the mechanisms and contrive appropriate countermeasures.

The goal of the present work is to analyze and compare the epigenomes of young adults and elderly adults by using chromatin accessibility (ATAC-seq) and gene expression profiling (RNA-seq).

Aims

Aim 1: To generate a comparative epigenomic and transcriptomic database in health young adults and the elderly using ATAC-seq and RNAseq technologies.

Currently, the resource of the ATAC-seq epigenomic database is limited and needs to be expanded. The epigenetic information extracted from healthy young adults and healthy elderly needs to be validated with transcriptomic data using RNA-seq.

Aim 2: Analyze and compare the epigenomic immune signature of young adults and the elderly.

Currently, the ATAC-seq data set can be pre-processed by public computational pipelines. However, tools to integrate ATAC-seq data leading to proper data interpretation are still missing. Developing non-biased proper computational analysis pipeline and applying it to the analysis are very important. We can determine whether ATAC-seq is a proper method for further study and whether a correlation exists between epigenomics and the transcriptomics obtained from blood-derived immune cells.

Aim 3: To understand the mechanisms that are involved in immune regulation in healthy young adults and healthy elderly adults.

The aim of this goal is to determine the cell population, cellular mechanisms, and signaling pathways play an important role in aging.

CHAPTER THREE

Materials and Methods

Subject Recruitment

All studies were conducted following approval by the Institutional Review Board of University of Connecticut Health Center (IRB Number: 14-194J-3). Following informed consent, blood samples were obtained from young (HY, 22-40 yrs.) and 26 old (HO, >= 65 yrs.) healthy volunteers residing in the Greater Hartford, CT, USA region using services of the UConn Center on Aging Recruitment and Community Outreach Research Core (http://health.uconn.edu/aging/research/research-cores/). Recruitment criteria were selected to identify healthy individuals who are experiencing "usual aging" and thus represent the health of the majority of the population within the different age groups. Selecting this type of cohort increases the generalizability of our studies and the likelihood that these findings can be translated to the general population.

Subjects were carefully screened in order to exclude potentially confounding diseases and medications, as well as frailty. All individuals who reported chronic or recent infections within the last two weeks were excluded. Subjects could have underlying chronic diseases, but were excluded if the following were present: congestive heart failure, known kidney disease (serum creatinine > 1.2 mg/dl in men and >1.1 mg/dl in women), diabetes requiring medications, immunosuppressive disorders or the use of immunosuppressive agents including oral prednisone in doses >10 mg daily.

Since declines in self-reported physical performance are highly predictive of frailty, subsequent disability and mortality (*13*), all subjects were questioned as to their ability to walk ¼ mile (or 2-3 city blocks). For those who self-reported an inability to walk ¼ mile (*13*), "Timed Up and Go" (TUG) test was performed and measured as the time taken to stand up from the sitting position, walk 10 feet and return to sitting in the chair (*14*). TUG > 10 sec represented an indication of increased frailty and was an exclusion from the study (*15*).

Primary Cell Isolation

Peripheral blood mononuclear cells (PBMCs) were purified from the blood of healthy donors collected in ACD tubes. Briefly, PBMCs were collected from the mononuclear cell enriched layer obtained by Ficoll (GE Cat no.17-1440-02) density centrifugation of the blood,

Cell Sorting and Flow Cytometry Analysis

For cell sorting, we used fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD27 (M-T271) (Biolegend) and CD4 (RPA-T4), CD45RO (UCHL1), CD45RA (HI100), CD19 (HIB19), CD16 (B73.1), IgD (IA6-2), and CD11c (S-HCL-3) (BD Biosciences), and CD8 (SCF121Thy2D3) and CD19 (J3-119) (Beckman-Coulter). Naïve CD4 (CD4+CD8-CD45RO-CD45RA+), naïve CD8 (CD4-CD8+CD45RO-CD45RA+), memory CD4 (CD4+CD8-CD45RO+CD45RA-), and memory CD8 (CD4-CD8+CD45RO+CD45RA-) T cells were sorted from the CD19-CD16-CD11c- fraction (DUMP channel). Naïve B cells (CD19+IgD+CD27-) were sorted from the CD3-CD16-

CD11c-) fraction (DUMP channel). Cell sorting was performed using the FACSAria Fusion instrument (BD). Monocytes were isolated from fresh PBMCs by positive selection using magnetic CD14 microbeads (Miltenyi Biotech). For phenotypic analysis, PBMCs were stained with fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD4 (RPA-T4), CD8 (SCF121Thy2D3), CD45RA (HI100), CD19 (HIB19), CD14 (MSE2), CCR7 (150503), and CD127 (HIL-7R-M21). For the analysis of the frequencies of naïve T cells (CD45RA+CCR7+), central memory T cells (CM; CD45RA-CCR7+), effector memory T cells (EM; CD45RA-CCR7-), and effector memory RA (EMRA; CD45RA+CCR7-), B cells and monocytes, PBMCs were stained with fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD4 (RPA-T4), CD8 (SCF121Thy2D3), CD45RA (HI100), CD19 (HIB19), CD14 (MSE2), CCR7 (150503), and CD127 (HIL-7R-M21). The stained cells were acquired with a BD Fortessa and analyzed with FlowJo software (TreeStar).

CMV-Seropositivity Measurements

Anti-CMV IgG titers were determined in frozen sera by commercially available enzyme-linked immunosorbent assay (ELISA) (Genway Biotech Inc. San Diego, CA) with an interassay coefficient of variance of 5.2%. A titer of 1.2 ELISA Units/ml or greater in a sample was predetermined by the manufacturer as CMV-seropositive.

Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq)

Approximately 50,000 cells were harvested and spun down at 500 g for 5 min, 4°C. Cells were washed once with 50 μL of cold 1x PBS buffer and pelleted by centrifugation at 500 g for 5 min, 4°C. Pellets were pipetted gently to resuspend the cell pellet in 150

μL of cold lysis buffer (50 mM Tris-HCl, pH 8.0, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% TritonX-100). The cell lysate was incubated on ice for 15 min. The lysate was centrifuged at 2500x g for 5 min, at 4°C and the supernatant was discarded.

Transposition reaction: Ice-cold transposition reaction mix was added to the cell pellet. Transposition reaction mix consisted of 25 μL 2x TD Buffer (Illumina Cat #FC-121-1030), 2.5μL Tn5 transposase (Illumina Cat #FC-121-1030), and 22.5 μL nuclease free water in a 50 μL total volume for reaction. The pelleted nuclei were resuspend in the transposition reaction mix and incubated at 37°C for 30 min. Immediately following transposition, DNA from the reaction was purified using a Qiagen MinElute Kit. Transposed DNA was eluted in 10 μL elution buffer (10mM Tris buffer, pH 8). The purified DNA was stored at -20°C.

PCR amplification: To amplify transposed DNA fragments, 10 μL of transposed DNA, 9.7 μL nuclease free water, 2.5 μL Nextera PCR Primer 1, 2.5 μL Nextera PCR Primer 2 (Barcode primer, table 1), 0.3 μL 100x SYBR Green I (Invitrogen Cat #S-7563), and 25 μL NEBNext High-Fidelity 2x PCR Master Mix (New England Bio Labs Cat #M0541) in a total 50 μL reaction volume were combined in a PCR tube. The PCR cycle as performed as follows: 72°C, 5 min, 98°C, 30 sec, 98°C, 10 sec, 63°C, 30 sec, 72°C, 1 min, repeat steps 3-5, 4x, and hold at 4°C. In order to reduce GC and size bias in the PCR reaction, the reaction was monitored using qPCR to stop amplification prior to saturation. The qPCR side reaction was performed as follows: 5 μL of PCR amplified DNA, 4.44 μL nuclease free water, 0.25 μL Nextera PCR Primer 1, 0.25 μL Nextera PCR

Primer 2, 0.06 μL 100x SYBR Green I, and 5 μL NEBNext High-Fidelity 2x PCR Master Mix, in a total reaction volume of 15μL. The qPCR reaction was conducted as follows: 98°C for 30 sec, 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min, repeat steps 3-5, 19x, and hold at 4°C.

To determine the additional number of cycles needed for the remaining 45 μL PCR reaction, linear Rn vs. Cycle was plotted and the RF threshold was set at 5000. The corresponding cycle number was calculated at a maximum fluorescent intensity for each individual sample. Then, the remaining 45 μL PCR reaction was run to the correct cycle number. The cycle was conducted as follows: 98°C, 30 sec, 98°C, 10 sec, 63°C, 30 sec, 72°C, 1 min, repeat steps 3-5, x times, and hold at 4°C. The amplified library was purified using Qiagen PCR Cleanup Kit and eluted in 20 μL elution buffer (10mM Tris Buffer, pH 8).

Library quantification. Real-time PCR based methods were used to quantify our ATAC-seq libraries. We found that other methods, such as Bioanalyzer and Qubit, gave misleading and inaccurate results due to the large distribution of insertion sizes. We used the commercially available library quantitation kit, KAPA Library Quant Kit for Illumina Sequencing Platforms (KAPABiosystems) for analysis.

ATAC-seq libraries were sequenced by using NextSeq series and HiSeq series from Illumina Sequencing Platforms. Pair-end sequencing was performed on an Illumina HiSeq 2500 with read lengths of 150 base pairs to a minimum depth of 30 million reads per sample.

ATAC-seq Data Pre-Processing

ATAC-seq sequences were quality-filtered using trimmomatic (*16*), and trimmed reads were mapped to the GRCh37 (hg19) human reference sequence using bwa-mem (*17*). After alignment, technical replicates were merged and all further analyses were carried out on these merged data. For peak calling, MACS2 (*18*) was used with no-model, 100bp shift, 200bp extension, and broad peaks options. Only peaks called with a peak score (q-value) of 1% or better were kept from each sample, and the selected peaks were merged into a consensus peak set using Bedtools *multiinter* tool.

Ad1_noMX:	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

Figure 6: Oligo designs. A list of ATAC-seq oligos used for PCR.

Only those consensus peaks overlapping 20 short reads or more in at least one sample were selected for further analyses. Finally, we excluded peaks overlapping blacklisted regions as defined by ENCODE mappability criteria, downloaded from *http://hgdownload.cse.ucsc.edu/ goldenpath/hg19/encodeDCC/wgEncodeMapability/* on July 2015.

An additional quality-control step was developed to filter out samples with a consistently poor signal, consisting of an algorithm to discover and characterize a series of relatively invariant *benchmark peaks*, defined as a set of peaks expected to be called in all samples. Samples that consistently miss calls for a significant portion of these benchmark peaks are flagged as having poor quality. A benchmark peak is defined based on three criteria, namely (1) that it remains approximately invariant between the two groups of interest (i.e., young and old samples), (2) that it captures a substantial number of reads, and (3) that it is called in most samples. For each peak, the absolute value of the log of the ratio of healthy old to healthy young mean normalized read counts (log fold change, logFC) was used to assess the first criteria, whereas the maximum read count over all samples (maxCt) is used to assess the second one. In this study, a peak was considered apt for benchmarking when (1) its absolute logFC was in the bottom decile of the distribution over all peaks, (2) its maxCt was in the top decile of the distribution over all peaks, and (3) the peak was called in at least 90% of the samples. Using these parameters, 1,491 (out of 238,004) peaks were selected as benchmark; only samples for which at least 95% of these peaks were called were selected for analyses, which excluded 10 samples from further.

We examined the effects of each of these parameter choices and found that the same samples were consistently chosen as poor quality for a wide range of values chosen to assess the benchmark criteria.

RNA-seq Library Generation and pre-processing

Total RNA was isolated from PBMCs using the Qiagen RNeasy (Qiagen) or Arcturus PicoPure (Life Technologies) kit following manufacturer's protocol. DNase treatment was additionally performed during RNA isolation using the RNase-free DNase set (Qiagen). RNA quality was checked using an Agilent 2100 Expert bioanalyzer (Agilent Technologies). RNA quality was reported as a score from 1 to 10, samples falling below threshold of 8.0 being omitted from the study. cDNA libraries were prepared using either the TruSeq Stranded Total RNA LT Sample Prep Kit with Ribo-Zero Gold (Illumina) or KAPA Stranded mRNA-Seq Library Prep kit (KAPA Biosytems) according to the manufacturer's instructions using 100ng or 500ng of total RNA. Final libraries were analyzed on a Bioanalyzer DNA 1000 chip (Agilent Technologies). Paired-end sequencing (2x100bp) of stranded total RNA libraries was carried out in either Illumina NextSeq500 using v2 sequencing reagents or the HiSeq2500 using SBS v3 sequencing reagents.

Quality control (QC) of the raw sequencing data was performed using the FASTQC tool, which computes read quality using summary of per-base quality defined using the probability of an incorrect base call (*19*). According to our quality criteria, reads with more than 30% of their nucleotides with a Phred score under 30 are removed, whereas samples with more than 20% of such low quality reads are dropped from analyses. None of the samples used in this study were dropped after QC. Reads from samples that pass

the quality criteria were quality-trimmed and filtered using trimmomatic (*16*). High-quality reads were then used to estimate transcript abundance using RSEM (*20*). Finally, to minimize the interference of non-messenger RNA in our data, estimate read counts were re-normalized to include only protein-coding genes. Table 2 summarizes the depth and alignment rate of our PBMC RNA-seq samples.

Differential Analysis

To identify differentially open chromatin regions from ATAC-seq and differentially expressed genes from RNA-seq data, the R package edgeR was used to fit a generalized linear model (GLM) to test for the effect of aging between healthy young and healthy old samples. In addition to age group (old vs. young), our models included sex and the season in which the sample was collected (summer vs. winter) as covariates (*21*), since it was determined, using Principal Variance Component Analysis (PVCA, (*22*)), that these factors account for a sizeable fraction of the variance in read counts. Furthermore, we used Surrogate Variable Analysis (SVA (*23*)) to capture unknown sources of variation (e.g., batch effects, subject-level heterogeneity) statistically independent from age group assignments. Using the built-in permutation-based procedure in the R package SVA, we choose to retain three SVs to include as covariates in the GLM model for PBMC ATAC-seq and RNA-seq data analyses (*24*). Within GLM models, a negative binomial link function was used, including both genome-wide and peak-specific dispersion parameters, estimated using edgeR's "common," "trended," and "tagwise" dispersion components, calculated using a robust estimation option. Benjamini-Hochberg P-value correction was used to select differentially open peaks at a False Discovery Rate (FDR) of 5%. To generate a set of model-adjusted peak estimates of chromatin accessibility (i.e., sex-,

season-, and SV-adjusted) for downstream analyses and visualization, we used edgeR to fit a "null" model excluding the age group factor, and then subtracted the resulting fitted values from this model from the original TMM-normalized reads.

An equivalent approach was used to analyze the effects of CMV seropositivity and seasonal variation (i.e., winter vs. summer-acquired samples) in PBMC data. For CMV analysis, the subset of samples for which this information was available (i.e., N=21, 12 HY and 9 HO) was fit to a model including a sex as a factor and CMV status (positive, negative) as a blocking factor. In this analysis, the season factor was not taken into consideration since all subjects for whom CMV status was available were collected in the same season. For seasonal analysis, we used season (summer, winter) as a blocking factor. In both analyses, we tested both separately and jointly for the significance of age group by CMV status or season. In addition, we fitted the converse models (CMV status or aging nested within age group) to test for and calculate fold change estimates for CMV+/CMV- and winter/summer stratified by age group.

Peak Annotation and Downstream Analyses

Multiple data sources were used to annotate ATAC-seq peaks with regard to functional and positional information. HOMER (*25*) was used to annotate peaks as promoter (i.e., within 1 kb of known TSS), intergenic, intronic, and other positional categories. A simplified version of 18-state ChromHMM-derived chromatin states obtained from Roadmap Epigenomics data for PBMC(*26*) was used for functional annotations. First, we intersected the Roadmap-generated states with our set of consensus peaks, and solved conflicting cases where multiple chromatin states overlap the same ATAC-seq peak so that each peak was assigned a single annotation, according to the following priority rules:

Active TSS > Active Enhancer 1 > Active Enhancer 2 > Genic Enhancer 1 > Genic Enhancer 2 > Weak Enhancer > Strong Transcription > Flanking Active TSS > Flanking Upstream Active TSS > Flanking Downstream Active TSS > Weak Transcription > Bivalent Poised TSS > Bivalent Enhancer > Weakly Repressed PolyComb > Repressed Polycomb > ZNF Genes and Repeats > Heterochromatin > Quiescent/Low signal. Finally, to facilitate interpretation and visualization, we simplified the set of 18 chromatin states to a scheme with 6 pooled meta-states containing: (1) active, flanking, and bivalent TSS states (2) active, weak and bivalent enhancer states, (3) both weak and strong PolyComb states, (4) both weak and strong transcription states, (5) the quiescent state, and (6) all others states combined together.

ATAC-seq peaks were also annotated using gene sets provided by curated immune function-related co-expression modules (*27*). We used these annotations to test for enrichment of modules in a variety of gene sets of interest, such as genes associated to closing/opening peaks. We assessed enrichment using the hypergeometric test, with a background defined by the set of genes that are expressed, as determined by RNA-seq data, or potentially expressed, as given by promoter accessibility, in the appropriate cell type. In addition, we summarized the representation of GO terms among gene annotations for all peaks, after solving for multiple GO annotations for the same gene by prioritizing terms according to the order: Immunity > Metabolic > Transcription, Translation > Migration > Mitochondria > Axon > Development.

Further functional enrichment analyses were carried out using ClueGO (*28*) to test for overrepresentation of GO:Immune System Process terms using GO term fusion option and Wikipathways pathways (*29*) among genes associated with differentially open peaks.

In addition to testing for enriched gene sets, ClueGO combines GO terms and pathways into functionally relevant meta-sets based on the rate of shared genes among terms, allowing for an efficient assessment of enriched categories, as well as their potential interactions, as inferred from sets of shared genes. We applied these methods separately to peaks significantly closing and opening between age groups to investigate the degree to which these two sets of peaks are associated to unique signatures. We only listed terms that are significant at p-value 0.05 after Bonferroni step down correction. In addition, we used ClueGO to annotate the aforementioned immunological co-expression modules that were originally associated to unknown function. Visualization of signaling pathways were generated ClueGO and PathVisio (*30*) tools.

Congruence between Chromatin Accessibility and Transcription Data

Gene expression (mRNA-seq, see above) data was generated for a subset of subjects with ATAC-seq profiles (n=39, 24 HY and 15 HO). These data were normalized to protein-coding transcripts, and annotated to ENSEMBL GRCh37 gene symbols. Genes for which at least three normalized reads per million were obtained in at least two samples were considered as expressed, all others removed prior to analysis. This resulted in a total estimate of 11,311 expressed genes in PBMCs. We built a data set comprising paired ATAC-seq and RNA-seq samples by matching promoter peaks to nearest gene (TSS) annotations. First, we retrieved the complete list of refSeq TSS coordinates from the hg19 genome reference (n=34,783), and defined promoters as the regions within 1000 bp flanks of each TSS. The final set of promoters was defined by merging overlapping flanked TSS regions annotated to the same gene (n=34,700). We then selected ATAC-seq peaks overlapping these promoters and annotated them to the corresponding gene.

Only the peak closest to the TSS was kept. Finally, the resulting data set was filtered as to only include promoter peaks for genes that were transcribed, as defined above. Whenever multiple expressed genes were matched to the same promoter peak, all of them were retained for analysis.

To study the concordance between promoter accessibility and gene expression, we subdivided the space defined by aging-related fold changes derived from ATAC-seq and RNA-seq data into gene sets defined by the direction and magnitude of change along both dimensions, such as genes with both up-regulated expression and increasing accessibility in elderly subjects, or genes for which expression is up-regulated but accessibility remains unchanged with aging. In order to capture enough genes to enable functional enrichment analysis of these gene sets, fold changes between healthy old and young subjects for matching promoter peaks and transcripts were estimated empirically as the difference between the mean normalized values of each group, and plotted against each other (see Figure 14A). For each gene set, we tested for enrichment in immune modules and Wikipathways pathways and compared these results to gene sets constructed based solely on significant differential accessibility or expression as determined using GLM as previously described. Specifically, we defined a gene or promoter as being significantly "up" or "down" if the empirical log fold change of the HO mean relative to HY mean was above or below zero, respectively, and if the adjusted empirical p-value < 0.01 for that gene. Empirical p-values were computed by randomly permuting the HO and HY sample labels 1,000 times for each promoter peak and gene. Genes for which $p < 0.01$ were considered significantly different between aging groups, whereas all others were considered to have "stable" expression and/or accessibility relative to aging. Here we

focus on a subset of the combined accessibility expression gene sets generated by this method, namely, (1) genes with both increased or (2) both decreased promoter accessibility and expression with aging, and (3) genes with increased or (4) decreased promoter accessibility but stable aging-related expression.

Transcription Factor (TF) Motif and Footprinting Analysis

ATAC-seq data from PBMCs and T cells were scanned for TF footprints using the PIQ algorithm (*31*). This method integrates genome-wide TF motifs (i.e., position weight matrices or PWMs) with chromatin accessibility estimates profiled at base pair resolution to generate a list of possible footprint matches for a motif. The method also produces a probability estimate for each footprint's reproducibility, termed "purity score". Here, we compiled a set of 1,273 distinct motifs comprising the curated (CORE) list available in the JASPAR 2016 database (n=466, http://jaspar.genereg.net) in addition to the complete set of HT-SELEX motifs made available in (*32*) (n=819). Altogether, these motifs represent binding sites for 381 distinct TF Prior to footprint calling, we merged samples belonging to the same cell type and age group to maximize our ability to find highly reproducible footprints. In addition, we used SAMtools v. 0.1.19 (*17*) to randomly subsample aligned reads from each merged data set to approximately match the mapped library depth of the least deeply sequenced sample, i.e., 113 Mb. This normalization step is included to minimize the impact of the high correlation observed between library depth and footprint purity scores. Only footprints with a purity score of 90% or more were retained for further analysis. Finally, footprint calls were further filtered to include in analyses only those associated to TFs determined that are expressed in immune cells.

To examine the regulatory landscape of IL7R, a potential aging biomarker, we focused on footprints called on the promoter region (+/-1 kb from TSS) of this gene separately by age group and cell type. To complement this set of footprints, we also carried out *de novo* motif discovery using HOMER by searching for motifs enriched in peaks annotated to IL7R relative to all peaks in PBMCs and T cell subsets. Each enriched motif was annotated to the best fitting known TF, as found by HOMER, with the added requirement that the annotated TF should be expressed in the appropriate cell type. We then used PIQ to call footprints of the enriched motifs, and combined those overlapping IL7R promoter with the previously selected footprints. Finally, in addition to footprint and motif enrichment analyses, known TF motifs were retrieved for the region around IL7R TSS (-10kb upstream, +1kb downstream) using MotifMap tool (*33*) at a 20% FDR.

CHAPTER FOUR

Results

Generating Epigenomic Datasets in Human Immune Cells

ATAC-seq is a relatively novel technology that enables generating open chromatin profiles from small cell numbers (as few as single cells) compared to other techniques. Moreover, ATAC-seq technique is also easier to manipulate than traditional DNA sequencing methods, such as DNase-seq. The key element of this novel technology is the use of hyperactive Tn5 transposase that can simultaneously fragment and tag a genome with adaptors for high-throughput DNA sequencing, However, ATAC-seq has also some drawbacks such as contamination of ATAC-seq libraries with mitochondrial DNA (mtDNA) and generation of excessive number of reads that can affect the computational analysis, in particular, alignment of sequences to the genome, Therefore, it is very critical to solve these issues in order to increase the power of computational analysis before we build our own epigenomic datasets and extract more accurate information from epigenetics of ageing.

Removal of mtDNA form Cell Lysates

In the preliminary experiments, we noticed that the majority of reads were aligned to mtDNA, due to the ineffective lysis method in the nuclear purification procedure. In order to optimize the original ATAC-seq protocol, we aimed to reduce the amount of mtDNA from the starting material.

The structure of mtDNA is a small circular chromosome, and is simpler than nuclear chromosomes, which is supercoiling structure. Its structure permits easier access of Tn5 transposase that cleaves mtDNA more readily than nuclear DNA, resulting in generation of more fragments of mtDNA than fragments of nuclear DNA. Data in Table 5 shows a high mtDNA duplicate rate among many of the libraries. Indeed, we found that more reads are specifically aligned to the mitochondrial rather than nuclear chromosomes. For example, data from two donors showed that of the total of 27M reads in an HO individual, 17.3M reads are occupied by mtDNA, which represents almost two-thirds of total number of reads (Figure 7). Likewise, the HY donor showed a similar pattern. Many mtDNA fragments were enriched around 100 to 200 bps region.

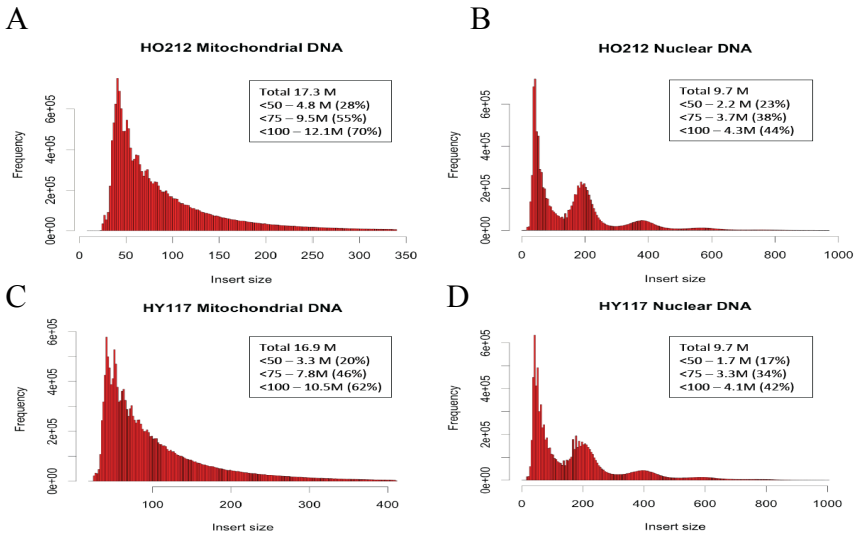


Figure 7. The proportion of nuclear DNA and mtDNA read counts in DNA sequencing. (A) and (B) The distribution of read counts in healthy elderly (HO). (C) and (D) The distribution of read counts in healthy young adult (HY).

Currently, samples are sequenced with millions of reads to compensate, for the reads aligning to the mtDNA, increasing the cost of experiments. We tested several different

methods to solve this issue including i. Targeted DNA size selection for Next-Gen sequencing using BluePippin system, ii. Avidin-biotin-mediated depletion of mtDNA , iii. Customized nuclear purification buffers, iv. collaboration with biotech company for mtDNA depletion.

First, the BluePippin systems use pre-cast and disposable agarose gel cassettes. DNA fractions are collected by electro-elution into a buffer-filled well using a branched channel configuration with switching electrodes (*48*). The timing of switching is determined by measuring the rate of DNA migration with optical detection of labeled markers. By taking the advantages of BluePippin system, we selected 7 ATAC-seq libraries enriched with high amount of mtDNA (4HY and 3HO), and tested whether mtDNA would be removed by selecting adequate size of DNA, from 120 bps to 250 bps and 250 bps to 500 bps. Results are indicated in Table 1, there are no differences after DNA size selection. Moreover, we might lose useful and interesting DNA information, DNA size smaller than 120 bps, if we apply this methodology to avoid mtDNA issue.

Table 1. BluePippin size selection result

Sample	Mapped pairs	chrM Mapped	Size selection mtDNA%	Original mtDNA%
HY114	983456	626150	63.39	63.9
HO212	1223965	639895	52.07	66
HY115	874045	515332	58.66	60.2
HY ND9	1523389	882894	57.43	69.5
HO215	1053904	541373	51.13	58.6
HY ND10	1518338	758000	49.74	59.2
HO213	489494	203934	41.17	54.3

Secondly, the interaction of biotin and avidin or streptavidin has been exploited for use in many protein and nucleic acid detection and purification methods (*49*). To utilize this

highly sensitive assay, we customized and designed specific avidin-biotin-mediated depletion to mtDNA sequences. We did not harvest any mtDNA products after avidin-biotin incubation.

Thirdly, we used the commercial cell lysis buffer provided by Sigma-Aldrich to separate mitochondria organelle from nuclei (*50*). The result indicated that (table 2) the average of mtDNA percentage using commercial lysis buffer and original lysis buffer were 29.2% and 34.7%. It did not provide a good outcome; however, using optimized nuclear lysis buffer to separate mitochondria organelle and nuclei would not interfere and loss any information.

Table 2. The commercial buffer and original buffer comparison

Sample	Mapped_pairs	chrM_Mapped	Percent_chrM
Commercial lysis buffer	1881395	490520	25.35009352
Commercial lysis buffer	1556876	534503	33.24639765
Original lysis buffer	1275844	525083	40.92851737
Original lysis buffer	1478263	423440	28.508642

Last, we collaborated with a biotech company for mtDNA depletion (*51*). We used the me290, a melanoma cancer cell line that contained high amount of mtDNA, and 5 PBMCs samples to test mtDNA depletion method provided by the biotech company. As result shown in table 3, the mtDNA was removed from samples successfully (me290 from 86% to 21% and PBMCs samples from 31.6% down to 4.2% on average). Even though mtDNA was successfully removed, it is still time-consuming issue. It took a long time to remove mtDNA in few samples, and outsourcing this step would generate another additional expense.

Table 3. The result from biotech company collaboration

Sample	Before depletion mtDNA(%)	After depletion mtDNA(%)
Me290	0.86	0.21
ND16_fnPBMC_1	0.24	0.03
ND16_fnPBMC_2	0.29	0.04
ND16_PBMC_1	0.29	0.03
ND16_PBMC_2	0.38	0.06
ND16_PBMC_3	0.38	0.05

Therefore, best results were obtained using optimized nuclear lysis buffer (*52*). To this end, we compared optimized new cell lysis buffer to the original cell lysis buffer, recipe provided by Greenleaf lab, using total 300 samples. As a result, the mtDNA duplicate rate dropped significantly from 56% to 26% (Figure 8A). In addition, the number of PCR cycles required for ATAC-seq library preparation also decreased from 6.25 cycles to 4.3 cycles (Figure 8B), a difference of 2 cycles, demonstrating that the new homemade cell lysis buffer made library preparation more efficient. The major difference in the new buffer is the addition of Triton-X100, which is a key component in the buffer to separate mitochondria organelle.

Furthermore, large amounts of raw reads were acquired after Next-seq sequencing. Our analysis pipeline filtered raw reads, eventually, retaining useful reads that could be aligned with the genome for further analysis. We noticed that 66.4% reads were useless using the original cell lysis buffer. In contrast, the protocol using the new homemade lysis buffer produced almost two-fold as many useful reads (i.e. from 33.6% to 59% more reads) (Figure 8C).

In sum, the new cell lysis buffer resulted in increased nuclear purity, lowered mtDNA duplicates (a 30% difference) and provided more reads which increased the power of analysis.

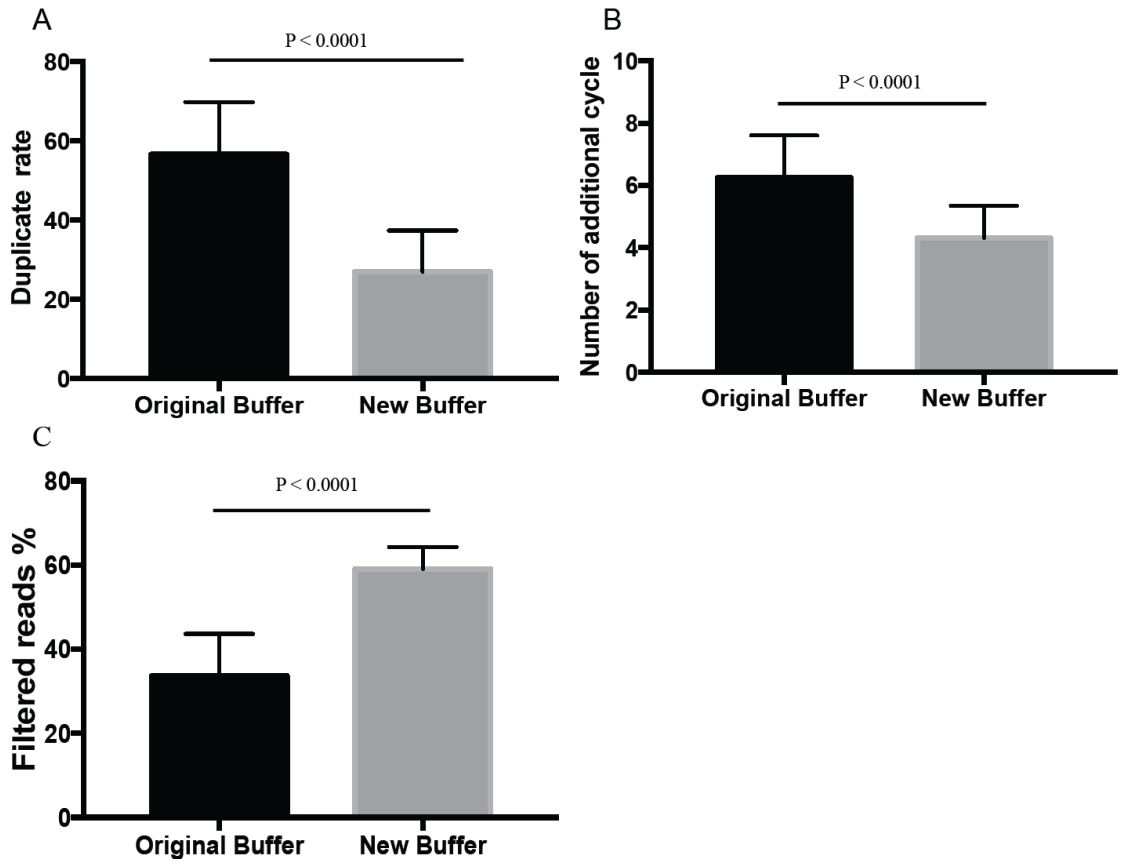


Figure 8. New homemade buffer improved ATAC-seq quality. (A) Duplicate rate decreases from 56% to 26%. (B) PCR cycle number decreases from 6.25 to 4.3 cycles in ATAC-seq library preparation. (C) After sequencing, filtered reads increase from 33.6% to 59%. P<0.0001 compared with original buffer group, as analyzed by two-tailed Student's t-test.

Taken together, we had tried five different methods, BluePipin size selection, biotin-avidin interaction, commercial lysis buffer, mtDNA blocking project, and new homemade lysis buffer to remove mtDNA. The conclusion is summarized as below (Table 4).

Table 4. Conclusion of mtDNA depletion methods

Method	Pros	Cons	Result
BluePippin system	Specific size selection	Undistinguished nuclear DNA and mtDNA	Negative
Biotin-avidin interaction	High affinity interaction	Unpredictable mtDNA fragments	Negative
Commercial lysis buffer	Time efficiency		Negative
mtDNA blocking project	Outsource	Time-consuming; pricy	Positive

Table 4—continued

Method	Pros	Cons	Result
New homemade lysis buffer	Time efficiency; data comparable		Positive

Improving the Number of Useful Reads

In our preliminary experiments, we found that after sequencing, many reads were filtered out as they were considered redundant primer dimers. As shown in Table 5, which indicates the general sequencing quality control information, around 60% of the raw reads were removed and only 40% reads could be used. As a result, the power of computational analysis was decreased not only due to mtDNA contamination interfering with the final readout, as shown above, but also due to the small portion of reads considered as useful reads (Table 5).

Table 5. The preliminary sequencing result from healthy donors, including young adults and elderly adults, were collected from UCHC aging center.

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
HO210-PBMC-50-1	37786837	15036982	39.79423311	50.099
HO211-PBMC-50-1	33694743	7693138	22.83186431	71.170
HO211-PBMC-50-2	35678556	11642061	32.63041531	58.791
VHY402_D0_PBMC_50_1	28235978	9644735	34.15760913	57.582
VHY402_D0_PBMC_50_2	30061027	12371955	41.15612883	49.707
HM501_PBMC_50_1	31849685	12083702	37.93978496	47.922
HM501_PBMC_50_2	31999636	14777007	46.17867216	38.117
HM502_PBMC_50_1	35634255	12246893	34.36831498	52.642
HM502_PBMC_50_2	29372000	11340409	38.60959077	48.767
HM503_PBMC_50_1	39192554	15365795	39.20590375	47.807
HM503_PBMC_50_2	30669338	12447092	40.58480819	44.936
ND9_PBMC_50_1	31817566	7111806	22.35182289	69.485

Table 5-- continued

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
ND9_PBMC_50_2	46424285	14694929	31.65353866	55.876
ND10_PBMC_50_1	33011817	10274516	31.12375184	58.240
ND10_PBMC_50_2	31832200	9671452	30.38260629	59.204
ND11_PBMC_50_1	26701073	10316367	38.63652596	46.819
ND11_PBMC_50_2	28353103	15478950	54.59349546	25.701
HO212_PBMC_50_1	51172337	13373090	26.13343612	64.437
HO212_PBMC_50_2	47107490	11948701	25.36475834	65.985
HY113_PBMC_50_1	30385486	10848445	35.70272004	51.790
HY113_PBMC_50_2	32918745	17395944	52.84510087	29.781
HY114_PBMC_50_1	33400809	9203209	27.55385057	63.856
HY114_PBMC_50_2	35434327	8103889	22.87016485	70.592
HO213_PBMC_50_1	28541117	5941472	20.81723711	71.485
HO213_PBMC_50_2	25767688	8508487	33.01998612	54.267
VHY402D1_PBMC_50_1	30034699	9547804	31.78924483	56.936
VHY402D1_PBMC_50_2	37558000	11692666	31.13229139	58.171
VHY402D7_PBMC_50_1	40597149	13340302	32.8601942	54.292
VHY402D7_PBMC_50_2	37923417	10139182	26.73593996	64.028
VHY402D28_PBMC_50_1	38887131	9353477	24.05288526	67.752
VHY402D28_PBMC_50_2	41562650	10664809	25.65959822	64.763
HY112_PBMC_50_1	35807400	16168626	45.15442618	41.404
HY112_PBMC_50_2	35099328	13238755	37.71797283	50.931
F303_PBMC_50_1	39603844	7278155	18.37739539	72.958
F303_PBMC_50_2	46340592	10336225	22.30490495	68.181
ND8_PBMC_50_1	36922628	8757347	23.71810316	68.002
ND8_PBMC_50_2	36922811	8612576	23.32589466	69.188
HO214_PBMC_50_1	33044657	8872071	26.84873079	63.959

Table 5--continued

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
HO214_PBMC_50_2	26064895	9628367	36.93998	51.884
HM504_PBMC_50_1	23460422	7531775	32.10417528	58.849
HM504_PBMC_50_2	23363486	7685049	32.89341753	58.828
HM505_PBMC_50_1	27232230	6621388	24.31452731	67.916
HM505_PBMC_50_2	22807220	6551444	28.72530716	63.591
HM506_PBMC_50_1	19095987	4796213	25.11633989	64.869
HM506_PBMC_50_2	24156274	8911492	36.89100397	53.756
HM507_PBMC_50_1	24481608	11897326	48.59699575	38.588
HO215_PBMC_50_1	24902449	8103249	32.53996826	58.605
HO215_PBMC_50_2	24539146	10675546	43.5041464	44.649
HO216_PBMC_50_1	24216437	6889293	28.44883002	63.061
HO216_PBMC_50_2	20499448	6712211	32.74337436	57.863
HY115_PBMC_50_1	23997273	7513017	31.30779485	60.216
HY115_PBMC_50_2	24401832	8882970	36.40288155	53.158
HY116_PBMC_50_1	23404746	11333246	48.42285407	37.959
HY116_PBMC_50_2	25837511	13592401	52.60723837	31.540
HM402_PBMC_50_1	33334606	6359476	19.07769961	73.658
HM403_PBMC_50_1	28580940	9386665	32.84239427	55.575
HM403_PBMC_50_2	24333909	6507547	26.74271117	64.833
HM404_PBMC_50_1	26235365	6621259	25.23791455	66.855
HM404_PBMC_50_2	23490515	4774021	20.32318576	72.704
HM508_PBMC_50_1	23442040	5569029	23.75658859	68.959
HM508_PBMC_50_2	25037853	5249926	20.967956	71.934
HY117_PBMC_50_1	33705027	8076663	23.96278454	68.813
HY117_PBMC_50_1	35099328	13238755	37.71797283	50.931
HY119_PBMC_50_1	25073261	8822561	35.18713023	52.426

Table 5--continued

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
HY119_PBMC_50_1	20574888	7291086	35.43681987	52.622
HY125_PBMC_50_1	22034043	7304625	33.15154191	56.208
HY125_PBMC_50_1	20533575	7958057	38.75631496	49.478
HY126_PBMC_50_1	21832155	7033569	32.21655856	57.183
HY126_PBMC_50_1	25880675	7654752	29.57709565	60.076
HY123-PBMC-50-1	37481447	20639973	55.0671723	31.898
HY123-PBMC-50-2	41061078	21553864	52.49220198	36.084
HY123-FROZEN-PBMC-50-1	37594892	21677554	57.66090244	30.587
HY123-FROZEN-PBMC-50-2	37232313	19311785	51.86834619	36.479
HY127-PBMC-50-1	41480758	22961382	55.35429705	30.451
HY127-PBMC-50-2	36830194	19060277	51.75176921	36.103
HY127-FROZEN-PBMC-50-1	42141867	21743349	51.59559969	36.830
HY127-FROZEN-PBMC-50-2	43299314	21088874	48.70486863	39.535
HY121_PBMC_50_1	30138080	10807638	35.8604065	52.388
HY121_PBMC_50_2	25806707	11016834	42.68980928	44.275
HY120_PBMC_50_1	29180795	10259435	35.15817509	51.956
HY120_PBMC_50_2	26310013	8371101	31.81716786	58.636
HO218_PBMC_50_1	25607626	8232215	32.1475134	58.040
HO218_PBMC_50_2	25883255	8221488	31.76373296	57.120
VHY402_D60_PBMC_50_1	25863745	6122570	23.6724032	68.138
VHY402_D60_PBMC_50_2	32388940	8699813	26.86044372	63.566
HO219_PBMC_50_1	33466836	15088614	45.08527188	39.697
HO219_PBMC_50_2	28246906	8338177	29.51890377	59.925
HY128_PBMC_50_1	25006849	13373418	53.47902089	29.792
HY128_PBMC_50_2	26735497	12062010	45.11608668	41.273
HM411_PBMC_50_1	22200689	11096229	49.98146229	36.000

Table 5--continued

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
HM411_PBMC_50_2	26623753	14646791	55.01399821	29.065
HM415_PBMC_50_1	25992273	11160672	42.93842251	42.206
HM415_PBMC_50_2	24554754	11362280	46.27323898	39.173
HM511_PBMC_50_1	31912618	8289858	25.97674061	64.675
HM511_PBMC_50_2	25858478	8213191	31.76208205	57.781
HO217_PBMC_50_1	30317856	11086001	36.5659135	51.639
HO217_PBMC_50_2	30600283	11679840	38.16905876	48.005
HM417_PBMC_50_1	27709743	9541525	34.43382712	54.251
HM417_PBMC_50_2	31854559	8591071	26.96967489	63.325
HO221_PBMC_50_1	31122483	5612646	18.03405596	76.034
HO221_PBMC_50_2	32623623	12768094	39.13757218	46.098
HM416_PBMC_50_1	26582060	8299063	31.22054122	56.014
HM416_PBMC_50_2	24591437	7433326	30.22729416	58.909
ND12_PBMC_50_1	28430928	8238663	28.97781951	61.962
ND12_PBMC_50_2	28391143	7437282	26.19578226	65.381
ND13_PBMC_50_1	32043492	5643908	17.61327386	75.865
ND13_PBMC_50_2	27550722	7051935	25.59618946	65.667
HM401_PBMC_50_1	40364624	14833770	36.74943188	50.026
HM401_PBMC_50_2	40990945	15294963	37.31302852	50.124
HM407_PBMC_50_1	39016651	7047614	18.06309311	74.880
HM407_PBMC_50_2	23565753	7134839	30.27630392	59.626
HM408_PBMC_50_1	26116800	5008095	19.17576043	73.969
HM408_PBMC_50_2	22465269	6946650	30.92173078	57.553
HM409_PBMC_50_1	23649161	5960118	25.20223868	66.194
HM409_PBMC_50_2	25051345	5807200	23.18119047	68.015
HM509_PBMC_50_1	28816457	6833381	23.71346693	68.021

Table 5--continued

Sample	Raw Reads	Aligned Reads	Filtered reads %	Percent of Duplication %
HM509_PBMC_50_2	26903188	6453857	23.98919043	67.895
HM510_PBMC_50_1	29398267	7428498	25.26848947	65.240
HM510_PBMC_50_2	26489867	7417120	27.99983858	61.993
HY129_PBMC_50_1	25093919	9763477	38.90774096	48.288
HY129_PBMC_50_2	32713650	6401361	19.56785929	73.587

While using the original protocol, we found that the library concentration around the 64 base pair (bps) region was too high and unknown (Figure 9A). Usually, the average size of fragments cut by Tn5 transposase is around 300 bps. We concluded that the index primer concentration was too high for the library preparation thus contributing to a large number of reads regarded as useless during the filtration process and therefore could not be aligned to the genome. To solve this, many labs use Ampure beads to remove redundant index primers contained in the libraries. However, the Ampure beads can potentially also remove useful genomic information from libraries. We, therefore, decided to test the effect of lowering 4 to 6-fold the concentration of index primers (Figure 9B-D). Thus, using four to six times lower concentration of index primer decreased the concentration of primer dimers near the 64 bps region. Moreover, the library complexity using the lower concentration of index primers was not affected.

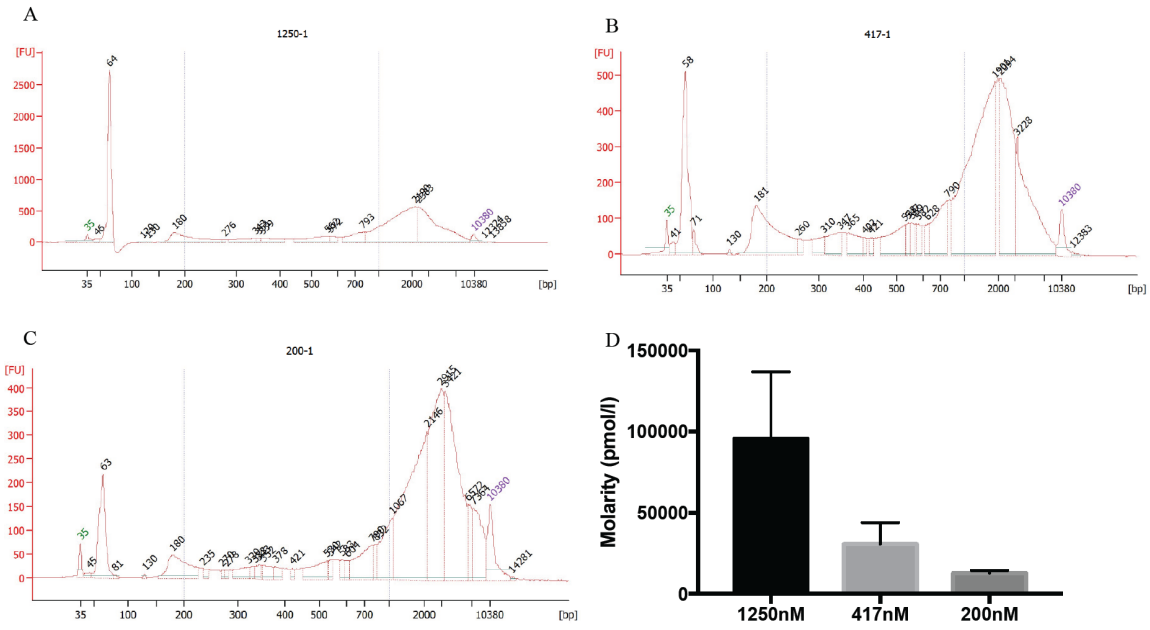


Figure 9. ATAC-seq library QC results from bioanalyzer. Shown are the profiles of ATAC-seq libraries prepared using (A) 1250nM index primer, according to original protocol (B) 417nM index primer, 3 fold less than the original concentration and (C) 200nM index primer, 6-fold less than original concentration. (D) Comparison of three different concentrations of index primer.

We next prepared ATAC-seq libraries using the lower concentration of index primers. We expected that more reads could be used to sequence the genome rather than sequencing the redundant index primers. Indeed, the control group, which contains the original index primer concentration, gained more than 3 million reads, (12,046,897 vs. to 9,427,182 from a total of 21,474,079 reads) than the experimental group using optimized index primer concentration (Table 6). We hypothesized that by using low concentration of index primers we will increase the number of reads and improve library quantification in the control group. However, we noticed in fact with the small size of library, it was difficult to conclude. There are additional factors that could potentially affect sequencing results, such as library quantification or sequencer programing. Table 2 shows that there is no difference in reads filtered, read genome alignment, and percent of duplication. Additional troubleshooting is needed to optimize the ATAC-seq protocol.

Table 6: Illumina Mi-seq raw data. Six ATAC-seq libraries were made by using two different concentrations of index primers.

Sample Name	Reads Total	Reads Filtered	Reads Filtered %	Reads Aligned Genome %	Percent of Duplication %
PBMCs 1250nM_1	2969808	1930992	65	99.5	4.8
PBMCs 1250nM_2	5193443	3331671	64.2	99.5	4.7
PBMCs 1250nM_3	3883646	2401657	61.8	99.5	8
PBMCs 417nM_1	4995248	3234701	64.8	99.4	8.2
PBMCs 417nM_2	4079444	2527034	61.9	99.5	9.3
PBMCs 417nM_3	352490	232415	65.9	99.5	2.4
Total Mi-seq reads	21474079				

Using the above-mentioned optimizations, we built the human immune epigenomic database. We used 474 subjects and we generated over 1350 ATAC-seq libraries from the total and sorted blood immune cell populations. Detailed information is summarized in Table 7.

Table 7. Construction of the epigenome database. Peripheral blood mononuclear cells (PBMCs); central memory (CM); effector memory (EM); follicular helper T cells (Tfh); helper 1 T cells (Th1), helper 2 T cells (Th2); RA positive effector memory T cells (EMRA); myeloid dendritic cells (mDC); plasmacytoid dendritic cells (pDC).

Cell type	Category	Donors	ATAC-seq samples
PBMCs	Healthy Infant	12	36
PBMCs	Healthy Young adult	71	213
PBMCs	Healthy Middle age	47	141
PBMCs	Healthy Elderly	41	123
PBMCs	Healthy Frail	8	24
CD14+	Healthy subject	33	99
BDCA1	Flu Vaccine	3	18
BDCA2	Flu Vaccine	3	18
BDCA3	Flu Vaccine	3	7

Table 7--continued

Cell type	Category	Donors	ATAC-seq samples
Total CD4 T	Healthy subject	13	39
Naïve CD4 T	Healthy subject	15	28
Memory CD4 T	Healthy subject	10	21
CD4 Tcm	Healthy subject	1	1
CD4 Tem	Healthy subject	1	2
Tfh	Healthy subject	7	14
Th1	Healthy subject	7	10
Th2	Healthy subject	7	11
Th17	Healthy subject	7	12
Total CD8 T	Healthy subject	10	30
Naïve CD8 T	Healthy subject	11	21
Memory CD8 T	Healthy subject	10	20
CD8 Tcm	Healthy subject	1	1
CD8 Tem	Healthy subject	1	2
CD8+CD28+	Healthy subject	9	27
CD8+CD28-	Healthy subject	9	27
Total CD19	Healthy subject	13	39
Naïve CD19	Healthy subject	11	21
Memory CD19	Healthy subject	11	14
Innate memory B	Healthy subject	11	16
CD56+	Healthy subject	6	17
CD3+CD56+	Healthy subject	6	15
CD8-NK-NKT-PBMC	Healthy subject	6	18
Naïve T cell	Healthy subject	8	18
Tcm_em	Healthy subject	8	24
Temra	Healthy subject	8	21

Table 7--continued

Cell type	Category	Donors	ATAC-seq samples
mDC	Healthy subject_treatment	3	45
pDC	Healthy subject_treatment	3	37
PBMCs	SLE	17	51
PBMCs	JDM	3	9
PBMCs	Inactive SLE	3	9
CD14+	SLE	14	42
PBMCs	SYS	3	9

Donor Recruitment

Seventy-seven healthy subjects were recruited by the Center for Aging at UConn Health Center to contribute the blood for this study. Of the 77 healthy subjects, 51 were young adult (22~40yrs) and 26 were elderly (>65yrs).

Epigenomic Datasets Generation

Of the 77 health subjects, 49 subjects were used to establish PBMCs epigenomic dataset, 8 subjects were used to established T lymphocyte epigenomic dataset, 7 subjects were used to established B lymphocyte epigenomic dataset, and 20 subjects were used to established monocyte epigenomic dataset. Once we successfully established the epigenomic dataset, the data were analyzed by using different computational tools (Figure 10). The clinical lab recorded the relevant information from each individual, such as age, gender, ethnicity, and personal disease history. More advanced information from each subject, such as cell compositions, was examined by flow cytometry for further analysis.

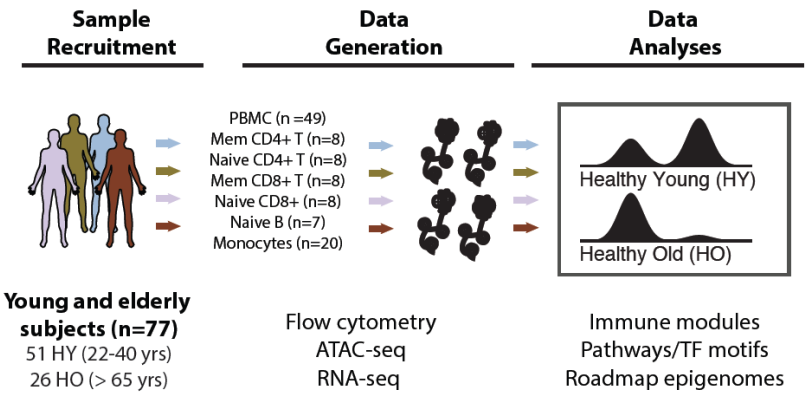


Figure 10. Epigenomic database generation flow chart. All studies were conducted following approval by the Institutional Review Board of University of Connecticut Health Center.

The Major Cell Populations Differ from Age Groups

We recorded cell compositions of PBMCs from different subjects using flow cytometry. As shown in Figure 11, the amount of CD8+ T cells and CD19+ B cells significantly decreased with aging, consistent with the age-related decline in thymus and bone marrow activity. Overall, CD4+ T cells showed non-significant reduction with aging, but the proportion of CD4+ T cell subsets changed, most notably naïve CD4+ T cells were significantly decreased with aging (Appendix A). The most significant aging-associated decline was observed in the naïve CD8+ T cell population (Appendix A), where the percentage of naïve CD8+ T cells in PBMCs decreased from ~7% to ~3% with age (p=1e-04, Wilcoxon rank-sum test).

Major cell population comparisons (%)

	Mean (HY)	Mean (HO)	SD (HY)	SD (HO)	Dif (HY-HO)	p-value (Wilcox)
CD14	15.7	15.6	7.4	5.4	0.1	0.993
CD19	7.5	5.4	3	3.4	2.1	0.014
CD4	30.7	28.5	7.8	10.4	2.2	0.526
CD8	15.6	10.5	4.7	7	5.1	0.006

Figure 11. Cell composition changes with aging. Changes in major cell compositions with aging. P-values are calculated using the Wilcoxon test. Note significant decrease in total CD19+ and CD8+ T cells.

An Epigenomic Signature of Aging in PBMCs

PBMCs, enriched with different immune cells, including T lymphocytes, B lymphocytes, monocytes, and dendritic cells, etc., represent a means by which to assess and monitor an individual's immune health. We examined the aging-associated chromatin accessibility profiles, after PBMC epigenomic database was successfully established using 49 subjects (28 HY, 21 HO). Only high-quality samples passing quality control criteria were used in downstream analysis (44 samples, 25 HY and 19 HO). We analyzed the adjusted reads estimated for all of the PBMC ATAC-seq peaks scored in this study (n=140,172 peaks). Using principal component analysis (PCA), we observed that age groups tend to cluster separately from total ATAC-seq peaks (Figure 12A). From differential analysis, total 12626 differential peaks were separated between age groups. 6977 differential peaks were defined as closing peaks, size of peaks or expression of peaks decreasing with age, and 5649 differential peaks were defined as opening peaks, size of peaks or expression of peaks increasing with age, respectively (Figure 12B). The heatmap result from hierarchical clustering analysis (Figure 12C), the differential peaks can be clearly separated into two groups. The blue cluster represented as HY group, and

the red cluster represented as HO group. The peaks opening with aging and the peaks closing with aging provides a clear picture of separation (Figure 12C).

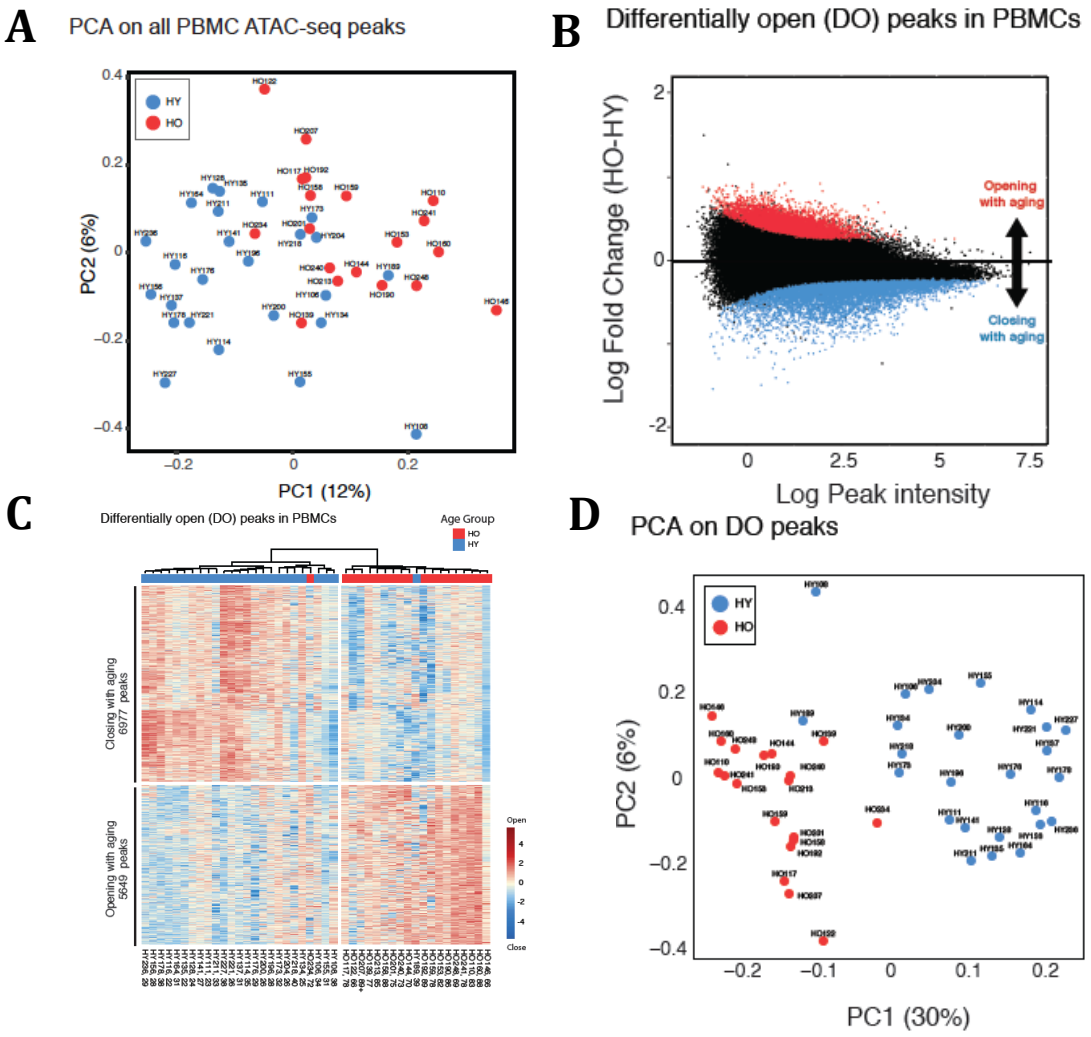


Figure 12. PBMCs epigenomic signature of aging. (A) Sample loadings on first and second principal component computed based on adjusted reads estimated for all of the PBMC ATAC-seq peaks scored in this study (n=140,172 peaks). Note that age groups tend to cluster separately, implying aging as a leading factor explaining epigenetic variation in PBMC. (B) Plot representing log2 fold change (old-young) versus average read count for ATAC-seq peaks. Peaks differentially opening (closing) with aging are represented in red (blue) (5% FDR). (C) Heatmap showing normalized (**z** -scores) chromatin profiles for differentially closing/opening peaks across PBMC samples. (D) Plot of first two Principal Components (PC) based on differential peaks confirms that PC1 accounts for the separation between age groups. Percent of variation among differential peaks accounted for by each PC is shown in parentheses. PC1 from this analysis accounts for ~7% of the variance in the complete data set (Appendix A).

PCA analysis uses differential peaks and age groups as parameters. We found that differential peaks were well separated between age groups in figure 12D. Two subjects (1 HY and 1 HO) were not clustered together with the age group they were supposed to be in (42 out of 44 hierarchical clustering), and considered as outliers.

Definition of Functional States of Differential Peaks

The Roadmap Epigenomics Project (*34*) profiled reference PBMC samples and defined functional states such as promoters, enhancers, and repressors in these cells. To determine where these differential peaks belong to on the genome, we annotated them using these Roadmap-defined chromatin states. Thus, we found that the most accessible peaks were shown at promoter sites and enhancer sites; however, the less accessible peaks were shown at quiescent and repressed sites (Figure 13A). The remarkable differences in the functional states of differential peaks, with closing peaks mostly found at promoters and enhancers, and with opening peaks mostly found at repressed and quiescent sites (Figure 13B). With closing peaks, 25% differential peaks were shown at promoter region, and around 50% peaks were shown at enhancer region. In contrast, with opening peaks, showing less promoter peaks, around 15% differential peaks were shown at repressed region, and 61.3% peaks were shown at quiescent region.

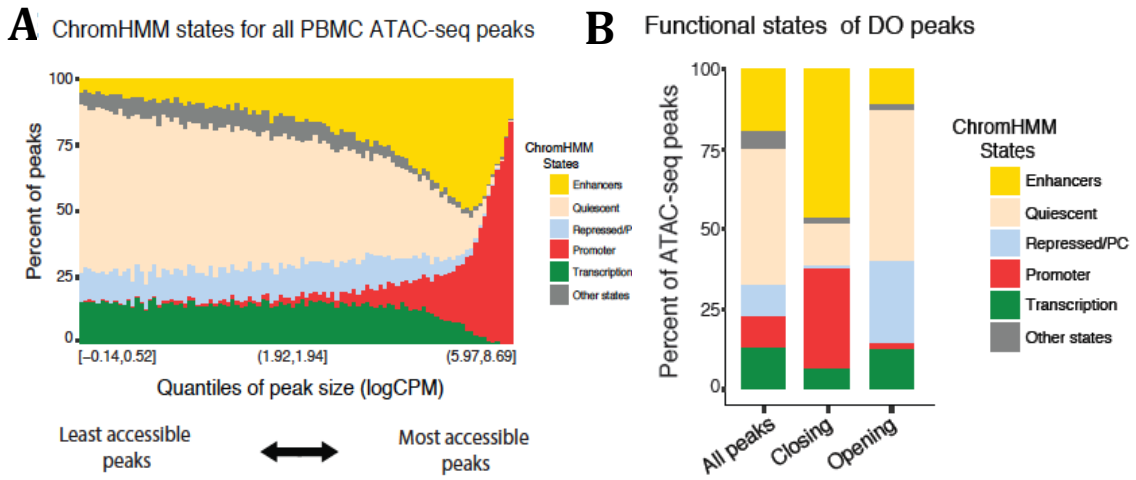


Figure 13. Definition of functional states of differential peaks. (A) Relationship between peak size (i.e., normalized read counts within peak) and functional annotations, as annotated using Roadmap chromHMM states for PBMC. Larger, more accessible peaks are more likely to be found at promoters (red bars) and enhancers (yellow), whereas small peaks are more likely to be called at inactive regions, represented by quiescent (salmon) and repressed sites (slate). (B) Relative to all peaks tested, differentially closing peaks are enriched in promoters and enhancers, whereas opening peaks are enriched in quiescent and repressed sites (Appendix A).

Gene Ontology Terms Definition for Differential Peaks

Differential peaks were annotated to the closing genes based on their distance to transcription start sites (TSS). Therefore, 3,987 closing peaks and 3069 opening peaks were linked to their closing genes respectively (Figure 14A). 622 out of 3987 closing peaks and 379 out of 3069 opening peaks were associated with immune-related genes. In addition, ClueGO (*28*) can compare clusters of genes and visualize their functional differences. ClueGO takes advantage of Cytoscape's versatile visualization framework and can be used in conjunction with the GOlorize plug (*28*). ClueGO enrichment analysis exhibited that a majority of 622 chromatin-closing immune-related genes were involved in T cell activation-related GO terms. In contrast, the majority of the 379 chromatin-opening immune-related genes were involved in myeloid leukocyte and osteoclast differentiation processes (Figure 14B).

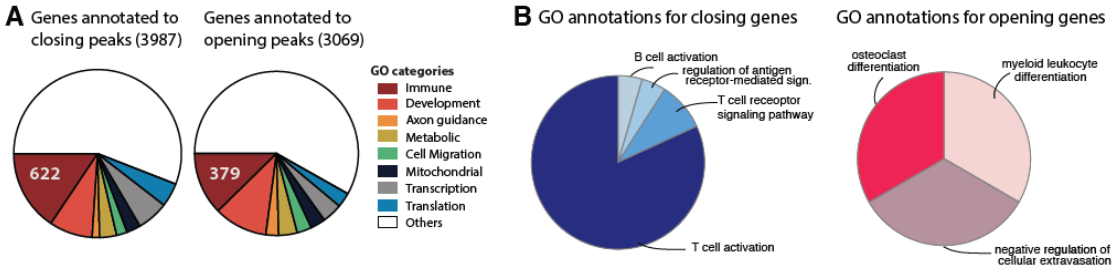


Figure 14. Epigenomic signature of aging at immune-related genes. (A) Major GO category annotations of genes associated with differentially closing and opening peaks. (B) GO terms associated with immune-related genes enriched among genes annotated to differentially closing (blue, left) and opening (red, right) peaks (Appendix A).

Epigenomic Signature of Aging Showing at T Cell Immune Module

To further interpret immunological implication of these chromatin changes, we applied transcription immune modules (*27*), where each module represents a coordinately expressed gene set across many PBMC expression profiles. These immune modules are functionally well characterized and linked to pathways or cell types involved in immune processes. When we applied ATAC-seq data to transcriptional modules analysis, we found significantly chromatin closing at promoter sites and enhancer sites in T cell gene module (Figure 15A). By looking into T cells module genes, we found genes in the T cells module were closing with aging systemically, such as CD28 (Figure 15B). Adopting a module-based data-mining strategy can enhance biomarkers and biological knowledge discovery. For example, in the inflammation I module showed promoter and enhancer peaks closing with aging, but repressed and quiescent peaks opened with aging. It might suggest that aging has dual effect on inflammation-related genes.

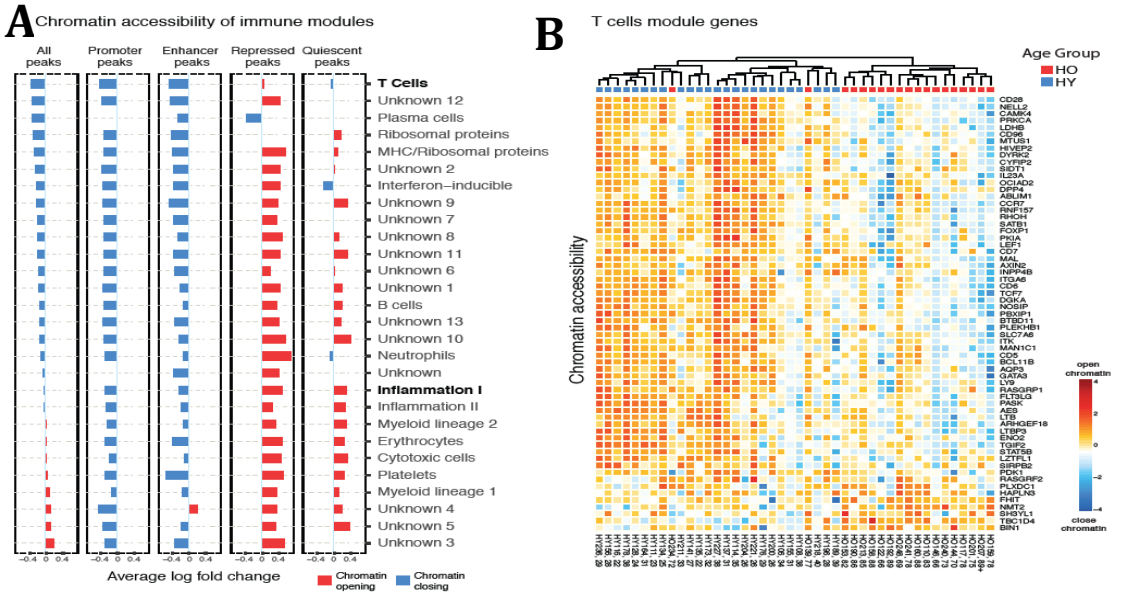


Figure 15 Epigenomic signature of aging showing at T cell immune module. (A) Average chromatin remodeling (log2 fold change) of genes listed in 28 immune co-expression modules, calculated based on all peaks (leftmost column) and separately using peaks annotated to specific chromHMM states. (B) Subject-specific normalized (**z** -scores) chromatin accessibility patterns of peaks annotated to genes in the T cell co-expression module reveals concerted aging-related variation across the cohort. Warmer (cooler) hues represent increased (decreased) chromatin accessibility relative to the cohort mean.

The Correlation between Epigenome and Transcriptome

To link aging-associated chromatin changes to transcription levels, we generated a PBMCs transcriptome database with 39 subjects (24 HY and 15 HO). 39 RNA-seq data were generated to match with ATAC-seq samples. The most significantly positive correlation between age-related changes in gene expression levels and chromatin accessibility at gene promoters was detected when we compared chromatin accessibility to gene expression data (Figure 16).

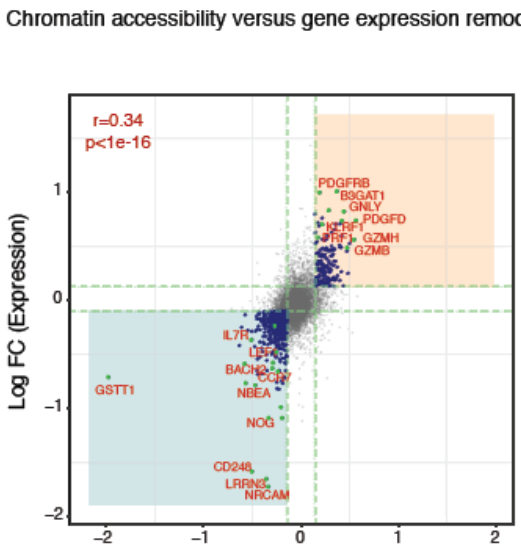


Figure 16. Aging-associated gene expression and chromatin accessibility changes. Chromatin remodeling at gene promoters correlates significantly with changes in expression of the co-located genes (Pearson r = 0.32, p-value < 2.2×10^{-16}). Dashed lines delineate the set of peaks (x-axis) and genes (y-axis) that are differentially accessible or expressed between young and old subjects with a permutation-based p value<0.01. Shaded quadrants define sets of genes showing congruent aging-related shifts in chromatin accessibility and expression (Appendix A).

Concordant Remodeling Together with Transcriptome Data and Epigenomic Data

Analysis of ATAC-seq data and RNA-seq data showed concordant data remodeling, increasing or decreasing together in the same direction. Next, we have to identify which transcriptional immune modules undergo transcriptional, epigenetic, and concordant changes with aging (Figure 17). We identified that the most significant concordant chromatin closing and decline of gene expression occurs in T cell module (Figure 17, Combined remodeling column and Figure 18). Furthermore, the most significant concordance between chromatin opening and increased in gene expression occurs in cytotoxic cell module (Figure 17).

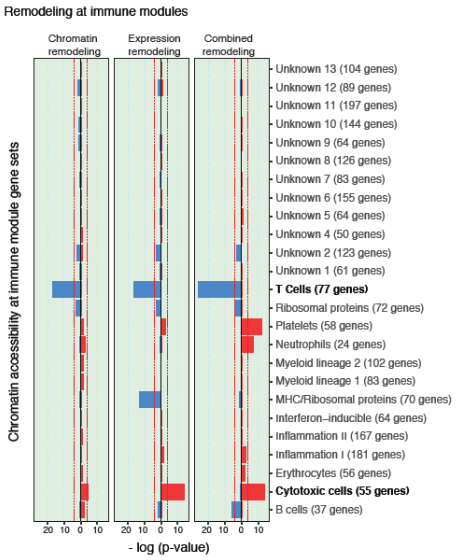


Figure 17. Concordant remodeling together with transcriptome data and epigenomic data. Enrichment of immune modules among genes sets associated to differentially accessible peaks (left), differentially expressed genes (center), and congruent (concordant) chromatin and expression remodeling (right). Plots show - log10 of hypergeometric test p-values, colored according to the direction of the observed change (blue for decrease and red for increase with age). Reference lines are drawn at the largest p-value for which a 5% FDR is attained (Appendix A).

In depth analysis of T cell module showed that gene expression decrease correlates with chromatin accessibility closing with aging. Many genes in this module are associated with T cell functions, including Transcription factors (TF), involved in lymphocyte development and activation. LEF1 and TCF7 are among the genes associated with T cell function (Figure 18B). Figure 18B indicates that ATAC-seq signal and RNA-seq signal were correlated, and gene expression and chromatin accessibility was decreased with aging.

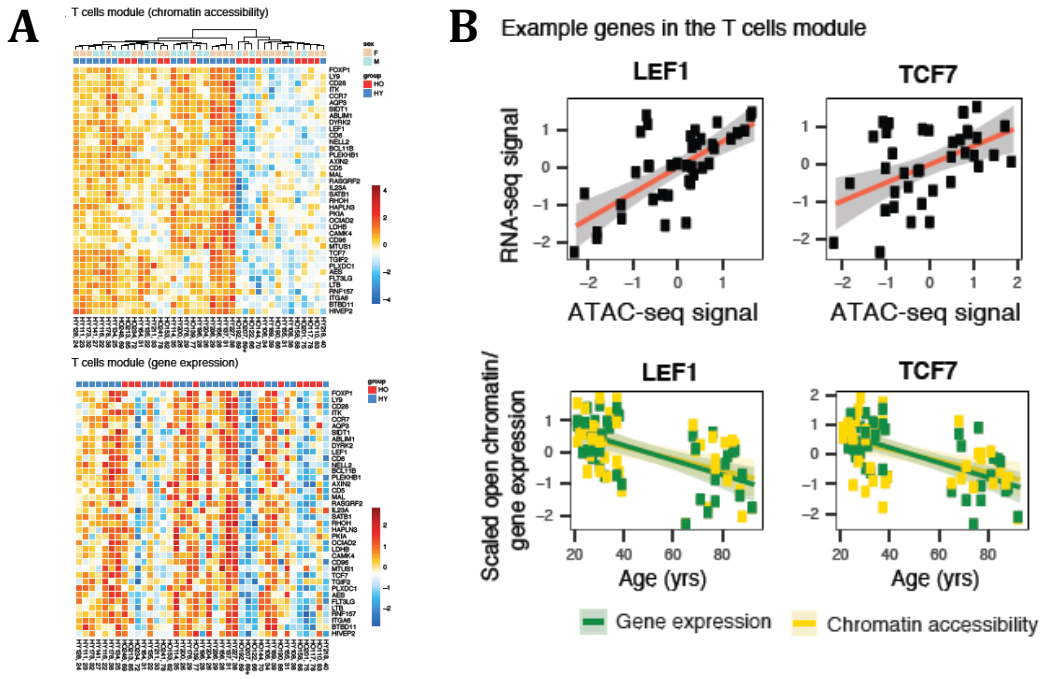


Figure 18. Concordant remodeling in T cells module. (A) Personal profiles of chromatin accessibility of promoter peaks associated to genes in the T cell module for which both chromatin accessibility and gene expression decrease with aging. Personal profiles of gene expression of genes in the T cell module, with samples and genes sorted to match the clustering computed based on chromatin accessibility. (B) Examples of concordantly remodeled genes from the T cells module. Top: chromatin accessibility and gene expression correlate among subjects. Bottom: both chromatin accessibility (yellow dots and lines) and gene expression (green dots and lines) decrease with aging (Appendix A).

In the cytotoxic cell module chromatin opening correlates with the activation of gene expression (Figure 19A). Among these are GNLY, GZMB, and GZMH, expressed by plasmacytoid dendritic cells (*53*), CD4+ T cells (*54, 55*), or plasma cells (*56*) (Figure 19B).

The epigenomic dataset is comparable with transcriptome dataset, showing concordant chromatin closing and decline of gene expression. Top 15 closing peaks were generated from ATAC-seq data and associated with chromatin closing (Figure 20).

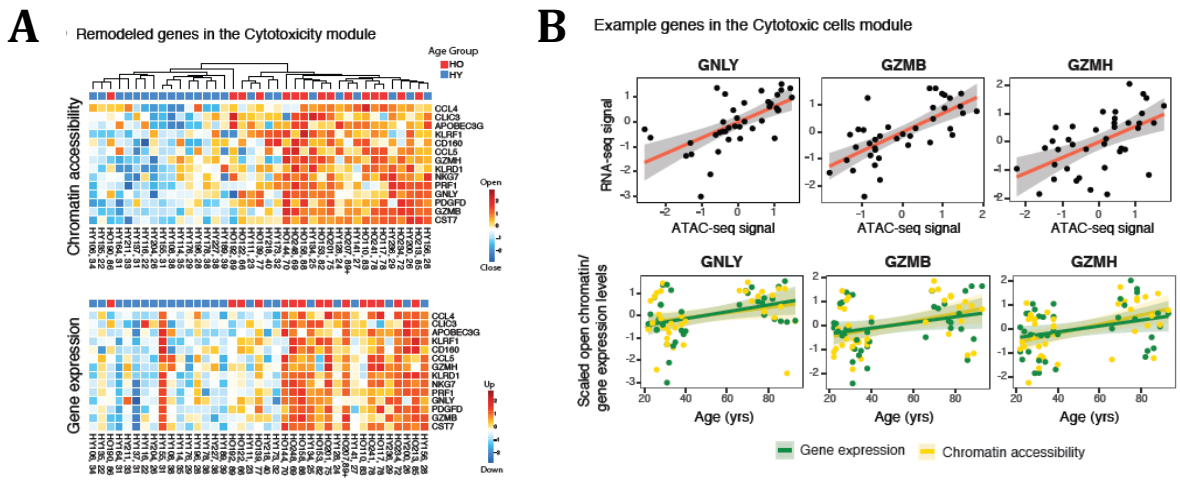


Figure 19. Concordant remodeling in cytotoxic cell module. (A) Promoter chromatin accessibility (top) and gene expression (bottom) of genes in the cytotoxic cells module that show congruent increased in accessibility and expression with aging. Warmer (cooler) hues represent increased (decreased) chromatin accessibility (expression) relative to the cohort mean, data shown as normalized (**z** -scores) values. (B) Examples of concordantly remodeled genes from the cytotoxic cells module. Top: chromatin accessibility and gene expression correlate among subjects. Bottom: both chromatin accessibility (yellow dots and lines) and gene expression (green dots and lines) increase with aging (Appendix A).

Many peaks were associated with immune-related function, such as TGFBR2 (*57*), BACH2 (*58*), FOXP1 (*59*), IL7R, etc. IL-7R has been reported to be associated with lymphocytes development, proliferation, and activation (*42*). Noteworthy, recent gerontology studies reported that IL-7R decreased with aging. Willemijn et. al. concludes that the IL7R network reflected by gene expression levels in blood may be involved in the rate of ageing and health status of elderly individuals (*42*). Deficiencies of other transcription factors that control IL-7Ra expression, such as Foxo1 and Ets1, also lead to a more profound reduction of peripheral naïve CD8+ T cells than CD4 T+ cells (*43, 44, 45*), strengthening reports that IL-7Ra controls development and survival of naïve CD8+ T cells more stringently than CD4+ naïve T cells (*46, 47*). Therefore, we selected IL-7R as a validation candidate for further analysis and validation experiments.

Top 15 peaks associated with chromatin closing

Gene Name	# of closing peaks	Average FDR	Average logFC	Average logCPM
ETS1	24	2.05E-02	-0.40	2.17
TGFBR2	15	8.75E-03	-0.49	2.40
FAM65B	13	5.95E-03	-0.48	2.21
BACH2	13	8.00E-03	-0.51	2.32
FOXP1	12	1.05E-02	-0.46	2.57
IL7R	12	4.89E-03	-0.60	2.12
BRINP3	11	3.03E-02	-0.41	0.92
CCR7	11	9.36E-03	-0.53	2.45
NCK2	11	9.11E-03	-0.43	2.81
CDC42SE2	11	1.15E-02	-0.38	2.75
LRRN3	11	5.26E-03	-0.58	1.96
IKZF1	11	1.25E-02	-0.44	2.86
PTPRC	10	2.05E-02	-0.35	2.88
CD44	10	1.03E-02	-0.47	2.56
SNX9	10	1.43E-02	-0.43	2.57

Figure 20. Top 15 genes associated with chromatin closing with aging. Genes are sorted with respect to the number of significantly closing peaks annotated to their promoters. IL7R is a top gene in this list with 13 closing peaks (Appendix A).

The Chromatin Accessibility Closing with Aging in IL-7R and IL-7 Signaling Pathway

We have visualized IL-7R peaks information using UCSD genome browser from 10 subjects (5 HY and 5 HO). Eight out of 12 highlighted peaks, located at either promoter or enhancer sites decreased with aging (Figure 12A). Moreover, ATAC-seq signal and RNA-seq signal were positively correlated; chromatin accessibility and gene expression decreased together with aging (Figure 21B, 21C). Furthermore, aging-related chromatin closing with aging affected not only IL-7R, but also disrupted other genes in the IL-7 signaling cascade. Thus, genes in the IL-7 signaling pathway, such as JAK1, JAK2, STAT5A, and STAT5B have their chromatin closed with aging (figure 21D) (*60*). The heatmap provided a more dynamic chromatin-closing pattern in IL-7 signaling pathway at subject level (Figure 21E).

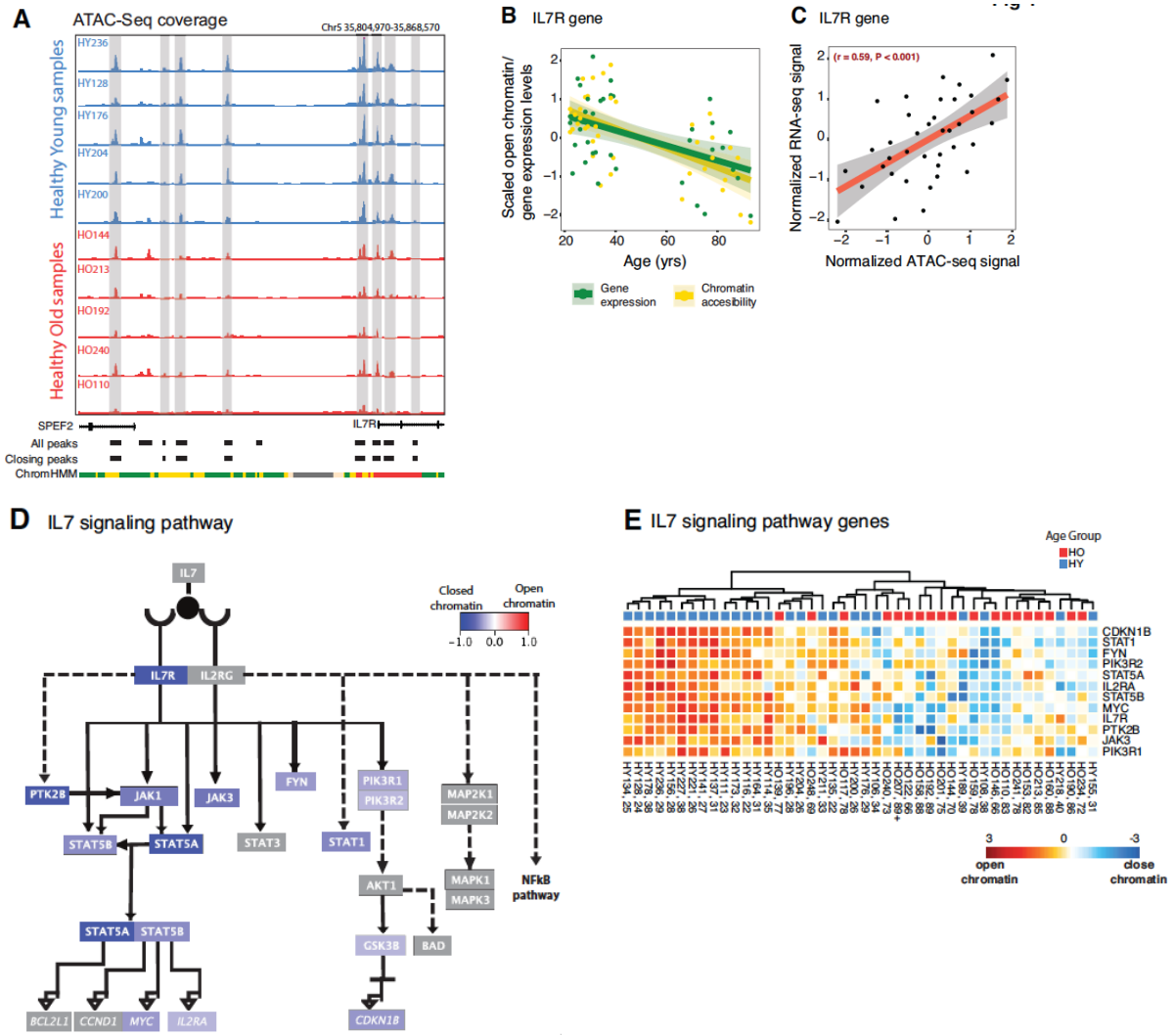


Figure 21. The chromatin accessibility is closing with aging in IL-7R and IL7 signaling pathway. (A) Genome browser view of IL7R locus highlighting 8 (out of 12) differentially closing peaks. Blue and red tracks represent open chromatin profiles of HY (n=5) and HO (n=5) samples, respectively. (B) IL7R expression and chromatin accessibility at its promoter decrease with aging. (C) Promoter chromatin accessibility and gene expression are highly correlated at among subjects. (D) Chromatin accessibility of peaks annotated to genes in the IL7 signaling pathway. Color represents the fold change of the most significant differential peak annotated to this gene. Genes marked in grey are not associated with a closing or opening peak. (E) Subject-specific chromatin accessibility of peaks significantly closing with aging and annotated to genes in the IL7 signaling pathway. Warmer (cooler) hues represent increased (decreased) chromatin accessibility relative to the cohort mean, data shown as normalized (z-scores) values (Appendix A).

IL-7R Protein Expression Decreases with Aging

Then we examined IL-7R protein expression between HY and HO in different immune cells, such as CD4+ T cells, CD8+ T cells, CD19+ B cells, and CD14+ monocytes.

Figure 22A indicated that IL-7R protein expression decreased specifically in CD8+ T

cells in HO compared to the expression in CD8+ T cells in HY. Other immune cells showed no difference between HY and HO groups. When the IL-7R protein expression was analyzed as a function of age, the IL-7R protein expression in CD8+ T cells showed a profound reduction with age (Figure 22B).

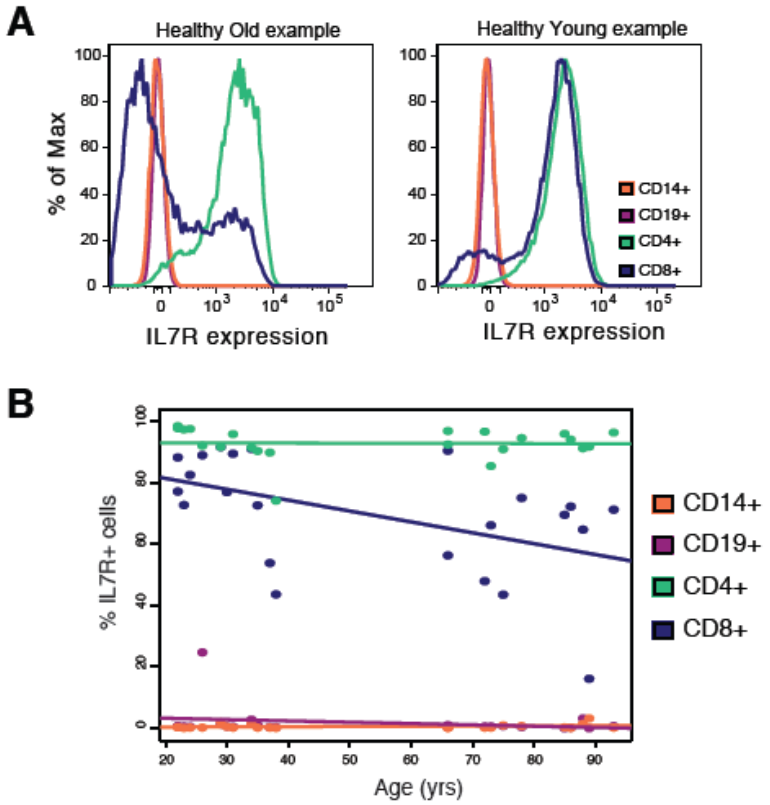


Figure 22. IL-7R protein expression decreases with aging. (A) Flow cytometry plots in representative young (left) and old (right) subjects illustrate the decrease in IL7R protein levels with aging in CD8+ T cells. (B) Flow cytometry results indicating that the aging-related decrease in IL7R levels is specific to CD8+ T Cells (Appendix A).

The Chromatin-Remodeling Signature Presented in CD8+ T cells

We next analyzed chromatin-remodeling signature on different T cell subpopulations such as naïve and memory CD4+ T cells and CD8+ T cells. The ATAC-seq dataset were generated after checking the purity of sorted cells. Using Generalized Linearized Model to distinguish the differential peaks from naïve and memory CD4+/CD8+ T cells between

HY and HO. Figure 23A indicated that naïve and memory CD4+ T cells showed minimal chromatin remodeling pattern; however, naïve and memory CD8+ T cells showed dramatic differences in chromatin remodeling. Naïve CD8+ T cells displayed 2,925 differential peaks, and memory CD8+ T cells displayed 8,503 differential peaks. Differential peak analysis in CD8+ T cells indicated that gene promoter closing is substantially higher on memory CD8+ T cells (Figure 23B). Closing peaks in memory CD8+ T cells were 4 times higher than in naïve CD8+ T cells. Further differential peaks information was identified with Roadmap T cell annotation ChromHMM states (*34*). The pattern of closing peaks in memory CD8+ T cells showing at promoter sites and enhancer sites was similar to PBMCs pattern. In contrast, more than 50% of opening peaks in naïve CD8+ T cells occurred at enhancer sites (Figure 23C).

We, next, inspected chromatin remodeling of selected gene promoters known to be expressed in CD4+ T cell subsets. Figure 24 indicates that surface molecules, such as IL7R, CD28 (*61*), and signaling molecule STAT4, chromatin-remodeling signature was mainly from memory CD8+ T cells, instead of the three T cell subsets.

Last, we compared age-induced chromatin remodeling in T cell subsets to that of PBMCs. Chromatin closing in PBMC around T cell signaling pathways, such as IL-7 signaling pathway and TCR signaling pathway, mostly stemmed from memory CD8+ T cells (Figure 25). Taken altogether, these results identify memory CD8+ T cells as the subpopulation with the most profound chromatin remodeling with aging.

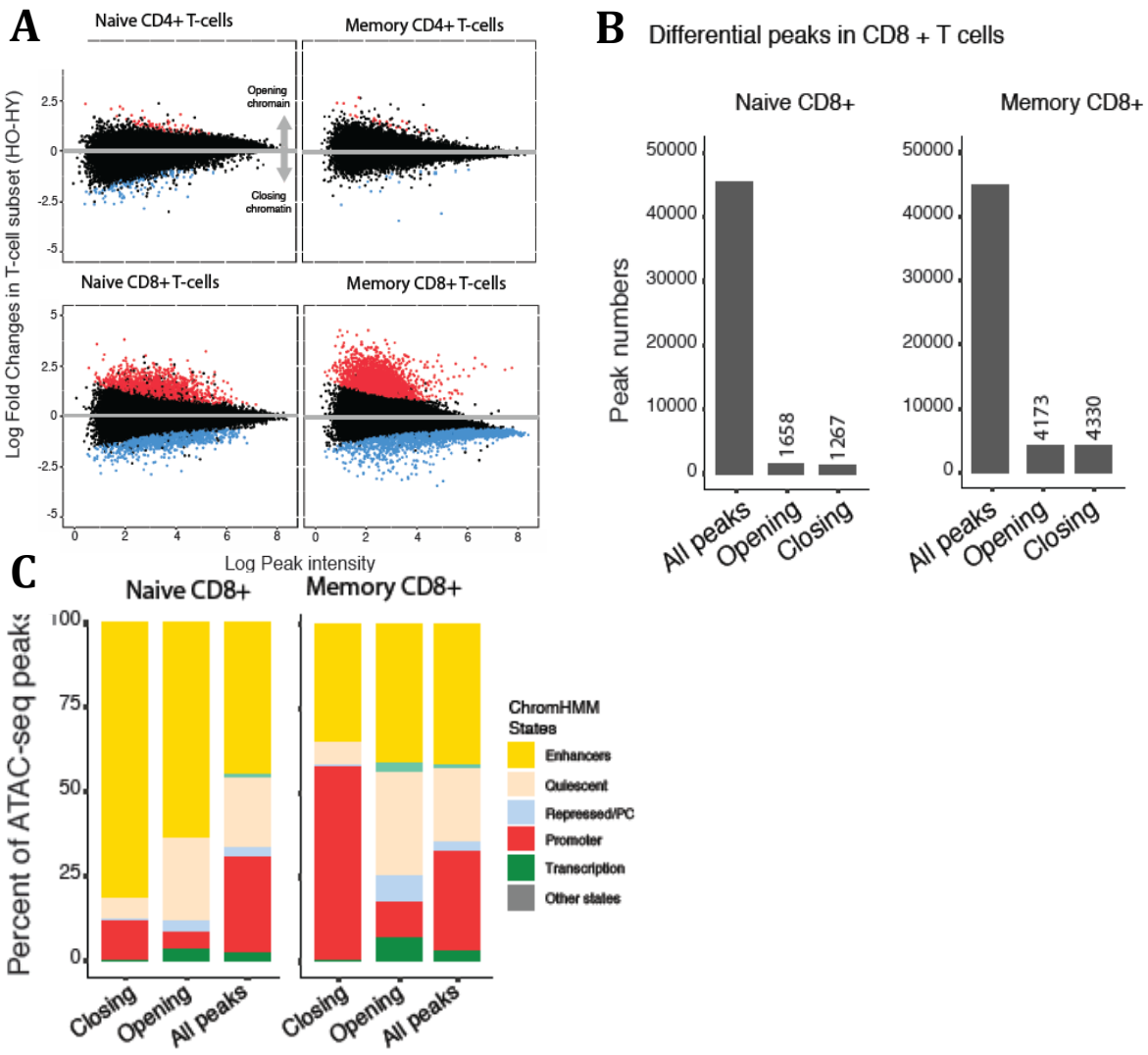


Figure 23. The chromatin-remodeling signature showed in CD8+ T cells. (A) Differential accessibility analyses in T cell subsets show that most significant aging-related remodeling occurs in CD8+ T cells, and particularly in memory CD8+ T cells. Plots representing log2 fold change (old-young) versus average read count for the corresponding ATAC-seq peaks in T cell subsets. Opening (closing) peaks are represented in red (blue) (5% FDR). (B) Number of peaks in memory and naïve CD8+ T cells showing the increased extent of remodeling associated to aging observed in memory CD8+ T cells. (C) Distribution of differential and all peaks classified by chromHMM state annotations (Roadmap T cell annotations) for memory and naïve CD8+ T cells. Promoters and enhancers close with aging in memory CD8+ T cells, similar to PBMCs (Appendix A).

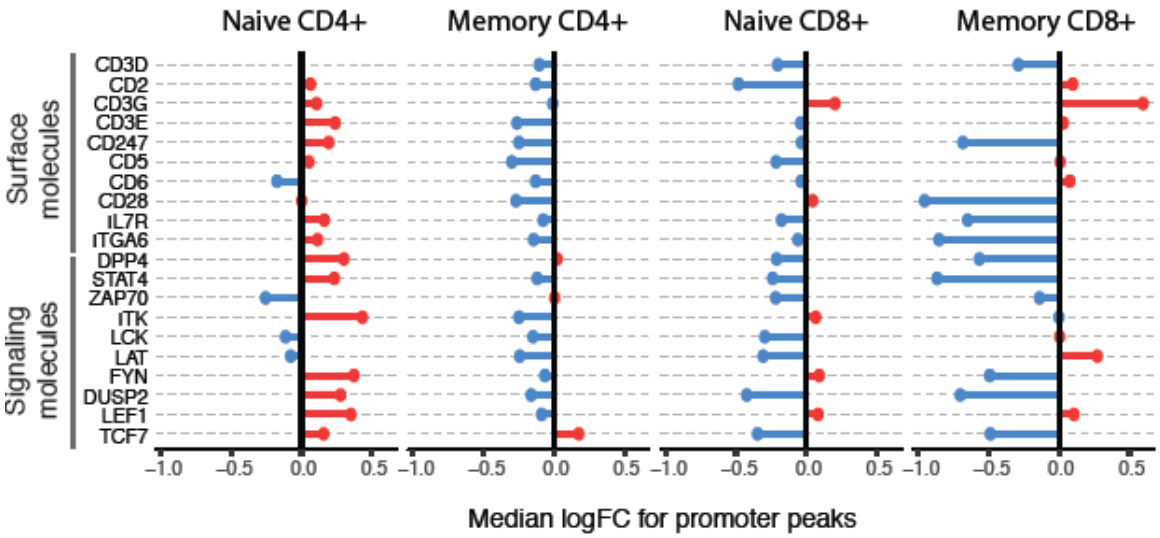


Figure 24. Selected gene promoters displayed chromatin accessibility remodeling in memory CD8+ T cell. Chromatin accessibility remodeling (median fold change) of promoters of selected functionally relevant signaling and surface molecules in naïve and memory CD4+ and CD8+ T cells. Red and blue bars represent positive (i.e., opening with aging) and negative (i.e., closing with aging) median fold change, respectively, aggregated over all peaks overlapping promoters of the corresponding gene (Appendix A).

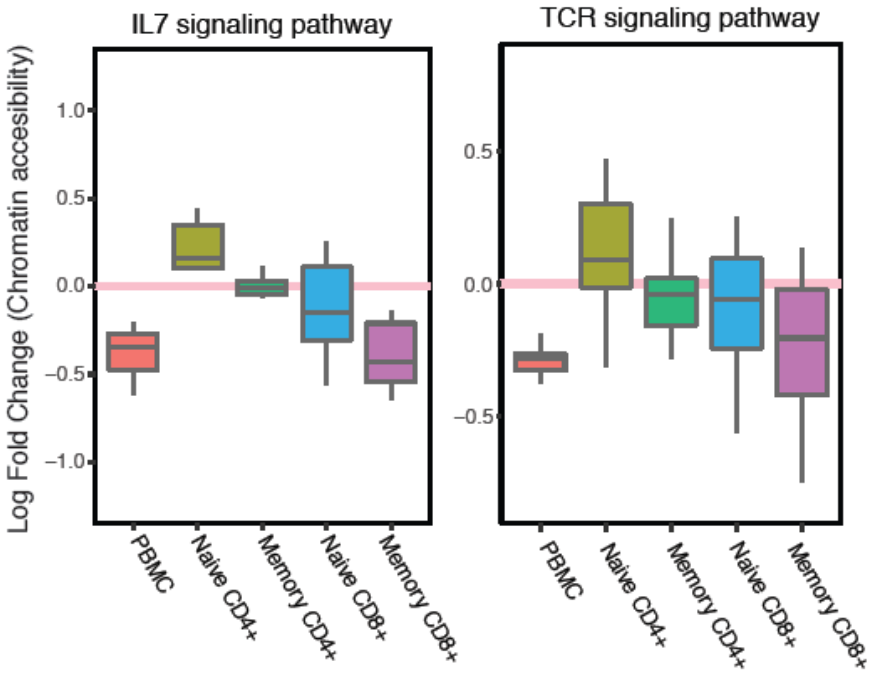


Figure 25. The comparison of age-induced chromatin accessibility in T subsets to PBMCs. Chromatin remodeling of closing PBMC regions associated to genes in the IL7 signaling pathway (left) and TCR signaling pathway (right) stems from the remodeling in memory CD8+ T cells. Boxplots for PBMC and T cell subsets represent distribution of log2 fold changes of peaks annotated to genes that are associated to closing peaks in PBMC (Appendix A).

CHAPTER FIVE

Conclusions

These results indicate the importance of optimizing the ATAC-seq protocol and the importance of studying the epigenetic landscape of aging in human PBMCs. Here, we provide an insight to the anatomy of immunosenscence associated with aging. The major observations from the epigenomic data generated from PBMC isolated from the blood of young and old healthy donors is as follows:

- Chromatin accessibility reduces systemically with aging. The majority of differential peaks closing at enhancer sites and promoter sites are associated with immune genes related T cell signaling pathways and T cell activation. In contrast, chromatin opening with aging is more stochastically distributed across the cohort, instead of systemically opening with aging. The majority of differential peaks opening at quiescent sites and repressed sites are associated with immune genes related myeloid leukocyte differentiation and osteoclast differentiation.
- Epigenomic signature in PBMCs mainly relates to memory and naïve CD8+ T cells.
- IL-7R gene and genes involved in the IL-7 signaling pathway may qualify as potential epigenomic biomarkers of aging.

From the technical standpoint, optimized ATAC-seq technology can provide a powerful tool to investigate the human epigenome. Among the major observations, the

reduction of mtDNA in the analyzed DNA and removal of redundant material from ATAC-seq libraries result in improved sequencing results:

- The use of improved buffer reduces between 30% and 50% the mtDNA contamination and increases the percentage of reads aligning to genomic DNA from 30% to 60%.
- Reduction of amplification cycles used in preparation of ATAC-seq libraries reduces redundant repetitive sequences, which substantially improves computational analysis.

CHAPTER SIX

Discussion

The major findings of the present study include, i. technical improvements to the ATAC-seq technique and ii. identification of the epigenomic landscape of the immune-related genes in elderly in particular memory CD8+ T cells.

A significant improvement to the current ATAC-seq technology is related to the optimization of the nuclear isolation buffer that is more selective in the isolation of genomic DNA and profoundly reduced the mtDNA extraction. The new buffer successfully separates mitochondrial organelle and nuclei from cells and does not alter the sequencing efficiency. The key component in this new lysis buffer is triton X-100. Proper concentration of this mild detergent, after 15 minutes incubation on ice, and 5 minutes centrifuge; then, we isolated good quality nuclei.

This method provided as with better results than any other tested methods such as size exclusion or using of blocking primers. Thus, the selection of 100-500 bps fragments using the BluePippin system poses the risk of excluding potential useful information. Moreover, we find that mtDNA also contains 100 bps to 200 bps fragments. Other methods, such as designing a set of mtDNA blocking primers, needs computational analysis to predict or to collect information about the frequency of mtDNA fragments randomly cut by Tn5 transposase. Outsourcing this methodology to the biotech company takes long time and increases the cost of experiments, even though it works successfully. But it will eventually bring up another issue, which is the comparison of different

techniques. Taken all together, new lysis buffer makes the ATAC-seq better in the most efficient. The major finding relative to the immune epigenome in elderly was that we find the chromatin accessibility closing with aging in PBMCs; the difference of chromatin accessibility between CD4+ and CD8+ T cells, CD8+ T cells display more epigenome changes with aging.

Of note, combined analysis of chromatin accessibility (ATAC-Seq) and gene expression (RNA-Seq) profiles revealed IL-7R and genes involved in IL-7 signaling pathway are dramatically reduced in elderly and may be potential biomarkers to monitor the status of immunity in elderly. Thus, factors that can regulate IL-7R gene, such as NF-kB and STATs, closing with aging are specifically associated with memory CD8+ T cells. Future experiments aimed at identification of poised enhancers, active enhancers, and promoters in chromatin immune-precipitates directed to specific histones (H3K4me1, H3K4me3, H3K27ac) will be crucial for validation. Combining these in vitro results and designing in vivo experiments will eventually lead us to clinical studies.

Recent studies mention that IL-7 therapy may improve the survival of $EM_{CD45RA+}$ CD8+ T cell with diverse TCR repertoire in the young but not in the elderly(*35*). It suggests that IL-7 therapy in the elderly may not be as effective and beneficial as in the young because a large proportion of CD8+ T cells in the elderly group has decreased IL-7Rα expression and limited TCR repertoire compared with the young group. Indeed, the clinical studies of recombinant human IL-7 (rhIL-7) to rejuvenate circulating T cells it is a matter of debate in the scientific community. Thus, we need to understand how to up-regulate IL-7R in aging CD8+ T cells.

Module-based data-mining strategy is very useful for ATAC-seq data analysis;

however, it was used for patient-based microarray transcriptional studies(*18*). Many genes in the module were well defined and were associated with different immune responses and different cell types in previous studies. But, in the literature, many genes and regulators that are associated with DNA replication, metabolism, non-coding elements, etc. are still undetermined. Further characterization of these modules using epigenomic data will be beneficial for the identification of potential biomarkers and regulators especially in the noncoding space (*97*).

Aging-associated chromatin accessibility profiles are stable and robust. It is neither associated with cytomegalovirus (CMV) seropositivity and season. Ageing-related increase in CMV seropositivity may affect chromatin signature in PBMCs. Analysis of either CMV+ cohort and CMV- cohort or CMV+ cohort shows no influence of CMV seropositivity. We divided our samples into two groups according to the collection season, from Dec to May as winter group, from Jun to Nov as summer group. The result shows that season does not affect aging-associated chromatin signature (Appendix A).

Last but not least, the epigenomic signature from other immune cells, such as monocytes or B lymphocytes, do not show any major difference between young adult group and elderly group (Appendix A). The phenotype of monocytes and B lymphocytes between young and elderly adults showed no difference, but the way these immune cells respond to stimuli may display difference. Frasca, D et al. explored that T-bet and CD11c expression was higher in memory than in naïve B cells, but no difference was observed between young and elderly individuals. After TLR7 stimulation, CD11c increases in all B cell subsets (especially in naïve and IgM) from the elderly (*41*). It explains why there are no epigenomic signature differences between young and elderly adults, and provides a

hint to add proper stimuli to monocytes, or B lymphocytes. Besides, we did not collect germinal center B lymphocytes. Possibly, germinal center B lymphocytes would show the difference of epigenomic signature in young and elderly adults. It may also be related to the small cohort size. We will try to increase cohort size to see what can be concluded. If not, looking for the proper stimuli would be another important task.

In sum, our computational analysis and advanced ATAC-seq technique established this systems immunology approach as an integral and powerful immune-monitoring tool for the study of epigenome in diverse clinical contexts.

APPENDICES

Appendix A

The Chromatin Accessibility Signature of Aging in Human Blood Leukocytes Stems from CD8+ T Cells (Accepted)

The chromatin accessibility signature of aging in human blood leukocytes stems from CD8+ T cells

Authors: Duygu Ucar[1†*], Eladio J. Márquez[1†], Cheng-han Chung[1,3†], Radu Marches[1], Robert J. Rossi[1], Asli Uyar[1], Te-Chia Wu[1], Joshy George[1], Michael Stitzel[1], Karolina Palucka[1], George A Kuchel[2*], and Jacques Banchereau[1*]

Affiliations:

[1]The Jackson Laboratory for Genomic Medicine, Farmington, CT.

[2]University of Connecticut Center on Aging, University of Connecticut, Farmington, CT.

[3]Department of Biomedical Studies, Baylor University, Waco, TX.

*To whom correspondence should be addressed: jacques.banchereau@jax.org, kuchel@uchc.edu, duygu.ucar@jax.org

†Co-first authors

One Sentence Summary: Aging is associated with epigenomic silencing of promoters and enhancers in CD8+ T cells.

Abstract:

Aging is linked to the functional alteration of the immune system associated with deficiency in responses to pathogens and increased systemic inflammation. Here we employed systems immunology approaches to establish chromatin accessibility and transcriptome profiles in peripheral blood mononuclear cells (PBMCs), and purified monocytes, B and T cells. Analysis of cells from healthy young and elderly subjects enabled the identification of an epigenomic signature of aging in PBMCs composed of (1) a systematic chromatin closing at promoters and enhancers associated with pathways involved in T cell function and (2) a potentially stochastic chromatin opening, mostly found at quiescent and repressed sites. Chromatin changes were concordant with gene expression changes and uncovered immune molecules activated or inactivated with aging, notably the silencing of *IL7R* gene – captured at the epigenome, transcriptome, and protein level - and the IL-7 signaling pathway genes. Chromatin accessibility profiles of examined subpopulations demonstrated that these aging-associated changes are borne by the CD8+ T cells. Among those, memory CD8+ T cells displayed the most profound aging-related changes, involving widespread chromatin closing at gene promoters including *IL7R* and transcription factors (TF) regulating immune functions. TF footprinting analyses suggested an aging-related loss in binding of NF-κB and STAT factors in memory CD8+ T cells. These aging-related chromatin accessibility changes are independent of cytomegalovirus status and seasonality. Thus, our study reveals that aging is associated with significant chromatin

accessibility alterations in CD8+ T cells – that can be captured via PBMC profiling- potentially explaining the impaired responses to infections and cancer.

Introduction

As we age, our immune system undergoes a broad range of functional changes including the two hallmarks: 1) immunosenescence, i.e., functional decline, especially impacting the adaptive arm of the immunity (1-3) and 2) "inflamm-aging", i.e., a persistent systemic inflammatory state (4, 5). These changes lead to diminished ability of the immune system to generate protective responses to immunological threats, predisposing older adults to infection, as well as raising the risk of many chronic diseases (6-8). Disruption of tightly regulated gene expression programs in immune cells is likely to underlie these functional changes.

Earlier studies have shown that epigenomic changes associated with aging lead to disruption of gene-expression patterns and transcriptional regulatory programs across diverse cell types and organisms (9-11). Transcriptomic profiling of human PBMCs and purified immune cells revealed genes that are differentially expressed with aging (12-14). Moreover, DNA methylation studies demonstrated that human immune system aging is associated with methylation changes at specific CpG sites (15-20). However, how chromatin accessibility patterns are disrupted with aging in human immune cells and which immune functions are affected by these disruptions remain relatively unknown.

ATAC-seq (21, 22) is a recent technology that enables genome-wide profiling of chromatin accessibility patterns at base pair resolution using limited cell numbers. This technology offers remarkable opportunity to define aging-associated disruptions to transcriptional regulatory programs in human immune cells with increased precision, including changes in non-coding cis-acting sequences (e.g., enhancers) and transcription factor (TF) activity. In this study, we profiled the chromatin accessibility (ATAC-seq) and gene expression (RNA-seq) in human PBMCs and purified subsets. This led to the identification of a chromatin accessibility signature of aging in PBMCs and the demonstration that it stems from CD8+ T cells.

Results

An epigenomic signature of aging in PBMCs

PBMCs, a composite of immune cells, represent a tissue resource to assess and monitor an individual's immune health and responses longitudinally. We have successfully applied PBMC profiling in earlier studies as means of identifying transcriptomic signatures of autoimmune diseases and of immune responses to infectious agents (23-25). To examine aging-associated chromatin accessibility profiles, we collected blood and isolated PBMCs from 77 healthy community-dwelling research volunteers: 51 healthy young (HY, 22-40 yrs) and 26 healthy old (HO, 65+ yrs) subjects (Fig.1A, table S1). As the changes captured in PBMC epigenomes could be attributable to both differences in the frequency of certain cell types and changes in genomic patterns that are intrinsic to specific cell subsets (26), we also examined the cell composition of PBMCs using flow cytometry from a subset of these subjects (table S2). Proportions of naïve CD4+ T cells, naïve CD8+ T cells and CD19+ B cells significantly decreased with aging, consistent with age-related decline in thymus and bone marrow activity (fig. S1, A, B, and C). The most significant aging-associated decline was observed in the naïve CD8+ T cell population (fig. S1C), where the percentage of naïve CD8+ T cells in PBMCs decreased from ~7% to ~3% with age ($p=1e^{-04}$, Wilcoxon rank-sum test).

ATAC-seq profiles were generated from 49 subjects (28 HY, 21 HO) by incubating the purified nuclei with Tn5 transposase to cut and "tag" accessible chromatin and sequencing the resulting "tags" to identify genome-wide open chromatin patterns. This approach identified 140,172 open chromatin sites (i.e., peaks) associated with 22,124 genes based on their distance to transcription start sites (TSS). Only high-quality samples passing quality control criteria were used in downstream analyses (44 samples, 25 HY and 19 HO) (Methods). Using a Generalized Linear Model, 12,626 differentially accessible peaks (9% of tested, FDR<0.05) were identified between age groups. Of these, 6,977 showed a decrease and 5,649 showed an increase in chromatin accessibility with aging, hereafter referred to as "closing" and "opening" peaks, respectively (Fig. 1B). Principal component analysis and hierarchical clustering analyses confirmed that differential peaks discriminated between age groups with high accuracy, where a majority of samples (42 out of 44 in hierarchical clustering) clustered based on the age group alone (Fig. 1, C and D, fig. S1D).

The Roadmap Epigenomics Project (27) profiled reference PBMC samples and defined functional states such as promoters and enhancers in these cells. To determine the location of ATAC-seq open chromatin sites (e.g., promoters, enhancers) we annotated them using these Roadmap-defined (i.e., ChromHMM) chromatin states. As expected the most accessible peaks mostly overlapped with promoter and enhancer states (fig. S1E). In contrast, less accessible peaks were annotated with repressed or quiescent states (fig. S1E). For example, the most accessible 10% among all peaks (14,222) were mostly at promoters (55.6%) and enhancers (36.1%), whereas the least accessible 10% (14,199) were mostly at quiescent sites (61.3%). As shown in Figure 1E, there was a remarkable difference in the functional annotation of differential peaks, with closing peaks mostly found at promoters and enhancers and opening peaks mostly found at repressed and quiescent sites. Moreover, closing peaks were more accessible (i.e., larger peaks) on the average than opening peaks and than all ATAC-seq peaks (fig. S1F).

Analysis of the frequency of differentially accessible chromatin regions across the cohort, revealed that closing peaks consist of open chromatin sites that are common between subjects (i.e., high-frequency peaks), whereas opening peaks consist of low-frequency and often subject-specific peaks (Fig. 1F). Thus, examination of chromatin accessibility led to identification of an epigenomic signature of aging in PBMCs composed of i) chromatin closing at the most accessible promoter/enhancer regions of the genome across the population and ii) a chromatin opening of less accessible regions of the genome in a subject-specific manner.

Chromatin closing at promoters and enhancers with aging

Differential peaks were annotated to the nearest gene based on their distance to transcription start sites (TSS), thereby linking 4,567 and 3,816 genes to "closing" and "opening" peaks, respectively (Fig. 2A, table S3 for genes associated with differentially open peaks). Genes annotated to differentially accessible peaks were further characterized using Gene Ontology (GO) terms, which revealed 622 and 379 immune-related genes that are closing and opening respectively (Fig. 2A). ClueGO(28) enrichment analyses revealed that chromatin closing is significantly associated with genes involved in T cell activation-related GO terms (n=161 genes) and T cell receptor signaling pathway (n=59 genes) (Fig. 2B, fig. S2A, table S4). In contrast, chromatin opening is associated significantly with genes involved in myeloid leukocyte (n=48 genes) and osteoclast differentiation (n=29 genes) processes (Fig. 2B, fig. S2B, table S4).

To further interpret the immunological implications of these chromatin changes, we applied transcriptional modules (25), where each module represents a coordinately expressed gene set across many PBMC expression profiles. These modules are functionally characterized and linked to pathways or cell types involved in immune processes which enabled us to systematically define transcriptional fingerprints of diverse immune diseases and responses (23, 29). We further characterized the modules with unknown functionality using ClueGO (28) annotations (Methods, table S5) and calculated the average fold change for immune module genes, where negative numbers represent chromatin closing with aging (blue bars in Fig. 2C) and positive ones represent chromatin opening with aging (red bars in Fig. 2C). This analysis revealed a systematic loss of chromatin accessibility at the promoters and enhancers of immune module genes. This loss was accompanied by a gain of chromatin accessibility around these genes at loci classified as repressed or quiescent by the Roadmap (27) profiles (Fig. 2C), consistent with observed genome-wide patterns (Fig. 1E). Modular analyses of the chromatin accessibility data revealed that the T cells module genes exhibit the most significant chromatin closing with aging (Fig. 2C, first row). Figure 2D shows that at the subject level, most genes in the T cells module exhibited chromatin silencing with aging, resulting in a striking separation of subjects into their respective age groups based on the chromatin accessibility of these genes. Focusing on the genes within the Inflammation-I module revealed that aging has a dual effect on inflammation-related genes (Fig. 2E). A set of inflammation-related genes are repressed via chromatin closing mostly at their enhancers, including hypoxia-inducible factor *HIF1A*, which modulates hypoxia responses in immune cells (30). Meanwhile, a mutually exclusive gene set was associated with chromatin opening mostly at quiescent sites, including *DUSP10*, a molecule that is known to play an essential role in local and systemic inflammation (31). A gain in chromatin accessibility was not detected around the IL-6 molecule, a key mediator of systemic inflammation (32), suggesting that age-associated increases in serum IL-6 levels age-associated increase might originate from cells other than PBMCs (33). Thus, our analysis shows a widespread chromatin closing at promoters and enhancers related to immune functions, especially T cell functions.

Aging-associated gene expression and chromatin accessibility changes

To link aging-associated chromatin changes to transcript levels, RNA-seq profiles of PBMCs from 39 subjects (24 HY, 15 HO) were generated to match the ATAC-seq samples. Figure 3A shows a significant positive correlation between age-related changes in gene expression levels and chromatin accessibility at gene promoters ($r=0.34$, $p < 1e^{-16}$). We identified immune modules that undergo transcriptional, epigenetic, and concordant (i.e., transcription and chromatin accessibility are remodeled together in the same direction) changes with aging (Fig. 3B). Analysis of genes showing concordant remodeling (Fig. 3B, third column) revealed the most significant chromatin closing with aging as well as declines in gene expression in T cells module (fig. S3A and B). Many genes associated with T cell functions, including transcription factors (TF) involved in lymphocyte development and activation, such as *LEF1* and *TCF7*, exhibited correlated decreases in chromatin and expression profiles (Fig. 3C, fig. S3C and D). Meanwhile, other immune modules, most notably cytotoxic cells, were activated both at the chromatin and gene expression levels with aging (25) (Fig. 3B**,** third column red bars). These included activation of granzymes (*GZMH* and *GZMB*) and granulysin (*GNLY*) (Fig. 3, D and E), and might originate from natural killer cells (34) though multiple other cell types can express *GZMB* including plasmacytoid dendritic cells (35), CD4+ T cells (36, 37) and plasma cells(38).

Our analyses also revealed chromatin remodeling that was not accompanied by changes in gene expression (fig. S3E). For example, chromatin opening at genes associated with inflammation was not accompanied by changes in gene expression, suggesting that transcriptional activation of these genes might depend on additional stimuli. Thus the combined analyses of epigenome and transcriptome increased the power of each assay and enabled us to identify immune effector molecules that are activated/inactivated with aging. Moreover, chromatin accessibility profiles provided more precise and genome-wide information than expression data alone including epigenomic changes at promoters and enhancers.

Epigenomic silencing of T cell signaling pathways with aging

IL7R, a gene critical for lymphocyte development and healthy immune responses (39), was among the top genes linked to multiple closing peaks (n=12; Fig. 4A, fig. S4A and B). The loss of chromatin accessibility around the *IL7R* locus was also accompanied by aging-associated decreases in *IL7R* expression (Fig. 4B). Moreover, the *IL7R* expression and the chromatin accessibility of its promoter were significantly correlated at the individual level ($r = 0.59$, $p < 0.001$, Fig. 4C). Additional genes in the IL-7 signaling cascade, including *JAK1*, *JAK3*, *STAT5A* and *STAT5B* also exhibited closing in aged individuals (Fig. 4D-E), possibly explaining the impaired signaling and responses to IL-7 in the elderly (40). Moreover, these results revealed *IL7R* and the members of the IL-7 signaling pathway as potential biomarkers of healthy aging. Wikipathways enrichment analyses (41) of genes annotated with differentially open peaks confirmed systematic and significant chromatin closing of IL-7 signaling pathway genes and other signaling pathways including TCR, IL-2, and IL-9 signaling (Fig. 4F, fig. S4C and D, table S4). In addition, 27 out of 70 genes in the "Histone Modifications" pathway were also associated with significant chromatin closing with aging (fig. S4C, table S4). These included histone genes (e.g., *HIST1H3D, HIST1H3E, HIST4H4*) as well as histone modifiers (e.g., *EZH1, SETD7*), in alignment with the known reduced expression of core histones and disruption of histone modification patterns associated with cellular aging (10). Taken together, our results suggest that aging is associated with the chromatin closing of multiple pathways related to T cell signaling that might explain impaired T cell responses in elderly. Moreover, these results establish *IL7R* and the members of the IL-7 signaling pathway as potential biomarkers of healthy aging.

CD8+ T cells account for the PBMC aging signature

Flow cytometry data from 23 subjects (12 HY and 11 HO) revealed that the age-related decrease in *IL7R* expression was limited to CD8+ T cells (Fig. 5A and B, table S6), indicating that chromatin alterations captured in PBMCs were not equal across profiled T cell subsets. Therefore, we profiled the chromatin accessibility of sorted naïve and memory CD4+ and CD8+ T cells from eight donors (4 HY, 4 HO) (Methods). A similar number of open chromatin sites were captured in these four subsets (~45-50K peaks). However, each subset exhibited different changes in their open chromatin sites with aging. CD4+ T cells showed minimal chromatin remodeling with aging: 44 peaks in memory and 216 peaks in naïve CD4+ T cells. In contrast, CD8+ T cells showed extensive chromatin remodeling with aging (Fig. 5C**,** fig. S5A). Specifically, memory CD8+ T cells displayed 8,503 (19.7% of tested) differential peaks, whereas naïve CD8+ T cells displayed 2,925 (6.4% of tested) differential peaks. A recent study reported aging-associated closing of chromatin accessibility at gene promoters in central memory

and naïve CD8+ T cells (42). Our data align with this observation and suggest that chromatin at gene promoters close with aging both in memory and naïve CD8+ T cells, although this pattern is more evident in memory CD8+ T cells. More specifically, functional state annotations from Roadmap T cell datasets (27) revealed that chromatin closing in memory CD8+ T cells occurred mostly at promoters (>50%) and enhancers (>30%), while chromatin opening was associated less with promoters (~10%) and more with quiescent (~30%) and enhancer sites (~40%) (Fig. 5D), similar to the PBMC signature (Fig. 1E). In naïve CD8+ T cells, chromatin remodeling was mostly observed at enhancers, including >75% of closing peaks and >50% of opening enhancer peaks.

Comparing aging-induced chromatin remodeling in T cell subsets to that of PBMCs indicated that chromatin remodeling in PBMCs correlates positively with the chromatin remodeling of CD8+ T cell subsets (fig. S5B). These results suggest that CD4+ and CD8+ T cell populations go through very different cellular changes with aging. Analysis of gene promoters known to be expressed in memory and naïve CD4+ and CD8+ T cells revealed that the chromatin closing of promoters of genes encoding certain surface (e.g., CD28, IL7R) and signaling (STAT4) molecules was more pronounced in memory CD8+ T cells (Fig. 5E, fig. S5C). Moreover, chromatin closing in PBMCs around T cell signaling pathways, including the IL7 signaling pathway, mostly stemmed from memory CD8+ T cells (Fig. 5F), including the *IL7R* locus itself (Fig. 6A). These results identify memory CD8+ T cells as the subpopulation with the most profound chromatin remodeling with aging. Moreover, the silencing of promoters and enhancers in PBMCs likely stems from CD8+ T cells, even though CD8+ T cells typically constitute <10-15% of PBMCs.

We also profiled the chromatin accessibility of purified monocytes (n=20) and naïve B cells (n=7) and found that these cell types do not display significant chromatin accessibility changes with aging (figs. S5D), further confirming that aging differentially affects distinct immune cell subsets. Moreover, chromatin changes detected in purified immune subsets (i.e., memory and naïve CD8+ T cells) suggest that the aging signature of PBMC is not merely a consequence of cell composition changes with aging, in agreement with similar observations based on single-cell transcriptomic profiling (26).

Potential regulators of IL7R are silenced in memory CD8+ T cells

Our data revealed that whereas chromatin closing around the *IL7R* locus (promoters and enhancers) was observed in CD8+ T cells, both memory and naïve, chromatin closing at the *IL7R* promoter was specific to memory CD8+ T cells (Fig. 6A, gray bar showing the *IL7R* promoter peak). To uncover potential regulators of these changes, we analyzed transcription factor (TF) binding motifs located around the *IL7R* transcription start site (10 kb upstream, 1 kb downstream). After filtering out the TFs based on their expression in PBMCs, several TFs and TF families emerged including LEF1, ETS2, BACH1/2, JUN, NF-κB family and the STAT family (Fig. 6B, table S7). Among these, NF-κB, JUN, and STATs are known as rapid acting factors, which can be present in an inactive state not requiring protein synthesis to be activated. Enrichment of these TFs at the *IL7R* promoter suggests a role for these rapid-acting TFs in ensuring the rapid activation of *IL7R* and modulating IL7 responsiveness in T cells. In fact, NF-κB directly controls the expression of *IL7R* gene in T cells through an enhancer control region close to the promoter (43). Furthermore, the chromatin around the promoters of these factors,

e.g., NF-κB and STAT family members, closed with aging specifically in memory CD8+ T cells (Fig. 6C and E, fig. S6A). Our data and analyses suggest that these TFs are likely to play a role in regulating the activity of *IL7R* in T cells and lose their chromatin accessibility –hence functionality- with aging specifically in memory CD8+ T cells. In alignment with these findings, Moskowitz et. al. (42) also reported aging-related disruptions in TF binding patterns in CD8+ T cells.

To determine whether silencing of TF promoters might correspond to changes in TF binding activity, we conducted TF footprinting analyses using PBMCs and T cell ATAC-seq samples. After pooling the ATAC-seq samples by cell type and age group, and normalizing with respect to the library depth, significant TF footprint calls were obtained using the PIQ algorithm (44). These analyses showed that several TFs with significant footprints around the *IL7R* promoter, including RXRA, NFKB1, ETS1, and TCF7 lost their footprints with aging specifically in memory CD8+ T cells (fig. S6B). Globally also there was an aging-related decrease in TF footprinting calls for all TFs in memory CD8+ T cells (Fig. 6F), including footprints for NF-κB factors, STATs, and TFs with important roles in T cell functions (Fig. 6G, table S8). Taken together, these results indicate that memory CD8+ T cells undergo an aging-associated silencing of regulatory elements and interactions, as evident from the loss of chromatin accessibility at TF gene promoters and decrease in TF binding estimates.

Aging-specific chromatin accessibility profiles are not linked with CMV

In chronic infections, most notably with cytomegalovirus (CMV), CD8+ T cells enter an "exhausted" state of reduced functionality and stop responding to further stimulation (45, 46) and aging-related increase in CMV seropositivity is associated with increased mortality in elderly (47, 48). To study whether the observed aging-associated chromatin signature in PBMCs is attributable to CMV seropositivity, we measured CMV IgG antibody status in a subset of our cohort (n=26, 17 HY, 9 HO), of which n=21 (12 HY, 9 HO) also had ATAC-seq profiles. As reported (47, 48), we observed that more elderly subjects were CMV-positive (70%) compared to young subjects (45%) (table S9). Moreover, there was a significant correlation between an individual's age and their CMV antibody levels (correlation coefficient = 0.65, Fig. 7A). Both aging and CMV status were correlated with changes in PBMC composition, most notably with changes involving CD8+ T cell subpopulations (fig. S7A-C). The decrease of naïve CD8+ T cell number is more dependent on aging than on the CMV status (fig. S7A-C), as observed earlier in large cohorts (49). Differential chromatin accessibility analyses between CMV-positive and CMV-negative subjects revealed that CMV seropositivity by itself is not associated with significant chromatin remodeling (Fig. 7B) even when differential analyses are stratified by age group (fig. S7D). Moreover, CMV-related and aging-related changes were not correlated (correlation coefficient = -0.02, Fig. 7C). Principal variance component analysis (PVCA(50)) confirmed that the variation in aging-associated ATAC-seq peaks cannot be explained by CMV status or sex, whereas age contributed ~30% of the variation in these data (fig. S7E). Thus while elderly subjects are more likely to have higher CMV infection rates, CMV seropositivity is neither associated with significant chromatin accessibility changes nor with the epigenomic aging signature observed in PBMCs.

Aging-associated chromatin accessibility profiles are stable over seasons

Seasonal variation affect immune cell counts, gene expression patterns (51), and immune responses (52, 53). PVCA analyses showed that season is among the biggest factors introducing variation in the chromatin accessibility data (~20% of the variation) (fig. S7F), therefore season is used as a covariate in our models (Methods). We separately analyzed PBMC samples collected in June–November ("Summer") and December–May ("Winter"). By using "season" as a blocking factor in the GLM models, we defined aging-associated chromatin accessibility changes in summer (10 HY, 7 HO) and winter samples (15 HY, 12 HO). This analysis resulted in 8,744 ATAC-seq peaks that are differentially accessible between HY and HO in summer and winter samples (Fig. 7D), 87% of which were also captured as differential in the combined cohort. Similar to the global signature, closing peaks are mostly found at promoters and enhancers, and opening peaks are at quiescent and repressed sites (fig. S7G). Fold changes between HY and HO samples that are obtained from two seasons correlated highly with the fold changes obtained from the whole cohort, though winter samples had a slightly higher correlation score (Fig. 7E, r = 0.81 for summer, r = 0.86 for winter). These results indicate that the chromatin accessibility-aging signature is robust and is not significantly impacted by seasonal variation.

Discussion

Systems immunology approaches combining chromatin accessibility (ATAC-Seq) and gene expression (RNA-Seq) profiles revealed an aging-specific signature in human blood leukocytes that is mainly borne by memory CD8+ T cells. Regions associated with chromatin opening were stochastically distributed across the cohort, where they are observed in a single subject or a small number of subjects. This chromatin opening might be associated with the aging-related loss of histone proteins leaving short strands of DNA accessible to the Tn5 transposase (54). However, the functional significance of this observation remains to be determined. In contrast, a widespread loss of chromatin accessibility at promoters and enhancers, especially affecting T cell signaling pathways was observed across all subjects. Matching chromatin accessibility and gene expression (RNA-seq) by genomic region, i.e. promoter and transcript, and by subject, helped us identify immune effector molecules that exhibit concordant alterations in their expression and chromatin accessibility levels with aging. Among these molecules, a substantial number of genes related to T cell functions were silenced epigenomically and transcriptionally, notably the *IL7R* gene and other genes encoding IL-7 signaling pathway.

Chromatin accessibility profiles of purified cells revealed that this aging-related signature in PBMCs stems from CD8+ T cells, which together account for 10-15% of PBMCs. We find that memory CD8+ T cells undergo a widespread silencing of promoters with aging whereas naïve CD8+ exhibit such a loss mostly at enhancers, a dichotomy which remains to be understood. Parallels between PBMCs and memory CD8+ T cells were particularly notable for genes in the T cell signaling pathways, including IL-7 signaling, where a strong chromatin closing around the promoters of these genes were observed both in PBMC and memory CD8+ T cells. The alteration of the IL7 pathway might explain the loss of homeostatic proliferation of CD8+T cells (55) as well as their reduced antigen-driven proliferation thus curtailing responses to infectious agents and cancer cells. Memory CD8+ T cells also exhibited an epigenomic silencing of TF promoters and a loss in TF footprints with aging. This aging-associated loss of TF footprints included fast-acting factors such as NF-κB and STATs, which might play a role in regulating rapid T cell responses. This loss in TF footprints was particularly noticeable around the IL7R locus. A recent study similarly reported the aging-associated erosion of chromatin accessibility

around gene promoters in naïve and central memory CD8+ T cells and showed that aging disrupts the TF binding patterns in these cells (42).

The epigenomic signature of aging in PBMCs described here were robust and neither associated with CMV positivity nor seasonal variation. As part of this signature, *IL7R* emerged as a potential biomarker of reduced immune responses, where aging-associated loss of this molecule is observed at the chromatin, transcriptome, and protein level, especially in memory CD8+ T cells. Such biomarkers could be instrumental in identifying individuals who might benefit the most from therapies to rejuvenate the declining immune functions due to aging or diseases, such as HIV (56, 57). Indeed, clinical studies of recombinant human IL-7 (rhIL-7) suggested a rejuvenation of circulating T cell profile upon administration of rhIL-7 especially in individuals with limited naïve T cells and diminished TCR repertoire diversity as in the case of elderly (58, 59). Chromatin accessibility profiling of immune cells for individuals would also be helpful in quantifying whether rejuvenation therapies administrated on individuals, such as rhIL-7, are effective and leading to measurable genomic changes around relevant genes/pathways.

The demonstration that aging has a profound impact on the epigenomes of human CD8+ T cells as presented here and in Moskowitz et. al.(42) opens the door to profiling epigenomes of these cells in other clinical situations such as diseases or response to therapeutic intervention including vaccination. Moreover, the ability to capture cell-specific epigenomic changes by leukocyte chromatin accessibility profiling, establishes this systems immunology approach as an integral and powerful immune-monitoring tool for future studies.

Materials and Methods

Human subject recruitment

All studies were conducted following approval by the Institutional Review Board of UConn Health Center (IRB Number: 14-194J-3). Following informed consent, blood samples were obtained from 75 healthy young (HY, 22-40 yrs) and 26 healthy old (HO, 65+ yrs) volunteers residing in the Greater Hartford, CT, USA region using services of the UConn Center on Aging Recruitment and Community Outreach Research Core (http://health.uconn.edu/aging/research/research-cores/). Recruitment criteria were selected to identify healthy individuals who are experiencing "usual aging" and are thus representative of average health conditions of the population within the corresponding age groups(60). Selecting this type of cohort increases the generalizability of our studies and the likelihood that these findings can be translated to the general population (60). Subjects were carefully screened in order to exclude potentially confounding diseases and medications, as well as frailty. Individuals who reported chronic or recent (i.e., within two weeks) infections were also excluded. Subjects could have chronic diseases, but were excluded if the following were present: congestive heart failure, kidney disease (serum creatinine >1.2 mg/dl in men and >1.1 mg/dl in women), diabetes mellitus requiring medications, immunosuppressive disorders or the use of immunosuppressive agents including oral prednisone in doses >10 mg daily. Since declines in self-reported physical performance are highly predictive of frailty, subsequent disability and mortality (61), all subjects were questioned as to their ability to walk ¼ mile (or 2-3 city blocks). For those who self-reported an inability to walk ¼ mile (61), the "Timed Up and Go" (TUG) test was performed and measured as the time taken to stand up from the sitting position, walk 10 feet and return to sitting in the chair (62). Scoring TUG > 10 sec was considered an indication of increased frailty and resulted in exclusion from the study (63).

Cell Sorting and Phenotypic Analysis

PBMCs were isolated from fresh whole blood using Ficoll-Paque Plus (GE) density gradient centrifugation. For cell sortings, we used fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD27 (M-T271) (Biolegend) and

CD4 (RPA-T4), CD45RO (UCHL1), CD45RA (HI100), CD19 (HIB19), CD16 (B73.1), IgD (IA6-2), and CD11c (S-HCL-3) (BD Biosciences), and CD8 (SCF121Thy2D3) and CD19 (J3-119) (Beckman-Coulter). Naïve CD4 (CD4+CD8-CD45RO-CD45RA+), naïve CD8 (CD4-CD8+CD45RO-CD45RA+), memory CD4 (CD4+CD8-CD45RO+CD45RA-), and memory CD8 (CD4-CD8+CD45RO+CD45RA-) T cells were sorted from the CD19-CD16-CD11c- fraction (DUMP channel). Naïve B cells (CD19+IgD+CD27-) were sorted from the CD3-CD16-CD11c-) fraction (DUMP channel). Cell sorting was performed using FACSAria Fusion (BD). Monocytes were isolated from fresh PBMCs by positive selection using magnetic CD14 microbeads (Miltenyi Biotech). For phenotypic analysis, PBMCs were stained with fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD4 (RPA-T4), CD8 (SCF121Thy2D3), CD45RA (HI100), CD19 (HIB19), CD14 (MSE2), CCR7 (150503), and CD127 (HIL-7R-M21). For the analysis of the frequencies of Naïve T cells (CD45RA+CCR7+), Central Memory T cells (CM; CD45RA-CCR7+), Effector Memory T cells (EM; CD45RA-CCR7-), and Effector Memory RA (EMRA; CD45RA+CCR7-), B cells and Monocytes, PBMCs were stained with fluorochrome-labeled antibodies specific for CD3 (UCHT1), CD4 (RPA-T4), CD8 (SCF121Thy2D3), CD45RA (HI100), CD19 (HIB19), CD14 (MSE2), CCR7 (150503), and CD127 (HIL-7R-M21). The stained cells were acquired with BD Fortessa and analyzed with FlowJo software (TreeStar).

CMV-seropositivity measurements

Anti-CMV IgG titers were determined in frozen sera by commercially available enzyme-linked immunosorbent assay (ELISA) (Genway Biotech Inc, San Diego, CA) with an interassay coefficient of variance of 5.2%. A titer of 1.2 ELISA Units/ml or greater in a sample was predetermined by the manufacturer as CMV-seropositive.

ATAC-seq Library Generation and pre-processing

ATAC-seq was performed as previously described (21). Fifty thousand unfixed nuclei were tagged using Tn5 transposase (Illumina, Nextera DNA sample prep kit) for 30 min at 37°C and the resulting library fragments were purified using Qiagen MinElute kit (Qiagen). Libraries were amplified by 10-12 PCR cycles, purified using a Qiagen PCR cleanup kit (Qiagen), and finally sequenced on an Illumina HiSeq 2500 with a minimum read length of 75 bp to a minimum depth of 30 million reads per sample. At least two technical replicates (average = 2.4 replicates) were processed per biological sample. Table S10 summarizes the depth, peak number, and fragments in reads (FrIP) scores for ATAC-seq samples.

ATAC-seq sequences were quality-filtered using trimmomatic (64), and trimmed reads were mapped to the GRCh37 (hg19) human reference sequence using bwa-mem (65). After alignment, technical replicates were merged and all further analyses were carried out on these merged data. For peak calling, MACS2 (66) was used with no-model, 100bp shift, 200bp extension, and broad peaks options. Only peaks called with a peak score (q-value) of 1% or better were kept from each sample, and the selected peaks were merged into a consensus peak set using Bedtools *multiinter* tool (67). Only peaks called on autosomal chromosomes were used in this study. We further filtered consensus peaks to avoid likely false positives by only including those peaks overlapping more than 20 short reads in at least one sample, and peaks for which the maximum read count did not exceed 500 counts per million (cpm) to account for regions that are potential artifacts. Finally, we excluded peaks overlapping blacklisted regions as defined by ENCODE mappability criteria developed for DNase assays (July 2015 version), downloaded from *http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeMapability/*.

An additional quality-control step was developed to filter out samples with a consistently poor signal, consisting of an algorithm to discover and characterize a series of relatively invariant *benchmark peaks*, defined as a set of peaks expected to be called in all samples. Samples that consistently miss calls for a significant portion of these benchmark peaks are flagged as having poor quality. A benchmark peak is defined based on three criteria, namely (1) that it remains approximately invariant between the two groups of interest (i.e., young and old samples), (2) that it captures a substantial number of reads, and (3) that it is called in most samples. For each peak, the absolute value of the log of the ratio of healthy old to healthy young mean normalized read counts (log fold change, logFC) was used to assess the first criteria, whereas the maximum read count over all samples (maxCt) is used to assess the second one. In this study, a peak was considered apt for benchmarking when (1) its absolute logFC was in the bottom decile of the distribution over all peaks, (2) its maxCt was in the top decile of the distribution over all peaks, and (3) the peak was called in at least 90% of the samples. Using these parameters, 273 (out of 169,636) peaks were selected as

benchmark; only samples for which at least 95% of these peaks were called were selected for analyses, which excluded 5 samples from further analyses. We examined the effects of each of these parameter choices and found that the same samples were consistently chosen as poor quality for a range of values chosen to assess the benchmark criteria. Prior to statistical analyses, ATAC-seq read counts were normalized to each sample's effective library size (i.e., the sum of reads overlapping peaks) using the trimmed mean of M-values normalization method (TMM) (68).

RNA-seq Library Generation and pre-processing

Total RNA was isolated from PBMCs using the Qiagen RNeasy (Qiagen) or Arcturus PicoPure (Life Technologies) kits following manufacturer's protocols. During RNA isolation, DNase treatment was additionally performed using the RNase-free DNase set (Qiagen). RNA quality was checked using an Agilent 2100 Expert bioanalyzer (Agilent Technologies). RNA quality was reported as a score from 1 to 10, samples falling below threshold of 8.0 being omitted from the study. cDNA libraries were prepared using either the TruSeq Stranded Total RNA LT Sample Prep Kit with Ribo-Zero Gold (Illumina) or KAPA Stranded mRNA-Seq Library Prep kit (KAPA Biosytems) according to the manufacturer's instructions using 100ng or 500ng of total RNA. Final libraries were analyzed on a Bioanalyzer DNA 1000 chip (Agilent Technologies). Paired-end sequencing (2x100bp) of stranded total RNA libraries was carried out in either Illumina NextSeq500 using v2 sequencing reagents or the HiSeq2500 using SBS v3 sequencing reagents.
Quality control (QC) of the raw sequencing data was performed using the FASTQC tool, which computes read quality using summary of per-base quality defined using the probability of an incorrect base call (69). According to our quality criteria, reads with more than 30% of their nucleotides with a Phred score under 30 are removed, whereas samples with more than 20% of such low quality reads are dropped from analyses. None of the samples used in this study were dropped after QC. Reads from samples that pass the quality criteria were quality-trimmed and filtered using trimmomatic (64). High-quality reads were then used to estimate transcript abundance using RSEM (70). Finally, to minimize the interference of non-messenger RNA in our data, estimate read counts were re-normalized to include only protein-coding genes. Table S11 summarizes the depth and alignment rate of our PBMC RNA-seq samples.

Differential analysis

To identify differentially open chromatin regions from ATAC-seq and differentially expressed genes from RNA-seq data, the *R* package edgeR was used to fit a generalized linear model (GLM) to test for the effect of aging between healthy young and healthy old samples. In addition to age group (old vs. young), our models included sex and the season in which the sample was collected (summer vs. winter) as covariates (71), since it was determined, using Principal Variance Component Analysis (PVCA, (50)), that these factors account for a sizeable fraction of the variance in read counts. Furthermore, we used Surrogate Variable Analysis (SVA (72)) to capture unknown sources of variation (e.g., batch effects, subject-level heterogeneity) statistically independent from age group assignments. Using the built-in permutation-based procedure in the R package sva, we choose to retain three SVs to include as covariates in the GLM model for PBMC ATAC-seq and RNA-seq data analyses (22). Within GLM models, a negative binomial link function was used, including both genome-wide and peak-specific dispersion parameters, estimated using edgeR's "common," "trended," and "tagwise" dispersion components, calculated using a robust estimation option. Benjamini-Hochberg P-value correction was used to select differentially open peaks at a False Discovery Rate (FDR) of 5%. To generate a set of model-adjusted peak estimates of chromatin accessibility (i.e., sex-, season-, and SV-adjusted) for downstream analyses and visualization, we used edgeR to fit a "null" model excluding the age group factor, and then subtracted the resulting fitted values from this model from the original TMM-normalized reads.

An equivalent approach was used to analyze the effects of CMV seropositivity and seasonal variation (i.e., winter- vs. summer-acquired samples) in PBMC data. For CMV analysis, the subset of samples for which this information was available (i.e., N=21, 12 HY and 9 HO) was fit to a model including a sex as a factor and CMV status (positive, negative) as a blocking factor. In this analysis, the season factor was not taken into consideration since all subjects for whom CMV status was available were collected in the same season. For seasonal analysis, we used season (summer, winter) as a blocking factor. In both analyses, we tested both separately and jointly for the significance of age group by CMV status or season. In addition, we fitted the converse models (CMV status or aging nested within

age group) to test for and calculate fold change estimates for CMV+/CMV- and winter/summer stratified by age group.

Peak annotation and downstream analyses

Multiple data sources were used to annotate ATAC-seq peaks with regard to functional and positional information. HOMER(73) was used to annotate peaks as "promoter" (i.e., within 1 kb of known TSS), "intergenic", "intronic", and other positional categories. For functional annotation of peaks, we implemented a simplified version of the 18-state ChromHMM-derived chromatin states obtained from Roadmap Epigenomics data for PBMC and T cell subsets(74): we first intersected the Roadmap-generated states with our set of consensus peaks, and solved conflicting cases where multiple chromatin states overlap the same ATAC-seq peak so that each peak was assigned a single annotation, according to the following priority rules: Active TSS > Active Enhancer 1 > Active Enhancer 2 > Genic Enhancer 1 > Genic Enhancer 2 > Weak Enhancer > Strong transcription > Flanking Active TSS > Flanking Upstream Active TSS > Flanking Downstream Active TSS > Weak Transcription > Bivalent Poised TSS > Bivalent Enhancer > Weakly Repressed PolyComb > Repressed Polycomb > ZNF Genes and repeats > Heterochromatin > Quiescent/Low signal. Then, to facilitate interpretation and visualization, we simplified the set of 18 chromatin states to a scheme with 6 pooled meta-states, namely (1) TSS, collecting active, flanking, and bivalent TSS states; (2) Enhancer, pooling active, weak and bivalent enhancer states; (3) Repressed PolyComb, combining both weak and strong PolyComb states; (4) Transcription, including both weak and strong transcription states, (5) the quiescent chromHMM state; and (6) other states (ZNF, heterochromatin) combined together.

ATAC-seq peaks were also annotated using gene sets provided by curated immune function-related co-expression modules(25). We used these annotations to test for enrichment of modules in a variety of gene sets of interest, such as genes associated to closing/opening peaks. We assessed enrichment using the hypergeometric test, with a background defined by the set of genes that are expressed, as determined by RNA-seq data, or potentially expressed, as given by promoter accessibility, in the appropriate cell type. In addition, we summarized the representation of GO terms among gene annotations for all peaks, after solving for multiple GO annotations for the same gene by prioritizing terms according to the order: Immunity > Metabolic > Transcription, Translation > Migration > Mitochondria > Axon > Development.

Further functional enrichment analyses were carried out using ClueGO (28) to test for overrepresentation of GO:Immune System Process terms using GO term fusion option and Wikipathways pathways(41) among genes associated to differentially open peaks. In addition to testing for enriched gene sets, ClueGO combines GO terms and pathways into functionally relevant meta-sets based on the rate of shared genes among terms, allowing for an efficient assessment of enriched categories, as well as their potential interactions, as inferred from sets of shared genes. We applied these methods separately to peaks significantly closing and opening between age groups to investigate the degree to which these two sets of peaks are associated to unique signatures. We only listed terms that are significant at p-value 0.05 after Bonferroni step down correction. In addition, we used ClueGO to annotate the aforementioned immunological co-expression modules that were originally associated to unknown function. Most salient enriched functional categories for these modules are listed in Sup Table 5. Visualization of signaling pathways were generated ClueGO and PathVisio (75) tools.

Congruence between chromatin accessibility and transcription data

Gene expression (mRNA-seq, see above) data was generated for a subset of subjects with ATAC-seq profiles (n=39, 24 HY and 15 HO). These data were normalized to protein-coding transcripts, and annotated to ENSEMBL GRCh37 gene symbols. Genes for which at least three normalized reads per million were obtained in at least two samples were considered as expressed, all others removed prior to analysis. This resulted in a total estimate of 11,311 expressed genes in PBMCs.We built a data set comprising paired ATAC-seq and RNA-seq samples by matching promoter peaks to nearest gene (TSS) annotations. First, we retrieved the complete list of refSeq TSS coordinates for the hg19 genome reference (n=34,783), and defined promoters as the regions within 1000 bp flanks of each TSS. The final set of promoters was defined by merging overlapping flanked TSS regions annotated to the same gene (n=34,700). We then selected ATAC-seq peaks overlapping these promoters and annotated them to the corresponding gene. Only the peak closest to the TSS was kept. Finally, the resulting data set was filtered as to only

include promoter peaks for genes that were transcribed, as defined above. Whenever multiple expressed genes were matched to the same promoter peak, all of them were retained for analysis.

To study the concordance between promoter accessibility and gene expression, we subdivided the space defined by aging-related fold changes derived from ATAC-seq and RNA-seq data into gene sets defined by the direction and magnitude of change along both dimensions, such as genes with both up-regulated expression and increasing accessibility in elderly subjects, or genes for which expression is up-regulated but accessibility remains unchanged with aging. In order to capture enough genes to enable functional enrichment analysis of these gene sets, fold changes between healthy old and young subjects for matching promoter peaks and transcripts were estimated empirically as the difference between the mean normalized values of each group, and plotted against each other (see Fig. 3A). For each gene set, we tested for enrichment in immune modules and Wikipathways pathways and compared these results to gene sets constructed based solely on significant differential accessibility or expression as determined using GLM as previously described. Specifically, we defined a gene or promoter as being significantly "up" or "down" if the empirical log fold change of the HO mean relative to HY mean was above or below zero, respectively, and if the adjusted empirical p-value < 0.01 for that gene. Empirical p-values were computed by randomly permuting the HO and HY sample labels 1,000 times for each promoter peak and gene. Genes for which $p < 0.01$ were considered significantly different between aging groups, whereas all others were considered to have "stable" expression and/or accessibility relative to aging. Here we focus on a subset of the combined accessibility-expression gene sets generated by this method, namely, (1) genes with both increased or (2) both decreased promoter accessibility and expression with aging, and (3) genes with increased or (4) decreased promoter accessibility but stable aging-related expression.

Transcription Factor (TF) motif and footprinting analysis

ATAC-seq data from PBMC and T cells were scanned for TF footprints using the PIQ algorithm (44). This method integrates genome-wide TF motifs (i.e., position weight matrices or PWMs) with chromatin accessibility estimates profiled at base pair resolution to generate a list of possible footprint matches for a motif. The method also produces a probability estimate for each footprint's reproducibility, termed "purity score". Here, we compiled a set of 1,273 distinct motifs comprising the curated (CORE) list available in the JASPAR 2016 database (n=466, http://jaspar.genereg.net) in addition to the complete set of HT-SELEX motifs made available in (76) (n=819). Altogether, these motifs represent binding sites for 381 distinct TF Prior to footprint calling, we merged samples belonging to the same cell type and age group to maximize our ability to find highly reproducible footprints. In addition, we used SAMtools v. 0.1.19 (65) to randomly subsample aligned reads from each merged data set to approximately match the mapped library depth of the least deeply sequenced sample, i.e., 113 Mb. This normalization step is included to minimize the impact of the high correlation observed between library depth and footprint purity scores. Only footprints with a purity score of 90% or more were retained for further analysis. Finally, footprint calls were further filtered to include in analyses only those associated to TFs determined that are expressed in immune cells.

To examine the regulatory landscape of *IL7R*, a potential aging biomarker, we focused on footprints called on the promoter region (+/-1 kb from TSS) of this gene separately by age group and cell type. To complement this set of footprints, we also carried out de novo motif discovery using HOMER by searching for motifs enriched in peaks annotated to *IL7R* relative to all peaks in PBMCs and T cell subsets. Each enriched motif was annotated to the best fitting known TF, as found by HOMER, with the added requirement that the annotated TF should be expressed in the appropriate cell type. We then used PIQ to call footprints of the enriched motifs, and combined those overlapping *IL7R* promoter with the previously selected footprints. Finally, in addition to footprint and motif enrichment analyses, known TF motifs were retrieved for the region around *IL7R* TSS (-10kb upstream, +1kb downstream) using MotifMap tool(77) at a 20% FDR.

Supplementary Materials

Sup Fig. 1. Cell composition changes with aging.
Sup Fig. 2. Immune genes are affected by chromatin remodeling.
Sup Fig. 3. T cell module is epigenetically and transcriptionally repressed with aging.
Sup Fig. 4. TCR signaling is epigenetically repressed with aging.
Sup Fig. 5. Memory CD8+ T cells are the largest contributors to the PBMC aging signature.

Sup Fig. 6. Aging-related loss of TF activity in memory CD8+ T cells.
Sup Fig. 7. CMV and season-related chromatin changes in PBMCs.

Table S1. Cohort information. Detailed sample metadata, including samples used in ATAC-seq and RNA-seq analyses, as well as samples used as source of other types of data (cell composition, CMV status).

Table S2. Cell composition. Proportion values of sorted T cell and major cell populations (CD4+, CD8+, CD14+, CD19+) in PBMC extracts from subjects included in this study.

Table S3. Genes associated with differentially (A) closing and (B) opening ATAC-seq peaks.

Table S4. Enrichment of immune GO terms and Wikipathways pathways in differentially closing and opening ATAC-seq peaks. Results from ClueGO analysis.

Table S5. Enrichment of immune GO, biological process GO, Wikipathways pathways, and KEGG pathways in immune modules annotated as unknown in Figs. 2C and 3B.

Table S6. Fraction of IL7R+ cells among CD4+, CD8+, CD19+, and CD14+ healthy young and old samples, showing that a decrease in this fraction correlates to age only in CD8+ compartment cells.

Table S7. TF motifs found near IL7R promoters. List built using MotifMap at a 20% FDR level.

Table S8. Genome-wide fraction of footprints called for each TF motif in healthy young and old subjects, in each T cell subset. Only footprints overlapping gene promoters are included in calculations. Proportion is computed over all footprint locations called in either or in both merged HY and HO samples. Footprints called separately in each cell subpopulation.

Table S9. Sample information used for determination of cytomegalovirus (CMV) status.

Table S10. ATAC-seq and RNA-seq data quality metrics for samples analyzed in this study.

REFERENCES

1. Pawelec, G., *Immunosenescence*. 2008: Springer Science & Business Media.
2. Goronzy, J.J., G. Li, Z. Yang, and C.M. Weyand, *The janus head of T cell aging - autoimmunity and immunodeficiency*. Front Immunol, 2013. **4**: p. 131.PMCID: 3671290
3. Goronzy, J.J. and C.M. Weyand, *Understanding immunosenescence to improve responses to vaccines*. Nature immunology, 2013. **14**(5): p. 428-436
4. Franceschi, C., M. Bonafe, S. Valensin, F. Olivieri, M. De Luca, E. Ottaviani, and G. De Benedictis, *Inflamm-aging. An evolutionary perspective on immunosenescence.* Ann N Y Acad Sci, 2000. **908**: p. 244-54
5. Pawelec, G., D. Goldeck, and E. Derhovanessian, *Inflammation, ageing and chronic disease.* Current opinion in immunology, 2014. **29**: p. 23-28
6. Dorshkind, K., E. Montecino-Rodriguez, and R.A. Signer, *The ageing immune system: is it ever too old to become young again?* Nat Rev Immunol, 2009. **9**(1): p. 57-62
7. Shaw, A.C., D.R. Goldstein, and R.R. Montgomery, *Age-dependent dysregulation of innate immunity*. Nat Rev Immunol, 2013. **13**(12): p. 875-87.PMCID: 4096436
8. Tchkonia, T., Y. Zhu, J. Van Deursen, J. Campisi, and J.L. Kirkland, *Cellular senescence and the senescent secretory phenotype: therapeutic opportunities.* The Journal of clinical investigation, 2013. **123**(3): p. 966-972
9. Rando, T.A. and H.Y. Chang, *Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock.* Cell, 2012. **148**(1-2): p. 46-57.PMCID: 3336960

10. Benayoun, B.A., E.A. Pollina, and A. Brunet, *Epigenetic regulation of ageing: linking environmental inputs to genomic stability.* Nature Reviews Molecular Cell Biology, 2015. **16**(10): p. 593-610
11. López-Otín, C., M.A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, *The hallmarks of aging.* Cell, 2013. **153**(6): p. 1194-1217
12. Harries, L.W., D. Hernandez, W. Henley, A.R. Wood, A.C. Holly, R.M. Bradley-Smith, H. Yaghootkar, A. Dutta, A. Murray, T.M. Frayling, J.M. Guralnik, S. Bandinelli, A. Singleton, L. Ferrucci, and D. Melzer, *Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing.* Aging Cell, 2011. **10**(5): p. 868-78.PMCID: 3173580
13. Reynolds, L.M., J. Ding, J.R. Taylor, K. Lohman, N. Soranzo, et al., *Transcriptomic profiles of aging in purified human immune cells.* BMC Genomics, 2015. **16**: p. 333.PMCID: 4417516
14. Cao, J.N., S. Gollapudi, E.H. Sharman, Z. Jia, and S. Gupta, *Age-related alterations of gene expression patterns in human CD8+ T cells.* Aging Cell, 2010. **9**(1): p. 19-31
15. Horvath, S., Y. Zhang, P. Langfelder, R.S. Kahn, M. Boks, K. van Eijk, L.H. van den Berg, and R.A. Ophoff, *Aging effects on DNA methylation modules in human brain and blood tissue.* Genome Biol, 2012. **13**(10): p. R97
16. Martino, D.J., M.K. Tulic, L. Gordon, M. Hodder, T.R. Richman, J. Metcalfe, S.L. Prescott, and R. Saffery, *Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans.* Epigenetics, 2011. **6**(9): p. 1085-1094
17. Rakyan, V.K., T.A. Down, S. Maslau, T. Andrew, T.-P. Yang, H. Beyan, P. Whittaker, O.T. McCann, S. Finer, and A.M. Valdes, *Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains.* Genome research, 2010. **20**(4): p. 434-439
18. Yuan, T., Y. Jiao, S. de Jong, R.A. Ophoff, S. Beck, and A.E. Teschendorff, *An integrative multi-scale analysis of the dynamic DNA methylation landscape in aging.* PLoS Genet, 2015. **11**(2): p. e1004996.PMCID: 4334892
19. Tserel, L., R. Kolde, M. Limbach, K. Tretyakov, S. Kasela, K. Kisand, M. Saare, J. Vilo, A. Metspalu, L. Milani, and P. Peterson, *Age-related profiling of DNA methylation in CD8+ T cells reveals changes in immune response and transcriptional regulator genes.* Sci Rep, 2015. **5**: p. 13107.PMCID: 4541364
20. Zheng, Y., B.T. Joyce, E. Colicino, L. Liu, W. Zhang, Q. Dai, M.J. Shrubsole, W.A. Kibbe, T. Gao, and Z. Zhang, *Blood Epigenetic Age may Predict Cancer Incidence and Mortality.* EBioMedicine, 2016
21. Buenrostro, J.D., P.G. Giresi, L.C. Zaba, H.Y. Chang, and W.J. Greenleaf, *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013. **10**(12): p. 1213-8.PMCID: 3959825
22. Qu, K., L.C. Zaba, P.G. Giresi, R. Li, M. Longmire, Y.H. Kim, W.J. Greenleaf, and H.Y. Chang, *Individuality and Variation of Personal Regulomes in Primary Human T Cells.* Cell Systems, 2015. **1**(1): p. 51-61
23. Berry, M.P., C.M. Graham, F.W. McNab, Z. Xu, S.A. Bloch, T. Oni, K.A. Wilkinson, R. Banchereau, J. Skinner, and R.J. Wilkinson, *An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.* Nature, 2010. **466**(7309): p. 973-977
24. Banchereau, R., S. Hong, B. Cantarel, N. Baldwin, J. Baisch, M. Edens, A.-M. Cepika, P. Acs, J. Turner, and E. Anguiano, *Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients.* Cell, 2016
25. Chaussabel, D., C. Quinn, J. Shen, P. Patel, C. Glaser, N. Baldwin, D. Stichweh, D. Blankenship, L. Li, and I. Munagala, *A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus.* Immunity, 2008. **29**(1): p. 150-164
26. Kowalczyk, M.S., I. Tirosh, D. Heckl, T.N. Rao, A. Dixit, B.J. Haas, R.K. Schneider, A.J. Wagers, B.L. Ebert, and A. Regev, *Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells.* Genome research, 2015. **25**(12): p. 1860-1872
27. Roadmap Epigenomics, C., A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30
28. Bindea, G., B. Mlecnik, H. Hackl, P. Charoentong, M. Tosolini, A. Kirilovsky, W.H. Fridman, F. Pages, Z. Trajanoski, and J. Galon, *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-3.PMCID: 2666812
29. Guiducci, C., M. Gong, Z. Xu, M. Gill, D. Chaussabel, T. Meeker, J.H. Chan, T. Wright, M. Punaro, and S. Bolland, *TLR recognition of self nucleic acids hampers glucocorticoid activity in lupus.* Nature, 2010. **465**(7300): p. 937-941

30. Palazon, A., A.W. Goldrath, V. Nizet, and R.S. Johnson, *HIF transcription factors, inflammation, and immunity.* Immunity, 2014. **41**(4): p. 518-28.PMCID: 4346319
31. Lang, R., M. Hammer, and J. Mages, *DUSP meet immunology: dual specificity MAPK phosphatases in control of the inflammatory response.* J Immunol, 2006. **177**(11): p. 7497-504
32. Ershler, W.B. and E.T. Keller, *Age-associated increased interleukin-6 gene expression, late-life diseases, and frailty.* Annu Rev Med, 2000. **51**: p. 245-70
33. Maggio, M., J.M. Guralnik, D.L. Longo, and L. Ferrucci, *Interleukin-6 in aging and chronic disease: a magnificent pathway.* The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2006. **61**(6): p. 575-584
34. Hayhoe, R.P., S.M. Henson, A.N. Akbar, and D.B. Palmer, *Variation of human natural killer cell phenotypes with age: identification of a unique KLRG1-negative subset.* Hum Immunol, 2010. **71**(7): p. 676-81
35. Matsui, T., J.E. Connolly, M. Michnevitz, D. Chaussabel, C.I. Yu, C. Glaser, S. Tindle, M. Pypaert, H. Freitas, B. Piqueras, J. Banchereau, and A.K. Palucka, *CD2 distinguishes two subsets of human plasmacytoid dendritic cells with distinct phenotype and functions.* J Immunol, 2009. **182**(11): p. 6815-23.PMCID: 2749454
36. Appay, V., J.J. Zaunders, L. Papagno, J. Sutton, A. Jaramillo, A. Waters, P. Easterbrook, P. Grey, D. Smith, A.J. McMichael, D.A. Cooper, S.L. Rowland-Jones, and A.D. Kelleher, *Characterization of CD4(+) CTLs ex vivo.* J Immunol, 2002. **168**(11): p. 5954-8
37. Namekawa, T., U.G. Wagner, J.J. Goronzy, and C.M. Weyand, *Functional subsets of CD4 T cells in rheumatoid synovitis.* Arthritis Rheum, 1998. **41**(12): p. 2108-16
38. Xu, W., P. Narayanan, N. Kang, S. Clayton, Y. Ohne, P. Shi, M.C. Herve, R. Balderas, C. Picard, J.L. Casanova, J.P. Gorvel, S. Oh, V. Pascual, and J. Banchereau, *Human plasma cells express granzyme B.* Eur J Immunol, 2014. **44**(1): p. 275-84
39. Schluns, K.S., W.C. Kieper, S.C. Jameson, and L. Lefrancois, *Interleukin-7 mediates the homeostasis of naive and memory CD8 T cells in vivo.* Nat Immunol, 2000. **1**(5): p. 426-32
40. Kim, H.R., M.S. Hong, J.M. Dan, and I. Kang, *Altered IL-7Ralpha expression with aging and the potential implications of IL-7 therapy on CD8+ T-cell immune responses.* Blood, 2006. **107**(7): p. 2855-62.PMCID: 1440715
41. Kelder, T., M.P. van Iersel, K. Hanspers, M. Kutmon, B.R. Conklin, C.T. Evelo, and A.R. Pico, *WikiPathways: building research communities on biological pathways.* Nucleic acids research, 2012. **40**(D1): p. D1301-D1307
42. Moskowitz, D.M., D.W. Zhang, B. Hu, S. Le Saux, R.E. Yanes, Z. Ye, J.D. Buenrostro, C.M. Weyand, W.J. Greenleaf, and J.J. Goronzy, *Epigenomics of human CD8 T cell differentiation and aging.* Science Immunology, 2017. **2**(8): p. eaag0192
43. Miller, M.L., M. Mashayekhi, L. Chen, P. Zhou, X. Liu, M. Michelotti, N. Tramontini Gunn, S. Powers, X. Zhu, C. Evaristo, M.L. Alegre, and L.L. Molinero, *Basal NF-kappaB controls IL-7 responsiveness of quiescent naive T cells.* Proc Natl Acad Sci U S A, 2014. **111**(20): p. 7397-402.PMCID: 4034246
44. Sherwood, R.I., T. Hashimoto, C.W. O'Donnell, S. Lewis, A.A. Barkal, J.P. van Hoff, V. Karun, T. Jaakkola, and D.K. Gifford, *Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape.* Nature biotechnology, 2014. **32**(2): p. 171-178
45. Sansoni, P., R. Vescovini, F.F. Fagnoni, A. Akbar, R. Arens, Y.-L. Chiu, L. Čičin-Šain, J. Dechanet-Merville, E. Derhovanessian, and S. Ferrando-Martinez, *New advances in CMV and immunosenescence.* Experimental gerontology, 2014. **55**: p. 54-62
46. Wherry, E.J., *T cell exhaustion.* Nat Immunol, 2011. **12**(6): p. 492-9
47. Fülöp, T., A. Larbi, and G. Pawelec, *HumanT cell aging and the impact of persistent viral infections.* How aging affects T lymphocyte-mediated immunity, 2015: p. 62
48. Savva, G.M., A. Pachnio, B. Kaul, K. Morgan, F.A. Huppert, C. Brayne, and P.A. Moss, *Cytomegalovirus infection is associated with increased mortality in the older population.* Aging cell, 2013. **12**(3): p. 381-387
49. Wertheimer, A.M., M.S. Bennett, B. Park, J.L. Uhrlaub, C. Martinez, V. Pulko, N.L. Currier, D. Nikolich-Žugich, J. Kaye, and J. Nikolich-Žugich, *Aging and cytomegalovirus infection differentially and jointly affect distinct circulating T cell subsets in humans.* The Journal of Immunology, 2014. **192**(5): p. 2143-2155
50. Boedigheimer, M.J., R.D. Wolfinger, M.B. Bass, P.R. Bushel, J.W. Chou, M. Cooper, J.C. Corton, J. Fostel, S. Hester, J.S. Lee, F. Liu, J. Liu, H.R. Qian, J. Quackenbush, S. Pettit, and K.L. Thompson,

Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. BMC Genomics, 2008. **9**: p. 285.PMCID: 2453529

51. Dopico, X.C., M. Evangelou, R.C. Ferreira, H. Guo, M.L. Pekalski, D.J. Smyth, N. Cooper, O.S. Burren, A.J. Fulford, B.J. Hennig, A.M. Prentice, A.G. Ziegler, E. Bonifacio, C. Wallace, and J.A. Todd, *Widespread seasonal gene expression reveals annual differences in human immunity and physiology.* Nat Commun, 2015. **6**: p. 7000.PMCID: 4432600
52. Aguirre-Gamboa, R., I. Joosten, P.C. Urbano, R.G. van der Molen, E. van Rijssen, et al., *Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits.* Cell Rep, 2016. **17**(9): p. 2474-2487.PMCID: 5130901
53. Ter Horst, R., M. Jaeger, S.P. Smeekens, M. Oosting, M.A. Swertz, et al., *Host and Environmental Factors Influencing Individual Human Cytokine Responses.* Cell, 2016. **167**(4): p. 1111-1124 e13
54. Pal, S. and J.K. Tyler, *Epigenetics and aging.* Science advances, 2016. **2**(7): p. e1600584
55. Briceño, O., A. Lissina, K. Wanke, G. Afonso, A. Braun, K. Ragon, T. Miquel, E. Gostick, L. Papagno, and K. Stiasny, *Reduced naïve CD8+ T–cell priming efficacy in elderly adults.* Aging cell, 2016. **15**(1): p. 14-21
56. Feinberg, A.P., *Phenotypic plasticity and the epigenetics of human disease.* Nature, 2007. **447**(7143): p. 433-40
57. Kennedy, B.K., S.L. Berger, A. Brunet, J. Campisi, A.M. Cuervo, E.S. Epel, C. Franceschi, G.J. Lithgow, R.I. Morimoto, J.E. Pessin, T.A. Rando, A. Richardson, E.E. Schadt, T. Wyss-Coray, and F. Sierra, *Geroscience: linking aging to chronic disease.* Cell, 2014. **159**(4): p. 709-13
58. Sportes, C., R.R. Babb, M.C. Krumlauf, F.T. Hakim, S.M. Steinberg, C.K. Chow, M.R. Brown, T.A. Fleisher, P. Noel, I. Maric, M. Stetler-Stevenson, J. Engel, R. Buffet, M. Morre, R.J. Amato, A. Pecora, C.L. Mackall, and R.E. Gress, *Phase I study of recombinant human interleukin-7 administration in subjects with refractory malignancy.* Clin Cancer Res, 2010. **16**(2): p. 727-35.PMCID: 2808195
59. Sportes, C., F.T. Hakim, S.A. Memon, H. Zhang, K.S. Chua, et al., *Administration of rhIL-7 in humans increases in vivo TCR repertoire diversity by preferential expansion of naive T cell subsets.* J Exp Med, 2008. **205**(7): p. 1701-14.PMCID: 2442646
60. Robertson, D. and G.H. Williams, *Clinical and translational science: principles of human research*. 2009: Academic Press.
61. Hardy, S.E., Y. Kang, S.A. Studenski, and H.B. Degenholtz, *Ability to walk 1/4 mile predicts subsequent disability, mortality, and health care costs.* J Gen Intern Med, 2011. **26**(2): p. 130-5.PMCID: 3019329
62. Podsiadlo, D. and S. Richardson, *The timed "Up & Go": a test of basic functional mobility for frail elderly persons.* Journal of the American geriatrics Society, 1991. **39**(2): p. 142-148
63. Rockwood, K., E. Awalt, D. Carver, and C. MacKnight, *Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging.* J Gerontol A Biol Sci Med Sci, 2000. **55**(2): p. M70-3
64. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014: p. btu170
65. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-1760
66. Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, and W. Li, *Model-based analysis of ChIP-Seq (MACS).* Genome biology, 2008. **9**(9): p. R137
67. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.PMCID: 2832824
68. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data.* Genome Biol, 2010. **11**(3): p. R25.PMCID: 2864565
69. Ewing, B., L. Hillier, M.C. Wendl, and P. Green, *Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment.* Genome research, 1998. **8**(3): p. 175-185
70. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC bioinformatics, 2011. **12**(1): p. 1
71. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-140
72. Leek, J.T., W.E. Johnson, H.S. Parker, A.E. Jaffe, and J.D. Storey, *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.PMCID: 3307112

73. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass, *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.PMCID: 2898526
74. Kundaje, A., W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, and M.J. Ziller, *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-330
75. Kutmon, M., M.P. van Iersel, A. Bohler, T. Kelder, N. Nunes, A.R. Pico, and C.T. Evelo, *PathVisio 3: an extendable pathway analysis toolbox.* PLoS Comput Biol, 2015. **11**(2): p. e1004085.PMCID: 4338111
76. Jolma, A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J.M. Vaquerizas, R. Vincentelli, N.M. Luscombe, T.R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale, *DNA-binding specificities of human transcription factors.* Cell, 2013. **152**(1-2): p. 327-39
77. Daily, K., V.R. Patel, P. Rigor, X. Xie, and P. Baldi, *MotifMap: integrative genome-wide maps of regulatory motif sites for model species.* BMC Bioinformatics, 2011. **12**: p. 495.PMCID: 3293935

Acknowledgements

This work was supported by National Institutes of Health (NIH) grants R01 AG048023, P01 AG021600, U19AI089987, and the Jackson Laboratory Director's Innovation Fund JAX-DIF-FY15-DU-JB. G.K also serves as the Citicorp chair in Geriatrics and Gerontology. JAX Genome Technologies provided help with the sequencing data generation.

Author contributions

J.B, D.U, and G.K designed the research. C.C., R.M., R.R., T.W. performed the experiments. D.U., E.M, J.B. analyzed the data. A.U, J.G. pre-processed the sequencing data. J.B., D.U., E.M., G.K. wrote the paper. M.S. and K.P. helped with data interpretation and paper writing.

Figures:

Fig 1

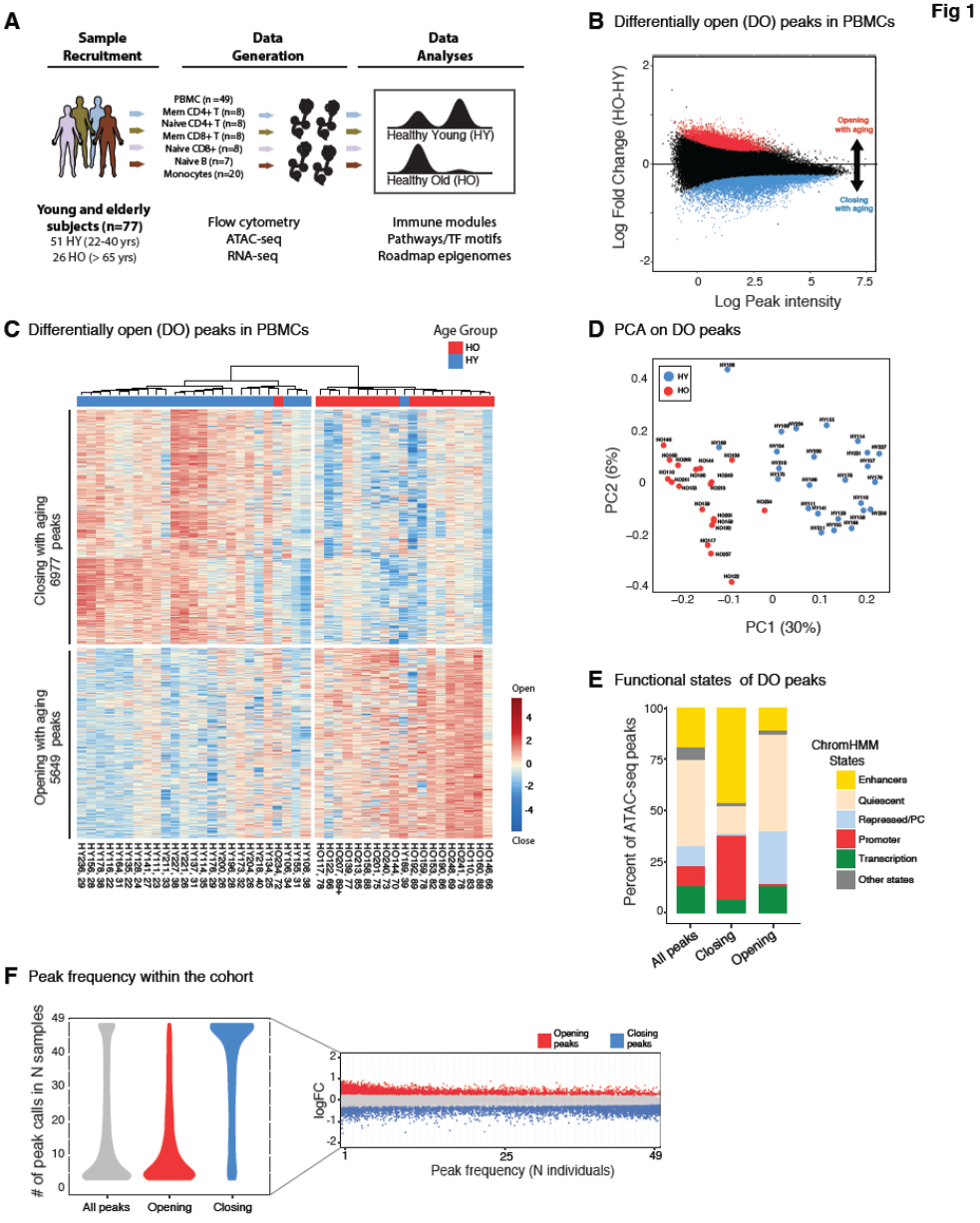

A
Sample Recruitment
Data Generation
Data Analyses
PBMC (n=49)
Mem CD4+ T (n=8)
Naive CD4+ T (n=8)
Mem CD8+ T (n=8)
Naive CD8+ (n=8)
Naive B (n=7)
Monocytes (n=20)
Healthy Young (HY)
Healthy Old (HO)
Young and elderly subjects (n=77)
51 HY (22-40 yrs)
26 HO (> 65 yrs)
Flow cytometry
ATAC-seq
RNA-seq
Immune modules
Pathways/TF motifs
Roadmap epigenomes
B Differentially open (DO) peaks in PBMCs
Log Fold Change (HO-HY)
Log Peak intensity
Opening with aging
Closing with aging
C Differentially open (DO) peaks in PBMCs
Age Group
HO
HY
Closing with aging 6977 peaks
Opening with aging 5649 peaks
Open
Close
D PCA on DO peaks
PC2 (6%)
PC1 (30%)
E Functional states of DO peaks
Percent of ATAC-seq peaks
ChromHMM States
Enhancers
Quiescent
Repressed/PC
Promoter
Transcription
Other states
All peaks
Closing
Opening
F Peak frequency within the cohort
# of peak calls in N samples
All peaks
Opening
Closing
Opening peaks
Closing peaks
logFC
Peak frequency (N individuals)

Fig. 1. Epigenomic signature of aging in PBMCs. (**A**) Schema summarizing our study. (**B**) Plot representing log2 fold change (old-young) versus average read count for ATAC-seq peaks. Peaks differentially opening (closing) with aging are represented in red (blue) (5% FDR). (**C**) Heatmap showing normalized (z-scores) chromatin profiles for differentially closing/opening peaks across PBMC samples. (**D**) Plot of first two Principal Components (PC) based on differential peaks confirms that PC1 accounts for the separation between age groups. Percent of variation among differential peaks accounted for by each PC is shown in parentheses. PC1 from this analysis accounts for ~7% of the variance in the complete data set. (**E**) Relative to all peaks tested, differentially closing peaks are enriched in promoters and enhancers, whereas opening peaks are enriched in quiescent and repressed sites. (**F**) Relationship between peak frequency (i.e., in how many samples/subjects a peak is called) and aging-related change in chromatin accessibility. Left: differentially closing peaks (in blue) are commonly found across the cohort, whereas opening peaks (in red) tend to be rare or private (all pairwise comparisons between shown distributions are significant after Wilcoxon test, $p<0.01$). Right: log2 fold change (old-young) as a function of peak frequency. Significantly closing and opening peaks are shown in blue and red respectively.

Fig 2

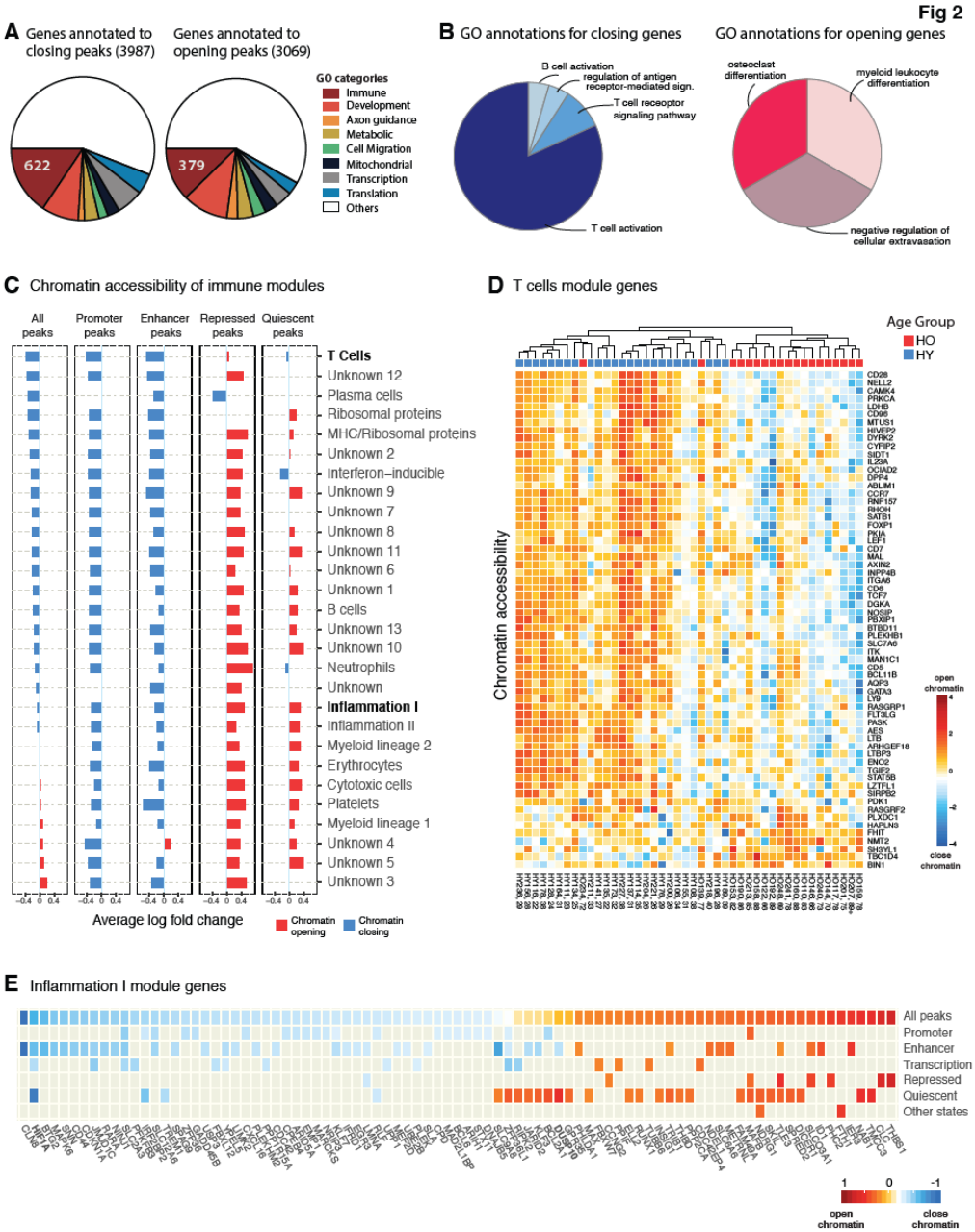

A Genes annotated to closing peaks (3987)
Genes annotated to opening peaks (3069)
622
379
GO categories
Immune
Development
Axon guidance
Metabolic
Cell Migration
Mitochondrial
Transcription
Translation
Others
B GO annotations for closing genes
B cell activation
regulation of antigen receptor-mediated sign.
T cell receoptor signaling pathway
T cell activation
GO annotations for opening genes
osteoclast differentiation
myeloid leukocyte differentiation
negative regulation of cellular extravasation
C Chromatin accessibility of immune modules
All peaks
Promoter peaks
Enhancer peaks
Repressed peaks
Quiescent peaks
T Cells
Unknown 12
Plasma cells
Ribosomal proteins
MHC/Ribosomal proteins
Unknown 2
Interferon–inducible
Unknown 9
Unknown 7
Unknown 8
Unknown 11
Unknown 6
Unknown 1
B cells
Unknown 13
Unknown 10
Neutrophils
Unknown
Inflammation I
Inflammation II
Myeloid lineage 2
Erythrocytes
Cytotoxic cells
Platelets
Myeloid lineage 1
Unknown 4
Unknown 5
Unknown 3
−0.4 0 0.4
Average log fold change
Chromatin opening
Chromatin closing
D T cells module genes
Age Group
HO
HY
Chromatin accessibility
open chromatin
close chromatin
E Inflammation I module genes
All peaks
Promoter
Enhancer
Transcription
Repressed
Quiescent
Other states
1
0
-1
open chromatin
close chromatin

Fig. 2. Epigenomic signature of aging at immune-related genes. (**A**) Major GO category annotations of genes associated with differentially closing and opening peaks. (**B**) GO terms associated with immune-related genes enriched among genes annotated to differentially closing (blue, left) and opening (red, right) peaks. (**C**) Average chromatin remodeling (log2 fold change) of genes listed in 28 immune co-expression modules, calculated based on all peaks (leftmost column) and separately using peaks annotated to specific chromHMM states. (**D**) Subject-specific normalized (z-scores) chromatin accessibility patterns of peaks annotated to genes in the T cell co-expression module reveals concerted aging-related variation across the cohort. Warmer (cooler) hues represent increased (decreased) chromatin accessibility relative to the cohort mean. (**E**) Average chromatin remodeling (log2 fold change) of peaks annotated to genes in the Inflammation I module, calculated based on all peaks (topmost row) and separately with respect to specific chromHMM state annotations.

Fig 3

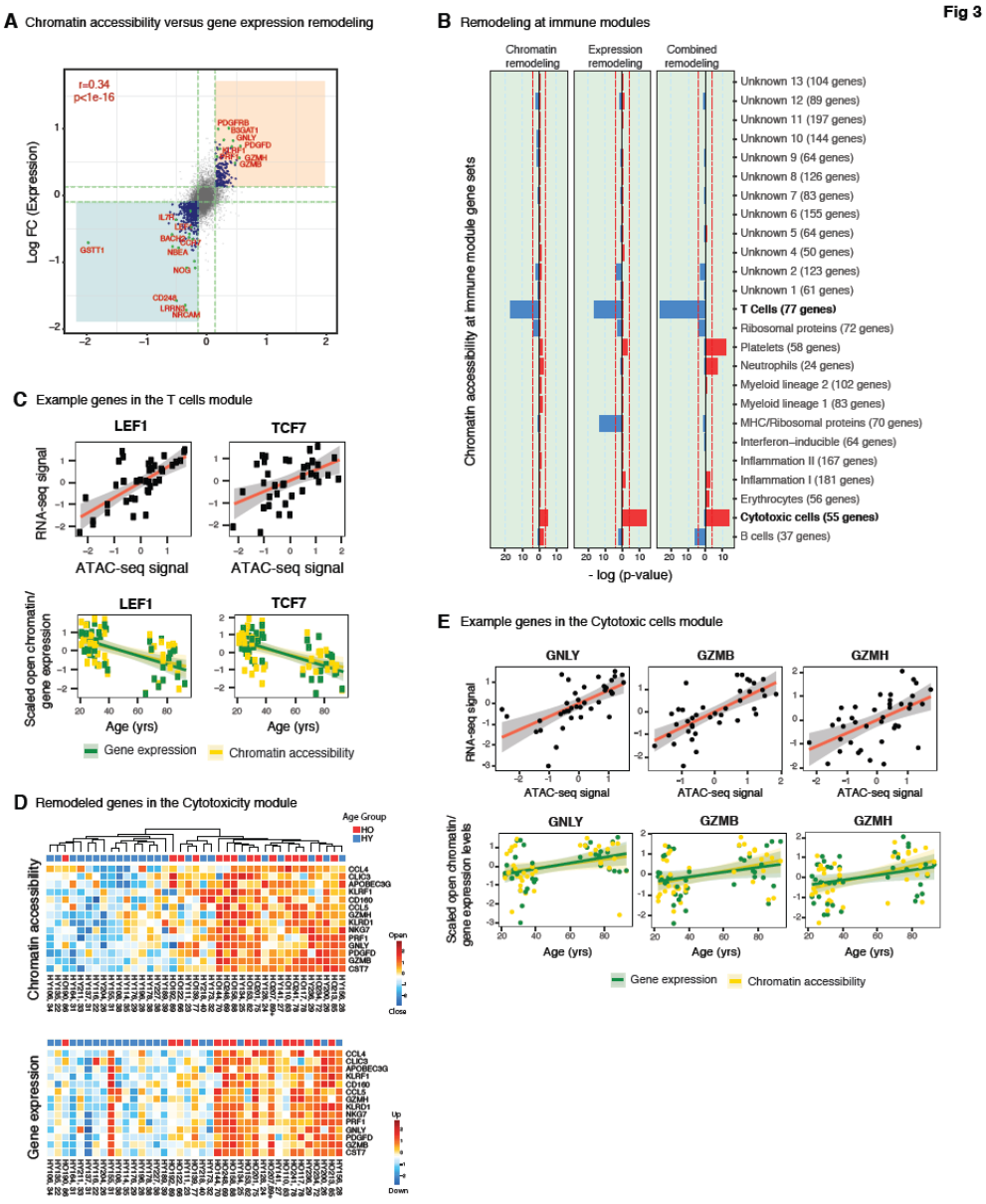

A Chromatin accessibility versus gene expression remodeling
r=0.34
p<1e-16
Log FC (Expression)
B Remodeling at immune modules
Chromatin remodeling
Expression remodeling
Combined remodeling
Chromatin accessibility at immune module gene sets
Unknown 13 (104 genes)
Unknown 12 (89 genes)
Unknown 11 (197 genes)
Unknown 10 (144 genes)
Unknown 9 (64 genes)
Unknown 8 (126 genes)
Unknown 7 (83 genes)
Unknown 6 (155 genes)
Unknown 5 (64 genes)
Unknown 4 (50 genes)
Unknown 2 (123 genes)
Unknown 1 (61 genes)
T Cells (77 genes)
Ribosomal proteins (72 genes)
Platelets (58 genes)
Neutrophils (24 genes)
Myeloid lineage 2 (102 genes)
Myeloid lineage 1 (83 genes)
MHC/Ribosomal proteins (70 genes)
Interferon–inducible (64 genes)
Inflammation II (167 genes)
Inflammation I (181 genes)
Erythrocytes (56 genes)
Cytotoxic cells (55 genes)
B cells (37 genes)
- log (p-value)
C Example genes in the T cells module
LEF1
TCF7
RNA-seq signal
ATAC-seq signal
Scaled open chromatin/ gene expression
Age (yrs)
Gene expression
Chromatin accessibility
D Remodeled genes in the Cytotoxicity module
Age Group
HO
HY
Chromatin accessibility
Open
Close
Gene expression
Up
Down
E Example genes in the Cytotoxic cells module
GNLY
GZMB
GZMH
Scaled open chromatin/ gene expression levels

Fig. 3. Concordant transcriptional and epigenomic changes associated with aging in PBMCs. (**A**) Chromatin remodeling at gene promoters correlates significantly with changes in expression of the co-located genes (Pearson r = 0.32, p-value < 2.2×10^{-16}). Dashed lines delineate the set of peaks (x-axis) and genes (y-axis) that are differentially accessible or expressed between young and old subjects with a permutation-based p-value<0.01. Shaded quadrants define sets of genes showing congruent aging-related shifts in chromatin accessibility and expression. (**B**) Enrichment of immune modules among genes sets associated to differentially accessible peaks (left), differentially expressed genes (center), and congruent (concordant) chromatin and expression remodeling (right). Plots show -log10 of hypergeometric test p-values, colored according to the direction of the observed change (blue for decrease and red for increase with age). Reference lines are drawn at the largest p-value for which a 5% FDR is attained. (**C**) Examples of concordantly remodeled genes from the T cells module. Top: chromatin accessibility and gene expression correlate among subjects. Bottom: both chromatin accessibility (yellow dots and lines) and gene expression (green dots and lines) decrease with aging. (**D**) Promoter chromatin accessibility (top) and gene expression (bottom) of genes in the cytotoxic cells module that show congruent increased in accessibility and expression with aging. Warmer (cooler) hues represent increased (decreased) chromatin accessibility (expression) relative to the cohort mean, data shown as normalized (*z*-scores) values. (**E**) Examples of concordantly remodeled genes from the cytotoxic cells module. Top: chromatin accessibility and gene expression correlate among subjects. Bottom: both chromatin accessibility (yellow dots and lines) and gene expression (green dots and lines) increase with aging.

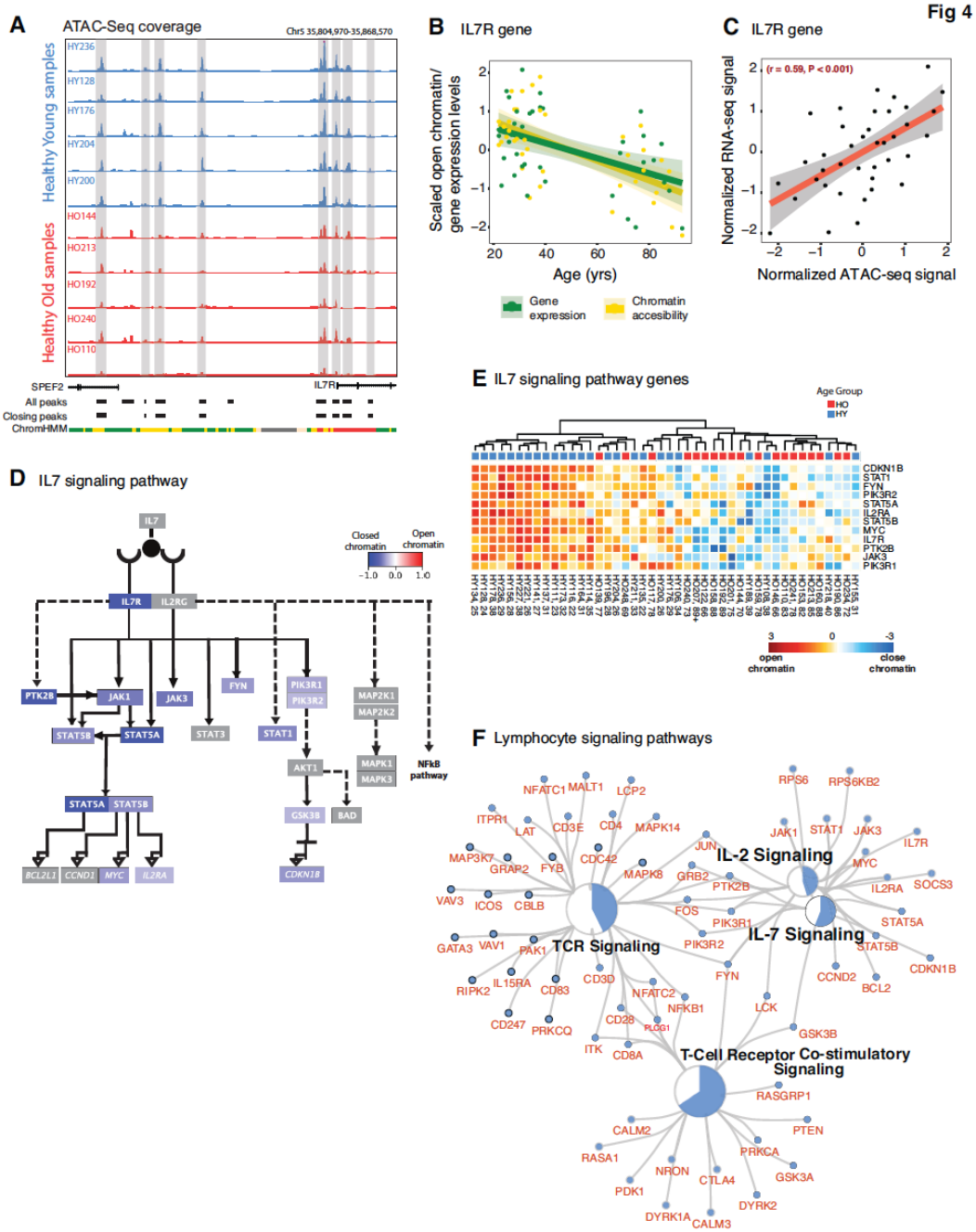


Fig. 4. T cell signaling pathways are affected with aging. (**A**) Genome browser view of *IL7R* locus highlighting 8 (out of 12) differentially closing peaks. Blue and red tracks represent open chromatin profiles of HY (n=5) and HO (n=5) samples, respectively. (**B**) *IL7R* expression and chromatin accessibility at its promoter decrease with aging. (**C**) Promoter chromatin accessibility and gene expression are highly correlated at among subjects. (**D**) Chromatin

accessibility of peaks annotated to genes in the IL7 signaling pathway. Color represents the fold change of the most significant differential peak annotated to this gene. Genes marked in grey are not associated with a closing or opening peak. (**E**) Subject-specific chromatin accessibility of peaks significantly closing with aging and annotated to genes in the IL7 signaling pathway. Warmer (cooler) hues represent increased (decreased) chromatin accessibility relative to the cohort mean, data shown as normalized (z-scores) values. (**F**) ClueGO figure representing the genes that are in T cell signaling pathways that annotate differentially closing peaks. Pie charts show in blue represent the percent of genes in that pathway associated with closing peaks.

Fig 5

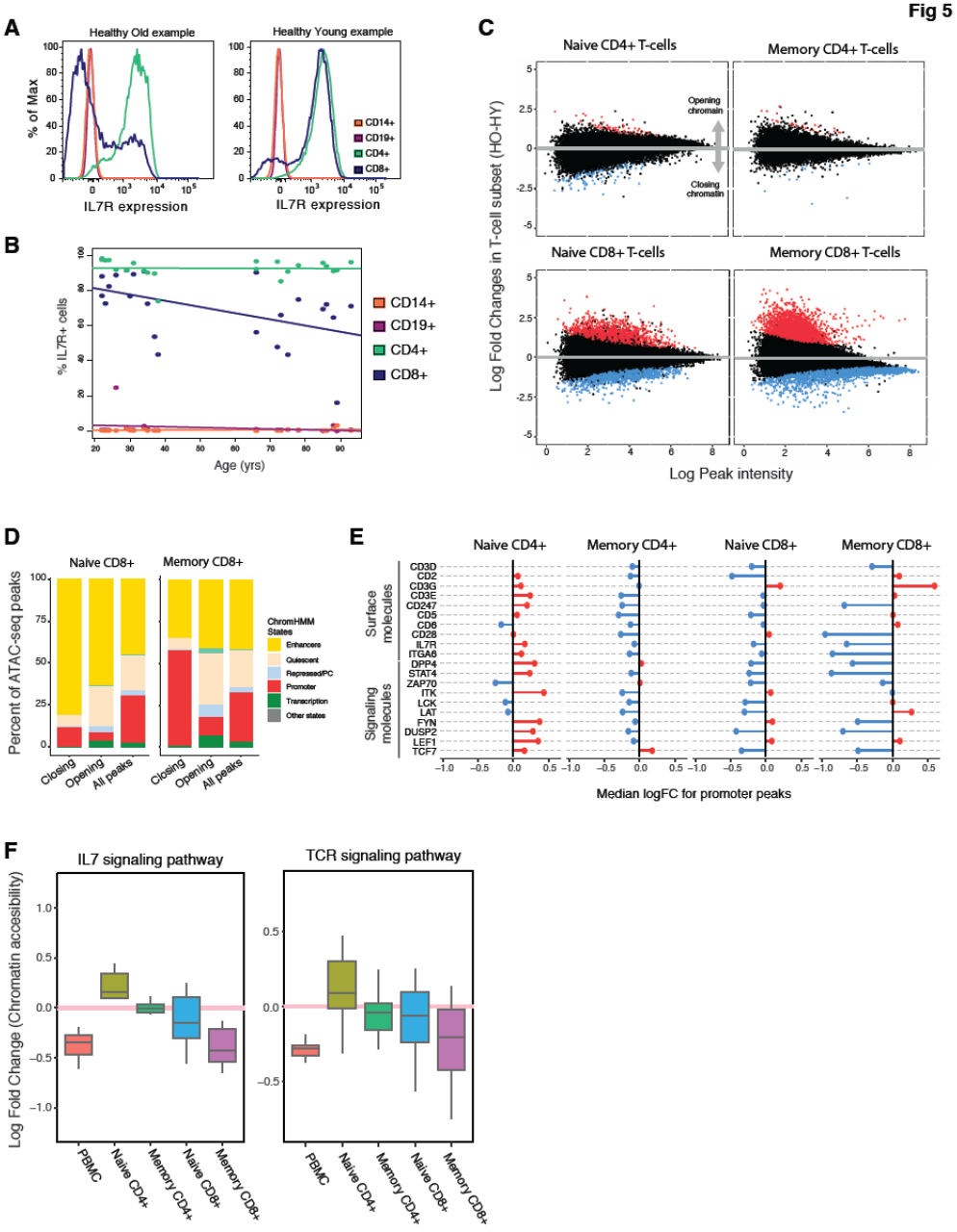
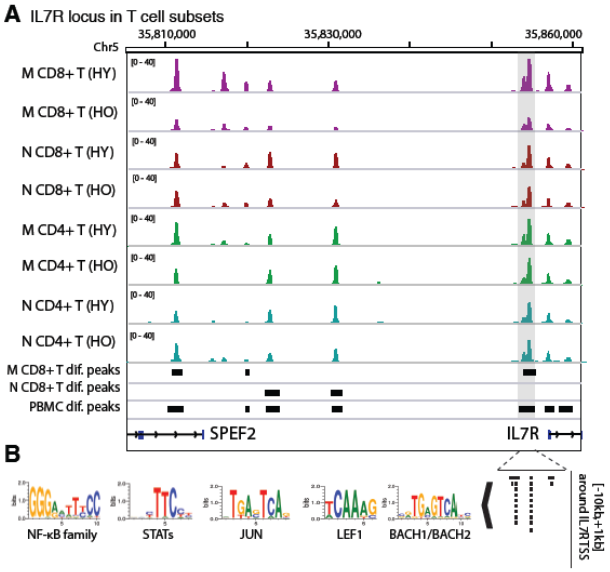

A
Healthy Old example
Healthy Young example
% of Max
IL7R expression
CD14+
CD19+
CD4+
CD8+
B
% IL7R+ cells
Age (yrs)
C
Naive CD4+ T-cells
Memory CD4+ T-cells
Naive CD8+ T-cells
Memory CD8+ T-cells
Opening chromatin
Closing chromatin
Log Fold Changes in T-cell subset (HO-HY)
Log Peak intensity
D
Naive CD8+
Memory CD8+
Percent of ATAC-seq peaks
ChromHMM States
Enhancers
Quiescent
RepressedPC
Promoter
Transcription
Other states
Closing
Opening
All peaks
E
Naive CD4+
Memory CD4+
Naive CD8+
Memory CD8+
Surface molecules
Signaling molecules
CD3D
CD2
CD3G
CD3E
CD247
CD5
CD6
CD28
IL7R
ITGA6
DPP4
STAT4
ZAP70
ITK
LCK
LAT
FYN
DUSP2
LEF1
TCF7
Median logFC for promoter peaks
F
IL7 signaling pathway
TCR signaling pathway
Log Fold Change (Chromatin accesibility)
PBMC
Naive CD4+
Memory CD4+
Naive CD8+
Memory CD8+

Fig. 5. The PBMC aging signature stems mostly from memory CD8+ T cells. (**A**) Flow cytometry plots in representative young (left) and old (right) subjects illustrate the decrease in *IL7R* protein levels with aging in CD8+ T cells. (**B**) Flow cytometry results indicating that the aging-related decrease in *IL7R* levels is specific to CD8+ T cells. (**C**) Differential accessibility analyses in T cell subsets show that most significant aging-related remodeling occurs in CD8+ T cells, and particularly in memory CD8+ T cells. Plots representing log2 fold change (old-young) versus average read count for the corresponding ATAC-seq peaks in T cell subsets. Opening (closing) peaks are represented in red (blue) (5% FDR). (**D**) Distribution of differential and all peaks classified by chromHMM state annotations (Roadmap T cell annotations) for memory and naïve CD8+ T cells. Promoters and enhancers close with aging in memory CD8+ T cells, similar to PBMCs. (**E**) Chromatin accessibility remodeling (median fold change) of promoters of selected functionally relevant signaling and surface molecules in naïve and memory CD4+ and CD8+ T cells. Red and blue bars represent positive (i.e., opening with aging) and negative (i.e., closing with aging) median fold change, respectively, aggregated over all peaks overlapping promoters of the corresponding gene. (**F**) Chromatin remodeling of closing PBMC regions associated to genes in the IL7 signaling pathway (left) and TCR signaling pathway (right) stems from the remodeling in memory CD8+ T cells. Boxplots for PBMC and T cell subsets represent distribution of log2 fold changes of peaks annotated to genes that are associated to closing peaks in PBMC.

Fig 6



C NF-κB family members

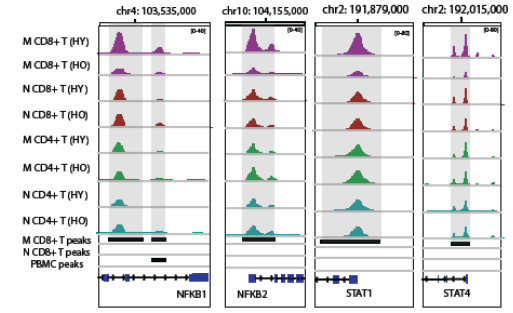
TF	Expression (PBMC)	Chromatin (M CD8+)	Chromatin (N CD8+)	Chromatin (PBMC)
NFKB1	↔	↓*	↑e	↓e
NFKB2	↔	↓*	↔	↔
REL	↔	↓	↔	↔
RELA	↔	↓	↔	↔
RELB	↔	↓	↔	↔

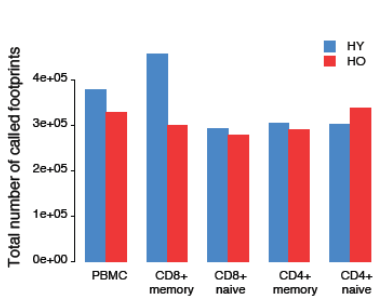
e : Significant enhancer peak
* : Significant at FDR 5%

D STAT factors

TF	Expression (PBMC)	Chromatin (M CD8+)	Chromatin (N CD8+)	Chromatin (PBMC)
STAT1	↓	↓*	↔	↓e
STAT2	↔	↓	↔	↔
STAT3	↔	↓*	↔	↔
STAT4	↑	↓*	↔	↔
STAT5A	↔	↓	↔	↓*
STAT5B	↔	↓	↔	↓e
STAT6	↔	↓	↔	↔

e : Significant enhancer peak
* : Significant at FDR 5%





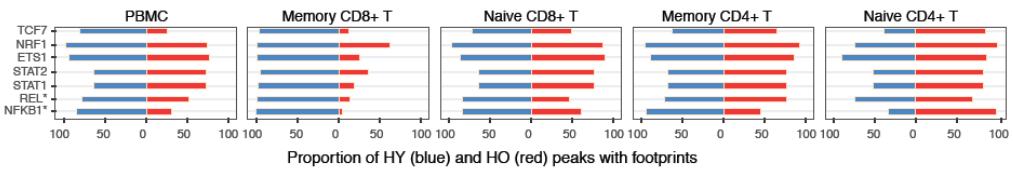


Fig. 6. TF activity is repressed specifically in memory CD8+ T cells. (**A**) Genome browser view of the IL7R locus. IGV tracks are generated after pooling 4 HO and 3HY samples. The same individuals are used for all T cell subsets. (**B**) TF motifs found at the IL7R promoter (-10,000bp upstream, +1000 bp downstream) at 20% FDR. TFs that are not expressed in PBMCs are filtered out. (**C**) Summary of TFs that belong to the NF-κB family in terms of expression changes in PBMCs and chromatin changes in PBMCs and CD8+ T cells. Note that these TFs are specifically affected with aging in Memory CD8+ T cells. (**D**) Summary of expression changes in PBMCs and chromatin changes in PBMCs and CD8+ T cells for TF in the STAT family. Note that these TFs are specifically affected with aging in Memory CD8+ T cells. (**E**) Chromatin accessibility profiles at NF-κB and STAT TF loci in T cell subsets. Note that these TFs lose the chromatin accessibility of their promoters specifically in memory CD8+ T cells with aging. The differential peaks (5% FDR) from PBMCs and naïve and memory CD8+ T cells are shown in black bars. (**F**) Total number of TF footprinting calls obtained in PBMC and T cell subsets from young (blue) and old (red) samples. All ATAC-seq samples are pooled and randomly down-sampled to the same depth prior to TF footprinting calls to eliminate potential biases due to depth of sequencing. Note that the number of footprint calls decrease with aging in memory CD8+ T cells. (**G**) Proportion of ATAC-seq peaks with a footprint for selected TFs. Notice that the decrease with aging in the proportion of peaks carrying footprints is specific to memory CD8+ T cells for NF-κB and STAT factors, as well as other TFs relevant for T cell functions.

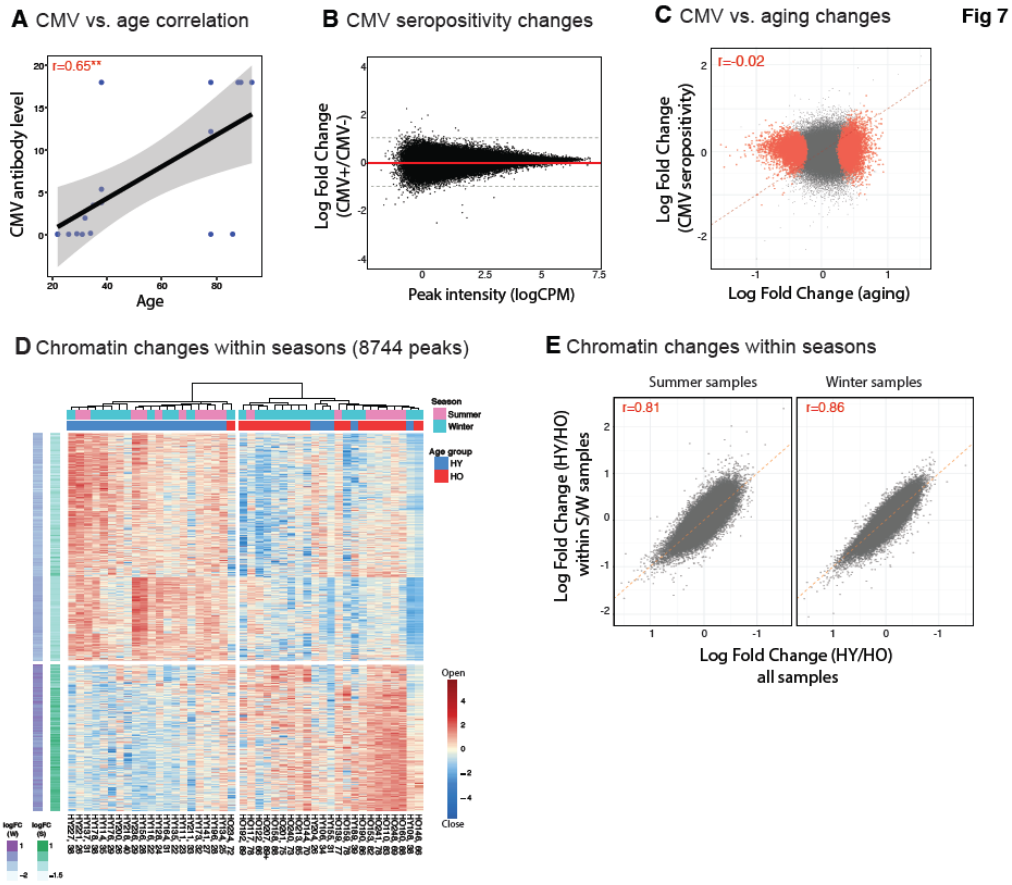

A CMV vs. age correlation
r=0.65**
CMV antibody level
Age
B CMV seropositivity changes
Log Fold Change (CMV+/CMV-)
Peak intensity (logCPM)
C CMV vs. aging changes
r=-0.02
Log Fold Change (CMV seropositivity)
Log Fold Change (aging)
D Chromatin changes within seasons (8744 peaks)
Season
Summer
Winter
Age group
HY
HO
Open
Close
E Chromatin changes within seasons
Summer samples
Winter samples
r=0.81
r=0.86
Log Fold Change (HY/HO) within S/W samples
Log Fold Change (HY/HO) all samples

Fig 7

Fig. 7. Relevance of CMV seropositivity and seasonality as factors influencing aging signature of chromatin accessibility in PBMC. (**A**) Correlation between CMV antibody level and age. Note the positive and significant correlation between CMV positivity and aging. (**B**) Plot representing log2 fold change (CMV+ vs. CMV-) versus average read count for ATAC-seq peaks. No differential peaks are obtained at 5% FDR. (**C**) Correlation between peak-specific log fold changes in chromatin accessibility associated with aging and log fold changes associated with CMV seropositivity. Peaks that are closing or opening with aging are shown in red. Note the weak correlation (r=-0.02) suggesting that CMV seropositivity does not explain aging-associated chromatin changes. (**D**) Heatmap showing normalized (*z*-scores) chromatin accessibility profiles of differentially closing and opening peaks across PBMC samples obtained in two seasons. Shades of purple and green on the left represents fold changes in winter and summer samples respectively. (**E**) Correlation between peak-specific log fold changes in chromatin accessibility associated with aging for all samples and changes associated with aging for summer (left) and winter (right) samples. Note that season-specific chromatin changes associated with aging are highly correlated with global changes. Changes in winter are slightly more strongly correlated with the global signature (0.86 vs. 0.81).

Supplementary Materials:

Fig S1

A Major cell population comparisons (%)

	Mean (HY)	Mean (HO)	SD (HY)	SD (HO)	Dif (HY-HO)	p-value (Wilcox)
CD14	15.7	15.6	7.4	5.4	0.1	0.993
CD19	7.5	5.4	3	3.4	2.1	0.014
CD4	30.7	28.5	7.8	10.4	2.2	0.526
CD8	15.6	10.5	4.7	7	5.1	0.006

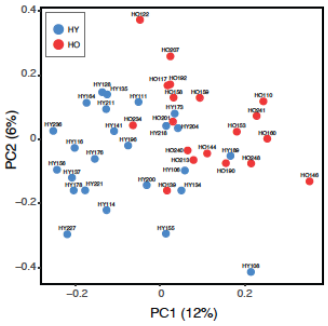
B CD4+ T cell subset comparisons (%)

	Mean (HY)	Mean (HO)	SD (HY)	SD (HO)	Dif (HY-HO)	p-value (Wilcox)
Total	30.4	29.7	7.9	9.1	0.7	0.875
Naive	11.1	7.8	6.3	5.4	3.4	0.047
CM	7.9	7.6	4.1	4.4	0.3	0.559
EM	9.1	11.8	4.8	4.5	-2.7	0.024
EMRA	2.4	2.5	2.4	2.1	-0.1	0.610

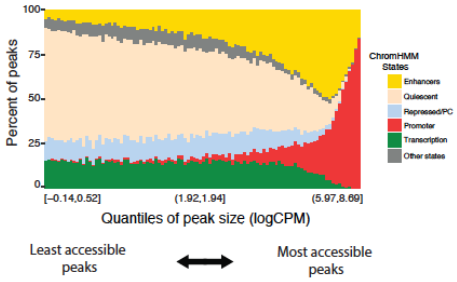
C CD8+ T cell subset comparisons (%)

	Mean (HY)	Mean (HO)	SD (HY)	SD (HO)	Dif (HY-HO)	p-value (Wilcox)
Total	15.6	10.9	4.6	7.4	4.7	0.022
Naive	7.2	3.2	3.2	4.1	4	1.E-04
CM	1.1	1.2	0.9	1.3	-0.1	0.441
EM	3.5	2.6	3.7	1.9	0.9	0.162
EMRA	3.7	4	2.3	4.1	-0.3	0.630

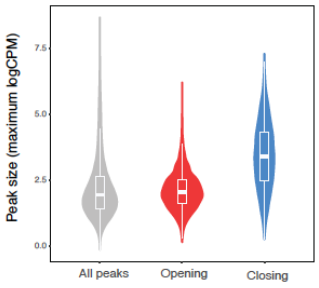
D PCA on all PBMC ATAC-seq peaks



E ChromHMM states for all PBMC ATAC-seq peaks



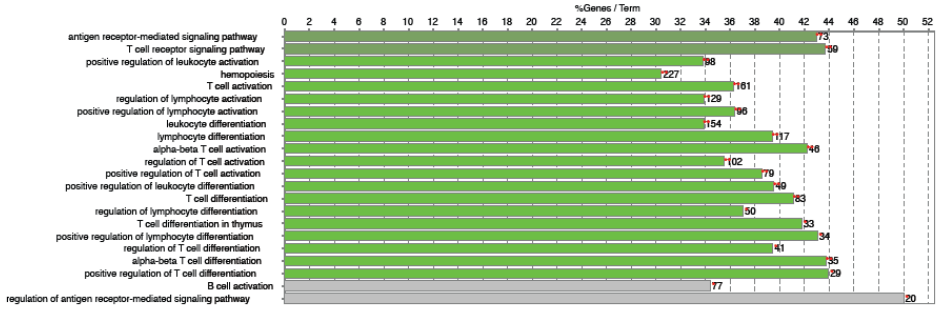
F Peak sizes for DO and all peaks



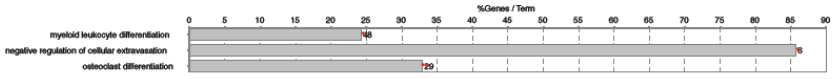
Sup Fig. 1. Cell composition changes with aging. (**A**) Changes in major cell compositions with aging. P-values are calculated using the Wilcoxon test. Note significant decrease in total CD19+ and CD8+ T cells. (**B**) Changes in CD4+ T cells subsets with aging. A significant increase in Effector Memory cells is observed. (**C**) Changes in CD8+ T cells subsets with aging. A significant decrease in Naïve CD8+ T cells is observed. (**D**) Sample loadings on first and second principal component computed based on adjusted reads estimated for all of the PBMC ATAC-seq peaks scored in this study (n=140,172 peaks). Note that age groups tend to cluster separately, implying aging as a leading factor explaining epigenetic variation in PBMC. (**E**) Relationship between peak size (i.e., normalized read counts within peak) and functional annotations, as annotated using Roadmap chromHMM states for PBMC. Larger, more accessible peaks are more likely to be found at promoters (red bars) and enhancers (yellow), whereas small peaks are more likely to be called at inactive regions, represented by quiescent (salmon) and repressed sites (slate). (**F**) Distribution of peak sizes (maximum logCPM of read counts among all samples) as a function of aging-related changes in peak accessibility for all, differentially closing, and differentially opening peaks.

Fig S2

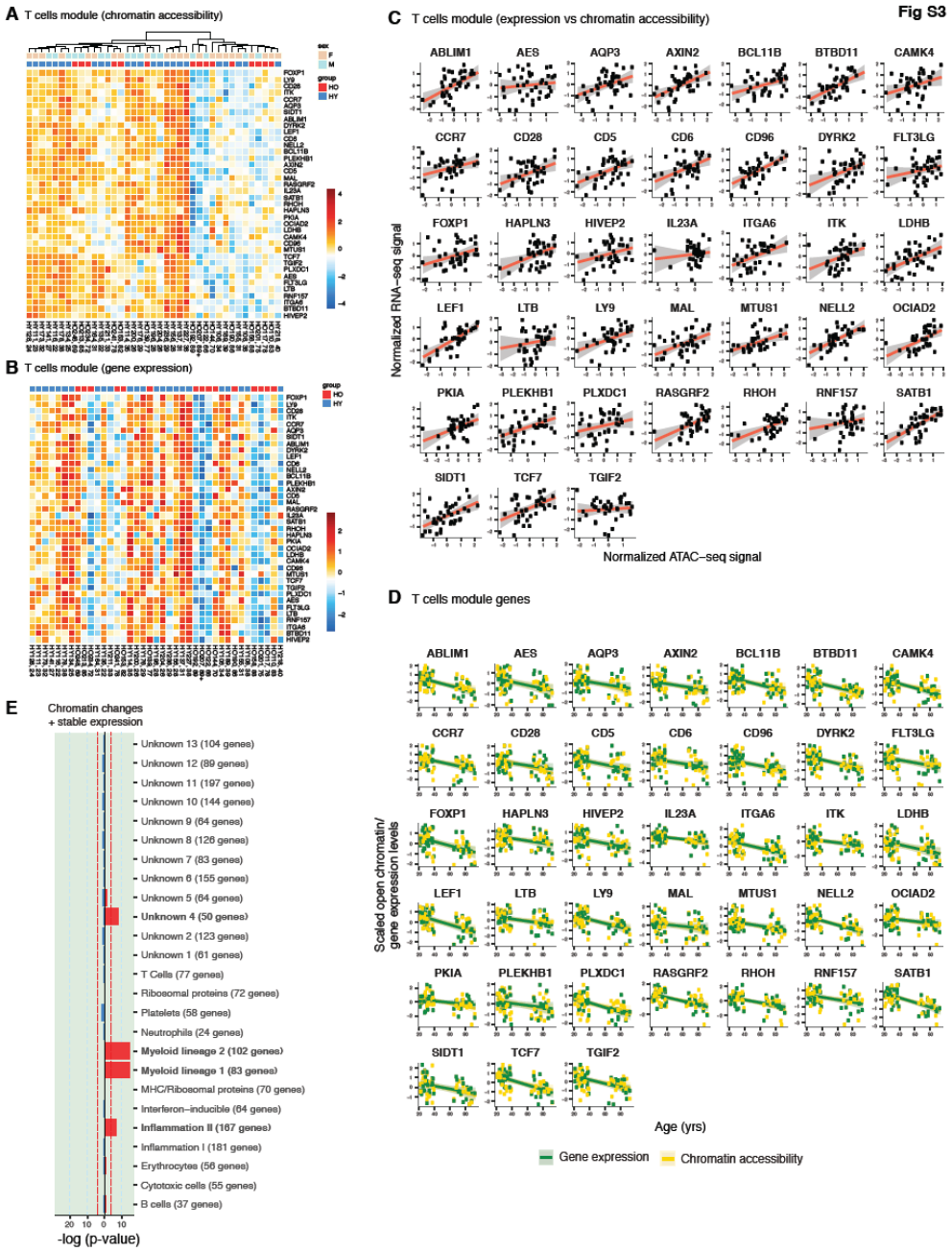
A GO Immune term annotations for HY-specific peaks, p < 0.05, bonferroni step down



B GO Immune term annotations for HO-specific peaks, p < 0.05, bonferroni step down



Sup Fig. 2. Immune genes are affected by chromatin remodeling. (**A**) ClueGO GO term enrichment results for HY-specific peaks, evaluated at a p-value<0.05 after Bonferroni step-down correction. Bars are proportional to the percent of genes in each enriched GO term that also present in the query set, whereas numbers next to the bars indicate the corresponding gene count. (**B**) ClueGO enrichment results for HO-specific peaks, evaluated at a p-value<0.05 (no correction). Bars and numbers next to them represent, respectively, the percent and total number of genes in each enriched GO term and also present in the query set.



Sup Fig. 3. T cell module is epigenetically and transcriptionally repressed with aging. (**A**) Personal profiles of chromatin accessibility of promoter peaks associated to genes in the T cell module for which both chromatin

accessibility and gene expression decrease with aging. (**B**) Personal profiles of gene expression of genes in the T cell module, with samples and genes sorted to match the clustering computed based on chromatin accessibility. (**C**) Plot of gene expression vs. chromatin accessibility for subject-matched samples for genes with aging-related promoter closing and lost expression in the T cell module. (**D**) Plot of chromatin accessibility and expression as a function of age for matched samples for genes in the T cell module with aging-related decrease in accessibility and expression. (**E**) Immune module enrichment of genes that undergo loss of accessibility with aging but which transcriptionally stable (reported as -log10 of p-values from a hypergeometric test; reference lines are drawn at the largest p-value for which a 5% FDR is attained).

Fig S4

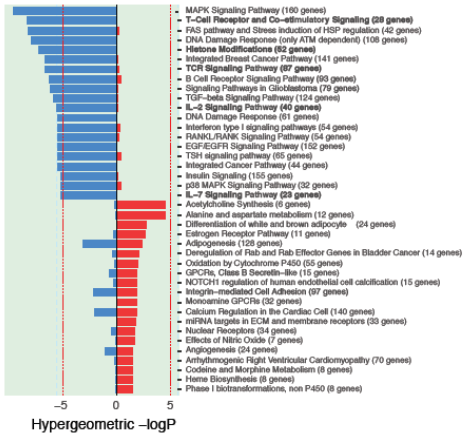
A Top 15 peaks associated with chromatin closing

Gene Name	# of closing peaks	Average FDR	Average logFC	Average logCPM
ETS1	24	2.05E-02	-0.40	2.17
TGFBR2	15	8.75E-03	-0.49	2.40
FAM65B	13	5.95E-03	-0.48	2.21
BACH2	13	8.00E-03	-0.51	2.32
FOXP1	12	1.05E-02	-0.46	2.57
IL7R	12	4.89E-03	-0.60	2.12
BRINP3	11	3.03E-02	-0.41	0.92
CCR7	11	9.36E-03	-0.53	2.45
NCK2	11	9.11E-03	-0.43	2.81
CDC42SE2	11	1.15E-02	-0.38	2.75
LRRN3	11	5.26E-03	-0.58	1.96
IKZF1	11	1.25E-02	-0.44	2.86
PTPRC	10	2.05E-02	-0.35	2.88
CD44	10	1.03E-02	-0.47	2.56
SNX9	10	1.43E-02	-0.43	2.57

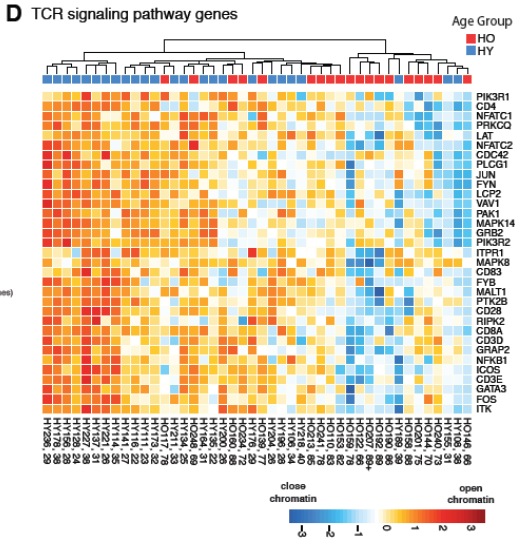
B Top 15 peaks associated with chromatin opening

Gene Name	# of closing peaks	Average FDR	Average logFC	Average logCPM
MAF	18	1.71E-02	0.42	1.25
FGFR2	9	1.64E-02	0.51	0.67
ELK2AP	9	1.63E-02	0.51	2.17
CPZ	9	2.11E-02	0.51	1.01
ZNF618	8	2.61E-02	0.40	1.24
KIF26A	7	2.49E-02	0.42	1.40
EGLN3	7	1.97E-02	0.44	0.85
APBA2	7	9.34E-03	0.43	1.71
NPIPA7	7	1.37E-02	0.59	0.77
VAC14	7	2.82E-02	0.34	2.09
ZNF469	7	1.61E-02	0.49	1.27
AGAP1-IT1	7	1.76E-02	0.44	0.88
CDH4	7	2.61E-02	0.45	1.22
DAB2IP	7	2.36E-02	0.42	1.30
ARHGEF10L	6	2.69E-02	0.39	1.22

C Wiki pathway enrichments

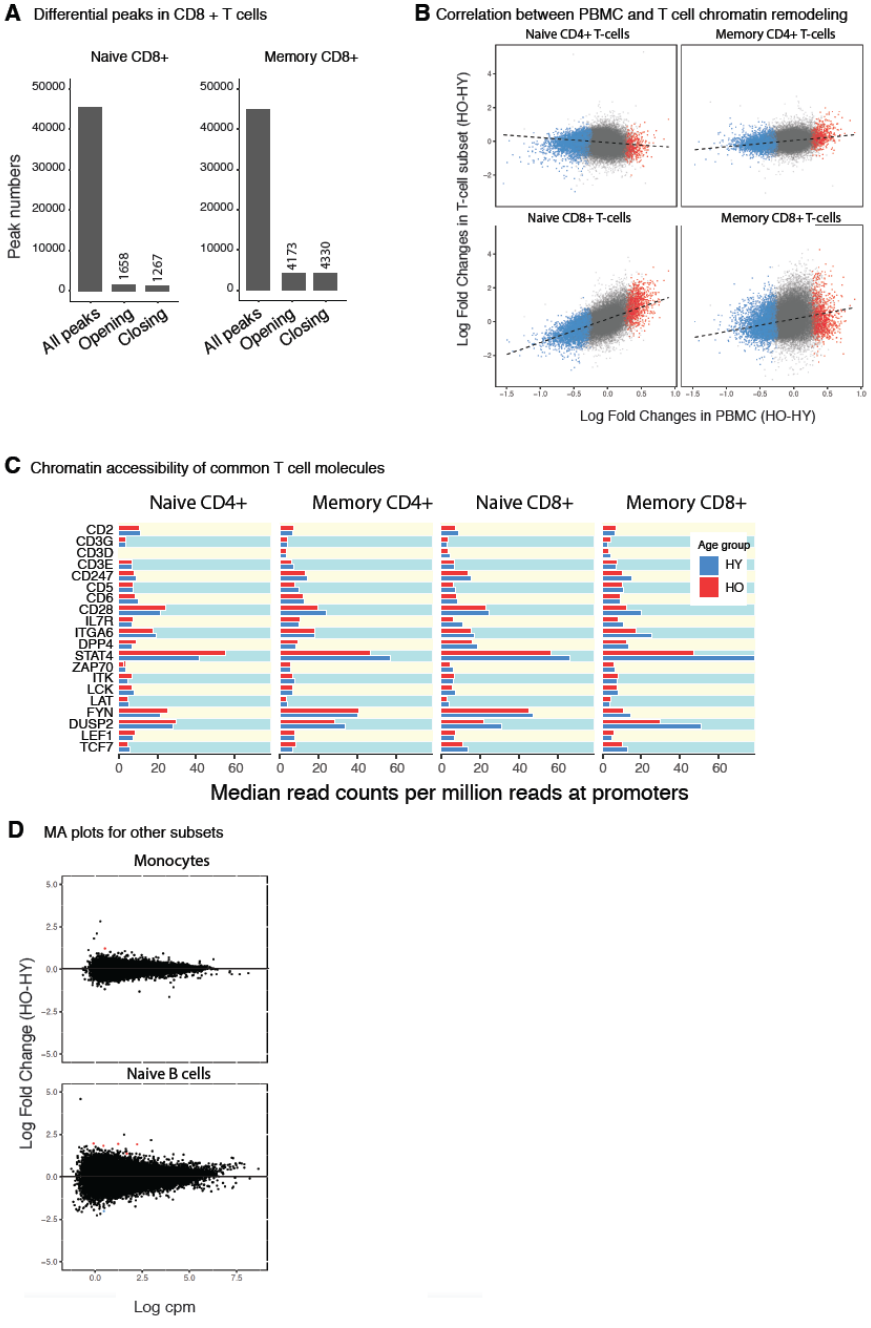


D TCR signaling pathway genes



Sup Fig. 4. TCR signaling is epigenetically repressed with aging. (**A**) Top 15 genes associated with chromatin closing with aging. Genes are sorted with respect to the number of significantly closing peaks annotated to their promoters. *IL7R* is a top gene in this list with 13 closing peaks. (**B**) Top 15 genes associated with chromatin opening with aging. Genes are sorted with respect to the number of significantly opening peaks annotated to their promoters. (**C**) Wikipathways enrichment results for genes associated with closing (blue bars) and opening (red bars) peaks. Reference lines are drawn at the largest p-value for which a 5% FDR is attained. (**D**) Normalized (*z*-scores) mean chromatin accessibility personal profiles based on significantly closing peaks annotated to genes in the TCR signaling pathway.

Fig S5


A Differential peaks in CD8 + T cells
Naive CD8+
Memory CD8+
Peak numbers
50000
40000
30000
20000
10000
0
1658
1267
4173
4330
All peaks
Opening
Closing
B Correlation between PBMC and T cell chromatin remodeling
Naive CD4+ T-cells
Memory CD4+ T-cells
Naive CD8+ T-cells
Memory CD8+ T-cells
Log Fold Changes in T-cell subset (HO-HY)
Log Fold Changes in PBMC (HO-HY)
C Chromatin accessibility of common T cell molecules
Naive CD4+
Memory CD4+
Naive CD8+
Memory CD8+
CD2
CD3G
CD3D
CD3E
CD247
CD5
CD6
CD28
IL7R
ITGA6
DPP4
STAT4
ZAP70
ITK
LCK
LAT
FYN
DUSP2
LEF1
TCF7
Age group
HY
HO
0
20
40
60
Median read counts per million reads at promoters
D MA plots for other subsets
Monocytes
Naive B cells
Log Fold Change (HO-HY)
Log cpm

Sup Fig. 5. Memory CD8+ T cells are the largest contributors to the PBMC aging signature. (**A**) Number of peaks in memory and naïve CD8+ T cells showing the increased extent of remodeling associated to aging observed in memory CD8+ T cells. (**B**) Correlation between chromatin remodeling in PBMCs and chromatin remodeling in T cell subsets. Pearson correlation values indicate that PBMCs correlate differently with different T cells, where this correlation is higher for CD8+ T cells than CD4+ T cells. (**C**) Median chromatin accessibility at the promoters of selected functionally relevant signaling and surface molecules in naïve and memory CD4+ and CD8+ T cells. (**D**) chromatin accessibility log2 fold change (old-young) versus average read count for ATAC-seq peaks for monocytes (top) and naïve B cells (bottom). Peaks differentially opening (closing) with aging are represented in red (blue) (5% FDR). Note that there are nearly zero aging differential peaks in these cell types.

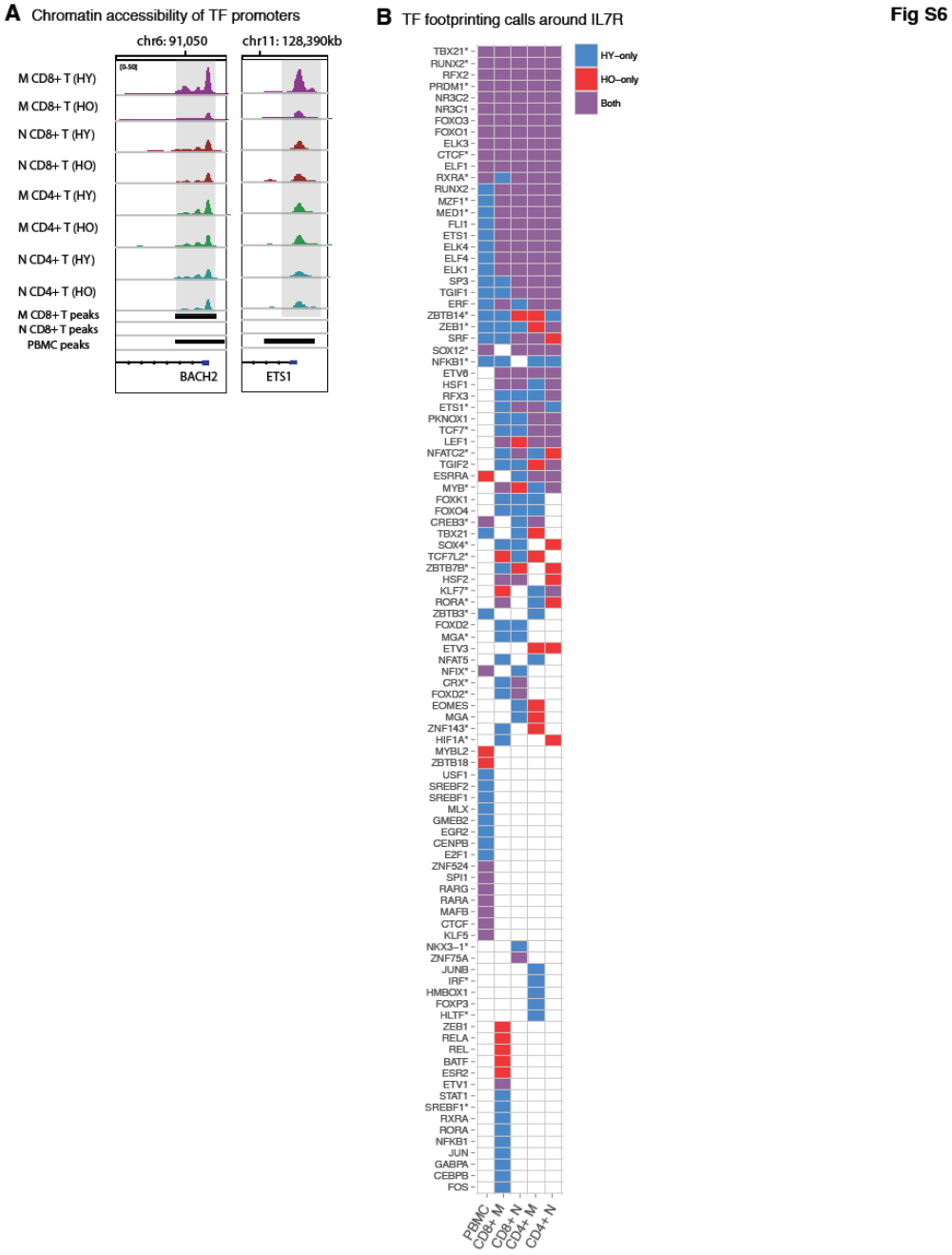
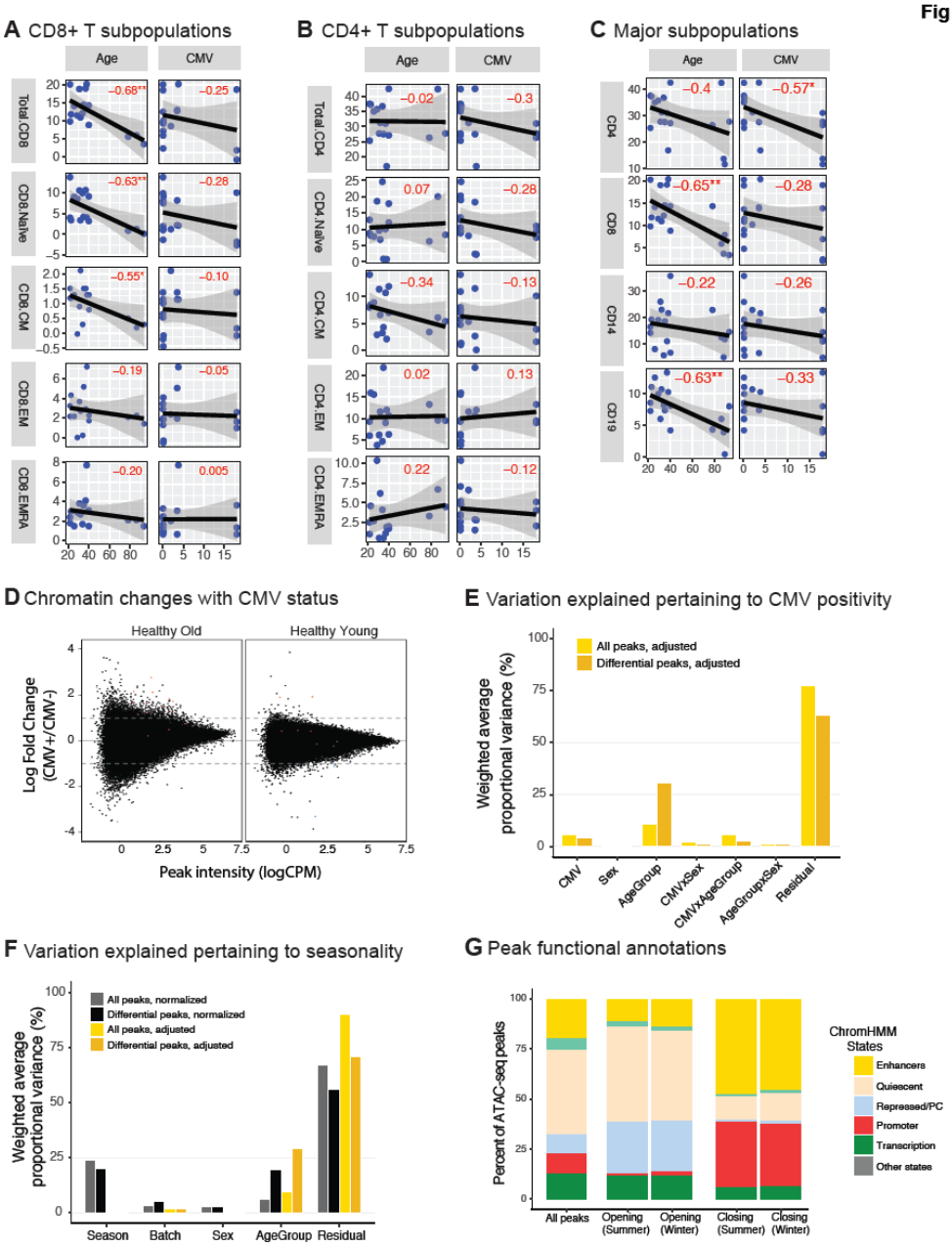

A Chromatin accessibility of TF promoters
chr6: 91,050
chr11: 128,390kb
M CD8+ T (HY)
M CD8+ T (HO)
N CD8+ T (HY)
N CD8+ T (HO)
M CD4+ T (HY)
M CD4+ T (HO)
N CD4+ T (HY)
N CD4+ T (HO)
M CD8+ T peaks
N CD8+ T peaks
PBMC peaks
BACH2
ETS1
B TF footprinting calls around IL7R
HY–only
HO–only
Both
TBX21*
RUNX2*
RFX2
PRDM1*
NR3C2
NR3C1
FOXO3
FOXO1
ELK3
CTCF*
ELF1
RXRA*
RUNX2
MZF1*
MED1*
FLI1
ETS1
ELK4
ELF4
ELK1
SP3
TGIF1
ERF
ZBTB14*
ZEB1*
SRF
SOX12*
NFKB1*
ETV6
HSF1
RFX3
ETS1*
PKNOX1
TCF7*
LEF1
NFATC2*
TGIF2
ESRRA
MYB*
FOXK1
FOXO4
CREB3*
TBX21
SOX4*
TCF7L2*
ZBTB7B*
HSF2
KLF7*
RORA*
ZBTB3*
FOXD2
MGA*
ETV3
NFAT5
NFIX*
CRX*
FOXD2*
EOMES
MGA
ZNF143*
HIF1A*
MYBL2
ZBTB18
USF1
SREBF2
SREBF1
MLX
GMEB2
EGR2
CENPB
E2F1
ZNF524
SPI1
RARG
RARA
MAFB
CTCF
KLF5
NKX3–1*
ZNF75A
JUNB
IRF*
HMBOX1
FOXP3
HLTF*
ZEB1
RELA
REL
BATF
ESR2
ETV1
STAT1
SREBF1*
RXRA
RORA
NFKB1
JUN
GABPA
CEBPB
FOS
PBMC
CD8+ M
CD8+ N
CD4+ M
CD4+ N

Fig S6

Sup Fig. 6. Aging-related loss of TF activity in memory CD8+ T cells. **(A)** Chromatin accessibility profiles of BACH2 and ETS1 in different T cell subsets. Note that these TFs are losing chromatin accessibility of their promoters with aging specifically in memory CD8+ T cells. **(B)** A heatmap representing whether a TF footprint is called near the IL7R gene promoter in a given cell type. Red (blue) represents that footprinting is found only in HO (HY) samples, whereas purple indicate that footprints for a TF motif are found in both HO and HY subjects. Note that about half of the listed TFs have footprints in all cell types.

Fig S7


A CD8+ T subpopulations
B CD4+ T subpopulations
C Major subpopulations
Age
CMV
Total.CD8
CD8.Naive
CD8.CM
CD8.EM
CD8.EMRA
Total.CD4
CD4.Naive
CD4.CM
CD4.EM
CD4.EMRA
CD4
CD8
CD14
CD19
D Chromatin changes with CMV status
Healthy Old
Healthy Young
Log Fold Change (CMV+/CMV-)
Peak intensity (logCPM)
E Variation explained pertaining to CMV positivity
All peaks, adjusted
Differential peaks, adjusted
Weighted average proportional variance (%)
CMV
Sex
AgeGroup
CMVxSex
CMVxAgeGroup
AgeGroupxSex
Residual
F Variation explained pertaining to seasonality
All peaks, normalized
Differential peaks, normalized
All peaks, adjusted
Differential peaks, adjusted
Season
Batch
Sex
AgeGroup
Residual
G Peak functional annotations
Percent of ATAC-seq peaks
All peaks
Opening (Summer)
Opening (Winter)
Closing (Summer)
Closing (Winter)
ChromHMM States
Enhancers
Quiescent
Repressed/PC
Promoter
Transcription
Other states

Sup Fig. 7. CMV and season-related chromatin changes in PBMCs. (**A**) Correlation between CD8+ T cell subset counts and age (left panels) and CMV antibody levels (right panels). Note that the most significant affect occurs on Naïve CD8+ T cells counts with aging. (**B**) Correlation between CD4+ T cell subset counts and age (left panels) and CMV antibody levels (left panels). (**C**) Correlation between major immune cell subset counts and age (left panels) and CMV antibody levels (right panels). (**D**) Plot showing log2 fold change of chromatin accessibility relative to CMV status (i.e., CMV+ vs. CMV-) versus average read count of ATAC-seq peaks when samples are stratified based on age groups. No significant chromatin remodeling associated with CMV status are obtained within either age group. (**E**) Variation explained by CMV and other factors of PBMC ATAC-seq peak counts, as quantified by PVCA. Note that neither CMV nor CMV*age interaction contribute much to the variance. Adjusted peak scores are obtained after fitting a generalized linear model (GLM) including CMV as a blocking factor. (**F**) Variation explained by aging, seasonality, sex, library preparation batch, and other factors of PBMC ATAC-seq peak counts. Note that seasonality, unlike sex or batch, contributes significantly to the variation, prompting its inclusion as a covariate in our models. Adjusted peak scores are obtained after fitting a GLM including season as a blocking factor. (**G**) ChromHMM annotations of all and differentially open/closed peaks obtained from summer and winter samples. Similar distributions are observed for both seasons, which also resembles the global pattern of variation.

BIBLIOGRAPHY

1. L. N. Booth, A. Brunet, The Aging Epigenome. *Molecular cell* 62, 728 (Jun 02, 2016).

2. C. Lopez-Otin, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The hallmarks of aging. *Cell* 153, 1194 (Jun 06, 2013).

3. B. K. Kennedy *et al.*, Geroscience: linking aging to chronic disease. *Cell* 159, 709 (Nov 06, 2014).

4. R. Zhang, H. Z. Chen, D. P. Liu, The Four Layers of Aging. *Cell systems* 1, 180 (Sep 23, 2015).

5. G. A. Poland, I. G. Ovsyannikova, R. B. Kennedy, N. D. Lambert, J. L. Kirkland, A systems biology approach to the effect of aging, immunosenescence and vaccine response. *Current opinion in immunology* 29, 62 (Aug, 2014).

6. T. Fulop *et al.*, Intracellular signalling pathways: targets to reverse immunosenescence. *Clinical and experimental immunology* 187, 35 (Jan, 2017).

7. J. J. Goronzy, F. Fang, M. M. Cavanagh, Q. Qi, C. M. Weyand, Naive T cell maintenance and function in human aging. *Journal of immunology* 194, 4073 (May 01, 2015).

8. D. B. Palmer, The effect of age on thymic function. *Frontiers in immunology* 4, 316 (Oct 07, 2013).

9. T. Fulop *et al.*, Cellular signaling in the aging immune system. *Current opinion in immunology* 29, 105 (Aug, 2014).

10. C. Franceschi, J. Campisi, Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *The journals of gerontology. Series A, Biological sciences and medical sciences* 69 Suppl 1, S4 (Jun, 2014).

11. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* 10, 1213 (Dec, 2013).

12. A. P. Boyle *et al.*, High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome research* 21, 456 (Mar, 2011).

13. S. E. Hardy, Y. Kang, S. A. Studenski, H. B. Degenholtz, Ability to walk 1/4 mile predicts subsequent disability, mortality, and health care costs. *Journal of general internal medicine* 26, 130 (Feb, 2011).

14. D. Podsiadlo, S. Richardson, The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society* 39, 142 (Feb, 1991).

15. K. Rockwood, E. Awalt, D. Carver, C. MacKnight, Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *The journals of gerontology. Series A, Biological sciences and medical sciences* 55, M70 (Feb, 2000).

16. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114 (Aug 01, 2014).

17. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754 (Jul 15, 2009).

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

19. B. Ewing, L. Hillier, M. C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research* 8, 175 (Mar, 1998).

20. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12, 323 (Aug 04, 2011).

21. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139 (Jan 01, 2010).

22. M. J. Boedigheimer *et al.*, Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. *BMC genomics* 9, 285 (Jun 12, 2008).

23. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882 (Mar 15, 2012).

24. K. Qu *et al.*, Individuality and variation of personal regulomes in primary human T cells. *Cell systems* 1, 51 (Jul 29, 2015).

25. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 38, 576 (May 28, 2010).

26. C. Roadmap Epigenomics *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317 (Feb 19, 2015).

27. D. Chaussabel *et al.*, A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 29, 150 (Jul 18, 2008).

28. G. Bindea *et al.*, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091 (Apr 15, 2009).

29. T. Kelder *et al.*, WikiPathways: building research communities on biological pathways. *Nucleic acids research* 40, D1301 (Jan, 2012).

30. M. Kutmon *et al.*, PathVisio 3: an extendable pathway analysis toolbox. *PLoS computational biology* 11, e1004085 (Feb, 2015).

31. R. I. Sherwood *et al.*, Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nature biotechnology* 32, 171 (Feb, 2014).

32. A. Jolma *et al.*, DNA-binding specificities of human transcription factors. *Cell* 152, 327 (Jan 17, 2013).

33. K. Daily, V. R. Patel, P. Rigor, X. Xie, P. Baldi, MotifMap: integrative genome-wide maps of regulatory motif sites for model species. *BMC bioinformatics* 12, 495 (Dec 30, 2011).

34. C. E. Romanoski, C. K. Glass, H. G. Stunnenberg, L. Wilson, G. Almouzni, Epigenomics: Roadmap for regulation. *Nature* 518, 314 (Feb 19, 2015).

35. H. R. Kim, M. S. Hong, J. M. Dan, I. Kang, Altered IL-7Ralpha expression with aging and the potential implications of IL-7 therapy on CD8+ T-cell immune responses. *Blood* 107, 2855 (Apr 01, 2006).

36. D. M. Moskowitz *et al.*, Epigenomics of human CD8 T cell differentiation and aging. *Science immunology* 2, (Feb, 2017).

37. C. I. Weidner *et al.*, Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome biology* 15, R24 (Feb 03, 2014).

38. S. Sauer *et al.*, T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR. *Proceedings of the National Academy of Sciences of the United States of America* 105, 7797 (Jun 03, 2008).

39. M. Maggio, J. M. Guralnik, D. L. Longo, L. Ferrucci, Interleukin-6 in aging and chronic disease: a magnificent pathway. *The journals of gerontology. Series A, Biological sciences and medical sciences* 61, 575 (Jun, 2006).

40. J. D. Buenrostro *et al.*, Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486 (Jul 23, 2015).

41. D. Frasca, A. Diaz, M. Romero, F. D'Eramo, B. B. Blomberg, Aging effects on T-bet expression in human B cell subsets. *Cellular immunology*, (Apr 24, 2017).

42. W. M. Passtoors *et al.*, IL7R gene expression network associates with human healthy ageing. *Immunity & ageing : I & A* 12, 21 (2015).

43. Y. M. Kerdiles *et al.*, Foxo1 links homing and survival of naive T cells by regulating L-selectin, CCR7 and interleukin 7 receptor. *Nature immunology* 10, 176 (Feb, 2009).

44. W. Ouyang, O. Beckett, R. A. Flavell, M. O. Li, An essential role of the Forkhead-box transcription factor Foxo1 in control of T cell homeostasis and tolerance. *Immunity* 30, 358 (Mar 20, 2009).

45. R. Grenningloh *et al.*, Ets-1 maintains IL-7 receptor expression in peripheral T cells. *Journal of immunology* 186, 969 (Jan 15, 2011).

46. S. Tani-ichi *et al.*, Interleukin-7 receptor controls development and maturation of late stages of thymocyte subpopulations. *Proceedings of the National Academy of Sciences of the United States of America* 110, 612 (Jan 08, 2013).

47. J. H. Park *et al.*, 'Coreceptor tuning': cytokine signals transcriptionally tailor CD8 coreceptor expression to the self-specificity of the TCR. *Nature immunology* 8, 1049 (Oct, 2007).

48. R. L. Collins *et al.*, Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome biology* 18, 36 (Mar 06, 2017).

49. T. R. Clinton *et al.*, Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. *Protein science : a publication of the Protein Society* 24, 446 (Apr, 2015).

50. M. E. Greenberg, T. P. Bender, Identification of newly transcribed RNA. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* Chapter 4, Unit 4 10 (Apr, 2007).

51. S. Kauppinen, B. Vester, J. Wengel, Locked nucleic acid (LNA): High affinity targeting of RNA for diagnostics and therapeutics. *Drug discovery today. Technologies* 2, 287 (Autumn, 2005).

52. J. L. Cotney, J. P. Noonan, Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. *Cold Spring Harbor protocols* 2015, 419 (Apr 01, 2015).

53. T. Matsui *et al.*, CD2 distinguishes two subsets of human plasmacytoid dendritic cells with distinct phenotype and functions. *Journal of immunology* 182, 6815 (Jun 01, 2009).

54. V. Appay *et al.*, Characterization of CD4(+) CTLs ex vivo. *Journal of immunology* 168, 5954 (Jun 01, 2002).

55. T. Namekawa, U. G. Wagner, J. J. Goronzy, C. M. Weyand, Functional subsets of CD4 T cells in rheumatoid synovitis. *Arthritis and rheumatism* 41, 2108 (Dec, 1998).

56. W. Xu *et al.*, Human plasma cells express granzyme B. *European journal of immunology* 44, 275 (Jan, 2014).

57. R. Ramalingam *et al.*, Dendritic cell-specific disruption of TGF-beta receptor II leads to altered regulatory T cell phenotype and spontaneous multiorgan autoimmunity. *Journal of immunology* 189, 3878 (Oct 15, 2012).

58. M. Kuwahara *et al.*, Bach2-Batf interactions control Th2-type immune response by regulating the IL-4 amplification loop. *Nature communications* 7, 12596 (Sep 01, 2016).

59. H. H. Nguyen *et al.*, Naive CD8(+) T cell derived tumor-specific cytotoxic effectors as a potential remedy for overcoming TGF-beta immunosuppression in the tumor microenvironment. *Scientific reports* 6, 28208 (Jun 16, 2016).

60. C. A. Goetz, I. R. Harmon, J. J. O'Neil, M. A. Burchill, M. A. Farrar, STAT5 activation underlies IL7 receptor-dependent B cell development. *Journal of immunology* 172, 4770 (Apr 15, 2004).

61. J. Krueger, C. E. Rudd, Two Strings in One Bow: PD-1 Negatively Regulates via Co-receptor CD28 on T Cells. *Immunity* 46, 529 (Apr 18, 2017).

62. S. L. Berger, P. Sassone-Corsi, Metabolic Signaling to Chromatin. *Cold Spring Harbor perspectives in biology* 8, (Nov 01, 2016).

63. B. A. Benayoun, E. A. Pollina, A. Brunet, Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nature reviews. Molecular cell biology* 16, 593 (Oct, 2015).

64. T. B. Kirkwood, Understanding the odd science of aging. *Cell* 120, 437 (Feb 25, 2005).

65. D. Gems, L. Partridge, Genetics of longevity in model organisms: debates and paradigm shifts. *Annual review of physiology* 75, 621 (2013).

66. J. Vijg, J. Campisi, Puzzles, promises and a cure for ageing. *Nature* 454, 1065 (Aug 28, 2008).

67. B. Liu *et al.*, Genomic instability in laminopathy-based premature aging. *Nature medicine* 11, 780 (Jul, 2005).

68. M. Armanios, E. H. Blackburn, The telomere syndromes. *Nature reviews. Genetics* 13, 693 (Oct, 2012).

69. K. L. Rudolph *et al.*, Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* 96, 701 (Mar 05, 1999).

70. P. Hainaut, A. Plymoth, Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Current opinion in oncology* 25, 50 (Jan, 2013).

71. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* 144, 646 (Mar 04, 2011).

72. T. Kuilman, C. Michaloglou, W. J. Mooi, D. S. Peeper, The essence of senescence. *Genes & development* 24, 2463 (Nov 15, 2010).

73. F. Rodier, J. Campisi, Four faces of cellular senescence. *The Journal of cell biology* 192, 547 (Feb 21, 2011).

74. P. V. Targonski, R. M. Jacobson, G. A. Poland, Immunosenescence: role and measurement in influenza vaccine response among the elderly. *Vaccine* 25, 3066 (Apr 20, 2007).

75. G. Gasteiger, A. Y. Rudensky, Interactions between innate and adaptive lymphocytes. *Nature reviews. Immunology* 14, 631 (Sep, 2014).

76. C. Munz, R. M. Steinman, S. Fujii, Dendritic cell maturation by innate lymphocytes: coordinated stimulation of innate and adaptive immunity. *The Journal of experimental medicine* 202, 203 (Jul 18, 2005).

77. R. Ostan *et al.*, Inflammaging and cancer: a challenge for the Mediterranean diet. *Nutrients* 7, 2589 (Apr 09, 2015).

78. D. Wilson, T. Jackson, E. Sapey, J. M. Lord, Frailty and sarcopenia: The potential role of an aged immune system. *Ageing research reviews* 36, 1 (Jul, 2017).

79. J. H. Hoeijmakers, DNA damage, aging, and cancer. *The New England journal of medicine* 361, 1475 (Oct 08, 2009).

80. D. Edgar *et al.*, Random point mutations with major effects on protein-coding genes are the driving force behind premature aging in mtDNA mutator mice. *Cell metabolism* 10, 131 (Aug, 2009).

81. A. Mackiewicz, H. Schooltink, P. C. Heinrich, S. Rose-John, Complex of soluble human IL-6-receptor/IL-6 up-regulates expression of acute-phase proteins. *Journal of immunology* 149, 2021 (Sep 15, 1992).

82. A. Harada *et al.*, Essential involvement of interleukin-8 (IL-8) in acute inflammation. *Journal of leukocyte biology* 56, 559 (Nov, 1994).

83. T. Andus, T. Geiger, T. Hirano, T. Kishimoto, P. C. Heinrich, Action of recombinant human interleukin 6, interleukin 1 beta and tumor necrosis factor alpha on the mRNA induction of acute-phase proteins. *European journal of immunology* 18, 739 (May, 1988).

84. C. Gabay, I. Kushner, Acute-phase proteins and other systemic responses to inflammation. *The New England journal of medicine* 340, 448 (Feb 11, 1999).

85. I. K. Chinn, C. C. Blackburn, N. R. Manley, G. D. Sempowski, Changes in primary lymphoid organs with aging. *Seminars in immunology* 24, 309 (Oct, 2012).

86. R. Aspinall, D. Andrew, Thymic atrophy in the mouse is a soluble problem of the thymic environment. *Vaccine* 18, 1629 (Feb 25, 2000).

87. J. Pido-Lopez, N. Imami, D. Andrew, R. Aspinall, Molecular quantitation of thymic output in mice and the effect of IL-7. *European journal of immunology* 32, 2827 (Oct, 2002).

88. D. H. Gray *et al.*, Developmental kinetics, turnover, and stimulatory capacity of thymic epithelial cells. *Blood* 108, 3777 (Dec 01, 2006).

89. X. Dou *et al.*, Short-term rapamycin treatment increases ovarian lifespan in young and middle-aged female mice. *Aging cell*, (May 22, 2017).

90. O. Lushchak *et al.*, The role of the TOR pathway in mediating the link between nutrition and longevity. *Mechanisms of ageing and development* 164, 127 (Jun, 2017).

91. C. Franceschi *et al.*, Inflamm-aging. An evolutionary perspective on immunosenescence. *Annals of the New York Academy of Sciences* 908, 244 (Jun, 2000).

92. C. consortium, Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523, 588 (Jul 30, 2015).

93. C. A. Meyer, X. S. Liu, Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature reviews. Genetics* 15, 709 (Nov, 2014).

94. A. P. Boyle *et al.*, High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome research* 21, 456 (Mar, 2011).

95. J. R. Hesselberth *et al.*, Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nature methods* 6, 283 (Apr, 2009).

96. S. Neph *et al.*, An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83 (Sep 06, 2012).

97. R. A. Dirks, H. G. Stunnenberg, H. Marks, Genome-wide epigenomic profiling for biomarker discovery. *Clinical epigenetics* 8, 122 (2016).