

**Influence of comorbidity burden and senescence on patterns of DNA methylation
in patients with Chronic Lymphocytic Leukemia**

by

Maria Shenna Fauni

**A Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of
the University of Manitoba
in partial fulfilment of the requirements of the degree of**

Master of Science

**Department of Biochemistry and Medical Genetics
University of Manitoba
Winnipeg**

Copyright © 2026 by Maria Shenna Fauni

Abstract

Chronic lymphocytic leukemia (CLL) is characterized by a heterogeneous clinical course, with some patients showing slowly progressive disease and others progressing rapidly, requiring early intervention. This variability complicates treatment decision-making and highlights the need for reliable prognostic indicators. One emerging factor is frailty, a multidimensional syndrome reflecting cumulative physiological decline across organ systems. In oncology, frailty is often estimated using the Cumulative Illness Rating Scale (CIRS), a weighted comorbidity index with prognostic value. However, the biological basis of CIRS and its relationship to molecular aging remain poorly understood. Because frailty shares physiological features with aging, biological aging markers, particularly epigenetic age derived from DNA methylation (DNAm) patterns, may offer mechanistic insight. Epigenetic age acceleration, where biological age exceeds chronological age, has been proposed as a predictor of frailty and disease risk. However, the accuracy of epigenetic clocks varies by tissue type, population, and training data. Cellular senescence, a hallmark of aging associated with a pro-inflammatory secretory phenotype, may also reshape the epigenetic landscape and influence these clocks. Yet whether senescence contributes to epigenetic age acceleration and clinical frailty in CLL remains unclear. To investigate this, we profiled DNAm in both malignant B cells and non-malignant immune cells from CLL patients aged 52–98 years, examining associations between epigenetic aging, CIRS scores, and known prognostic markers, including IGHV mutation status and CD38 expression. We also analyzed T cell subsets to determine whether cellular senescence affected epigenetic age estimates or their predictive accuracy. Our findings show that epigenetic age acceleration does not correlate with CIRS scores or poor prognostic markers in CLL. Epigenome-wide association studies (EWAS) identified no significant CpG sites associated with high CIRS scores. In addition, cellular senescence markers did not influence epigenetic clock accuracy or affect estimates of cell composition. These results suggest that DNAm-based epigenetic clocks, while valuable for studying aging, may have limited utility in capturing clinical frailty, at least as defined by CIRS, in CLL. Overall, this study highlights the complexity of aging and frailty in hematological malignancies and underscores the importance of integrating molecular, clinical, and immunological data when developing biomarkers for risk stratification and planning.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Meaghan Jones, for her guidance, patience, and the tremendous help she has provided throughout this journey. Her support has meant a great deal to me, and I feel very fortunate to have had her mentorship.

I would also like to thank my committee members, Dr. Aaron Marshall and Dr. Britt Drögemöller, for their thoughtful feedback and continued encouragement. I am equally grateful to the many previous and current members of the Jones Lab, who so willingly offered technical help and created such a welcoming environment, one filled not only with generosity but also with shared laughs and enjoyable moments along the way.

I would like to acknowledge the Marshall Lab, our CancerCare collaborators, and all others who contributed to this work. I am also very thankful to the staff in the BMG office, as well as the department chair, for their encouragement and for helping me navigate the many administrative stages of the program and stay on track. I am likewise grateful for the support I received from student support services, which has been indispensable. Outside of the program, I am deeply grateful to my supervisors and coworkers at work for their flexibility and understanding. Finally, I would like to thank my parents for their unwavering support, care, and love.

This journey has been challenging in many ways, and I would not have reached this point without the help of so many people. I will always be grateful for the academic, professional, and personal support I have received.

Table of Contents

List of Figures	7
List of Tables	8
List of abbreviations	9
1. Introduction	10
1.1. Chronic Lymphocytic Leukemia	10
1.2. B cell dysfunction in CLL	10
1.3. Non-modifiable risk factors of CLL.....	11
1.4. Etiology and Pathophysiology of CLL	12
1.5. CLL trajectory and the factors that affect prognosis.....	15
1.5.1. Complications	19
1.6. Aging.....	21
1.7. Hallmarks of Biological Aging	22
1.7.1. Cellular senescence's potential for elucidating biological aging.....	23
1.8. Frailty	25
1.8.1. Comorbidity burden as a proxy measure of frailty in CLL.....	26
1.8.2. Measuring comorbidity with the Cumulative Illness Rating Scale	27
1.9. DNA Methylation	28
1.9.1. Factors contributing to variation in DNAm.	29
1.10. DNAm patterns in common human diseases	31
1.10.1. Methods for analyzing DNAm and its successes.	33
1.10.2. Cell population heterogeneity compounds DNAm studies.	34
1.11. DNAm changes with age	35
1.11.1. Epigenetic Clocks.	37
1.11.2. Constructing Epigenetic Clocks	37
1.11.3. Different epigenetic clocks and their relevance in CLL studies.....	40
2. Rationale, Hypothesis, and Objectives	45
2.1 Cell type predictions and senescence	45
2.1.1 Rationale	45
2.1.2 Research Questions.....	45
2.1.3 Hypothesis.....	46
2.2 Epigenetic Age and Epigenetic Clock.....	46
2.2.1 Rationale	46
2.2.2 Research Questions.....	47
2.2.3 Hypothesis.....	47

2.3 Epigenome-wide association study	48
2.3.1 Rationale	48
2.3.2 Research Question	48
2.3.3 Hypothesis	48
3. Methods	49
3.1. Sample Selection and Preparation	49
3.2. CIRS Cut-off for Comorbidity Group Classification	50
3.3. DNA Extraction, Bisulfite Conversion and PCR	50
3.4. Measuring DNAm levels in the PBMCs	51
3.5. Preprocessing, Normalization, and Quality Control	53
3.6. Epigenetic Clock Output and Epigenetic Age Acceleration	53
3.7. Cell Type Estimations	54
3.8. EWAS	55
3.9. Statistical Analyses	56
4. Results	58
4.1 Study population characteristics	58
4.2 Cell type predictions	62
4.2.1 Prediction of cell types in B and non-B cells	62
4.2.2 Predictive accuracy of cell-type composition by deconvolution tools	67
4.2.3 Influence of senescent cell proportions on the accuracy of cell type predictions	74
4.3 Epigenetic age	79
4.3.1 Epigenetic age calculation and epigenetic age acceleration	79
4.3.2 Association between epigenetic age acceleration and CLL risk factors (comorbidity burden, sex, and smoking history)	84
4.2.3 Association between epigenetic age acceleration and CLL risk factors (IGHV mutation, p53 expression, Zap-70 expression, and CD38+ level of expression)	88
4.2.4 Association between epigenetic age acceleration and CLL risk factors (IGHV mutation, p53 expression, Zap-70 expression, and CD38+ level of expression)	92
4.2.5 Association between epigenetic age acceleration and senescent cell proportion	96
4.3 Epigenome-wide association study	102
5. Discussion	107
5.2. Disrupted epigenetic aging in CLL reflects cancer-driven methylation reprogramming	112
5.3. Reevaluating the link between epigenetic aging and senescence	115
6. Conclusion	121

7. References.....	123
---------------------------	------------

List of Figures

Figure 1. Distinguishing between biological and chronological age..	18
Figure 2. Relationship between epigenetic age and chronological age..	36
Figure 3. Comparison of DNA methylation–predicted and flow cytometry–measured CD4 ⁺ and CD8 ⁺ T-cell proportions.	59
Figure 4. Inverse relationship between DNAm-estimated B cell proportion and flow cytometry–derived non-B cell purity.....	60
Figure 5. Comparison of DNA methylation-predicted and flow cytometry–measured CD4 ⁺ and CD8 ⁺ T-cell proportions.....	71
Figure 6. Compared to CD8 ⁺ T cells, CD4 ⁺ T cells exhibited greater difference between flow-cytometry measured and DNAm-based predictors of cell proportions..	69
Figure 7. Relationship between senescent T-cell markers and deconvolution error in CD4 ⁺ and CD8 ⁺ T-cell proportion estimates.....	72
Figure 8. Relationship between naive CD8 ⁺ T-cell proportions and deconvolution prediction error.....	74
Figure 9. Relationship between chronological age and DNA methylation age in B-cell and non-B-cell fractions.....	79
Figure 10. Influence of sex, comorbidity burden, and smoking history on age acceleration in B and non-B cells of CLL patients.....	83
Figure 11. Comparison of epigenetic age between IGHV/ZAP-70 molecular subgroups in CLL B cells.....	85
Figure 12. Comparison of pcgtAge between IGHV/ZAP-70 molecular subgroups in CLL B cells... ..	91
Figure 13. Association between non-B-cell age acceleration and proportions of senescent T cells.....	94
Figure 14. Association between B-cell age acceleration and proportions of senescent T cells.....	95
Figure 15. Association between senescent T-cell proportions and pcgtAge in B and non-B cells..	97
Figure 16. Volcano plot of EWAS results for comorbidity burden in non-B cells.....	101
Figure 17. Manhattan plot of EWAS results for comorbidity burden in non-B cells....	102

List of Tables

Table 1. Key features of select epigenetic clocks: type, training characteristics, and performance metric.....	43
Table 2. Clinical and demographic characteristics of the CLL patient cohort.	60
Table 3. CpG sites highlighted in the volcano plot from the EWAS of comorbidity burden in non-B cells from CLL patients.	105

List of abbreviations

ACTN4	Actinin alpha 4
AHRR	Aryl hydrocarbon receptor repressor
BEpiC	Breast cancer-specific epigenetic clock
BMIQ	Beta-mixed quantile normalization
bp	Base pair
CCCP	Carbonyl cyanide m-chlorophenylhydrazone
CD	Cluster of differentiation
CIRS	Cumulative illness rating scale
CLL	Chronic lymphocytosis leukemia
CpG	Cytosine-guanine dinucleotides
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EAA	Epigenetic age acceleration
epiTOC	Epigenetic timer of cancer
EWAS	Epigenome wide association studies
FCR	Fludarabine-cyclophosphamide-rituximab
FDR	False discovery rate
hTERT	Human telomerase reverse transcriptase
IDAT	Intensity data
IDOL	Identifying optimal libraries
IGHV	Immunoglobulin heavy chain variable region
MBL	Monoclonal B lymphocytosis
NK	Natural killer
NKG2D	Natural killer group 2 member D
PBMC	Peripheral blood mononuclear cells
PCA	Principal component analysis
PCGT	Polycomb group target
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
ROS	Reactive oxygen species
RSME	Root mean square error
TMEM63C	Transmembrane protein 63C
TP53	Tumor protein p53
ZAP-70	Zeta-chain-associated protein kinase 70

1. Introduction

1.1. Chronic Lymphocytic Leukemia

Chronic Lymphocytic leukemia (CLL) is the most common form of leukemia among adults in Western countries, with an incidence rate of 4.2 cases per 100000 individuals in a year¹. In Canada, more than 2000 patients are diagnosed with CLL per year, with approximately 550 patients dying from its complications in 2022². This disease is highly prevalent among older individuals, with the median age of diagnosis ranging from 67 to 72 years old^{2,3}.

1.2. B cell dysfunction in CLL

CLL is characterized by the excessive proliferation of clonally related, immunologically dysfunctional B lymphocytes in the bone marrow and peripheral blood^{4,5}. To gain insights into the biology of CLL, it is important to first establish the primary function and developmental stages of B cells under normal circumstances.

B lymphocytes are integral components of the adaptive immune system, which primarily functions to produce a large and diverse repertoire of antibodies, serve as professional antigen-presenting cells, and secrete cytokines that contribute to inflammatory processes. Development of B lymphocytes occurs in the bone marrow, where hematopoietic stem cells differentiate into pro-B cells before differentiating into pre-B cells, during which the cells undergo a rearrangement of genes for immunoglobulin heavy chain (IGHV) and immunoglobulin light chains, respectively^{6,7}. Assembly of the heavy and light chains leads to the formation of B cell receptors (B-cell receptors) on the cell surface, which marks the transition of pre-B cells into immature B

cells^{6,8}. These cells undergo a selection process that ensure B cells can bind to their ligand to survive and that self-reactive B cells are eliminated⁹. Immature B cells leave the bone marrow for further differentiation and become naive B cells that circulate the immune periphery in search of antigens¹⁰. After encountering an antigen via the B-cell receptor, B cells migrate to secondary lymphoid tissues (e.g. spleen and lymph nodes), where the B cells are activated through a series of signalling events that lead to differentiation and proliferation¹¹. In the germinal centres of the secondary lymphoid tissues, B cells undergo somatic hypermutation in which deliberate point mutations are introduced into the gene encoding the variable region of the heavy and light chain of the immunoglobulins¹². The accumulation of mutations diversifies the B cell receptors to produce high-affinity antibodies that can effectively recognize antigens¹³.

Defects that occur during one of these stages in B cell development have significant repercussions for the B cell function. In fact, cytogenetic components that play a crucial role in normal B cell development are implicated in CLL development and provide prognostic information for clinicians, such as the mutation status of the IGHV gene^{1,14}. Further details will be discussed later.

1.3. Non-modifiable risk factors of CLL

The higher prevalence of CLL in older adults highlights the important role that advanced age plays as a risk factor. However, advanced age is not the sole determinant. Other non-modifiable risk factors for developing CLL include being male, of European descent, and having a family history of the disease.

The male-to-female incidence ratio for CLL is roughly 1.5 to 2, suggesting that men are diagnosed with CLL more often than women^{4,15,16}. Additionally, CLL is predominantly diagnosed among the white population. In the United States, the incidence rate of CLL among White Americans is reported to be 5.1 cases per 100,000 individuals¹⁷. In contrast, these rates are lower among African Americans (3.2 cases per 100,000) and Asian Americans (1.1 cases per 100,000)¹⁶. Furthermore, there is evidence indicating a familial link to CLL; first-degree relatives, such as parents, siblings, and children, of individuals diagnosed with CLL have an eightfold increased likelihood of developing the disease themselves¹⁸.

Recognizing and understanding the impact of non-modifiable factors such as sex, ethnicity, and familial history in the context of CLL pathophysiology is crucial for elucidating the disease's pathophysiology. These factors not only influence disease risk and progression but also inform clinical practice and targeted treatment strategies aimed at improving outcomes for CLL patients.

1.4. Etiology and Pathophysiology of CLL

In healthy adults, B cells comprise approximately 10% of all lymphocytes in the peripheral blood, and the absolute B cell count is approximately 200 cells per microliter^{19,20}. A formal diagnosis of CLL requires B-cell count of over 5000 cells per microliter for over 3 months, as confirmed by flow cytometry^{21,22}. Accumulation of these cells in the bone marrow may crowd out other leukocytes and interfere with blood cell development, which can lead to immune suppression^{23,24}. Aside from the excess lymphocyte count, CLL is characterized by the expression of the following set of surface

markers on B cells: CD5, CD19, CD20 (low), and CD23 (high)²¹. By comparison, normal B cells express CD19 (a universal marker of all B cells), CD20 (relatively higher), CD23 (low), and no CD5²⁵.

The origins of CLL and the mechanism by which it acquires malignancy have yet to be elucidated. Nonetheless, its pathogenesis appears to be marked by two primary events: the development of the precursor condition called Monoclonal B lymphocytosis (MBL) and the progression of MBL to CLL²⁶. Studies have found that the accumulation of genetic abnormalities leading to autonomous antigen-independent signalling through the B cell receptor pathway (causing an augmented B cell response observed in CLL) and malignant B cell interactions with the microenvironment have been implicated in the development of MBL and its progression to CLL^{27–29}. The CLL microenvironment is primarily found in lymph nodes, spleen, and bone marrow, where B cells interact with various functionally altered cells like nurselike cells, stromal cells, follicular dendritic cells, and T cells^{30–32}. These cells are involved in the release of chemokines to recruit CLL cells and adhesion molecules to maintain cell interaction and activate survival and proliferation pathways through CLL contact with its receptors³⁰. Due to the significant role played by the tumour microenvironment in the unregulated proliferation and survival of CLL cells, treatments have been developed to target one of these interactions. One example of such a therapeutic drug is ibrutinib, an inhibitor of Bruton's tyrosine kinase essential for B-cell receptor signalling^{33,34}.

Although the dysfunction of B-cells primarily defines CLL, T cells are also implicated in the pathogenesis of this disease. Similar to B cells, T cells play several crucial roles in the adaptive immune system, including destroying cancer cells,

damaged cells, and cells infected with pathogens; coordinating immune response through cytokine signals, developing immunological memory for more effective secondary immune response, and regulating immune cells to maintain self-tolerance³⁰. The most relevant function is that helper T cells stimulate the proliferation of B cells upon recognition of antigen peptides presented by the B cell. This, along with the B-cell receptor-mediated pathway, are essential components for B-cell activation. As such, aberrations like T cell exhaustion and immunological synapse defects can lead to a diminished ability to kill CLL cells and negatively impact B cell response that also relies on the cytokine signals from T cells^{35–38}. T cell exhaustion is a phenotype highly associated with the increased expression of inhibitory molecules like PD-1, leading to reduced proliferative capacity and cytokine production^{37,39}.

Epigenetic profiling has refined current models of CLL pathogenesis by clarifying the likely cell of origin. Beyond genetic lesions and microenvironmental influences, DNA methylation studies indicate that CLL does not originate from a single, uniform B-cell stage across all patients. Instead, CLL methylation patterns can be positioned along the same developmental continuum observed in normal B-cell maturation, suggesting that the malignant clone retains a methylation imprint of the B-cell differentiation state from which it arose²⁶⁰.

These maturation-linked methylation patterns are relatively stable and remain detectable after transformation, allowing CLL cases to be classified by their epigenetic similarity to less differentiated (naïve-like) versus more differentiated (memory-like) B-cell states. This “cell-of-origin” signal provides a biologically grounded explanation for inter-patient heterogeneity in CLL and has been linked to differences in clinical behavior.

For example, cases with methylomes that more closely resemble memory B cells have been associated with slower disease progression and more favorable outcomes, whereas more naïve-like profiles tend to align with more aggressive disease courses^{260, 261}. This maturation-linked epigenetic heterogeneity provides important context for prognostic markers such as IGHV mutation status and ZAP-70 and CD38 expression, which track with distinct biological subtypes.

1.5. CLL trajectory and the factors that affect prognosis.

CLL is known to have a heterogeneous trajectory for each patient, which complicates treatment decisions. For approximately one-third of diagnosed patients with CLL, the disease is stable or progressing very slowly. Patients in this group are likely to be asymptomatic, thus requiring no treatment at any point after their diagnosis⁴⁰. However, clinicians continue to monitor the status of this disease by conducting blood cell counts and clinical examinations every three to six months to monitor for physical symptoms of progression, such as enlarged spleen, swollen lymph nodes, a 50% increase in lymphocyte count over two months, bone marrow failure leading to anemia, abnormally low platelet count and lymphoproliferative disease symptoms like extreme fatigue, weight loss, fevers, and night sweats. The second one-third may show slow progression over several years before needing treatment. At this point, the patients are likely to exhibit indicators of CLL progression. If one of these signs has been documented, therapeutic intervention can be initiated. Lastly, the remaining one-third will follow an aggressive disease course in which the patient presents with signs of CLL progression, thus requiring immediate treatment^{41,42}. A thorough assessment of

prognostic markers will help determine high-risk patients for aggressive CLL and the optimal treatment options for them. It is important to add that the choice of treatment is dependent not only on prognostic markers but also on patient age and fitness, the latter measured by the degree of comorbidity.

There are four main prognostic molecular markers: IgHV mutational status, CD38 expression, ZAP-70 expression, and TP53 aberrations, that are tested prior to starting treatment.

1. Among the most informative biomarkers is the mutation status of the immunoglobulin heavy chain variable region (IGHV) genes, which encode the variable domain of the antibody's heavy chain component that play a crucial role in antigen-binding¹⁴. Therefore, under normal circumstances, the IGHV gene of B cell is expected to be mutated or more clinically defined as having a 2% or greater difference from the germline reference⁴³. Meanwhile, clinical studies have found that unmutated IGHV is associated with poorer prognosis, such as a more aggressive clinical course and a shorter time to first treatment for patients^{43–45}. Epigenetic profiling further suggests that this prognostic subgroup more often aligns with less differentiated, naïve-like methylation subtypes, whereas IGHV-mutated cases more often align with more differentiated, memory-like methylation subtypes, consistent with a maturation-linked cell-of-origin imprint^{260,261}. The exact reasons for this association are not fully understood, but it is proposed that it may be related to augmented

signalling facilitated by the B cell receptors of CLL cells with an unmutated IGHV⁴⁶.

2. Aberrations in the TP53 gene, whether through deletions on the short arm of chromosome 17 or direct mutations on the gene, are one of CLL's strongest prognostic markers^{1,14,47}. The TP53 gene encodes the tumour suppressor protein, p53, which serves a critical role in preventing cancer development by inducing a number of cellular responses including but not limited to apoptosis, DNA repair mechanism activation, and cell cycle arrest in response to DNA damage⁴⁸. Mutations in TP53s have been detected in about 50% of all cancers, making it the most frequently mutated gene in human cancers^{47,49}. CLL is no exception, as mutations in TP53 are strongly correlated with shorter progression-free survival and overall survival among patients with CLL^{47,50}.
3. Expression of CD38 in CLL cells was correlated with aggressive progression of the disease^{51,52}. CD38 is a transmembrane glycoprotein that functions as a receptor and signalling molecule in lymphocytes. CD38-mediated signalling was reported to influence B-cell receptor responses that foster abnormal B-cell proliferation⁵³. Furthermore, CD38+ promotes the migration of CLL cells to tumour microenvironments in proliferation centers of secondary lymphoid tissues^{53,54}. The percentage of CD38+ distinguishes the clinical outcomes of CLL patients as those with a high expression of CD38+ in CLL cells (20-30%) will require treatment earlier and have a higher risk of mortality⁵¹.

4. Zeta-chain-associated protein kinase (Zap-70) is a tyrosine kinase that is primarily expressed in T and NK cells and mainly functions by mediating signal transduction pathways to initiate T cell response^{55,56}. In CLL, Zap-70 was reported to have anti-apoptotic properties, thereby promoting CLL cell survival through enhanced B-cell receptor signalling and indirectly regulating the recruitment of T cells⁵⁶. Positive Zap-70 expression in tandem with CD38 expression was more receptive to B-cell receptor-mediated stimulation and correlated with greater migratory potential towards secondary lymphoid organs^{57,58}. Accordingly, Zap-70 expression is highly correlated with aggressive progression of CLL⁵⁹. Consistent with methylation-defined maturation subtypes described earlier, ZAP-70 positivity is more frequently observed in cases with less mature or naïve-like epigenetic profiles, whereas ZAP-70-negative cases more often align with more mature/memory-like epigenetic states^{260,261}.

Two of these prognostic markers, IGHV and TP53 gene mutation status, are strongly predictive of overall survival and have been used extensively to predict treatment response among CLL patients. The importance of these two markers for the determining choice of therapy is emphasized in the clinical guidelines for treating CLL released by the European Society of Medical Oncology in 2021¹. Patients with TP53 mutations are precluded from receiving chemoimmunotherapy (e.g. fludarabine, cyclophosphamide and rituximab; and bendamustine plus rituximab) due to poor response⁶⁰ 1. Meanwhile, chemoimmunotherapy drugs may be offered to CLL patients with mutated IGHV status but are not recommended for patients with unmutated IGHV.

CD38 and Zap70 expressions have been shown to predict the prognosis of CLL and have been used to stratify patients according to their prognostic outcomes (e.g. Good, poor, and intermediate)⁶¹. Although these two markers may not be powerful treatment predictors compared to IGHV and TP53, analysis of CD38 and Zap-70 still offer complementary information for assessing CLL progression in patients⁵⁸.

1.5.1. Complications

Among the leading causes of mortality among patients with CLL are disease progression, infections, and secondary cancers, which account for more than 50% of mortality cause in several cohort studies^{22,62}. Although the cited studies do not explicitly define disease progression in the context of mortality, it can be implied that it refers to complications that include the crowding out of leukocytes, which interferes with the blood development within the bone marrow and infiltration of the malignant cells into the liver and kidney, which impairs organ function. Disease progression is also linked to the onset of hypogammaglobulinemia, a condition characterized by low levels of antibodies and is observed in as high as 85% of CLL patients^{63,64}. This high prevalence is typically observed in groups of patients who have been diagnosed with CLL for multiple years, indicating that the severity of hypogammaglobulinemia tends to increase as CLL progresses and as more time passes since the initial diagnosis⁶⁵. One of the primary effects of hypogammaglobulinemia in CLL patients is the increased susceptibility to infections with studies reporting that up to 70% of patients experience at least one significant infection during their disease course^{66,67}. Specifically, patients with low immunoglobulin levels have a higher frequency of bacterial infections, particularly those

caused by encapsulated organisms such as *Streptococcus pneumoniae*⁶⁸. In addition, CLL patients are at a higher risk of developing secondary cancers compared to the general population which can be attributed to the compromised immune system in CLL patients. CLL is characterized by a dysfunctional immune response, particularly involving T-cells, which impairs the body's ability to surveil and eliminate emerging tumors^{69,70}. This immune dysregulation is further exacerbated by the disease itself, leading to an increased incidence of infections and malignancies, most including skin cancers and lymphoproliferative disorders^{71,72}. The emergence of secondary cancers in CLL patients poses several challenges. First, the presence of a secondary malignancy can complicate the management of CLL, as treatment regimens may need to be adjusted to address both conditions simultaneously⁷³. This can lead to increased morbidity and mortality, as the prognosis for patients with concurrent malignancies is often poorer than for those with CLL alone^{22,73}. Furthermore, the psychological burden of facing multiple cancer diagnoses can significantly impact the quality of life for these patients^{69,70}.

Given the highly variable course of CLL, identifying biomarkers that can accurately predict patient outcomes is of particular interest. Measures of patient fitness are among the necessary evaluations performed to maximize survival outcomes. However, the biological underpinnings of such clinical assessments have yet to be elucidated. Various biomarkers involved in inflammation, metabolism, telomere length, and DNA methylation offer a promising framework for understanding the pathophysiological mechanisms that underlie the origins and progression of CLL. Interestingly, these aforementioned biochemical changes are characteristic of the aging

process, which aligns with the observation that CLL is a cancer primarily found in older individuals. Therefore, novel insights into the mechanisms of CLL development and progression may be garnered by exploring the aging process and its relationship with CLL.

1.6. Aging

The risk of cancers such as CLL rises dramatically with age, pointing to potential shared mechanisms between the aging process and CLL pathophysiology. However, the fact that not all individuals of advanced age develop CLL underscores a crucial gap in our understanding of aging. This calls for deeper investigation into the biological underpinnings of aging, with a focus on identifying the specific pathological changes that differentiate those who age healthily from those who are at risk of developing CLL. Such investigative efforts start with the definition of aging, which has been colloquially

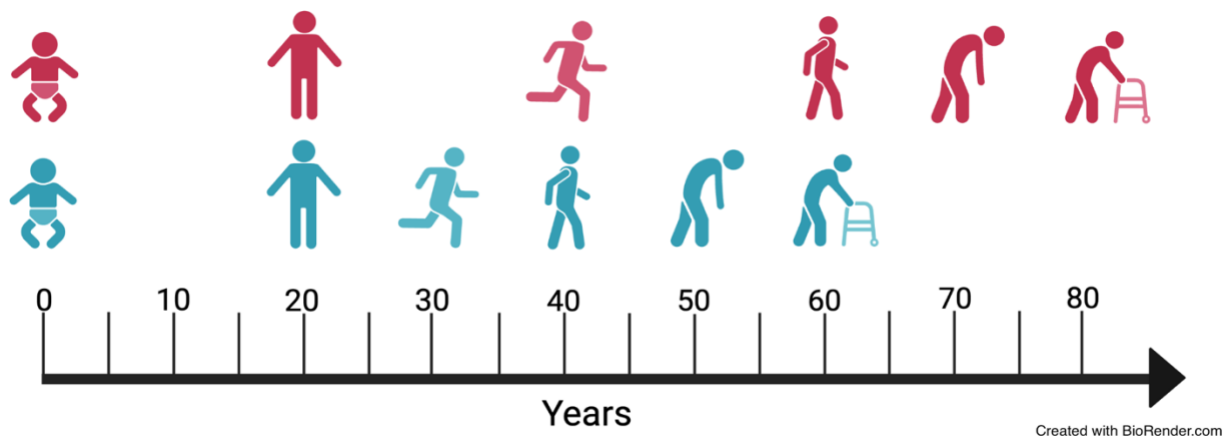


Figure 1. Distinguishing between biological and chronological age. Two individuals of the same chronological age may exhibit different phenotypic outcomes as they age. At the age of 60, one individual may still be partaking in marathons without the burden of the typical morbidities, while the other may already need a cane to assist with mobility. The combination of genetic and environmental factors can influence the rate of phenotypic aging, which is a reflection of biological aging.

referred to as chronological age or the time elapsed since birth. Aging is a multifaceted

natural process involving biological, psychological, and social changes that affect the individual's function^{74–76}. This complexity is a contributing factor to the lack of consensus over its formal definition in academic circles^{76,77}. For our purposes, we focus on the biological aspects of aging, which is commonly understood in the geriatrics field as the gradual decline of biological function across all levels of the human body resulting from cumulative cell damage over time. Biological aging is often discussed in the same context as age-related diseases like cancer and syndromes like frailty due to the significant overlap in the conditions and mechanisms that underlie them^{75,77}.

1.7. Hallmarks of Biological Aging

Since the publication of the landmark review *Hallmarks of Aging*, researchers have been equipped with a comprehensive framework to understand how complex biological processes contribute to aging phenotypes and aging itself. Lopez-Otin et al. (2013) listed nine hallmarks of aging: altered intercellular communication, cellular senescence, deregulated nutrient-sensing, epigenetic alterations, genomic instability, loss of proteostasis, mitochondrial dysfunction, stem cell exhaustion, and telomere attrition⁷⁸. The ever-expanding knowledge of the aging mechanism driven by continuous research has led to the addition of disabled macro autophagy, chronic inflammation, and dysbiosis, thus increasing the number of aging hallmarks from nine to twelve in the paper's most recent update⁷⁹. Of the 12 hallmarks, we focus on cellular senescence and epigenetic alterations for two main reasons. Firstly, changes in the epigenetic profile, particularly in DNAm patterns, has been emerging as the most robust markers of biological age^{76,80}. This attribute positions it as an ideal medium for elucidating the

complex interplay between the aging process and CLL development. Secondly, the relationship between cellular senescence, aging, and DNAm alterations remains largely unexplored. While cellular senescence is recognized as a key feature of biological aging, there has been minimal research into the interactions between senescence and DNAm-based measures of aging, as well as its potential contribution to CLL development.

1.7.1. Cellular senescence's potential for elucidating biological aging.

Cellular senescence is a prolonged state of cell division arrest characterized by phenotypic changes, including chromatin remodelling and the senescence-associated secretome⁸¹. While it is primarily understood to function as an anti-cancer mechanism by preventing the proliferation of damaged cells, cellular senescence is also implicated in aging and age-related diseases. In younger individuals, senescent cells are typically cleared by immune cells such as macrophages and T cells⁸². However, as individuals age, the ability to eliminate these cells diminishes due to their acquired resistance to apoptosis and evasion of immune surveillance, such as downregulating NKG2D ligands to avoid detection by NK and T cells^{82–86}. In the absence of effective mechanisms to remove the accumulating senescent cells, increased pro-inflammatory signals sustain chronic inflammation, which is believed to accelerate the aging process^{82,84}.

While there is a recognized link between cellular senescence and aging, the question remains: how well is senescence correlated with other molecular changes associated with aging, particularly with DNA methylation? After all, they share similarities. Both senescence and aging are associated with specific alterations in DNA

methylation patterns, which may make it difficult to pinpoint whether DNAm alteration are specifically due to aging itself, cellular senescence, or other external factors or biological processes. For example, Sidler et al. reported that senescent human fetal lung fibroblast cells exhibited global and site-specific DNA hypomethylation that are also observed in aging cells, indicating that age-associated methylation patterns may not solely correlate with chronological age but also with the senescent state of the cells⁸⁷. Furthermore, DNA methylation patterns can serve as biomarkers to predict biological age and reflect the senescence status of cells^{88,89}. Although these DNAm changes occur at different genomic sites, they share similarities in that both age- and senescence-associated DNAm are completely reset during reprogramming into induced pluripotent stem cells, implying that both processes are regulated by similar epigenetic mechanisms⁹⁰.

To explore this connection further, we investigated the association of age-related DNA methylation changes with cellular senescence, which is characterized by a variety of markers that can be categorized into morphological, biochemical, and molecular indicators. For this project, we used a combination of two surface markers, CD28⁻ and CD57⁺, to identify senescent T cells. CD28 is a costimulatory receptor expressed on CD4⁺ and CD8⁺ T cells, and it plays a significant role in T cell activation and function⁹¹. Aging and chronic infections, such as HIV and chronic hepatitis B, are associated with an impaired immune response attributed to the accumulation of CD28⁻ T cells^{92,93}. These cells result from repeated antigenic stimulation, where continuous exposure leads to a decrease in costimulatory signals provided by CD28, which is crucial for T cell activation and differentiation. Studies indicate that persistent stimulation, whether by viral antigens

or proinflammatory signals, causes CD4⁺ and CD8⁺ T cells to progressively downregulate CD28 expression, resulting in a population of dysfunctional T cells that are less capable of responding to new antigens and increased susceptibility to persistent infections^{92,93,94}. In addition to the downregulation of CD28, senescent T cells are characterized by increased expression of CD57, a recognized marker of terminal differentiation in both NK and T cells⁹⁵. The accumulation of CD57⁺ cells with age reflects the decline in the proliferative capacity of T cells, which is a key feature of senescence^{96,97}.

1.8. Frailty

Here, we pivot from biochemical markers to physiological manifestations of aging, such as frailty, taking in a more interdisciplinary perspective to understand how the aging process leads to diseases associated with advanced age. Frailty has become the umbrella term for describing the accumulation of physiological function declines leading to reduced ability to cope with everyday stressors and increased vulnerability to adverse health outcomes like falls, disability, hospitalization, and mortality^{98–100}. Due to the lack of a consensus definition, studies operationalize and measure frailty differently^{101–105}. Two of the most widely used tools for assessing frailty in older adults are the Fried frailty criteria and Rockwood's frailty index.

The Fried frailty criteria are based on the phenotype model, defining frailty in terms of compromised energetics specifically by satisfying three of the following phenotypes: weight loss, exhaustion, low physical activity, slowness, and weakness¹⁰⁵. Meanwhile, Rockwood's frailty index, based on the cumulative deficit model, measures

70 clinical deficits, and its interpretation depends on the clinician and the nature of the patient's illness¹⁰⁴. Several studies rely on these two tools underscore the importance of frailty as a significant predictor of adverse clinical outcomes, including increased hospitalization use and mortality^{99,100,106,107}. Nevertheless, both have limitations. Fried's frailty criteria require objective measurements of gait speed and grip strength to be assessed in clinics, while the Rockwood frailty index evaluates a broad range of health deficits, including physical, social, and cognitive functioning, which requires data collection from multiple sources (e.g., medical history, lab tests, and clinical exams). Considering the need for additional resources and specialized healthcare providers to conduct assessments, oncological clinics opt to measure surrogate measures of frailty like comorbidity burden, which offers a simpler and more efficient way of assessing frailty using the patient's medical records^{108–110}.

1.8.1. Comorbidity burden as a proxy measure of frailty in CLL.

Comorbidity, the concurrent presence of two or more medical conditions or diseases, and frailty are interrelated and are often observed among older individuals^{111,112}. Although comorbidity burden and frailty are distinct, there is a strong overlap between the two, which is evident in their association with adverse health outcomes and their impact on treatment decisions. For instance, Fernandez et al. showed that non-frail patients with end-stage renal disease with a high comorbidity burden have an increased risk of waitlist mortality similar to their frail counterparts¹¹³. Meanwhile, Huang et al. found that the presence of comorbidities significantly affects

the frailty risk of patients with chronic obstructive pulmonary disease, which has clinical implications for tailoring preventative and treatment interventions¹¹².

These studies suggest that comorbidity captures similar aspects of health deterioration as frailty. This, along with the frequent co-occurrence of the two conditions, presents a stronger argument for comorbidity to be a proxy measure of frailty¹¹¹. By utilizing comorbidity as a surrogate measure, clinicians can effectively assess frailty and its implications for patient care in oncology settings.

1.8.2. Measuring comorbidity with the Cumulative Illness Rating Scale. Patients newly diagnosed with CLL often present with comorbidities, of which the most common are diabetes mellitus, congestive heart failure, other forms of cancers, and chronic lung disease^{44,114}. The most common measure of comorbidities in CLL is the Cumulative Illness Rating Scale (CIRS) in leukemia clinics. CIRS rates the impairments of 13 organ systems on a five-point severity scale, and the resulting scores are associated with patient outcomes following treatment^{115,116}. Validation studies found a strong association between CIRS and adverse long-term outcomes such as mortality, acute hospitalizations, and functional disabilities among older and frail populations^{117,118}. Although CIRS is not a validated measure of frailty, it is considered a proxy indicator of frailty since it evaluates health aspects (i.e. comorbidities) that share similarities with the concept of frailty.

When CLL patients have progressed to the point of requiring treatment, one of the required assessments is to determine the patient's fitness to tolerate toxic chemotherapy. Fitness determination primarily relies on the patient's chronological age

and comorbidity burden¹. Standard chemoimmunotherapy for CLL such as fludarabine-cyclophosphamide-rituximab (FCR), are reserved for younger individuals (less than age 65) who are more likely to tolerate the toxicity associated with FCR^{119,120}. Patients older than 65 with significant comorbidities are more likely to be prescribed alternatives like the combination of chlorambucil and obinutuzumab, which is reportedly to improve patients' treatment response rate without the adverse side effects of FCR^{1,121}.

Functional impairments, presence of comorbidities, and organ function are also important components of accurately evaluating fitness, especially for older individuals¹.

A combined personalized assessment of prognostic markers and fitness status allows for an optimized treatment plan aiming to maximize survival time and quality of life among CLL patients¹⁰⁹. Nonetheless, the biological underpinnings of comorbidity and other proxy measures of frailty have yet to be elucidated. Given the strong correlation between frailty symptoms and advanced aging, which share similarities with the loss of homeostatic regulation, a thorough examination of DNA methylation, one of the most robust markers of biological aging, and its mechanistic implications could offer novel perspectives^{122,123}.

1.9. DNA Methylation

One of the most intensely studied epigenetic marks is DNAm, which involves the addition of a methyl group to the 5th carbon of the cytosine's pyrimidine ring adjacent to the guanine nucleotide known as cytosine-guanine dinucleotides (CpG) sites^{124,125}. Of the 28 million CpGs distributed unevenly across the human genome, most are found in non-coding regions that are typically methylated^{126,127}. However, a small fraction of the

total CpGs occur in clusters called CpG islands. CpG islands are formally defined as DNA regions of at least 500 bp long with over 55% GC content and greater than 65% observed to the expected ratio of CpG¹²⁸. CpG islands carry functional significance due to their proximity to promoters and important role in gene regulation. In fact, more than 70% of gene promoters harbour CpG islands that are generally unmethylated^{124,129}. Methylation of CpG islands located in promoters of normal cells is known to play a role in gene regulation and lead to gene silencing through direct and indirect means^{124,130}. Mechanisms of direct suppression of gene expression involve inhibiting transcription by blocking the binding of transcription factors to the DNA^{131,132}. Meanwhile, indirect mechanisms of gene repression include recruiting repressive proteins like methyl-binding proteins to cooperate with corepressors to modify the chromatin structure, making it less accessible to transcription factors^{132,133}.

1.9.1. Factors contributing to variation in DNAm.

DNAm has traditionally been viewed as a gene repression mechanism, especially in the promoter regions. However, further studies have come to challenge this view. Positive correlations between DNAm in gene bodies and gene expression have been reported in normal and malignant cells^{134–136}. This positive relationship appears to be more common in proliferating cells, where genes are actively transcribed. In contrast, in slowly dividing cells or tissues, such as the brain, gene body methylation is often inversely correlated with gene expression.^{137–139} Overall, the complex regulatory nature of DNAm leads to significant differences in both global and specific DNAm patterns

across various tissue and cell types. These variations contribute to the distinct gene expressions profiles, which define the identity and function of each cell or tissue type.

DNAm patterns do not just differ by cell type or tissue but also by sex. Czamara et al. identified 3031 CpGs located on autosomes that were differentially methylated between newborn males and females¹⁴⁰. In a later study using improved methylation assays with more stringent cutoffs, Grant et al. identified 396 strongly differentially methylated CpGs between adult males and females¹⁴¹. These sex-specific DNAm patterns can have profound implications for understanding differences in the risk and progression of various health conditions as seen in cardiovascular diseases and neurological disorders^{142,143}.

Additionally, variation in DNAm patterns between different ethnic groups and between ethnic subgroups have been identified. A study comparing first-generation South Asian immigrants and European-origin individuals living in the UK found 16 433 differentially methylated CpGs attributed to ethnic differences¹⁴⁴. In another study comparing Hispanic subgroups (e.g. Puerto Rican, Mexican, Other Latino, and Mixed Latinos), 194 differentially methylated CpGs were found to be associated with ethnicity, most of which are reported to be influenced by genetic ancestry and environmental factors¹⁴⁵. Although ethnicity is primarily a social construct with contentious biological significance, specific ethnic groups can be associated with certain genetic markers inherited from common ancestors, indicating shared ancestry. Moreover, members of ethnic groups often share cultural elements (e.g. dietary habits) and social experiences (e.g. psychosocial stress), which can influence their environmental exposures and susceptibility to diseases^{145–148}. For example, traditional diets emphasizing vegetables,

fruits, and whole grains, often associated with the Mediterranean diet, correlate with methylation patterns observed in cancer-free individuals^{149,150}. Moreover, prolonged chronic stress experienced by disadvantaged ethnic groups may significantly alter methylation levels at select CpG sites^{146,151}.

Differentially methylated CpGs have also been associated with exposure to specific environmental chemicals, such as heavy metals (e.g., cadmium, lead, and mercury), tobacco smoke, and ambient air pollution, which are said to affect the enzymatic activities involved in maintaining methylation patterns^{152,153}. While environmental factors have been independently shown to explain variations in numerous studies, it is important to acknowledge that the effects of these factors can also be modified by ethnic, social, and cultural influences. For instance, marginalized populations or ethnic groups associated with disadvantaged socioeconomic status are often at a higher risk of exposure to environmental pollutants due to the characteristics of their residential neighborhoods^{154–157}.

Given the significant methylation differences by sex, cell type or tissue distribution, ethnicity, and environmental exposures, it is crucial for epigenomic studies to consider these factors as potential confounders when investigating differential methylation attributed to biology^{158,159}.

1.10. DNAm patterns in common human diseases

The regulation of gene expression underscores the importance of DNAm's role in various cellular processes including embryonic development, tissue differentiation, maintenance of chromosomal stability, and X chromosome inactivation¹²⁴. As such,

perturbations in DNAm and its regulatory mechanisms have been implicated in the pathogenesis of various cancers (e.g. colon, breast, prostate), atherosclerosis, and inflammatory bowel diseases.

Several clinical studies report aberrant DNAm patterns in cancerous phenotypes compared to those classified as healthy or normal. For instance, global hypomethylation, defined as the decrease in DNA methylation levels across the genome relative to control, has been confirmed in malignant prostate, breast, and colon tissues^{160–164}. Conversely, hypermethylation, or higher methylation levels relative to control, at specific genomic sites, were found in promoters of tumour suppressor genes in colorectal cancer as well as in the non-promoter regions (e.g. gene bodies) in breast cancer^{165–167}. In addition to malignant phenotypes, DNAm variants have also been associated with adverse health outcomes such as (i.e. diseases progression, treatment resistance, shorter survival, and higher mortality)^{168–172}

Overall, these studies not only underscores the significance of CpGs in understanding the dysregulated mechanisms leading to disease but also suggest the potential of DNAm variants to serve as reliable biomarkers for prognostic, predictive, and therapeutic purposes^{170,171}. While the discovery of candidate DNAm biomarkers have profound clinical implications, it is important to note that these breakthroughs are largely supported by the emergence of high-throughput sequencing technologies and complex algorithms to analyze high-volume epigenomic data.

1.10.1. Methods for analyzing DNAm and its successes.

As genetics research advances, more DNAm loci that are strongly associated with diseased traits or adverse health outcomes are identified by epigenome-wide association studies (EWAS). EWAS is a research approach that can be divided into two main components: bench work and computational analysis¹⁷³. The bench work component involves interrogating hundreds of thousands of CpGs across the genome of a given sample using arrays like the Infinium methylation assays or high resolution techniques like next-generation sequencing¹⁷⁴. Meanwhile, the computational analysis segment creates a predictive model from the profiled epigenome to identify DNAm variants with the most robust association with the phenotypes of interest¹⁷⁴. Over the last decade, EWAS has made several achievements in identifying significant associations while investigating various diseases and exposures^{175–179}. Among the notable early works on EWAS is the analysis to determine differentially methylated CpGs associated with smoking. Breitling et al. discovered a significantly differentiated CpG located on a gene that had not been linked to smoking prior to the study^{180,181}. Further applications of EWAS in conjunction with improved Infinium methylation assays have expanded the number of identified differentially methylated CpG linked to smoking into the thousands. Some of these findings (e.g. cg05575921 in *AHRR*) have been validated in subsequent studies, with some evidence suggesting its potential clinical relevance for assessing risk of smoking-related morbidities (e.g. lung cancer) and mortality^{182–184}. For complex diseases, EWAS has identified CpGs with significant associations with other types of cancers (e.g. breast, colorectal), autoimmune (e.g.

rheumatoid arthritis), metabolic (e.g. Type II diabetes), and neurodegenerative (e.g. Alzheimer's) disorders^{185–189}.

1.10.2. Cell population heterogeneity compounds DNAm studies.

Despite the successes of EWAS, some challenges arise with this approach; thus, caution must be exercised when interpreting the results. Many epigenome-wide association analyses were performed on easily accessible bulk tissues like whole blood, which is composed of different cell types. Seven major leukocyte cell types (neutrophils, eosinophils, basophils, monocytes, B cells, CD4 +T and CD8 + T cells) can be identified from whole blood, each cell type exhibiting unique DNA methylation profiles^{190,191}. Because DNAm data resulting from microarray is the average methylation across a mixed population of cells, it can be challenging to distinguish DNAm alterations attributed to biology (i.e. phenotype of interest) from DNAm alterations due to shifts in cell proportions^{144,179,192,193}.

To address these issues, correction methods were developed to distinguish true differential DNAm from differential DNAm attributed to cell composition. One such tool is cell deconvolution. Cell deconvolution is a computational technique for predicting cell type proportions from heterogeneous samples using DNA methylation profiles. Several of these tools use references of cell-type-specific methylation signatures derived from purified cell types of young and healthy populations to differentiate between cell types^{191,194–197}. Integrating cell deconvolution algorithms in EWAS is crucial for ensuring the accuracy and reliability of the findings. This necessity is highlighted in the study by Liu et al., who investigated rheumatoid arthritis cases compared to controls using whole

blood samples¹⁷⁹. Prior to adjusting for cell type distribution and other confounding variables such as age, sex, and smoking status, the study identified approximately 51,000 CpG sites associated with the disease. However, following these adjustments, significant associations were markedly reduced to around 9,400 CpG sites¹⁷⁹. Similarly, Bauer et al. discovered that the differentially methylated expression in a CpG site that has been reported to be significantly associated with smoking was actually influenced by variations in T cell subtypes within the peripheral blood samples¹⁹⁸. These studies highlight the critical role that proper adjustment for cell type composition and other confounders plays in obtaining accurate and reliable results in EWAS.

Thus far, we have introduced one approach for analyzing DNAm data: identifying methylation markers associated with diseases, traits, or exposures using EWAS methods. Although EWAS serves as a powerful tool for understanding epigenetic variations and disease risk, it is important to take precautions to minimize bias by adjusting for major confounders, such as cell type composition. In subsequent sections, we explore an alternative method of analyzing DNAm data, which focuses on understanding inter-individual differences in aging and their implications for disease development and progression risk.

1.11. DNAm changes with age

Previously, we discussed briefly that there are DNA methylation alterations associated with aging. Another important feature of epigenetic aging is that inter-individual DNA methylation variability increases with age. This is well illustrated by monozygotic twin studies, which show that twins exhibit virtually identical methylomes

but become progressively more epigenetically discordant over time, consistent with accumulating stochastic drift and divergent environmental and lifestyle exposures¹⁹⁹. Moreover, a comparison study of newborn and centenarian methylomes shows that the latter had lower methylated CpGs in promoters, exons, introns, and intergenic regions²⁰⁰. It is important to note that age-associated methylation changes are not limited to hypomethylations. Methylation gains were found in several CpG islands, many of which are located in the promoters of developmental genes that also encode tumour suppressors and transcription factors for differentiation^{201,202}.

Evidence suggests that these methylation changes may be linked to gradual deregulation of epigenetic machinery that maintains DNAm patterns may be involved^{178,202–204}. A study by Ciccarone et al. reported that the expression of DNMT1, an enzyme mainly responsible for maintaining methylation patterns, decreases with age²⁰². Meanwhile, Xiao et al. found that expression levels of DNMT3a and DNMT3b, both of which are enzymes similar to DNMT1s that are known to catalyze methylation of unmethylated DNA during early development, are higher in older individuals compared to their younger counterparts²⁰⁴. While the results of these studies have provided valuable insights into the dynamic methylation process, the precise biological mechanisms regulating methylation alterations over time are still not well understood²⁰⁵. Nonetheless, this phenomenon of widespread changes in DNA methylation (DNAm) patterns throughout the genome, often referred to as epigenetic drift in the literature, is believed to underlie the predictive DNAm changes captured by mathematical models of age predictors, such as epigenetic clocks.

1.11.1. Epigenetic Clocks.

The discovery of differentially methylated CpGs that strongly correlate with chronological age and are well-validated across multiple samples led to the development of epigenetic clocks, models that accurately predict age based on methylation profiles. The first generation of epigenetic clocks was considered chronological clocks since the models only contained the parameters for chronological age^{206,207}. These clocks generally perform well overall; however, prediction error and bias are not necessarily uniform across the lifespan, and several studies have reported larger deviations at the extremes of age (e.g., younger adults and older adults), with the direction and magnitude of bias varying by clock and tissue²⁵⁷⁻²⁵⁹. Although chronological clocks accurately predict chronological age, these tools do not fully capture the age-related decline in cell or tissue function that is reflected in biological age^{208,209}. Since biological age is more closely tied to health outcomes, disease risk, and mortality risk, chronological clocks were reported to have relatively limited clinical relevance.

1.11.2. Constructing Epigenetic Clocks.

Epigenetic clocks are derived from linear regression algorithms trained by a large dataset of DNAm profiles and corresponding chronological ages of the sample donors. The algorithm learns the relationship between DNAm patterns and the sample donors' chronological age, resulting in a model that can predict age based on the methylation profiles of a given sample²⁰⁶. To improve model performance, the number of parameters (i.e. number of CpGs) in a model must be reduced from 27,000-800,000 CpG,

depending on the array used for interrogation, to only the most informative CpGs for age prediction. Therefore, feature selection techniques are applied to select only the CpGs with the strongest and most consistent correlation with chronological age^{222,223}.

These steps are sufficient to produce chronological clocks that predict age based on DNAm using only chronological age as their parameter. Meanwhile, biological clocks consider not only chronological age but also molecular and physiological changes that occur with age. Biological clock parameters may also include clinical measurements pertaining to diseases, making this clock more informative of disease prognosis and mortality.

Improving on the shortcomings of chronological clocks ushered in a new generation of epigenetic clocks that reflect the physiological changes during aging. Biological clocks, such as phenotypic or mitotic clocks, are more precise predictors of individuals' health outcomes and disease risk than chronological clocks since the models account for biological health measurements. For example, PhenoAge incorporates nine clinical biomarkers of aging, such as albumin, creatinine, glucose serum, c-reactive protein, lymphocyte percent, mean cell volume, red cell distribution width, alkaline phosphatase, and white blood cell count, in addition to chronological age²⁰⁸. As a result, PhenoAge outperformed chronological clocks (e.g. Horvath and Hannum) in predicting morbidity and mortality risk^{188,210–212}. Other biological clocks measure specific molecular changes (e.g. stem cell division in epiTOC clocks) related to aging and output a predictive score indicating disease risk²¹³.

Epigenetic clocks capture the changes in DNAm patterns that occur with normal aging. The risk of developing adverse health conditions are said to be associated with

accelerated aging, which is also captured by epigenetic clocks. Epigenetic age acceleration is the residuals of regressing DNAm age on chronological age. Regardless of which epigenetic clock is used (whether you choose to use biological or chronological), these tools provide an estimate of biological age (i.e. DNAm age) that is compared to chronological age to assess the rate of aging.

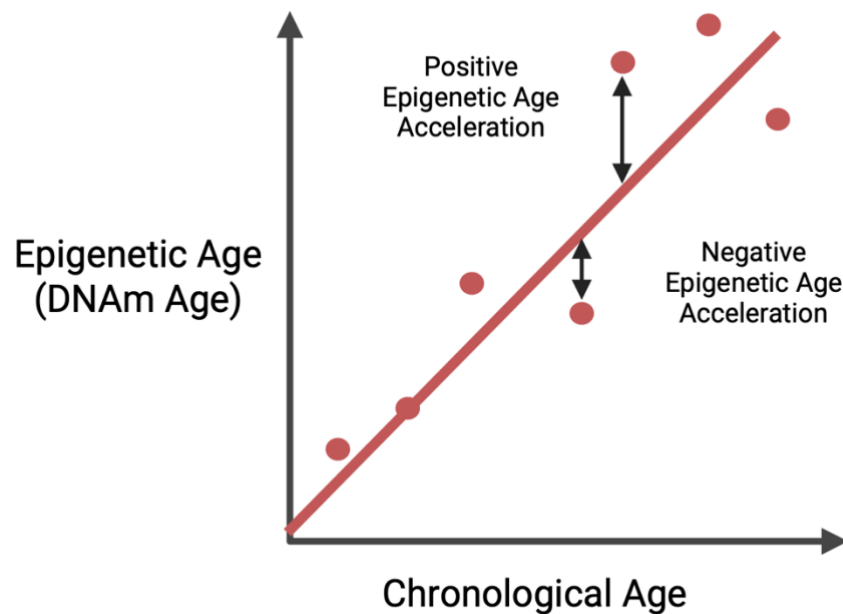


Figure 5. Relationship between epigenetic age and chronological age. In ideal circumstances, epigenetic age closely tracks chronological age, following a one-to-one line relationship. Double-headed arrows indicate residuals derived from regressing epigenetic age onto chronological age. Positive age acceleration occurs when epigenetic age exceeds chronological age, suggesting faster biological aging. Conversely, negative age acceleration reflects a lower epigenetic age relative to chronological age, indicating slower biological aging.

If the estimated DNAm age is higher than the chronological age, this indicates accelerated aging or epigenetic age acceleration (EAA). EAA has been linked to increased risk of adverse health outcomes such as cardiovascular diseases, some cancers (e.g. breast, and head and neck) and overall mortality^{214–217}. Conversely, if

estimated DNAm age is lower than chronological age, this indicates slower aging or aging deceleration. Although much of the literature focuses on EAA, it has been posited that age deceleration may be associated with longevity and healthier lifestyle habits (e.g. diet, alcohol consumption, and smoking)^{218–221}.

1.11.3. Different epigenetic clocks and their relevance in CLL studies.

To our knowledge, more than 20 epigenetic clocks for humans have been published. Each clock exhibits unique features, but in general, the differences include the types of tissues used to train the clock, and the age of the samples. We chose four clocks for our studies: Hannum, Horvath, EpiTOC, and PhenoAge.

Introduced in 2013, the Hannum clock was developed using 71 CpG sites from adult whole blood samples²⁰⁷. Hannum clock correlates well with chronological age in blood samples and has been used to predict adverse health outcome risks in psychiatric disorders such as schizophrenia, cancer (colon, colorectal, breast, pancreatic), and overall mortality^{188,216,224}. However, because Hannum was specifically trained on blood samples, it is not a highly accurate age predictor in non-blood tissues, thereby limiting its generalizability across different populations and conditions.

In contrast, the Horvath clock, which was also published in 2013, is a pan-tissue epigenetic clock that uses 353 CpG sites to predict chronological age across various tissues and cell types²⁰⁶. Its development was considered significant in the epigenetic field as it was trained in various tissues and cell types, making it robust against variations in tissue and cell type composition. Due to its versatility, the Horvath clock enabled researchers to explore the effects of aging in various contexts. While

investigating the relationship between epigenetic aging and risk of selected cancers, Durso et al. showed that accelerated biological aging, as measured by the Horvath clock, is associated with an increased risk of breast and colorectal cancer, highlighting the clock's potential as a clinical biomarker²²⁵. The Horvath clock has also been applied to assess biological aging in the context of neurodegenerative disorders and psychosocial stressors. Kular et al. reported that patients with progressive multiple sclerosis exhibited accelerated epigenetic aging as measured by the Horvath clock, providing insights into the mechanisms underlying the disease's progression²²⁶. Additionally, research has shown that adverse childhood experiences are associated with accelerated epigenetic aging, as measured by the Horvath clock, indicating that early life stress can have long-lasting impacts on biological health^{227–229}. The Horvath clock, while a significant advancement in the field of epigenetics for estimating biological age, has some limitations that can affect its applicability and accuracy in certain contexts. Although it correlates well with chronological age, the Horvath clock does not specifically account for the biological factors that contribute to health and disease, which can vary significantly among individuals¹⁸⁸.

To address these shortcomings, phenotypic clocks like PhenoAge clocks were developed to incorporate health-related biomarkers and clinical data into their algorithms. The PhenoAge clock, developed by Levine et al., is designed to predict biological age based on a combination of chronological age and nine clinical biomarkers, including measures of inflammation and metabolic health²⁰⁸. This approach allows PhenoAge to provide a more nuanced understanding of biological aging that is directly linked to health outcomes, making it particularly useful in clinical

settings^{188,210,212}. For example, findings from Vetter et al. suggested that PhenoAge may be a promising tool for identifying patients who are at greater risk of diabetes complications. Moreover, some studies have noted that the PhenoAge clock can guide clinical interventions. Reduction in predicted age by PhenoAge clocks was significantly associated with physical functional capacity, making it valuable for evaluating the effectiveness of assessing weight management interventions to improve healthy aging^{210,230}.

Many clocks like PhenoAge were designed to provide a measure of biological age that is strongly predictive of various health outcomes, including chronic diseases like cancers and all-cause mortality, contributing to the broader understanding of the interplay between biological aging and various adverse health conditions. Conversely, clocks can be designed to generate risk scores for more specific conditions like cancer; one example of such a clock is EpiTOC. The EpiTOC clock has been used to study the relationship between aging and cancer risk by focusing on the epigenetic changes at 385 CpG sites associated with cellular turnover²¹³. The output of the EpiTOC clock, known as the pcgtAge score, reflects the accumulated DNAm errors as the cell divides over time, which is an indicator of cellular aging and is also significantly involved in cancer development. Notably, blood cancers were not used in the clock's cancer validation analyses; instead, the authors demonstrated cancer-associated acceleration primarily across multiple solid tumour types²¹³. Due to its emphasis on mitotic activity, the EpiTOC clock holds significant relevance in cancer research, with studies utilizing it to identify individuals at high risk of developing cancer^{213,231–233}.

The selection of these four distinct epigenetic clocks for this study is grounded in their robust validation within the literature, demonstrating strong predictive accuracy for various health outcomes. Utilizing these clocks enabled a more nuanced understanding of biological aging and its intricate interactions with age-related conditions like frailty and senescence, thereby enhancing the depth of our analysis. Given that each epigenetic clock possesses inherent limitations, including variations in accuracy across different tissues and populations, employing multiple clocks allows us to mitigate these individual constraints and enrich our findings. Furthermore, the primary technical assumptions required by each clock to ensure the reliability and validity of our results were met by our samples.

Table 1. Key features of select epigenetic clocks: type, training characteristics, and performance metric.

Epigenetic Clock Name	Epigenetic Clock Type	Training samples (tissue types)	Training age range	Reported Correlation (Error)
Hannum ²⁰⁷	Chronological	Whole blood	19-101	0.96 (3.9)
Horvath (Pan-Tissue) ²⁰⁶	Chronological	Multi-tissue, including whole blood	0-100	0.98 (2.7 for whole blood)
Levine (PhenoAge) ²⁰⁸	Biological/Phe notypical	Whole blood	21-100	0.94
EpiTOC ²¹³	Biological/Mito tic	Whole blood	19-101	--

These technical assumptions include the chronological age of our samples fall within the chronological age range of samples used to train the Hannum, Horvath, PhenoAge, and EpiTOC clocks²⁰⁸. Moreover, our samples are peripheral blood mononuclear cells

(PBMCs) isolated from whole blood, representing the leukocyte fraction (T cells, B cells, NK cells, and monocytes) similar to those utilized in training the clocks. This is crucial because epigenetic clocks are developed based on specific DNA methylation patterns that are characteristic of particular tissues. These patterns reflect age-associated changes that occur in the epigenome, and their accuracy in predicting biological age is contingent upon the similarity between the training samples and the samples being analyzed^{206,234}. Furthermore, our samples contain various clinical information that may allow for a more comprehensive assessment of biological age that reflects the individual's health status by PhenoAge. Lastly, our samples were derived from patients with cancer, specifically CLL; thus cancer-related epigenetic changes may be uniquely captured by EpiTOC.

2. Rationale, Hypothesis, and Objectives

2.1 Cell type predictions and senescence

2.1.1 Rationale

While DNAm-based predictors of epigenetic age and cell type proportions are recognized for their accuracy in the literature, the biological factors or conditions that may affect this accuracy remain underexplored. There is speculation that senescence may be linked to age-associated alterations in the methylation landscape that these age predictors rely on. Therefore, it is necessary to determine how senescence impacts the accuracy of DNAm-based measurements.

2.1.2 Research Questions

In the final phase of my project, we aimed to investigate three key questions pertaining to the relationship between cellular senescence and DNAm in the context of age and cell type estimation.

1. First, we sought to determine whether the degree of senescence in a sample is significantly associated with the magnitude of predictive errors in DNAm age.
2. Second, we assessed whether the degree of senescence in a sample is significantly associated with the magnitude of predictive errors in DNAm cell type proportion.
3. Lastly, we examined whether differences in the senescence markers examined affect the relationship between senescence and predictive accuracy of DNAm-based estimators of age and cell types.

Through addressing these questions, we explored how senescence may affect the accuracy of DNAm-based estimators, which could shed light on the biological mechanisms that regulate precise ticking of epigenetic clocks.

2.1.3 Hypothesis

We expected to observe significant positive correlation between specific senescent markers, such as positive expression of CD57 and negative expression of CD28, and age acceleration. Furthermore, I expect to find a significant positive correlation between these senescent markers and discrepancies in cell type proportions.

2.2 Epigenetic Age and Epigenetic Clock

2.2.1 Rationale

The CIRS offers a comprehensive framework for assessing the severity and complexity of illnesses across multiple bodily systems. This tool has been extensively employed in the treatment planning of patients with CLL due to its strong predictive capacity regarding patient survival following therapeutic interventions. Nevertheless, the biological basis of the CIRS remains to be fully elucidated. Among the various biomarkers currently under investigation as potential predictors of CIRS scores, epigenetic age, specifically DNAm age, has garnered particular interest. Consequently, it is important to investigate the relationship between CIRS scores and epigenetic age to assess the validity of epigenetic age as a biomarker for CIRS.

2.2.2 Research Questions

In this part of the project, we addressed two crucial questions that explore the interplay between CLL and DNAm age.

1. First, we determined whether epigenetic age is more closely associated with the CIRS score than chronological age. This inquiry is crucial as it will provide valuable insights into the biological mechanisms that underpin the CIRS score.
2. Secondly, we investigated the extent to which cell types, epigenetic clocks, and various CLL risk factors explain the variations in DNAm age in CLL patients. This question will guide us in identifying other associations with DNAm age of our CLL patients.

2.2.3 Hypothesis

We hypothesized that CLL patients with a high comorbidity burden, defined as a CIRS score exceeding six, will demonstrate greater age acceleration compared to those with a low comorbidity burden, characterized by a CIRS score below six. Furthermore, CLL patients presenting with poor prognostic markers, such as unmutated IGHV genes, a history of smoking, positive expression of ZAP-70, and low expression of CD28, are also expected to exhibit heightened age acceleration. I also expect that there will be differences in epigenetic age acceleration between B and non-B cells.

2.3 Epigenome-wide association study

2.3.1 Rationale

EWASes have been instrumental in identifying promising epigenetic biomarkers for various diseases, serving clinical purposes such as disease risk assessment, prognosis and treatment outcomes. To date, no CpGs have been identified in the literature that can distinguish between high and low comorbidity burden in CLL patients. Therefore, an epigenome-wide association analysis was conducted on the CIRS comorbidity burden.

2.3.2 Research Question

In the second phase of my project, we focused on a central research question: Are there differentially methylated sites that are significantly associated with the CIRS score? This inquiry is designed to guide our exploration in identifying novel methylation patterns that may serve as robust predictors of frailty measurement scores. Such findings could ultimately enhance the optimization of treatment strategies for patients with CLL.

2.3.3 Hypothesis

We hypothesized that performing EWAS on our samples will identify differentially methylated positions or CpGs that are significantly associated with CIRS comorbidity burden, which can prompt further investigation into its potential as a proxy biomarker of frailty among CLL patients.

3. Methods

3.1. Sample Selection and Preparation

Peripheral blood mononuclear cells from 64 CLL patients aged 52-98 were obtained from the Manitoba Tumour Bank, which houses a collection of tumour samples, initially collecting breast cancer cells but has since expanded to leukemic samples voluntarily taken from patients diagnosed with cancer in the province. The samples were selected from a list of 68 patients. Four non-B samples with the lowest purity (13%-46% non-B cells) were removed. All measurements of the dead and B cells were performed using flow cytometry. Quality of the non-B samples were prioritized due to the project's initial focus on the non-cancerous non-B cells of CLL.

All samples used in this project were taken during the period between diagnosis and first treatment, so none of the samples were exposed to treatment. Samples were transferred to the Marshall Lab to isolate B cells via negative selection using the EasySep Human B cell Enrichment Kit II (STEMCELL Technologies, Vancouver, Canada) according to manufacturer's directions. Marshall Lab former members, Xun Wu and Jane Maksimova primarily performed the separation of B and non-B cells from the CLL PBMCs. I assisted in the separation of the first few samples. With negative selection, non-B cells adhere to the antibodies attached to the magnetic beads, which were then magnetized to pin the tagged non-B cells to the tube walls. The untouched B cells remained in suspension and were poured off into the new tube. After centrifuging and draining the supernatant, the B cells in the precipitate were ready for downstream analysis. The non-B cells that were attached to the tube wall held by the magnetic

beads are resuspended and collected before centrifuging then snap freezing them into pellets for DNA analysis.

3.2. CIRS Cut-off for Comorbidity Group Classification

The cut-off used to define comorbidity burden in this study was based on the Cumulative Illness Rating Scale (CIRS), with scores ≥ 6 classified as high comorbidity and scores < 6 classified as low comorbidity. This threshold was selected because it has been used in the literature to distinguish patients with greater comorbidity burden from those considered more clinically fit, particularly in the context of treatment stratification in CLL²⁶³. The same cut-off was applied in this study to define case and control groups on the basis of clinically meaningful differences in overall health status. This approach was also consistent with the judgment of our CancerCare clinical collaborators, who considered a CIRS score of 6 to be an appropriate threshold for identifying patients with higher comorbidity burden and reduced fitness for standard chemotherapy in this cohort.

3.3. DNA Extraction, Bisulfite Conversion and PCR

DNA extraction from the B and non-B cell pellets were performed using the DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Germany). To discriminate methylated from unmethylated CpGs, bisulfite conversion was performed with the EZ DNA Methylation-Gold™ (ZymoResearch, Irvine, CA, USA) Kit protocol. In this step, the unmethylated cytosine residues are converted to uracil, then later, to thymine after DNA amplification with polymerase chain reaction (PCR). The methylated cytosine residues

remain unaltered. Microarray assays require approximately 750 ng of DNA for optimum performance, thus samples with low DNA yield were evaporated using a vacuum concentrator. DNA concentration and quality of the samples were assessed with the NanoDrop Spectrophotometer. Standard DNA methylation quality control procedures, described below, were applied to all samples, including those that underwent concentration. The Omixer package in R was used to randomize the position of each sample during its loading onto the microarray chips. Randomization was performed to reduce batch effects, which is a phenomenon of introducing non-biological variations into microarray data stemming from differences in sample processing and measurement and thus, may hinder the detection of true biological signals.

3.4. Measuring DNAm levels in the PBMCs

In this project, DNAm was measured using the Illumina EPIC array, which is capable of profiling DNAm modifications in over 850 000 DNA methylation sites. Intensity Data (IDAT) files are the raw Illumina methylation data that store that intensity data captured by each probe. The array measures the methylated and unmethylated intensities of each CG site, both of which are used to calculate the proportion of methylation at each CG site. There are two ways to quantify methylation levels are individual CpG sites: B-values and M-values

$B = M / (M + U)$ [Ratio of methylated probe intensity and the overall probe intensity]

Beta values fall between zero and one or 0% and 100%. Ideally, beta value of zero indicates that all probes of a given CG site are unmethylated. Conversely, beta values of one or 100% indicates that all probes for that CG site are methylated. The bounded range of beta values allows direct interpretation of the results since it approximates that proportion of methylation of a CG site. However, this very statistical feature makes beta values unsuitable for further methylation analysis and regression modelling.

$B = \log_2(M / U)$ [the log₂ ratio of the intensities of methylated probe and unmethylated probe]

M values closer to zero are indicative of an equal ratio of methylated and unmethylated probes. Higher amounts of methylated molecules relative to unmethylated probes are indicated by more positive M-values whereas the inverse is indicated by more negative M-values. Both M-values and B-values are metrics used to quantify DNA methylation levels and are frequently reported together. However, the appropriateness of each method depends on the specific context. M-values exhibit homoscedasticity, where the residual variations of regression models are consistent across all values of the independent variables. This feature is generally regarded as important when performing statistical analyses in order to produce reliable and valid results. On the other hand, B-values are helpful for interpretation since B-values directly reflect the proportion of methylation at a specific CpG. This makes B-values easier to interpret biologically.

3.5. Preprocessing, Normalization, and Quality Control

Raw IDAT files were imported and processed using the *minfi* package in R. Sample-level quality control was performed by assessing detection p-value distributions, signal intensities, and overall array performance to identify poorly performing or outlier samples. Exclusion decisions were based on the combined assessment of these quality control metrics, and no samples showed consistent evidence of poor performance across measures; therefore, no samples were excluded after DNA methylation data generation. Initial quality control steps included the removal of CpG probes with detection p-values >0.01 in more than 1% of the samples and as well as those with a bead count (the number of beads that successfully report a signal for a given CpG probe in a given sample with low counts indicating unreliable measurement) of less than three. Non-CpG probes, probes located on the X and Y chromosomes, and probes with known cross-reactivity or probes with common single nucleotide polymorphisms were filtered out. Following functional normalization to correct for technical variation like batch effects, beta values were further normalized using the *wateRmelon* R package's beta-mixed quantile normalization (BMIQ) function. BMIQ corrects for the probe type bias inherent in the Illumina Infinium platforms, which uses two types of probes that were designed differently by aligning the distribution of beta values across probe types.

3.6. Epigenetic Clock Output and Epigenetic Age Acceleration

Outputs from Horvath, Hannum, PhenoAge, and EpiTOC clocks are available through the Clock Foundation (<https://dnamage.clockfoundation.org/>). This online clock only requires the beta values from a list of specific CG probes (csv file of the necessary

probes is provided by Horvath and available in the Clock Foundation website) and the sample annotation file specifying the chronological age, sex, and tissue of a sample. Therefore, no probe filtering or probe-type normalization was done performed prior to the upload of subset of DNAm data from the CLL samples. The following estimates were obtained from the clock output: Epigenetic age from Hannum (DNAmAgeHannum), Horvath (DNAmAge), and PhenoAge (DNAmPhenoAge), epigenetic age acceleration from Hannum, Horvath, and PhenoAge (AgeAccelPheno), and pcgtAge from EpiTOC. For each clock, epigenetic age acceleration was calculated as the residual from a linear regression of DNAm age on chronological age. Residuals greater than zero indicate higher DNAm age than expected for a given chronological age, whereas residuals less than zero indicate lower DNAm age than expected.

3.7. Cell Type Estimations

To obtain DNAm-predicted cell type estimations, the *Flow.Sorted.Blood.EPIC* reference dataset along with optimal IDOL probes from the raw IDAT data was used to deconvolute CD8T, CD4T, NK, Monocytes, Neutrophils, and B cell proportions.

Due to the inherent challenges in working with compositional data such as cell type proportions that sum to 1 or 100%, we applied a combined isometric log ratio transformation, followed by robust Principal Component Analysis (PCA) using the *robCompositions* package (v.2.3.1), to identify independent components from the inter-correlations in our cell composition data. The first four components, which accounted for greater than 90% of the variance in our cell-type mixtures, were then used for

subsequent EWAS. The composition of each cell type was independently assessed in a linear regression with age and comorbidity burden to evaluate which had a larger effect.

For the senescence analysis of T cells in CLL patients, flow-cytometry measured proportions of CD4⁺T and CD8⁺T cells along with senescent markers, CD28⁺ and 57⁺, and were obtained. Following isolation of PBMCs, the cells were stained with anti-CD45 antibody to gate on leukocytes. Data acquisition was done on a Beckman Coulter Cytoflex flow cytometer (Beckman Coulter, Brea, USA), and populations were gated using FlowJo software v10 (BD Biosciences, San Jose, USA). The proportion of senescent T cells were identified as CD4⁺T and CD8⁺T subsets expressing the markers CD57⁺ or CD28⁺. In this case, CD4⁺CD57⁺ cell are the sum of CD4⁺CD57⁺PD1⁺ and CD4⁺CD57⁺PCD1⁻, whereas CD8⁺CD57⁺ cells are the sum of CD8⁺CD57⁺PD1⁺ and CD8⁺CD57⁺PCD1⁻. CD4⁺ T cells with the senescent marker, CD28⁺ have the sum of CD4⁺CD27⁺CD28⁺ and CD4⁺CD27⁻CD28⁺ while CD8⁺ T cells with the senescent marker, CD28⁺ have the sum of CD8⁺CD27⁺CD28⁺ and CD8⁺CD27⁻CD28⁺. Percentages were calculated relative to the total live CD4⁺T cells or CD8⁺T cells.

3.8. EWAS

Differential DNA methylation between high and low comorbidity burden in the non-B cell fraction of CLL samples was assessed using epigenome-wide association analysis (EWAS) with the *limma* R package. Linear models were fitted for each CpG site using ComBat-corrected M-values as the outcome. To account for unmeasured sources of technical and biological variation, surrogate variable analysis (SVA) was performed, and the estimated surrogate variables were included in the models. Models

were additionally adjusted for chronological age, sex, smoking status, and cell composition, the latter represented by the first four principal components derived from cell type estimates, which together explained more than 90% of the variance.

To reduce bias and inflation in the EWAS test statistics, bacon correction was applied to the model output, and corrected p-values were used for downstream interpretation. Because CpG sites were tested on an epigenome-wide scale, p-values were further adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure. CpG sites with FDR-adjusted p-values < 0.05 were considered statistically significant. For visualization and descriptive interpretation, volcano plots were generated using nominal significance thresholds of $p < 1 \times 10^{-5}$ and $p < 1 \times 10^{-6}$, corresponding to $-\log_{10}(p) \geq 5$ and $-\log_{10}(p) \geq 6$, respectively. CpG sites were highlighted in the volcano plot if they met $p < 1 \times 10^{-5}$ and an absolute methylation difference of $|\Delta\beta| \geq 0.01$, where $\Delta\beta$ represents the difference in mean beta values between the high and low comorbidity groups. CpG sites exceeding $p < 1 \times 10^{-6}$ were additionally annotated on the plot.

3.9. Statistical Analyses

Analyses were performed using R statistical software version 4.4.1. Plots were created using the ggplot package. Pearson correlation coefficient was used to assess the strength and direction of the association between DNAm methylation age and chronological age, and cancer risk score (pcgtAge) and chronological age. Pearson correlation was also used to examine the relationship between epigenetic age acceleration, and proportions of senescent T cells and naive T cells as well as the

relationship between the magnitude of epigenetic clock errors and senescent T cell proportions. Epigenetic clock errors were defined by taking the absolute values of the residuals or the differences between the epigenetic age and chronological age of each CLL sample. Differences in age acceleration between adverse prognostic variables and less adverse prognostic variables in non-B and B cell were assessed with a paired sample T-test, with significance defined at $p < 0.05$.

4. Results

4.1 Study population characteristics

The study included 64 patients diagnosed with CLL, with a median age of 77 years (range, 52–98), with a mean age of 75.6 years (SD 9.6), and 33 patients (51.6%) were female. At the time of sampling, all patients had not undergone treatment.

Participants were categorized into two groups based on their comorbidity burden, as assessed by the CIRS. Thirty-five patients had a CIRS score ≥ 6 and were classified as having a high comorbidity burden, while 29 patients had a CIRS score < 6 and were classified as having a low comorbidity burden. The distribution was not balanced due to prioritization of sample quality, which resulted in a greater number of patients in the high comorbidity group.

Patients in the high comorbidity group were older than those in the low comorbidity group, with a median age of 79 years (range, 52–98; mean 77.8, SD 8.9) versus 74 years (range, 53–97; mean 73.0, SD 9.9), respectively, and this difference was statistically significant ($p = 0.023$). No statistically significant differences were observed between the high and low comorbidity groups for sex ($p = 0.625$), smoking history ($p = 0.917$), IGHV status ($p = 0.779$), p53 status ($p = 0.933$), ZAP-70 status ($p = 0.530$), or CD38 expression ($p = 0.945$). In terms of smoking history, approximately 48% of the cohort had never smoked, while 47% were either current or former smokers. IGHV mutation status was available for most participants, with 45% harbouring mutated IGHV and 42% unmutated; the remainder had missing data. The distribution of p53 mutation status was skewed, with only 6% confirmed to have a mutated p53, 55% unmutated, and the rest unknown. ZAP-70 expression was negative in 45% of the

cohort. CD38 expression was below 20% in 52% of patients, while 30% exhibited CD38 positivity ($\geq 20\%$).

Additional clinical and demographic variables, including del(17p) mutation status, race, and ethnicity, were excluded from analysis due to extensive missingness. These limitations, along with the need to prioritize high-quality DNA samples, restricted our ability to conduct further subgroup matching. Table 2 summarizes the full clinical and demographic characteristics of the cohort.

Table 2. Clinical and demographic characteristics of the CLL patient cohort.

Comorbidity Burden				
	High ^a (N=35)	Low ^b (N=29)	Total (N=64)	p-value ^e
CIRS ^c score				
1	0 (0%)	5 (17.2%)	5 (7.8%)	
2	0 (0%)	3 (10.3%)	3 (4.7%)	
3	0 (0%)	4 (13.8%)	4 (6.3%)	
4	0 (0%)	11 (37.9%)	11 (17.2%)	
5	0 (0%)	6 (20.7%)	6 (9.4%)	
6	11 (31.4%)	0 (0%)	11 (17.2%)	
7	4 (11.4%)	0 (0%)	4 (6.3%)	
8	7 (20.0%)	0 (0%)	7 (10.9%)	
9	7 (20.0%)	0 (0%)	7 (10.9%)	
10	2 (5.7%)	0 (0%)	2 (3.1%)	
11	3 (8.6%)	0 (0%)	3 (4.7%)	
12	1 (2.9%)	0 (0%)	1 (1.6%)	
Sex				0.625
Female	17 (48.6%)	16 (55.2%)	33 (51.6%)	
Male	18 (51.4%)	13 (44.8%)	31 (48.4%)	
Chronological Age (years)				0.023
Mean (SD)	77.8 (8.9)	73.0 (9.9)	75.6 (9.6)	
Median [Min, Max]	79.0 [52.0, 98.0]	74.0 [53.0, 97.0]	77.0 [52.0, 98.0]	
Smoking History				0.917
No	16 (45.7%)	15 (51.7%)	31 (48.4%)	
Unknown	2 (5.7%)	1 (3.4%)	3 (4.7%)	

Yes	17 (48.6%)	13 (44.8%)	30 (46.9%)	0.779
IgHV^d Status				
Mutated	15 (42.9%)	14 (48.3%)	29 (45.3%)	
Unknown	4 (11.4%)	4 (13.8%)	8 (12.5%)	
Unmutated	16 (45.7%)	11 (37.9%)	27 (42.2%)	
p53 Status				0.933
Negative	20 (57.1%)	15 (51.7%)	35 (54.7%)	
Positive	3 (8.6%)	3 (10.3%)	6 (9.4%)	
Unknown	12 (34.3%)	11 (37.9%)	23 (35.9%)	
Zap 70 Status				0.530
Negative	14 (40.0%)	15 (51.7%)	29 (45.3%)	
Positive	13 (37.1%)	7 (24.1%)	20 (31.3%)	
Unknown	8 (22.9%)	7 (24.1%)	15 (23.4%)	
CD38+ Expression				0.945
<20%	18 (51.4%)	15 (51.7%)	33 (51.6%)	
>= 20%	11 (31.4%)	8 (27.6%)	19 (29.7%)	
Unknown	6 (17.1%)	6 (20.7%)	12 (18.8%)	

^{a-b} High comorbidity burden indicates CIRS score of 6 or higher. Low comorbidity burden indicates a CIRS score of below 6.

^cCIRS means Cumulative Illness Rating Scale.

^dIgHV means immunoglobulin heavy chain variable region.

^eFisher's exact test; Wilcoxon rank sum test (for Chronological Age).

4.2 Cell type predictions

4.2.1 Prediction of cell types in B and non-B cells

Different blood cell types have distinct DNAm profile. Heterogenous samples like the whole blood extracted from CLL patients were expected to be a composite of distinct DNAm patterns, which would be problematic for EWAS analysis. If cell type proportions are not accounted for, it could confound results and make it difficult to pinpoint which cell type(s) are strongly associated with observed variable of interest. Furthermore, variation in cell-type composition can obscure relevant signals or produce associations driven by cell-type differences rather than by the variable of interest itself. Since we do not have complete measured proportions of cell types in our CLL whole been separated into B and non-B cells using their respective DNAm data.

Following the estimation of the cell type composition for each sample, two observations suggested that the cell separation process was successful. First, the predicted proportion of B cells was more than 50% greater in the B cell fraction than in the non-B cell fractions. Second, cell type estimations of the non-B samples were composed predominantly of non-B immune population including T cells, natural killer (NK) cells and monocytes. Small predicted neutrophil proportions were also detected by the deconvolution algorithm in some samples. Although neutrophils are not typically expected in PBMC-derived fractions, low levels may reflect minor contamination during sample processing or limitations of DNA methylation-based deconvolution rather than true substantial neutrophil content (Figure 3).

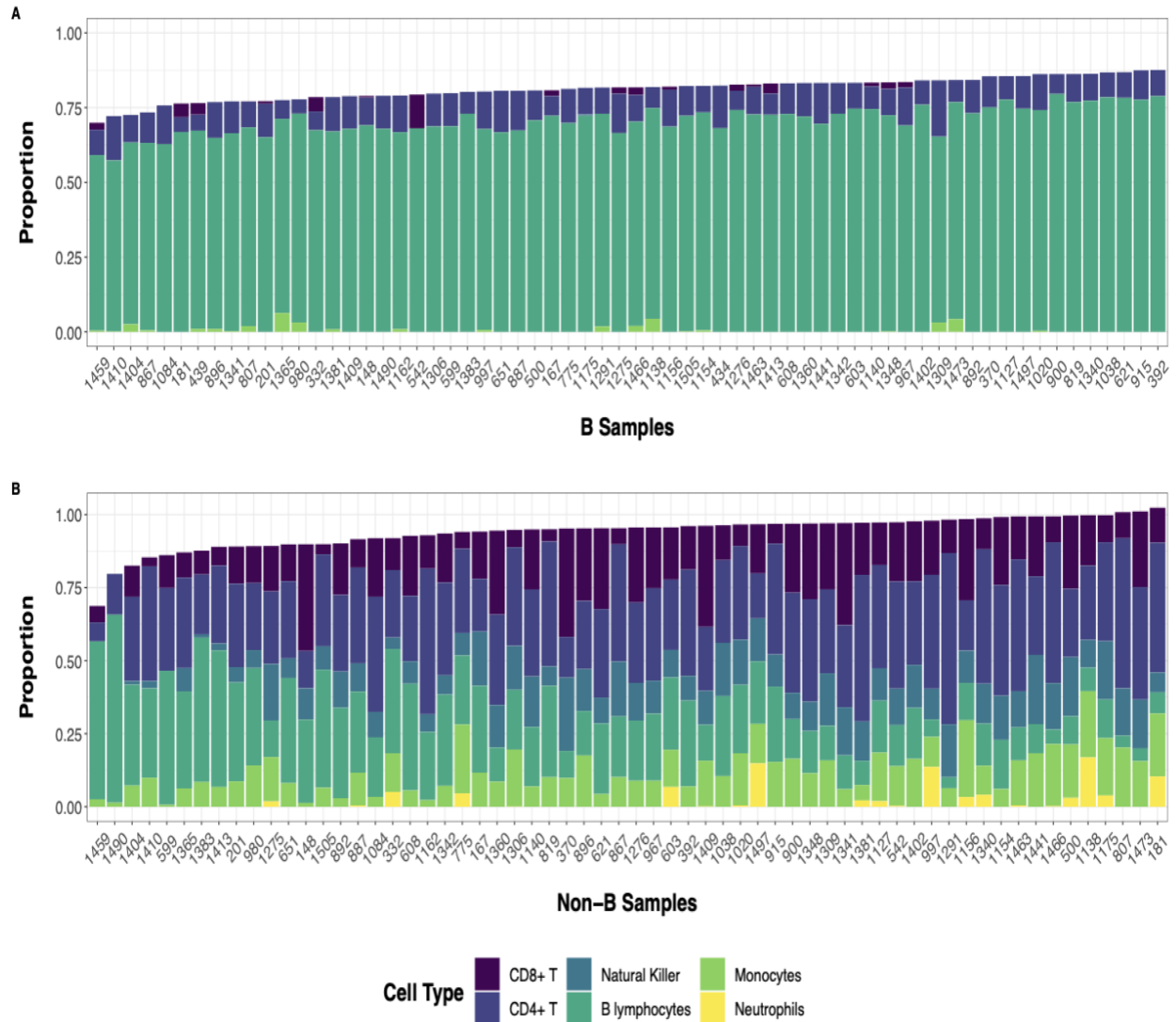


Figure 3. Comparison of DNA methylation–predicted and flow cytometry–measured CD4⁺ and CD8⁺ T-cell proportions. Cell type proportions in B and non-B samples estimated using DNA methylation profiles. More than 50% of estimated proportions within the B fractions were identified to be B cells. However, non-B samples still contained a proportion of B lymphocytes, ranging from 3.8% to 64.4%, despite undergoing magnetic separation.

These observations are in line with what we expected. However, there were findings that suggest our physical separation of the B and non-B cells was not perfect. For one, our B samples contained small proportions of non-B cells (CD8⁺T, CD4⁺T, and monocytes). Secondly, the non-B samples still contained a considerable proportion of B

cells, and in some cases (e.g. sample 1459 and 1490) they comprised more than 50% of the estimated cell proportion. This is verified by purity estimates of non-B cells courtesy of flow cytometry.

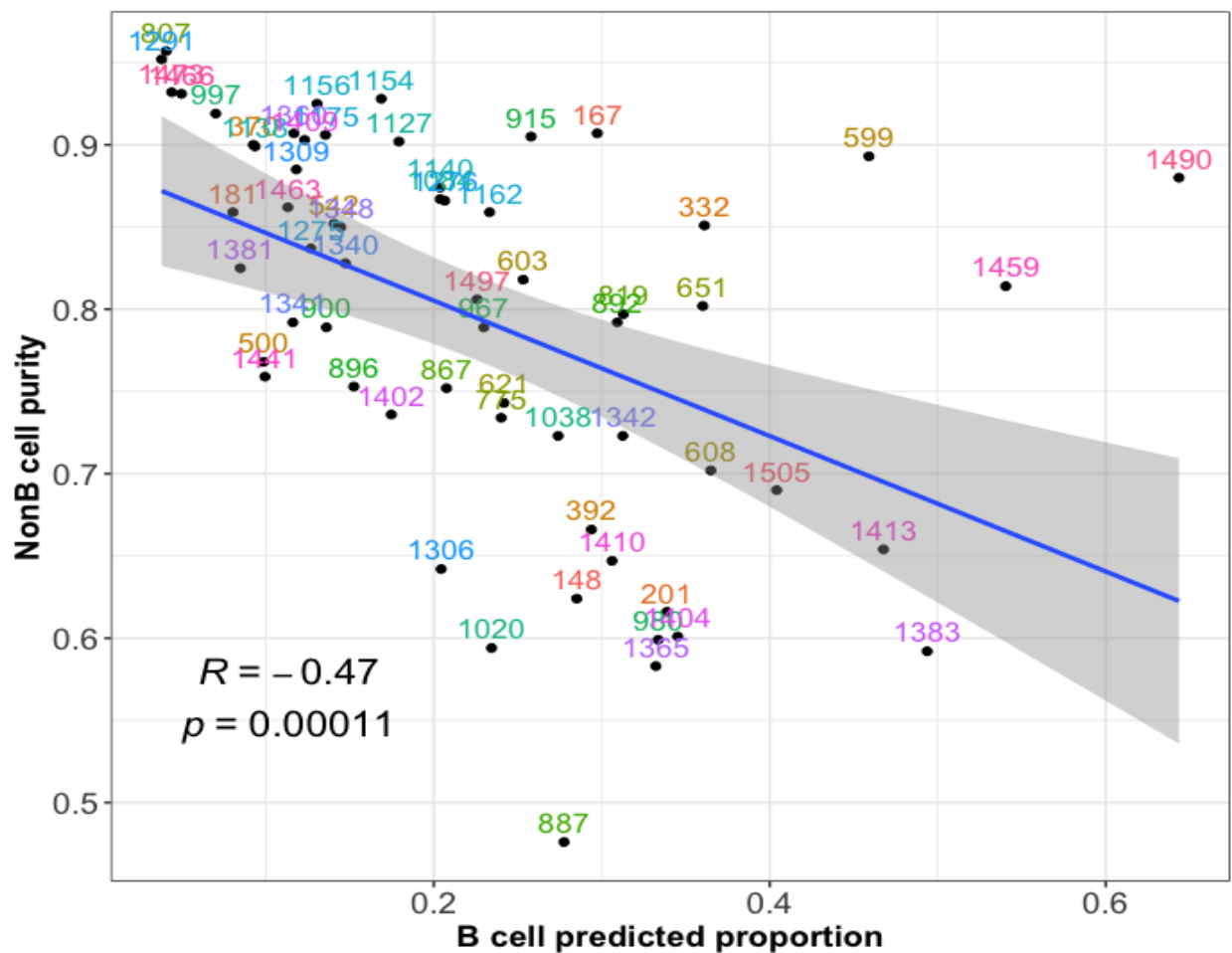


Figure 4. Inverse relationship between DNAm-estimated B cell proportion and flow cytometry–derived non-B cell purity. Each point represents an individual sample. The solid blue line represents the fitted linear regression, with the shaded region indicating the 95% confidence interval. Each point represents an individual sample. The solid blue line represents the fitted linear regression, with the shaded region indicating the 95% confidence interval.

Flow cytometry–based purity estimates for the non-B fractions ranged from 58% to 95%. To quantitatively assess concordance between physical cell separation and DNAm-based deconvolution, we examined the relationship between flow cytometry–derived non-B cell purity and DNAm-estimated B cell proportions within the non-B fraction (Figure 4). As expected, higher estimated B cell proportions were associated with lower non-B cell purity, demonstrating a moderate negative correlation ($R = -0.47$, $p = 0.00011$). These findings indicate that DNAm-based deconvolution captures relative differences in sample purity and suggest that residual B cell signal in the non-B fraction is largely attributable to incomplete physical separation rather than solely limitations of the deconvolution method.

These indicators suggest that the magnetic separation methods of PBMCs did not yield completely purified B cells in our B cell groups and purified non-B cells in our non-B cell groups. One possible contributing factor is that cell clumping during magnetic separation may have led to non-specific binding, which affects the purity of the isolated cell populations. Another observation is that several of the estimated cell proportions do not add up to one (or 100%). In theory, computational techniques that estimate the proportion of different cell types in a mixed sample known as deconvolution methods, would use an exhaustive reference of all cell types present in the bulk samples and constrain the estimated proportions to add up to 100%^{262,263}. Although our analysis focused on only six leukocyte subtypes, the absence of reference profiles for additional cell type, such as dendritic cells and macrophages, not included in the cell-type prediction algorithm may partially account for the unassigned DNA methylation signals. However, a more plausible explanation for the observed sample impurities is the

substantial predominance of CLL B cells within the PBMCs. The excessive abundance of B cells likely hindered effective magnetic separation, resulting in contamination of the non-B cell fraction and obscuring their DNAm signatures.

Nonetheless, one sample that drew my attention is sample 1459. As seen in Figure 1, Sample 1459 had cell proportions that only add up to less than 75% for both B and non-B cells. Although my presumptions about the missing cell type references applies to sample 1459, a considerable portion of the DNAm signal is unassigned led me to look into quality control checks, we investigated the percentage of live and dead cells from flow cytometry (i.e. cell viability) in sample 1459 compared to the other samples since high ratio of dead cells to live cells increases the chance of non-specific bindings and contaminate the sorted cell population. We found that sample 1459 has a live cell percentage of 81.40% and dead cell percentage of 1.05%, indicating a high cell viability. In comparison, one of the samples with relatively low unassigned DNAm signals, sample 392 has a live and dead cell percentage of 66.60% and 5.10%, respectively. In addition, collaborators in the Marshall laboratory independently identified sample 1459 as atypical in their immunophenotypic analyses, reporting the lowest T-cell counts in the cohort by far and a markedly skewed CD4 T-cell phenotype.

In addition, we assessed the quality controls during DNAm preprocessing, such as sex prediction, which is one of the primary quality controls during DNAm age estimation and visualized quality control plots during the preprocessing of raw DNAm data with *minfi*. Sex prediction based on DNAm of sample 1459 aligns with the reported sex of the patient from whom the sample was extracted. Quality control plots also did not reveal any profound differences between sample 1459 with the rest of the samples.

Since the sample did not fail our quality control checks, we did not discard sample 1459 however, the sample was subjected to further scrutiny in subsequent analysis.

The finding that cell type proportions differ between B and non-B samples is crucial for adjusting cell composition in EWAS. Without such adjustments, observed differences in DNAm could be misinterpreted as within-cell type epigenetic alterations or disease-associated methylation changes, when they may actually reflect discrepancies in cell type composition. We have flow cytometry counts for B, CD4+T, and CD8+T cells, but lack direct measurements for other non-B cell subtypes, such as monocytes, neutrophils, and NK cells. In the absence of directly measured proportions for these non-B cell subtypes, the cell composition predicted by our tools will be integrated into the interpretation of the results, with the predicted cell proportions being used in the subsequent EWAS analysis. Nonetheless, some of the predictions yielded odd findings like proportions not summing up to one. Moreover, DNAm patterns can vary between populations due to disease-related factors. If a reference dataset is based on normal cell population as is the case for many deconvolution tools like Horvath and IDOL-Salas, the accuracy of cell estimations for cancerous samples (e.g. CLL cells) may be impacted. As such, it is important to evaluate the accuracy of the tools used for estimating the B and non-B cell proportions.

4.2.2 Predictive accuracy of cell-type composition by deconvolution tools

Accurate predictions of cell type composition help control for confounding variables, leading to more reliable associations between DNAm and traits of interest. Although the reliability of DNA methylation-based predictors of cell type proportions has

been well established in the literature, this may not necessarily apply to our samples, which are comprised of cancerous cells. Significant limitations arise when deconvolution tools are trained exclusively on normal cells, potentially impairing its cell type predictive capabilities in pathological contexts. This limitation is particularly pertinent given the complex nature of tumour microenvironments, which often exhibit altered cellular compositions and gene expression profiles compared to normal tissues.

Here, we compared the accuracy of cell type proportion estimates derived from non-B cell DNAm data to their measured counterparts obtained by flow cytometry. Specifically, we evaluated estimates generated by two reference-based approaches: the Horvath clock and the IDOL algorithm using the Salas reference. The online DNAm age calculator by Horvath incorporates a supplementary function that estimates cell composition by using Houseman's method of cell deconvolution as described earlier under Methods. The Houseman method uses leukocyte subtypes, including B cells, monocytes, granulocytes, natural killer cells, CD4+T and CD8+T cells, that were profiled from the 27K array as reference.

Meanwhile, the IDOL algorithm utilizes the Houseman approach as framework for cell deconvolution, but feature selection techniques have been applied, which enhances the prediction of cell type proportions, especially in complex or heterogeneous samples. Additionally, IDOL uses an expanded reference library of 12 leukocyte subtypes, which includes five cell types that Houseman uses. Improving upon the earlier Housman's reference dataset, the Salas dataset has higher cell profile coverage (i.e. profiled 12 leukocytes as opposed to six) and offers more accurate cell-type resolution (i.e. it can discern granulocytes into eosinophils and basophils).

Considering that more optimization processes are integrated into the IDOL-Salas algorithm compared to the Houseman method in Horvath's online calculator in addition to more leukocyte subtype DNAm profiles, we expected to see more accurate cell type estimations with the IDOL and Salas's method.

To test this, we compared the cell type proportions of CD4+T and CD8+T cells predicted by the IDOL algorithm using Salas' reference library and the Houseman-based cell type estimator featured in Horvath's clock to the flow-cytometry measured CD4+T and CD8+T cell proportions as the reference.

Our analysis shows a moderate correlation between the IDOL-Salas predicted proportions of CD4+T and CD8+T cells and their flow cytometry-measured counterparts ($R = 0.61$, $p < 0.001$ for both) (Figure 5). Similarly, the Horvath-predicted proportions exhibited a significantly moderate correlation with flow cytometry measurements ($R = 0.70$, $p < 0.001$ for CD4+T; $R = 0.71$, $p < 0.001$ for CD8+T). These moderate yet significant associations suggest that both deconvolution tools accurately estimate CD4+T and CD8+T proportions in CLL non-B samples.

However, contrary to our hypothesis, IDOL-Salas's methods did not outperform Horvath as the latter showed a stronger correlation with measured values. This indicates that Houseman's predictions align more closely with flow cytometry results. Nevertheless, the root mean square error (RMSE) values, which calculates the help magnitude of errors between predicted and measured cell proportions, were identical for both methods (RMSE = 0.27) across T cell types. This suggests that neither method was superior to the other. Our measurements, including R, p-value, RMSE, and trend, suggest that both deconvolution tools performed well, although not as optimally as

those reported in the literature, which typically show correlations of 0.84 to 0.9 and RMSE values of 0.07 to 0.19. However, it is important to note that these studies often used normal or non-diseased whole blood samples, unlike our CLL samples. Given that pathophysiological changes can influence cells and their methylation patterns, this could explain the differences in predictive performance between our regression model and those reported in the literature.

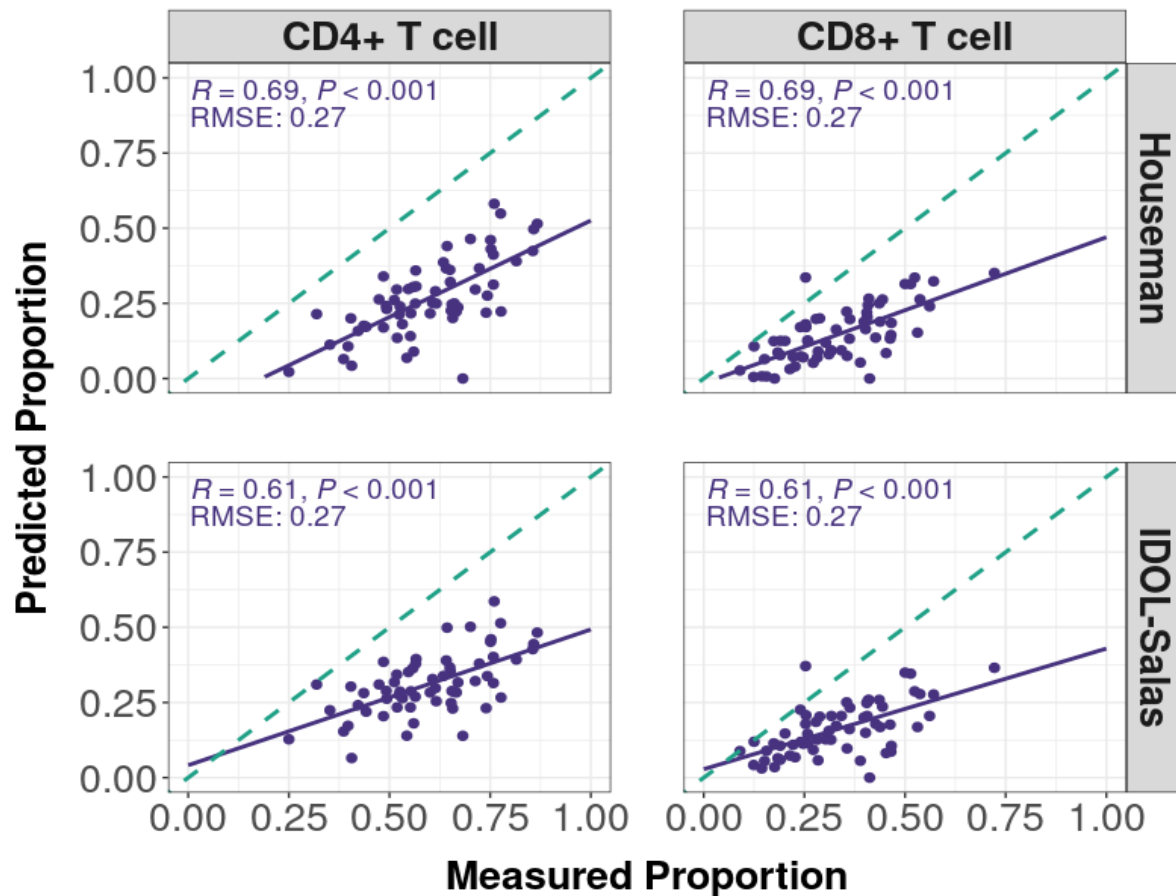


Figure 5. Comparison of DNA methylation-predicted and flow cytometry-measured CD4+ and CD8+ T-cell proportions. DNAm-based predictors of cell type composition underestimate the proportion of CD4+T and CD8+T cells in CLL cells. The accuracy of the predictors appears to differ between the T cell type and the proportion values. Purple line indicates the regression line after regressing the DNAm-predicted proportions of the T cells onto its flow-cytometry measured counterpart. Dotted green line indicates the optimal regression line if there is a one-to-one relationship between the predicted and measured proportions. RMSE means root mean square error. Significance is measured at $p < 0.05$.

5

One interesting observation is that our deconvolution tools seem to underestimate the CD4⁺T and CD8⁺T proportions, as indicated by the predicted proportions being smaller than the measured proportions in Figure 6. There is no difference in the degree of underestimation between the Houseman and IDOL-Salas'

predictor tools when looking at either CD4⁺T cells and CD8⁺T cells. However, there were notable discrepancies between the predicted and measured cell proportions for CD4⁺T cells and CD8⁺ T cells. These differences may reflect limitations in methylation-based deconvolution for estimating specific T-cell subsets but could also arise from technical factors affecting the flow cytometry measurements, including the issue of bead adherence to cells during separation. The average difference between the Houseman-predicted and measured proportions of CD4⁺T is CD8⁺T cells are 0.33 and 0.19, respectively. Meanwhile, the average difference between the IDOL-Salas-predicted and measured proportions of CD4⁺T and CD8⁺T cells are 0.29 and 0.17, respectively. Under ideal conditions, differences between flow cytometry and deconvolution estimates would primarily reflect deconvolution prediction error. However, in this dataset, bead-based cell separation likely introduced additional compositional variability and may have affected the accuracy of flow cytometry measurements. As such, discrepancies between measured and predicted proportions likely reflect a combination of deconvolution error and measurement uncertainty and should be interpreted with caution.

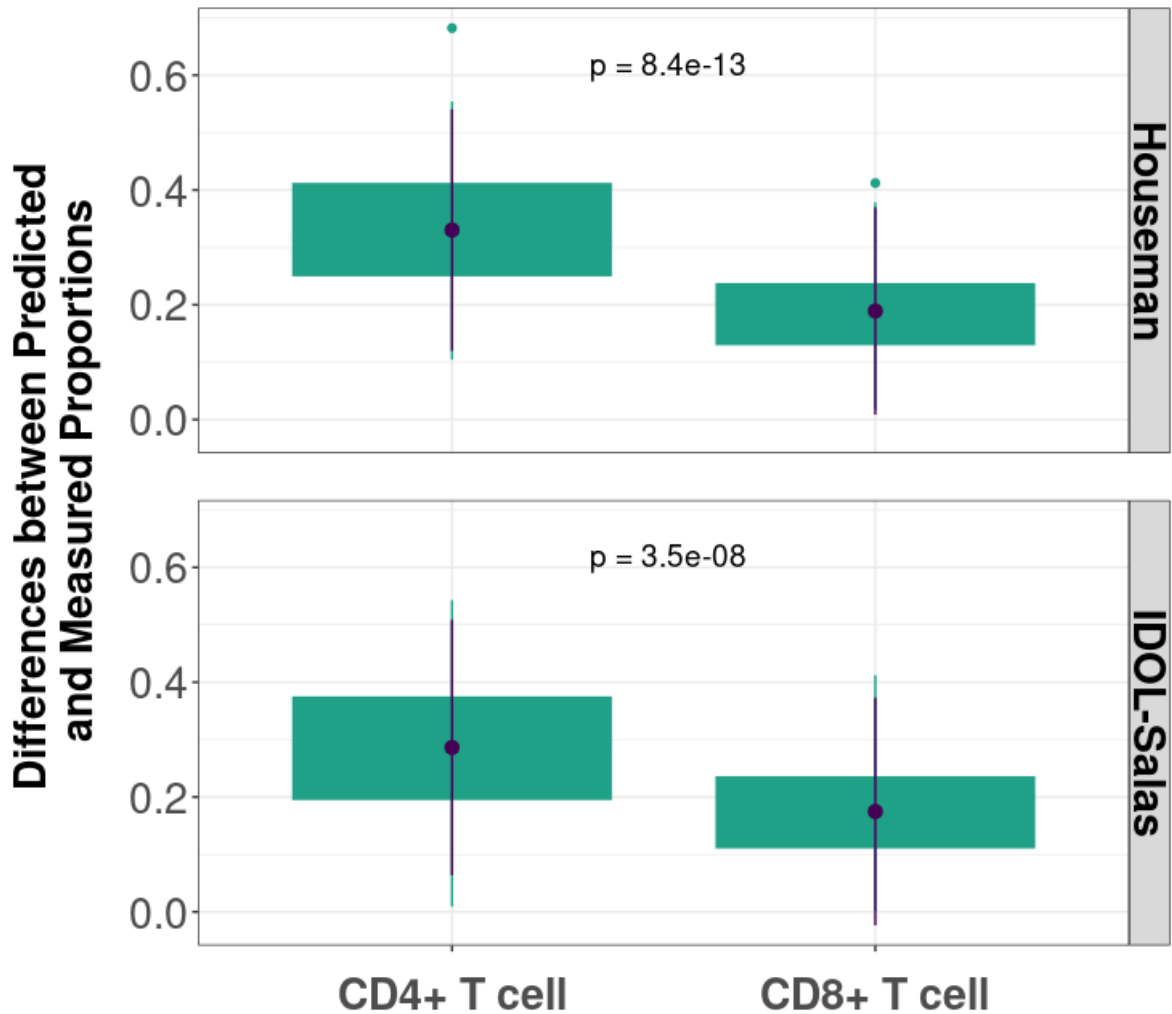


Figure 6. Compared to CD8+T cells, CD4+T cells exhibited greater difference between flow-cytometry measured and DNAm-based predictors of cell proportions. In addition, the differences between predicted and measured cell proportions on average were smaller in CD8+T cells compared to CD4+T cells, implying the accuracy of DNAm based cell type predictors varies by cell type. Significance is measured at $p < 0.05$.

Our findings indicate that predictive errors are significantly higher when estimating CD4+ T cell proportions compared to CD8+ T cells, that is, CD8+ T cell estimates more closely matched their measured counterparts regardless of the deconvolution method used (Houseman or IDOL-Salas). The range of predictive errors for CD8+ T cells was 0.0016–0.41, while for CD4+ T cells, it ranged from 0.10–0.68.

This is rather surprising, considering we expected that IDOL-Salas method would yield less predictive errors. Overall, the differences in predictive errors between the CD4⁺T and CD8⁺T cells seem to imply that biological factors altering methylation patterns in CpGs involved in cell type estimation affect these cell types differently. Notably, since our T cells were derived from patients with CLL, this raises the question of whether the physiological changes associated with cancer are altering the methylation patterns that deconvolution algorithms rely on to estimate cell proportions.

4.2.3 Influence of senescent cell proportions on the accuracy of cell type predictions

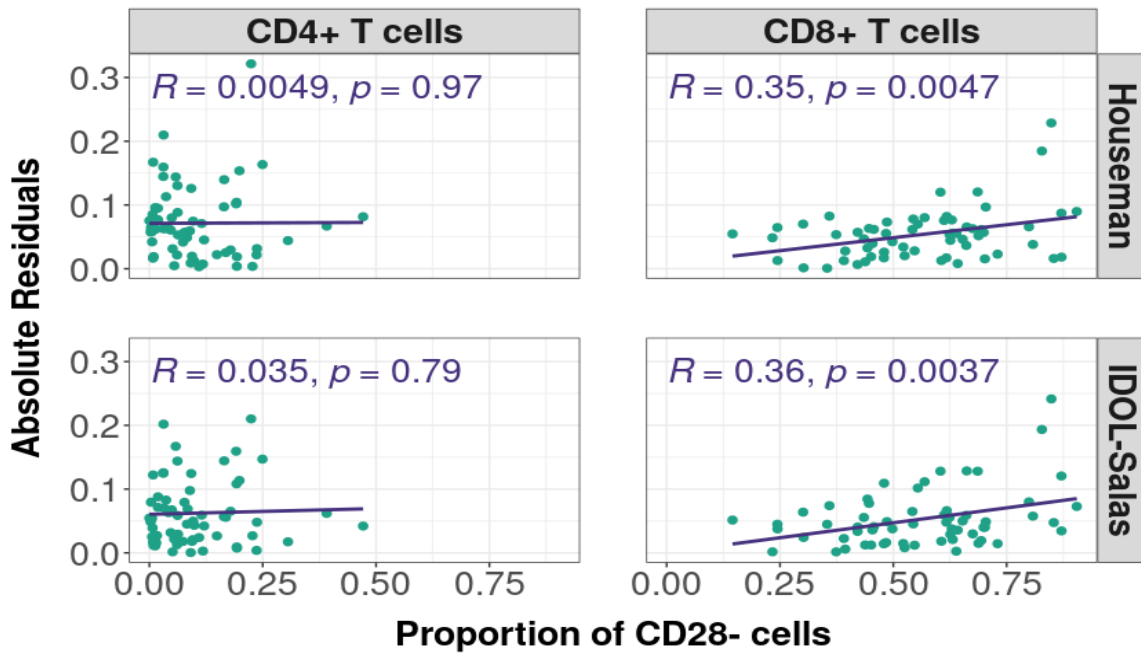
So far, our comparisons reveal that that neither the cell type (CD4⁺ T vs. CD8⁺ T) nor the predictor tool (IDOL-Salas vs. Houseman) account for the variations in our predicted T cell proportions. In fact, our previous analysis suggests that the cancerous nature of our samples may influence the DNAm patterns measured by deconvolution algorithms, which in turn affects the accuracy of cell composition estimates by deconvolution methods, which are often trained in normal tissues.

Differential DNAm patterns are not solely influenced by cell type but are also shaped by changes in the cellular environment resulting from physiological and pathological stressors such as cancer. In the following analysis, we shifted focus to explore the nature of our predicted T cell proportions and examine whether the accuracy (or inaccuracy) of these predictions is linked to cancer-related biological processes, particularly senescence.

Here, we assessed the relationship between measured proportions of senescent markers CD28⁻ and CD57⁺ and the predictive errors in the proportions of CD4⁺ and

CD8+ T cells, as measured by absolute residuals. In line with observations of senescence affecting DNAm patterns, we expected that there would be a significant correlation between the senescent T cell proportions and the magnitude of errors in the predicted cell proportion. Although our results refute that hypothesis, there were a few interesting findings as seen in Figure 7.

A



B

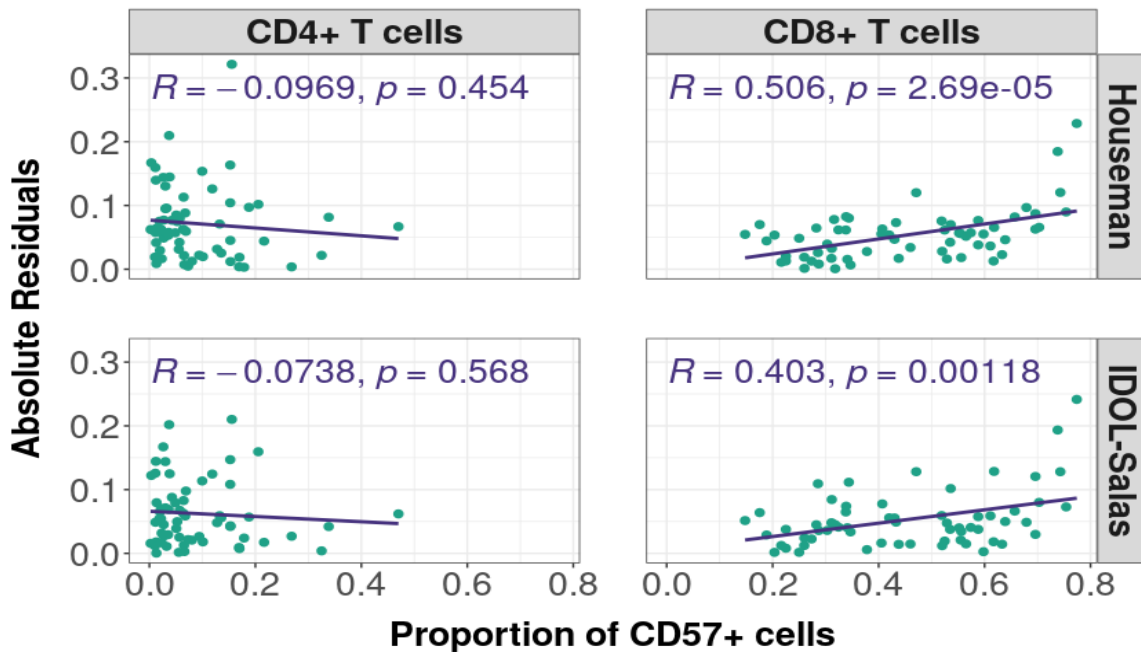


Figure 7. Relationship between senescent T-cell markers and deconvolution error in CD4+ and CD8+ T-cell proportion estimates. The influence of accumulating senescent cells on the accuracy of cell composition estimators is more pronounced in CD8+ T cells than in CD4+ T cells, highlighting cell type-specific effects. Absolute residuals refer to the difference between the predicted proportions of CD4+ T and CD8+ T cells (from the Horvath and IDOL-Salas models) and the proportions measured by flow cytometry. CD57+ and CD28- are senescent markers examined in the T cells. Significance is measured at $p < 0.05$.

In CD4⁺T cells, the proportion of cells with the senescence marker CD28⁻ did not significantly correlate ($R < 0.01$, $p = 0.970$ for Houseman; $R = 0.03$, $p = 0.789$ for IDOL-Salas) with errors in the cell proportion prediction. Similar observations were made for the senescence marker CD57⁺ ($R = -0.10$, $p = 0.454$ for Horvath; $R = -0.07$, $p = 0.568$ for Salas). The predictor tool used to derive the absolute residuals does not make a difference between Horvath and IDOL-Salas' methods.

However, in CD8⁺ T cells, there is a weak yet positive correlation ($R = 0.35$, $p = 0.005$ for Houseman $R = 0.36$, $p = 0.004$ for IDOL-Salas) between the proportion of cells with the senescence marker CD28⁻, and the magnitude of predictive errors. If we look into the proportion of senescent marker, CD57⁺, in CD8⁺T cells, we see that its correlation with predictive errors (i.e. absolute residuals) is stronger ($R = 0.51$, $p < 0.001$ for Houseman; $R = 0.40$, $p = 0.001$ for IDOL-Salas). While no differences in predictive error magnitude were observed between the two deconvolution methods for CD4⁺ T cells, Houseman-derived absolute residuals in CD8⁺ T cells showed a stronger relationship with CD8⁺ senescent T-cell proportions than with CD57⁺ senescent cell proportions.

For comparison, we also examined the relationship between the measured proportions of naïve CD4⁺ and CD8⁺ T cells and the corresponding prediction errors in CD4⁺T and CD8⁺ T cell estimates. We hypothesized that prediction errors would decrease as the proportion of naïve T cells increased, given that naïve T cell levels typically decline with age, a process associated with the accumulation of senescent cells. Our findings suggest a weak inverse relationship between CD8⁺ T cell proportions and prediction errors, although the associations were not statistically significant

(Houseman: $R = -0.16$, $p = 0.216$; IDOL-Salas: $R = -0.18$, $p = 0.152$) as shown in Figure 8. No significant or meaningful association was observed between naïve CD4⁺ T cell proportions and their corresponding prediction errors by Horvath and IDOL-Salas.

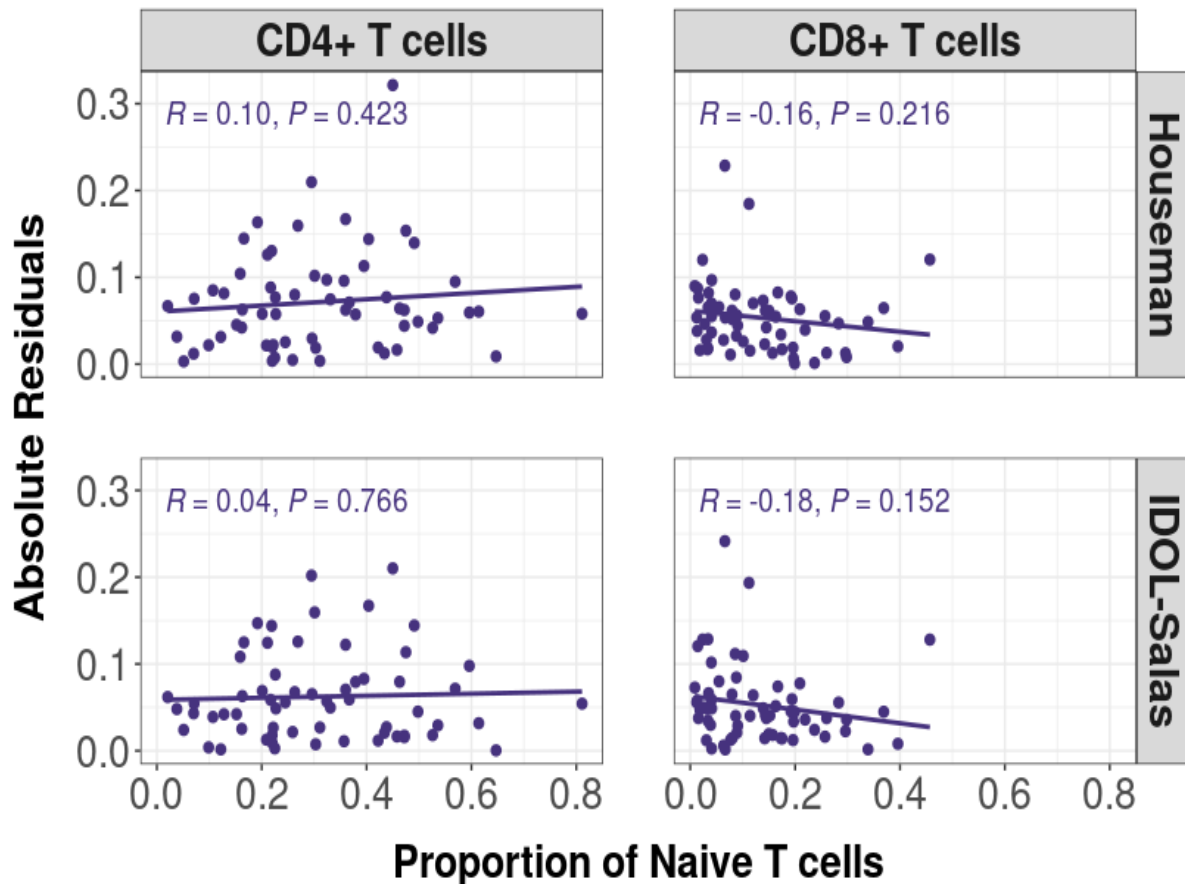


Figure 8. Relationship between naïve CD8⁺ T-cell proportions and deconvolution prediction error. As the proportion of naïve CD8⁺ T cells increase, the errors in cell type estimators appears decrease, although the relationship is not significant. Absolute residuals refer to the predictive errors or the difference between the predicted proportions of CD4⁺ T and CD8⁺ T cells and their measured counterparts. Significance is measured as $p < 0.05$.

Overall, our findings imply two major takeaways. First, prediction accuracies of deconvolution methods differ by cell type. Senescence is more strongly related to the predictive errors by cell deconvolution methods in CD8⁺ T cells whereas the relationship

between the two is not pronounced in CD4⁺T cells. Second, the type of senescence marker does seem to have a slight moderating effect on the relationship between proportion of senescent cells and the predictive errors, but this effect is specific to certain cell types. In CD8⁺T cells, the relationship between senescent cell proportion and predictive error is stronger when senescence was defined using CD57⁺ rather than CD28⁻, suggesting that CD57⁺ marker is more strongly associated with the changes in the methylation patterns involved in cell proportion prediction. In contrast, this pattern was not observed in CD4⁺T cells, further supporting that senescence-associated markers may relate to predictive error in the same across T-cell lineages.

These findings highlight that the influence of senescence on methylation-based cell proportion prediction is not uniform across cell types as evidenced by the distinct effects of CD28⁻ and CD57⁺ markers. However, it is important to note that our study focused on only two senescent markers, while several others have been reported in the literature, indicating that further investigation is needed.

4.3 Epigenetic age

4.3.1 Epigenetic age calculation and epigenetic age acceleration

Epigenetic age, determined by DNAm patterns at specific genomic sites, strongly correlates with chronological age and serves as a biomarker of biological aging. Deviations from expected epigenetic age, particularly when epigenetic age is higher than chronological age (i.e. epigenetic age acceleration), are linked to increased risks of age-related conditions, including cancer and mortality.

In the context of our study, we hypothesized that CLL patients' epigenetic ages would exceed their chronological age, as previous research suggests cancer-related

DNA methylation alterations and epigenetic drift may contribute to this effect. If we want to be more specific, since CLL samples have been divided into B and non-B cells, I expected epigenetic age acceleration differences between these two groups, with the B cell samples exhibiting greater acceleration due to the malignancy's primary origin in B cells.

One important clinical feature of CLL relevant to our study is the comorbidity burden, which has been reported to adversely affect patient outcomes, treatment decisions, and overall survival. Many CLL patients, who tend to be diagnosed at advanced age, suffer from chronic conditions such as cardiovascular diseases and hypertension, which are associated with epigenetic age acceleration in younger and non-cancerous populations. We hypothesized that CLL patients with a higher comorbidity burden would exhibit greater epigenetic age acceleration than individuals without CLL, even when comorbidities are present. Although the absence of normal control samples precluded a direct comparison with non-CLL individuals, we still expected that, within the CLL cohort, patients with more severe comorbidity burden (CIRS score ≥ 6) would demonstrate greater epigenetic age acceleration than those with lower comorbidity burden (CIRS score < 6).

In our analysis, we observed a moderate and statistically significant positive correlation between DNAm age and chronological age in non-B cells (e.g. Horvath ($r = 0.46$, $p < 0.001$), Hannum ($r = 0.55$, $p < 0.001$), PhenoAge ($r = 0.48$, $p < 0.001$) across all three chronological clocks. These findings indicate that the relationship between DNAm age and chronological age was more preserved in non-B cells than in B cells within our cohort (Figure 9). In contrast, the relatively weaker and non-significant

correlation between DNAm age and chronological age in B cells (e.g. Horvath ($r = 0.24$, $p = 0.055$), Hannum ($r = 0.19$, $p = 0.139$), suggesting a less consistent linear relationship and that there is no evidence to support a reliable relationship between the two. The only significant correlation between chronological age and DNAm age is one where the latter is measured by PhenoAge ($r = 0.28$, $p = 0.026$).

We also observed relatively low variability between DNAm age and chronological age in our non-B cell samples, as evidenced by the data points clustering tightly around the regression line. In contrast, B cells exhibited a greater spread of data points away from the regression line, some of which can be seen as having a higher DNAm age than their chronological age. This means that the B cells of some of the CLL patients are aging at a faster rate than their actual or chronological age. No significant differences were observed among the three epigenetic clocks in the relationship between DNAm age and chronological age within cell groups. In non-B cells, the correlations between chronological age and DNAm age derived from the three clocks are all positive and significant. However, when ranking the correlation coefficients by strength, Hannum clock ranks first, followed by Horvath, and PhenoAge, in that order. In B cells, the association between DNAm age and chronological age is weaker compared to non-B cells and are non-significant regardless of the epigenetic clock used to derive DNAm age.

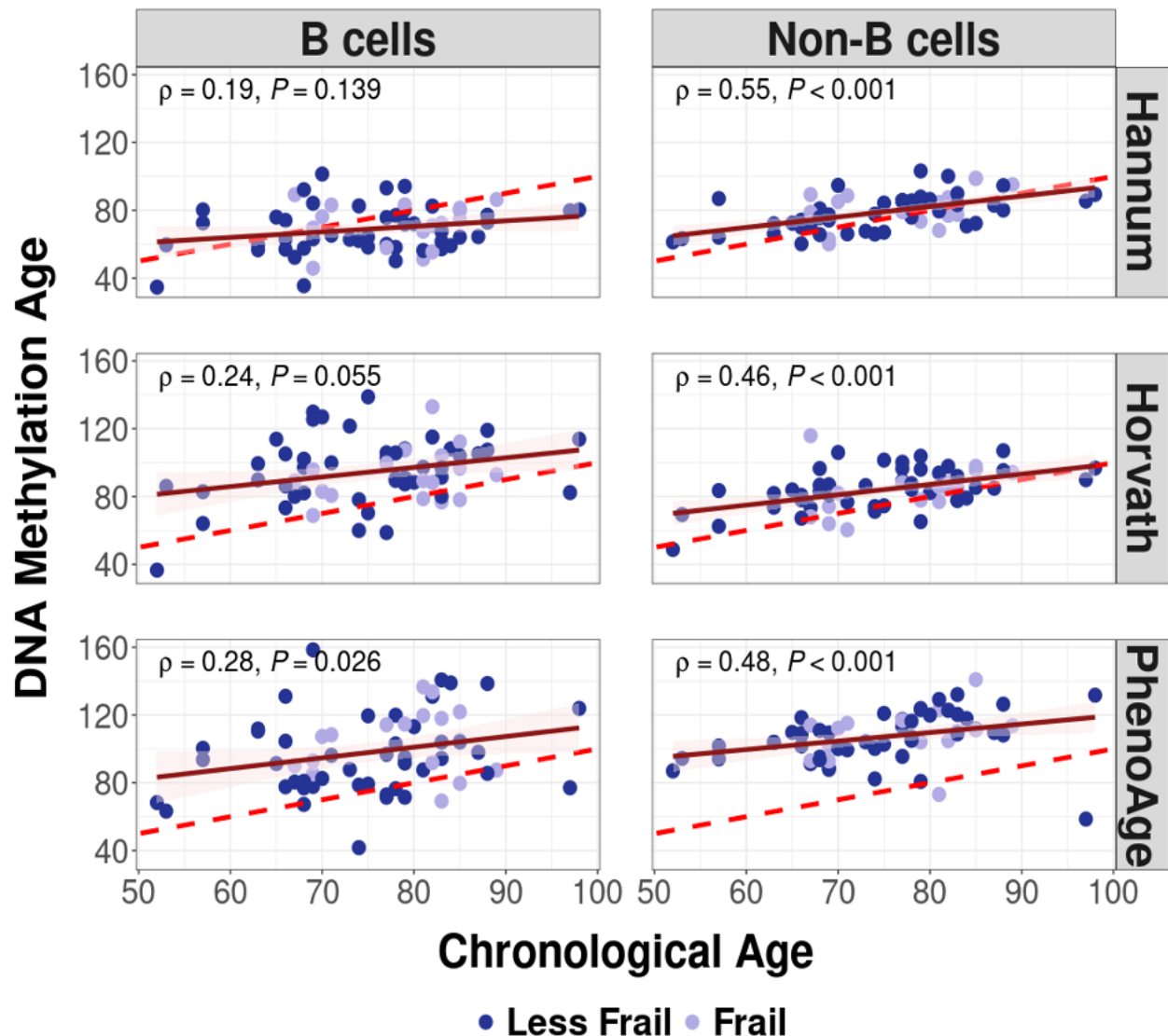


Figure 9. Relationship between chronological age and DNA methylation age in B-cell and non-B-cell fractions. DNA methylation age in non-B cells more strongly correlates with chronological age than with B cells, implying the two cell type groups are aging differently. Three different sets of DNAm age were calculated by three clocks: Horvath, Hannum, and PhenoAge. All three epigenetic clocks were trained in samples that include the cell types and age range of our CLL cohort. Dashed red lines indicates a perfect relationship between DNAm age and chronological age. Solid brown line indicates the regression line. Less frail is defined as CIRS score of less than 6. Frail is defined as CIRS of 6 or greater.

Although the differences in correlation metrics may be driven by variations between cell groups (B versus non-B cell groups), there are noticeable distinctions

between the epigenetic clocks regarding how closely the estimated or predicted DNAm age matches chronological age. The DNAm ages of non-B cells derived from the Hannum clock most closely match their chronological age counterparts, as shown by the regression line's near alignment with the one-to-one relationship line. This is followed closely by the Horvath clock, in which the regression line moves towards the one-to-one line as the non-B cells age. PhenoAge also follows this pattern of DNAm age getting closer to its chronological age, but most of the non-B cells had higher DNAm age than their chronological age

In summary, our results highlight distinct DNA methylation age patterns between malignant B cells and non-malignant non-B cells in CLL patients, underscoring the importance of cell type and specific analysis in heterogeneous samples. Notably, DNAm age tracked more closely with chronological age in non-B cells, while this relationship was weaker in CLL B cells across all three clocks. This discrepancy suggests that the malignant state of B cells may disrupt normal age-related methylation changes, reflecting cancer-associated reprogramming of the epigenome. While DNAm age provides a snapshot of the biological aging state of each cell population, further analysis using age acceleration metrics may offer deeper insights into how cancer affects epigenetic aging in CLL, and how this may relate to clinical outcomes such as treatment response and prognosis.

4.3.2 Association between epigenetic age acceleration and CLL risk factors (comorbidity burden, sex, and smoking history)

Epigenetic age alone does not inform us of the rate at which the sample is aging biologically compared to its chronological age. On the other hand, epigenetic age acceleration (EAA), when DNAm age is greater than chronological age, has been associated with a variety of adverse health conditions. A person with accelerated epigenetic aging might be at greater risk for developing age-related diseases earlier than expected including cardiovascular disease, cancer, diabetes, and neurodegenerative diseases, thus making EAA more informative of disease risks and other health complications.

For the following analysis, we conducted t-tests to examine the difference in age acceleration between B and non-B cells and whether those differences are stratified by various clinically relevant variables that can influence risk, prognosis, and treatment of CLL. We looked into seven risk factors in CLL that have been divided into two sets: prognostic and non-prognostic set of variables. First, we examined the non-prognostic variables such as sex, comorbidity burden, and smoking history. We expect the male patients, those with a high comorbidity burden or CIRS score of 6 or greater, and patients who had a history of smoking would exhibit greater age acceleration. These factors have been linked to poorer health outcomes. Notably, the incidence of CLL is higher in men, with a male-to-female ratio of 2:1. Comorbidities are prevalent among CLL patients and some of the common ones include cardiovascular diseases and diabetes, both of which have been significantly linked to age acceleration. Meanwhile, smoking has been found to accelerate the aging of multiple tissues. Similar to our

previous expectations, we also anticipate differences in EAA across cell types, and that these differences will vary depending on the epigenetic clock used. Given that epigenetic ages were derived from the Horvath, Hannum, and PhenoAge clocks, each sample has three corresponding EAA estimates.

Our results demonstrated no significant differences in age acceleration between the following groups: high (CIRS ≥ 6) and low (CIRS < 6) comorbidity burden groups (Figure 10 top panel), male and female CLL patients (Figure 10 middle panel), and individuals with a smoking history compared to those without (Figure 10 bottom panel). Regardless of whether B or non-B cells were examined, these findings suggest that comorbidity burden, sex, and smoking history do not influence the DNAm patterns associated with age prediction in these cell types. Despite the absence of EAA differences between dichotomized groups on non-prognostic CLL risk factors, contrary to our initial expectations, DNAm pattern disparities between B and non-B cells persist.

Notably, the EAA values in non-B cells were relatively uniform, exhibiting a narrower range, whereas B cells showed a substantially wider spread in EAA values. This broader variability among B cells suggests that their cancerous nature in CLL patients may have a more pronounced and heterogeneous effect on epigenetic aging of these cells, accelerating it in some patients while appearing decelerated in others. Both PhenoAge and Horvath clocks detected substantial variability in age acceleration between B and non-B cells. In contrast, Hannum clock-derived age acceleration showed more subtle variation between B and non-B cells compared to the other two clocks.

This analysis demonstrates that non-prognostic CLL variables, including sex, comorbidity burden, and smoking history, have minimal to no impact on the aging of B

and non-B cells in CLL patients. Consequently, we now shift our focus to the prognostic CLL variables to investigate whether these factors influence the rate of aging in the B and non-B cells of CLL patients.

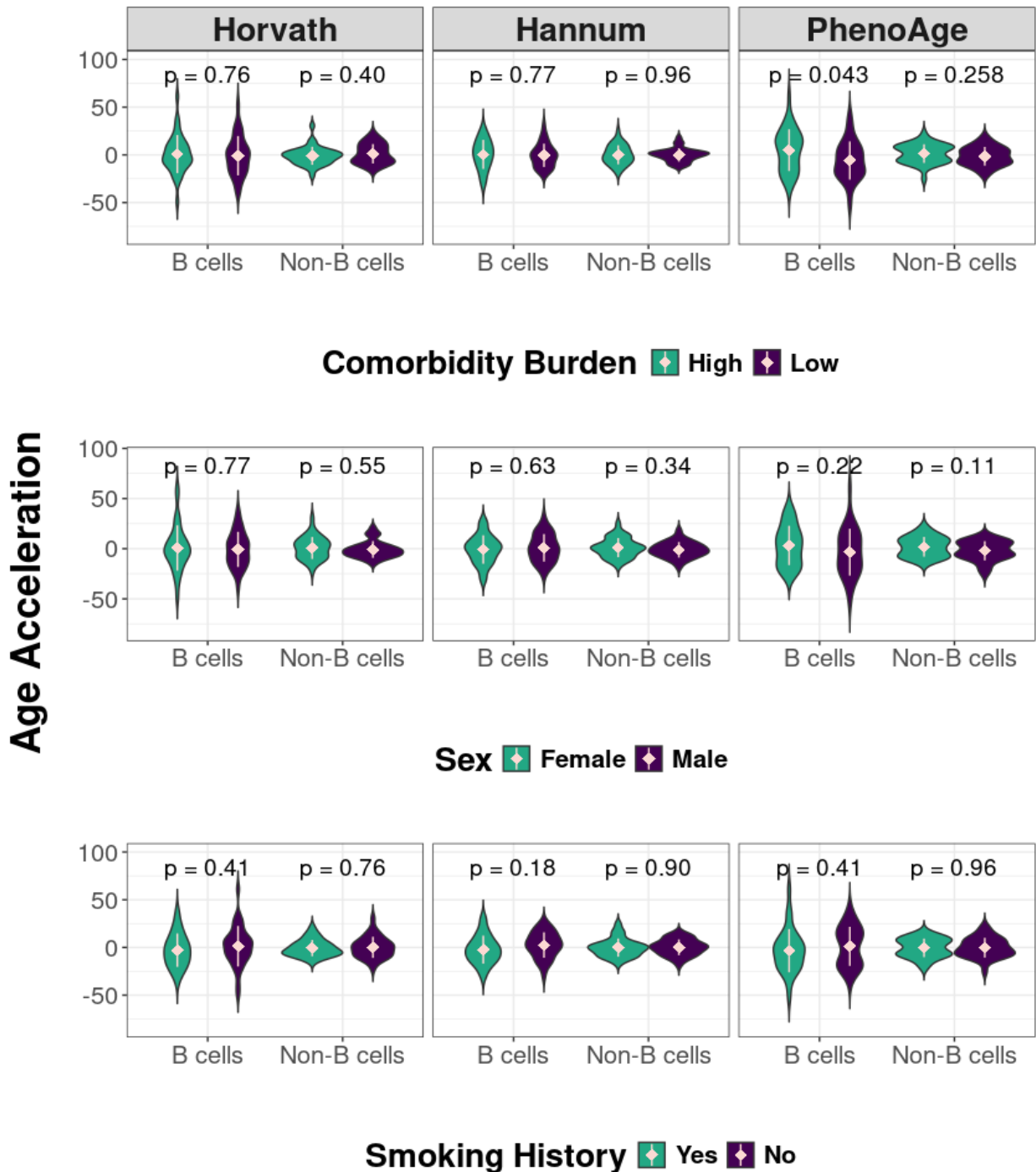


Figure 10. Influence of sex, comorbidity burden, and smoking history on age acceleration in B and non-B cells of CLL patients. The male sex, high comorbidity burden, and history of smoking do not significantly influence the rate of aging in B and non-B cells in CLL patients. Comorbidity burden refers to the CIRS score of CLL patients, with CIRS of 6 or greater is considered high comorbidity burden and CIRS below 6 is considered low burden. Smoking history encompasses patients who reported to be current smokers at the time of CLL diagnosis or previously smoked at some point in their life before CLL diagnosis. Age acceleration was calculated from Horvath, Hannum, and PhenoAge clocks. Dichotomized group differences were assessed with a t-test. P value > 0.05 is considered significant.

4.2.3 Association between epigenetic age acceleration and CLL risk factors (IGHV mutation, p53 expression, Zap-70 expression, and CD38+ level of expression)

Although our data do not support the influence of comorbidity burden or smoking history on DNAm patterns for age prediction, known prognostic markers of CLL emerge as potential candidates that could explain variation in DNAm patterns linked to age acceleration. Four markers were examined, including IGHV mutation status (e.g., unmutated vs. mutated), P53 mutation status (e.g., present vs. absent), CD38+ expression (e.g., $\geq 20\%$ vs. $< 20\%$), and ZAP-70 expression (e.g., negative vs. positive). Unmutated IGHV, mutated P53, 20% expression of CD38+, and positive expression of Zap-70 are associated with poorer outcomes in CLL patients. Since the literature indicates that several adverse health outcomes are associated with age acceleration, we hypothesized that poor prognostic markers in CLL would also associate with age acceleration. Among the four markers examined, IGHV and P53 mutation status are the most informative in predicting CLL treatment outcomes and disease aggressiveness. Therefore, we anticipated that age acceleration would be more pronounced and significant in relation to these two markers compared to CD38+ and ZAP-70 expression. To test our hypotheses, we conducted twelve t-tests similar to the one we performed with comorbidity burden and history of smoking. In addition to age acceleration comparisons between adverse and less adverse prognostic markers, we also compared age acceleration between cell types (B and non-B cells), and chronological epigenetic clocks (Hannum, Horvath, and PhenoAge).



Figure 11. Comparison of epigenetic age between IGHV/ZAP-70 molecular subgroups in CLL B cells. B cells that have mutated IGHV gene and negative Zap-70 expression are epigenetically older compared to B cells with unmutated IGHV gene and positive Zap-70 expression. Each panel refers. * indicates statistical significance of $p < 0.05$.

Our analysis revealed that the age acceleration difference between unmutated and mutated IGHV mutation status meets the significance threshold in B cells ($p = 0.031$). Furthermore, the age acceleration difference between the negative and positive Zap-70 expression is significant ($p = 0.017$) in B cells. The age acceleration in both of these cases were derived from the Hannum clock. In particular, mutated IGHV and negative Zap-70 expression have greater age acceleration than unmutated IGHV and positive Zap-70 expression, respectively. Age acceleration differences between negative and positive p53 expression and between 20% or more expression of CD38+ and less than 20% were not significant, regardless of the cell type and epigenetic clock used.

In contrast to the previous age acceleration analysis involving sex, comorbidity burden, and smoking history, our findings suggest that the choice of epigenetic clock plays a role in detecting differences in age acceleration between adverse and less adverse prognostic markers. Among the three clocks used, only Hannum yielded significant results, highlighting the influence of clock selection on the observed associations.

Overall, our findings suggested that B cells from CLL patients with mutated IGHV seem to age faster according to the Hannum clock. This finding was unexpected, as we initially hypothesized that unmutated IGHV, typically associated with a poorer prognosis, would show greater age acceleration. However, this association was observed only with the Hannum clock and was not replicated in the other two epigenetic clocks. Similarly, we found that CLL patients with negative ZAP-70 expression had significantly greater age acceleration than those with positive ZAP-70 expression, but again, this result was

only significant when using the Hannum clock. These unexpected results seem to imply that relationship between age acceleration and markers of poor prognosis is not always consistent, highlighting how aging in relation to adverse health outcomes is complex.

Contrary to our hypothesis, the CLL B cells did not consistently show accelerated aging among patients with poorer prognostic indicators. In some cases, individuals with unfavorable prognostic features even exhibited slower epigenetic aging compared to those with more favorable profiles. This paradox highlights the complex and sometimes counterintuitive ways in which CLL may influence biological aging, particularly within the malignant B cell population.

In contrast, P53 mutation status and CD38+ expression ($\geq 20\%$) did not show significant differences in age acceleration in either B cells or non-B cells. Moreover, neither the choice of epigenetic clock nor the cell type used in these comparisons had a significant effect on the observed differences in age acceleration for P53 or CD38+ expression, implying that P53 mutation status and CD38+ expression do not strongly influence the DNAm aging signals captured by epigenetic clocks.

The fact that only the Hannum clock yielded significant results further supports the notion that associations between prognostic markers and age acceleration are also clock-dependent. Since different epigenetic clocks capture different aspects of biological aging, it is worth considering other types of epigenetic clocks to clarify the relationship between poor prognostic outcomes and changes in age-dependent DNAm patterns. So far, we examined chronological clocks that focuses on DNAm changes that correlate with chronological age. One specific type of epigenetic clock that we seek to explore focuses is the mitotic clock, which estimates age based on the DNAm changes that

accumulate as a result of a cell's replicative history.

4.2.4 Association between epigenetic age acceleration and CLL risk factors (IGHV mutation, p53 expression, Zap-70 expression, and CD38+ level of expression)

Mitotic clocks measure the accumulation of DNAm changes over time as a result of mitotic divisions. In rapidly dividing cells, such as cancer cells, these changes may be more pronounced and result in malignant transformations. By examining mitotic clocks, researchers can track how this shift occurs and understand how it may contribute to CLL's progression or poor treatment response. In this analysis, we focused on EpiTOC clock, which yields a score, pcgtAge, that reflects the cumulative increase in DNAm in the polycomb group target (pcgt) promoter CpGs, as a result of cell replication errors over time. PCGTs are typically unmethylated in many fetal tissues. Higher scores of pcgtAge, indicative of greater methylation errors due to increased mitotic activity, has been associated with accelerated aging in several solid tumors and is said to be predictive of cancer risk in precancerous tissues. We hypothesized that similar observations would be seen with our CLL samples in that we expected that pcgtAge scores would be higher in all seven CLL markers (the male sex, high comorbidity burden, smoking history, unmutated IGHV, negative P53 expression, positive Zap-70 expression, 20% or more CD38+ expression) that are associated with more aggressive CLL course and higher proliferation rates. We also expected differences in pcgtAge scores between B and non-B cells would be apparent and significant. To test our hypothesis, we conducted t-tests on seven dichotomized risk factors of CLL.

What we found is that there is that CLL patients with mutated IGHV exhibit greater pcgtAge score in B cells compared to those with unmutated IGHV ($p = 0.0029$) as shown in Figure 12. Another significant finding is that CLL patients with a negative Zap-70 expression also exhibit greater pcgtAge scores in B cells than those with positive Zap-70 expression ($p = 0.0023$). Meanwhile, there is no significant difference in pcgtAge scores between dichotomized groups of sex, comorbidity burden, smoking history, p53 expression, and CD38+ expression.

Our expectation that pcgtAge scores would be profound and significant between B and non-B cells have been met. B cells exhibit greater pcgtAge scores compared to non-B cells, lending support to the presumption that primary malignant cells in CLL, the B cells, exhibit DNAm changes that are associated with greater risk of developing the disease. Despite the significant difference in cancer risk attributed cell types, this was not replicated when comparing other CLL risk factors.

Similar to our results with age acceleration using the three chronological clocks, our findings do not support our hypothesis that seven CLL risk factors would have greater pcgtAge scores. Notably, IGHV-unmutated and ZAP-70–positive cases exhibited lower pcgtAge than IGHV-mutated and ZAP-70–negative cases. Because pcgtAge is derived from DNA methylation levels at PCGT promoter CpGs and was developed to capture a replication or cell turnover-linked methylation drift signal, this finding suggests that pcgtAge do not map straightforwardly onto conventional CLL prognostic risk groupings. One plausible interpretation is that pcgtAge in CLL B cells is influenced by maturation-linked epigenetic subtype differences described in CLL methylome studies where cases fall along a normal B-cell maturation continuum from

naïve-like to memory-like states^{260,261}. Under this framework, IGHV-mutated (memory-like) and IGHV-unmutated (naïve-like) CLL can differ in baseline methylation programs carried from their developmental context^{260,261}. If methylation at PCGT promoter CpGs contributing to pcgtAge differ with maturation state, then it is possible that pcgtAge may differ between IGHV-defined subtypes for reasons related to epigenetic subtype-associated methylation differences. However, this would need to be tested directly. Future analyses should test this directly by examining whether the individual PCGT promoter CpGs used to compute pcgtAge show differential methylation by IGHV and ZAP-70 subtype and whether IGHV-associated differentially methylated sites are enriched among the epiTOC PCGT CpG set. In summary, five of the CLL risk factors (the male sex, high comorbidity burden, smoking history, negative TP53 expression, and 20% or more CD38+ expression) do not significantly influence the DNAm changes in CLL cells that reflect the methylation maintenance errors during replication, while two of the CLL risk factors (unmutated IGHV and positive Zap-70 expression) appear to exhibit DNAm patterns changes associated with decreased cell proliferation in several cancer cells. On the other hand, cells of CLL patients appear to exhibit DNAm patterns in line with what is seen in cancer cells (i.e. rapidly proliferating cells), regardless of the prognostic markers.



Figure 12. Comparison of pcgtAge between IGHV/ZAP-70 molecular subgroups in CLL B cells. B cells that have mutated IGHV gene and negative Zap-60 expression are epigenetically older compared to B cells with unmutated IGHV gene and positive Zap-70 expression. Each panel refers. * indicates statistical significance of $p < 0.05$.

4.2.5 Association between epigenetic age acceleration and senescent cell proportion

Aging is accompanied by profound immunological remodeling, including the gradual accumulation of senescent T cells that exhibit altered function, impaired proliferative capacity, and pro-inflammatory secretory profiles. At the molecular level, these processes parallel changes in the epigenome, where DNA methylation patterns progressively drift from their youthful state. Epigenetic clocks such as those developed by Hannum, Horvath, and Levine (PhenoAge) quantify this drift and generate measures of epigenetic age acceleration, capturing the degree to which an individual appears epigenetically older, or younger, than expected from their chronological age. Because immunosenescence is a hallmark of biological aging, we hypothesized that individuals with higher epigenetic age acceleration would also exhibit a greater proportion of senescent T-cell phenotypes. In this study, we therefore examined the relationship between established epigenetic clock–derived age acceleration metrics and T-cell senescence markers, reasoning that a divergence between chronological and epigenetic age would parallel an increased burden of senescent immune cells and reflect broader systemic aging dynamics.

Despite our initial hypothesis, no significant associations were observed between epigenetic age acceleration in non-B cells and the proportion of senescent T-cell subsets, irrespective of whether age acceleration was derived from B-cell or non-B–cell methylation profiles (Figure 13-14). This lack of correlation suggests that epigenetic age acceleration and T-cell senescence may reflect distinct biological processes, rather than tightly coupled manifestations of a unified aging axis. In other words, age-related methylation drift captured by clocks such as Hannum, Horvath, and PhenoAge does not

appear to be the primary driver of senescent T-cell accumulation, implying that immune senescence in CLL may be shaped more strongly by disease-specific factors, microenvironmental stressors, or chronic antigenic stimulation rather than systemic epigenetic aging alone.

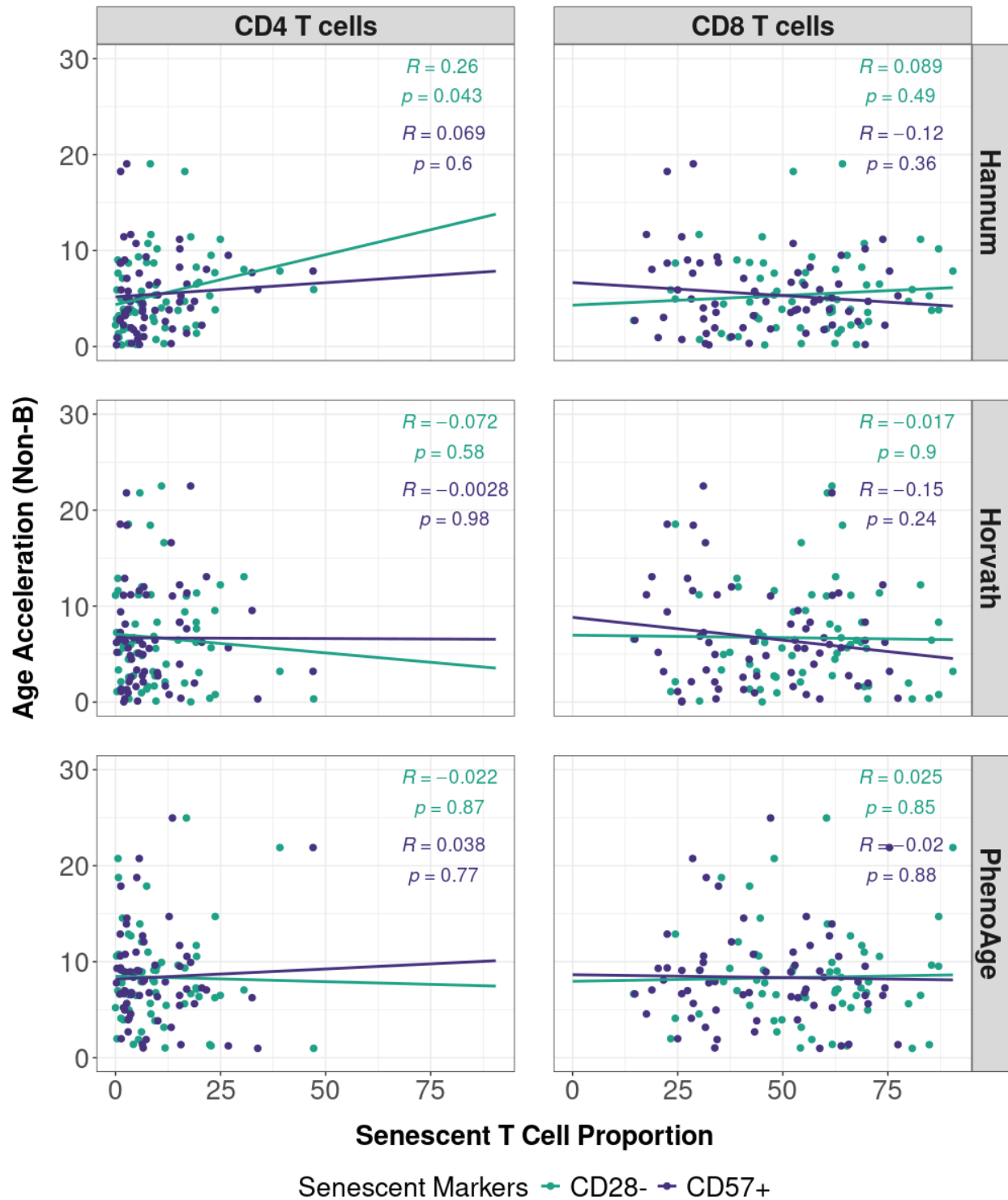


Figure 13. Association between non-B-cell age acceleration and proportions of senescent T cells. No significant association between age acceleration in non-B and senescent T cell proportions were found. Pearson correlation coefficients (R) and corresponding p -values are shown for each relationship. Statistical significance is defined as $p < 0.05$.

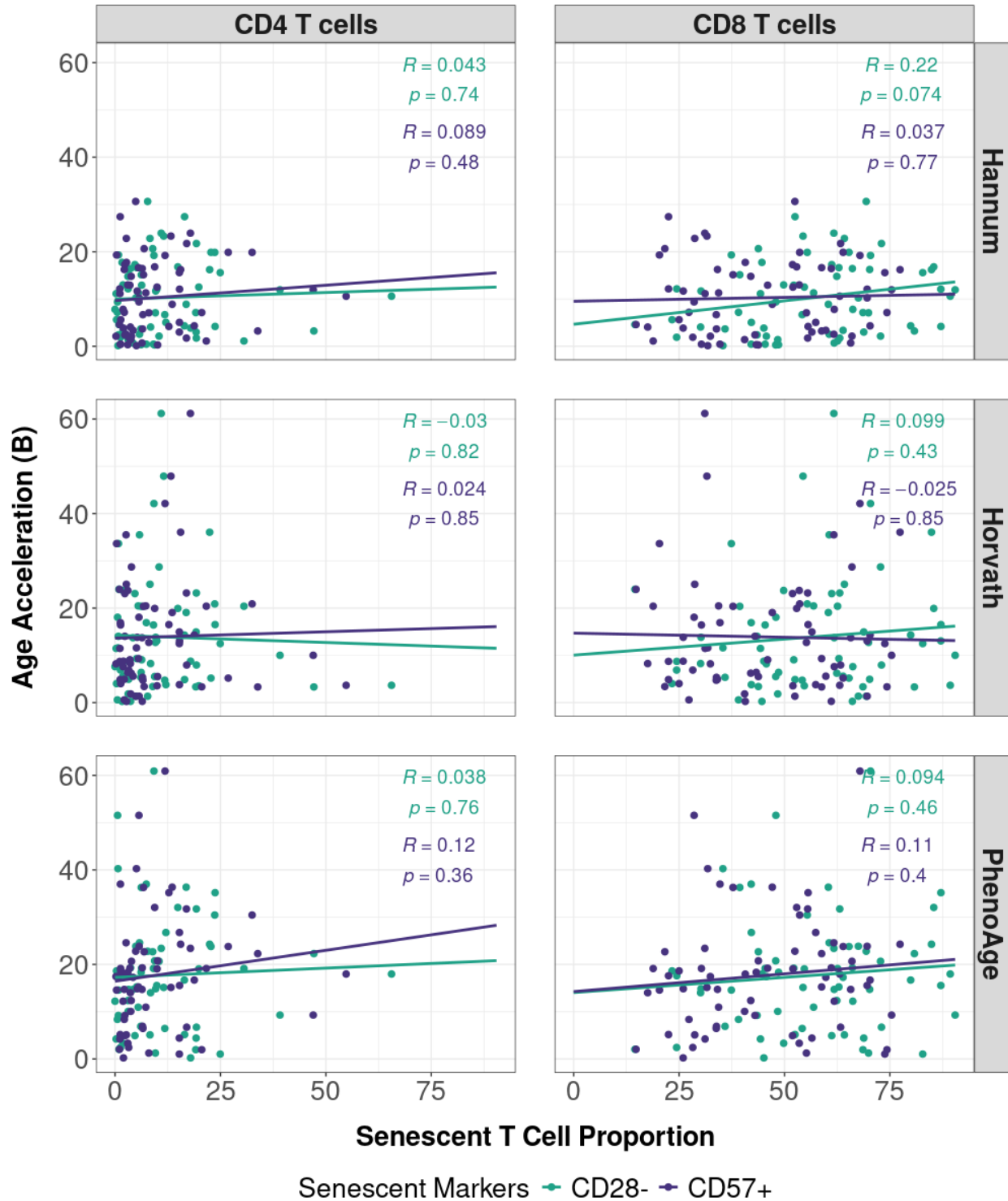


Figure 14. Association between B-cell age acceleration and proportions of senescent T cells. Age acceleration in B cells is not significantly influenced by the proportion of senescent markers, CD28- and CD57+. Pearson correlation coefficients (R) and corresponding p-values are shown for each relationship. Statistical significance is defined as $p < 0.05$.

In contrast, when examining DNA methylation-based cancer risk using pcgtAge in non-B cells, we observed statistically significant positive associations with both CD4⁺ senescent T-cell proportions (Figure 15). (CD28⁻: $r = 0.33$, $p = 0.0084$; CD57⁺: $r = 0.32$, $p = 0.011$) and CD8⁺ senescent T-cell proportions (CD28⁻: $r = 0.38$, $p = 0.0025$; CD57⁺: $r = 0.35$, $p = 0.0051$). Corresponding associations using B-cell-derived pcgtAge were weaker and did not reach statistical significance in this cohort (CD4⁺CD28⁻: $r = 0.18$, $p = 0.15$; CD4⁺CD57⁺: $r = 0.24$, $p = 0.052$; CD8⁺CD28⁻: $r = 0.081$, $p = 0.53$; CD8⁺CD57⁺: $r = 0.15$, $p = 0.24$). Collectively, these patterns suggest that methylation alterations linked to polycomb-regulated cancer risk in non-B hematopoietic compartments may be more closely intertwined with systemic immune aging, particularly T-cell senescence, than conventional epigenetic clocks.

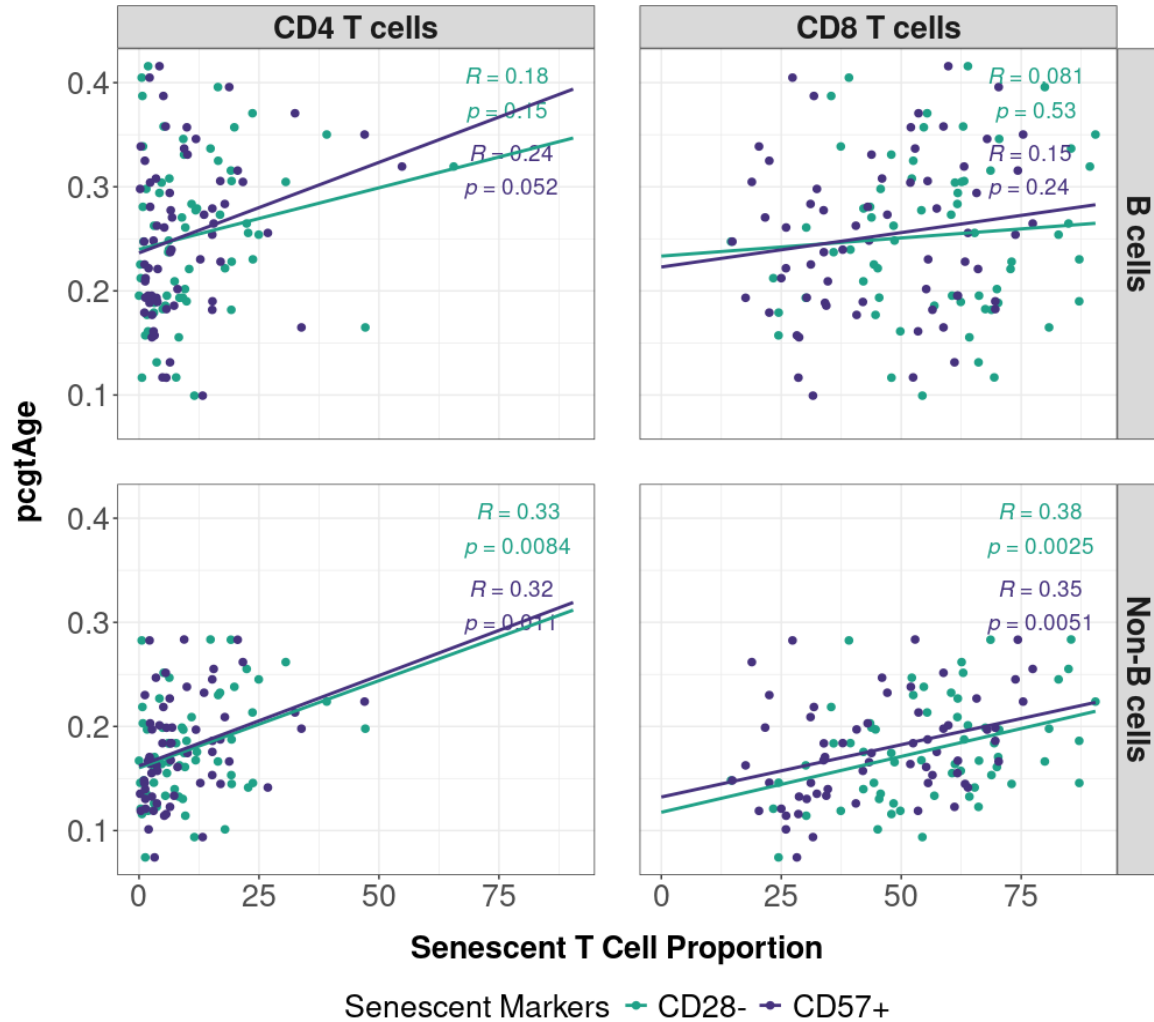


Figure 15. Association between senescent T-cell proportions and pcgtAge in B and non-B cells. Increased proportions of senescent T cells are more strongly correlated with DNA methylation changes associated with cancer risk, as measured by pcgtAge, in non-B cells of CLL patients compared to their B cells. Pearson correlation coefficients (R) and corresponding p-values are shown for each relationship. Statistical significance is defined as $p < 0.05$.

4.3 Epigenome-wide association study

To examine whether comorbidity burden was associated with altered DNA methylation in the non-B cell fraction of CLL patients, we compared patients with high versus low comorbidity burden across the epigenome. Overall, the EWAS showed a largely null pattern, with only a small subset of CpG sites standing out against the background distribution. Thirty-six CpG sites showed more than a 1% methylation difference ($|\Delta\beta| \geq 0.01$) between the high and low comorbidity groups and met the nominal threshold of $p < 1 \times 10^{-5}$, while 9 of these exceeded the more stringent threshold of $p < 1 \times 10^{-6}$ (Figure 16). These findings indicate that only a limited number of loci showed measurable methylation differences associated with comorbidity burden, and that these differences were modest in magnitude. The signals were observed in both directions, with 20 CpGs showing higher methylation and 16 showing lower methylation in patients with higher comorbidity burden, indicating no consistent global shift toward either hypermethylation or hypomethylation. These loci mapped to multiple genes, including FAM133B/FAM133DP, PTPRU, LHX1, GNAQ, CBS, and SENP5, while some had no clear gene annotation (Table 3). However, no CpG sites remained significant after multiple testing correction (0 CpGs with $FDR < 0.05$). This indicates that, although several loci showed suggestive nominal signals, none met the threshold for statistically robust epigenome-wide significance after accounting for the large number of CpG sites tested. Taken together, these results suggest that higher comorbidity burden in CLL patients is associated with subtle, localized methylation differences in non-B cells, rather than widespread remodeling of the methylome.

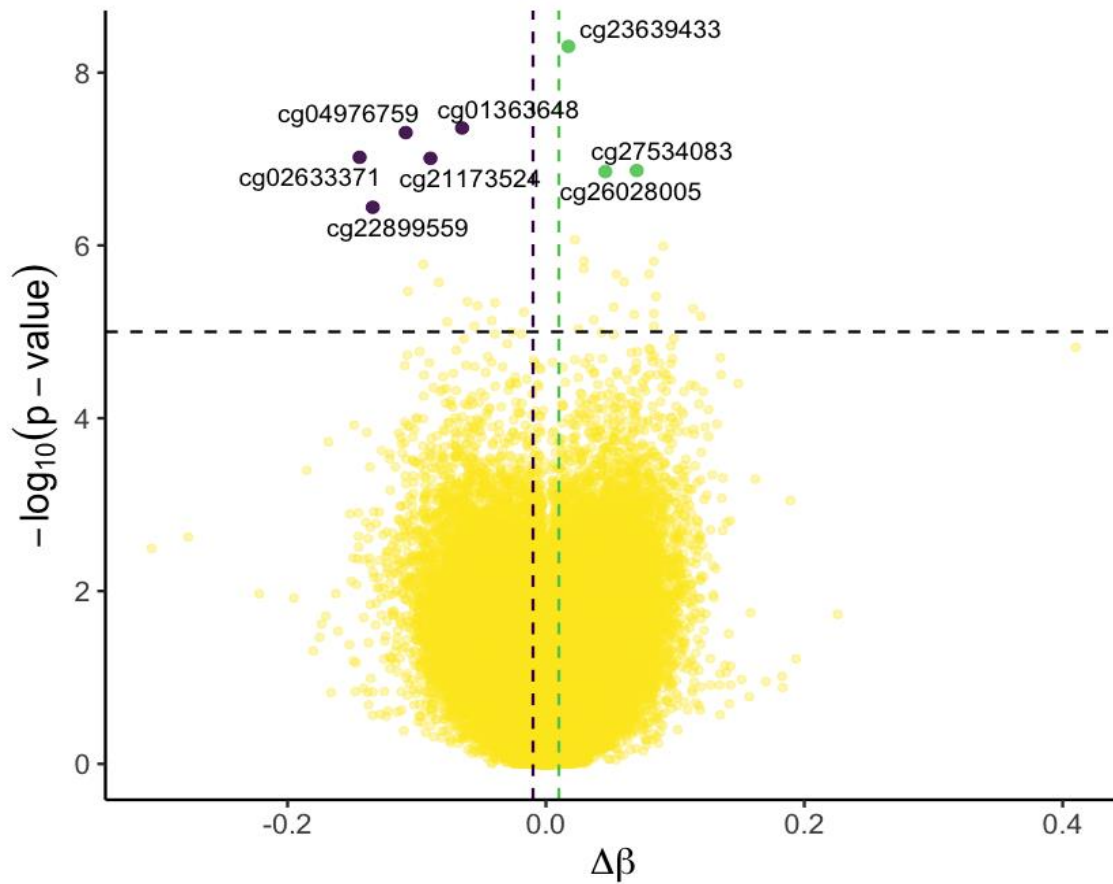


Figure 16. Volcano plot of EWAS results for comorbidity burden in non-B cells. 36 CpG sites that met the volcano plot highlighting criteria of nominal $p < 1 \times 10^{-5}$ (6 of which met the threshold of $p < 1 \times 10^{-6}$) and absolute methylation difference $|\Delta\beta| \geq 0.01$. Positive $\Delta\beta$ values indicate higher methylation in patients with high comorbidity burden compared to patients with low comorbidity burden, whereas negative $\Delta\beta$ values indicate lower methylation. Purple points indicate CpGs with lower methylation in the high comorbidity group, while green points indicate CpGs with higher methylation. Dashed lines denote significance thresholds, with the horizontal line representing $p = 1 \times 10^{-5}$ and vertical lines representing $\Delta\beta = -0.01$ and 0.01 . Labeled points correspond to CpG sites exceeding both thresholds.

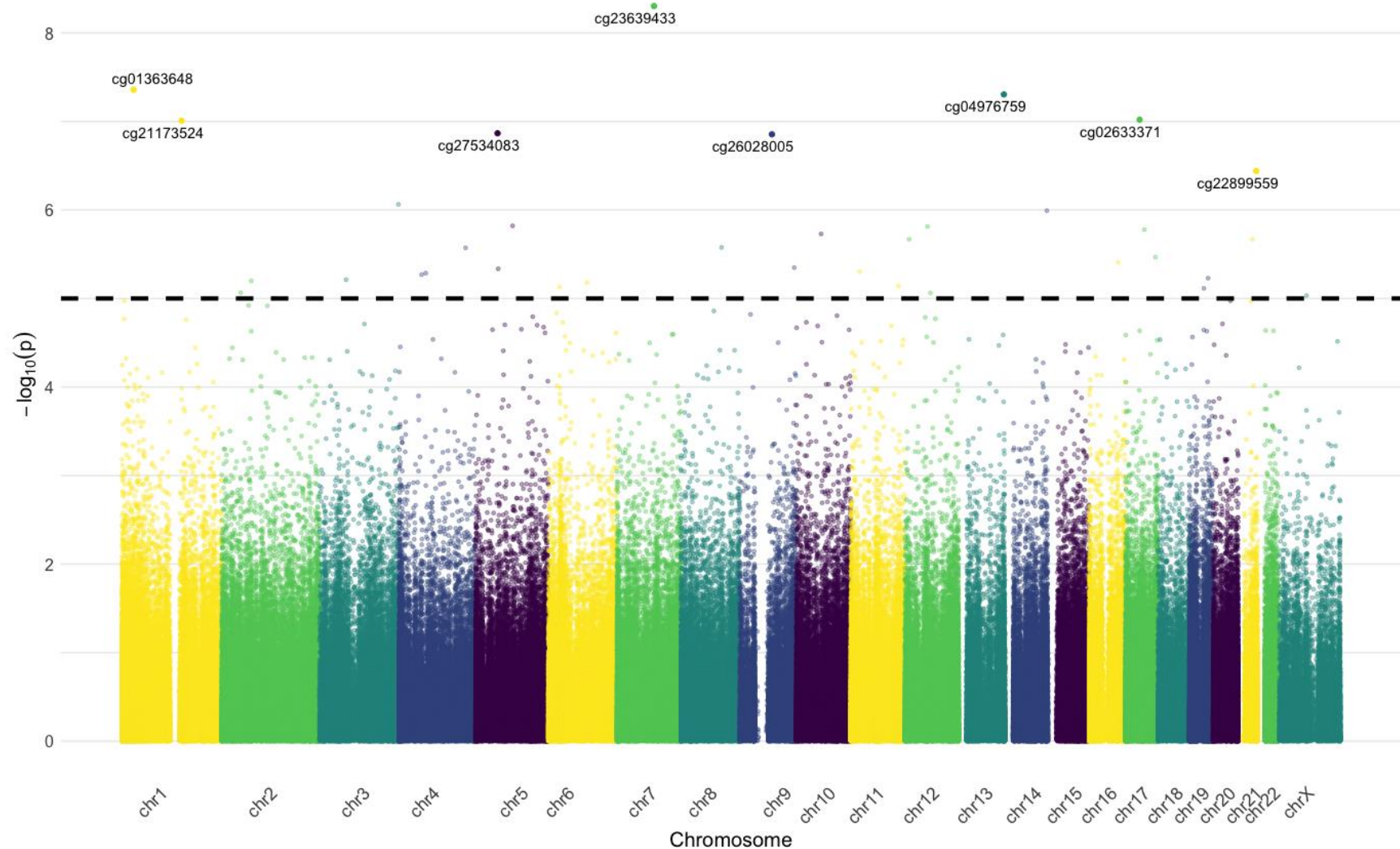


Figure 17. Manhattan plot of EWAS results for comorbidity burden in non-B cells. Significant CpG sites are sparsely distributed across the genome, with the strongest association observed at cg23639433. Additional signals are observed on chromosomes 1, 5, 9, 13, 16, and 22, although no clear clustering of signals is evident within a single genomic region. The dashed horizontal line indicates a nominal significance threshold of FDR-adjusted p-value < 0.05).

Table 3. CpG sites highlighted in the volcano plot from the EWAS of comorbidity burden in non-B cells from CLL patients.

CpG	Chromosome	Genomic Position	Methylation Difference	P value	FDR	Mapped Gene
cg23639433	7	92207069	0.01741542	4.98E-09	0.00339093	FAM133B
cg01363648	1	29589154	-0.0649215	4.38E-08	0.01124131	PTPRU;
cg04976759	13	112669393	-0.1085182	4.95E-08	0.01124131	
cg02633371	17	35298618	-0.1442301	9.56E-08	0.0133816	LHX1
cg21173524	1	149144820	-0.0894076	9.82E-08	0.0133816	
cg27534083	5	54515036	0.07028218	1.36E-07	0.01361007	
cg26028005	9	80378603	0.04599594	1.40E-07	0.01361007	GNAQ
cg22899559	21	44483554	-0.1340723	3.62E-07	0.0308314	CBS
cg08370082	3	196616876	0.02239674	8.66E-07	0.06555699	SENP5
cg07022764	14	104569519	0.09081687	1.02E-06	0.06930412	ASPG
cg24885828	5	91997234	0.02962542	1.51E-06	0.08720556	
cg09485707	12	56222623	0.08350003	1.54E-06	0.08720556	TMEM198B
cg11383850	17	47209546	-0.0949589	1.67E-06	0.08729825	B4GALNT2
cg02871615	10	61824717	0.02933922	1.87E-06	0.09074693	ANK3
cg17483598	12	10538268	0.07991262	2.14E-06	0.09160907	LOC101928100; KLRK1; KLRC4-KLRK1
cg09844354	21	35212282	0.05448239	2.15E-06	0.09160907	ITSN1
cg17292100	8	101250595	0.06048006	2.65E-06	0.10142522	SPAG1
cg00915818	4	165875543	-0.0830028	2.68E-06	0.10142522	
cg03742250	17	74550235	-0.1067687	3.41E-06	0.12223888	
cg04305977	16	72211722	0.0851609	3.90E-06	0.13295437	

cg14468055	9	136003437	-0.0607697	4.48E-06	0.14300466	RALGDS
cg24322161	5	56214362	-0.0394573	4.62E-06	0.14300466	SETD9
cg05811998	11	22480945	-0.0529179	5.00E-06	0.14669978	LINC01495
cg23161350	4	66655798	0.05257632	5.17E-06	0.14669978	
cg20260841	4	56420110	0.11393522	5.38E-06	0.14669978	
cg17803298	19	46849276	-0.0171437	5.90E-06	0.15356782	PPP5C
cg25637390	3	66474569	0.08355251	6.13E-06	0.15356782	LRIG1
cg02757152	2	73485819	0.06823909	6.31E-06	0.15356782	FBXO41
cg04645031	6	96913503	0.1199559	6.62E-06	0.1554678	
cg20546265	11	118892438	0.03698768	7.25E-06	0.16269504	TRAPPC4
cg27131663	6	27806196	-0.0418705	7.40E-06	0.16269504	HIST1H2BN; HIST1H2AK
cg14634089	19	36302777	-0.0764273	7.67E-06	0.16333265	PRODH2
cg24431141	2	47126073	-0.0554252	8.60E-06	0.17331235	
cg10862431	12	63544783	0.08354785	8.65E-06	0.17331235	AVPR1A
cg02890529	X	70335118	0.02511697	9.27E-06	0.18044726	
cg13431910	2	127834822	-0.0262944	1.00E-05	0.18306502	BIN1;

5. Discussion

5.1. Refining the definition of frailty in CLL remains essential

In this project, we compared age-associated DNAm patterns in CLL patients with high and low comorbidity to help elucidate the biological basis linking frailty, using comorbidity as a proxy. Our results suggested that age-associated DNAm, as captured by epigenetic clocks, are not influenced by the comorbidity burden status of CLL patients. When we initiated this investigation, we hypothesized that higher comorbidity burden would be associated with accelerated epigenetic aging, based on numerous studies linking epigenetic age acceleration to age-related morbidities and cancer outcomes in population-based cohorts and solid tumor-focused analyses. Although a smaller number of studies have examined hematologic malignancies, mostly acute myeloid leukemia and acute lymphoblastic leukemia, the relationship between epigenetic aging and CLL has not been systematically evaluated^{106,254}.

For instance, a study by Jain et al., demonstrated that higher accelerated aging, quantified by the Horvath clock, is associated with a greater number of chronic diseases, such as diabetes and lung disease, in the whole blood of older women aged 90 and above²³⁵. This relationship is further supported by another study using a diverse cohort, which included both men and women aged 60 and above, as well as studies involving cancer populations. Head and neck cancer patients, ranging from 37 to 96 years of age, were found to exhibit higher age acceleration, quantified by the Horvath, Hannum, and PhenoAge clocks, in their leukocytes that are associated with comorbidities beyond cancer such as cardiovascular disease and diabetes²¹⁷.

Despite similarities between the cited literature and our study, including the epigenetic clocks used to estimate age acceleration and a comparable age range of

CLL patients, we did not observe a robust association between epigenetic age acceleration and comorbidity burden in our cohort. There were some nominal signals, however given documented technical issues. One methodological consideration is sample purity. Technical limitations during bead-based magnetic separation may have reduced the efficiency of B-cell and non-B-cell fractionation in some samples. In particular, the non-B-cell fraction contained residual B cells in varying proportions, which could have introduced additional heterogeneity into the DNAm profiles and influenced DNAm-based clock estimates and their associations with clinical variables. Given that DNAm-based clock outputs and downstream associations can be sensitive to immune cell composition, this difference in sample type could influence epigenetic age estimates and their associations with clinical variables.

Additionally, variation in how comorbidity burden is defined and measured across studies may further contribute to the lack of concordance in observed associations. Jain et al.'s assessment of multimorbidity was based on the number of self-reported chronic conditions within the cohort. The conditions were limited to a set of 12 that are most prevalent among older women, including cardiovascular disease, cancer, cognitive impairment, depression, osteoarthritis, history of falls, chronic obstructive pulmonary disease, hypertension, diabetes, hip fracture, and cerebrovascular disease. Similarly, the study that used diverse cohorts quantifies comorbidities from the self-reports of diabetes, heart disease, stroke, chronic lung disease, and non-skin cancers²³⁵. These conditions were included in a composite measure of multimorbidity that range from 0 to 5. Meanwhile, the authors of the head and neck cancer patient studies used the Charlson comorbidity index, which assessed patients on 17 comorbidity categories with

associated weights based on adjusted mortality risk²¹⁷. The sum of all these weights results in a single comorbidity score. The higher the score, the more comorbidities the patient has and the higher the risk of mortality.

On the other hand, our study used CIRS which quantifies the severity of impairments in each of the 14 organ systems in a five-point scale. Unlike the comorbidity measures of the described literature which focuses only on the presence or absence of select medical diagnosis, CIRS employs a more comprehensive and systematic way of evaluating comorbidity by taking into account various conditions that burden each body system. For example, a patient does not necessarily have to be diagnosed with congestive heart failure to be labelled with comorbidity, but if they present occasional chest pain symptoms, this will be accounted for in CIRS.

The differences in population characteristics and the methods used to assess comorbidities may partially explain why our results did not align with expectations. Another potential factor contributing to the null findings is the composite nature of the CIRS, which underlies the dichotomous classification of comorbidity burden. The CIRS aggregates health ratings across 14 organ systems into a single score, providing a broad, holistic measure of overall health. While this approach captures general comorbidity burden, it may obscure the influence of specific conditions more directly linked to epigenetic aging.

It is widely utilized in clinical settings to guide treatment decisions for CLL patients, with a score of 6 or higher indicating a high comorbidity burden. A key challenge in using CIRS is the difficulty in determining the relative contribution of each

component to the overall score. In the case of the CIRS, the question arises as to which specific body systems contribute most to the overall score.

Given the extensive data encompassed within the CIRS scoring tool, it may be worth performing multiple correlation analyses to identify the most influential components of CIRS in relation to age acceleration. One organ system measured in the CIRS that warrants particular attention is the vascular system, due to its relevance to epigenetic aging. Hypertension, a key condition assessed in the vascular component of the CIRS, is one of the most common comorbidities among CLL patient cohorts, with reported prevalence rates ranging from 40% to 70%^{236,237}. In addition, several studies have linked cardiovascular diseases, including hypertension, to epigenetic age acceleration^{238,239}. As such, future research can benefit from performing an ordinal logistic regression model that compares the vascular system severity scores with EAA, which will help determine if said organ system dysfunction is strongly linked with EAA. Where sufficient data on the assessment of different organ system impairments in CLL patients is available, logistic regression models can be applied.

However, conducting these analyses will require a sufficiently large sample size to ensure adequate statistical power and stable estimation of the relationships between age acceleration and the CIRS components. Unfortunately, comprehensive clinical data required to determine CIRS is not readily accessible. For our study, we have no access to the comprehensive assessment of the patient's comorbidities used to determine the CIRS score, which has its own ethical and legal considerations. Additionally, obtaining DNAm data is resource-intensive from obtaining blood samples from CLL patients, to purifying them to B and non-B cells to generating DNAm data. For these reasons, our

initial sample size was set at 64; however, only 62 samples passed the DNAm data quality checks. Given that several of the correlation analyses yielded non-significant results, it is plausible that the effect of comorbidity burden may be too small to detect with the current sample size, despite the cancer context²⁴⁰. Unless other ways to obtain more CLL patient samples become feasible, we will have to explore other avenues for improving the study.

In addition to the challenges presented by the composite nature of CIRS when it comes to interpreting our project's results, CIRS as a measure of comorbidity burden is not a validated measure of frailty, rather it is a proxy measure of frailty. Frailty, along with age, is a key determinant of patient fitness for certain CLL treatments, such as FCR that may exclude frail patients over 65. Comprehensive geriatric assessments are complex and time-consuming, making CIRS a quicker alternative. Since comorbidity and frailty often co-occur in older patients and both predict treatment outcomes in CLL, comorbidity is used as a proxy for frailty in treatment assessments. If one wishes to explore the relationship between EAA and frailty, rather than comorbidity, using a validated frailty measure such as the frailty index may be of interest like the Fried Frailty Criteria²⁴¹ and FRAIL Scale Score. Recently published papers have reported that higher scores of frailty is associated with increasing age acceleration in older individuals. The relationship between validated frailty scores and age acceleration has not been explored in patients with CLL, presenting an exciting avenue for future research.

5.2. Disrupted epigenetic aging in CLL reflects cancer-driven methylation reprogramming

Differences in epigenetic age and acceleration between B cells and non-B cells in CLL likely reflect a combination of factors intrinsic to the malignant transformation of B cells. Given that CLL is a malignancy of B cells, it is not surprising that the epigenetic clocks based on DNAm exhibit sensitivity to both cell lineage and transformation events. The disruption of normal age-associated DNAm patterns in CLL suggests a complex interplay between oncogenesis and epigenetic aging mechanisms.

Previous research by Lin and Wagner demonstrated that in various cancer types, the correlation between predicted epigenetic age and chronological age often deteriorates²⁴². Most cancers have been associated with accelerated epigenetic aging, especially in tumor tissue such as in the case of colorectal cancer study conducted by Zheng et al. who reported an estimated age of 6-to-24-year acceleration, or the colorectal cancer tumor had an estimated age of 6 to 24 years older than the patients chronological age²⁴³. In Horvath's analysis of almost 6000 pan-solid tumour samples found an average of accelerated aging of 36 years. Despite the general trend, there have been documented exceptions where cancer tissues exhibit negative age acceleration such as reports of significant deceleration or younger DNAm ages in stomach adenocarcinoma and endometrial cancer, compared to its normal counterpart²⁴⁴. Interestingly, our findings in CLL cells align with this observation, had lower age acceleration, with the effect being more pronounced in B cells than in non-B cells. These results suggest that malignant transformation significantly alters the

expected epigenetic aging trajectory, potentially due to extensive DNAm reprogramming in cancerous cells.

Additional insights from Zheng et al. further emphasize the role of epigenetic age discrepancies in cancer prognosis. Their analysis of blood-derived epigenetic age measures revealed that deviations between epigenetic and chronological age are associated with cancer outcomes, though the magnitude and direction of these discrepancies vary depending on tissue context²⁴³. Such findings reinforce the notion that the profound methylome reprogramming occurring during malignant transformation disrupts canonical age-associated DNAm changes. Consequently, epigenetic clocks, which are predominantly trained on normal tissues, may not accurately reflect the epigenetic dynamics of cancer cells, underscoring the need to recalibrate existing models or develop new cancer-specific clocks.

Recent efforts to refine epigenetic clocks for oncological applications have led to the development of models that aim to capture cancer-specific patterns of epigenetic aging. Some of these clocks are designed to detect deviations between epigenetic and chronological age, allowing researchers to distinguish between different tumor phenotypes. For example, the breast cancer-specific epigenetic clock (BEpiC) was proposed as a tool to classify tumor aggressiveness. Although still in the preprint stage and not yet peer-reviewed, BEpiC illustrates the kinds of approaches currently being explored²⁴⁵. Alongside such cancer-specific models, pan-cancer clocks have also been developed to estimate DNA methylation age across a range of tumor types, reflecting a broader effort to understand and characterize epigenetic aging in the context of oncogenesis. Alongside such cancer-specific models, pan-cancer clocks have also

been developed to estimate DNA methylation age across a range of tumor types, reflecting a broader effort to understand and characterize epigenetic aging in the context of oncogenesis.

Despite the research progress involving epigenetic clocks, limitations remain in our current analysis. Specifically, the lack of DNAm data on normal (non-cancerous) B and non-B cells prevents direct comparisons that could strengthen the argument that cancer phenotypes drive the deregulation of age-associated DNAm patterns. However, a published study from our lab that looked into age acceleration as measured by Horvath and Hannum clocks, of healthy adults may provide partial context in the absence of healthy (no CLL diagnosis) controls²⁵². In that study, healthy adult B cells from individuals aged 19–59 years exhibited relatively young epigenetic ages, with age acceleration values ranging from -18 to 7 years across the Horvath, Hannum, and PhenoAge clocks. The negative age acceleration values (e.g. -18) mean that one of the adult participants was predicted to be 18 years younger than their chronological age, which means that that B cells in some of the healthy adult participants are epigenetically younger than their chronological age. In contrast, B cells from our CLL cohort, which ranged in age from 52 to 98 years, showed a much broader distribution of age acceleration values, spanning approximately -35 to 90 years. Although this difference is consistent with greater disruption of epigenetic age regulation relative to healthy adult B cells, the comparison should be interpreted with caution because the healthy reference group was substantially younger than our cohort. As such, these findings provide useful context but do not allow us to fully conclude that the broader range observed in our CLL samples is attributable solely to disease-related biological differences. Nevertheless,

the notable variability observed in CLL B cells further highlight the impact of malignant transformation on epigenetic age regulation and reinforce the necessity for refined epigenetic aging models tailored to cancer biology.

5.3. Reevaluating the link between epigenetic aging and senescence

Direct investigations into the relationship between age-related DNAm changes and cellular senescence remain relatively limited. However, earlier studies point to a potential connection between these processes. Notably, research involving the reprogramming of somatic cells, such as those from mice, into induced pluripotent stem cells demonstrated that both age-associated and senescence-associated DNAm patterns can be reversed. This observation suggests that aging and senescence may be linked through shared epigenetic regulatory mechanisms^{90,246}.

To further explore this relationship, we examined T-cell senescence, as indicated by the expression of CD57⁺ and loss of CD28⁻ expression, in the context of accelerated aging. Contrary to our hypothesis, our findings did not support the notion that senescence accounts for variations in epigenetic age. In other words, age acceleration in T cells is independent of the senescent markers, CD57⁺ and CD28⁻.

Our findings align with previous studies exploring the impact of different senescence inducers on epigenetic aging^{247,248}. One such study evaluated human dermal fibroblasts that were driven into senescence through various mechanisms: DNA damage via X-ray irradiation, oncogene-induced senescence through RAS activation, and replicative senescence via progressive telomere shortening during extended culturing²⁴⁷. The researchers observed that fibroblasts rendered senescent by X-ray or

RAS activation exhibited little to no increase in DNAm age, often measuring close to zero. In contrast, fibroblasts undergoing replicative senescence showed increased epigenetic aging. However, the study ruled out telomere attrition as the underlying cause. When fibroblasts were transduced with the telomerase enzyme, hTERT, to bypass replicative senescence, they still exhibited increases in DNAm age comparable to non-hTERT controls. This strongly suggests that the observed epigenetic aging in replicative senescent cells is not a direct consequence of telomere shortening, but rather a reflection of other aging-related processes occurring during cell proliferation in culture²⁴⁷.

Despite our findings, and those of other studies suggesting that cellular senescence does not directly drive epigenetic aging, it would seem premature to conclude that there is no link between the two. After all, the literature remains mixed, with a few studies suggesting that a more robust association between epigenetic age acceleration and cellular senescence may emerge if alternative senescence biomarkers are utilized²⁴⁸⁻²⁵¹.

While our findings, along with those of previous studies, support the view that senescence does not directly drive epigenetic aging, they also highlight how the link between senescence and accelerated aging is context dependent.

For instance, Smitherman and colleagues conducted a study that investigated frail cancer survivors found that elevated levels of the senescence-associated cell-cycle kinase inhibitor p16^{INK4a} were linked to accelerated aging²⁵¹. Notably, the researchers in this study did not employ DNAm-based methodologies for age estimation; rather, they converted differences in p16^{INK4a} expression into chronological years to conceptualize

the degree of age acceleration. More importantly, all participants were under the age of 29 and had received chemotherapy, a known accelerator of aging, which likely played a central role in the observed association. While we did not include p16^{INK4a} as a marker of senescence, nor were our study samples derived from individuals who had undergone cancer treatment, future research could benefit from incorporating a broader range of senescence markers alongside epigenetic aging measures, while also accounting for the cellular environment, such as exposure to chemotherapy.

Although it would be valuable to conduct further analyses utilizing alternative senescence markers, existing literature also highlights another key hallmark of aging that may more effectively explain the observed DNAm alterations: mitochondrial function. Mitochondrial activity is crucial in regulating metabolic cofactors essential for epigenetic modifications, and its dysfunction, whether through diminished energy production or increased reactive oxygen species (ROS) generation, can disrupt normal epigenetic maintenance mechanisms. Research on cellular senescence has demonstrated a strong association between mitochondrial dysfunction and accelerated aging, underscoring the potential central role of mitochondria in driving age-related changes in DNAm. In a comparative study by Kabacik et al., human dermal keratinocytes were treated with carbonyl cyanide m-chlorophenylhydrazone (CCCP), a chemical compound that disrupts membrane potential and reduces ATP production and were compared to its untreated counterparts²⁴⁷. The results showed that CCCP-treated cells exhibited a significantly accelerated epigenetic aging relative to untreated controls. To further explore the relationship between mitochondrial function and epigenetic aging, an additional experimental group of human dermal cells were treated with Bezafibrate,

which is a pharmacological agent known to promote mitochondrial biogenesis and function²⁴⁷. The findings reveal that Bezafibrate-treated cells exhibited a slower rate of epigenetic aging compared to both untreated and CCCP-treated cells. Collectively, these findings provide compelling empirical support for mitochondrial function as a key regulator of epigenetic aging. Future studies integrating mitochondrial biomarkers alongside DNAm-based age estimation could provide deeper insights into the interplay between metabolic regulation and epigenetic aging in both normal and malignant B and non-B cell contexts.

Although senescence markers (CD57 expression and CD28 loss) did not explain variation in epigenetic age acceleration (EAA) in our data, this finding likely reflects the specific biology captured by the EAA clocks applied. Conventional epigenetic clocks estimate chronological or composite physiological age using selected CpG sets, which may not correspond to the methylation features associated with immunosenescence-related differentiation and immune history. As such, a null association between CD28⁻ and CD57⁺ and EAA does not indicate a lack of substantial DNA methylation remodeling in senescent T-cell states but rather suggests that this senescence-associated variation may not align with the specific CpG features captured by the clocks used here. .

Consistent with this metric-specific interpretation, large numbers of age-associated CpG sites have been reported even within purified naïve CD4⁺ T cells (11,431 sites in adults aged 19–66; 57% hypermethylated with age), indicating age-related methylation changes in T cells that are not even inherently senescent. In addition, epigenetic age estimates can differ markedly across T-cell subsets: applying the Horvath clock, naïve T cells from healthy adults in their 50s–60s were estimated

near 0 epigenetic years, whereas memory CD8⁺ T-cell populations ranged roughly 20–50 years, supporting the idea that DNAm age outputs can also reflect differentiation or immune history instead of just chronological age²⁵⁴.

In this context, the significant association of senescent T-cell proportions with pcgtAge in the non-B fraction is informative rather than contradictory. The pcgtAge score is derived from DNA methylation at PCGT loci and was developed as a mitotic-like measure intended to capture replication-associated methylation drift relevant to turnover and cancer-related epigenetic change²⁴³. Because CD28 loss and CD57 expression often emerge after prolonged antigen exposure and repeated clonal expansion, these phenotypes may align more closely with a replication or turnover-linked methylation axis than with chronological-age clocks²⁵⁶. This pattern is also consistent with the sample composition: the non-B fraction contains the T-cell populations on which senescence markers are defined, whereas the B-cell fraction is enriched for B cells and contains relatively few T cells²⁵⁵. Importantly, the association does not imply a one-directional causal pathway; it is compatible with (i) mixture effects, where shifts toward late-differentiated T-cell states influence the average methylation at PCGT CpGs in bulk non-B profiles, and/or (ii) shared immune-history biology, where chronic stimulation and repeated proliferation increase both senescence-associated phenotypes and replication-linked methylation drift captured by mitotic clocks.

An important direction for future work is to clarify how immune senescence relates to different DNAm-based aging measures in heterogeneous immune compartments, including analyses that separate mixture effects from within-cell

methylation changes and that test whether adjusting for immune composition attenuates the pcgtAge–senescence association.

6. Conclusion

Disentangling the complex interplay between aging, frailty, and their respective roles in shaping the DNAm landscape presents a significant challenge. Although these processes are conceptually distinct, they are biologically and clinically interrelated in ways that complicate efforts to isolate their individual contributions. Aging is not a uniform or linear process; it may culminate in either healthy longevity or pathological decline. However, differentiating between “normal” and pathological aging remains a central challenge in the biology of aging.

CLL provides a particularly relevant model for investigating these relationships. Situated at the intersection of immunosenescence, chronic disease, and age-related physiological decline, CLL reflects many of the complexities inherent to aging itself. This makes it a valuable system for studying how malignancy interacts with age-associated biological processes.

In this study, we aimed to contribute to the broader effort to delineate the roles of aging and frailty in epigenetic remodeling, with a specific focus on DNA methylation dynamics. Our findings indicate that while cellular senescence contributes to epigenomic changes, a substantial proportion of age-related DNAm alterations occur independently of senescent cell accumulation. This suggests that senescence is not the sole driver of epigenetic aging and highlights the multifactorial nature of age-associated methylation patterns.

Notably, we observed that even non-malignant cells in CLL patients exhibit DNAm profiles influenced by cancer-associated processes. This observation underscores the extent to which malignancy can distort the molecular hallmarks of

aging, complicating the interpretation of DNAm as a biomarker for biological age in disease contexts. These findings support the view that CLL can disrupt the normative aging trajectory at the molecular level.

Contrary to our initial hypothesis, DNAm patterns did not reliably distinguish between major prognostic subtypes in CLL. This result suggests that, in certain clinical contexts, DNAm-based biomarkers may have limited prognostic utility. Additionally, we detected minimal evidence for differential epigenetic age acceleration among CLL patients, reinforcing the notion that frailty is tightly integrated within the aging process and may not be easily parsed as a separate biological entity using current epigenetic tools and methodologies.

Collectively, these findings refine our understanding of the utility and limitations of DNAm as a biomarker of aging in the context of chronic disease. They also highlight key areas where further research is needed to improve the resolution and interpretability of epigenetic aging measures.

In summary, this work underscores the complexity of aging biology, emphasizes the need for integrative and disease-aware analytical frameworks, and provides novel insights into the interplay between aging, frailty, and malignancy. By clarifying the contexts in which DNAm is a meaningful indicator of biological aging, this study contributes to the ongoing effort to develop robust, clinically relevant measures of aging and disease risk.

7. References

1. Eichhorst, B. *et al.* Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **32**, 23–33 (2021).
2. Yang, J. *et al.* Clinical Outcomes in a Large Canadian Centralized CLL Clinic Based on Treatment and Molecular Factors over a Decade. *Curr. Oncol.* **30**, 6411–6431 (2023).
3. Huang, S. J. *et al.* Impact of age and treatment institution type on outcomes of patients treated for chronic lymphocytic leukemia in British Columbia, Canada. *Leuk. Res.* **103**, 106538 (2021).
4. Kipps, T. J. *et al.* Chronic lymphocytic leukaemia. *Nat. Rev. Dis. Primer* **3**, 16096 (2017).
5. Mukkamalla, S. K. R., Taneja, A., Malipeddi, D. & Master, S. R. Chronic Lymphocytic Leukemia. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).
6. Gan, T., Li, B. E., Mishra, B. P., Jones, K. L. & Ernst, P. MLL1 Promotes IL-7 Responsiveness and Survival during B Cell Differentiation. *J. Immunol.* **200**, 1682–1691 (2018).
7. Pieper, K., Grimbacher, B. & Eibel, H. B-cell biology and development. *J. Allergy Clin. Immunol.* **131**, 959–971 (2013).
8. Janeway, C. A., Travers, P., Walport, M. & Shlomchik, M. J. The rearrangement of antigen-receptor gene segments controls lymphocyte development. in

Immunobiology: The Immune System in Health and Disease. 5th edition (Garland Science, 2001).

9. Fuentes-Pananá, E. M., Bannish, G. & Monroe, J. G. Basal B-cell receptor signaling in B lymphocytes: mechanisms of regulation and role in positive selection, differentiation, and peripheral survival. *Immunol. Rev.* **197**, 26–40 (2004).
10. Batten, M. *et al.* Baff Mediates Survival of Peripheral Immature B Lymphocytes. *J. Exp. Med.* **192**, 1453–1466 (2000).
11. Levine, M. H. *et al.* A B-cell receptor-specific selection step governs immature to mature B cell differentiation. *Proc. Natl. Acad. Sci.* **97**, 2743–2748 (2000).
12. Harris, R. S., Kong, Q. & Maizels, N. Somatic hypermutation and the three R's: repair, replication and recombination¹. *Mutat. Res. Mutat. Res.* **436**, 157–178 (1999).
13. Kwak, K. *et al.* Intrinsic properties of human germinal center B cells set antigen affinity thresholds. *Sci. Immunol.* **3**, eaau6598 (2018).
14. Hussmann, D. *et al.* IGHV-associated methylation signatures more accurately predict clinical outcomes of chronic lymphocytic leukemia patients than IGHV mutation load. *Haematologica* **107**, 877–886 (2022).
15. Hallek, M. Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* **92**, 946–965 (2017).
16. Cantú, E. S. *et al.* Male-To-Female Sex Ratios of Abnormalities Detected by Fluorescence in Situ Hybridization in a Population of Chronic Lymphocytic

- Leukemia Patients. *Hematol. Rep.* **5**, 13 (2013).
17. Vardell, V. A. *et al.* Influence of racial and ethnic identity on overall survival in patients with chronic lymphocytic leukemia. *Am. J. Hematol.* **98**, E172–E174 (2023).
 18. Goldin, L. R. *et al.* Common occurrence of monoclonal B-cell lymphocytosis among members of high-risk CLL families. *Br. J. Haematol.* **151**, 152–158 (2010).
 19. Holl, E. K. *et al.* Examining Peripheral and Tumor Cellular Immunome in Patients With Cancer. *Front. Immunol.* **10**, 1767 (2019).
 20. Koffman, B. Diagnosis of Chronic Lymphocytic Leukemia. *CLL Society* <https://cllsociety.org/2016/03/diagnosis-chronic-lymphocytic-leukemia/> (2016).
 21. Rawstron, A. C. *et al.* Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytometry B Clin. Cytom.* **94**, 121 (2018).
 22. Strati, P. & Shanafelt, T. D. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood* **126**, 454–462 (2015).
 23. American Cancer Society. If You Have Chronic Lymphocytic Leukemia (CLL). <https://www.cancer.org/cancer/types/chronic-lymphocytic-leukemia/if-you-have-chronic-lymphocytic-leukemia.html> (2018).
 24. Canadian Cancer Society / Société canadienne du. What is chronic lymphocytic

leukemia? *Canadian Cancer Society* <https://cancer.ca/en/cancer-information/cancer-types/chronic-lymphocytic-leukemia-cll/what-is-chronic-lymphocytic-leukemia>.

25. Marrero, Y. T. *et al.* Immunophenotypic Characterization by Flow Cytometry of Chronic Lymphoid Leukemia. *J. Cancer Immunol.* **Volume 3**, 196–205 (2021).
26. Zhang, S. & Kipps, T. J. The Pathogenesis of Chronic Lymphocytic Leukemia. *Annu. Rev. Pathol.* **9**, 103 (2013).
27. Minden, M. D. *et al.* Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. *Nature* **489**, 309–312 (2012).
28. Quinten, E. *et al.* Autonomous B-cell receptor signaling and genetic aberrations in chronic lymphocytic leukemia-phenotype monoclonal B lymphocytosis in siblings of patients with chronic lymphocytic leukemia. *Haematologica* **109**, 824 (2023).
29. Sepulveda Yanez, J. H. *et al.* Autonomous BCR Signaling and Genetic Aberrations in CLL-Phenotype Monoclonal B Lymphocytosis in Siblings of CLL Patients. *Blood* **140**, 1526–1527 (2022).
30. Peters, F. S., Strefford, J. C., Eldering, E. & Kater, A. P. T-cell dysfunction in chronic lymphocytic leukemia from an epigenetic perspective. *Haematologica* **106**, 1234–1243 (2021).
31. Dubois, N. *et al.* Importance of Crosstalk Between Chronic Lymphocytic Leukemia Cells and the Stromal Microenvironment: Direct Contact, Soluble Factors, and Extracellular Vesicles. *Front. Oncol.* **10**, (2020).

32. Svanberg, R., Janum, S., Patten, P. E., Ramsay, A. G. & Niemann, C. U. Targeting the tumor microenvironment in chronic lymphocytic leukemia. *Haematologica* **106**, 2312 (2021).
33. García-Muñoz, R. & Llorente, L. Chronic lymphocytic leukaemia: could immunological tolerance mechanisms be the origin of lymphoid neoplasms? *Immunology* **142**, 536 (2014).
34. Chavez, J. C., Sahakian, E. & Pinilla-Ibarz, J. Ibrutinib: an evidence-based review of its potential in the treatment of advanced chronic lymphocytic leukemia. *Core Evid.* **8**, 37 (2013).
35. Ramsay, A. G. *et al.* Chronic lymphocytic leukemia T cells show impaired immunological synapse formation that can be reversed with an immunomodulating drug. *J. Clin. Invest.* **118**, 2427–2437 (2008).
36. Vlachonikola, E., Stamatopoulos, K. & Chatzidimitriou, A. T Cells in Chronic Lymphocytic Leukemia: A Two-Edged Sword. *Front. Immunol.* **11**, (2021).
37. Roessner, P. M. & Seiffert, M. T-cells in chronic lymphocytic leukemia: Guardians or drivers of disease? *Leukemia* **34**, 2012–2024 (2020).
38. Taghiloo, S. *et al.* Frequency and functional characterization of exhausted CD8⁺ T cells in chronic lymphocytic leukemia. *Eur. J. Haematol.* **98**, 622–631 (2017).
39. Blank, C. U. *et al.* Defining ‘T cell exhaustion’. *Nat. Rev. Immunol.* **19**, 665 (2019).
40. Wright, K. B. *et al.* Caring for an Individual with Chronic Lymphocytic Leukemia (CLL): Understanding Family Caregivers’ Perceptions of Social Support, Caregiver

- Burden, and Unmet Support Needs. *J. Cancer Educ. Off. J. Am. Assoc. Cancer Educ.* **39**, 180–185 (2024).
41. Hampel, P. J. & Parikh, S. A. Chronic lymphocytic leukemia treatment algorithm 2022. *Blood Cancer J.* **12**, 161 (2022).
 42. Wierda, W. G. *et al.* Multivariable Model for Time to First Treatment in Patients With Chronic Lymphocytic Leukemia. *J. Clin. Oncol.* **29**, 4088 (2011).
 43. Crombie, J. & Davids, M. S. IGHV Mutational Status Testing in Chronic Lymphocytic Leukemia. *Am. J. Hematol.* **92**, 1393 (2017).
 44. Rotbain, E. C. *et al.* The CLL comorbidity index in a population-based cohort: a tool for clinical care and research. *Blood Adv.* **6**, 2701–2706 (2022).
 45. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. *Blood* **94**, 1848–1854 (1999).
 46. Rozovski, U., Keating, M. J. & Estrov, Z. Why Is the Immunoglobulin Heavy Chain Gene Mutation Status a Prognostic Indicator in Chronic Lymphocytic Leukemia? *Acta Haematol.* **140**, 51–54 (2018).
 47. Zenz, T. *et al.* TP53 Mutation and Survival in Chronic Lymphocytic Leukemia. *J. Clin. Oncol.* **28**, 4473–4479 (2010).
 48. Ozaki, T. & Nakagawara, A. Role of p53 in Cell Death and Human Cancers. *Cancers* **3**, 994–1013 (2011).
 49. Vaddavalli, P. L. & Schumacher, B. The p53 network: cellular and systemic DNA

- damage responses in cancer and aging. *Trends Genet.* **38**, 598–612 (2022).
50. Campo, E. *et al.* TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. *Haematologica* **103**, 1956 (2018).
 51. Ibrahim, S. *et al.* CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. *Blood* **98**, 181–186 (2001).
 52. Del Poeta, G. *et al.* Clinical significance of CD38 expression in chronic lymphocytic leukemia. *Blood* **98**, 2633–2639 (2001).
 53. Burgler, S. Role of CD38 Expression in Diagnosis and Pathogenesis of Chronic Lymphocytic Leukemia and Its Potential as Therapeutic Target. *Crit. Rev. Immunol.* **35**, 417–432 (2015).
 54. Malavasi, F. *et al.* CD38 and chronic lymphocytic leukemia: a decade later. *Blood* **118**, 3470–3478 (2011).
 55. Wang, H. *et al.* ZAP-70: An Essential Kinase in T-cell Signaling. *Cold Spring Harb. Perspect. Biol.* **2**, a002279 (2010).
 56. Chen, J. *et al.* ZAP-70 constitutively regulates gene expression and protein synthesis in chronic lymphocytic leukemia. *Blood* **137**, 3629–3640 (2021).
 57. Deaglio, S. *et al.* CD38 and ZAP-70 are functionally linked and mark CLL cells with high migratory potential. *Blood* **110**, 4012–4021 (2007).
 58. Basabaeen, A. A. *et al.* Combined analysis of ZAP-70 and CD38 expression in sudanese patients with B-cell chronic lymphocytic leukemia. *BMC Res. Notes* **12**, 282 (2019).

59. D'Arena, G. *et al.* Prognostic significance of combined analysis of ZAP-70 and CD38 in chronic lymphocytic leukemia. *Am. J. Hematol.* **82**, 787–791 (2007).
60. Kutsch, N., Fink, A. M. & Fischer, K. Management of front line chronic lymphocytic leukemia. *Am. J. Hematol.* **97**, S3–S10 (2022).
61. Hus, I. *et al.* The clinical significance of ZAP-70 and CD38 expression in B-cell chronic lymphocytic leukaemia. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **17**, 683–690 (2006).
62. Wang, Y. *et al.* Cause of death in patients with newly diagnosed chronic lymphocytic leukemia (CLL) stratified by the CLL-International Prognostic Index (CLL-IPI). *J. Clin. Oncol.* **38**, 8026–8026 (2020).
63. Noto, A. *et al.* Should treatment of hypogammaglobulinemia with immunoglobulin replacement therapy (IgRT) become standard of care in patients with chronic lymphocytic leukemia? *Front. Immunol.* **14**, 1062376 (2023).
64. Hamblin, A. D. & Hamblin, T. J. The immunodeficiency of chronic lymphocytic leukaemia. *Br. Med. Bull.* **87**, 49–62 (2008).
65. Dearden, C. Disease-Specific Complications of Chronic Lymphocytic Leukemia. *Hematology* **2008**, 450–456 (2008).
66. Hilal, T., Gea-Banacloche, J. C. & Leis, J. F. Chronic lymphocytic leukemia and infection risk in the era of targeted therapies: Linking mechanisms with infections. *Blood Rev.* **32**, 387–399 (2018).
67. Murru, R. *et al.* Prediction of severe infections in chronic lymphocytic leukemia: a

- simple risk score to stratify patients at diagnosis. *Ann. Hematol.* **103**, 1655 (2024).
68. Pasiarski, M. *et al.* Antibody and Plasmablast Response to 13-Valent Pneumococcal Conjugate Vaccine in Chronic Lymphocytic Leukemia Patients – Preliminary Report. *PLOS ONE* **9**, e114966 (2014).
69. Anand, K., Cingam, S. & Peddi, P. Recurrent Malignant Melanoma Presenting as Isolated Pleural Metastases in a Patient with Chronic Lymphocytic Leukemia. *Case Rep. Oncol.* **10**, 86–90 (2017).
70. Landego, I. *et al.* PD-1 Inhibition in Malignant Melanoma and Lack of Clinical Response in Chronic Lymphocytic Leukemia in the Same Patients: A Case Series. *Curr. Oncol.* **27**, 169–172 (2020).
71. Chatzikonstantinou, T. *et al.* Other malignancies in the history of CLL: an international multicenter study conducted by ERIC, the European Research Initiative on CLL, in HARMONY. *eClinicalMedicine* **65**, (2023).
72. Popovic, L. *et al.* Concurrent Chronic Lymphocytic Leukemia and Merkel Cell Carcinoma in Primary Skin Tumor and Metastatic Lymph Node. *Indian J. Hematol. Blood Transfus.* **30**, 422–424 (2014).
73. Brewer, J. D. *et al.* Chronic Lymphocytic Leukemia Is Associated With Decreased Survival of Patients With Malignant Melanoma and Merkel Cell Carcinoma in a SEER Population-Based Study. *J. Clin. Oncol.* (2012)
doi:10.1200/JCO.2011.34.9605.
74. Dziechciaż, M. & Filip, R. Biological psychological and social determinants of old

- age: Bio-psycho-social aspects of human aging. *Ann Agric Env. Med* **21**, 835–838 (2014).
75. Fedarko, N. S. The Biology of Aging and Frailty. *Clin. Geriatr. Med.* **27**, 27–37 (2011).
 76. Ferrucci, L. *et al.* Measuring biological aging in humans: A quest. *Aging Cell* **19**, e13080 (2020).
 77. Lipsitz, L. A. Physiological Complexity, Aging, and the Path to Frailty. *Sci. Aging Knowl. Environ.* **2004**, pe16–pe16 (2004).
 78. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153**, 1194–1217 (2013).
 79. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).
 80. Jylhävä, J., Pedersen, N. L. & Hägg, S. Biological Age Predictors. *eBioMedicine* **21**, 29–36 (2017).
 81. Dodig, S., Čepelak, I. & Pavić, I. Hallmarks of senescence and aging. *Biochem. Medica* **29**, 030501 (2019).
 82. Song, P., An, J. & Zou, M.-H. Immune Clearance of Senescent Cells to Combat Ageing and Chronic Diseases. *Cells* **9**, 671 (2020).
 83. Hu, L. *et al.* Why Senescent Cells Are Resistant to Apoptosis: An Insight for Senolytic Development. *Front. Cell Dev. Biol.* **10**, 822816 (2022).
 84. Kale, A., Sharma, A., Stolzing, A., Desprez, P.-Y. & Campisi, J. Role of immune

- cells in the removal of deleterious senescent cells. *Immun. Ageing* **17**, 16 (2020).
85. Ryu, S. J., Oh, Y. S. & Park, S. C. Failure of stress-induced downregulation of Bcl-2 contributes to apoptosis resistance in senescent human diploid fibroblasts. *Cell Death Differ.* **14**, 1020–1028 (2007).
86. Muñoz, D. P. *et al.* Targetable mechanisms driving immunoevasion of persistent senescent cells link chemotherapy-resistant cancer to aging. *JCI Insight* **4**, e124716 (2019).
87. Sidler, C., Woycicki, R., Kovalchuk, I. & Kovalchuk, O. WI-38 senescence is associated with global and site-specific hypomethylation. *Aging* **6**, 564 (2014).
88. Belsky, D. W. *et al.* Quantification of biological aging in young adults. *Proc. Natl. Acad. Sci.* **112**, E4104–E4110 (2015).
89. Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844–1859 (2016).
90. Wagner, W. The Link Between Epigenetic Clocks for Aging and Senescence. *Front. Genet.* **10**, 303 (2019).
91. Weng, N., Akbar, A. N. & Goronzy, J. CD28– T cells: their role in the age-associated decline of immune function. *Trends Immunol.* **30**, 306–312 (2009).
92. Effros, R. B. Loss of CD28 expression on T lymphocytes: A marker of replicative senescence. *Dev. Comp. Immunol.* **21**, 471–478 (1997).
93. Czesnikiewicz-Guzik, M. *et al.* T cell subset-specific susceptibility to aging. *Clin. Immunol.* **127**, 107–118 (2008).

94. Pangrazzi, L. & Weinberger, B. T cells, aging and senescence. *Exp. Gerontol.* **134**, 110887 (2020).
95. Kared, H., Martelli, S., Ng, T. P., Pender, S. L. & Larbi, A. CD57 in human natural killer cells and T-lymphocytes. *Cancer Immunol. Immunother. CII* **65**, 441 (2016).
96. Brenchley, J. M. *et al.* Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8⁺T cells. *Blood* **101**, 2711–2720 (2003).
97. Merino, J. *et al.* Progressive decrease of CD8^{high}+ CD28⁺ CD57[–] cells with ageing. *Clin. Exp. Immunol.* **112**, 48–51 (1998).
98. Chu, W., Chang, S.-F. & Ho, H.-Y. Adverse Health Effects of Frailty: Systematic Review and Meta-Analysis of Middle-Aged and Older Adults With Implications for Evidence-Based Practice. *Worldviews Evid. Based Nurs.* **18**, 282–289 (2021).
99. Hao, Q. *et al.* The role of frailty in predicting mortality and readmission in older adults in acute care wards: a prospective study. *Sci. Rep.* **9**, 1207 (2019).
100. Fan, L. *et al.* Frailty Predicts Increased Health Care Utilization Among Community-Dwelling Older Adults: A Longitudinal Study in China. *J. Am. Med. Dir. Assoc.* **22**, 1819–1824 (2021).
101. Cesari, M., Gambassi, G., Abellan van Kan, G. & Vellas, B. The frailty phenotype and the frailty index: different instruments for different purposes. *Age Ageing* **43**, 10–12 (2014).
102. Hafizoğlu, M. *et al.* Which comorbidity index is more appropriate for geriatric patients from the frailty perspective? *Eur. Geriatr. Med.* **15**, 115–125 (2024).

103. Ekdahl, A. W. *et al.* Frailty and comprehensive geriatric assessment organized as CGA-ward or CGA-consult for older adult patients in the acute care setting: A systematic review and meta-analysis. *Eur. Geriatr. Med.* **6**, 523–540 (2015).
104. Rockwood, K. *et al.* A global clinical measure of fitness and frailty in elderly people. *CMAJ Can. Med. Assoc. J.* **173**, 489–495 (2005).
105. Fried, L. P. *et al.* Frailty in Older Adults: Evidence for a Phenotype. *J. Gerontol. Ser. A* **56**, M146–M157 (2001).
106. Taborda, C. C. *et al.* Fried's Frailty Phenotype Predicts Mortality and Survival for Newly Diagnosed Older Patients with Acute Myeloid Leukemia or High-Risk Myelodysplastic Syndrome. *Blood* **140**, 10825–10827 (2022).
107. Pamoukdjian, F. *et al.* Frailty Parameters, Morbidity and Mortality in Older Adults with Cancer: A Structural Equation Modelling Approach Based on the Fried Phenotype. *J. Clin. Med.* **9**, 1826 (2020).
108. Bieniek, J., Wilczyński, K. & Szewieczek, J. Fried frailty phenotype assessment components as applied to geriatric inpatients. *Clin. Interv. Aging* **11**, 453–459 (2016).
109. Pandey, Y. *et al.* Evaluation of Agreement Among Frailty Assessment Tools in Head and Neck Surgery. *Otolaryngol. Neck Surg.* **168**, 32–38 (2023).
110. Pearce, J. *et al.* Frailty and treatment outcome in advanced gastro-oesophageal cancer: An exploratory analysis of the GO2 trial. *J. Geriatr. Oncol.* **13**, 287–293 (2022).

111. Schoenborn, N. L., Blackford, A. L., Joshu, C. E., Boyd, C. M. & Varadhan, R. Life expectancy estimates based on comorbidities and frailty to inform preventive care. *J. Am. Geriatr. Soc.* **70**, 99–109 (2022).
112. Huang, F. *et al.* Development and validation of a predictive risk model for frailty in elderly patients with multimorbidity. *Geriatr. Gerontol. Int.* **22**, 471–476 (2022).
113. Pérez Fernández, M. *et al.* Comorbidity, Frailty, and Waitlist Mortality among Kidney Transplant Candidates of All Ages. *Am. J. Nephrol.* **49**, 103–110 (2019).
114. Villavicencio, A. *et al.* Comorbidities at Diagnosis, Survival, and Cause of Death in Patients with Chronic Lymphocytic Leukemia: A Population-Based Study. *Int. J. Environ. Res. Public. Health* **18**, 701 (2021).
115. Hudon, C., Fortin, M. & Soubhi, H. Abbreviated guidelines for scoring the Cumulative Illness Rating Scale (CIRS) in family practice. *J. Clin. Epidemiol.* **60**, 212.e1-212.e4 (2007).
116. Osborn IV, K. P. *et al.* Cumulative Illness Rating Scale (CIRS) can be Used to Predict Hospital Outcomes in Older Adults. *2017* **3**,.
117. Gordon, M. J. *et al.* The chronic lymphocytic leukemia comorbidity index (CLL-CI): a three-factor comorbidity model. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **27**, 4814–4824 (2021).
118. Gordon, M. J. *et al.* Comorbidities predict inferior outcomes in chronic lymphocytic leukemia treated with ibrutinib. *Cancer* **124**, 3192–3200 (2018).
119. Obrtlíkova, P. *et al.* Fludarabine, Cyclophosphamide and Rituximab (FCR) Related

- Prolonged Cytopenia Is Frequent and Adverse Factor Affecting Survival of Patients with Chronic Lymphocytic Leukemia (CLL). *Blood* **120**, 1790 (2012).
120. Hill, B. T. & Kalaycio, M. Profile of obinutuzumab for the treatment of patients with previously untreated chronic lymphocytic leukemia. *OncoTargets Ther.* **8**, 2391–2397 (2015).
121. Długosz-Danecka, M. *et al.* Efficacy and safety of the obinutuzumab-chlorambucil combination in the frontline treatment of elderly CLL patients with comorbidities - Polish Adult Leukemia Group (PALG) real-life analysis. *Pol. Arch. Intern. Med.* **128**, 421–426 (2018).
122. Kretzmer, H. *et al.* Preneoplastic Alterations Define CLL DNA Methylome and Persist through Disease Progression and Therapy. *Blood Cancer Discov.* **2**, 54–69 (2021).
123. Tsagiopoulou, M. *et al.* DNA methylation profiles in chronic lymphocytic leukemia patients treated with chemoimmunotherapy. *Clin. Epigenetics* **11**, 177 (2019).
124. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38**, 23–38 (2013).
125. Jin, B., Li, Y. & Robertson, K. D. DNA Methylation. *Genes Cancer* **2**, 607–617 (2011).
126. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
127. Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and Non-CpG Methylation in

- Epigenetic Gene Regulation and Brain Function. *Genes* **8**, 148 (2017).
128. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3740 (2002).
129. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1412 (2006).
130. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
131. Kribelbauer, J. F. *et al.* Quantitative Analysis of the DNA Methylation Sensitivity of Transcription Factor Complexes. *Cell Rep.* **19**, 2383–2395 (2017).
132. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
133. Woo, H. R., Dittmer, T. A. & Richards, E. J. Three SRA-Domain Methylcytosine-Binding Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in Arabidopsis. *PLoS Genet.* **4**, e1000156 (2008).
134. Yang, X. *et al.* Gene Body Methylation can alter Gene Expression and is a Therapeutic Target in Cancer. *Cancer Cell* **26**, 577 (2014).
135. Ball, M. P. *et al.* Targeted and genome-scale methylomics reveals gene body signatures in human cell lines. *Nat. Biotechnol.* **27**, 361 (2009).
136. Rauluseviciute, I., Drabløs, F. & Rye, M. B. DNA hypermethylation associated with

- upregulated gene expression in prostate cancer demonstrates the diversity of epigenetic regulation. *BMC Med. Genomics* **13**, 6 (2020).
137. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes.
 138. Maunakea, A. K. *et al.* Conserved Role of Intragenic DNA Methylation in Regulating Alternative Promoters. *Nature* **466**, 253 (2010).
 139. Xie, W. *et al.* Base-resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* **148**, 816 (2012).
 140. Yousefi, P. *et al.* Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics* **16**, 911 (2015).
 141. Grant, O. A., Wang, Y., Kumari, M., Zabet, N. R. & Schalkwyk, L. Characterising sex differences of autosomal DNA methylation in whole blood using the Illumina EPIC array. *Clin. Epigenetics* **14**, 1–16 (2022).
 142. Shen, E. *et al.* Epigenetics and sex differences in the brain: A genome-wide comparison of histone-3 lysine-4 trimethylation (H3K4me3) in male and female mice. *Exp. Neurol.* **268**, (2015).
 143. Hartman, R. J. G., Huisman, S. E. & den Ruijter, H. M. Sex differences in cardiovascular epigenetics—a systematic review. *Biol. Sex Differ.* **9**, 19 (2018).
 144. Elliott, H. R. *et al.* Characterisation of ethnic differences in DNA methylation between UK-resident South Asians and Europeans. *Clin. Epigenetics* **14**, 130 (2022).

145. Galanter, J. M. *et al.* Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *eLife* **6**, e20532 (2017).
146. Hing, B. *et al.* Chronic social stress induces DNA methylation changes at an evolutionary conserved intergenic region in chromosome X. *Epigenetics* **13**, 627 (2018).
147. Fujimura, J. H. & Rajagopalan, R. Different differences: The use of 'genetic ancestry' versus race in biomedical human genetic research. *Soc. Stud. Sci.* **41**, 5 (2011).
148. Maugeri, A. & Barchitta, M. How Dietary Factors Affect DNA Methylation: Lesson from Epidemiological Studies. *Medicina (Mex.)* **56**, 374 (2020).
149. Barchitta, M. *et al.* Dietary Patterns are Associated with Leukocyte LINE-1 Methylation in Women: A Cross-Sectional Study in Southern Italy. *Nutrients* **11**, 1843 (2019).
150. Zhang, F. F. *et al.* Dietary Patterns Are Associated with Levels of Global Genomic DNA Methylation in a Cancer-Free Population. *J. Nutr.* **141**, 1165 (2011).
151. Opsasnick, L. A. *et al.* Epigenome-wide association study of long-term psychosocial stress in older adults. *Epigenetics* **19**, 2323907 (2024).
152. Elkin, E. R., Higgins, C., Aung, M. T. & Bakulski, K. M. Metals exposures and DNA methylation: Current evidence and future directions. *Curr. Environ. Health Rep.* **9**, 673 (2022).
153. Stepanyan, A. *et al.* Long-term environmental metal exposure is associated with

- hypomethylation of CpG sites in NFKB1 and other genes related to oncogenesis. *Clin. Epigenetics* **15**, 126 (2023).
154. Cerutti, J., Lussier, A. A., Zhu, Y., Liu, J. & Dunn, E. C. Associations between indicators of socioeconomic position and DNA methylation: a scoping review. *Clin. Epigenetics* **13**, 221 (2021).
 155. Reuben, A. *et al.* Association of Neighborhood Disadvantage in Childhood With DNA Methylation in Young Adulthood. *JAMA Netw. Open* **3**, e206095 (2020).
 156. Martin, C. L. *et al.* Neighborhood environment, social cohesion, and epigenetic aging. *Aging* **13**, 7883–7899 (2021).
 157. Smith, J. A. *et al.* Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: The Multi-Ethnic Study of Atherosclerosis. *Epigenetics* **12**, 662–673 (2017).
 158. Houseman, E. A., Kim, S., Kelsey, K. T. & Wiencke, J. K. DNA Methylation in Whole Blood: Uses and Challenges. *Curr. Environ. Health Rep.* **2**, 145–154 (2015).
 159. Tesfaye, M. *et al.* Sex effects on DNA methylation affect discovery in epigenome-wide association study of schizophrenia. *Mol. Psychiatry* **29**, 2467–2477 (2024).
 160. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22**, 246–258 (2012).
 161. Szigeti, K. A. *et al.* Global DNA hypomethylation of colorectal tumours detected in tissue and liquid biopsies may be related to decreased methyl-donor content. *BMC*

- Cancer* **22**, 605 (2022).
162. Bedford, M. T. & van Helden, P. D. Hypomethylation of DNA in pathological conditions of the human prostate. *Cancer Res.* **47**, 5274–5276 (1987).
163. Gama-Sosa, M. A. *et al.* The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Res.* **11**, 6883 (1983).
164. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1**, 239 (2009).
165. Ng, J. M.-K. & Yu, J. Promoter Hypermethylation of Tumour Suppressor Genes as Potential Biomarkers in Colorectal Cancer. *Int. J. Mol. Sci.* **16**, 2472 (2015).
166. Rauscher, G. H. *et al.* Exploring DNA methylation changes in promoter, intragenic, and intergenic regions as early and late events in breast cancer formation. *BMC Cancer* **15**, 816 (2015).
167. Park, S. Y. *et al.* Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch.* **458**, 73–84 (2011).
168. Colicino, E. *et al.* Blood DNA methylation sites predict death risk in a longitudinal study of 12,300 individuals. *Aging* **12**, 14092 (2020).
169. Jelinek, J. *et al.* Aberrant DNA Methylation Is Associated with Disease Progression, Resistance to Imatinib and Shortened Survival in Chronic Myelogenous Leukemia. *PLoS ONE* **6**, e22110 (2011).
170. Lin, Y.-L. *et al.* Aberrant Protocadherin17 (PCDH17) Methylation in Serum is a Potential Predictor for Recurrence of Early-Stage Prostate Cancer Patients After Radical Prostatectomy. *Med. Sci. Monit.* **21**, 3955–3690 (2015).

171. Li, P., Liu, J., Li, J. & Liu, P. DNA methylation of CRB3 is a prognostic biomarker in clear cell renal cell carcinoma. *Mol. Biol. Rep.* **46**, 4377–4383 (2019).
172. Luo, Z.-G., Li, Z.-G., Gui, S.-L., Chi, B.-J. & Ma, J.-G. Protocadherin-17 promoter methylation in serum-derived DNA is associated with poor prognosis of bladder cancer. *J. Int. Med. Res.* **42**, 35–41 (2014).
173. Campagna, M. P. *et al.* Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin. Epigenetics* **13**, 214 (2021).
174. Tanić, M. Chapter Twenty-Nine - Epigenome-wide association study (EWAS): Methods and applications. in *Epigenetics Methods* (ed. Tollefsbol, T.) vol. 18 591–613 (Academic Press, 2020).
175. Flanagan, J. M. Epigenome-Wide Association Studies (EWAS): Past, Present, and Future. in *Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis* (ed. Verma, M.) 51–63 (Springer, New York, NY, 2015). doi:10.1007/978-1-4939-1804-1_3.
176. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).
177. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81–86 (2017).
178. Markunas, C. A. *et al.* Identification of DNA Methylation Changes in Newborns Related to Maternal Smoking during Pregnancy. *Environ. Health Perspect.* **122**, 1147–1153 (2014).

179. Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in Rheumatoid Arthritis. *Nat. Biotechnol.* **31**, 142 (2013).
180. Wei, S. *et al.* Ten Years of EWAS. *Adv. Sci.* **8**, 2100727 (2021).
181. Breitling, L. P., Yang, R., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-Smoking-Related Differential DNA Methylation: 27K Discovery and Replication. *Am. J. Hum. Genet.* **88**, 450 (2011).
182. Barcelona, V. *et al.* Novel DNA methylation sites associated with cigarette smoking among African Americans. *Epigenetics* **14**, 383 (2019).
183. Gao, X., Jia, M., Zhang, Y., Breitling, L. P. & Brenner, H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* **7**, 113 (2015).
184. Bojesen, S. E., Timpson, N., Relton, C., Smith, G. D. & Nordestgaard, B. G. AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality. *Thorax* **72**, 646–653 (2017).
185. Roubroeks, J. A. *et al.* An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiol. Aging* **95**, 26 (2020).
186. Johansson, A. & Flanagan, J. M. Epigenome-wide association studies for breast cancer risk and risk factors. *Trends Cancer Res.* **12**, 19 (2017).
187. Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven

- new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508 (2010).
188. Wang, T. *et al.* Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk. *Clin. Epigenetics* **12**, 5 (2020).
189. Patel, C. J., Bhattacharya, J. & Butte, A. J. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLoS ONE* **5**, e10746 (2010).
190. Adalsteinsson, B. T. *et al.* Heterogeneity in White Blood Cells Has Potential to Confound DNA Methylation Measurements. *PLOS ONE* **7**, e46705 (2012).
191. Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* **19**, 64 (2018).
192. Teschendorff, A. E. & Zheng, S. C. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* **9**, 757–768 (2017).
193. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15**, R31 (2014).
194. Teschendorff, A. E., Zhu, T., Breeze, C. E. & Beck, S. EPISCORE: cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data. *Genome Biol.* **21**, 221 (2020).
195. Salas, L. A. *et al.* Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat. Commun.* **13**, 761 (2022).
196. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing

- epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
197. Kapourani, C.-A., Argelaguet, R., Sanguinetti, G. & Vallejos, C. A. scMET: Bayesian modeling of DNA methylation heterogeneity at single-cell resolution. *Genome Biol.* **22**, 114 (2021).
198. Bauer, M. *et al.* A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood. *Clin. Epigenetics* **7**, 81 (2015).
199. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10604–10609 (2005).
200. Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10522–10527 (2012).
201. Lee, T. I. *et al.* Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* **125**, 301 (2006).
202. Ciccarone, F. *et al.* Age-dependent expression of DNMT1 and DNMT3B in PBMCs from a large European population enrolled in the MARK-AGE study. *Aging Cell* **15**, 755 (2016).
203. Cole, J. J. *et al.* Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol.* **18**, 58 (2017).
204. Xiao, Y. *et al.* Age and gender affect DNMT3a and DNMT3b expression in human liver. *Cell Biol. Toxicol.* **24**, 265–272 (2008).

205. Teschendorff, A. E., West, J. & Beck, S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum. Mol. Genet.* **22**, R7–R15 (2013).
206. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
207. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **49**, 359–367 (2013).
208. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
209. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
210. Petersen, C. L., Christensen, B. C. & Batsis, J. A. Weight management intervention identifies association of decreased DNA methylation age with improved functional age measures in older adults with obesity. *Clin. Epigenetics* **13**, 46 (2021).
211. Vetter, V. M. *et al.* DNA methylation age acceleration is associated with risk of diabetes complications. *Commun. Med.* **3**, 1–10 (2023).
212. Peterson, M. D., Collins, S., Meier, H. C. S., Brahmsteadt, A. & Faul, J. D. Grip strength is inversely associated with DNA methylation age acceleration. *J. Cachexia Sarcopenia Muscle* **14**, 108–115 (2023).
213. Yang, Z. *et al.* Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* **17**, 205 (2016).
214. Faul, J. D. *et al.* Epigenetic-based age acceleration in a representative sample of

- older Americans: Associations with aging-related morbidity and mortality. *Proc. Natl. Acad. Sci.* **120**, e2215840120 (2023).
215. Rozenblit, M. *et al.* Evidence of accelerated epigenetic aging of breast tissues in patients with breast cancer is driven by CpGs associated with polycomb-related genes. *Clin. Epigenetics* **14**, 30 (2022).
216. Kresovich, J. K. *et al.* Methylation-Based Biological Age and Breast Cancer Risk. *JNCI J. Natl. Cancer Inst.* **111**, 1051 (2019).
217. Xiao, C. *et al.* Epigenetic age acceleration, fatigue, and inflammation in patients undergoing radiation therapy for head and neck cancer: A longitudinal study. *Cancer* **127**, 3361–3371 (2021).
218. Kim, Y. *et al.* Higher diet quality relates to decelerated epigenetic aging. *Am. J. Clin. Nutr.* **115**, 163–170 (2022).
219. Fiorito, G. *et al.* Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. *Aging* **11**, 2045 (2019).
220. Jain, P. *et al.* Analysis of Epigenetic Age Acceleration and Healthy Longevity Among Older US Women. *JAMA Netw. Open* **5**, e2223285 (2022).
221. Galow, A.-M. & Peleg, S. How to Slow down the Ticking Clock: Age-Associated Epigenetic Alterations and Related Interventions to Extend Life Span. *Cells* **11**, 468 (2022).
222. Field, A. E. *et al.* DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol. Cell* **71**, 882–895 (2018).

223. Choi, H., Joe, S. & Nam, H. Development of Tissue-Specific Age Predictors Using DNA Methylation Data. *Genes* **10**, 888 (2019).
224. Chung, M. *et al.* DNA methylation ageing clocks and pancreatic cancer risk: pooled analysis of three prospective nested case-control studies. *Epigenetics* **16**, 1306–1316 (2021).
225. Durso, D. F. *et al.* Acceleration of leukocytes' epigenetic age as an early tumor and sex-specific marker of breast and colorectal cancer. *Oncotarget* **8**, 23237–23245 (2017).
226. Kular, L. *et al.* Epigenetic clock indicates accelerated aging in glial cells of progressive multiple sclerosis patients. *Front. Aging Neurosci.* **14**, 926468 (2022).
227. Tang, R. *et al.* Adverse childhood experiences, DNA methylation age acceleration, and cortisol in UK children: a prospective population-based cohort study. *Clin. Epigenetics* **12**, 55 (2020).
228. Kim, K. *et al.* Association of Adverse Childhood Experiences With Accelerated Epigenetic Aging in Midlife. *JAMA Netw. Open* **6**, e2317987 (2023).
229. Joshi, D., Gonzalez, A., Lin, D. & Raina, P. The association between adverse childhood experiences and epigenetic age acceleration in the Canadian longitudinal study on aging (CLSA). *Aging Cell* **22**, e13779 (2023).
230. Foster, C. A. *et al.* Epigenetic age acceleration correlates with BMI in young adults. *Aging* **15**, 513 (2023).
231. Widayati, T. A. *et al.* Open access-enabled evaluation of epigenetic age

- acceleration in colorectal cancer and development of a classifier with diagnostic potential. *Front. Genet.* **14**, (2023).
232. Teschendorff, A. E. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. *Genome Med.* **12**, 56 (2020).
233. Ambatipudi, S., Inchanalkar, M. & Mahimkar, M. B. Role of DNA methylation-based mitotic ageing indices in oral cancer development and recurrence. *Oral Dis.* **30**, 2929–2939 (2024).
234. Liu, W. *et al.* Senescent Tumor CD8⁺ T Cells: Mechanisms of Induction and Challenges to Immunotherapy. *Cancers* **12**, 2828 (2020).
235. Jain P., Binder A., Chen B. *et al.* The association of epigenetic age acceleration and multimorbidity in older women: a study of the Women's Health Initiative. *J. Gerontol. A Biol. Sci. Med. Sci.* **77**, 1239-1249 (2022).
236. Pottinger T.D., Khan S.S., Zheng Y. *et al.* Association of cardiovascular health and epigenetic age acceleration. *Clin. Epigenetics* **13**, 42 (2021).
237. Ferrajoli A., Follows G., Marmor Y. *et al.* Cumulative review of hypertension in patients with chronic lymphocytic leukemia (CLL) and other hematologic malignancies treated with acalabrutinib. *Blood* **142** (Suppl. 1), 1917 (2023).
238. Joyce B.T., Carnethon M.R., Guan W. *et al.* Epigenetic age acceleration reflects long-term cardiovascular health. *Circ. Res.* **129**, 770-781 (2021).
239. Roetker N.S., Guan W., Xu X. *et al.* Prospective study of epigenetic age acceleration and incidence of cardiovascular disease outcomes in the ARIC

- Study. *Circ. Genom. Precis. Med.* **11**, e001937 (2018).
240. Deering K. L., Sundaram M., Harshaw Q. et al. Health-related quality of life and treatment satisfaction in Chronic Lymphocytic Leukaemia (CLL) patients on ibrutinib compared to other CLL treatments in a real-world US cross-sectional study. *PLoS ONE* **17**, e0270291 (2022).
241. Carneiro T.X., Sales J.L.S., Tavares J.S., Wanderley I.J.M., Santos J.K.D., Pimentel L.D.S., Moraes N.S. Frailty assessment in patients with chronic lymphocytic leukaemia treated with the combination of venetoclax and ibrutinib. *Hematol. Transfus. Cell Ther.* **45**, 101636 (2023).
242. Lin Q. & Wagner W. Epigenetic ageing signatures are coherently modified in cancer. *PLoS Genet.* **11**, e1005334 (2015).
243. Zheng C. et al. Association of epigenetic clock with consensus molecular subtypes and overall survival of colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* **28**, 1345-1354 (2019).
244. Hao J., Liu T., Xiu Y., Yuan H. & Xu D. High DNA methylation age deceleration defines an aggressive phenotype with immuno-exclusion environments in endometrial carcinoma. *Front. Immunol.* **14**, 1208223 (2023).
245. Shang S., Gao X., Li J. et al. Epigenetic clock detected a breast cancer mitosis subtype with improved immunotherapy. *NPJ Breast Cancer* **7**, 58 (2021).
246. Boland M.J., Hazen J.L., Nazor K.L. et al. Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91–94 (2009).

247. Kabacik S., Lowe D., Fransen L., Leonard M., Ang S.-L., Whiteman C., Corsi S., Cohen H., Felton S., Bali R., Horvath S., Raj K. The relationship between epigenetic age and the hallmarks of ageing in human cells. *Nature Aging* **2**, 484–493 (2022).
248. Lowe, D., Horvath, S. & Raj, K. Epigenetic clock analyses of cellular senescence and ageing. *Oncotarget* **7**, 8524–8531 (2016).
249. Hillary R. F., Stevenson A. J., Cox S. R. et al. An epigenetic predictor of death captures multi-modal measures of ageing. *bioRxiv* pre-print (2019) — published version: *Aging Cell* **18**, e12877 (2019).
250. Qin N., Burkhardt S. C., Hudson M. M. et al. Epigenetic Age Acceleration and Chronic Health Conditions Among Adult Survivors of Childhood Cancer. *J. Natl. Cancer Inst.* **113**, 1485–1492 (2021).
251. Smitherman, A. B. et al. Accelerated aging among childhood, adolescent, and young adult cancer survivors is evidenced by increased expression of p16INK4a and frailty. *Cancer* **126**, 4975–4983 (2020).
252. Jones, M.J., Konwar, C., Asimwe, R. et al. DNA methylation differences between cord and adult white blood cells reflect postnatal immune cell maturation. *Commun Biol* **8**, 237 (2025).
253. Dozmorov, M. G., Coit, P., Maksimowicz-McKinnon, K. & Sawalha, A. H. Age-associated DNA methylation changes in naive CD4+ T cells suggest an evolving autoimmune epigenotype in aging T cells. *Epigenomics* **9**, 429–445 (2017).

254. Mi, T. et al. Conserved epigenetic hallmarks of T cell aging during immunity and malignancy. *Nat. Aging* **4**, 1053–1063 (2024).
255. Tomusiak, A. et al. Development of an epigenetic clock resistant to changes in immune cell composition. *Commun. Biol.* **7**, 934 (2024).
256. Strioga, M. et al. CD8⁺ CD28[–] and CD8⁺ CD57⁺ T cells and their role in health and disease. *Immunology* **134**, 17–32 (2011).
257. Bergsma, T. & Rogaeva, E. DNA methylation clocks and their predictive capacity for aging phenotypes and healthspan. *Neurosci. Insights* **15**, 2633105520942221 (2020).
258. El Khoury, L. Y. et al. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biol.* **20**, 283 (2019).
259. McEwen, L. M. et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc. Natl Acad. Sci. USA* **117**, 23329–23335 (2020).
260. Oakes, C. C. et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* **48**, 253–264 (2016).
261. Queirós, A. C. et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia* **29**, 598–605 (2015).
262. Dietrich, A. et al. omnideconv: a unifying framework for using and benchmarking single-cell-informed deconvolution of bulk RNA-seq data. *Genome Biol.* **27**, 6 (2026).
263. Nguyen, H. et al. Fourteen years of cellular deconvolution:

methodology, applications, technical evaluation and outstanding challenges.

Nucleic Acids Res. **52**, 4761–4783 (2024).

263. Gribben, J. G. Chronic lymphocytic leukemia: planning for an aging population.

Expert Rev. Anticancer Ther. 10, 1389–1394 (2010).