DECIPHERING EPIGENETIC SIGNATURES OF AGING AND NEURODEGENERATION IN ALZHEIMER'S DISEASE WITH MACHINE LEARNING

Lasya P. Sreepada

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2025

Supervisor of Dissertation	Co-Supervisor of Dissertation
Corey T. McMillan	David A. Wolk
Associate Professor of Neurology	Professor of Neurology

Graduate Group Chairperson

Yale E. Cohen, Gabriel Tucker Professor of Otorhinolaryngology

Dissertation Committee

Paul A. Yushkevich, Professor of Radiology
Russell T. Shinohara, Professor of Biostatistics
Wanding Zhou, Assistant Professor of Pathology, Children's Hospital of Philadelphia

DECIPHERING EPIGENETIC SIGNATURES OF AGING AND NEURODEGENERATION IN ALZHEIMER'S DISEASE WITH MACHINE LEARNING

This dissertation is dedicated to the little girls and young women who dream big, defy limits, and dare to discover the unknown—science needs you.

ACKNOWLEDGEMENT

This dissertation was made possible by the support and contributions of my advisors, thesis committee, research collaborators, lab mates, department, university administration, friends & mentors, and most importantly, my family and God. Thank you all for supporting my dream.

First and foremost, I am deeply grateful to my PhD advisors, Dr. Corey McMillan and Dr. David Wolk, for taking a chance on me and welcoming me into their labs. Their mentorship shaped every stage of this dissertation and they always encouraged me to push the boundaries of my knowledge. Corey and Dave make the best dynamic duo and their witty humor and storytelling made even the hardest days lighter.

Next, I would like to thank my thesis committee, Profs. Paul Yushkevich, Taki Shinohara, and Wanding Zhou, who have been instrumental in shaping my dissertation and research direction. Thanks to Wanding for his leadership in the Methylation Working Group and his thoughtful insights on my work. Taki's course on *Statistical Methods in Neuroimaging* was the first class I attended at Penn and became a foundation for my work. Equally instrumental was Paul's course on *Biomedical Image Analysis*. Thanks to Taki for his enthusiasm and insight during project discussions. I am grateful to Paul for having me on board as a Teaching Assistant in his class and helping me pursue the Graduate Teaching Certificate under his mentorship.

I thank the research participants, their families, and caregivers for making my work possible. This research was supported by funding from the National Institutes of Health T32 Fellowship (AG076411) and the Penn Alzheimer's Disease Research Center, or ADRC (AG072979). Through the T32, I gained mentorship from Dr. John Detre, who shared key insights on neuroimaging and the ethics of translational research. I also participated in the Christopher Clark Scholars Program at the ADRC, through which I gained mentorship from Dr. Jason Karlawish. I learned a lot in his seminal class, *The Problem of Alzheimer's Disease*. I thank the broader ADRC and Frontotemporal Degeneration Center (FTDC) for providing an unparalleled environment for me to grow as a scientist.

During my PhD, I was fortunate to work across two exceptional research groups: the Penn Bioinformatics in Neurodegenerative Disease (BiND) Lab led by Corey, and the Penn ADRD Trajectories, Copathologies, and Heterogeneity (PATCH) Lab, affectionately known as the "HippoGang", led by Dave, Paul, and Dr. Sandhitsu (Sandy) Das. I am grateful for the friendly, collaborative lab environment where I was encouraged to share my work and receive feedback. Thanks to Drs. Rory Boyle, Nadia Dehghani, David Goldberg, and Anil Wadhwani for their collaboration on my work, and to Sandy and Ms. Emily McGrew for their guidance with neuroimaging analysis. Thanks also to Drs. Christopher Brown, Lyles Clark, Katie Cousins, and Melanie Matyi for their guidance and their kindness. Thanks to Dr. Christos Davatzikos for giving me the opportunity to come to Penn and mentoring me in my initial years. From his lab, I am grateful to Drs. Sindhuja Tirumalai Govindarajan and Mathilde Antoniades for always lending a kind ear, sharing resources, and encouraging me.

I'm proud to be part of the Bioengineering Department, which is *objectively* the best at Penn as declared by our graduate group chair, Dr. Yale Cohen. Through my leadership in the Graduate Association of Bioengineers (GABE), I've had the privilege of working closely with Yale, whose mentorship has been invaluable since Day 1. I thank Dean Boon Thau Loo for his advocacy on behalf of students and for the opportunity to serve on the Dean's Doctoral Advisory Board. I am grateful for the bridge funding from Penn Engineering during my lab transition, as well as continued guidance from administrators, Ms. Kathleen Venit, Ms. Tori Frew, and Ms. Alyse Edwards. Thanks to Drs. Susan Davidson, Sharath Chandra Guntuku, and Ms. Staci Kaplan from the Data Science program for their guidance during my Master's practicum, which led to my participation in the Women in Data Science Conference and opened valuable opportunities.

My heartfelt thanks to Dr. Alexander Lin at Harvard Medical School, who inspired me to pursue my PhD. As a confused college sophomore uncertain about coursework and career paths, I found clarity and inspiration in Alex's lab. His enthusiasm for research and deep commitment to mentoring students motivated me to become a scientist and mentor future students like myself. Alex always treated me as a valued colleague, encouraged my ideas, and fostered curiosity and confidence. I'm

deeply indebted to Alex for his belief in me. I also thank Dr. Katie Breedlove for her collaboration during my summers in the lab, and for her continued support throughout my PhD.

As I reflect on my personal journey, I'm grateful to those who have stood beside me over the past 6 years. To my dear friends, Drs. Apurva Singh and Madhura Nijsure, thanks for your unwavering support as we achieved this milestone together. Thanks to Melanie Hilman for indulging my nostalgia for our hometown of Boston and for being a great coworking buddy. Thanks to Ludwig Zhao, the best GABE co-president, who is always up for an adventure and never fails to make me laugh. Thanks to Joanna Fisch and Coach Joe Feeney from Penn Women's Ice Hockey for sharing your love of the game and taking interest in my research. Finally, thanks to Dr. Pulkit Khandelwal for inspiring me with his love of science over chai sessions and walks through Rittenhouse.

I'm grateful to have found community in weekly *pujas* and meditation led by Umesh Uncle and the Omkar Foundation. I appreciate the Chaplain's Office for providing a welcoming space for students of all faiths to come together. Thanks to Dr. Murali Balaji at Penn, his wife, Ms. Devi Ramkissoon, and their 5 year old son Siddarth, for welcoming me to their home like family. I've cherished playing, coloring, and learning all about insects and dinosaurs with Siddarth on random Friday afternoons. A special thank you to Drs. Asha Shipman and Gurpreet Kaur, whose wisdom and spiritual care have been a guiding light for many years.

I'm grateful to have formed friendships beyond Penn that truly made Philadelphia feel like home. Thank you to Suvarna Murali for her energy and contagious love of Philly (Go Birds!). Thanks to Kayleigh Sprague and Simmi Kalsi for their thoughtfulness and memorable walks around the city. Finally, thanks to Allie Diamond, Fran Hackett, and Zero for their warmth and kindness.

In the Spring of 2020, just 6 months into starting my PhD in a new city, the world shut down. During this difficult and isolating time, my school and college friends became an anchor. Through phone calls and game nights on Discord, they brought joy and connection when it was most needed. Thanks to Anna Blech, Sara Tabin, David Rubio, and "El Cuerpo": Jake Gertie, Sarah Wagner, Eugine Szeto, and Parthib Samadder, for staying connected and cheering me on. To my dear friends

Shelly Mishra and Brooke Yang, thanks for standing by me since 6th grade through many seasons of life. I'm grateful for the many phone calls, texts, and virtual work sessions over the years.

Heartfelt thanks to Dr. Shawn Kenner (DK), who is a tireless advocate for women in STEM and my first computer programming (and AP Chemistry) teacher in high school. Although I didn't quite excel in AP Chem, I earned my stripes as a reliable lab minion. I've come a long way from prepping chemicals and cleaning glassware, but I will always be grateful to DK for helping me find my way in science. I've also come to appreciate that communication is one of the most important skills in becoming a great scientist. Special thanks to Mr. Glenn Shiebler, my high school Humanities and English teacher, whose legendary writing advice has stayed with me and proved invaluable as I wrote my dissertation. Shiebs, thanks for encouraging us not only to keep writing, but to use our voices for social good and to speak up when it truly matters.

This journey would not have been possible without the love, support, and blessings of my family. Thanks to my *mamayyas* (uncles), Bhanu and Radha Kuchibhotla for always encouraging me. I am especially grateful for my loving sister Sri Sreepada (aka Tham, Goose, Frog, and various other nicknames I've come up with). She has been a tremendous source of support, the best FaceTime work buddy, and the only person I can be truly silly and goofy with. I carry with me the blessings and wisdom of my grandparents: *Tatagaru* and *Mamma* (S. Sitaram and S. Suryaprabhavathi), and *Thathayya* and *Ammamma* (K. Venkateswarlu and K. Seetha Mahalakshmi). Their immense love, prayers, and pride in my work have been pillars of support throughout my life. Above all, I am forever grateful and indebted to my parents, *Amma* and *Nanna* (Durga and Venkata Sreepada), for their sacrifices, wisdom, boundless love, and belief in my dreams even when the path was uncertain. Whenever I felt like I couldn't keep going, they dropped everything to drive 8 hours each way to Philly, to spend just 24 hours with me before traveling home ahead of a busy workweek. They revived me with home-cooked food and the biggest bear hugs. My family has been my greatest source of strength and energy, and this PhD is as much their achievement as it is mine. Finally, I thank the Divine for giving me the opportunity to complete my PhD at this prestigious university with courage, resilience, and integrity.

ABSTRACT

DECIPHERING EPIGENETIC SIGNATURES OF AGING AND NEURODEGENERATION IN ALZHEIMER'S DISEASE WITH MACHINE LEARNING

Lasya P. Sreepada

Corey T. McMillan

David A. Wolk

Aging of the human brain is marked by considerable heterogeneity, with specific regions such as the medial temporal lobe (MTL) exhibiting increased vulnerability to neurodegeneration in disorders like Alzheimer's disease (AD). Emerging evidence suggests that epigenetic mechanisms, particularly DNA methylation, may contribute to this regional susceptibility and offer potential as peripheral biomarkers of brain aging. In this dissertation, we investigate the hypothesis that blood-based epigenetic modifications reflect and predict patterns of brain aging and neurodegeneration, as measured by *in vivo* structural MRI.

To address this, we pursue two aims using a combination of statistical modeling and machine learning (ML) techniques. In **Aim 1**, we assess whether biological aging—quantified via epigenetic clocks applied to blood-derived DNA methylation—explains individual differences in cortical and MTL neurodegeneration beyond chronological age. We introduce a novel imaging-derived biomarker, the Cortico-Medial Temporal (CoMeT) index, to capture regional atrophy along the cortical-MTL axis. Associations between CoMeT and epigenetic age acceleration are evaluated through hypothesis testing and subgroup analyses, revealing that faster biological aging correlates with greater vulnerability in AD-sensitive brain regions.

In **Aim 2**, we develop predictive models of brain age and hippocampal volume from genome-wide methylation data using Elastic Net regression. Within a nested cross-validation framework, we compare feature selection strategies and identify robust epigenetic signatures of neurodegeneration. These features are annotated and interpreted using gene ontology and enrichment analyses. We

further assess the convergence of methylation markers across independent datasets and examine their correspondence between blood and brain, as well as associations with blood gene expression. Results indicate partial overlap in peripheral and central signatures, suggesting that while blood methylation captures biologically relevant information, it incompletely reflects brain-specific processes.

Together, these studies demonstrate that epigenetic markers are associated with regional patterns of brain aging and neurodegeneration, and that blood-based methylation profiles offer promising, albeit limited, surrogates for brain imaging phenotypes. This work provides new insights into the molecular correlates of neurodegeneration and highlights the potential of epigenetics for advancing early detection and mechanistic understanding of age-related cognitive decline.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 Biological Background

1.1.1 Aging

Aging is a universal biological process characterized by the gradual decline of physiological systems and the reduced capacity to maintain homeostasis. At the cellular level, aging is characterized by progressive decline in physiological integrity, leading to senescence and vulnerability to death (López-Otín et al., 2013). Cellular senescence refers to a state of cell cycle arrest accompanied by phenotypic changes that accumulate in tissues over time, contributing to age-related tissue dysfunction and chronic inflammation (Campisi and d'Adda di Fagagna, 2007). Genetic and epigenetic variations impacting cellular and molecular pathways play a central role in modulating the pace and presentation of aging. Pathways involved in DNA repair, mitochondrial function, and stress resistance have been shown to influence life- and and health-span across diverse organisms. Moreover, age-associated epigenetic modifiers, including DNA methylation changes and histone modifications, regulate gene expression and in turn drive cellular aging and tissue dysfunction (Liu et al., 2020).

At a macroscopic scale, cellular and molecular processes manifest as coordinated changes across multiple organ systems, resulting in increased susceptibility to disease, functional impairment, and mortality (Gladyshev, 2016). For example, in the cardiovascular system, aging leads to arterial stiffening, endothelial dysfunction, and impaired cardiac output, contributing to the elevated incidence of hypertension, atherosclerosis, and heart failure in older adults (Vakka et al., 2023). The immune system also undergoes profound changes that impair adaptive immune responses and promote chronic low-grade inflammation, which is implicated in a range of age-related diseases (Liu et al., 2023). In the musculoskeletal system, aging is associated with loss of muscle mass and strength and reductions in bone mineral density, leading to frailty and an increased risk of falls and fractures. These systemic changes collectively diminish resilience to stressors and reduce the capacity for tissue repair and regeneration (Leser et al., 2021). Aging thus acts as a primary driver

of many life-threatening conditions, including cancer, cardiovascular disease, metabolic dysfunction, and neurodegenerative diseases (López-Otín et al., 2013). Among these, Alzheimer's disease (AD) is one of the most prevalent and devastating neurodegenerative disorders, affecting over 30 million individuals worldwide and representing a leading cause of morbidity and mortality in older adults (Gustavsson et al., 2023). The incidence of AD rises sharply with age, underscoring the profound interplay between aging processes and neurodegeneration. Emerging research suggests that genetic and epigenetic factors not only modulate susceptibility to AD, but also shape its progression by influencing neuroinflammation and neuropathological pathways such as protein aggregation (Smith et al., 2021).

1.1.2 Brain Aging

The brain is particularly vulnerable to the effects of aging. Over the lifespan, a wide range of structural, functional, and molecular changes occur that influence cognitive performance and increase the risk of neurodegenerative disease (Gonzales et al., 2022). Even in the context of healthy aging, individuals commonly experience subtle but progressive declines in cognitive domains such as memory, processing speed, and executive function. These changes are thought to reflect a combination of synaptic loss, reductions in grey and white matter volume, and alterations in functional brain networks (Jost and Kujach, 2025). Despite these changes, many adults maintain relatively stable cognitive function into late life, suggesting individual variability in aging trajectories.

On the other hand, brain aging also sets the stage for pathological processes that underlie neurodegeneration. For example, biological processes such as the accumulation of oxidative stress, chronic inflammation, and mitochondrial dysfunction contribute to neuronal vulnerability and eventually, apoptosis. These mechanisms are especially pertinent to the onset and progression of AD and related dementias (Lin and Beal, 2006; Wang et al., 2025). AD is the most common cause of dementia worldwide and is characterized by the progressive accumulation of amyloid-β plaques and tau neurofibrillary tangles, accompanied by widespread synaptic and neuronal loss. These pathologies begin years, if not decades, before clinical symptoms emerge, highlighting the need for early detection and prevention strategies (Jack Jr et al., 2018). The convergence of age-related physiological decline

with disease onset that occurs long before our ability to detect it clinically underscores the importance of studying brain aging not only as an accumulation of damage, but also as an active biological process with identifiable and potentially modifiable mechanisms. To this end, many groups have developed methods to capture and quantify various aspects of brain aging, ranging from structural and functional imaging to molecular and cellular biomarkers (Komleva et al., 2025). Quantitative measures are essential for capturing the continuous changes that characterize brain aging, enabling earlier detection and more precise characterization of dynamic age-related processes.

1.1.3 Heterogeneity of Aging and Implications for Quantifying Brain Age

Aging is not a uniform process; rather, it is highly individualized, shaped by a combination of genetic, environmental, and lifestyle factors. One of the most intuitive illustrations of this heterogeneity lies in the human face. With the exception of identical twins, each person's face is uniquely identifiable, and these differences persist and evolve with age. Observable features such as skin texture, pigmentation, wrinkle distribution, facial contour, and other dermal markers offer visible cues to an individual's age. Yet these features do not change uniformly across individuals. One person may exhibit early facial aging marked by deep wrinkles and pigmentation changes, while another of the same chronological age may retain a youthful appearance (Guyuron et al., 2009). As such, while it is often possible to approximate someone's age based on their facial features, these estimates are inherently imprecise and reflect variability in how aging manifests externally.

This variability has inspired the development of computational, data-driven approaches that can estimate age from various data modalities, including facial images, with increasing accuracy. Section 1.2 provides a detailed treatment of the foundations and applications of modern engineering methods that can estimate biological measures of age, including brain age. Importantly, the increasing interest in these approaches underscores a broader point: individuals of the same chronological age can exhibit strikingly different biological profiles, physical capabilities, and levels of independence. For example, some older individuals remain fully active and cognitively unimpaired, while others require substantial assistance with activities of daily living, whether due to typical aging or underlying neuropathological processes. This divergence highlights the critical difference between

chronological age, or the amount of time lived, and biological age, which reflects the cumulative physiological burden experienced by an individual. This variability complicates efforts to define "normal" brain aging and encourages precision medicine approaches to diagnosis and treatment of neurological disorders.

1.1.4 Neuroimaging as a Window into the Heterogeneity of Brain Aging and AD

Understanding heterogeneity in aging is particularly crucial when considering AD, and there is growing recognition that AD is both clinically and biologically heterogeneous. There is substantial variability not only in the clinical presentation but also in temporal onset, anatomical spread, and underlying pathology of AD. Historically, the gold standard for AD diagnosis relied on *post-mortem* histopathological analysis and quantification of amyloid plaques and neurofibrillary tangles in brain tissue. These pathological hallmarks defined the disease and anchored definitive diagnosis for decades. However, clinical diagnosis had limited accuracy in prediction of pathology, particularly in early disease stages, hindering the development of interventions targeting these early disease stages.

The advent of *in vivo* neuroimaging, beginning in the mid-20th century, has transformed our understanding of the aging brain and provided essential tools to assess the heterogeneity inherent in both normal cognitive aging and AD. Structural magnetic resonance imaging (MRI) enabled detailed visualization of brain anatomy, allowing researchers to track patterns of atrophy across aging and disease states. Later, the development of positron emission tomography (PET) imaging facilitated the direct measurement of pathological proteins such as amyloidβ and tau, thus bridging the gap between clinical presentation and neuropathology. Together, these imaging modalities offered an unprecedented opportunity to observe the living brain and assess the presence and distribution of AD pathology even before the onset of clinical symptoms. This has enabled newfound insights into the spatial and temporal dynamics of AD pathology in the brain during life, facilitating a more nuanced understanding of the disease's progression and heterogeneity across individuals.

This new paradigm was formalized in the A/T/N biomarker framework proposed by Jack and colleagues in Jack Jr et al. (2018), which used biomarker-based criteria, specifically amyloid (A),

tau (T), and neurodegeneration (N), to define the AD continuum *in vivo*. The A/T/N biomarkers now serve as the foundation for research definitions of AD, moving diagnosis toward an *ante mortem*, biologically grounded model. The guidelines were updated in 2024 (Jack et al., 2024), reaffirming and expanding this framework to incorporate newer imaging tracers and integrate longitudinal assessments.

Longitudinal neuroimaging studies have shown that the regional buildup of amyloid and tau precedes and predicts downstream neurodegeneration, which can be visualized on an MRI as cortical thinning or regional volume loss. Typical AD is characterized by widespread amyloid accumulation and more regionally focal tau deposition, particularly within the medial temporal lobe (MTL) or "limbic" regions, including the entorhinal cortex (ERC) and hippocampus. This distribution aligns with the hallmark symptom of memory impairment, as these regions are essential for episodic memory encoding and consolidation.

1.1.5 Atypical Presentations of AD

Neuroimaging has revealed considerable anatomical and functional variability among individuals with AD, particularly in atypical cases. Structural MRI and PET frequently identify regional patterns of neurodegeneration and pathology that deviate from the canonical MTL-dominant trajectory. For instance, occipital-predominant atrophy and lateral temporal degeneration are common neuroimaging features of two established atypical clinical syndromes associated with AD pathology. These patterns often emerge early and persist longitudinally, suggesting that heterogeneity is not merely a byproduct of disease stage but reflects fundamental differences in disease biology. Yet, interpreting such imaging data remains complex due to overlapping pathologies, inter-individual anatomical variability, and limited resolution of current biomarkers in distinguishing mixed disease states. This leaves us with an incomplete understanding of why certain brain regions are more susceptible in some individuals than others. Integrating neuroimaging findings with multimodal data is essential for explaining the biological mechanisms driving AD heterogeneity.

While the aforementioned changes contribute towards a common endpoint in AD, the pathways to this endpoint are diverse. The observation of biomarker accumulation outside typical regions un-

derscores the mechanistic heterogeneity of AD and suggests that multiple biological subtypes may exist within the AD spectrum. For example, amyloid and tau pathology may be concentrated in lateral temporal, parietal, or frontal regions, sometimes in the absence of significant MTL involvement. These "atypical" patterns (i.e., cortical-predominant or MTL-predominant) may correspond to distinct clinical phenotypes, such as posterior cortical atrophy, or suggest the presence of co-pathologies like limbic-predominant age-related TDP-43 encephalopathy (LATE). By enabling the direct measurement of disease processes across time and individuals, *in vivo* neuroimaging offers a powerful tool for characterizing heterogeneity in AD.

1.1.6 Genetic and Epigenetic Explanations of Heterogeneity in Brain Aging and AD

While neuroimaging provides a powerful means to visualize and quantify variability in brain structure, pathology, and function, it does not, on its own, explain why individuals diverge so markedly in their aging trajectories or responses to neuropathological burden. A growing body of research points to a constellation of biological and environmental factors that also interact in complex, dynamic ways, as contributors to heterogeneity in brain aging and AD.

Genetic variability plays a foundational role in shaping individual susceptibility to aging-related processes and neurodegenerative disease. For instance, the APOE $\varepsilon 4$ allele is the strongest known genetic risk factor for sporadic AD, influencing both the likelihood and age of onset. Beyond APOE, genome-wide association studies (GWAS) have identified numerous loci linked to AD risk, However, the effect sizes of individual variants are typically modest, and the interaction between genetic background and other risk modifiers contributes substantially to the observed variability in disease expression.

With the completion of the Human Genome Project in the early 21st century and the recognition that the genome sequence is quite similar across humans and even species, scientists are now focused on so-called "post-genome" era studies of gene regulation and expression in bulk tissue and single cells. Epigenomics, literally meaning "above the genome", refers to the study of heritable modifications in gene expression that do not involve changes to the DNA sequence itself. These modifications, including DNA methylation, histone modifications, and non-coding RNA activity, play

essential roles in regulating gene function and cellular identity. Epigenomics presents a particularly compelling area of study due to its unique intersection between genetic regulation, environmental influence, and disease susceptibility. Unlike the static nature of the genome, epigenetic patterns are dynamic, modifiable across the lifespan, and responsive to a broad range of physical, biological, and psychosocial stimuli.

However, a significant challenge in studying brain aging with epigenetics lies in the inaccessibility of brain tissue throughout the human lifespan. The brain cannot be biopsied or monitored longitudinally in a minimally invasive manner, limiting opportunities to investigate dynamic molecular changes *in vivo*. In contrast, peripheral tissues such as blood can be collected repeatedly, safely, and at relatively low cost. These practical advantages have motivated the use of peripheral epigenetic data as proxies for brain aging processes. While epigenetic signatures are known to be highly tissue- and cell-specific, reflecting the roles of epigenetic regulation in cellular differentiation, emerging research has identified patterns of concordance between blood and brain methylation profiles (C. Silva et al., 2022). Notably, several studies have demonstrated that peripheral epigenetic markers can, to some extent, capture age-related molecular changes relevant to the central nervous system, supporting their utility as surrogate indicators of brain health (Milicic et al., 2022, 2023).

This growing body of evidence suggests that peripheral epigenetic data may offer a valuable, non-invasive window into brain-related aging processes. Ongoing studies are refining our understanding of cross-tissue methylation patterns, with a focus on identifying specific loci and pathways that show consistent regulation across blood and brain. As a result, there is increasing interest in integrating epigenomic data into studies of neurodegeneration and dementia to improve early detection, stratify individual risk, and elucidate molecular mechanisms that underlie the clinical and biological heterogeneity of brain aging and AD.

1.1.7 Engineering Approaches to Estimate Biological Age

In the past decade, there has been increasing interest to develop advanced engineering methods to leverage large-scale multimodal data to predict biological measures of age. Among these are "epigenetic clocks", which are machine learning (ML) algorithms that have emerged as promising

tools to quantify biological age in individuals using DNA methylation profiles from tissues of interest (Horvath, 2013). In Chapter 2, Section 2.4, we provide more details on three "generations" of epigenetic clocks, including their construction and utility in predicting a variety of age-associated phenotypes (Teschendorff and Horvath, 2025). Epigenetic clocks have been associated with various age-related phenotypes, cognitive decline, and mortality risk, offering a powerful framework for examining individual differences in aging trajectories. Recent studies have demonstrated that epigenetic modifiers are not only influenced by lifestyle and environmental exposures, including diet, exercise, stress, and social and structural determinants of health (SSDOH), but also serve as robust predictors of biological age.

In parallel with the development of epigenetic clocks, brain age estimation has emerged as a complementary approach for quantifying biological aging using neuroimaging data. Brain age estimation methods leverage ML algorithms trained on large datasets of structural and functional brain imaging to predict an individual's chronological age. The discrepancy between predicted brain age and actual chronological age, known as the "brain age gap", has been associated with a range of neurological, psychiatric, and systemic health outcomes, including cognitive decline, dementia risk, and all-cause mortality. These models are particularly sensitive to factors such as neuroinflammation, vascular burden, and lifestyle exposures, reflecting the brain's vulnerability to both intrinsic and extrinsic aging processes. The next section introduces the foundational principles of ML, delving into the computational mechanisms underlying biological age estimation models.

1.2 Principles of Machine Learning

The renowned American scholar Herbert Simon described learning as "any process by which a system improves performance from experience." While learning has traditionally been associated with intelligent life forms, artificial intelligence (AI) focused on the development of computing systems that automatically learn to make predictions or decisions. The term "machine learning" was popularized by Arthur Samuel, a pioneer of computer gaming and AI. Samuel famously noted that ML, which is one of the foundational subfields of AI, "gives computers the ability to learn without being explicitly programmed". At its core, ML is concerned with generalization, or the

capacity to infer patterns in observed data that support accurate predictions on new, previously unseen examples. This capability makes ML especially valuable in complex domains such as aging research, where large-scale, high-dimensional datasets present intricate, often nonlinear relationships that challenge traditional statistical modeling.

ML encompasses a broad family of approaches, typically organized into three primary paradigms based on the nature of the data and the type of learning task. In **supervised learning**, models are trained on labeled datasets that provide both input features and corresponding target outcomes. The goal is to learn a mapping from inputs to outputs that generalizes well to new data. On the other hand, **unsupervised learning** deals with unlabeled data and seeks to discover hidden patterns, groupings, or lower-dimensional representations, such as clusters or latent factors, without predefined targets. A third paradigm, **reinforcement learning**, involves training agents to make sequences of decisions by interacting with an environment, receiving feedback in the form of rewards or penalties, and gradually learning a policy to optimize task performance. While reinforcement learning currently plays a less prominent role in current biomedical research, both supervised and unsupervised learning have proven indispensable for tasks such as biomarker discovery and disease progression modeling.

The remainder of this section focuses on *supervised learning*, the most widely used paradigm in biological age estimation models. We examine its fundamental components, including data representations, model architectures, training processes, and evaluation metrics, to lay the groundwork for understanding how ML models are developed and validated in aging research.

1.2.1 Supervised Learning Paradigm

Most biological age estimation frameworks rely on supervised learning, a paradigm in which models learn a mapping between a set of input features (e.g., parcellated gray matter volumes or methylation levels) and a known output (e.g., chronological age). The training dataset is composed of pairs, $(x_1, y_1), ..., (x_n, y_n)$ where $x_i \in \mathbb{R}^d$ denotes a d-dimensional input vector for the i-th individual, and $y_i \in \mathbb{R}$ represents the corresponding target variable. The objective of supervised learning is to identify a function $f : \mathbb{R}^d \to \mathbb{R}$ that accurately predicts the output y for a given input x, even

when applied to previously unseen data. In regression tasks, such as predicting chronological age, the labels y are continuous, real-valued variables. Model performance is commonly optimized by minimizing the error of a loss function. Minimizing the loss guides the model to approximate the underlying relationship between input features and the output variable as closely as possible.

Supervised learning algorithms can be broadly categorized into several families, each with distinct modeling assumptions and tradeoffs. In the next section, we provide an overview of different ML model families and loss functions, which comprise two of the design choices in ML model architecture.

1.2.2 Model Architecture and Design

Designing a supervised ML model involves three core decisions that critically shape its ability to learn patterns from data and generalize to unseen examples: choosing a model family, defining an appropriate loss function, and selecting an optimization algorithm. The model family refers to the class of functions the model can learn. A central concept in evaluating model families is their capacity, or the complexity of functions they can represent. High-capacity deep learning (DL) models can capture intricate, nonlinear relationships in the data, while lower-capacity models such as linear regression impose stronger assumptions. However, greater model capacity increases the risk of "overfitting", where the model learns spurious noise or random fluctuations specific to the training data, rather than capturing the underlying signal. Conversely, models with insufficient capacity may underfit, failing to learn meaningful patterns and yielding poor predictive performance. This tension is formally captured by the "bias-variance trade-off", a fundamental principle in ML. Models with high bias rely on overly simplistic assumptions and tend to underfit the data, systematically missing relevant structure. In contrast, high-variance models are overly sensitive to the training data and prone to overfitting, exhibiting large fluctuations in predictions when exposed to new data. Optimal generalization performance requires finding a balance between these two sources of error, which is especially challenging in biological settings characterized by high-dimensional features and relatively small sample sizes. The second major design element is the loss function, which is essentially an error metric that the model seeks to minimize during training. In regression-based age prediction tasks, the most common loss functions are the mean squared error (MSE) and mean absolute error (MAE).

The choice of loss function reflects assumptions about the error distribution and directly influences the optimization landscape. Finally, the learning process is guided by the third design choice—the optimizer, which adjusts model parameter weights to minimize loss. Optimization algorithms such as gradient descent iteratively update parameter weights based on the gradient of the loss function. Together, these three design choices determine the representational capacity, learning behavior, and performance of an ML model. The next section turns to the critical role of data preprocessing and feature extraction, steps that directly impact model capacity and the bias-variance tradeoff by shaping the input representation before learning begins.

1.2.3 Data Preprocessing and Feature Engineering

Extensive data preprocessing and feature engineering are necessary to ensure the input data is structured, reliable, and informative prior to model training. This process involves extracting and integrating features from diverse sources and handling missing data. Common approaches to missing value handling include elimination of incomplete features or samples or imputation of missing values using statistical methods or model-based prediction. Missingness could be caused by numerous reasons, from random human error and measurement faults, to mismatches during data merging, to inherent properties of the data and feasibility of acquisition (e.g., patient dropout due to disease severity or death and accessibility to healthcare) In neuroimaging and epigenomic datasets, the number of input features often far exceeds the number of observations. For example, structural MRI scans can yield thousands of anatomical measures, and DNA methylation arrays now profile almost 1 million cytosine-guanine dinucleotide (CpG) sites per sample (Kaur et al., 2023). This leads to multicollinearity and an increased risk of overfitting, making traditional statistical methods unreliable. Dimensionality reduction and regularization are two key strategies for addressing this issue. Regularization introduces a prior into the loss function to constrain model complexity and stabilize learning. L1 (Lasso) regularization promotes sparsity by shrinking some coefficients to zero, effectively performing feature selection, while L2 (Ridge) regularization shrinks all coefficients toward zero without eliminating any entirely. Elastic Net regression combines both L1 and L2 penalties, making it a favorable model for high-dimensional datasets such as epigenomic profiles.

Data-driven dimensionality reduction strategies, such as principal component analysis (PCA), provide another alternative. PCA transforms the original feature space into a smaller set of uncorrelated components. Recent generations of epigenetic clocks, known as "PC clocks," rely on PCA to condense epigenome-wide methylation signals into a limited number of principal components that capture dominant sources of variation (Higgins-Chen et al., 2022). While this improves model stability and reduces overfitting, it also diminishes interpretability by obscuring the contributions of individual CpG sites and biological pathways. As a result, PC clocks tend to prioritize predictive performance over biological insight, underscoring the trade-off between model simplicity and interpretability that often arises in high-dimensional biological data.

1.2.4 Training, Optimization, and Evaluation

The training phase of a supervised ML model involves adjusting model weights and hyperparameters to minimize prediction error on a labeled dataset. However, optimizing performance on the training set alone is insufficient: the true goal is to develop models that generalize, i.e., maintain predictive accuracy on unseen data. To this end, cross-validation (CV) is essential. Nested CV is highly recommended when performing both feature selection and hyperparameter tuning, as it prevents data leakage and yields more reliable performance estimates. CV simulates the deployment of the model in independent samples, providing a robust assessment of its practical utility.

In addition to optimizing performance, modern evaluation frameworks incorporate stratified sampling to ensure balanced representation of subgroups, e.g., those defined by sex, clinical diagnosis, or ancestry, which helps to reduce bias and improve fairness. Model performance is typically assessed using standard regression metrics such as MAE and the Pearson correlation coefficient. In aging studies, derived metrics such as the biological age gap provide biologically meaningful interpretations of model output. These derived measures are frequently used as surrogate aging biomarkers and linked to health outcomes, underscoring the importance of rigorous and transparent evaluation in the development of ML models for aging research.

1.2.5 Interpretability, Fairness, and Bias

Interpretability of AI models has emerged as a crucial concern, not only for understanding how models make predictions, but for ensuring that those predictions are meaningful, trustworthy, and ethically sound. The embedded feature selection and weighting provided by Elastic Net makes it a popular method when interpretability is prioritized, despite the lower capacity of linear regression models. In DL settings where full transparency is not feasible, tools such as SHapley Additive Explanations (SHAP values) enable researchers to decompose predictions into feature-level contributions (Lundberg and Lee, 2017), offering insights into which features most influence a given prediction. These explanations are particularly useful for debugging and model validation, even. Coupling interpretability techniques with domain expertise can reveal biologically plausible associations, uncover data preprocessing issues, and support the generation of new hypotheses about aging mechanisms.

Beyond interpretability, ensuring fairness and mitigating model bias are critical challenges, especially when working with biomedical datasets. Supervised models trained on biased datasets, such as those with imbalanced representations of racial, ethnic, and socioeconomic groups, are prone to systematic error disparities and biased predictions. These biases can be inherited from historical inequalities embedded in the data, including tainted labels or differential measurement quality. To reduce such risks, models should be trained on large and diverse datasets, using data augmentation or transfer learning when necessary, However, fairness is not a purely technical problem. Rather, it is shaped by societal norms and longstanding ethical debates. Decisions about which metrics define a "fair" model must involve close collaboration with domain experts, recognizing that imposed fairness constraints may not always align with clinical or biological validity. Transparency, careful evaluation, and human-in-the-loop approaches are essential to ensure that ML models contribute to equitable and scientifically robust insights.

1.2.6 Applications to Biological Age Estimation

Quantifying biological age measures, including brain age, has become a prominent focus of computational neuroscience research with ML/DL playing an increasingly critical role. Traditionally,

brain aging was studied using scalar neuroimaging markers such as cortical thickness, hippocampal volume, or white matter hyperintensity burden. However, recent advances have enabled the development of multivariate models that estimate an individual's brain age from high-dimensional neuroimaging data, most commonly T1-weighted structural MRI, by training regression algorithms to predict chronological age from imaging features. The predicted brain age is then compared to actual chronological age to compute the brain age gap, which is interpreted as a biomarker of accelerated or decelerated aging. Models have ranged from linear regressions on hand-crafted features to more complex random forests, support vector regressors, and DL models including convolutional neural networks (CNNs) applied directly to voxel-wise data. CNNs promise higher predictive accuracy and minimal reliance on preprocessing or feature engineering. However, they also introduce challenges around interpretability, data leakage, overfitting, and reproducibility.

Many studies evaluate brain age gap as a biologically useful metric of its own, positing that the discrepancy between observed and predicted age is reflective of biological variability that may be linked to resilience or vulnerability. A central controversy in this field surrounds residual-based approaches to calculate brain age gap. Often intended to correct for regression-to-the-mean effects, this practice can introduce statistical artifacts and undermine biological interpretability. For instance, residuals may reflect model variance rather than true biological deviations, and their distribution can be distorted by age-skewed training samples. Moreover, the use of age as a training target inherently conflates chronological and biological aging, raising questions about whether ML models are capturing meaningful pathophysiological signals or merely age-correlated anatomical variance. Cross-site generalizability is another significant engineering barrier, as site-specific scanner effects, demographic variation, and preprocessing pipelines can substantially degrade model performance. Recent efforts to address these issues include harmonization techniques such as ComBat, domain adaptation, and federated learning frameworks that aim to reduce site bias without compromising privacy.

Given the parallels between brain age and epigenetic age estimation, many of these challenges and considerations are directly applicable to epigenetic clocks. Ultimately, while ML-based biological

age models hold considerable promise as non-invasive tools for early detection and monitoring, their clinical and mechanistic validity hinges on resolving these methodological challenges, integrating multimodal data, and developing models that are both accurate and interpretable.

1.3 Overview of Dissertation Aims

Building upon recent advances in neuroimaging, epigenetics, and ML approaches to biological age estimation, this dissertation examines how epigenetic measures relate to *in vivo* neuroimaging biomarkers of brain aging and neurodegeneration, offering a biologically grounded structure for understanding disease progression and heterogeneity of aging and AD. Within this context, we explore the potential for peripheral blood-based epigenetic markers to offer insights into brain aging processes and their variability across individuals.

At the core of this dissertation is the central hypothesis that epigenetics contributes meaningfully to the heterogeneity of brain aging and neurodegeneration, as quantified by structural MRI. We postulate that the intra-individual variation in vulnerability of specific brain regions to AD pathology and neurodegeneration, such as in the MTL, may be at least partially explained by differences in biological age as quantified by epigenetic clocks. We leverage both statistical and data-driven approaches in the following two aims to explore this central question.

Aim I: Determine how biological age by blood-based epigenetic profiles influences atypical neurodegenerative patterns in AD

Chronological age has been associated with distinct patterns of AD pathology, including variations in the regional distribution and severity of neurodegeneration. To understand the heterogeneity in AD, particularly in cases exhibiting atypical patterns of neurodegeneration, it is essential to consider factors beyond chronological age that may contribute to divergent clinical and anatomical presentations. One such factor is biological aging, which can be estimated using DNA methylation-based epigenetic clocks that reflect cumulative molecular damage and regulatory drift over time. We apply established epigenetic clocks to whole-blood DNA methylation profiles to compute biological age gap (BAG), which quantifies the discrepancy between an individual's biological and chronological age. We will use BAG to classify individuals into accelerated or decelerated biological

age groups, thereby allowing us to probe whether biological aging status relates to distinct neurodegenerative patterns in AD. Using established neuroimaging signatures of AD, we will define a novel MRI-derived Cortico-Medial Temporal index (CoMeT) as the signed difference of the cortical and MTL structural measures. We will statistically examine the relationship between BAG and CoMeT using hypothesis testing, effect sizes, and power assessments. In summary, we will quantify atypical neurodegenerative pattern along the corticolimbic axis and assess whether biological age explains regional vulnerability beyond the known effects of chronological age. This will allow us to uncouple the association between atypical neurodegenerative patterns and age of onset, focusing more on biological explanations for AD heterogeneity than chronological age alone. Working Hypothesis: Decelerated BAG is associated with a more cortical-predominant anatomical presentation of AD, while an accelerated BAG corresponds to the more typical MTL-dominant pattern, thereby providing a mechanistic link between biological aging and phenotypic heterogeneity in AD.

Aim II: Identify blood-based epigenetic signatures and associated biological pathways underlying brain aging and neurodegeneration in the MTL

To develop novel epigenetic clocks that specifically predict brain age and hippocampal volume, traditional ML algorithms (Elastic Net) will be applied to whole-blood DNA methylation profiles covering 700,000 CpGs across the epigenome. Various feature selection methods using nested cross validation with hyperparameter tuning will be implemented to preselect CpGs that have the greatest association with the target feature of interest (i.e., brain aging or hippocampal volume), and subsequent prediction performance will be assessed to identify the best feature selection method and number of features to input to an Elastic Net model. Training and validation be done with an initial discovery cohort and the final clock will be tested in an independent dataset to evaluate predictive performance. The CpGs selected by Elastic Net will be annotated using existing databases. Gene set enrichment and gene ontology analyses will be conducted to assess the biological implications of these features. Finally, the relationship between methylation of the selected CpGs and brain methylation, as well as the relationship to blood and brain gene expression, will be analyzed to infer the ability of blood-based methylation profiles to inform brain-related processes. Working Hypothesis: Epigenetic modifiers measured in blood are predictive of brain aging and neurodegenerative

processes affecting the MTL.

Figure 1.1 illustrates the conceptual framework guiding the two aims of this dissertation, which seek to unify two distinct but complementary domains of research: (1) epigenetic models of biological aging derived from accessible peripheral tissue and (2) neuroimaging-based models of structural and pathological changes in brain aging and AD. By bridging these domains, the dissertation investigates whether epigenetic variation helps explain the divergent patterns of brain aging and variability in focal neurodegenerative patterns observed across individuals.

1.4 Outline of Chapters

Below, we describe the chapters of this dissertation in the furtherance of the aims outlined above.

In **Chapter2**, we discuss the background and recent methodological developments that motivate this thesis. First, we summarize the various clinical and neuropathological presentations of AD with a particular focus on identified atypical variants of the disease. Then, we summarize the various current frameworks for explaining this observed heterogeneity and atypicality, also identifying limitations of each framework, and motivating the study of epigenetics as a potentially unifying explanation. We then delve into blood-based epigenetic studies of the brain and comprehensively review studies focused on assessing cross-tissue methylation patterns and using blood as a surrogate for brain. Finally, we motivate the empirical chapters of this thesis based on the reviewed background, postulating epigenetics as a unifying explanation for regional vulnerability in the brain and hypothesizing that these changes can be measured in blood.

In **Chapter3**, we investigate how biological age determined by blood-based epigenetic profiles influences atypical neurodegenerative patterns in AD (Aim I). We demonstrate the utility of our CoMeT measure in capturing the spectrum of neurodegeneration along the corticolimbic axis and its implications for atypical variants of AD by validating it using established results from literature associating this pattern with chronological age. We then further determine the influence of biological age on this pattern. This chapter aligns with the first central question regarding the ability of epigenetics to explain region-specific neurodegenerative patterns, as well as the second question of

whether these epigenetic markers can be assessed in blood, linking it to the next chapter.

In **Chapter 4**, we identify blood-based epigenetic signatures and associated biological pathways underlying brain aging and neurodegeneration (Aim II). We discuss the ability of blood based epigenetic clocks to inform brain related processes, especially when those processes are highly associated with chronological aging. We identify and interpret the meaning of potential pathways implicated in the vulnerability of the MTL, identified by the data driven methods employed in this aim.

In **Chapter 5**, we discuss the results from each of the empirical chapters in the context of the central questions and hypotheses of this thesis. We summarize the key contributions of this dissertation and offer a roadmap for future studies to continue this work. We identify limitations, challenges, and alternative approaches that future studies can address, ideally with unlimited resources and time. This focuses on both engineering and methodological innovations as well as biological advances, concluding with potential implications for clinical practice and translational research.

The engineering methods and biological findings of this dissertation may improve the exploration of biological mechanisms underlying aging and neuropathological processes underlying neurodegenerative diseases such as AD. This may better inform patient stratification, therapeutic targeting, and further basic research into the prevention or even reversal of these processes.

CHAPTER 2

LITERATURE REVIEW

This chapter lays the foundation for the empirical chapters of this dissertation. In Section 2.1, we provide an overview of heterogeneity in AD, focusing on regional vulnerability to disease pathology and downstream effects on cognition. We then delve into current frameworks to explain heterogeneity in AD and identify limitations of each in Section 2.2, motivating the need to develop a complementary explanation. This sets the stage for Section 2.3, which is a précis on the case for epigenetics as a potentially unifying framework for characterizing heterogeneity and regional vulnerability in AD. Section 2.4 provides a literature review of peripheral blood-based epigenetic studies of brain aging and associated phenotypes. Finally, Section 2.5 provides a summary of the chapter and ongoing research challenges, offering a roadmap for future work and a contextualized preview of empirical Chapters 3 and 4, which begin to address some of these open questions.

2.1 Heterogeneity and Atypicality in AD

Recognized globally as a devastating neurodegenerative disorder, AD was first described in the early 20th century by its namesake, Dr. Alois Alzheimer, who was relating a case which had been observed in the insane asylum of Frankfurt am Main in Germany. Dr. Alzheimer noted at the time, "Considering everything, it seems we are dealing here with a peculiar illness" (Stelzmann et al., 1995), emphasizing that the case was so clinically unusual that it could not be classified as one of the recognized illnesses. Despite the rapidly growing prevalence of AD since it was first documented just over 100 years ago, Dr. Alzheimer's description remains apt, albeit in a different sense than what he may have meant. The word "peculiar" is defined as "characteristic of only one person, group, or thing", alternatively, "distinctive" (Merriam-Webster, 2025). Today, there is increasing recognition of the extensive heterogeneity in AD leading to vastly different clinical presentations, pathologies, and rates of progression among individuals. This makes it indeed a "peculiar" disease. Heterogeneity in AD has significant implications for both research and clinical practice, particularly in the current era, as disease-modifying therapies become available and precision medicine approaches are prioritized.

2.1.1 Typical presentations of AD

Typically, AD is characterized by a gradual and progressive decline in cognition, with the earliest and most salient impairment occurring in anterograde episodic memory. During the prodromal phase of AD, in which individuals are often diagnosed with mild cognitive impairment (MCI) due to AD, memory concerns are reported most frequently. Individuals often face difficulty recalling recent events or conversations (Wilson et al., 2011). As the disease progresses, additional cognitive domains become affected, including executive function, attention, semantic memory, visuospatial processing, and praxis (Grady et al., 1988). Despite this evolution, episodic memory impairment remains the clinical hallmark of the "typical", amnestic presentation of AD. The onset is usually insidious, and both patients and family members may attribute early symptoms to normal aging, often delaying clinical recognition until deficits become pronounced (Grady et al., 1988). Memory dysfunction reflects the early involvement of MTL structures, particularly the hippocampus and ERC, which are critical for encoding and consolidating new episodic memories.

Neuropathology

The neuropathology of typical AD is defined primarily by two protein aggregates: amyloid-beta (Aβ) peptides in the form of senile plaques and hyperphosphorylated tau protein forming neurofibrillary tangles (NFTs). The pathological deposition of (Aβ) is believed to initiate in the ERC, subsequently involving the hippocampus and other MTL structures, and finally extending to neocortical areas including the lateral temporal, parietal, and frontal cortices. This stereotyped progression of tau pathology has been formalized into the Braak staging system (Braak and Braak, 1991), which strongly correlates with clinical severity. However, some individuals exhibit tau distributions that deviate from classical staging schemes, highlighting the complex biological heterogeneity underpinning clinical phenotypes (Jack Jr et al., 2018). Notably, Aβ deposition often plateaus early in the disease and shows weak correlations with symptom severity, whereas the spread of tau pathology more closely tracks with cognitive decline and neuronal loss (Jack et al., 2024).

Neuroimaging features

Neuroimaging studies have substantially contributed to our understanding of the topography and progression of AD-related neurodegeneration, providing *in vivo* correlates of both neuropathology and clinical symptoms. Structural MRI reliably demonstrates early atrophy of the MTL, particularly the hippocampus, ERC, and parahippocampal gyrus. This pattern of atrophy mirrors the regions initially affected by tau pathology and aligns with the hallmark memory impairment seen in typical AD. As the disease progresses, greater neurodegeneration is observed in the lateral and posterior temporal lobes, inferior parietal cortex, and medial posterior cingulate, correlating with impairments in semantic memory, visuospatial abilities, and attention. Fluorodeoxyglucose positron emission tomography (FDG-PET) reveals reduced glucose metabolism in the posterior cingulate cortex, precuneus, and temporoparietal junctions. Hypometabolism in these areas has been associated with deficits in episodic memory retrieval, orientation, and attentional modulation. Amyloid PET imaging, using tracers such as Pittsburgh Compound B, florbetapir, or florbetaben, can detect fibrillar Aβ deposition years before the onset of clinical symptoms (Landau et al., 2013). However, amyloid burden shows only modest associations with disease severity and is often widespread even in asymptomatic individuals, suggesting it is necessary but not sufficient for symptom manifestation. Finally, the development of PET tracers that allow visualization of tau neurofibrillary pathology is a more recent development that shows greater spatial correlation with neurodegeneration and cognitive decline. In typical AD, tau PET reveals that accumulation is most prominent in MTL structures in the early stages and spreads to lateral temporal and parietal cortices (Ossenkoppele et al., 2016).

2.1.2 Atypical Presentations of AD

Atypicality in AD refers to deviations from the typical presentation of the disease as described in Section 2.1.1. Atypical presentations include variations in age of onset, initial deposition and spread of pathology, and neuroanatomical changes, often leading to challenges in accurate diagnosis and management. These presentations disproportionately affect individuals with earlier disease onset, with symptoms emerging before the age of 65 years. Table 2.1 provides an overview of the recognized subtypes and atypical variants of AD.

Table 2.1: Comparative Overview of Alzheimer's Disease Subtypes and Variants

Subtype	Atrophy Pattern	Clinical Features	Pathology
Typical Amnestic	Severe hippocampal and MTL atrophy; additional temporoparietal cortical loss.	Prominent episodic memory impairment and spatial disorientation.	Relatively balanced tau distribution affecting limbic and association cortices.
Cortical-Predominant (Hippocampal Sparing)	Relative preservation of hippocampus; pronounced atrophy in parietal, temporal, and frontal association areas.	Non-amnestic symptoms (aphasia, apraxia, visuospatial or executive dysfunction); younger onset; faster progression.	High neocortical, low hippocampal tau.
Limbic-Predominant	Disproportionate hippocampal and MTL atrophy with minimal cortical involvement	Profound memory loss in very old patients; amnestic presentation	Dense tau in hippocampus and entorhinal cortex; relatively spared neocortex (Braak stage VI)
lvPPA	Focal atrophy in left posterior perisylvian (temporoparietal) cortex.	Fluent aphasia with word-finding pauses and impaired repetition; preserved grammar and early memory.	Asymmetric AD pathology in language regions; hippocampal-sparing tau pattern
PCA	Occipital and parietal atrophy, often right-lateralized; MTL relatively spared.	Visual cognitive deficits (reading, object agnosia); spatial difficulties.	Concentrated in posterior cortex; delayed hippocampal involvement.
Frontal / Behavioral	Frontal-predominant cortical atrophy.	Behavioral changes (apathy, disinhibition, poor judgment); resembles frontotemporal dementia.	Pathology focused on frontal networks; confirmed via biomarkers or autopsy.

Clinically defined variants

The two most well-documented atypical phenotypes are Posterior Cortical Atrophy (PCA) and logopenic variant Primary Progressive Aphasia (lvPPA). PCA is a clinical syndrome of progressive visuospatial dysfunction characterized by difficulties with spatial awareness and perception and worsening ability to interpret visual information. For example, individuals with PCA may struggle with reading, object recognition, spatial navigation, or simultaneous perception. First described by (Benson et al., 1988; Crutch et al., 2012), PCA exemplifies how AD pathology can target non-MTL cortical regions, leading to a clinical syndrome distinct from that of typical AD. Structural MRI of PCA typically demonstrates pronounced atrophy of the parietal and occipital cortices, which can be bilateral but is often asymmetric and more pronounced in the right hemisphere. A notable feature is the relative sparing of the hippocampus and MTL in PCA compared to typical AD. This preservation of MTL structures correlates with the observation that episodic memory is relatively intact in earlier

stages and is maintained longer during the disease course. Correspondingly, FDG-PET reveals hypometabolism in occipitoparietal regions, reflecting degeneration of posterior cortical networks. AD is the neuropathological cause of the majority of PCA cases, with abundant amyloid plaques and tau tangles accumulating in posterior cortical regions (Singh et al., 2015). Less commonly, other etiologies such as dementia with Lewy bodies (DLB) or corticobasal degeneration can also result in a PCA-like syndrome (Crutch et al., 2012; Chapleau et al., 2024).

In contrast, lvPPA presents primarily with language deficits, notably non-fluent speech and agrammatism. This syndrome is characterized by progressive language impairment, notably difficulties with word-finding and impaired sentence repetition, with relative preservation of grammar and motor speech (Gorno-Tempini et al., 2011). Like in PCA, other cognitive functions such as memory remain relatively preserved in early stages although individuals may develop more widespread cognitive deficits later. Neuroimaging in lvPPA classically reveals predominant atrophy in the left posterior superior temporal and inferior parietal cortices, corresponding to the posterior perisylvian language network. FDG-PET studies similarly show hypometabolism in left lateral temporoparietal regions and posterior cingulate cortex (Matías-Guiu et al., 2015). Clinically, lvPPA presents as a fluent aphasia with pauses (i.e., logopenia) rather than the amnestic memory loss observed in typical AD. Frequently, lvPPA is underpinned by AD pathology, evidenced by histology and amyloid PET studies. In one clinicopathological study, 84% of PPA cases with AD pathology exhibited a logopenic language phenotype (Henry and Gorno-Tempini, 2010). Thus, lvPPA represents a focal presentation of AD in which amyloid and NFTs are concentrated asymmetrically in left perisylvian cortices while MTL structures are relatively spared earlier in the disease course.

Neuropathologically-defined variants

Beyond clinical syndromes, *post mortem* studies have defined distinct neuropathological subtypes of AD based on the regional distribution of neurofibrillary tangles. In a landmark analysis of hundreds of AD brains, Murray and colleagues identified a "hippocampal-sparing" subtype – essentially a cortical-predominant pattern – in which tau pathology is disproportionately high in the neocortex but relatively low in the MTL and particularly the hippocampus (Murray et al., 2011). Conversely,

a limbic-predominant (i.e., MTL predominant) subtype showed the opposite pattern, with heavy NFT burden in the hippocampus and MTL with comparatively sparse neocortical involvement. The remaining "typical" AD cases had a more balanced distribution of tangles in both hippocampal and cortical regions. Quantitatively, about 75% of individuals fell into the typical category, whereas roughly 15% of AD cases were classified as either hippocampal-sparing or limbic-predominant. These neuropathological subtypes help explain some of the clinical variability in AD. Notably, they do not necessarily correspond one-to-one with the named clinical syndromes, but they provide a framework for extremes of regional vulnerability within AD. It is also notable that the "-sparing" nomenclature may be potentially misleading, given that almost all individuals along the AD spectrum experience global neuropathological changes, including in the hippocampus. It is helpful to instead consider these terms as relative rather than absolute for the purpose of biologically informed AD subtyping.

Neuroimaging correlates of these subtypes have been demonstrated *in vivo*. MRI studies and visual rating scales can reliably identify individuals with predominant cortical involvement versus those with predominant MTL involvement. Cortical-predominant AD shows relatively preserved hippocampal volume but disproportionate atrophy in cortical areas, including the temporoparietal and frontal regions. This pattern aligns with a non-amnestic clinical profile in which individuals are far less likely to present with memory loss. Approximately 42% of individuals begin with an amnestic syndrome and are more likely to present with language, visuospatial, or executive deficits (Graff-Radford et al., 2021). By contrast, MTL predominant presentations produce the most severe hippocampal and MTL atrophy with relatively lower cortical atrophy. Consistent with this, most MTL-predominant individuals present with the classic amnestic syndrome—in one study, 94% had memory-predominant concerns at disease onset (Whitwell et al., 2008).

Emerging atypical variants and subtyping methods

The frontal variant of AD is a rare subtype associated with prominent frontal cognitive and behavioral symptoms, which may include executive dysfunction and changes in personality. This variant is characterized by a significant increase in neurofibrillary tangles within frontal regions compared to typical AD pathology. Additionally, some individuals exhibit symptoms akin to frontotemporal

dementia in what is known as the behavioral variant of AD. This syndrome is characterized by changes in behavior and personality rather than memory loss. Another presentation is the dysexecutive variant, where individuals predominantly experience difficulties with executive functions such as working memory and cognitive flexibility. Recent studies have also identified various cognitive subtypes of AD through data-driven clustering methods applied to cross-sectional neuropsychological data. These studies have highlighted the difficulties of capturing the temporal nature of disease progression solely through cross-sectional assessments, as they often fail to reflect the dynamic changes in cognitive function over time (Scheltens et al., 2017). More recent studies have utilized innovative statistical methods to explore cognitive subtypes within AD, revealing diverse trajectories of cognitive decline among these variants.

2.1.3 Implications of AD heterogeneity on research and clinical practice

The phenotypic diversity observed in atypical AD underscores the complexity of clinical presentations, as these variants often exhibit overlapping symptoms that defy conventional classifications. Research indicates that only about 25% of atypical AD displays a "pure" clinical presentation, suggesting that most individuals experience a continuum of symptoms rather than discrete variants (Whitwell, 2024). Diagnosing AD is inherently complex due to the overlap of symptoms with other medical conditions, such as depression and metabolic disorders. This complexity is heightened in atypical cases, where symptoms may deviate significantly from the standard memory loss typically associated with the disease. Atypical forms of AD, particularly non-amnestic subtypes, frequently lead to misdiagnosis, complicating access to appropriate treatments and support services which are critical for improving quality of life. Integrating multi-dimensional data from neuropsychological assessments, neuroimaging, and genetic analyses may provide a more comprehensive understanding of atypical AD and inform the development of individualized treatment plans. As our understanding of atypical AD variants evolves, ongoing research should aim to refine diagnostic criteria and improve recognition among healthcare providers.

Understanding the heterogeneity of AD is also crucial for improving treatment strategies. The advent of disease-modifying therapies underscores the need for more precision medicine-based ap-

proaches to diagnosing and managing atypical AD, particularly as evidence suggests that initiating treatment in presymptomatic stages could yield better outcomes (Devi, 2023). Moreover, recognizing the unique trajectories of cognitive decline and the mechanisms behind them may facilitate the development of novel therapies that address the specific needs of patients. The diversity of atypical variants demonstrates that AD pathology can produce vastly different clinical syndromes depending on the regions affected, suggesting that additional modulators of disease heterogeneity are at work. Together, these variants highlight the importance of recognizing clinicopathological heterogeneity and spur further investigation into why different brain regions are vulnerable to neurodegenerations in different individuals. This sets the stage for developing frameworks to identify disease mechanisms and understanding how they interact to modulate patterns of regional vulnerability.

The variability in regional neurodegenerative patterns seen in atypical AD raises fundamental questions: why do certain brain regions succumb to AD pathology and neurodegeneration in some individuals, but not others? What drives the initial deposition of pathology in certain brain regions prior to others?

In the next section, we will discuss current frameworks for explaining atypicality and regional vulnerability in AD. Emerging evidence suggests that a complex interplay of factors may modulate regional vulnerability to neurodegeneration. Understanding the spectrum of AD variants and their neuroimaging signatures provides a crucial foundation for exploring how various biological and environmental factors, such as epigenetic modification, might influence these heterogeneous patterns of neuroanatomical change.

2.2 Frameworks Explaining Heterogeneity in Neurodegenerative Patterns

2.2.1 Chronological age

Association of age with atypical clinical presentations

Younger individuals more frequently exhibit atypical, non-amnestic AD variants, whereas older individuals more often have a "typical" amnestic presentation. In sporadic early-onset AD (EOAD), memory loss may not be the dominant symptom; instead, individuals can present with focal cortical

syndromes such as PCA (visual-spatial impairments), lvPPA (language deficits), frontal/dysexecutive AD, or an AD-related corticobasal syndrome. Amnestic AD increases in prevalence with older age, whereas at least a third of AD cases under 65 are non-amnestic variants. These atypical variants often begin in the 50s or early 60s. cortical-predominant AD tends to occur at a younger age of onset (often early 60s to 70s), whereas limbic-predominant AD is more common in older adults (late 70s to 80s) and predominantly female individuals. Cortical-predominant cases had a mean age 72 at death (over 10 years younger than typical AD) and were mostly male, whereas limbic-predominant cases averaged 86 years old and were mostly female. The relative degree of MTL versus cortical neurodegeneration can vary widely in AD, contributing directly to differences in clinical presentation and progression (Murray et al., 2011). Clinically, sporadic EOAD cases progress faster, with more global cognitive impairment (attention, language, visuospatial, executive) compared to late-onset cases at a similar stage (Sirkis et al., 2022). Thus, chronological age not only modulates overall risk of AD, but also influences the regional atrophy patterns and syndrome variants, with early-onset cases providing key evidence of AD's potential to target specific functional networks beyond the medial temporal memory system.

Neuropathological evidence corroborating the age-atypicality association

Younger individuals with AD tend to have a higher burden of tau neurofibrillary tangles and more pronounced cortical atrophy at disease onset than older individuals with AD. Especially in atypical cases, EOAD shows "hypothesis-free" patterns of tau deposition that correspond to the clinical syndrome. For example, high occipitoparietal tau is observed in PCA and likewise, left perisylvian tau in lvPPA. On the other hand, individuals with later-onset AD often have less neocortical tau and more mixed pathologies such as TDP-43 or vascular changes confounding the picture. Post-mortem and tau-PET imaging studies have demonstrated that younger AD onset is associated with a "cortical-predominant" pattern of tau pathology (more intense tau in neocortical association areas with relative sparing of the MTL), whereas individuals with later-onset AD show a more limbic-predominant tau pattern (heavy involvement of hippocampus and ERC). This aligns with classic Braak staging (Braak and Braak, 1991), in which tau pathology in typical late-onset AD (LOAD) progresses from ERC to hippocampus and then neocortex, but in younger individuals the neocortical

involvement is disproportionately high.

Limitations of age-based explanations of atypical neurodegenerative patterns

While chronological age is a useful heuristic for anticipating clinical and neuropathological patterns in AD, it has clear limitations in fully accounting for the observed regional heterogeneity across individuals. Although EOAD is more often associated with atypical variants and LOAD with typical amnestic presentations, this is not always the case. Some younger individuals present with an MTL-predominant pattern of neurodegeneration and classic amnestic symptoms, and conversely, some older adults exhibit cortical-predominance and atypical clinical symptoms. These mismatch cases challenge the notion that age alone determines the topography of pathology and the resulting clinical phenotype. Such variability suggests that other factors likely modulate disease expression beyond chronological age. This underscores the need for individualized assessments incorporating biomarker and imaging data rather than relying solely on age-based expectations to understand and predict AD subtype expression.

Genetic models

Investigations into the genetic underpinnings of atypical AD have revealed that common genetic risk factors, such as the APOE $\varepsilon 4$ allele, are primarily associated with LOAD. Emerging studies suggest that AD heterogeneity may be influenced by heritable mutations in specific genes as well as rare variants. Furthermore, imaging-genetic endophenotype studies aim to connect genetic markers with specific disease phenotypes, potentially elucidating the biological mechanisms driving the heterogeneity observed in atypical AD presentations. Ongoing research is also examining how factors like genetic polymorphisms impact heterogeneity along the spectrum of relative cortical to MTL network involvement. For example, the tau-associated MAPT H1 haplotype was significantly more frequent in limbic-predominant AD than in cortical-predominant AD, potentially reflecting network-level factors favoring cortical spread of tau, whereas limbic-predominant AD suggests a biology that confines pathology to MTL circuits.

Familial versus sporadic cases

Approximately 5 to 10% of AD cases show familial clustering and a small subset are caused by autosomal-dominant mutations, comprising less than 1% of all AD cases (Sirkis et al., 2022). These cases are often associated with early onset of disease and demonstrate more aggressive progression. Familial AD cases are usually caused by mutations in one of three genes: amyloid precursor protein (APP), PSEN1, or PSEN2. These mutations lead to increased production of Aβ-42 peptide, which is prone to aggregation, or otherwise alter the Aβ-42/Aβ-40 ratio. These known variants account for a majority of familial and autosomal dominant AD cases and approximately 10-15% of early-onset cases overall, indicating that most EOAD is not due to known Mendelian variations.

By contrast, over 90% AD cases are late-onset sporadic AD, which is not caused by a single gene but is influenced by a polygenic risk architecture. The strongest genetic risk factor for sporadic AD is the APOE ε4 allele, which is common at approximately 20% frequency in the general population and dose-dependently increases AD risk at about 3-fold per allele. It is found in a large proportion of AD individuals and is estimated to account for roughly a quarter of the genetic predisposition to LOAD. Notably, however, atypical EOAD variants are less likely to carry APOE4 compared to typical amnestic AD – for example, individuals with PCA or other nonmemory presentations often do not have APOE4, suggesting that APOE-associated pathways preferentially lead to the typical hippocampal pathology. Beyond APOE, recent genome-wide association studies have identified over 40 genetic risk loci for LOAD. These include genes involved in lipid metabolism (e.g., APOE, ABCA7), microglial immune response (e.g., TREM2, CR1, CLU, ABI3), synaptic function, and other pathways. Notably, rare variants in the microglial receptor gene TREM2 confer an approximately 2 to 3-fold increase in AD risk, similar in magnitude to a single APOE4 allele.

Limitations of genetic models

Genetics alone cannot fully explain regional vulnerability in aging and neurodegeneration. Even individuals with the same mutation can have variations in clinical presentation and known genes do not dictate which brain networks will bear the brunt of pathology in each individual. Furthermore, most sporadic cases have no single genetic cause, and risk alleles like APOE only partially predict

disease, indicating that other factors contribute to why certain brain regions degenerate in some individuals more than others.

2.2.2 Network-based models of pathological propagation

AD pathology follows a characteristic spatiotemporal sequence, suggesting brain network-based vulnerability and disease spread as a potential explanation for regional vulnerability. As previously mentioned, the Braak staging scheme establishes that tau pathology starts in the ERC and hippocampus and then progresses to higher-order association cortices in a stepwise manner. Modern network models build on this staging scheme by proposing that misfolded proteins, especially tau, propagate trans-synaptically from neuron to neuron along the brain's connectivity graph. In other words, the neurodegenerative process may "travel" via anatomical connections, causing connected regions to degenerate in tandem, rather than solely spreading contiguously through tissue. This hypothesis is supported by neuropathological evidence of misfolded tau and Aβ seed propagation and by the observation that syndromic variants of AD correspond to network-specific patterns of atrophy rather than random lesions. For example, the default mode network (DMN) is a highly connected network involving the posterior cingulate, precuneus, hippocampus, and temporoparietal junction. The DMN is heavily impacted in typical amnestic AD, whereas the visual network is targeted in PCA as are language networks in lvPPA. These patterns align better with functional connectivity maps than with just blood flow and vasculature or anatomical boundaries, hinting that network connectivity mediates vulnerability. Neuroimaging research has provided empirical support for network-based models.

Within the network paradigm, two complementary theoretical models have been proposed. The first is the trans-synaptic spread model, where the emphasis is on the agent (i.e., misfolded tau or Aβ) moving along pathways in the connectome. This model accounts for the experimental evidence that injected pathological tau can propagate to synaptically connected neurons. The second is the cascading network failure theory, which focuses on the vulnerability of network topology itself. This theory posits that certain networks and hubs are inherently more susceptible to failure in AD. For example, highly connected hub nodes such as those in the DMN may accumulate amyloid

and undergo functional failure, which then causes connected nodes to degenerate due to loss of network support. In this view, Aβ has a more global distribution, often relatively uniform across the cortex early on, reflecting uptake in hub-rich networks that support widespread connectivity. Tau pathology, on the other hand, marks the specific network modules associated with clinical deficits, leading to the diverse phenotypes.

Functional MRI (fMRI) studies show that AD-related tau and neurodegeneration tend to appear in regions that are functionally connected at rest. For instance, regions with strong connectivity to the ERC are the same regions that develop tau tangles later in the disease. Likewise, tau PET scans in atypical AD variants reveal that tau deposition respects functional network boundaries. Individuals with language predominant AD have tau concentrated in the left frontal-temporal language network, whereas PCA individuals have tau in occipital and dorsal parietal areas of the visual network. Diffusion-weighted MRI studies of the connectome further suggest that the progression of atrophy can be modeled by a network diffusion process starting from an epicenter such as the ERC, with the pathology spreading along white matter tracts to connected regions over time.

The cascading network model elegantly explains why amyloid PET tends to show diffuse cortical uptake even in atypical AD. Amyloid targets hub-rich regions that are common to many networks, yielding a uniform "substrate" for disease, perhaps impairing hub connectivity or triggering tau. On the other hand, tau and neurodegeneration are more heterogeneous and reflect the specific network that fails in each individual. Generally, amyloid and tau distributions can be dissociated, such that tau and resultant neurodegeneration correlate better with clinical phenotype than amyloid. Building on emerging models of tau-neurodegeneration (T-N) mismatch, network-based frameworks of AD heterogeneity suggest that regional vulnerability and resilience are shaped not solely by tau deposition, but by the proportion of tau pathology observed relative to neurodegeneration, i.e., T-N mismatches (Lyu et al., 2024). This highlights the variable susceptibility of neural systems to tau-induced injury. For example, in some individuals, pronounced MTL tau burden may result in minimal neurodegeneration, potentially reflecting resilience within the limbic network, whereas in others, modest cortical tau can lead to substantial neurodegeneration in association areas, indicating

heightened vulnerability. These network-based susceptibilities likely interact with co-pathologies, age-related neurobiological changes, and individual differences in cognitive reserve. Thus, mapping T-N mismatch patterns across networks provides a framework for understanding how disease expression deviates from age-based expectations and may offer more precise, biologically informed subtyping of AD.

Limitations of network models

Overall, network-based frameworks have substantially advanced our understanding of how AD might propagate through neuronal circuitry by accounting for connectome architecture rather than just anatomical proximity. These models align with the observation that AD targets networks, not just isolated regions. However, network models still leave some open questions, particularly surrounding the initial targeting of vulnerable regions or networks. For example, they do not fully explain why certain networks, such as the DMN, are initially vulnerable to AD pathology. Additionally, they cannot nor why misfolded proteins like tau begin aggregating in specific epicenters such as layer II ERC neurons. Network models must also be reconciled with biochemical factors driving the initiation of tau misfolding and interactions between amyloid and tau interactions that modulate network spread. These gaps point to the possibility that other biological mechanisms, including those at the molecular and epigenetic level, are at play in conferring regional susceptibility to pathology.

2.2.3 Other factors influencing AD heterogeneity

Co-pathologies

Older AD individuals frequently have additional age-related pathologies such as cerebrovascular disease or TDP-43 protein inclusions in limbic regions resulting in LATE. These co-pathologies might exacerbate neurodegeneration in some regions. For example, TDP-43 in the hippocampus leads to worsening memory loss in older individuals, thereby skew the clinical presentation toward "typical AD" in older individuals. In contrast, younger AD individuals may have relatively "purer" brains terms of pathology, which might allow the idiosyncratic network-specific effects of tau to manifest more clearly, such as isolated visual syndrome in PCA). This perspective suggests that

part of regional vulnerability is contextual, depending on what other pathologic processes are present to interact with Aβ or tau.

In younger-onset cases, the absence of traditional vascular risk factors suggests that atypical AD may arise from distinct underlying processes, possibly related to specific genetic factors, such as the absence of the APOE4 allele. A newer concept is that different conformations of misfolded proteins might cause different patterns of pathology. In other proteinopathies such as prion disease or alpha-synucleinopathies, distinct strains can target different cell types. One study suggested that the diversity of clinical phenotypes and atrophy patterns in AD could arise from different fibrillar strains of pathogenic tau, as well as from genetic predispositions. However, clear evidence of strain-specific tau in AD variants is still being investigated. This model attempts to link molecular heterogeneity at the protein level to regional vulnerability.

Neuroinflammatory and Glial Responses

There is growing evidence that non-neuronal cells, especially microglia and astrocytes, play a role in regional vulnerability. Regions with high baseline inflammation or an exaggerated glial response might suffer more damage once AD pathology begins. Studies comparing atypical and typical AD variants have found some intriguing differences – for instance, one PET study showed that individuals with PCA had disproportionately increased microglial activation in the parietal-occipital cortex and lower activation in the MTL, relative to typical AD. This suggests the immune response in PCA was concentrated in the regions that degenerated (visual cortices) and comparatively blunted in usually vulnerable areas (hippocampus), possibly contributing to the shift in vulnerability profile. Likewise, genetic studies indicate that certain immune-related gene variants (including TREM2) confer different risk levels for atypical presentations. The concept of regional "neuroinflammatory signatures" is emerging: some brain regions might be primed to respond to pathology with an aberrant immune reaction, accelerating local neurodegeneration. Glial cells also undergo aging; epigenomic studies show that microglia in aged brains can develop "immune tolerance" or primed states that might differ between brain regions. Such differences could underlie why, for example, the hippocampus and ERC (which show early AD changes) might have a distinct inflammatory milieu

or protein clearance capacity compared to relatively spared regions like the cerebellum.

Cognitive Reserve and Environmental Exposures

Cognitive reserve refers to the concept that individuals with higher education, occupational complexity, or intellect may cope longer with pathology before showing symptoms. As such, cognitive reserve may play a role in heterogeneous presentations of AD. For example, it could lead to atypical clinical presentations: a person with high reserve might sustain significant hippocampal pathology without obvious memory loss, only to present later when the disease affects another network, resulting in an atypical non-memory syndrome. Conversely, someone with lower reserve might become symptomatic earlier in the course, perhaps with the classic memory complaints. Thus, two individuals with a similar distribution of pathology could show different primary symptoms. Cognitive reserve does not explain regional vulnerability to the pathology itself, but it influences how that pathology translates to clinical heterogeneity.

In addition to cognitive reserve, lifestyle and environmental factors significantly modulate AD heterogeneity by interacting with brain health, resilience, and vulnerability. Exercise, diet, regular and restorative sleep, and engagement in cognitively and socially stimulating activities are all associated with reduced risk of AD and may buffer the effects of tau and neurodegeneration on cognition. These protective behaviors often correlate with higher education and income, highlighting the influence of broader social and structural determinants of health. Socioeconomic status and environmental exposures such as air pollution, neighborhood safety, access to green spaces, clean water, and nutritious food shape opportunities for healthy lifestyle choices. In turn, these exposures influence biological aging and vulnerability to neurodegenerative disease. Disparities in access to healthcare, diagnostic delays, and historical and ongoing structural inequities further compound risk in marginalized racial and ethnic groups, particularly in the United States. As such, the clinical manifestation of AD cannot be disentangled from the lived environments and social conditions that shape reserve, exposure, and resilience across the life course. Understanding these interdependent factors is critical to addressing heterogeneity in AD and developing equitable approaches to prevention, diagnosis, and care.

Each framework provides valuable insight yet leaves important gaps. It is important to note that these theories are not mutually exclusive and likely interact. This opens the door for additional explanations that support existing frameworks and offer a way to bridge the gaps.

2.2.4 The case for epigenetics as a potential modulator of AD heterogeneity

Epigenetics might play a key role as a complementary explanation. Epigenetic changes, such as DNA methylation and histone modification, are dynamically influenced by aging and environment and regulate gene expression in a cell- and region-specific manner. There is evidence that in AD, neurons in vulnerable regions show different epigenetic marks compared to those in resistant regions. For instance, one study found large-scale changes in histone acetylation in AD brains that correlated with tau pathology load. Epigenetic dysregulation could conceivably make certain regions, like the hippocampus, less resilient to Aβ or tau accumulation effectively predisposing those regions to earlier or more severe degeneration. Similarly, microglia in aging brains develop "innate immune memory" via epigenetic reprogramming. If microglia in the hippocampus accrue a pro-inflammatory or primed epigenetic state with age, that region might mount a more toxic response to amyloid plaques than a region whose microglia remain quiescent. Region-specific epigenetic landscapes could help tie together multiple pieces of the puzzle by integrating nature and nurture: epigenetics can be influenced by genetics, environmental exposures, and developmental history of brain networks. Considering epigenetic mechanisms is thus a logical next step, offering a framework that is both dynamic and regional. The epigenome can change over time and differ across brain regions and cell types. This could explain, for example, why the same disease process manifests differently across individuals or how lifestyle changes could modulate disease trajectory through epigenetic pathways. It could also explain the "latent" period of AD: decades of mid-life exposures and aging might epigenetically prime certain brain regions for vulnerability to pathology buildup in alignment with the prevailing staging schemes and hypotheses.

2.2.5 Summary

Current frameworks have significantly advanced our understanding of heterogeneity and regional vulnerability in AD leading to atypical clinical variants. Genetic models highlight inherited sus-

ceptibilities but leave much of the sporadic variation unexplained. Network-based models account for the patterned spread of pathology across connected brain regions, matching the clinical syndromes, yet they do not fully answer why those networks are initially targeted. Other theories of region-specific vulnerability draw attention to intrinsic cellular factors, comorbid pathologies, and systemic influences like neuroinflammation or cognitive reserve, each shedding light on pieces of the puzzle. However, each framework in isolation has limitations and heterogeneity likely arises from a convergence of these factors.

This sets the stage for exploring epigenetic mechanisms as a complementary and potentially unifying explanation. Epigenetics offers a dynamic framework linking aging, genetics, cellular and regional network biology, and environmental factors. By influencing gene expression in a region- and cell-type-specific manner, epigenetic changes could underlie the differential vulnerability of brain regions, the impact of age and life experiences, and even the network-level propagation of pathology. Furthering research in epigenetic models may yield a more holistic model of AD heterogeneity and inspire new avenues for individualized therapeutic strategies targeting epigenomic modulation of vulnerable brain regions.

The next section delves into established and emerging research on epigenetic modifications underlying brain-related phenotypes associated with aging and neurodegenerative disease, with a particular focus on peripheral blood-based signatures. We examine how epigenetics might interface with genetic, network, cellular, and environmental frameworks to produce the observed tapestry of heterogeneity and regional specificity in AD.

2.3 Epigenetics as a Complementary Framework for Modeling Heterogeneity

As discussed in Section 2.2, existing frameworks for explaining heterogeneity in AD remain insufficient to fully capture the complexity of its clinical and anatomical diversity. Genetic risk factors explain only a fraction of sporadic AD cases, while models based on connectivity and spread do not adequately account for initial vulnerability of specific brain regions. Epigenetics offers a promising integrative framework capable of reconciling these gaps by providing a mechanistic basis for region-specific susceptibility. Positioned at the intersection of genetic potential, developmental history,

and lifelong environmental exposures, epigenetic regulation may underlie the divergent trajectories of AD pathology across individuals and brain networks.

Reframing epigenetics as one of the drivers of initial pathology deposition and spatial distribution, rather than solely where it spreads, shifts the modeling paradigm. Epigenetic factors may dictate the regional "permissiveness" that enables initial pathological seeding. This permissiveness refers to enduring molecular configurations, such as DNA methylation patterns, histone modifications, and chromatin accessibility, that influence cellular identity, gene expression, and the intrinsic vulnerability or resilience of specific brain regions to neurodegenerative processes. The focal atrophy observed in atypical AD variants such PCA and lvPPA cannot be fully explained by pathological spread alone. Rather, these patterns may reflect developmental epigenetic programming that leaves certain cortical regions predisposed to neurodegeneration under pathological stress. Early-life adversity and prenatal conditions may also leave lasting epigenetic imprints on specific brain areas, shaping long-term susceptibility. These imprints persist and evolve in interaction with environmental exposures, physiological stressors, and the aging process, contributing to the emergence of atypical phenotypes later in life.

Unlike the genome, the epigenome is dynamic and potentially reversible. It encodes an individual's developmental history and environmental interactions in a region- and cell-type-specific manner (De Jager et al., 2014; Gjoneska et al., 2015). As such, epigenetic regulation provides a powerful framework for understanding both common and atypical forms of AD, capturing the intersection of biological determinism and environmental modulation more wholly than existing models.

This theoretical synthesis positions epigenetics as a unifying model for explaining heterogeneity in AD and lays the conceptual foundation for identifying biomarkers of regional vulnerability. The next section builds on this framework by reviewing emerging research into the epigenetics of brain aging and AD, focusing on how age-related epigenomic changes may shape trajectories of neurodegeneration. We place a particular emphasis on detecting epigenetic modifications in peripheral blood tissues, which are more stable and easily accessible than brain tissue and can be acquired longitudinally throughout the lifespan.

2.4 Peripheral Blood DNA Methylation as a Surrogate Marker of Brain Change

There is increasing recognition that epigenetic mechanisms play a central role in modulating region-specific vulnerability to structural brain changes in AD. Epigenetic modifications are heritable yet reversible alterations to DNA and chromatin that regulate gene expression without changing the underlying DNA sequence. They operate in cell- and region-specific manners to orchestrate gene regulatory programs. These mechanisms serve as an interface between genetic predisposition, environmental exposures, and developmental processes, offering a plausible explanation for the multifactorial nature and phenotypic heterogeneity of AD.

One key epigenetic process implicated in this context is DNA methylation, which refers to the addition of a methyl group on the 5' cytosine at a phosphorylated cytosine-guanine dinucleotide along the genome. Greater methylation indicates tighter regulation and therefore reduced expression of the associated gene. This review focuses on epigenetic studies in humans that explore molecular mechanisms underlying regional brain vulnerability in AD, with particular emphasis on cross-tissue analysis between blood and brain methylation. We survey recent advances in understanding how peripheral blood-based DNA methylation changes contribute to brain aging and AD-related phenotypes. Additionally, we discuss tissue- and cell-type specificity of epigenetic regulation, region-specific epigenetic landscapes in the brain, and how these relate to structural and functional network vulnerabilities. Finally, we also explore how epigenetics interact with AD co-pathologies and cognitive changes.

2.4.1 Regional epigenetic atlases and landscapes

Over the past decade, research has increasingly demonstrated that epigenetic regulation is essential for normal neuronal processes, including synaptic plasticity and memory formation, and that these regulatory systems become disrupted during aging and in AD. Epigenetic alterations have been implicated across nearly all major pathological pathways associated with AD, including amyloid-β accumulation, tau pathology, neuroinflammation, and metabolic dysfunction, through their capacity to modulate gene transcription at scale. Because these mechanisms mediate interactions between the genome and environmental factors, they provide a compelling framework for understanding how

modifiable lifestyle factors such as diet, education, and environmental toxins can influence AD risk. Moreover, they may help explain the vulnerability of certain brain regions and cell types to AD pathology, while others remain comparatively resilient.

Comprehensive methylation atlases from human cohorts (e.g., ROSMAP) show regionally distinct epigenetic landscapes correlating with known AD vulnerability patterns. The ERC and hippocampus—early targets of tau pathology—exhibit unique methylation profiles compared to more resilient areas like the cerebellum, indicating a potential molecular basis for regional vulnerability across individuals (De Jager et al., 2014; Ng et al., 2021). Integrative analyses using human brain tissue have identified epigenetically regulated gene networks associated with tau burden, amyloid deposition, and cognitive decline (Mostafavi et al., 2018; Ng et al., 2021). These networks highlight roles for immune-related genes in microglia, synaptic genes in excitatory neurons, and metabolic pathways in astrocytes, linking cell-specific epigenetic dysregulation with clinical phenotypes.

2.4.2 Associations between Blood and Brain Methylation

Epigenetic patterns are highly tissue specific. Human studies demonstrate significant differences in epigenetic landscapes between peripheral tissues (blood) and central nervous system tissues. DNA methylation in peripheral blood associates modestly with AD status, cognitive decline, and aging (Lunnon et al., 2014; Levine et al., 2015). However, peripheral epigenetic changes often do not reflect brain-specific regulation. Post-mortem brain analyses reveal extensive methylation differences between AD-affected and unaffected regions, underscoring the need for brain-specific epigenetic investigations (De Jager et al., 2014). This has important implications for AD research, since many studies seek biomarkers in peripheral blood that reflect brain changes. In general, most DNA methylation marks are not shared between blood and brain – inter-individual methylation variability in whole blood poorly predicts methylation variability in cortical brain tissue. The brain also exhibits unique chromatin states and histone modification patterns consistent with its long-lived post-mitotic cells and complex circuitry. These fundamental differences mean that epigenetic dysregulation in AD brain may not be detectable in blood, and conversely, peripheral epigenetic changes (for example, due to systemic inflammation or aging) might not mirror those occurring

in neurons. Indeed, epigenome wide association studies in blood have identified some methylation changes associated with AD, but their overlap with brain-specific changes is limited. Thus, the tissue specificity of epigenetic marks necessitates cautious interpretation of peripheral epigenetic biomarkers and underscores the importance of examining the brain directly to understand AD epigenetics.

A fundamental question in neuroepigenetics is to what extent DNA methylation patterns in easily accessible tissues like blood reflect those in the brain. Direct brain samples are difficult or impossible to obtain in living individuals, so many studies have investigated whether peripheral blood can serve as a surrogate for brain epigenetic changes. Overall, most CpG sites show poor correlation between blood and brain methylation levels, indicating that peripheral markers generally do not mirror brain methylation one to one. For example, a within-subject analysis of paired blood and brain samples found that only 7.9% of CpGs had a statistically significant strong correlation across the two tissues. This means that most DNA methylation changes in blood do not reliably predict methylation status in the brain. However, a small subset of CpGs does exhibit cross-tissue concordance, providing a potential opportunity to use those specific sites as proxies for brain changes.

Recent efforts have leveraged cross-tissue DNA methylation datasets to identify thousands of informative CpGs with high correlation ($r > 0.7$) between blood and various brain regions (BECon, Hannon et al., 2014). Notably, a 2024 study selected approximately 18,293 correlated CpGs and showed that some of these markers can distinguish neurodegenerative disease patients from controls using blood samples. In that study, nine blood–brain correlated CpGs showed significantly altered blood methylation in individuals with AD compared to controls, and those same CpGs displayed parallel differences in AD brain tissue. These concordant CpGs were in or near genes previously implicated in AD, lending biological plausibility to their disease association. Such findings support the rationale for using blood DNA methylation as a minimally invasive surrogate for certain brain epigenetic changes in aging and neurodegeneration. It is hypothesized that circulating immune cells such as T lymphocytes may acquire brain-specific methylation signatures and carry them in peripheral blood. In summary, while blood and brain methylomes are largely tissue-specific, a cross-

tissue comparison approach can isolate a subset of CpGs that meaningfully reflect brain processes, providing a foundation for blood-based biomarkers of brain aging and disease.

2.4.3 Association of blood-based Epigenetic clocks with Brain Changes

One of the most prominent applications of epigenetics in aging research is the development of epigenetic clocks, which are composite scores of DNA methylation that estimate biological age. First-generation clocks including Horvath's multi-tissue clock and Hannum's blood clock were trained to predict chronological age from methylation data and are broadly applicable across tissues (Horvath 2013, Hannum et al., 2013). Subsequent second-generation clocks were designed to capture aspects of biological aging more closely linked to morbidity and mortality risk. These newer clocks, including PhenoAge (Levine et al., 2018) and GrimAge (Lu et al., 2019, 2022) incorporate DNA methylation changes associated with the "health span" and "lifespan", rather than age alone. Notably, second- and third-generation clocks generally show stronger associations with age-related health outcomes than the original clocks. However, early studies yielded mixed results on whether epigenetic aging in blood relates to brain aging and dementia. A recent systematic review of mostly cross-sectional studies found no strong evidence that blood DNA methylation age acceleration is associated with dementia or mild cognitive impairment risk (2022). Similarly, a very recent expert review noted that neither second- nor third-generation clocks have shown robust associations with clinical dementia to date. These observations suggested that global epigenetic aging might have limited utility as a proxy for brain aging, at least when measured cross-sectionally (2025).

More recent studies, however, indicate that epigenetic clock measures can predict brain-related outcomes when considering longitudinal data. Two second-generation clocks, PhenoAge and GrimAge, significantly predicted progression to AD and faster cognitive deterioration in follow-up assessments. Strikingly, higher epigenetic age was also strongly associated with MRI evidence of brain aging, including greater cortical thinning in vulnerable AD-related regions, and increased white matter lesion burden. These findings suggest that blood-based epigenetic clocks can capture aspects of brain aging and neurodegeneration, especially when using clocks calibrated to biological aging and when analyzing changes over time.

Beyond traditional clocks, emerging "Pace of Aging" metrics have shown even more promise in linking blood methylation to brain health. The DunedinPACE clock is a third-generation blood-based epigenetic biomarker of the rate of aging that was tested in a large 2022 study of older adults from ADNI and the Framingham Heart Study. Notably, DunedinPACE was significantly associated with poorer cognitive test performance and with clinical AD diagnosis in the ADNI cohort, whereas first- and second-generation clock ages (Horvath, Hannum, PhenoAge, and GrimAge) showed inconsistent or no associations in that cross-sectional analysis. Moreover, participants with a faster DunedinPACE had higher risk of developing dementia during longitudinal follow-up. In an independent sample of 2,200 individuals, each standard deviation increase in DunedinPACE conferred a 27% higher hazard of future dementia onset. These results were confirmed in replication and highlight that the pace of biological aging, defined epigenetically, may be more relevant to brain aging than static age acceleration at a cross-sectional time point. In summary, blood-derived epigenetic aging metrics do reflect brain aging to a degree, especially when using more recent clocks and when examining longitudinal changes. These measures have been linked to higher likelihood of subsequent cognitive impairment, AD conversion, and even measurable brain atrophy.

Still, epigenetic clock readouts are influenced by a host of systemic factors, and initial reports caution that these markers are not yet robust dementia predictors on an individual basis. Ongoing research is working to improve brain-specific epigenetic clocks and to understand how blood-based age acceleration relates to neuropathology. The hope is that such biomarkers can eventually monitor brain aging and AD risk in a minimally invasive manner.

2.4.4 Performance of Brain-Based Epigenetic Clocks in Blood

Shireby et al. developed a cortex-specific epigenetic clock, henceforth referred to as the Cortical Clock, trained exclusively on DNA methylation data from human cortical tissue across a broad age range (1–108 years). Using Elastic Net, they identified 347 CpG sites that robustly predicted chronological age in brain tissue. In both a test dataset ($n = 350$) and an independent cortical validation cohort ($n = 1221$), the clock outperformed widely used epigenetic age estimators, including Horvath's multi-tissue clock (Pearson $R = 0.83$ vs. $R = 0.65$), Zhang's blood clock (Pearson $R = 0.52$),

and Levine's PhenoAge clock (Pearson $R = 0.32$). These results confirmed that the Cortical Clock provided a highly accurate and unbiased estimate of chronological age in brain tissue.

To assess its generalizability, the authors applied the cortex-trained clock to peripheral blood samples ($n = 1175$, aged 28–98 years). While the Cortical Clock still showed a relatively strong correlation with chronological age ($R = 0.88$) and an RMSE of approximately 10.8 years, its performance was inferior to blood-trained clocks, particularly Zhang's model ($R = 0.97$, RMSE = 3.9 years approximately). Moreover, the cortex-trained clock systematically underestimated age in older adults, indicating a non-linear deviation that likely reflects the divergence of age-related methylation patterns between brain and blood tissues. This underlines a key limitation: although brain-trained clocks can capture some aspects of biological age in peripheral blood, they introduce systematic error when used outside their tissue of origin.

Overall, the findings by Shireby et al. (2020) highlight the importance of tissue-specific training for epigenetic clocks. While a brain-derived model retains some utility in peripheral tissues, its age predictions are biased and less accurate compared to clocks trained directly on blood methylation data. These results emphasize the necessity of using tissue-appropriate or cross-tissue-calibrated clocks for applications in aging and neurodegenerative disease research, particularly when using blood as a surrogate for brain-related biological processes.

2.4.5 Blood-based Epigenetic Studies of AD

Numerous studies have interrogated blood DNA methylation differences in AD. Human postmortem brain tissue analyses have revealed widespread methylation alterations in AD, particularly in cortical regions. The question is whether any of these disease-associated methylation changes manifest in peripheral blood, which could enable early detection or risk assessment without invasive procedures.

Over the past decade, a plethora of case–control studies have compared blood methylomes of AD patients and cognitively normal controls. A 2018 systematic review identified 36 such studies and reported that about 67% found significant DNA methylation differences associated with dementia.

This indicates that while not every study detects changes, a majority do report some AD-related

differential methylation in blood. However, specific findings have varied across cohorts, reflecting moderate reproducibility. Early candidate-gene studies focused on loci with known roles in AD and often found methylation differences in peripheral blood at AD-related genes. For example, hypomethylation of promoters in APP and PSEN1 was observed in some individuals, consistent with their overexpression in familial AD. More recently, epigenome-wide approaches have highlighted novel loci including hypermethylation in HOXB6, which was robustly associated with AD in multiple cohorts. DNA methylation changes at certain loci may be detectable years before symptom onset. Fransquet et al. (2020) found distinct methylation signatures in blood that could predict dementia up to 3 years prior to clinical diagnosis. Specific CpGs in genes like HOXB6, and APP showed altered methylation in individuals who later developed AD, compared to controls. These candidate markers hint at pathophysiologic changes measurable peripherally during the prodromal phase of disease.

Large-scale EWAS have attempted to systematically discover blood based CpGs associated with AD. Notably, a large meta-analysis in 2022 combined the Alzheimer's Disease Neuroimaging Initiative (ADNI) from North America and the Australian Imaging Biomarkers and Lifestyle study (AIBL) for a total sample of several hundred cases and controls. This meta-EWAS identified five CpG sites that were significantly associated with AD diagnosis in blood. While a handful of hits may seem modest, it likely reflects the reality that peripheral methylation differences in AD are subtle and require large samples to detect. To improve power and relevance, the authors also performed a cross-tissue analysis by integrating the blood EWAS results with methylation data from four cohorts with over 1000 postmortem brain samples. This cross-tissue meta-analysis highlighted 97 CpGs and 10 genomic regions that showed consistent methylation associations with AD in both blood and brain. In other words, by focusing on overlapping signals, they prioritized methylation changes that occur in the diseased brain and are also detectable in blood. Some of these cross-tissue markers were located near genes involved in synaptic function and immune pathways, aligning with known AD mechanisms. This study also created a weighted "methylation risk score" from the top blood–brain overlapping CpGs, like a polygenic risk score. In an independent cohort, this score modestly predicted AD status when combined with age, sex, and cell counts (AUC 0.70),

outperforming models without epigenetic information. Although not yet clinically applicable, this score demonstrates the potential of blood methylation profiles to contribute to AD risk prediction or diagnosis when aggregated appropriately.

Several smaller EWAS have reported AD-related methylation hits in blood that have since been replicated. For instance, hypomethylation at CpGs in B3GALT4 and ZADH2 was observed in AD compared to healthy individuals. Additionally, multiple studies on Japanese and European cohorts found significantly lower methylation at immune-related AD risk genes that were already known from GWAS as genetic risk factors for AD. The methylation evidence suggests an epigenetic component possibly mediating their effect. It is noteworthy that many AD-associated methylation changes in blood involve genes tied to the immune system or inflammation. AD pathogenesis has a strong neuroinflammatory component, and systemic immune dysregulation is often observed in patients. For example, HOXB6 is implicated in hematopoietic differentiation. Such findings support the idea that blood methylation changes in AD may reflect peripheral immune responses to the disease. Immune cells that infiltrate the brain, such as T-cells and monocytes, could undergo epigenetic reprogramming in the CNS environment and then carry those marks in circulation.

In 2023, an innovative approach was taken by Sun et al., who integrated large-scale genetics and methylation data to identify blood DNA methylation biomarkers related to AD risk. Instead of measuring methylation in patients directly, the authors used genetic predictors of methylation, known as methylation quantitative trait loci (meQTL) in about 111,000 AD cases and 677,000 controls from GWAS data. This methylation-GWAS integration identified 1,168 CpGs whose genetically predicted methylation levels were associated with AD risk (FDR < 0.05). Many of these CpGs were linked via expression quantitative trait loci to nearby genes, and in total 32 genes showed a consistent pattern where the risk-associated methylation change corresponded to altered gene expression and higher AD risk. Gene enrichment analysis confirmed that several of the implicated genes fall in known AD pathways or a network of neurological disease genes. This study demonstrates a powerful strategy to uncover methylation biomarkers, and it reinforces that many loci with methylation changes in blood are indeed involved in AD pathogenesis.

2.4.6 Blood-based epigenetic modifiers of brain structure and co-pathologies

Regional atrophy patterns in AD phenotypes may be driven partly by epigenetic regulation in neurons and glia. DNA hypermethylation of genes involved in cytoskeletal organization and axon guidance correlates with synaptic loss in AD temporal cortex (De Jager et al., 2014). Furthermore, neuronal chromatin accessibility changes relate directly to cortical atrophy severity (Corces et al., 2020). EWAS in peripheral tissues may provide non-invasive biomarkers correlating with these structural brain changes. A recent example comes from the ENIGMA-Epigenetics consortium which combined almost 3500 samples from 11 cohorts (Jia et al., 2021). The authors examined epigenome-wide associations with three regional brain volumes and identified two CpGs linked to the hippocampus, each explaining 0.9% of the phenotypic variance. These results point to small effect sizes in the relationship between individual peripheral DNA methylation markers and regional volumes, which might indicate limited value of single CpGs as biomarkers for brain structure. However, small effects might still be of mechanistic relevance, providing evidence for causality through pathway and enrichment analyses (Walton et al., 2023).

A model integrating epigenetic receptivity proposes that pathological proteins preferentially propagate to brain regions primed by specific epigenetic states, such as open chromatin and pro-inflammatory gene expression profiles (De Jager et al., 2014; Corces et al., 2020). This hypothesis complements connectivity-based propagation theories by introducing molecular specificity to vulnerability. For example, epigenetic signatures accompany tau pathology in vulnerable regions. In hippocampal and entorhinal tissues, tau accumulation associates with methylation changes in genes related to microtubule stability and tau clearance, alongside reduced histone acetylation at neuroprotective loci (Gjoneska et al., 2015; Smith et al., 2021). TDP-43 pathology as seen in LATE also features distinct epigenetic profiles in hippocampal and amygdalar neurons, implicating altered methylation of RNA-processing and stress-response genes as mediators of regional vulnerability (Yang et al., 2022).

2.4.7 Associations between blood-based epigenetic mechanisms and cognitive change

Cognitive trajectories in AD reflect both pathology and molecular resilience mechanisms. EWAS in brain tissues identified methylation changes linked to cognitive performance, involving genes for synaptic function and inflammation (Levine et al., 2015; Ng et al., 2021). Large-scale analyses indicate that blood DNA methylation has a modest association with cognitive function. In the first meta-analysis of EWAS on cognitive performance (2018), which pooled 11 cohorts (2,500–6,800 participants each) in middle to late life, only two CpG sites reached genome-wide significance. One site was associated with general cognitive ability and another with verbal fluency, located in an intergenic region on chromosome 12 and in the gene INPP5A on chromosome 10, respectively. These loci showed moderate correlation (Pearson $R = 0.4$) between blood and brain tissue methylation, suggesting some cross tissue relevance. However, neither CpG was linked to white matter volumes from brain MRI, and they did not associate with postmortem AD pathology or cognitive decline trajectories in supplemental analyses. This underscores the point that strong blood DNA methylation has limited signal associated with cognition, and any detected signal may not directly translate to structural brain changes or AD-related outcomes.

Several smaller cohort studies, often using twin designs to control for genetics, have reported additional candidate methylation–cognitive links. For example, an EWAS in 486 middle-aged monozygotic twins (average age 66 years) identified multiple CpGs where differential methylation correlated with cognitive test scores. The top signals for cross-sectional cognitive ability were in genes like ZBTB46 and TAF12 ($p = 5 \times 10^{-}7$), and pathways analysis of the top-ranked sites highlighted neuroactive ligand–receptor interaction pathways. When looking at 10-year cognitive change in the same twin study, the strongest CpGs were in AGBL4 and SORBS1 ($p = 1 \times 10^{-}6$), genes involved in neuronal survival and previously implicated in AD and stroke. Though these did not reach strict epigenome-wide significance, they hint that methylation changes in genes related to synaptic function and neurodegeneration may accompany cognitive aging. Another longitudinal study in Sweden of 535 twins found six CpGs significantly associated with baseline cognitive level at age 65, located in or near genes such as PPP1R13L, NRXN3, POGZ, and ENTPD8. These genes are involved in

neuroinflammation and synaptic function. Notably, that study found no epigenome-wide significant associations with 20-year cognitive decline, and the cross-sectional associations were roughly halved in within-pair analyses. This suggests that genetic or familial factors confound many links between methylation and cognition.

In general, blood methylation signatures explain only a small fraction of variance in cognitive aging. There is also substantial heterogeneity between studies, including different cognitive domains of focus, age ranges, tissues, and analytic methods, which makes it challenging to identify consistent biomarkers. A systematic review in 2020 surveyed 60 studies of DNA methylation vs. brain structure or function and concluded that results were modest and variable, with no standardized approach in the field. Key sources of bias included differing tissue sources, lack of cell-type adjustment, and small sample sizes in many reports. Therefore, while specific CpGs in blood have been linked to cognitive performance, these findings need replication and often have small effect sizes. Increasingly, researchers are exploring composite methylation scores rather than single-site biomarkers. These scores use weighted combinations of methylation levels at sites to create a surrogate measure associated with a phenotype such as cognitive ability. These "EpiScores" are analogous to epigenetic clock ages but assess age-associated phenotypes rather than biological age itself. Early evidence suggests that "EpiScores" can track risk factors for neurodegeneration, like chronic inflammation, and even directly correlate with cognitive test performance. In summary, blood-based methylation changes do relate to cognitive aging, but single CpG associations are weak and often entangled with genetic influences. Future studies leveraging large consortia, longitudinal data, and multi-CpG epigenetic scores may yield more robust blood biomarkers for cognitive decline.

2.4.8 Summary

Overall, the literature supports that while peripheral blood DNA methylation in AD exhibits consistent alterations at certain sites, these changes are relatively small in magnitude and influenced by genetic and environmental heterogeneity. No single CpG has emerged as a definitive blood test for AD. Instead, multiple subtle signals, often related to immune function, metabolic regulation, or known AD genes, distinguish AD from unimpaired individuals. Combining information across

many such CpGs is likely needed for any clinically useful biomarker. Another consideration is that age is the strongest risk factor for sporadic AD, and many AD-related methylation differences could be accelerated aging marks. Indeed, studies find that AD patients often show accelerated epigenetic aging in brain tissue, as measured by the Cortical Clock. The latest evidence using longitudinal designs and second or third-generation clocks suggests that methylation measures do capture increased aging and disease burden in incipient AD. Integrating epigenetic clocks, site-specific methylation markers, and other omics such as proteomics or transcriptomics may provide a multi-faceted blood biomarker panel for brain aging and AD.

Human-focused epigenetic research offers crucial insights into the molecular basis of regional vulnerability in AD. Studies leveraging brain tissue, peripheral biomarkers, and single-cell methods illuminate the complexity of epigenetic dysregulation driving AD heterogeneity. Peripheral DNA methylation is a particularly attractive tissue source due to the relative stability of DNA marks and reduced invasiveness of blood draws. While challenges remain, the field has made substantial progress in showing that blood methylation reflects aspects of brain health. With continued large-scale studies and cross-tissue validation, blood-based DNA methylation profiles could become valuable tools for monitoring neurodegenerative disease risk and the efficacy of interventions in the future.

2.5 Chapter Summary

This chapter reviewed the emerging evidence that epigenetic mechanisms, particularly DNA methylation, may play a critical role in modulating regional vulnerability in the brain and heterogeneity in AD. Human studies demonstrate that different brain regions and cell types maintain distinct epigenetic signatures, which may influence their susceptibility to hallmark AD pathologies such as tau and amyloid-beta. These regionally specific molecular environments help explain why certain brain areas degenerate earlier or more severely in both typical and atypical AD phenotypes.

However, the field faces several notable challenges. Firstly, there is limited availability of region-specific epigenomic data from *post mortem* human brains, especially in individuals with atypical AD variants such as PCA or lvPPA. Studies are also constrained by a lack of consistent longitu-

dinal sampling, and technical variability across platforms complicates longitudinal interpretation. Finally, atypical presentations of AD remain difficult to diagnose accurately, reducing the power and generalizability of phenotype-specific studies.

To address these challenges, future research should prioritize (1) regional epigenomic mapping in atypical AD brains, using emerging technologies such as single cell methylation and spatial epigenetics to resolve cell-type-specific vulnerability, (2) development of peripheral blood-based biomarkers that reflect central epigenetic states, enabling non-invasive early detection and stratification of patients based on predicted progression or therapeutic responsiveness, and (3) clinical translation, especially in the context of amyloid-clearing therapies, where epigenetic profiling may help identify patients vulnerable to atypical pathology or progression and consequently those less likely to respond optimally to amyloid-directed therapeutic interventions.

These directions align with a broader shift toward molecularly informed, precision approaches in neurodegenerative disease research.

2.5.1 Contextualized Preview of Empirical Chapters

This dissertation is motivated by the central hypothesis that epigenetic modifiers modulate the structural heterogeneity and regional vulnerability to atrophy in AD. While classical models have focused on genetic risk, molecular pathology, and network-based spread, they do not fully account for the consistent divergence in regional atrophy seen in atypical AD. Epigenetics offers a compelling unifying framework by capturing both intrinsic cellular properties and the accumulated impact of environmental exposures, shaping a region's receptivity to pathology. This view reframes heterogeneity in AD as an outcome of epigenetically encoded regional susceptibility. In doing so, it advances a novel, testable model for understanding and eventually intervening in the divergent trajectories observed across AD phenotypes.

Chapter 3 investigates the role of biological age, defined by peripheral epigenetic measures, in modulating the relative involvement of the cortex and MTL in AD. Established epigenetic clocks are applied to blood-based DNA methylation profiles to estimate biological age in individuals. We

then test the association of biological age acceleration (i.e., the discrepancy between biological and chronological age) with a novel index derived from structural MRI that quantifies relative neurodegeneration in the cortex and MTL. As such, this chapter characterizes the role of epigenetic modifications in modulating regional vulnerability of the brain, providing a novel explanation for heterogeneity and atypicality in AD.

Chapter 4 examines epigenetic predictors of regional structural atrophy using human methylation and MRI data. It identifies specific epigenetic markers that significantly predict cortical thickness and subregional volumes, particularly in the MTL. This chapter also explores shared biological pathways among predictive loci, revealing possible mechanistic links between epigenetic regulation and structural degeneration in AD.

Together, these empirical chapters test the central proposition that epigenetic mechanisms underlying biological aging and regional vulnerability to pathology underlie the spatial and phenotypic heterogeneity observed in AD and its atypical variants.

CHAPTER 3

EPIGENETIC AGE ACCELERATION MODULATES ATYPICAL NEURODEGENERATIVE PATTERN IN ALZHEIMER'S DISEASE

This chapter was adapted from Sreepada et al. (2025), "Biological age acceleration in Alzheimer's disease modulates relative cortical to medial temporal lobe neurodegeneration" published in *Neurobiology of Aging* with contributions from the following co-authors: Christopher A. Brown, MD, PhD; Sandhitsu R. Das, PhD; Paul A. Yushkevich, PhD; David A. Wolk, MD; and Corey T. McMillan, PhD. As first-author of this publication, I made key contributions in advancing the following aspects: conceptualization, methodology, Data curation, investigation, data visualization, formal analysis, writing the original manuscript draft, and review and editing of the manuscript.

The work presented in this chapter builds upon and expands findings previously published. The chapter presents novel insights into the epigenetic factors underlying the heterogeneity of neurodegenerative patterns in AD and individual vulnerability to less typical presentations of AD. In the original article, we describe our innovative approach focusing on the discordance between chronological age and atypicality in AD, hypothesizing that measures of biological age capture additional variance in this relationship. Here, we provide a more detailed treatment of the methodology and results. We also contextualize this work within the overall scope of the dissertation, linking it to preceding and subsequent chapters.

3.1 Introduction

Alzheimer's disease (AD) is an age-related neurodegenerative condition, and there is increasing recognition of the extensive heterogeneity in its clinical presentation. While clinical AD typically presents after 65 years of age, approximately 15% of individuals with AD present atypically and between 5% and 10% (Sirkis et al., 2022) have a younger age of onset (Koedam et al., 2010; Graff-Radford et al., 2021; Sirkis et al., 2022). AD is most often considered an amnestic, multi-domain disorder with memory loss most prominent due to early accumulation of tau and neurodegeneration in the medial temporal lobe (MTL). However, other less common phenotypes

are known to occur, including predominant language, visuospatial, or executive impairments that are often referred to as cortical-predominant presentations of AD, due to disproportionate atrophy in cortical regions relative to the MTL (Galton, 2000; Murray et al., 2011; Dickerson et al., 2017; Ferreira et al., 2017; Graff-Radford et al., 2021; Polsinelli and Apostolova, 2022). Although these presentations occur more frequently in individuals with early-onset disease, the contributions of aging mechanisms to AD heterogeneity are still poorly understood.

While chronological age is a known driver of AD pathogenesis (Brookmeyer et al., 1998; Swerdlow, 2007), there has been a surge in studies of biological aging in the past decade. There is particular interest in developing markers of biological aging using the epigenome or other "omics" data that are highly dynamic over the lifespan (Rutledge et al., 2022).

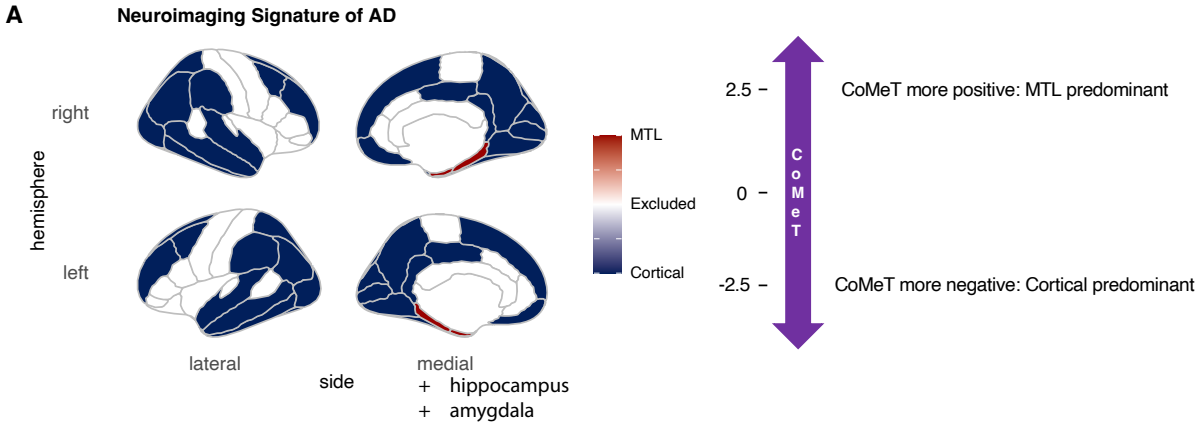
While biological age has been quantified using various methods and data modalities (Jylhävä et al., 2017), recent statistical approaches have focused on machine learning algorithms trained to predict age, mortality risk, or other age-associated phenotypes from DNA methylation profiles, termed epigenetic clocks (Hannum et al., 2013; Horvath, 2013; Levine et al., 2018; Lu et al., 2019; Belsky et al., 2022; Lu et al., 2022). Further, these epigenetic clocks have been used to measure the discordance between chronological age and biological age, known as the "biological age gap" (BAG), which represents the degree of age acceleration (biological age $>$ chronological age) or deceleration (biological age $<$ chronological age) in an individual (Salih et al., 2023). As such, two individuals with the same chronological age may show phenotypic differences that could be explained by differences in their BAGs. For instance, accelerated biological age has been associated with age-related outcomes such as mortality risk, frailty, and neurodegenerative conditions such as AD (Chen et al., 2016; Jain et al., 2022; Gao et al., 2024). Consistent with standard nomenclature, we adopt the terms "individuals with accelerated BAG" and "individuals with decelerated BAG" to describe groups who are estimated to be biologically older or younger on average than their chronological age, respectively.

In this chapter we evaluate the potential contributions of biological aging, defined by DNA methylation, to AD heterogeneity. We hypothesize that age-related biological factors are the mechanism by

which chronological age modulates AD presentation, thus we predict that biological age acceleration modulates atypicality in AD beyond chronological age alone. Given the evidence that less typical neurodegenerative patterns in AD are associated with earlier disease onset (Koedam et al., 2010; Graff-Radford et al., 2021; Sirkis et al., 2022), this study tests the *a priori* hypothesis that, for a given chronological age, individuals with decelerated BAG (i.e., biologically younger) are likely to have less typical, cortically predominant presentations and conversely that individuals with accelerated BAG (i.e., biologically older) are likely to have more typical, limbic-predominant presentations.

To investigate this, we developed the Cortico-Medial Temporal index (CoMeT): a novel score of disease atypicality in AD that is based on relative cortical to MTL involvement (Fig 3.1). CoMeT is a continuous measure derived from *in vivo* structural MRI and compares structures of signature brain regions involved in AD, which were previously established by neuroimaging studies (Bakkour et al., 2009; Dickerson et al., 2009; Möller et al., 2013; Touroutoglou et al., 2023). Specifically, CoMeT is defined as the signed difference of cortical thickness in non-MTL signature regions and of thickness or volume in MTL regions. Lower values of CoMeT indicate relative cortical predominance (i.e., less typical presentation), while higher values indicate relative MTL predominance (i.e., more typical presentation). Not reported in Sreepada et al. (2025), to relate structure to cognition, we also developed a cognitive CoMeT using harmonized cognitive domain scores, such that the executive function, language, and visuospatial domains represent cortical involvement and memory represents MTL involvement.

We then apply established epigenetic clocks (Horvath, 2013; Hannum et al., 2013; Shireby et al., 2020)to DNA methylation profiles to estimate biological age and BAG in symptomatic individuals. After adjusting for chronological age, we test whether CoMeT is significantly lower in individuals with decelerated BAG compared to those with accelerated BAG. In other words, we expect biological aging to influence the neurodegenerative pattern beyond the known effect of chronological age. By exploring the interplay between biological aging and disease presentation, we aim to deepen our understanding of the biological mechanisms underlying heterogeneity and atypicality in AD.



B

w-score	**MTL +**	**MTL -**
Cortex +	CoMeT ~ 0	MTL Predominant (higher CoMeT)
Cortex -	Cortical Predominant (lower CoMeT)	CoMeT ~ 0

Figure 3.1: **The Cortico-Medial Temporal index (CoMeT).** This study assesses the spectrum of neurodegenerative pattern in AD, which varies along a continuum from cortical-predominant to limbic-predominant (MTL focused). We define CoMeT to quantify the relative involvement of cortex to MTL. **(A)** Brain map (left) providing a visualization of the bilateral DKT atlas ROIs considered as part of cortex (blue): inferior temporal, temporal pole, superior parietal, precuneus, supramarginal, superior frontal, middle frontal, fusiform, inferior parietal, superior temporal, posterior cingulate, isthmus cingulate, lateral occipital, middle temporal, medial orbitofrontal, cuneus, lingual, and pericalcarine regions; and MTL (red): ERC and parahippocampal gyrus. Volumes of hippocampus and amygdala were also included in the MTL. Schematic diagram (right) providing interpretation of CoMeT, which quantifies relative cortical and MTL atrophy. **(B)** Table providing interpretation of CoMeT. Lower values of CoMeT indicate greater cortical involvement relative to MTL and correspond to less typical presentation. Conversely, higher values of CoMeT indicate greater MTL involvement relative to cortex, corresponding to more typical presentation. DKT = Desikan-Killiany-Tourville; MTL = Medial Temporal Lobe. ROI = Region of Interest.

3.2 Methods

3.2.1 Participants and study design

We included 1011 individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) who had an MRI scan and known amyloid status (Petersen et al., 2010; Weiner et al., 2017). Amyloid status was determined using the nearest PET or CSF measure within 2 years of baseline MRI, which we defined as the earliest timepoint with the smallest interval between MRI and DNA methylation, or simply the earliest timepoint for participants who did not have DNA methylation. We defined amyloid positivity using previously defined PET standardized uptake value ratios (SUVR) cut-offs for florbetapir (18F-AV45 SUVR $\geq$ 1.1) or Pittsburgh Compound B (PiB SUVR $\geq$ 1.47) (Landau et al., 2013), when available, or CSF Aβ peptide (A-β-42 $<$ 980) otherwise (Hansson et al., 2018).

Participants were further divided into two groups based on clinical consensus diagnosis closest to baseline MRI: cognitively unimpaired (CU) individuals (n = 329) who are amyloid-negative, and symptomatic individuals with either mild cognitive impairment (MCI) or dementia due to AD (i.e., amyloid-positive, n = 682). Among these individuals, a subset of 448 individuals had DNA methylation data (Vasanthakumar et al., 2020), which was required for analyses of biological aging, including 163 CU individuals and 285 symptomatic individuals. The Clinical Dementia Rating scale Sum-of-Boxes (CDRSB) (Morris, 1993; O'Bryant, 2008) was used as a measure of global function and cognition. A schematic diagram of the cross-sectional study design and inclusion/exclusion criteria is provided in Fig 3.2.

3.2.2 Estimating biological age using epigenetic clocks

DNA methylation data from whole blood were available in ADNI for 448 individuals who satisfied the study criteria as described above. DNA methylation was assayed using the Illumina Infinium Human Methylation EPIC v1.0 BeadChip (Pidsley et al., 2016), which assesses methylation levels at over 850,000 phosphorylated cytosine-guanine dinucleotide (CpG) sites across the epigenome. Raw intensity files were processed using the *SeSAMe* R package (Zhou et al., 2018) according to standard protocols, including sample filtering, CpG probe filtering, signal correction, and normalization.

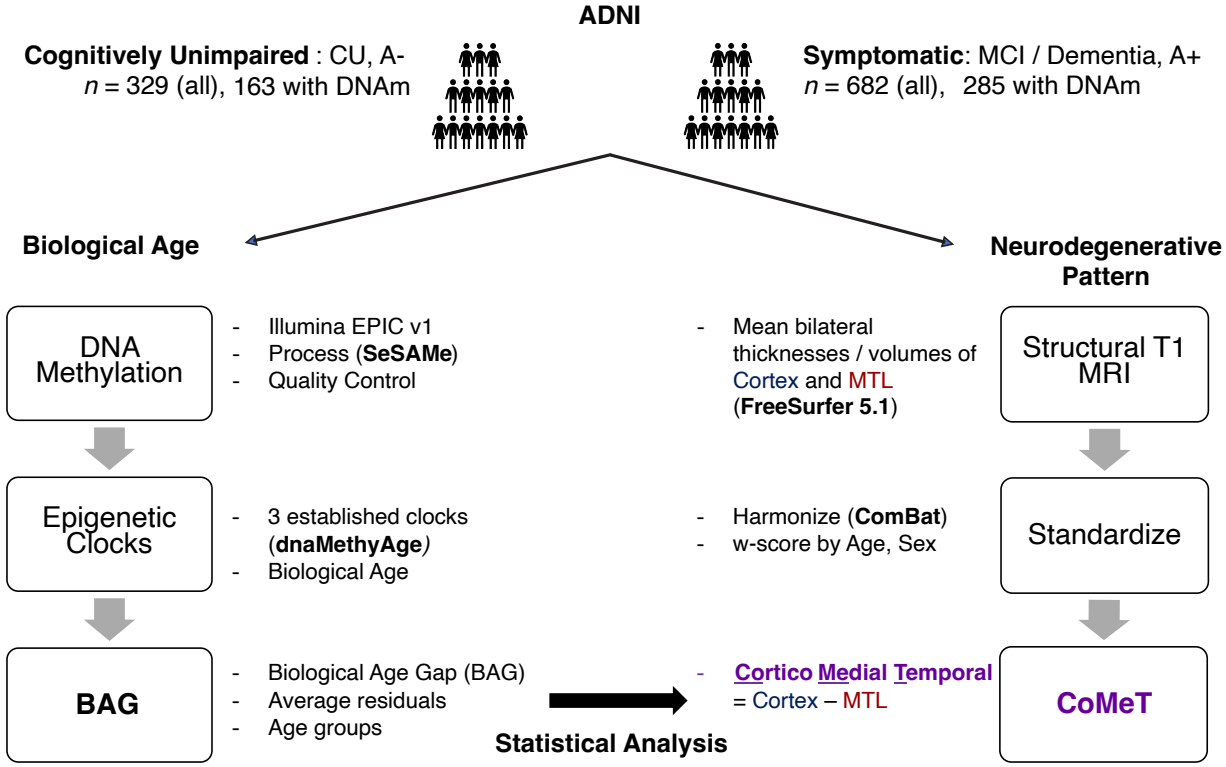


Figure 3.2: **Schematic diagram of the study design** This is a retrospective, cross-sectional case-control study. Participants from ADNI with MRI and amyloid status were selected, a subset of which also had whole-blood DNA methylation data. Participants were matched for age, sex, and education level and defined as CU individuals, who were amyloid-negative, or symptomatic individuals with MCI or dementia due to AD, who were amyloid-positive. Biological age was estimated using established epigenetic clocks for participants with DNA methylation. Age acceleration was computed for each clock age by regressing biological age against chronological age and extracting residuals. To mitigate noise from individual epigenetic clocks, residuals were averaged across clocks, giving the biological age gap (BAG) used throughout the study. We define CoMeT, an index to quantify relative cortical to MTL involvement along a continuum, using regional thickness measures from in vivo structural MRI. Wilcoxon rank-sum tests and Pearson correlation were used for statistical analysis of the association between BAG and CoMeT. A+/- = amyloid positive/negative; AD = Alzheimer's disease; ADNI = Alzheimer's Disease Neuroimaging Initiative; CU = Cognitively Unimpaired; CoMeT = Cortico-Medial Temporal index; MCI = Mild Cognitive Impairment.

Quality control procedures ensured at least 95% signal detection at each CpG (see Figure 3.3a), and Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2020) was used to check for sample separation by covariates (see Figure 3.3b).

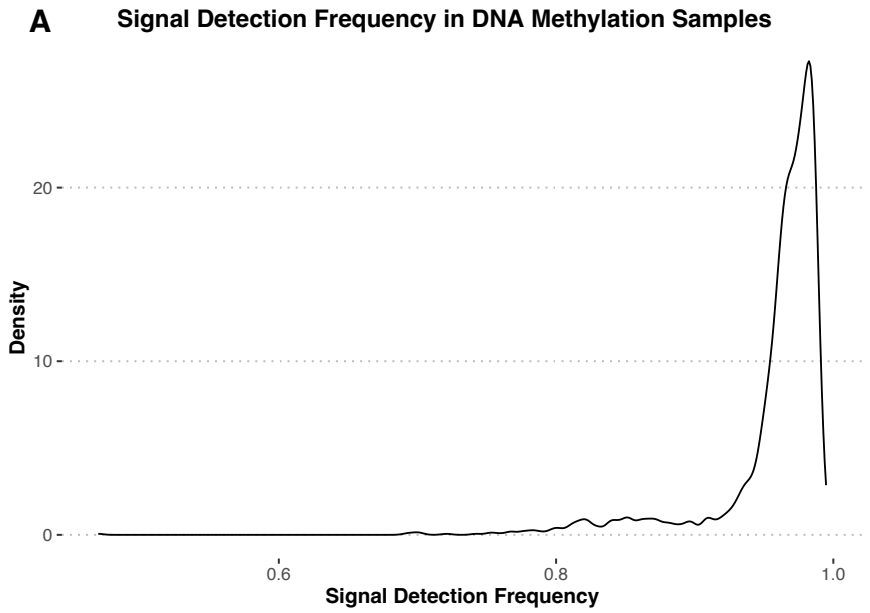
Three established epigenetic clocks (Horvath, 2013; Hannum et al., 2013; Shireby et al., 2020) were applied to processed DNA methylation data to compute biological age and BAG. The *dnaMethyAge* R package (Wang et al., 2023) was used to compute biological ages and the regression-based BAGs using the clocks referenced above. We opted for the regression-based measure because it reduces systematic deviations in BAG across datasets (Dhingra et al., 2018). There were no adjustments for covariates in the BAG calculation, as they were accounted for in downstream analyses. For timepoints with technical replicates, one replicate was randomly selected for further analysis.

While there is inherent noise in epigenetic clock measures due to variation in the populations and tissues used to train epigenetic clocks (Higgins-Chen et al., 2022), there is also high covariance amongst clock measures. To mitigate noise from sample variance, we averaged BAGs from the Horvath (2013), Hannum et al. (2013), and Shireby et al. (2020) clocks. This average is referred to simply as "BAG" throughout the rest of the paper and is the measure used in analyses. We report sensitivity analyses within individual clocks in Figures 3.4a and 3.4b.

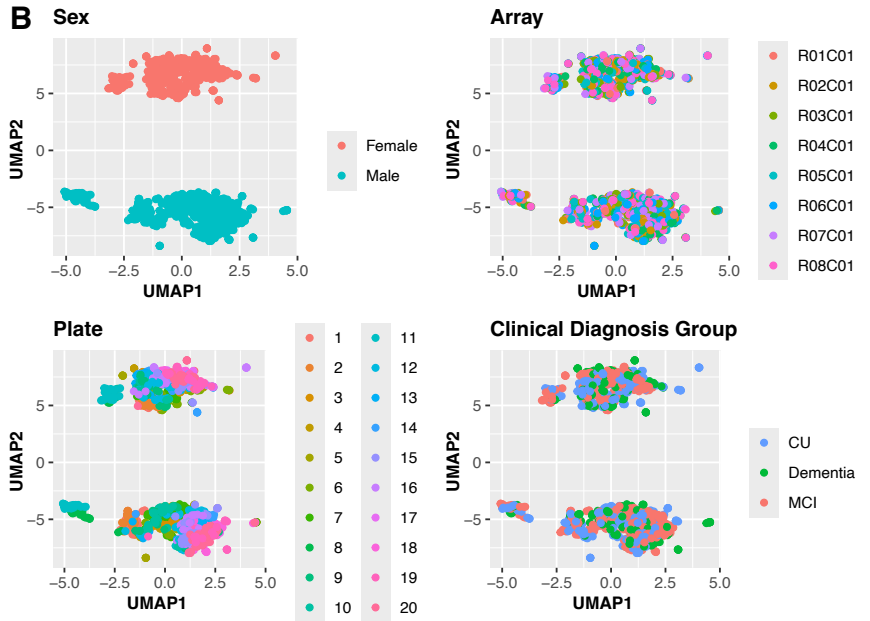
Biological and chronological age groups

Participants with DNA methylation data were stratified into biological age groups (accelerated, decelerated, and neutral) based on BAG as follows (see Fig 3.5). Accelerated: BAG was at least 0.5 standard deviation (SD) above the mean; Decelerated: BAG was at least 0.5 SD below the mean; Neutral: BAG was within 0.5 SD of the mean. The threshold of 0.5 SD, as opposed to a more extreme cutoff, was chosen to optimize the sample and effect sizes between the accelerated and decelerated groups.

All participants, regardless of DNA methylation availability, were grouped by chronological age to replicate results from literature (Dickerson et al., 2017) and serve as a reference for comparison: Younger (65 years and younger); Older (80 years and older); Average (between 65 and 80 years).

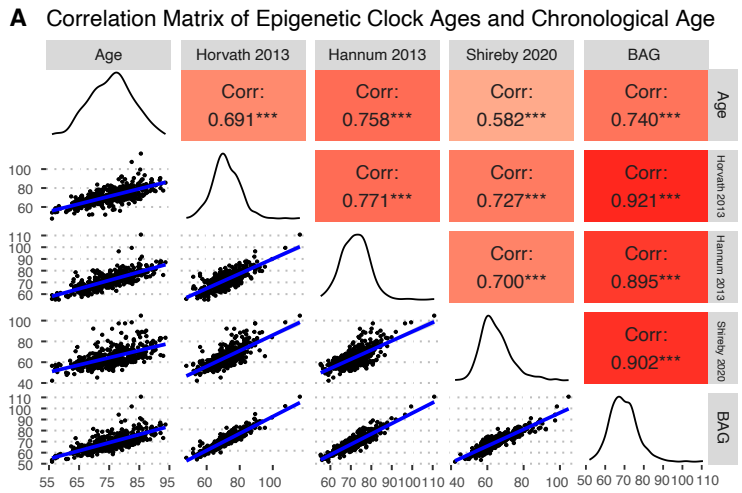


(a) Density plot of the signal detection frequency across all CpGs, which was 95.8% on average.

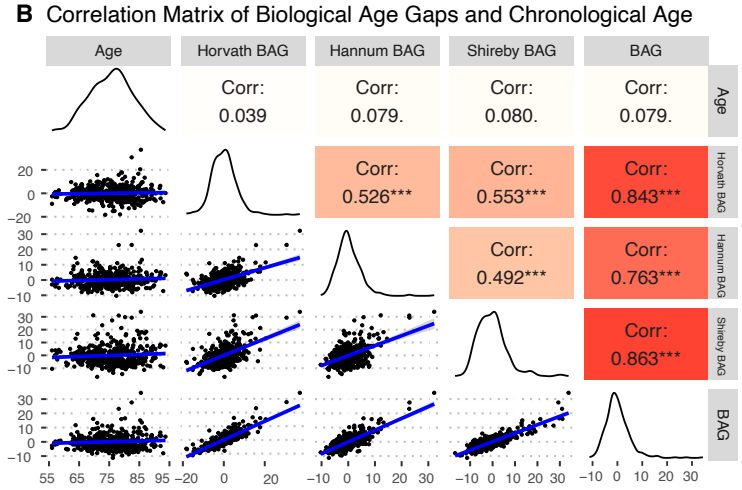


(b) Scatterplots of UMAP dimensions 1 and 2 by covariates.

Figure 3.3: **UMAP of processed DNA methylation data.** DNA methylation was assayed using the Illumina Infinium Human Methylation EPIC v1.0 BeadChip. Raw intensity files were processed using the *SeSAMe* R package.



(a) Correlation matrix of chronological age; epigenetic clock ages from three clocks: Horvath 2013, Hannum 2013, and Shireby 2020; and the average clock age. All correlations are significant. The residuals of each clock age regression against chronological age (first column) are the clock BAGs.



(b) Correlation matrix of chronological age, clock BAGs, and the average clock BAG. BAG and chronological age are not correlated. Clock BAGs correlated modestly with each other and highly with the average.

Figure 3.4: **Covariance heatmap of epigenetic clock measures.** Correlation matrices including all study participants with DNA methylation data ($n = 448$) illustrating the high covariance among epigenetic clocks. The upper triangle is a heatmap with the Pearson R value and significance level (*** = $p < 0.001$) between each pair of age measures. Lighter red indicates lower correlation, approaching 0, and darker red indicates higher correlation, approaching +1. The lower triangle shows a scatterplot of each pair of age measures with a linear regression line in blue. Density plots of each measure are along the diagonal.

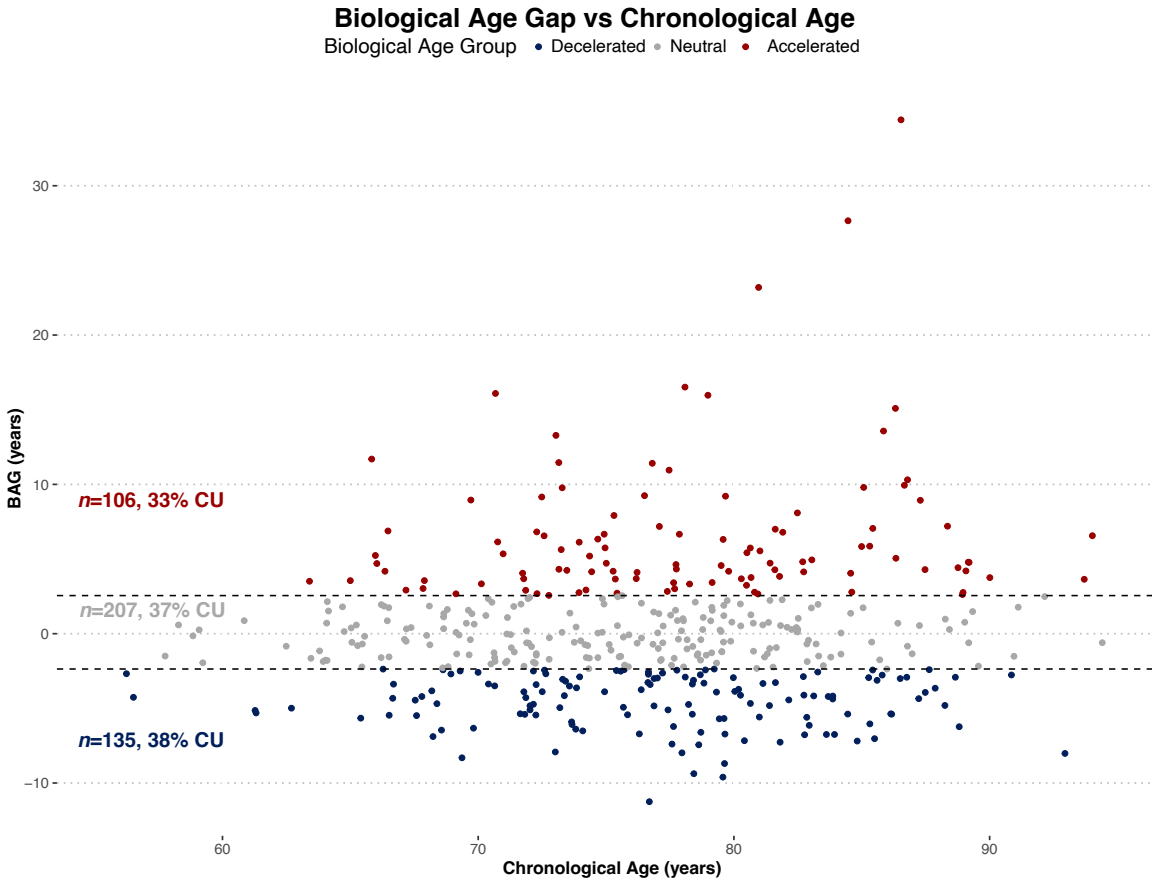


Figure 3.5: **Classification of participants by biological age group.** Participants with DNA methylation ($n = 448$) were classified into biological age groups based on their biological age gap averaged across clocks (BAG). Scatterplot of BAG versus chronological age, colored by biological age group, with a solid linear regression line in black and dashed lines showing the ± 0.5 SD cutoffs: Accelerated (top, red) Decelerated (bottom, blue), Neutral (center, green). Sample sizes and % CU for each group are provided in text. There is no correlation between BAG and chronological age, suggesting that BAG likely captures additional variance encapsulating biological aspects of aging.

3.2.3 Neuroimaging measures of cortical structures

The participants in this study were enrolled across multiple phases of ADNI, namely ADNI1 and ADNIGO/2. Structural T1-weighted MRI acquired at 1.5 Tesla (T) or 3T through ADNI were processed and quantified using the FreeSurfer 5.1 cross-sectional pipeline (Fischl, 2004) to yield regional cortical thickness measures and volumes. Statistical harmonization methods were necessary to account for scanner differences affecting acquisition and to remove associated batch effects. The ComBat family of harmonization methods is frequently used to correct for batch effects in neuroimaging data (Fortin et al., 2018). ComBat uses Empirical Bayes models to harmonize means and variances of neuroimaging measures across batches. We applied ComBat to harmonize cortical thickness and volume measures at the ROI level using the neuroHarmonize package (Pomponio et al., 2020) in Python version 3.10.4. The batch variable was defined as the magnetic field strength of the scanner (either 1.5T or 3T). Prior and observed distributions by batch are illustrated in Figure 3.6. As established by prior studies (Fortin et al., 2018; Pomponio et al., 2020), the harmonization model was fit to all CU individuals and applied to symptomatic individuals. The effects of biological covariates, namely age, sex, and diagnosis, were preserved. Boxplots of cortical thickness and volume before and after harmonization in CU individuals are shown in Figure 3.7.

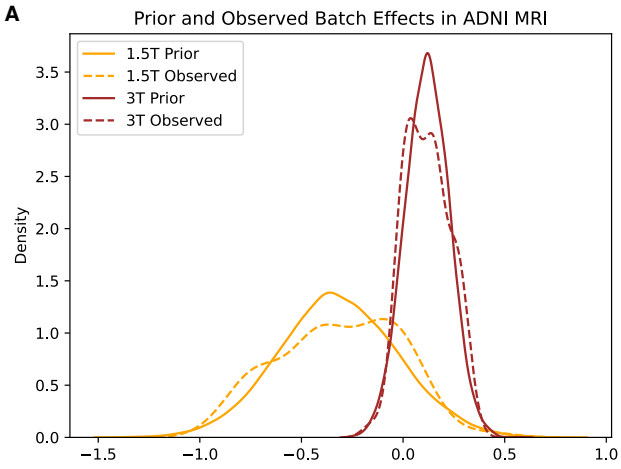


Figure 3.6: Density plots of prior (dashed) and observed (solid) distributions of batch effects for each batch, i.e., scanner field strength, with 1.5T shown in orange and 3T shown in brown.

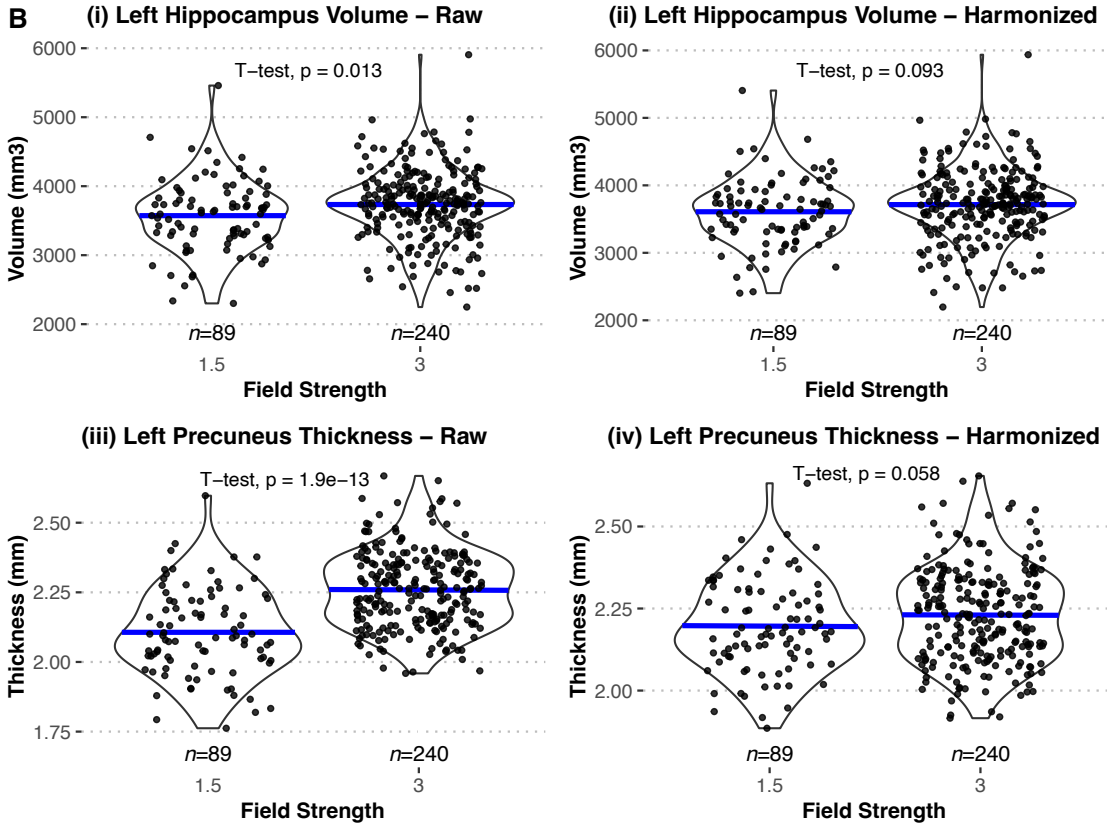


Figure 3.7: **Harmonization of Neuroimaging Thickness Measures.** Cross-sectional thickness measures derived from MRI were harmonized to adjust for batch effects. The batch variable was defined as scanner field strength, which was either 1.5 Tesla (T) or 3T. The ComBat harmonization model (neuroHarmonize Python package) was fit to all CU individuals, preserving the effects of age and sex. The fitted model was then applied to symptomatic individuals. We present violin plots of left hippocampal volume (i and ii) and thickness measures in left precuneus (iii and iv), which is one of the cortical signature brain regions in AD. Plots show all CU individuals ($n = 329$). Blue crossbars indicate the means of each distribution. Raw thickness measures (left, i and iii) are significantly different across batches, while harmonized thickness measures (right, ii and iv) are no longer significantly different across batches.

Neuroimaging Cortico-Medial Temporal index (CoMeT)

To quantify the relative involvement of cortex to MTL as a continuous measure, we defined the Cortico-Medial Temporal index (CoMeT) as follows. First, we identified brain regions of interest (ROIs) by approximately mapping neuroimaging signatures of AD (Bakkour et al., 2009; Dickerson et al., 2009; Möller et al., 2013; Touroutoglou et al., 2023) to the neuroanatomically-defined ROIs in the Desikan-Killiany-Tourville atlas (Desikan et al., 2006). Composite structural measures for cortex and MTL were computed by averaging bilateral ROIs across the respective brain areas (see Fig 3.1 for the list of ROIs). We then calculated w-scores by adjusting composite structural measures for age and sex (or age and intracranial volume for volumetric measures) using the full CU cohort as reference (see Table 1). Finally, CoMeT was calculated by subtracting the MTL composite w-score from the cortical composite w-score. A mathematical representation of the CoMeT calculation is provided in Equation 3.1. We define the "cortical" composite as comprising all non-MTL regions, aligning with established literature discussing the "corticolimbic" distribution of pathology (Murray et al., 2011; Kouri et al., 2024) i.e., cortical-predominant versus limbic-predominant or MTL-predominant presentations.

$$CoMeT = \frac{1}{n}\sum_{i=1}^{n} W_{x_i} - \frac{1}{m}\sum_{j=1}^{m} W_{y_j} \tag{3.1}$$

where:

- W_{x_i} is the age- and sex-adjusted w-score of the bilateral mean structural measure in cortical region x_i, referenced to a control cohort.
- W_{y_j} is the age- and sex-adjusted w-score of the bilateral mean structural measure in MTL region y_j, referenced to a control cohort.
- n is the number of cortical regions.
- m is the number of MTL regions.

3.2.4 Neuropsychological Measures by Cognitive Domain

Cognitive domain scores encompassing executive function, visuospatial ability, language, and memory were sourced from the Alzheimer's Disease Sequencing Project – Phenotype Harmonization Consortium (ADSP-PHC) (Consortium, 2021), following harmonization protocols developed by at University of Washington (Mukherjee et al., 2022). Given the variability in neuropsychological test batteries across studies and cohorts, cognitive phenotypes needed to be harmonized to allow for the integration of data from diverse sources by calibrating individual tests to a common metric, thereby supporting direct comparison across cohorts. Briefly, composite scores for each cognitive domain were generated using advanced factor-analysis methods. These techniques were co-calibrated across ten large-scale cohort studies, including ADNI, and the resulting scores were standardized to a unified scale. We then further w-scored these harmonized composite scores to control for age and sex effects using the CU cohort as reference, as with the neuroimaging measures.

The memory domain score provided by the PHC integrates a range of memory-related tasks, including both recall and recognition, which may dilute its specificity for MTL-dependent processes (Crane et al., 2012). For this project, a more focused memory score emphasizing recall performance would provide a purer measure of episodic memory and better align with our goal of mapping domain-specific neurocognitive changes to focal neuroanatomical regions. Therefore, we derived a new composite score using the Rey Auditory Verbal Learning Test (RAVLT) (Rey, 1964; Taylor, 1959) in ADNI (Crane et al., 2012) to represent a purer measure of episodic memory . The RAVLT emphasizes delayed free recall accuracy while penalizing intrusion errors. Our composite measure integrates the number of correctly recalled words after a delay (AVDELTOT), subtracts intrusion errors (AVDELERR2), and rescales the composite metric by adding the maximum test score of 15 (see Equation 3.2), which yields a more nuanced index of memory performance and better isolates MTL-dependent recall processes than the broader PHC harmonized memory composite. Finally, this episodic memory score was w-scored to account for age and sex, using the CU cohort as reference.

$$AVDELTOT + 15 - AVDELERR2 \tag{3.2}$$

Cognitive CoMeT

In alignment with the neuroimaging CoMeT framework, we conceptualized executive function, visuospatial ability, and language as "cortical-centric" cognitive domains, while memory was considered to reflect MTL function. Cognitive CoMeT scores were then computed by subtracting the episodic memory domain w-score (i.e., MTL-centric) from each of the cortical domain w-scores, resulting in three cognitive CoMeT indices: executive-memory, visuospatial-memory, and language-memory. These scores capture the relative integrity of cortical versus MTL-associated cognitive performance. We subsequently evaluated the associations between these cognitive CoMeT measures and BAG to investigate whether epigenetically informed markers of brain aging correspond to domain-specific patterns of cognitive vulnerability.

3.2.5 Statistical analysis

To study the association between CoMeT and BAG as continuous traits, we performed a correlation analysis using Pearson's correlation coefficient. Wilcoxon rank-sum tests were used to statistically compare CoMeT across symptomatic individuals with decelerated or accelerated BAGs as well as younger and older symptomatic individuals. Effect sizes were computed as rank-biserial correlations and interpreted using established guidelines (Cohen, 1992, 2013). Chi-squared tests were conducted to assess the association between biological age groups and APOE $\varepsilon 4$ alleles. Since all measures were adjusted for chronological age and sex, we do not additionally covary for chronological age or sex in our statistical analyses. All analyses were conducted in R version 4.4.0 using R Studio.

3.3 Results

3.3.1 Study population characteristics

Cross-sectional characteristics of the study participants by cohort are shown in Table 3.1. The CU and symptomatic cohorts were comparable for age, sex, and education level. We observed median CDRSB values of 0 for CU and 3 for symptomatic individuals. Among symptomatic individuals with MCI or dementia due to AD, the median CDRSB scores were slightly higher (1.5 and 5, respectively). The time between DNA methylation and the nearest MRI was less than 1 year on average among participants with both measures. Symptomatic individuals showed a higher

frequency of carrying at least one APOE $\varepsilon 4$ allele (68.5%) compared to CU individuals (21.0%). Chi-squared tests showed no significant differences by biological age groups across individuals with 0, 1, or 2 APOE $\varepsilon 4$ alleles among symptomatic individuals, although these differences were significant in CU individuals ($\chi^2 = 11.521$, $p = 0.021$), see Table 3.2.

Characteristic	**Cognitively Unimpaired**	**Symptomatic**
N	329 (163 with DNAm)	682 (285 with DNAm)
Clinical Diagnosis	–	MCI (378, 170 with DNAm) Dementia (304, 115 with DNAm)
Age at MRI (years)	74.84 (6.96)	75.57 (7.43)
Sex (% female)	50.8%	41.3%
APOE4 (% carrier)	21.0%	68.5%
CDRSB	0 (0)	3 (3.5)
Education (years)	16 (3)	16 (4)
BAG (years)	0.21 (5.69)	0.03 (4.42)
MRI to DNAm (years)	-0.98 (1.55)	-0.64 (1.36)
Biological Age Groups (%)	Accelerated (22.1%) Neutral (46.6%) Decelerated (31.3%)	Accelerated (24.6%) Neutral (46.0%) Decelerated (29.5%)

Table 3.1: **Demographics and clinical information of the cross-sectional study participants by cohort.** All participants have MRI and amyloid status. The "Cognitively Unimpaired" (CU) cohort includes all amyloid-negative, CU individuals. The "Symptomatic" cohort includes all amyloid-positive individuals with either MCI or dementia due to AD. The interval between MRI and DNAm as well as frequencies by biological age groups are provided for the subset of participants with DNA methylation (448 total, $n = 163$ for CU, $n = 285$ for symptomatic). All values are in specified units in mean (SD) format or as percentages except for CDRSB and education, which are in median (IQR) format. APOE4 = Apolipoprotein $\varepsilon 4$ allele. BAG = Biological Age Gap. CDRSB = Clinical Dementia Rating scale Sum of Boxes. DNAm = DNA methylation. IQR = Inter Quartile Range. MCI = Mild Cognitive Impairment.

3.3.2 Validation of CoMeT by chronological age groups in symptomatic individuals

We first explored how CoMeT associates with chronological age and whether our results aligned with established findings (see Fig 3.8). Overall, CU individuals had mean structural w-scores of 0 in cortex and MTL as expected given that these measures were derived from the CU cohort. Symptomatic individuals had significantly atrophied structures compared to CU individuals in both cortex and the MTL ($p < 0.001$). There was a significant correlation between CoMeT and chronological age among all symptomatic individuals, including participants in the "average" group between 65 and

	(a)			(b)		
Group / Alleles	0	1	2	0	1	2
Decelerated	48	3	0	30	41	13
Neutral	54	20	2	42	66	23
Accelerated	30	6	0	22	40	8

Table 3.2: χ^2 tests of biological age groups (decelerated, neutral, and accelerated) by the number of APOE $\varepsilon 4$ alleles (0, 1, or 2) in **(a)** all CU individuals with DNA methylation ($n = 163$) revealed a significant association between the two variables, $\chi^2 = 11.521$, df=4, $p = 0.021$. **(b)** all symptomatic individuals with DNA methylation ($n = 285$) revealed no significant association between the two variables, $\chi^2 = 1.9665$, df=4, $p = 0.742$.

80 years of age ($n = 682$, Pearson $R = 0.15$, $p < 0.001$). Among individuals with dementia only, this correlation was quantitatively stronger ($n = 304$, Pearson $R = 0.27$, $p < 0.001$). Categorical analyses of chronological age groups revealed that younger symptomatic individuals had significantly lower cortical thickness compared to older symptomatic individuals with a medium effect size ($W = 4598$, $p < 0.001$, rank-biserial correlation$= 0.31$) but no significant difference in MTL structure ($W = 6235$, $p = 0.430$, rank-biserial correlation$= -0.06$). This emphasizes the importance of the relative difference (CoMeT) as opposed to cortical or MTL structures individually. As expected, CoMeT was significantly lower in younger individuals compared to older individuals within the symptomatic cohort with a small effect size ($W = 4751$, $p < 0.001$, rank-biserial correlation$=-0.29$), reflecting greater cortical neurodegeneration relative to MTL in younger individuals. There were no associations between chronological age and CoMeT in the CU or MCI cohorts (see Figure 3.9).

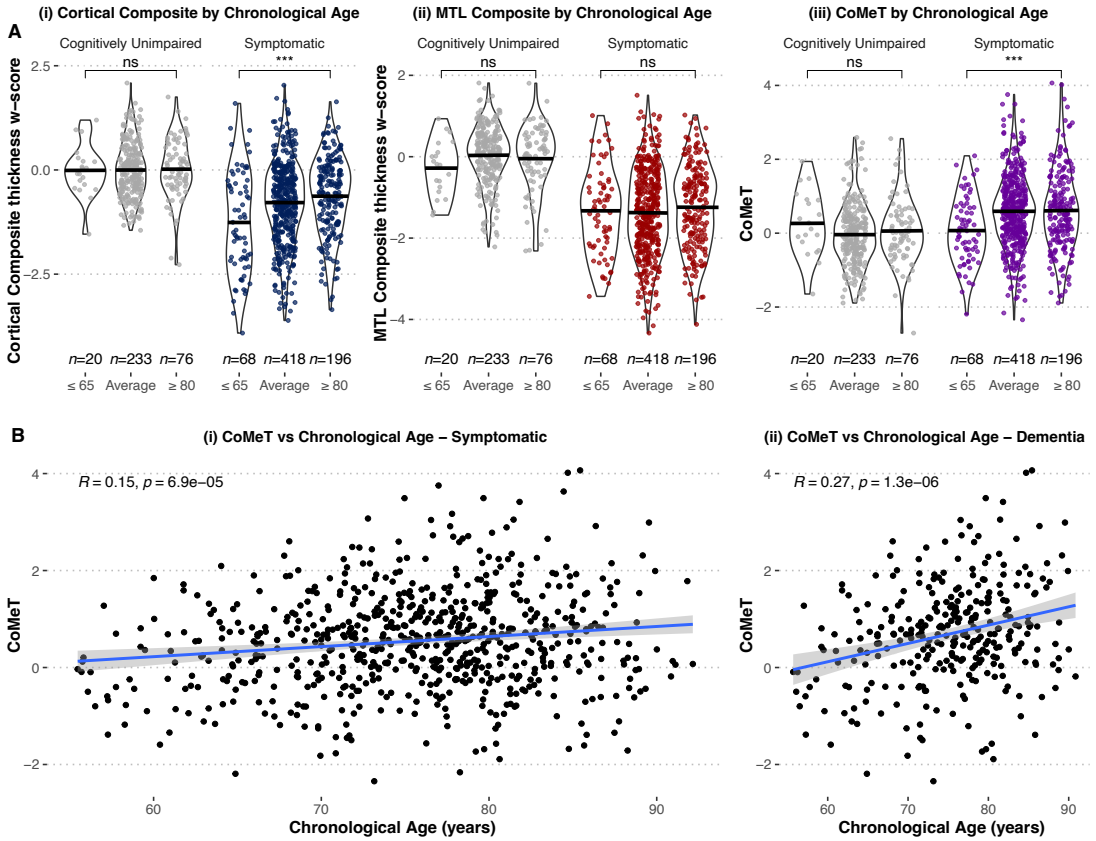


Figure 3.8: **(A)** Violin plots of (i) cortical composite thickness w-scores (ii) MTL composite thickness w-scores (iii) CoMeT by chronological age groups (younger, $\leq$ 65 years and older, $\geq$ 80 years) in CU individuals (left, grey) and symptomatic individuals (right, in color). Black crossbars show the mean of each distribution. CU individuals are all centered at z̃ero mean after adjusting for age and sex to remove confounding effects due to healthy aging. Overall, symptomatic individuals show lower thickness in both cortex and MTL than CU individuals ($p < 0.001$). Younger symptomatic individuals show significantly lower cortical thickness but no difference in MTL thickness compared to older symptomatic individuals. Critically, younger symptomatic individuals show significantly lower CoMeT scores than older symptomatic individuals ($p < 0.001$), indicating greater cortical involvement relative to MTL. **(B)** Scatterplots of CoMeT versus chronological age with blue regression lines in (i) all symptomatic individuals, i.e., including ages 65–80 (Pearson $R = 0.15$, $p < 0.001$) and (ii) dementia subgroup (Pearson $R = 0.27$, $p < 0.001$)

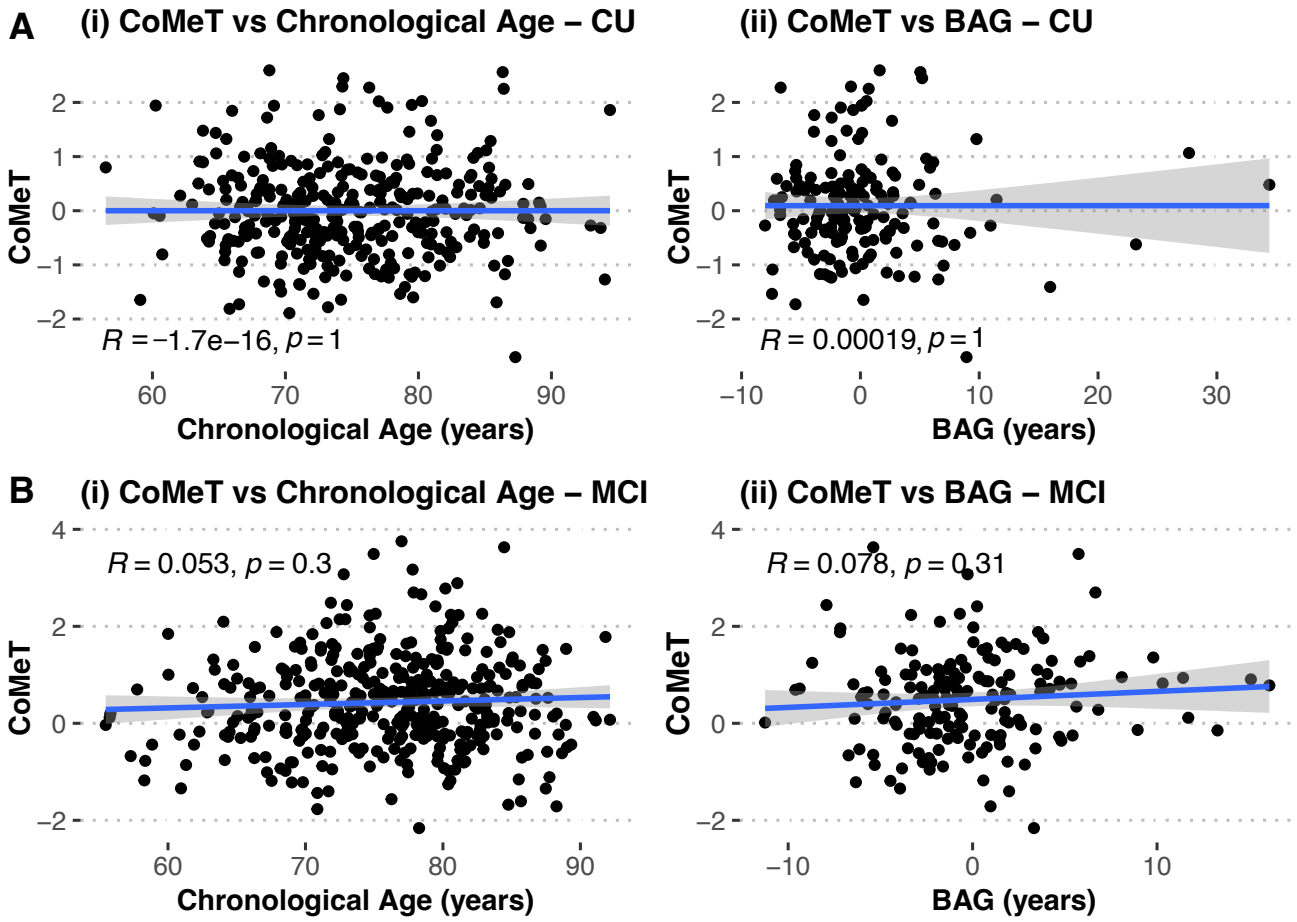


Figure 3.9: **Association of CoMeT with Age and BAG in CU and MCI individuals.** Scatterplots of CoMeT in **(A)** CU individuals and **(B)** MCI individuals by (i) chronological age ($n = 329$, CU and $n = 378$, MCI) and (ii) BAG ($n = 163$, CU and $n = 170$, MCI), with blue regression lines overlaid. CoMeT was not significantly correlated with chronological age or BAG in neither CU nor MCI individuals.

Across all four cognitive domains, CU individuals again had mean w-scores of 0. Symptomatic individuals had significantly poorer cognition across domains compared to CU individuals ($p < 0.001$). Younger symptomatic individuals were more impaired than older symptomatic individuals in all domains, with the difference in mean w-scores between these groups reaching traditional significance levels for executive function, visuospatial, and episodic memory domains but not language. Table 3.3 provides the mean w-scores for each cognitive domain across chronological age groups. In contrast to the neuroimaging findings, we did not observe significant associations between chronological age and any cognitive CoMeT measures in the symptomatic cohort. Again, there were no associations between chronological age and cognitive CoMeT in the CU or MCI cohorts.

Domain	**Younger ($\leq$ 65)**	**Average**	**Older ($\geq$ 80)**	**p-value**
Executive Function	-2.78 (2.26)	-1.95 (1.62)	-1.63 (1.31)	< 0.001
Language	-1.80 (1.44)	-1.61 (1.39)	-1.45 (1.12)	0.127
Visuospatial	-1.64 (2.15)	-1.08 (1.61)	-0.94 (1.43)	0.009
Memory	-2.68 (1.59)	-2.34 (1.73)	-2.00 (1.77)	0.011

Table 3.3: **Cognition in symptomatic individuals by chronological age groups.** Harmonized measures from the PHC were adjusted for age and sex relative to the CU cohort to obtain the executive function, language, and visuospatial w-scores. The memory w-score is a composite derived from the RAVLT and adjusted for age and sex relative to the CU cohort, representing a purer measure of episodic memory (as described in the Methods, Section 3.2.4). The w-scores for each cognitive domain are provided in mean (SD) format.

3.3.3 Association of CoMeT with biological aging in symptomatic individuals

Analyses using continuous measures revealed a positive correlation between BAG and CoMeT ($n = 285$, Pearson $R = 0.13, p = 0.023$), consistent with deceleration relating to relatively greater cortical involvement and acceleration relating to greater MTL involvement (see Fig 3.10). Categorical analyses revealed that CoMeT was significantly lower in individuals with decelerated BAG compared to individuals with accelerated BAG in the symptomatic cohort, with a large effect size ($W = 2314$, $p = 0.023$, rank-biserial correlation$=-0.98$). Therefore, individuals with decelerated BAG show a less typical, more cortical-predominant disease pattern compared to individuals with accelerated BAG. CoMeT scores did not associate with BAG in the CU or MCI cohorts (see Figure 3.9). We additionally performed a subgroup analysis of symptomatic individuals with dementia

only, i.e., excluding MCI (see Fig 3.10) that also revealed a significant correlation between CoMeT and BAG ($n = 115$, Pearson $R = 0.19, p = 0.048$), indicating relatively greater cortical involvement in individuals with decelerated BAG compared to individuals with accelerated BAG. While CoMeT scores were numerically lower in individuals with decelerated BAG compared to individuals with accelerated BAG, this result was not significantly different despite a large effect size, likely due to lower power and smaller sample sizes ($W = 445, p = 0.214$, rank-biserial correlation=-0.98).

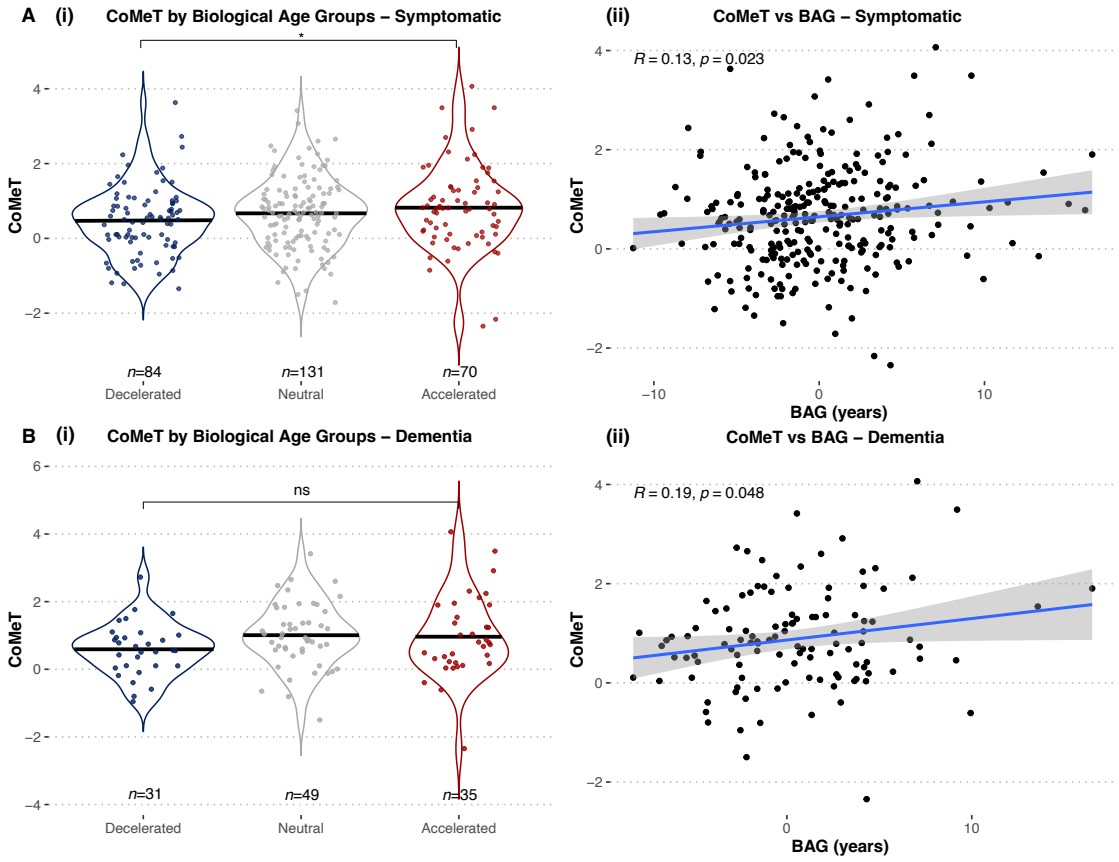


Figure 3.10: **Association of BAG with CoMeT**. Violin plots of CoMeT by biological age groups in **(A, i)** all symptomatic individuals and **(B, i)** dementia-only subset. Symptomatic individuals with decelerated BAG have significantly lower CoMeT scores than those with accelerated BAG ($W = 2314$, $p = 0.023$, rank-biserial correlation= -0.98). This difference is not significant in the dementia subgroup, likely due to power constraints, but maintains a large effect size ($W = 445$, $p = 0.214$, rank-biserial correlation= -0.98). Scatterplots of CoMeT versus BAG with blue linear regression lines in **(A, ii)** all symptomatic individuals (Pearson $R = 0.13$, $p = 0.023$) **(B, ii)** dementia only subset (Pearson $R = 0.19$, $p = 0.048$).

3.4 Discussion

3.4.1 Interpretation of Results

Our results support established findings reporting an association between younger age of onset and disproportionately more cortical presentations of AD (Dickerson et al., 2017). As hypothesized, our results also demonstrate a significant association between biological aging and the spectrum of cortical to MTL atrophy in AD. Specifically, symptomatic individuals with decelerated BAGs exhibit significantly lower CoMeT scores, reflecting greater cortical involvement, compared to those with accelerated BAGs. Moreover, the mean CoMeT score of symptomatic individuals with "neutral" BAGs fell in between the means of their decelerated and accelerated counterparts. This suggests that the relationship between biological aging and atypical neurodegenerative pattern exists across a continuum. Indeed, CoMeT and BAG correlate when examined across all symptomatic individuals. This association appears at least quantitatively stronger when limited to individuals with dementia due to AD (i.e., excluding those with MCI due to AD). We observe no significant association in the MCI subgroup, suggesting that the manifestations of biological age on atypicality may emerge more prominently with later disease progression. It is also worth noting that since the neutral group represents the largest proportion of individuals and has CoMeT scores between the two extreme groups, one might also consider the biologically accelerated group as having a more limbic-predominant "atypical" presentation, as described in the literature, while the biologically decelerated group has more cortical-predominant, atypical cases (Murray et al., 2011; Ferreira et al., 2020).

These results align with the effects of chronological age on disease presentation as frequently noted in prior reports (Swerdlow, 2007; Dickerson et al., 2017) and suggest that biological age captures additional variance. Therefore, biological age may be a core factor modulating the topography of AD pathology and may further explain some cases of discordance between chronological age and degree of atypicality, for instance, older-onset cases with more cortical involvement.

In contrast to the structural neuroimaging findings, our analysis using cognitive domain measures mapped to cortical or MTL regions did not reveal significant associations between BAG and CoMeT.

When stratifying by chronological age, however, we observed that younger-onset individuals were generally more cognitively impaired than their older-onset counterparts, particularly in the executive function domain. This aligns with prior reports indicating greater impairment in EOAD relative to LOAD (Koedam et al., 2010; Sirkis et al., 2022). Interestingly, while memory impairment was more pronounced in younger individuals, structural measures did not reflect this difference. This could be due to recruitment bias in the ADNI cohort: specifically, younger individuals may have entered the study with prominent memory complaints, leading to a sample enriched for memory dysfunction and thus reducing the relative salience of corticolimbic mismatch captured by CoMeT. Additionally, the lack of association between BAG and cognitive CoMeT may reflect differences in the temporal dynamics of aging-related changes. While structural decline may accumulate gradually and reflect broader neurobiological aging processes, cognitive measures are more susceptible to inter-individual variability, compensation, and floor effects in symptomatic stages. Furthermore, the spatial mapping of cognitive domains onto specific cortical and MTL regions may not fully capture the distributed network interactions underlying cognitive dysfunction in AD, limiting the sensitivity of CoMeT to detect biologically age-related disparities. Future studies should aim to replicate these findings in more diverse cohorts with broader recruitment criteria or with greater focus on younger individuals, such as the Longitudinal Early-onset Alzheimer's Disease Study (LEADS) cohort (Touroutoglou et al., 2023; Vandekar et al., 2021, 2023). Additionally, leveraging network measures of cognition-structure mappings may improve our ability to characterize the relationship between biological aging, brain structure, and cognition.

Biological aging may also interact with other factors to influence AD presentation. Given that MTL tau deposition is commonly associated with healthy aging (Braak and Braak, 1991), older adults may already exhibit an MTL-localized epicenter driving the clinical manifestations of AD. This predisposition could result in a more limbic-predominant presentation as the disease progresses. As our findings suggest, individuals with accelerated BAG tend to present with higher CoMeT scores, indicating greater MTL involvement.

As with patterns of atrophy, neuropathological studies have also consistently shown that the dis-

tribution of tau pathology is not uniform across patients (Arezoumandan et al., 2022). While most individuals exhibit a mix of MTL and cortical involvement, some exhibit more "atypical" burdens with disproportionately heavy burden isolated in the MTL or in cortical regions (Murray et al., 2011; Kouri et al., 2024). These patterns are often linked to differences in clinical symptoms, age of onset, and the presence of co-pathologies (Ossenkoppele et al., 2020). For example, AD cases with more cortical tau accumulation tend to exhibit earlier disease onset and atypical cognitive symptoms, such as language or visuospatial deficits, rather than the more classic memory impairments seen with MTL-centered pathology (Frontzkowski et al., 2022). Some studies have suggested that certain genetic factors, such as APOE4 genotype, might influence whether tau pathology predominantly affects the cortex or the MTL (Therriault et al., 2020; Steward et al., 2023).

The corticolimbic index (CLix) (Kouri et al., 2024) is a recently developed neuropathological measure derived from ex vivo brain tissue analysis that quantifies the relative burden of tau pathology in the cortex versus the MTL. CLix was developed to explore corticolimbic vulnerability, conceptualizing tau pathology as a continuous trait that varies across individuals, thus providing a spectrum of disease topography from predominantly cortical involvement to predominantly MTL involvement. CoMeT is analogous to CLix as a continuous measure that, in contrast, can be assessed with structural MRI during life. Furthermore, CoMeT can be used in conjunction with other *in vivo* biomarkers, including amyloid- and tau-PET, CSF, plasma, and multiomics data to explore how different factors contribute to the corticolimbic distribution of atrophy. This allows for the dynamic tracking of pathology and its clinical correlates over time, facilitating future longitudinal studies that could elucidate the progression of AD and its variants.

Recent studies have highlighted that EOAD, often clinically labeled as "amnestic", may manifest with non-amnestic, dysexecutive presentations more frequently than previously recognized (Townley et al., 2020; Corriveau-Lecavalier et al., 2024; Putcha et al., 2025). Notably, analyses of the LEADS cohort demonstrated that a significant proportion of individuals under 65 diagnosed with "amnestic AD" exhibited prominent dysexecutive symptoms rather than memory impairment (Putcha et al., 2025). Our findings of greater cortical relative to MTL involvement in individuals

with decelerated biological aging may align with these emerging clinical observations, suggesting that some early-onset, MTL-sparing patients within the ADNI dataset could indeed reflect a dysexecutive phenotype. This highlights the importance of considering biological aging in clinical phenotypic variability, especially in the context of structural neuroimaging findings in AD. While this paper focuses on pathological and neurodegenerative heterogeneity in this paper, future research could expand on our work by studying associations between biological aging and heterogeneity in clinical presentation.

3.4.2 Limitations and Future Work

There are a few limitations in this study that may also be overcome through future research. Firstly, we utilized DNA methylation data from blood rather than brain tissue. Since epigenetic modifiers of gene expression are tissue specific, it would be ideal to study methylation from the cortex and MTL to understand these age effects more directly. However, this type of data is very rare and previous studies have shown a high correlation among predicted outcomes from clocks trained on multiple tissues and applied to blood methylation data (Horvath, 2013; Shireby et al., 2020; De Lima Camillo et al., 2022). For this reason, we suspect that leveraging other accessible peripheral tissues, such as salival or buccal cells, would reflect similar associations capturing systemic aging effects that impact neurodegenerative pattern. Peripheral tissues also allow for *in vivo* estimation of biological age, as opposed to brain tissue which would necessitate *ex vivo* analysis. Future studies combining multi-tissue epigenetic profiling, possibly including cerebrospinal fluid, buccal swabs, or postmortem brain regions, will be critical for teasing apart shared vs tissue-specific aging signals, and for refining models of vulnerability to neurodegeneration in the brain.

Our study population had limited representation of individuals under 65 years of age and was constrained due to the incorporation of multi-modal data including DNA methylation, imaging, and amyloid status. Additionally, our sample was primarily comprised of highly educated participants who self-identified as White (Sreepada et al., 2025). This homogeneity is crucial to address for broader applicability of these results. Future work is therefore necessary to evaluate the generalizability of these findings and the application of biological aging measures in more diverse datasets

across the life-course. Additionally, development of new epigenetic clocks tailored to aging populations and accounting for co-pathologies may better inform our understanding of age-related disease trajectories, given that predictors of neurodegeneration differ between CU older adults who subsequently become impaired (Armstrong et al., 2019).

In this study, we interpreted the residual from a linear regression of predicted biological age on chronological age, namely biological age gap (BAG), which captures individual-level deviation from normative aging, allowing us to quantify accelerated or decelerated brain aging relative to peers. The current analysis is grounded in the assumption that BAG is a continuous marker of brain aging severity, not necessarily tied to discrete diagnostic thresholds. Importantly, we applied subtype-informed biological age models to capture how deviation from typical aging trajectories differs by regional predominance (Sreepada et al., 2025). Thus, rather than relying on categorical comparisons, we aimed to exploit individual-level variation to reflect continuous processes of neurodegeneration.

While this approach is simple and interpretable, we acknowledge that an ordinary regression framework assumes linear relationships and that the residual-based approach does not fully separate variance due to age from variance due to regional patterns such as relative cortical- or MTL-predominance. In theory, a 2×2 factorial design or a Bayesian hierarchical model could more explicitly model these dimensions by treating regional neurodegenerative subtype and age category as partially independent factors. Specifically, we could incorporate subtype-by-age interaction terms or leverage multilevel models that treat individuals nested within subtypes while estimating age-related trajectories separately. This might allow us to test for interaction effects or shared variance structures in a more flexible manner.

Regarding variance as a function of age, we did examine whether BAG variance changes across the age spectrum. The variance of residuals is largely stable across chronological age, consistent with prior findings that biological age acceleration measures, when derived as residuals from regression against age, exhibit homoscedasticity across age ranges (Tong et al., 2024). In our data, there was no significant association between age and the variance of BAG, supporting the validity of applying a fixed SD threshold across the full age range.

We stratified participants into biological age groups using a ±0.5 SD threshold from the mean BAG distribution. We selected the ±0.5 SD threshold *a priori* to maximize statistical power while retaining interpretability and ensuring balanced group sizes across accelerated, neutral, and decelerated groups (Sreepada et al., 2025). However, we also tested alternative thresholds (e.g., ±0.75, 1 SD) to evaluate robustness. Across these sensitivity analyses, the main effects and directionality of results were preserved. As expected however, increasing the threshold reduced sample size, which slightly attenuated statistical significance but did not reverse any findings. In stratified subsamples by age tertiles, the associations between BAA and outcomes of interest persisted within each age group, further confirming that the effects of BAA are not confounded by age range. Nevertheless, we acknowledge that residual-based stratification does not completely remove age-related information, particularly in cases where regional atrophy patterns correlate with both age and subtype. While BAG is orthogonal to age by construction, subtle age-related structure may still exist due to interactions between age, subtype, and disease severity.

Together, these analyses demonstrate that BAG explains unique variance in brain structure and cognitive function that is not reducible to chronological age. This supports the core rationale of using BAG as a biomarker, since it reflects accelerated or decelerated neurobiological aging processes after accounting for the effects of chronological age.

This is the first study, to our knowledge, investigating the role of biological aging in driving heterogeneity in the neuroanatomic distribution of AD. We define a novel continuous measure, CoMeT, to operationalize and quantify the spectrum of neurodegenerative patterns. As hypothesized, we observed that decelerated BAG is associated with a less typical disease presentation with relatively greater cortical involvement, while accelerated BAG is linked to relatively greater MTL involvement. Future studies involving pathway and enrichment analyses may identify specific epigenetic modifiers and genetic variants driving this heterogeneity.

We conclude that biological aging modulates atypical disease presentations in AD in conjunction with and beyond chronological age. This study suggests that fundamental effects of the aging process contribute to the neuroanatomic presentation of AD, and our work contributes to a better

understanding of the biological basis and mechanisms underlying AD heterogeneity.

3.5 Chapter Summary

The findings presented in this chapter build directly on prior literature linking atypical AD presentations to regional vulnerability and epigenetic mechanisms, discussed in Section 2.3. Briefly, previous studies have highlighted how atypical neurodegenerative patterns, such as disproportionate cortical versus medial temporal lobe involvement, may stem from intrinsic molecular features of intra-individual variability in region-specific neurodegenerative vulnerability. Emerging evidence from cross-tissue blood-brain methylation studies, also comprehensively reviewed in Chapter 2, Section 2.4, demonstrates partial concordance between blood and brain epigenetic signatures, supporting the use of peripheral markers to infer central nervous system processes. By leveraging blood-based DNA methylation to estimate biological measures of aging, this study provides new evidence that biological aging is associated with neuroanatomic patterns of disease expression beyond what is explained by chronological age alone. These results suggest that systemic aging processes, captured epigenetically, may play a causal or modulatory role in the emergence of atypical AD phenotypes. This chapter thus extends the framework of regional vulnerability to incorporate biological aging as a cross-cutting factor, linking peripheral aging signals to spatially specific neurodegeneration. Chapter 4 builds on this foundation and aims to capture brain-relevant aging and neurodegeneration more precisely by developing blood-based epigenetic signatures of aging and MTL-focused neurodegeneration. This will further advance our understanding of the epigenetic basis of regional susceptibility in AD.

3.6 Acknowledgements

This work was supported by funding from the NIH T32 Program in Translational Neuroimaging of Alzheimer's Disease and Related Dementias (AG076411) and the Alzheimer's Disease Research Center (AG072979) at the University of Pennsylvania.

We thank the participants of ADNI for their contribution to this study. The ADNI data is shared without embargo through the Laboratory of Neuroimaging (LONI) Image and Data Archive, a secure research data repository. Scientists may obtain access to imaging, clinical, genomic (including

DNA methylation), and biomarker data for scientific investigation, teaching, or planning clinical research studies via application. Access is contingent on adherence to the ADNI Data Use Agreement and the publication policies. More information is available at the ADNI webpage. In addition, the analysis code and documentation for this chapter are available publicly on GitHub.

We also thank members of the Penn Bioinformatics in Neurodegenerative Disease Laboratory (BiND Lab), Penn Frontotemporal Degeneration Center (Penn FTDC), and Penn Image Computing and Science Laboratory (PICSL) at the Perelman School of Medicine, University of Pennsylvania, for their review and feedback on this chapter.

CHAPTER 4

BLOOD-BASED EPIGENETIC SIGNATURES AND BIOLOGICAL PATHWAYS UNDERLYING BRAIN AGING AND NEURODEGNERATION

4.1 Introduction

Neurodegeneration is a ubiquitous feature of normative aging and is observed to varying extents in nearly all individuals over the age of 50. It is also a key aspect of neuropathological changes observed in many neurological disorders, of which AD and related dementias are among the most common. Neurodegenerative diseases are often characterized by regional aggregation of proteins such as amyloid, tau, alpha-synuclein, or transactive response DNA-binding protein 43 (TDP-43). These protein aggregates interfere with normal cellular function and contribute to neuronal damage, resulting in downstream neurodegeneration. While the hallmark pathologies of many neurodegenerative diseases have been identified, the spatial distribution and inter-regional severity of neurodegenerative changes demonstrate considerable inter-individual variability that cannot be solely explained by chronological aging or the presence of co-pathologies. In particular, the mechanisms underlying the vulnerability of specific brain regions to neuropathology remain poorly understood.

The MTL is a hotspot for AD pathology and is among the earliest brain regions to exhibit amyloid and tau accumulation, resulting in regional neurodegeneration. Several structures frequently implicated in AD, such as the hippocampus and amygdala, are within the MTL. The hippocampus plays a critical role in memory and is particularly susceptible to early neurodegenerative processes. MTL atrophy is a characteristic feature of the typical amnestic presentation of AD and serves as a widely recognized in vivo biomarker of the disease. However, similar patterns of structural decline also occur as part of normative aging, albeit at a slower and more gradual rate. Longitudinal neuroimaging studies consistently demonstrate age-related volumetric reductions in MTL subregions, including the hippocampus and ERC, in cognitively unimpaired individuals. The convergence of structural brain changes in both healthy aging and AD complicates efforts to differentiate normative aging from early pathological processes. The biological mechanisms underlying aging that confer

vulnerability to neuropathology and structural changes in the MTL remain poorly understood. This highlights the need to identify additional factors that mechanistically link systemic aging to regional neurodegeneration observed in diseases.

One promising avenue to decipher these mechanisms is the study of epigenetic modifications, which have been recently studied with increasing frequency as measures of biological aging. Unlike alterations in the DNA sequence, epigenetic modifiers do not alter the genomic sequence, instead influencing how genes are expressed. Epigenetics may mediate the complex interplay between environmental exposures, lifestyle factors, and gene regulation throughout the lifespan. In particular, DNA methylation patterns are shaped by developmental and environmental factors and play important roles in brain development and disease by regulating gene expression. Mechanism: addition of a methyl group to the 5' cytosine at a phosphorylated cytosine guanine dinucleotide (CpG) regulates up or downregulation of the associated gene. Epigenetic markers have become an attractive area of study in the context of brain aging because they reflect cumulative environmental and lifestyle influences on the genome and may reveal how such influences contribute to neurodegeneration. In contrast to fixed genetic risk factors, epigenetic modifications are dynamic and potentially reversible, suggesting that they could be targets for future therapeutic intervention. Recent studies have also studied epigenetic reprogramming in stem cells and modifiable lifestyle factors such as diet, exercise, and treating existing medical conditions like blood pressure could naturally reprogram the epigenetic switch board regulating genetic expression. This suggests it may be possible to slow aging processes and thereby mitigate neurodegenerative risk.

Epigenetic modifiers, including DNA methylation, are highly tissue-specific and recent studies have documented differences between blood and brain methylation patterns. However, blood DNA methylation shows some promise as a surrogate indicator for brain-related phenotypes in humans, including cognitive function and neurological disease risk. For instance, epigenetic clock algorithms now provide an estimate of biological age from an individual's blood DNA methylation profile. Numerous studies have documented the association between accelerated biological age (biological age $>$ chronological age) and diffuse neurodegeneration, leading to downstream cognitive impairment.

Biological age acceleration has also been associated with greater risk of cognitive decline and progression to mild cognitive impairment (MCI) or AD in initially healthy, unimpaired individuals, as well as with greater brain atrophy in AD-vulnerable regions, such as cortical thinning in the ERC. A recent study assessed the relationship between accelerated vs decelerated biological aging with focal neurodegeneration in the cortex relative to MTL regions, suggesting that variations in epigenetic profiles might make an individual more prone to one neurodegenerative pattern vs the other which could have downstream implications on cognitive function (i.e., less typical, language or behavior or visuospatial issues or issues with executive function compared to a typical, primarily amnestic profile). These findings underscore the potential of peripheral epigenetic markers to predict and reflect brain health. Finally, the ability to measure epigenetic modifiers in vivo from peripheral tissues like blood offers significant practical advantages. Blood-based epigenetic markers can be obtained minimally invasively and longitudinally, allowing close monitoring of progression during life. This is especially promising in an era of precision medicine where disease modifying therapies are becoming available, and sensitive / specific biomarkers for patient stratification are becoming paramount.

Several analytical approaches and frameworks can be leveraged to probe the link between peripheral epigenetics and brain structure. Firstly, an epigenome-wide association study (EWAS) tests for differential methylation linked to a phenotype of interest, in this case structural brain differences, allowing for pinpointing of epigenetic markers and associated genomic loci where blood methylation reflects brain atrophy or preservation. Jia et al performed a large EWAS of subcortical volumes from the ENIGMA consortium, identifying few CpGs associated with hippocampal volume which explained 1% of the variance in this phenotype. This suggests that blood has limited ability to explain brain related phenotypes, but also opens the question of whether newer, more data-driven techniques than univariate statistical analysis might capture more variance and predictive power.

Another approach that has gained popularity in the past decade is the epigenetic clock, which is an ML algorithm that learns patterns of DNA methylation from epigenome-wide methylation profiles to estimate biological age by predicting age or an age-associated phenotype of interest. There

are a few generations of clocks, with first-generation clocks predicting age and second-generation clocks predicting age-related phenotypes. Residual approaches are applied to clock predictions to assess the discrepancy between the clock prediction and the observed phenotype, whether age or an associated phenotype. It is understood that clock predictions $>$ observed value reflects accelerated biological aging, while clock predictions $<$ observed value reflects decelerated biological aging. Milicic and colleagues comprehensively reviewed existing first and second-generation epigenetic clocks and observed that disproportionate biological aging, as measured by the clocks, is associated with hippocampal volume. This aligns with the findings discussed in Chapter 3, namely that accelerated aging is associated with greater MTL atrophy relative to cortex. Traditionally, Elastic Net models have been used for clock models because of their balance of the L1 and L2 penalties on parameter weights, effectively driving the weights of many parameters to zero and thereby selecting a few features that are predictive of the phenotype of interest. However, recent studies have also applied more complex ML algorithms such as XGBoost and RandomForest.

Brain age models are analogous to established first-generation epigenetic clocks in that they model chronological age and are trained on multivariate neuroimaging data as opposed to DNA methylation profiles. In other words, brain age measures complement methylation-based biological age indices by estimating an imaging-derived correlate of age based on brain structure. We hypothesized that a novel epigenetic clock could predict an imaging-derived measure of brain-based biological age, akin to previously established second-generation clocks predicting various age-associated indices derived from multimodal biomedical datasets, such as PhenoAge (Levine et al. 2018) and GrimAge (Lu et al 2020, Lu et al. 2022).

One of the issues with epigenetic clocks is the limited interpretability of CpGs selected by the models. Often, CpGs that are highly correlated and close to each other on the genome are selected at random by linear models which do not handle collinearity well. Recent approaches have leveraged deep neural networks for explainable AI models of biological age where CpG selection and information scores provide more information about why certain CpGs were selected. Another issue is that the most recent methylation arrays assay over 850,000 CpG sites along the epigenome, which leads to a stark

imbalance between the number of features and the sample size. This can lead to severe overfitting or underfitting of the model to the training dataset and reduce generalizability. For this reason, pre-selecting features using various classes of feature selection methods can help prioritize features to learn from and reduce the risk of overfitting. Recent studies have focused on implementing rigorous feature selection schemes prior to epigenetic clock model development (Doherty et al. 2023), comparing methods in different classes such as filters (univariate, regression-based, or ensembles), embedded selection, and transformative. These methods often tradeoff between interpretability, biological relevance, and predictive performance. In this study, we seek to maximize biological relevance to be able to trace back the relevance of identified CpGs to aging and neurodegenerative processes in the brain.

Upon selecting a subset of features through model development, the next step is to interpret the biological pathways implicated by the clocks, by performing gene set enrichment and pathway analyses using identified epigenetic candidate markers from EWAS or clock analyses. These are downstream statistical analyses that interpret biological significance by annotating CpGs along the genome, identifying linked genes, and examining whether the genes associated with significant methylation sites converge on biological pathways or gene networks. Using gene ontology enrichment and pathway analyses, we can uncover whether certain pathways are over-represented in the epigenetic signatures identified. This helps link methylation changes to underlying cellular and biomolecular processes.

To further bridge epigenetic findings with known biology, we can correlate methylation with gene expression profiles from blood to explore whether differential methylation might be associated with altered gene expression in pathways relevant to neurodegeneration. This integrative approach provides a holistic view of how epigenetic variation in peripheral tissue corresponds to multiomic factors linked to brain structure.

This chapter aims to identify blood-based epigenetic signatures associated with aging and brain structure, with particular focus on the MTL given its known vulnerability to AD pathology. We leverage the approaches described above in a multi-level analysis approach with large-scale methylation data and quantitative analysis of large-scale structural MRI to uncover molecular signatures

in blood that reflect the state of the aging brain. The overarching goal is to link these epigenetic patterns to the underlying biological processes of neurodegeneration in conjunction with and beyond the effects of aging. In doing so, we hope to understand how systemic aging processes, influenced by genetics, environment, and lifestyle, converge to effect regional vulnerability and resilience in the brain. This may point to novel blood-based biomarkers of structural brain change or intervention targets, serving as accessible predictors of brain aging and AD-vulnerability, with future implications in early identification of at-risk individuals and opening new avenues for preventative strategies to promote healthy brain aging.

4.2 Methods

4.2.1 Study Design and Cohort Selection

The study data comprised two independent cohorts from ADNI and the Aging Brain Cohort (ABC) from the Alzheimer's Disease Research Center at the University of Pennsylvania. Participants with both DNA methylation samples and structural MRI were selected. A cross-sectional dataset was selected within each cohort by matching longitudinal DNA samples to the nearest MRI within 3 years and selecting the time point with the smallest interval between DNA and MRI. Participants who were identified as CU based on clinical consensus diagnosis closest to MRI were selected. The CDRSB (Morris, 1993) was used as a measure of global function and cognition. Amyloid status was not considered based on the *a priori* hypothesis that amyloid does not significantly impact structural brain change in preclinical individuals.

The ADNI dataset (Weiner et al., 2017) was used for training and discovery purposes and included all individuals who met the criteria outlined above ($n = 225$). All participants provided written informed consent, and study procedures were approved by the Institutional Review Board (IRB) of participating sites per the Declaration of Helsinki. The ABC dataset included all individuals who met the criteria outlined above ($n = 215$). All ABC participants provided informed consent for participation, and the study received approval for ethical human subjects research from the University of Pennsylvania IRB. All experiments were conducted in compliance with institutional and national ethical standards for the collection and analysis of biospecimens and neuroimaging.

4.2.2 Epigenomics Data

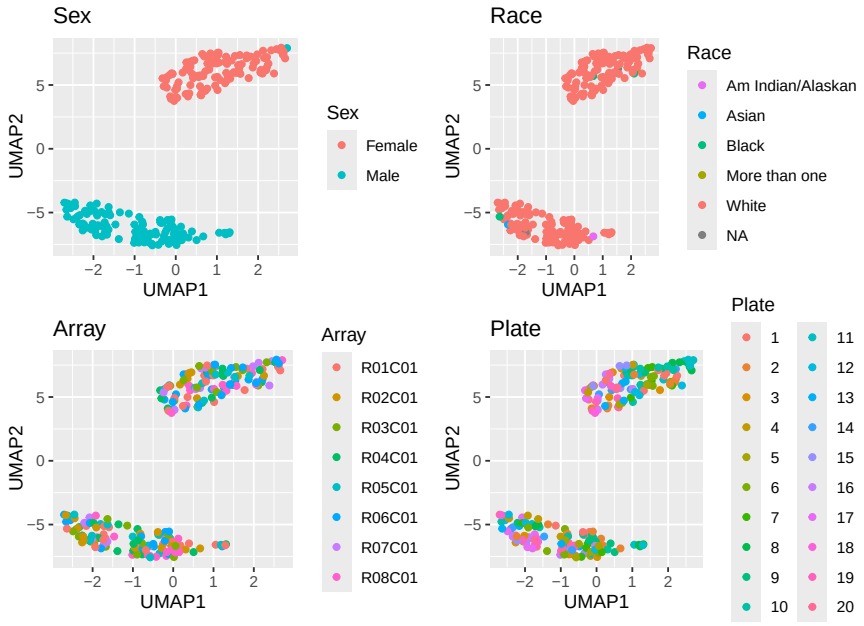
DNA methylation data were derived from whole blood samples and assayed on the Illumina Infinium Methylation EPIC BeadChip. All samples were assayed on EPIC v1.0 in ADNI and either EPIC v1.0 or EPIC v2.0 in ABC.

Data Processing and Quality Control

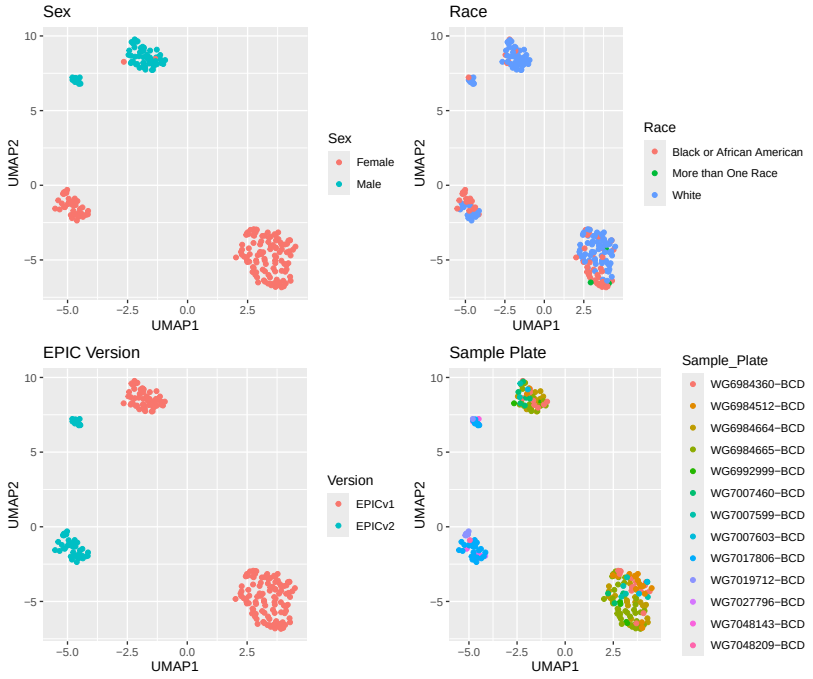
Raw intensity files were processed to yield beta matrices for each cohort using the *SeSAMe* R package (Zhou et al., 2018) according to standard protocols, including sample filtering, CpG probe filtering, signal correction, and normalization. Quality control (QC) procedures ensured at least 95% signal detection at each CpG. Unsupervised clustering analysis was used to check for potential technical or biological sources of variation using the UMAP algorithm (McInnes et al., 2020) in the *umap* package in R. CpGs with greater than 50% missingness across all samples in each cohort were removed, and remaining missing values were imputed using the beta value mean from non-missing samples. Sample separation by covariates is illustrated in Figure 4.1.

Estimation of Cell-type Proportions

Cell type proportions were estimated from DNA methylation data using a reference-based cell type deconvolution algorithm from the *EpiDISH* package in R (Teschendorff et al., 2017). The reference matrix was obtained from the *FlowSorted.Blood.EPIC* package (Salas et al., 2018), which includes methylation signatures for CD4+ T cells, CD8+ T cells, B cells, natural killer cells, monocytes, neutrophils, eosinophils, basophils, and other blood cell types. Beta matrices for ADNI and ABC cohorts and the reference matrix were filtered to include only CpG sites present in both datasets. Deconvolution was performed using the Robust Partial Correlations method (Wang et al., 2020b), which is suitable for blood-based methylation and provides estimates of the relative proportions of each cell type in each sample. Cell type proportion estimates were extracted from the deconvolution output and integrated with the main dataset using sample identifiers. This step enabled downstream analyses that control for cellular heterogeneity in bulk DNA methylation data, which is essential for understanding how cell type proportions vary across samples and differentially associate with phenotypes of interest.



(a) ADNI. Samples cluster primarily by Sex, minimally by Plate.



(b) ABC. Samples cluster primarily by Sex and EPIC version, and therefore by Plate. There is some evidence of clustering by Race.

Figure 4.1: Scatterplots of UMAP dimensions 1 and 2 in **(a)** ADNI and **(b)** ABC DNA methylation samples, colored by covariates.

Data Pooling and Missing Value Handling

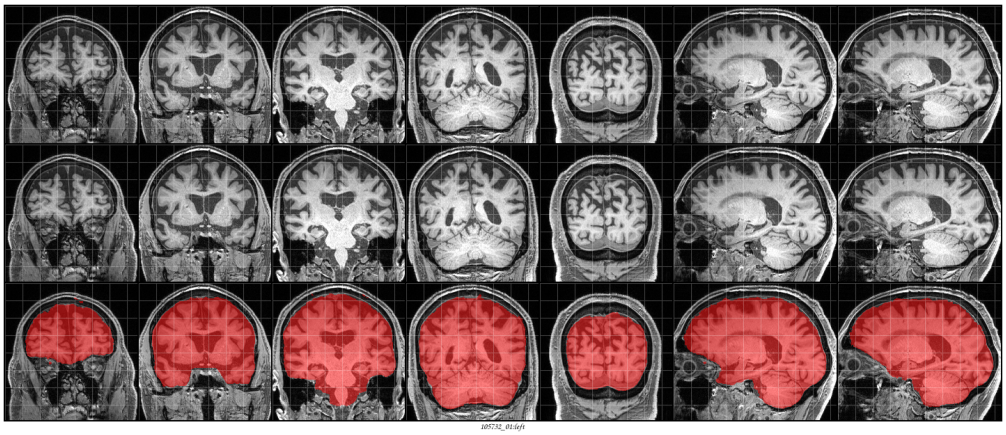
Processed DNA methylation data were limited to 721,802 CpGs in common between EPIC v1.0 (Pidsley et al., 2016) and EPIC v2.0 (Kaur et al., 2023) arrays to ensure compatibility across different versions. We further excluded CpG probes flagged as cross-reactive or prone to off-target binding using the *maxprobes* package in R (McCartney et al., 2016; Pidsley et al., 2016). A systematic approach to missing data was implemented using two-tier filtering: samples with more than 25% missing values were excluded. Subsequently, remaining CpGs with any missing values were removed. This approach ensured data quality while maximizing the number of samples retained for analysis and avoiding mean-based imputation, which may obscure biological signals by adding noise to the data distribution. CpG sites that showed insufficient variation across categorical covariates were identified and removed to ensure adequate statistical power for downstream analyses.

4.2.3 Neuroimaging Data

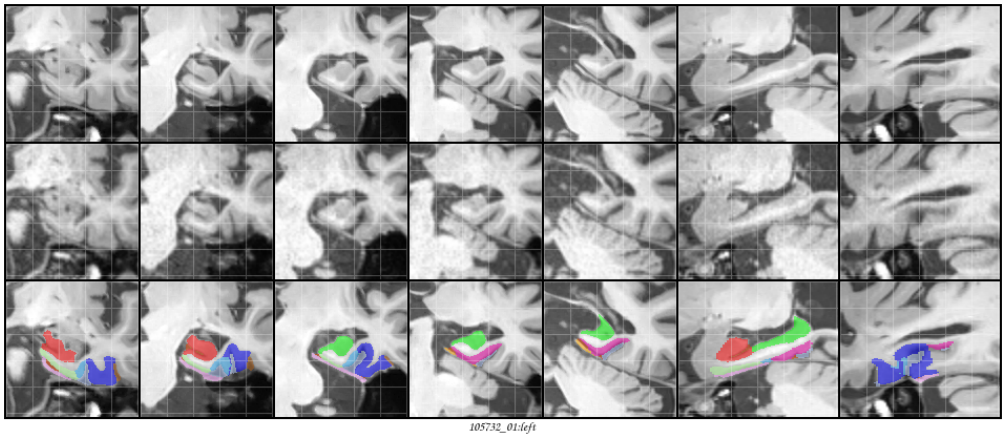
All participants underwent *in vivo* T1-weighted structural MRI scans. The ADNI cohort was enrolled across multiple phases of ADNI, namely ADNI1 and ADNIGO/2, for which MRI were acquired at either 1.5T or 3T in ADNI. Participants in ABC had MRI scans at 3T only. Scanner protocols were harmonized within cohorts to ensure consistency in acquisition parameters, including voxel resolution, pulse sequence type, and orientation. Detailed imaging protocols for ADNI are publicly available online. The ABC study protocols were closely aligned with standardized guidelines for neuroimaging acquisition and were comparable to those of ADNI.

Image Processing and Harmonization

MRI scans were processed and quantified using the Automated Segmentation of Hippocampal Subfields (ASHS) cross-sectional pipeline (Yushkevich et al., 2015) to yield regional volumes and cortical thickness measures in MTL subregions, as defined by the ASHS atlas. Researchers trained in neuroimaging and neuroanatomy assessed the ASHS segmentations for completeness and accuracy of anatomical boundaries, providing both qualitative evaluations and numerical ratings for QC. Only scans with acceptable QC ratings were retained for further analysis. Figure 4.2 provides examples of acceptable segmentations.



(a) Coronal (bilateral) and sagittal (left hemisphere) slices of T1 MRI with ICV segmentation boundaries indicated in red.



(b) ASHS segmentation boundaries in left hemisphere for anterior hippocampus (red) and posterior hippocampus (bright green), among other MTL subregions.

Figure 4.2: **Manual QC procedure for ASHS pipeline**. Acceptable segmentation outputs are shown for **(a)** ICV and **(b)** MTL subregions.

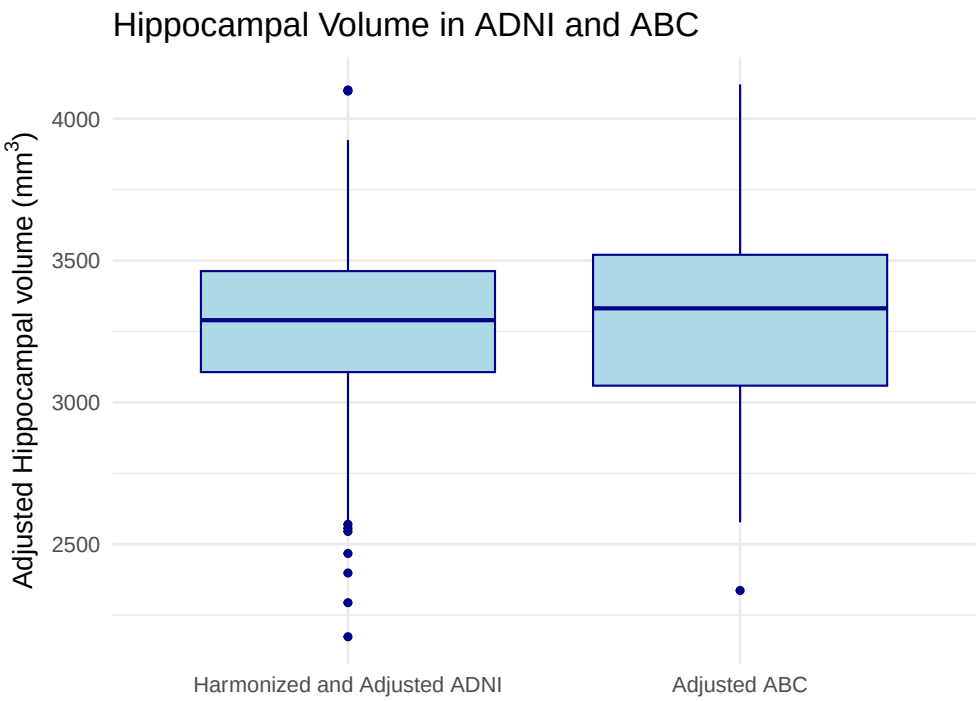


Figure 4.3: **Hippocampal volume in ADNI (left) and ABC (right)**. Volumes were harmonized to account for scanner field strength in ADNI and subsequently adjusted for ICV in both cohorts. There is no significant difference in the means of each cohort ($t = -1.3494$, df $= 437.8$, $p = 0.178$)
.

This study focused on bilateral measures of hippocampal volume averaged across the anterior and posterior regions. Statistical harmonization methods were necessary to account for varying acquisition field strengths in ADNI, but not in ABC. The ComBat family of harmonization methods is frequently used to correct for batch effects in neuroimaging measures (Fortin et al., 2018). ComBat uses Empirical Bayes models to harmonize means and variances across batches while preserving biological variability related to covariates of interest. We applied ComBat to harmonize total intracranial volume (ICV) and mean hippocampal volume using the *neuroCombat* package in R (Fortin et al., 2018). Magnetic field strength was modeled as the batch variable (i.e., 1.5T or 3T), while age, sex, and intracranial volume were retained as biological covariates. Finally, hippocampal volumes were adjusted to account for ICV in both ADNI and ABC . See Figure 4.3 for boxplots of hippocampal volume in ADNI versus ABC after harmonization and/or adjustment.

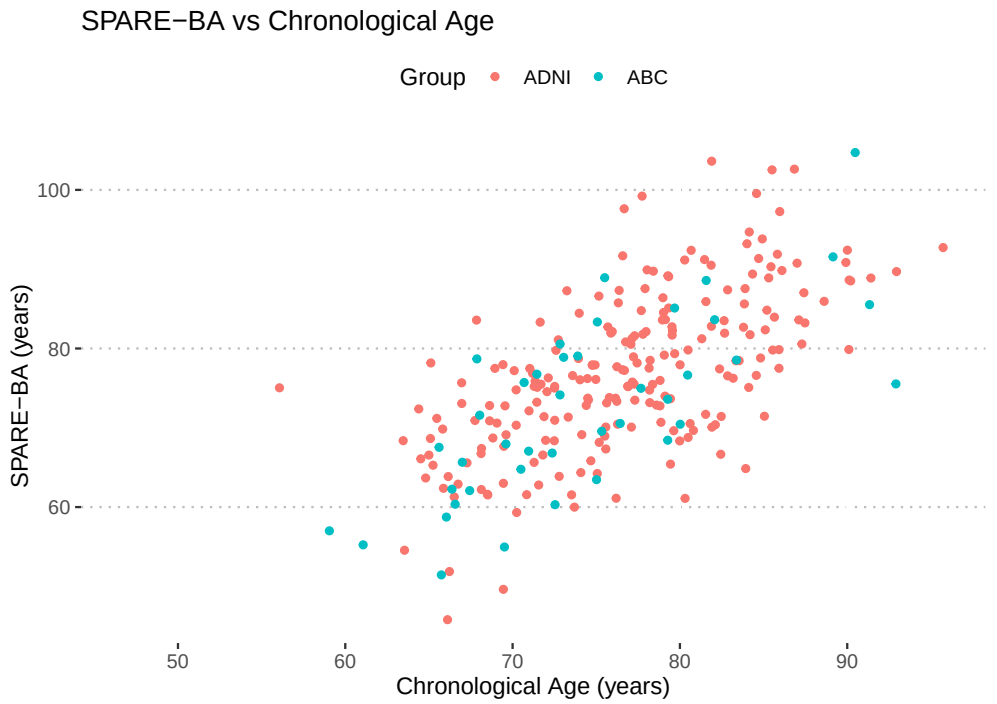


Figure 4.4: **Scatterplot of SPARE-BA versus Chronological Age.** SPARE-BA has a comparable data distribution in both ADNI and ABC. Additionally, there is a strong positive correlation between SPARE-BA and age across cohorts (Pearson $R = 0.65$, $p < 0.001$).

Brain Age Estimation from Structural Neuroimaging

The Spatial Pattern of Atrophy for Recognition of Brain Aging (SPARE-BA) is a popular brain age index that provides individualized estimates of brain-based biological age (Habes et al., 2016). The SPARE-BA model was trained on structural MRI from a large-scale, multi-study dataset pooled and harmonized by the Imaging-based coordinate SysTem for Aging and NeurodeGenerative diseases (iSTAGING) Consortium at the University of Pennsylvania (Habes et al., 2021). The SPARE-BA index has been validated as a useful biomarker of brain aging using clinical measures and neuropsychological test scores. SPARE-BA was calculated for almost all individuals in the ADNI cohort ($n = 223$) and a much smaller proportion of the ABC cohort ($n = 41$). Akin to epigenetic clock ages, higher SPARE-BA scores indicate relatively older-appearing brains (brain age $>$ chronological age), while lower scores reflect a younger neuroanatomical profile (brain age $<$ chronological age). Figure 4.4 illustrates the correlation between SPARE-BA and chronological age.

4.2.4 Machine Learning and Modeling Framework

DNA methylation levels at CpG sites formed the input feature set for the ML models, while the target variable was age or an age-associated phenotype relating to brain structure, i.e., SPARE-BA (Habes et al., 2016) or hippocampal volume. The initial set of CpG sites was restricted to CpGs common to both ADNI and ABC datasets after QC and missing value handling, forming the final set of CpGs used in feature selection and ML modeling as described below. The *scikit-learn* package in Python was used for the development and application of both feature selection algorithms and ML models.

Feature Selection

In this study, we used two types of feature selection: filter-based and embedded selection. Table 4.1 provides a description of each of the 4 methods, grouped by type (Doherty et al., 2023).

Filter-based selection typically requires a specified threshold on a metric calculated on the feature set that describes its association with the target variable and thereby its potential "utility" in a predictor model. We used three such metrics, namely the F-test, mutual information gain, and the Pearson correlation coefficient (Pearson R). Using each metric, we ranked the association of each CpG with the target variable and then applied a filter to include only the top n CpGs based on this ranking. As such, we not only compare different feature selection metrics, but also determine the optimal size of the feature subset for each method. We investigated models with successively larger feature subset sizes (50, 100, 150, 200, 250, etc. in steps of 50 up to 500 features, thereafter in steps of 100 up to 1000 features, and thereafter by multiplicative or exponential factors to test varying orders of magnitude). After initial filter-based feature selection, Elastic Net models were trained with hyperparameter tuning to find the subset size yielding the minimum error and maximum correlation between predicted and observed values of the target variable.

The embedded feature selection approach in the Elastic Net algorithm was used as a baseline model for comparison, i.e., no filter was applied to the feature set prior to training the Elastic Net, which leverages regularization penalties to "select" a final set of features by adjusting their weights during

the training process. More information on Elastic Net is provided in the following section.

All feature selection procedures were implemented strictly within the inner loop of a nested cross-validation (CV) framework, as described in (Doherty et al., 2023) and below. This design ensures that no information from the outer test fold is used during feature selection or model tuning, thereby avoiding data leakage and producing an unbiased estimate of model performance. Rankings are recalculated for each inner fold, and model performance is evaluated across those folds to determine the best-performing feature selection strategy and number of features. These optimal parameters are then applied to the full outer training set, and performance is assessed on the independent test set. This nested CV approach ensures that both feature importance and model hyperparameters are estimated independently of the final test data in each fold, aligning with current best practices for machine learning in high-dimensional biological data.

Table 4.1: **Feature Selection Methods.** In this study, we test 3 filter-based feature selection methods against a baseline method with embedded feature selection.

Type	**Method**	**Description**
Filter — Univariate	Correlation - Pearson's R	Calculates the Pearson correlation coefficient: $\rho(a,b) = \frac{\text{cov}(a,b)}{\sigma_a \sigma_b}$, which quantifies linear association between feature and response. Features below a defined correlation threshold are excluded.
	Multiple Hypothesis Testing (F-test)	Uses F-test statistics to evaluate feature-response association. Features are ranked based on p-value. CpG sites with p-value below 0.05 were retained.
	Mutual Information Gain (MIG)	Measures information gain between feature and response: $MI = H(x) + H(x\|y)$, where H is entropy. Features are ranked by MI gain.
Embedded	Elastic Net Regression	Regularized regression method with embedded feature-selection. Mixes both the L1 and L2 penalties in the objective function and tunes the bias towards one of the norms using a hyperparameter.

Elastic Net Regression and Hyperparameter Tuning

Given its frequent application in both first- and second-generation epigenetic clocks, this study uses Elastic Net regression as a baseline model, pairing it with the various feature selection approaches discussed previously. Elastic Net is a popular ML algorithm integrating two different regularization (i.e., penalty) terms, namely L1 (LASSO) and L2 (Ridge). An L1 penalty also reduces coefficient magnitudes but can shrink some to exactly zero, effectively removing associated predictors and performing feature selection. An L2 penalty reduces the magnitude of all coefficients without setting any to zero. It addresses collinearity and contributes to dimensionality reduction by shrinking coefficients. Hyperparameters control the mixture of L1 and L2 penalties (α) and the overall strength of regularization (λ). Hyperparameter tuning ensures that the optimal model is selected for each feature set and reduces the risk of overfitting. A mathematical representation of the Elastic Net algorithm is provided in Equation 4.1.

$$\hat{\beta} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^T \beta\right)^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \right\} \tag{4.1}$$

where:

- y_i: response variable for the i^{th} observation
- $\mathbf{x}_i$: predictor vector for the i^{th} observation
- β: coefficient vector
- n: number of observations
- λ: regularization parameter controlling the overall strength of penalty
- α: mixing parameter between L1 (LASSO) and L2 (Ridge) regularization terms, where:
  - $\alpha = 1$: Pure LASSO regression
  - $\alpha = 0$: Pure Ridge regression

- $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$: L1 norm of the coefficient vector (L1 penalty)
- $\|\beta\|_2^2 = \sum_{j=1}^{p} \beta_j^2$: Squared L2 norm of the coefficient vector (L2 penalty)

Two-Stage Model Development using Cross-Validation

We applied a two-stage modeling framework to develop epigenetic clocks of brain age and hippocampal volume. An illustration of the modeling framework, adapted from (Doherty et al., 2023), is presented in Figure 4.5.

Stage 1: The first stage of the analysis involved comparing multiple feature selection approaches in conjunction with Elastic Net using a nested 5 × 3 CV framework. The outer loop consisted of a 5-fold CV, effectively splitting the training dataset using an 80-20 split for training (X_{train}) and validation (X_{val}). The inner loop consisted of a 3-fold CV for hyperparameter tuning. Each feature-selection method was applied wholly to X_{train}, yielding a subset of features used to train an Elastic Net model in the inner CV loop with a randomized hyperparameter search, evaluated based on mean absolute error (MAE). Upon completion of the inner loop, the tuned ElasticNet model for the given feature subset size and chosen hyperparameters (i.e., estimator) was applied for prediction in X_{val}. A range of performance scores were used to evaluate estimator efficacy within nested CV. This constitutes one iteration of the 5-fold outer CV that is part of the 5×3 nested CV loop and is repeated for the remaining iterations in the outer loop. Performance metrics for each of the 5 folds were averaged to assess overall estimator performance.

The nested CV process was repeated for each combination of feature selection method (i.e., 4 methods comprising the baseline and 3 filter-based approaches) and feature subset size. Overall performance curves for each estimator were plotted against feature subset size, illustrating the effect of increasing feature subset size on model performance. Using the "elbow method", we identified the point of diminishing returns, namely the smallest number of features that yielded performance within 1% of the maximum observed across all subset sizes. This point represents a balance between parsimony and predictive accuracy and was selected as the optimal subset size for that feature selection method. This implementation ensured that the selected model was both interpretable and

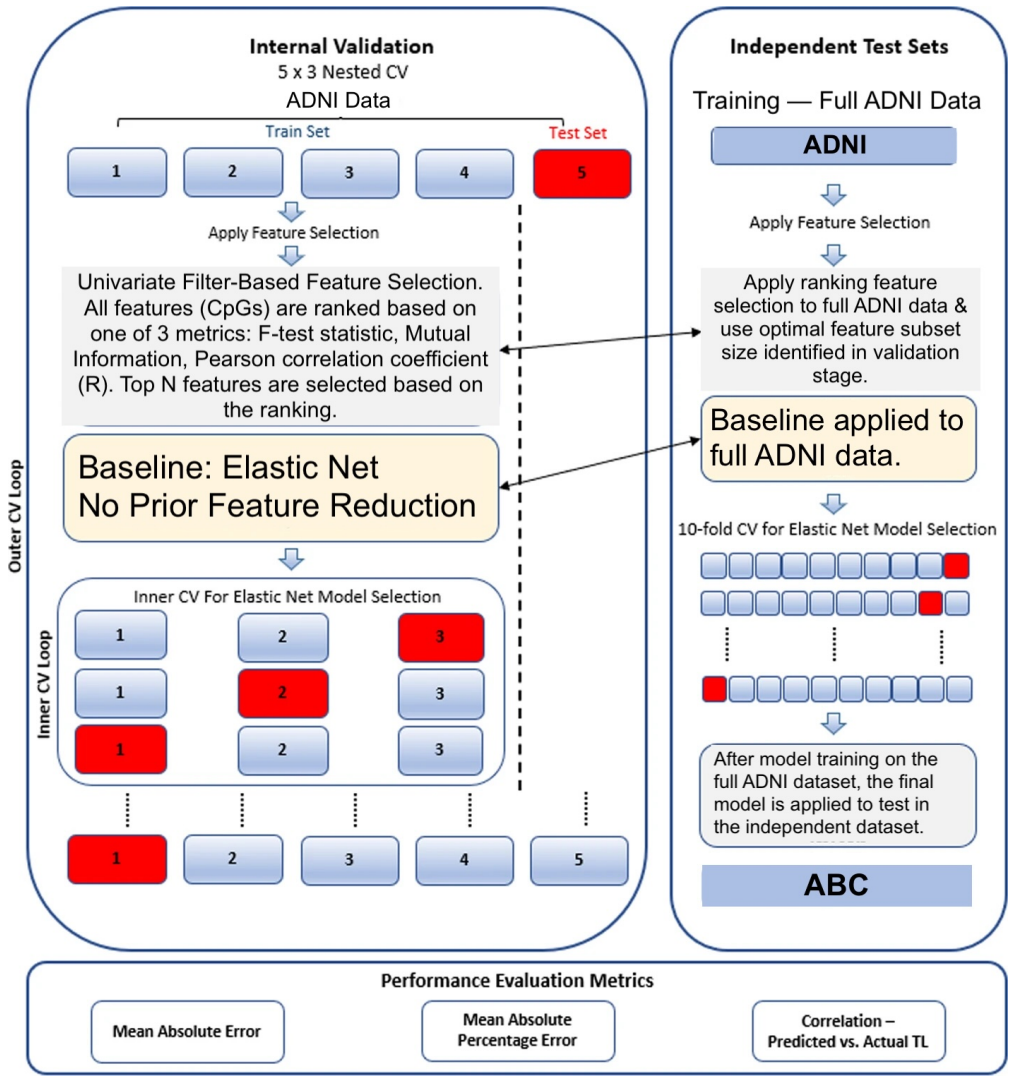


Figure 4.5: **Two-stage ML Framework for Epigenetic Clock Engineering**. In the first stage (internal validation, left), a nested 5×3 CV loop is used for feature selection and hyperparameter tuning. The outer loop (5-fold) performs feature selection on 80% of the data while the inner loop (3-fold) performs hyperparameter tuning within-fold. The tuned models are then tested on the remaining 20% of the data. In this manner, 4 models are developed for each target phenotype to enable comparison of feature selection methods (i.e., three with coupled feature selection and Elastic Net, and a baseline naïve Elastic Net). In the second stage (independent testing, right), the best approach is selected, based on internal validation performance, to train a final model using the full training dataset. Performance is evaluated in both stages using MAE, MAPE, and Pearson's R.

generalizable, while minimizing overfitting. Finally, the performance of the best estimators for each feature selection method were compared to pick the optimal feature selection algorithm. Stage 1 was repeated for each target variable, yielding 3 estimators for final training and evaluation in the second stage.

Stage 2: Based on the best estimators selected from the first stage, final epigenetic clocks were constructed for each target variable. Each approach was applied to the full training dataset (i.e., $X_{train} + X_{val}$) using 10-fold CV with exhaustive hyperparameter tuning using Grid Search. These final clocks were then tested on the held out, independent dataset.

Reciprocal Train-Test Procedure and Independent Test Set Evaluation

To evaluate model generalizability, we applied the trained epigenetic clocks to an independent dataset. Predicted values were compared to observed phenotypes using the same metrics as in the training procedure. We also implemented a train-test reversal approach in this study to assess the convergence of hippocampal volume clock pathways trained in different datasets. In one analysis, models were trained in the ADNI cohort and tested in ABC. In the reverse analysis, models were trained in ABC and tested in ADNI. This bidirectional validation approach ensured that clock performance was not driven by cohort-specific artifacts and that methylation-phenotype relationships were replicable across independent populations. Feature selection and performance consistency across training directions served as an additional validation of model robustness. As such, the whole pipeline (i.e., feature selection and hyperparameter tuning within nested CV, final training using the full training dataset, and independent test set evaluation) was performed twice for the hippocampal volume clock, i.e., training in ADNI or ABC and testing in the opposite dataset.

Statistical Analysis

We implemented a comprehensive statistical evaluation framework in this study, reporting multiple performance metrics including the mean absolute error (MAE), mean absolute percentage error (MAPE), and correlation (Pearson's R). Overall model performance was determined by averaging performance metrics from each fold of the 5-fold outer CV during training, providing stable estimates of out-of-sample accuracy. The same performance metrics were calculated for independent test set

evaluation. This allowed for direct comparison of within-sample and out-of-sample accuracy and provided insight into model generalizability. Residuals for all models were examined to ensure that assumptions of linearity and normality were met.

Gene Set Enrichment Analyses and Gene Ontology

We employed a pathway convergence approach to identify overlapping features across the independently trained clocks for chronological age, brain age, and hippocampal volume. CpG sites included in multiple models were considered convergent features, and their associated genes were flagged for downstream annotation and enrichment analysis. CpG sites selected for each model were annotated using the Know Your CpG (KYCG) database, which provides gene-level mapping and functional context for methylation loci, through the *kycg* package in R (Zhou et al., 2018, 2022; Ding et al., 2023; Kaur et al., 2023; Lee et al., 2024).

To further interpret the biological significance of the convergent gene set, we conducted gene ontology (GO) enrichment analysis using the *enrichR* package in R (Kuleshov et al., 2016). This platform allows comprehensive annotation of gene sets across multiple ontologies. We specifically queried three GO databases: Biological Processes, Molecular Function, and Cellular Component. Enrichment results were evaluated for statistical significance after correction for multiple comparisons, and the most strongly enriched terms were used to interpret potential molecular pathways underlying age- and brain-related methylation signatures.

Associations with Brain Methylation and Blood Gene Expression

To investigate the relevance of peripheral DNA methylation features for brain biology, we assessed cross-tissue correlations of methylation profiles. We first assessed the blood–brain correspondence of prioritized CpG sites using two publicly available databases: Blood–Brain Epigenetic Concordance (BECon) (Edgar et al., 2017) and the Blood Brain DNA methylation Comparison Tool from the University of Exeter (Hannon et al., 2015). These databases were developed based on matched blood and brain tissue samples from 16 individuals in BECon and between 71 to 75 individuals at Exeter. Each database had brain methylation samples from at least 3 different brain regions with minimal overlap of regions between databases. For each CpG of interest, we queried the databases

to determine whether blood methylation levels reflected brain methylation status.

To further explore the functional implications of candidate methylation sites, we integrated gene expression data from ADNI acquired through RNA sequencing (RNA-seq) of whole blood samples. RNA-seq data were processed using a standardized pipeline including transcript quantification, normalization, and log-transformation. For CpG–gene pairs identified in the pathway convergence analysis, we tested for associations between methylation levels and gene expression using Pearson's R without any covariate adjustments.

4.3 Results

4.3.1 Description of Study Cohorts

Cross-sectional characteristics of the study participants by cohort are shown in Table 4.2. Both ADNI and ABC are comparable for age, education level, and cognitive performance. Specifically, we observed median CDRSB values of 0 in both cohorts. In both cohorts, the time between DNA methylation and the nearest MRI was less than 1 year on average, with a slightly higher average interval in ABC compared to ADNI. ADNI is primarily composed of individuals identifying White, with minimal representation of Black individuals. In contrast, at least 30% of the ABC cohort identifies as Black. Finally, ADNI has a slightly higher percentage of amyloid-positive individuals and a slightly lower percentage of APOE $\varepsilon 4$ carriers compared to ABC.

4.3.2 Feature Selection Enhances Predictive Performance Over Baseline Models

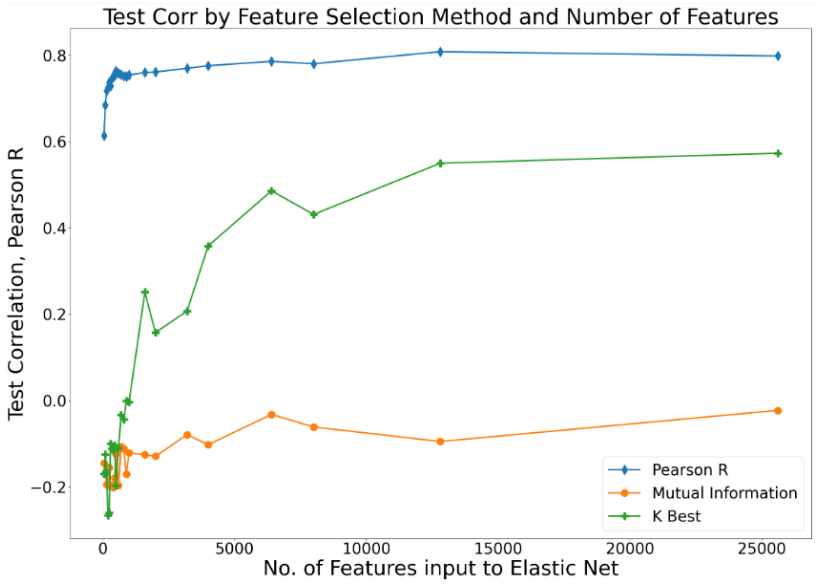
In the internal validation stage, we compared the performance of models coupling feature selection with Elastic Net against the baseline model using all features. For all target phenotypes, the baseline model exhibited poorer performance compared to models with prior feature selection. Table 4.3 provides a snapshot of the validation performance of age models, which we expected would perform well, based on *a priori* hypotheses rooted in previous studies of age-based epigenetic clocks.

For models with feature selection, we systematically varied the number of top-ranked features, n input to Elastic Net. As n increased, performance (Pearson's R) improved initially but reached a plateau or increased slightly beyond an optimal range, indicating potential overfitting to noise. Ad-

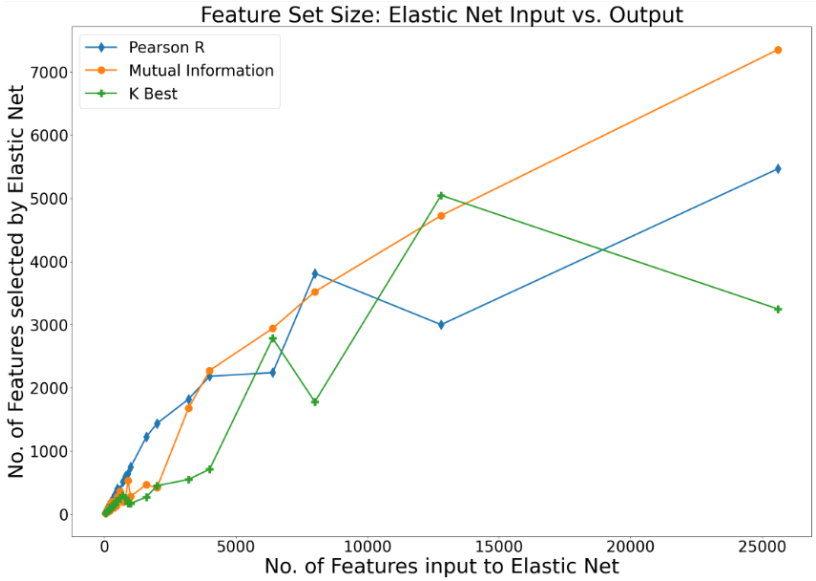
Characteristic	ADNI	ABC
N	225	215
Sex (% Female)	50.2%	69.0%
Age at DNA sample (years)	76.95 ± 6.82	71.33 ± 6.69
DNA to MRI (years)	0.21 ± 0.74	0.52 ± 1.33
Education (years)	16 ± 2	16 ± 2
CDRSB	0.0 ± 0.0	0.0 ± 0.0
Race (%)	White (97.3%) Black (1.3%) Other (1.2%)	White (68.1%) Black (30.1%) Other (1.9%)
Ethnicity (% not Hispanic/Latino)	97.8%	99.1%
Amyloid Status (% positive)	34.0%	25.5%
APOE4 Alleles (% carrier)	24.4%	33.8%
APOE Genotype (%)	E2/E2 (0%) E2/E3 (12.4%) E2/E4 (0.9%) E3/E3 (63.1%) E3/E4 (22.2%) E4/E4 (1.3%)	E2/E2 (0.5%) E2/E3 (8.3%) E2/E4 (1.4%) E3/E3 (57.4%) E3/E4 (28.2%) E4/E4 (4.2%)

Table 4.2: **Cross-sectional characteristics of the study cohorts**. All participants were assigned CU status based on clinical consensus diagnosis. Participants with both DNA methylation and structural MRI were selected. Longitudinal DNA samples were matched to the nearest MRI within 3 years and selecting the time point with the smallest interval between the DNA and MRI. All values are specified with units in "mean ± SD" format or as percentages, except for CDRSB and Education, which are in "median ± IQR" format.

ditionally, the final number of features selected by Elastic Net after feature selection (i.e., assigned non-zero coefficients) increased with n, however, the proportion of features selected remained relatively small compared to the size of the feature set, demonstrating a plateau-like behavior. Figure 4.6 plots performance metrics and number of features selected as a function of n in the ADNI-trained hippocampal volume models. Table 4.4 provides detailed performance metrics by each input feature subset size. Based on the internal validation stage, the best feature selection method and input feature subset size were selected for each target phenotype to develop final models using the full training dataset. The following subsections discuss performance and insights from the final, fully-trained models.



(a) Performance (Pearson's R) plotted against the number of ranked input features n for each feature selection method. The best performing method was Pearson's R.



(b) Final feature set size as a function of Input Feature Set Size for each feature selection method. While the size of the final feature set continues to increase with n, it remains a small proportion of the total number of available features.

Figure 4.6: **Internal Validation of Hippocampal Volume Clock.** In **(a)**, line plots of performance as a function of input feature set size, n. In **(b)**, line plots of final feature set size as a function of n. Performance metrics are provided for all feature selection methods as Pearson's R values, i.e., correlation between predicted and observed hippocampal volumes.

Table 4.3: **Internal Validation Performance of Epigenetic Clocks predicing Age**. Baseline Elastic Net exhibits poorer performance compared to models combining filter-based feature selection and Elastic Net. Performance is evaluated using Pearson's R in the held out validation set in ADNI. Pearson's R achieves better performance compared to F-test and MIG.

# Features	Baseline	F-test	MIG	Pearson's R
100	–	0.684	0.681	0.701
1000	–	0.627	0.682	0.694
492374 (all)	0.519	–	–	–

4.3.3 Validation of Modeling Framework with Chronological Age in Older Adults

The blood-based age prediction model trained in ADNI demonstrated reasonable generalizability in the test dataset (ABC). The model achieved a MAE of 3.86 years and MAPE of 5.55%. The correlation between predicted and actual chronological age was $R = 0.74$, indicating a strong linear relationship. This validates our modeling and feature selection framework using an established phenotype that is known to be predictable from blood-based methylation. While these results confirm that age-related methylation patterns can be detected in our data, overall performance was lower than that of established epigenetic clocks trained on large, well-curated cohorts. In direct comparison, the Horvath et al. (2018) skin-blood clock and the Hannum et al. (2013) blood-based clock both outperformed our model, consistently achieving > 0.8 correlation when tested in unseen blood methylation datasets. This discrepancy likely reflects the relatively smaller training set, differences in preprocessing pipelines, and the limited representation of CpGs known to contribute most strongly to age prediction in these prior models. Notably, individuals in ADNI and ABC were older on average than populations previously used to train first-generation epigenetic clocks.

4.3.4 Blood Methylation has Modest Predictive Power for Brain Structure

The brain age clock demonstrated modest generalizability in ABC. It achieved a MAE of 8.7 years and a MAPE of 12.89%. The correlation between predicted and true SPARE-BA was $R = 0.48$, suggesting a moderate but clearly weaker association compared to the chronological age clock.

On the other hand, the hippocampal volume clock performed poorly in ABC. Test MAE was 275.29 mm^3 with a MAPE of 8.59% and a low test correlation ($R = 0.071$), suggesting limited predictive power. We then assessed whether testing in subsets of ABC that better matched the ADNI training

Table 4.4: **Performance of Hippocampal Volume Clock in Internal Validation Stage.** Ranking and feature selection using Pearson's R yielded the best performing model. Validation performance is quantified by Pearson's R, ordered by number of ranked input features.

# Ranked Input Features	Performance
50	0.610
100	0.680
150	0.717
200	0.725
250	0.737
300	0.728
350	0.744
400	0.751
450	0.752
500	0.763
600	0.758
700	0.755
800	0.752
900	0.751
1000	0.753
2000	0.761
4000	0.775
8000	0.780
12800	0.808
25600	0.798

cohort would improve test performance. Specifically, we tested in individuals with EPIC v1.0 samples ($n = 161, R = 0.08$) and individuals identifying as White ($n = 147, R = 0.13$). Finally, we observed comparable performance when testing the ABC-trained hippocampal volume clock in ADNI. Table 4.5 provides a summary of performance metrics for all clocks and test sets.

4.3.5 Pathway Convergence in Hippocampal Volume Signatures

We assessed pathway convergence by evaluating CpGs selected by both ADNI- and ABC-trained models. Tables 4.6 and 4.7 provide a snapshot of the resulting signatures, including summary statistics for the top 25% of CpGs selected by each model, ranked by absolute value of the coefficients. Coefficients with higher magnitude indicate a stronger relationship between the associated CpG and hippocampal volume. Between the two signatures, one common CpG was identified (cg03526776), located on chromosome 6 and linked to TREML2. There was a modest but significant correlation between methylation at cg03526776 with adjusted hippocampal volume ($R = 0.21, p = 0.002$).

Table 4.5: Independent test set performance metrics for epigenetic clock predictors of age, brain age (SPARE-BA), and adjusted hippocampal volume. MAE is in years for age and brain age, and mm^3 for hippocampal volume.

Feature	**Training**	**Testing**	**# Features**	**MAE**	**MAPE (%)**	**Pearson R**
Age	ADNI	ABC	269	3.86	5.55	0.74
		ABC – EPICv1.0		4.23	6.11	0.75
		ABC – White only		4.10	5.94	0.76
SPARE-BA	ADNI	ABC	356	8.737	12.889	0.475
		ABC – EPICv1.0		8.91	12.79	0.40
		ABC – White only		9.07	13.59	0.59
Hippocampal Volume	ADNI	ABC	491	275.292	8.589	0.071
		ABC – EPICv1.0		282.48	8.89	0.08
		ABC – White only		269.64	8.42	0.13
	ABC	ADNI	196	277.53	8.94	0.05

Annotations of the full set of CpGs selected by each hippocampal volume clock yielded 14 unique linked genes in common between clocks. Among these genes, TREML2 and BCL11B were notable for their known implications in the nervous system.

We performed separate pathway enrichment analyses on the linked gene sets from each model. We identified 1468 enriched pathways in common between the two models, across three query databases: Biological Processes, Molecular Functions, and Cellular Components. Most pathways had at least two associated genes. While none of the gene ontology terms associated with these pathways passed significance levels after adjustment for multiple comparisons, many met traditional levels of significance. A few relevant, notable terms include "Intrinsic Apoptotic Signaling Pathway", and "Regulation of Neurotransmitter Secretion". Figure 4.7 shows pathway enrichment plots from *enrichR* for both ADNI and ABC and provides unadjusted p-values. Finally, we performed a pathway enrichment analysis on the set of 14 linked genes in common between the ADNI and ABC models, yielding 104 unique ontology terms across the three databases. Again, these terms did not pass multiple comparisons correction in terms of statistical significance, but included biologically relevant pathways including the two mentioned above.

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg06432119	305.816354
cg03871866	301.693131
cg02101882	287.277900
cg21809927	286.319090
cg07724670	273.407408
cg16418734	247.272360
cg06395414	246.263850
cg05212359	245.062956
cg16996788	230.994720
cg03673989	228.408320
cg17610144	223.652498
cg02184147	212.881568
cg14920334	211.700583
cg13262366	211.675444
cg04956949	211.038366
cg16457351	199.600489
cg20645010	197.003691
cg24881255	191.991480
cg09185587	190.615665
cg14671926	189.208618
cg08301503	188.394841
cg17581174	184.838430
cg19768241	183.315395

Continued on next page

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg01438090	179.892574
cg06933824	178.047591
cg08262378	176.593095
cg15492461	176.038597
cg21243717	173.665503
cg17425144	172.132671
cg01653772	165.621740
cg08576794	164.798869
cg13723190	164.296846
cg24107270	163.020128
cg22225577	161.043843
cg19666287	159.715456
cg14084907	158.087785
cg16679630	157.546415
cg07893949	155.511516
cg26976576	155.356048
cg15567340	154.125109
cg11473614	149.225610
cg22120446	144.750185
cg14799392	143.424883
cg25564800	143.403329
cg01440916	140.965659
cg01176433	139.584024
cg03487897	139.157651

Continued on next page

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg21498235	138.296980
cg02534450	137.321367
cg01596495	136.264956
cg13630250	134.074620
cg04464316	129.707872
cg07583744	127.888034
cg24477886	126.226248
cg21277266	125.480382
cg22787191	123.477680
cg05715828	121.974456
cg06502071	121.930354
cg12450316	121.280392
cg15022400	121.061320
cg19698976	119.923856
cg06037848	119.885210
cg00582194	119.122977
cg25160899	118.618028
cg14091493	118.406458
cg05385881	118.091879
cg11313460	117.838283
cg00902691	117.077977
cg02535223	116.641159
cg26603050	115.241135
cg11978824	112.886094

Continued on next page

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg17089162	112.879827
cg20337196	112.509322
cg14952605	109.851923
cg15832029	109.412321
cg22375689	109.230698
cg10342447	108.928240
cg10087412	108.865640
cg14177418	108.592757
cg01180684	107.969000
cg11443170	107.744330
cg07898974	107.284549
cg00383808	105.195520
cg08822897	103.199157
cg25883219	103.112977
cg09076662	101.779942
cg09751342	100.915522
cg07260592	100.794829
cg00025211	100.415551
cg25136644	100.127347
cg03432275	98.389516
cg02893407	97.572683
cg26864691	97.552540
cg17482714	97.379048
cg04374813	96.729363

Continued on next page

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg14532755	96.626232
cg27476074	96.132153
cg18996808	93.502208
cg12600184	93.405854
cg02855836	92.789377
cg00030591	92.241440
cg14286292	90.567339
cg04834502	90.418947
cg01233922	90.328096
cg10891352	88.818522
cg21292909	88.293027
cg06055249	88.032510
cg07741162	87.266586
cg07987287	87.213625
cg25614726	87.043765
cg02532672	86.320721
cg09861977	84.478893
cg21506819	83.574314
cg02607447	83.186250
cg18246518	82.600866
cg21721649	82.273763
cg17184477	81.473091
cg20332645	81.466333
cg16894439	81.373304

Continued on next page

Table 4.6: Top predictive features from ADNI model.

Feature name	Coefficient
cg02317055	81.073946
cg01538305	80.524725
cg05044994	80.123285
cg17870959	79.129003
cg14982096	78.691222
cg16725984	78.410082
cg14078092	76.643559
cg10688516	75.843869
cg25331593	74.027634
cg03526776	73.941390
cg22261895	73.494368
cg02593958	71.962675
cg10068105	71.622205

Table 4.7: Top predictive features from ABC model.

Feature name	Coefficient
cg08128007	580.674329
cg01041810	496.189368
cg27415945	470.135383
cg05237015	381.974908
cg01153287	376.916388
cg26018272	367.178319
cg13091431	364.677093
cg00919468	336.509408
cg21200703	325.193626
cg16370875	324.226062
cg19773510	310.143952
cg25955359	307.254202
cg08308806	298.599441
cg17248729	287.194293
cg03440930	270.443585
cg03526776	262.040969
cg02479816	257.177500
cg13228576	257.129474
cg02208848	248.946598
cg09323728	248.160594
cg11908736	246.856841
cg23815702	245.948144
cg04277677	245.586824

Continued on next page

Table 4.7: Top predictive features from ABC model.

Feature name	Coefficient
cg18316254	243.847787
cg00364357	242.984878
cg23906738	239.680811
cg10207277	238.832896
cg24393212	232.401756
cg15240733	226.663281
cg26198488	215.844450
cg24200059	215.021595
cg15519382	214.910147
cg04275947	214.846592
cg00210702	209.713335
cg19105961	205.117861
cg14841875	203.603199
cg19311206	202.147474
cg03134809	199.566176
cg08718017	197.516538
cg25387677	196.986819
cg18425118	196.533795
cg09841704	195.685017
cg00358132	191.566030
cg06259025	186.041367
cg09826019	180.841535
cg13798289	179.849144
cg18775606	177.439877

Continued on next page

Table 4.7: Top predictive features from ABC model.

Feature name	Coefficient
cg01890404	172.257274
cg11594343	171.105588
cg16503980	168.443586
cg07033494	158.207910
cg25579307	155.808040

Blood-Brain Methylation Concordance and Association with Gene Expression

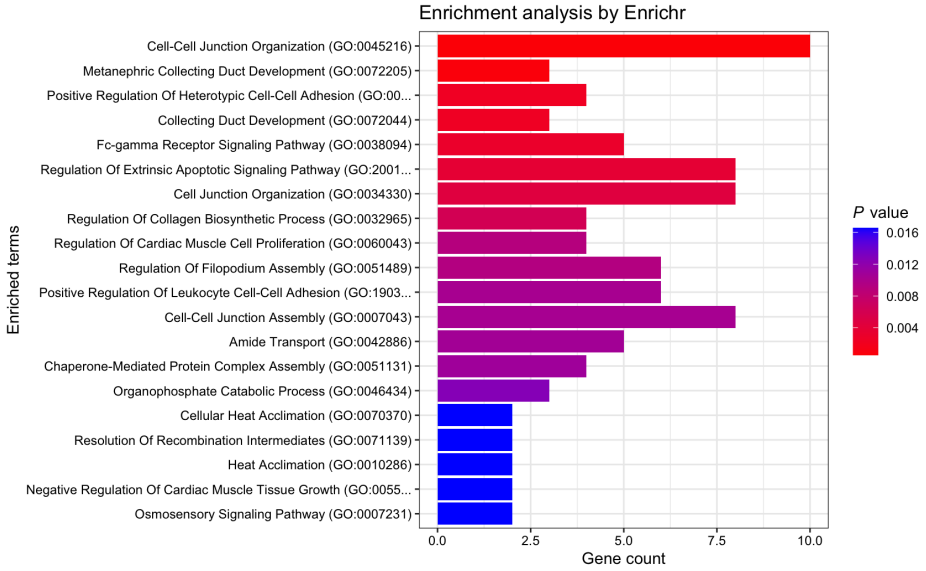
We assessed the correlation between methylation levels of the convergent CpG (cg03526776) in blood and various brain regions using two established datasets. The strongest concordance observed using the BECon dataset (Edgar et al., 2017) was in Brodmann Area 10 ($R = 0.8, p < 0.01$). Likewise, the strongest concordance observed using the Exeter dataset (Hannon et al., 2015) was in the superior temporal gyrus ($R = 0.456, p < 0.001$). There was also a significant correlation with methylation in the ERC, which is notably part of the MTL ($R = 0.358, p = 0.002$).

Integration with matched bulk RNA-seq data from blood revealed no correlation between methylation at cg03526776 and blood expression levels of TREML2 ($R = 0.04$). However, methylation at cg23479730 showed a strong inverse correlation with BCL11B expression in blood ($R = -0.5336097, p < 0.001$) and a modest yet significant correlation with adjusted hippocampal volume, ($R = -0.233703, p < 0.001$), consistent with a potential transcriptional repression mechanism.

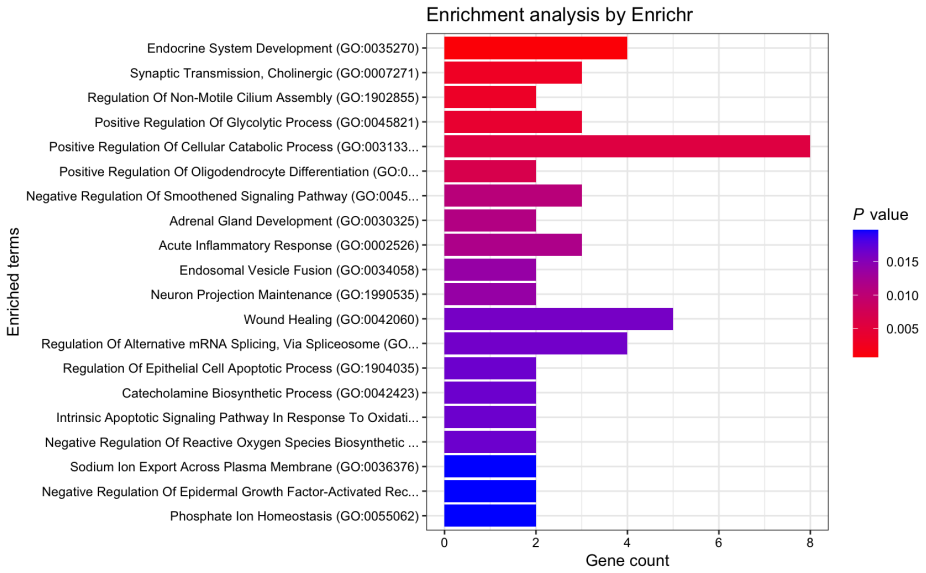
4.4 Discussion

4.4.1 Feature Selection is Critical for Next-Generation Epigenetic Clock Engineering

Feature selection played a crucial role in enhancing the internal validation performance of epigenetic clocks compared to baseline models that utilized the full feature set without pre-selection. By reducing the dimensionality of the input data, these methods helped mitigate noise and irrelevant information, leading to more robust and generalizable predictions. The performance gains highlight



(a) ADNI Pathway Enrichment in Biological Processes



(b) ABC Pathway Enrichment in Biological Processes

Figure 4.7: Pathway enrichment was performed on each linked gene set from the hippocampal volume models trained in (a) ADNI (912 genes) and (b) ABC (375 genes). Illustrated here are the top ontology terms from each model, including the number of overlapping genes for each term and the unadjusted χ^2 test p-value, meeting traditional significance levels. The p-values did not maintain significance after adjusting for multiple testing. These queries were performed for each of three databases, as described in the text. The figures here provide enriched terms from the Biological Processes database.

the value of targeted feature inclusion in modeling age from methylation data.

The application of the elbow method provided valuable insights into model optimization. As the number of input features increased, initial improvements in model performance were observed. However, a plateau was soon reached, suggesting diminishing returns and the potential onset of overfitting. This pattern reinforces the importance of identifying an optimal subset of features, beyond which additional CpG sites contribute marginal benefits while increasing computational complexity and risk of model overfitting.

Interestingly, while the total number of selected features grew with larger input sets, the relative proportion of selected features remained consistently low and demonstrated a plateau-like trend. This finding further underscores the principle that only a small fraction of CpG sites contribute meaningfully to predictive performance, emphasizing the redundancy inherent in high-dimensional methylation datasets. These observations also reflect the broader challenge of the curse of dimensionality, where the high feature-to-sample ratio can degrade model performance if not adequately addressed. Feature selection thus becomes essential, not only for improving predictive accuracy but also for maintaining model interpretability and stability.

4.4.2 Limited Predictive Power of Blood-Based Methylation in Brain

The use of blood-based DNA methylation profiles as proxies for brain-related phenotypes presents both opportunities and limitations due to the inherent tissue specificity of epigenetic marks. Numerous studies, including the large-scale ENIGMA consortium analysis (Jia et al., 2021), have demonstrated associations between blood methylation signatures and structural brain characteristics such as subcortical volumes. These findings suggest that peripheral methylation may carry some signal reflective of central nervous system processes. However, the strength and consistency of these associations are often modest, highlighting the challenge of inferring complex brain phenotypes from accessible but biologically distinct tissue types.

DNA methylation patterns are highly specific to tissue and cell types. Blood and brain differ substantially in their cellular composition and functional roles, which is reflected in their epigenetic

landscapes. In blood, immune cell heterogeneity adds another layer of complexity, as methylation profiles can vary widely depending on the relative abundance of different leukocyte subtypes. Similarly, the brain contains distinct cellular populations across regions, neurons, glia, and other cell types, with region-specific expression and epigenetic regulation. These biological differences limit the extent to which blood-based measures can reliably reflect brain-specific methylation dynamics.

This limitation is also evident in brain-based methylation samples. A recent study developed an epigenetic clock from brain tissue to predict tau in hippocampus and frontal cortex (Goldberg et al., 2024), demonstrating region-specific sensitivity with stronger associations observed in the frontal cortex compared to hippocampal regions. Such regional specificity further complicates efforts to use blood methylation data to draw inferences about heterogeneous brain processes, especially when those processes vary across anatomical substructures.

Given these complexities, the results linking peripheral epigenetic markers to brain aging should be interpreted with caution. Many of the reported associations remain tentative due to the absence of rigorous validation across independent datasets and the inherent limitations of using surrogate tissues. While blood offers a practical and non-invasive source of DNA for methylation analysis, its utility for capturing nuanced, brain-specific biological processes remains constrained, emphasizing the need for multi-tissue validation and integrative approaches in future research.

While the overall cross-validated test performance of the hippocampal volume model was modest, we believe that discussing the underlying biological signals remains justified for several reasons:

- **Consistency Across Datasets**: Although the aggregate predictive performance was low, individual datasets showed higher within-cohort predictive performance, suggesting that signal quality varies by cohort. In future studies, we could evaluate pathway enrichment using the fully-trained within-cohort models.

- **Biological Relevance**: As discussed in Doherty et al. (2023) and similar studies using high-dimensional epigenetic predictors, low R^2 values are common when predicting complex phenotypes like hippocampal volume from peripheral blood methylation, due to tissue specificity

and measurement noise. Nonetheless, even modest models can recover biologically meaningful features, especially when convergence is observed across models and datasets.

- **Pathway Convergence and Enrichment Interpretation**: Rather than relying on a single model's feature set, we used a reciprocal train-test and convergence approach to identify markers that appeared consistently across datasets. These converging features were then interpreted using pathway and gene set enrichment analyses. The resulting annotations are supported by independent literature and do not hinge on any single model's predictive power.

We clarify that the goal was not solely high predictive accuracy, but rather the identification of reproducible and biologically interpretable epigenetic markers of brain structure. Future work should place emphasis on inter-cohort reproducibility and enrichment-based interpretation. Given this, we discuss the genetic and pathway annotations of the converging epigenetic signatures below.

4.4.3 Functional Relevance of Converging CpGs and Linked Genes

The identification of CpG sites near genes with established roles in neurodevelopment and neurodegeneration provides an opportunity to explore potential biological convergence in epigenetic clock models. However, these findings should be interpreted cautiously, as they are preliminary and exploratory in nature. The inclusion of CpGs in intergenic regions, while sometimes lacking clear gene annotations, may reflect underlying regulatory activity,such as enhancer elements,that influences gene expression. Their recurrent appearance in predictive models suggests possible relevance, but definitive functional interpretations remain speculative without rigorous experimental validation.

Several gene-associated loci highlighted in this study, such as TREML2 and BCL11B, align with pathways known to be involved in brain aging and pathology. For instance, TREML2, a member of the TREM gene family, has been implicated in AD through its role in microglial function. While TREM2 is more directly involved in effector functions like phagocytosis, TREML2 may play a modulatory and potentially protective role. Some genetic variants of TREML2 have been associated with reduced AD risk, which raises the possibility that methylation near this gene could reflect relevant biological processes. However, these associations are preliminary and derived from blood

methylation data, which may not fully capture central nervous system dynamics due to the tissue-specific nature of epigenetic regulation (Wang et al., 2020a).

Similarly, CpG sites mapping near BCL11B, a gene critical for neurodevelopment, suggest potential links to neuronal differentiation and maintenance. BCL11B is active in multiple brain regions, including the neocortex and hippocampus, where it supports progenitor cell proliferation and axonal development. In the adult brain, it helps regulate the survival and function of mature neurons (García-Aznar et al., 2024). While these pathways are compelling, the methylation patterns observed in blood may not accurately reflect BCL11B activity in the brain, and their significance in the context of aging remains to be validated in independent, tissue-relevant datasets.

Taken together, these observations suggest a degree of convergence on biologically plausible pathways involved in neurodevelopment and neurodegeneration. However, given the absence of replication across independent cohorts and the known challenges of inferring brain-specific processes from peripheral methylation data, these results should be viewed as hypothesis-generating rather than confirmatory. Future studies integrating brain tissue methylation profiles, gene expression data, and functional assays will be necessary to clarify the roles of these loci in brain aging and to substantiate their inclusion in biologically informed epigenetic clocks.

4.4.4 Limitations and Future Directions

This study provides an exploratory framework for the development of biologically informed, blood-based epigenetic clocks with relevance to brain structure. However, several limitations must be acknowledged, motivating key areas for future work.

Tissue Specificity

One of the central challenges in interpreting these findings lies in the issue of tissue specificity. DNA methylation patterns are highly dependent on tissue and cell type, and blood derived CpGs may not fully capture methylation dynamics occurring in the brain. Although some convergence with genes known to influence brain structure, such as TREML2 and BCL11B, was observed, it remains unclear to what extent blood-based CpGs are directly involved in or reflective of brain processes.

They may be markers of more generic biological pathways, such as inflammation, immune signaling, or systemic aging processes, rather than specific drivers of structural brain change.

To address this, future studies should prioritize the use of cross-tissue methylation datasets and leverage tools that can evaluate the concordance of methylation changes between blood and brain. Feature selection approaches informed by known biological pathways or gene expression relationships across tissues may enhance the relevance and interpretability of selected features. Further, a clearer distinction is needed between features that relate to chronological age, those that align with biological aging biomarkers, and those that are uniquely associated with brain structural integrity.

Modeling Strategy and Feature Selection

The use of linear models such as Elastic Net in high-dimensional methylation data carries several implications for feature stability and interpretability, particularly due to collinearity among CpG sites and the stochastic nature of feature selection under regularization. We discuss these asThese linear frameworks may not be sufficient to capture complex, non-linear interactions between CpGs that influence brain phenotypes such as hippocampal volume. The biology of brain aging likely involves interconnected networks of CpGs and gene regulatory elements, suggesting that single-CpG analyses may underestimate multivariate or synergistic effects.

Many CpG sites are highly correlated due to shared genomic context, such as proximity within a regulatory region or shared chromatin domains. In such settings, Elastic Net may arbitrarily select one feature over another from a set of collinear predictors, depending on small fluctuations in the training data. As a result, different cross-validation folds or datasets may select different CpGs that carry redundant biological information. This phenomenon can contribute to discordance at the feature level across models or "clocks," even if the underlying biological signal is shared. While we used nested cross-validation and multiple iterations to mitigate overfitting, there is still inherent stochasticity in model training due to fold assignment and hyperparameter tuning. This can influence which features are retained in the final model, particularly when many CpGs are near the regularization threshold.

To address this, we emphasized feature convergence across models and datasets rather than relying solely on any single model's output, which reduces the influence of random or dataset-specific selections. Observed concordance or discordance across methylation clocks may reflect both true biological divergence and model-induced variability. Our reciprocal train-test convergence strategy was designed to prioritize consistently selected features across datasets, thereby distinguishing stable biological markers from artifacts of collinearity or regularization paths. Nonetheless, we acknowledge that some degree of discordance may arise from the model's sensitivity to correlated inputs, especially in the absence of prior biological constraints.

In future work, methods such as stability selection, ensemble modeling, or graph-regularized regression could further reduce selection variability and promote robust identification of co-regulated methylation modules. In particular, advanced Deep Learning models could incorporate CpG-Gene-Pathway interactions as priors to potentially improve prediction accuracy, and model robustness, enabling biologically interpretable insight generation.

Sample Size and Generalizability

The relatively small sample size used for model training poses constraints on generalizability and may limit the stability of feature selection and model estimates. Future research should aim to pool data across studies using harmonized pipelines or federated learning techniques to increase power. Furthermore, the generalizability of the current models to diverse ancestral backgrounds is limited. Most epigenetic studies, including this one, are based on samples of predominantly European ancestry. Expanding the diversity of study populations and applying statistical methods that account for ancestry-related methylation variation will be crucial for ensuring equitable and accurate predictive models across populations.

The limited study power necessitated *post hoc* analyses which reflect an exploratory approach, aimed at identifying patterns and potential pathways of interest rather than establishing definitive mechanisms. Limited convergence across different analytical strategies and feature selection methods is unsurprising given the small sample size, methodological variation, and the underlying complexity of the epigenetic landscape. This lack of consistency highlights the need for more rigorous,

pre-registered analyses in future work, supported by larger training and replication datasets.

Disentangling Age from Structural Brain Change

Finally, the relationship between age and brain structure is a significant confound in the modeling process. While controlling for age helps isolate brain-specific signals, it also reduces the variance attributable to aging and may obscure meaningful biological associations. Conversely, including age in the model may risk conflating general age-related change with specific structural features. Future research should explore strategies that model both shared and distinct components of age and brain structure, potentially through multi-task learning or residual-based frameworks that explicitly disentangle these effects. While the present work offers promising insights, it remains an exploratory effort. Rigorous validation, more biologically informed modeling, and improved data diversity will be essential for translating these findings into reliable tools for studying brain aging and related disorders.

4.5 Chapter Summary

This study offers tentative but intriguing evidence for the potential of blood-based DNA methylation markers to reflect structural brain changes. However, the limitations imposed by tissue specificity underscore the need for brain-based or cross-tissue analyses to enhance biological relevance and interpretability. While predictive power remains modest, exploratory analyses revealed pathway-level convergence at loci such as TREML2 and BCL11B, pointing to biologically plausible mechanisms that may underlie neurodegeneration and cognitive decline.

Moving beyond traditional epigenetic clocks, our findings support a shift toward biologically grounded models that prioritize mechanistic insight over chronological age prediction. Future research should leverage larger, more diverse datasets and adopt advanced DL techniques to unravel the epigenetic networks influencing brain aging and neurodegeneration. By focusing on functional pathways rather than singular associations, the field can make meaningful progress toward early detection and intervention.

4.6 Acknowledgements

This work was supported by funding from the Alzheimer's Disease Research Center at the University of Pennsylvania. We thank the ADNI and ABC participants, their families, and care partners, for their contributions to this study. ADNI data are shared without embargo through the LONI Image and Data Archive. Requests for neuroimaging, genetic and epigentic data, gene expression, biomarker and biomarker data for scientific investigation, teaching, or planning clinical research may be made via application. Access is contingent on adherence to the ADNI Data Use Agreement and the publication policies. More information is available at the ADNI webpage. ABC data are stored in the Integrative Neurodegenerative Disease Database (INDD) at the University of Pennsylvania and are not publicly available, but may be accessed upon request. In addition, the analysis code and documentation for this chapter are available publicly on GitHub.

We would like to acknowledge the following collaborators without whom this work would not have been possible: Drs. Sandhitsu Das, David Goldberg, Stanislau Hrybouski, Taki Shinohara, Anil Wadhwani, Paul Yushkevich, and Wanding Zhou. Special thanks to Emily McGrew and Gaylord Holder for their assistance with image data processing and compute cluster management. We also thank members of the Penn BiND Lab, Penn FTDC, PICSL, and "Neuroimaging of Neurodegenerative Disease" journal club at the Perelman School of Medicine, University of Pennsylvania, who contributed thoughtful insights and provided feedback on this chapter.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 Dissertation Summary

In this dissertation, we set out to investigate the central hypothesis that epigenetic modifiers, specifically DNA methylation, influence the heterogeneity of brain aging and regional vulnerability to neurodegeneration as quantified by *in vivo* structural MRI. Within this context, we also characterized the predictive power of blood-based epigenetic profiles and their potential as a surrogate for brain-derived phenotypes of aging and AD. To this end, we leveraged both statistical analysis and data-driven ML approaches across two aims:

In **Aim 1** (Chapter 3), we postulated that biological aging, defined epigenetically, explains the relative involvement of cortical and MTL structures in AD beyond the known effects of chronological age. Established ML algorithms, i.e., epigenetic clocks, were applied to estimate biological age acceleration from blood-based DNA methylation profiles in individuals. The Cortico-Medial Temporal index (CoMeT) was developed as a novel quantitative measure of neurodegeneration along the cortical-MTL axis, derived from *in vivo* structural MRI. Finally, the association between biological aging and CoMeT was evaluated using statistical hypothesis testing in both continuous and stratified subgroup analyses.

In **Aim 2** (Chapter 4), we developed blood-based epigenetic signatures of brain aging and neurodegeneration using ML and identified their association with underlying biological pathways. To achieve this, we trained Elastic Net regression models to predict brain age and hippocampal volume using methylation levels across the entire epigenome. Within a nested CV loop with hyperparameter tuning, a variety of feature selection methods were evaluated to identify features most strongly associated with the target phenotypes. The prediction performance of these models was assessed to determine the optimal feature selection strategy and number of features for the final model training. We leveraged a pathway convergence approach, performed by conducting reciprocal train-test iterations on discovery and validation datasets, to extract common markers from the epigenetic

signatures identified by the trained models. These markers were annotated using publicly available databases and interpreted using gene ontology and gene set enrichment analyses. Finally, we examined the concordance between methylation levels in blood and brain at the converging markers, along with their associations with blood gene expression. This enabled me to characterize the extent to which peripheral blood methylation reflects brain-specific biological processes.

Together, both empirical chapters bolster the theoretical arguments based on established literature that were developed in Chapter 2, positioning epigenetics as a complementary approach for deciphering heterogeneity and atypicality in brain aging and AD. The findings of this dissertation demonstrate that epigenetic markers modulate regional vulnerability to neurodegeneration in conjunction with and beyond chronological age. Additionally, the results highlight both the promises and the current limitations of blood-based epigenetic biomarkers with respect to organ-specific phenotypes, especially in the context of the brain. In the following subsections, we expand further on the biological and engineering insights from this dissertation and integrate these novel results with existing knowledge before outlining the limitations of this work.

5.1.1 Biological Insights and Conceptual Advances

This dissertation takes a novel approach to deciphering heterogeneity in aging and AD by framing epigenetics as modulator of regional neurodegenerative patterns reflected in both typical and atypical presentations of disease. Historically, age of onset and atypicality have been conflated based on the observation that most EOAD cases present with more atypical clinical syndromes. The advent of *in vivo* neuroimaging and fluid biomarkers corroborated this observation, demonstrating greater neuropathological burden in the cortex relative to the MTL in these cases. With the increasing recognition of "mismatch" cases, where younger onset cases present more typically and older onset cases present atypically, it started to become evident that chronological age was only one piece of the puzzle. Given the rising popularity of epigenetic clocks in estimating biological measures of age, it seemed fitting to apply this approach to investigate biological mechanisms underlying age-associated heterogeneity.

Indeed, we discovered in Chapter 3 that the spectrum of relative cortical to MTL neurodegen-

eration in AD is partially regulated by biological age, as estimated from blood-based epigenetic profiles Specifically, individuals appearing biologically younger were more likely to have cortical-predominant neurodegenerative patterns, whereas those who appeared biologically older tended towards MTL-predominant presentations. Importantly, these associations persisted after adjusting for chronological age and sex and did not have an association with the number of APOE $\varepsilon 4$ alleles. This suggests that epigenetic clocks provide information not captured by traditional demographic or genetic variables. However, these findings did not translate into the cognitive domains, consistent with established knowledge positing that structural brain change precedes detectable clinical decline. This also cautions against the assumption that any biomarker linked to neurodegeneration will immediately translate into cognitive prognostication.

The insights gained in Chapter 3 provide fertile ground for Chapter 4. Given that patterns of regional vulnerability to neurodegeneration were shown to be linked to blood-based epigenetic measures of biological age, the next logical step was to survey the whole methylome to identify specific signatures of focal neurodegeneration in the hippocampus and brain age estimates derived from structural neuroimaging (i.e., SPARE-BA). To overcome the limited sample size relative to the vast feature space, we leveraged a pathway convergence approach in which we reciprocally trained and tested ML models in two independent datasets and identified converging epigenetic markers from each model associated with the target phenotypes. Therefore, the pathway analyses are best viewed as hypotheses for future validation, ideally in larger, diverse datasets.

As the association between the target phenotype and age waned, the predictive performance of the trained ML models also decreased. Specifically, the epigenetic clock trained to predict age achieved the best performance and generalized in an independent dataset, achieving a peak test correlation of Pearson $R = 0.74$. On the other hand, the clock predicting brain age performed more modestly, with a maximum test correlation of Pearson $R = 0.48$. Finally, we determined that it was not possible to construct an "episcore" of hippocampal volume from a blood-based, cross-sectional epigenomic dataset of this sample size given the null test performance (Pearson $R = 0.07$), which improved marginally when testing in only White individuals (Pearson $R = 0.13$). These results reinforce the

notion that blood epigenetic markers chiefly reflect systemic aging processes.

While brain age may capture some biology associated with neurodegeneration, it is effectively a noisier estimate of chronological age. The underlying model is trained to predict chronological age from 3D structural MRI and inevitably introduces error into the predictions. Further, while it is known that hippocampal atrophy is a prominent feature in both healthy aging and age-related neurodegenerative processes, it is an even noisier surrogate of chronological age. Notwithstanding, the pathway convergence approach identified 14 shared genes linked to the epigenetic signatures captured by the models predicting hippocampal volume. Among these, TREML2 and BCL11B are particularly interesting. TREML2 is a gene that affects immune system function and has been implicated in studies of AD [REF]. It may modulate disease progression by regulating microglial, polarization, thereby influencing the inflammatory response in AD. BCL11B is expressed in the neocortex and hippocampus, among other brain regions, where it regulates development and function of axonal projection of corticospinal motor neurons () Garcia-Aznar et al., 2024). It plays a key role in the neurogenesis of the hippocampus, where it promotes proliferation of progenitor cells and regulates cell differentiation. The absence of BCL11B leads to significant impairment of hippocampus and motoneurons (). Both TREML2 and BCL11B have been implicated in previous genetic studies of AD.

5.1.2 Engineering and Methodological Advances

As part of the approach to uncover the findings summarized in section 5.1.1, this dissertation also made advances in engineering and computational methods. First, we developed a novel index (CoMeT) that operationalizes atypicality in AD along a spectrum by quantifying relative cortical to MTL involvement, as measured by structural MRI. The final version of CoMeT presented in Sreepada et al. (2025) was achieved after several iterations. To decide which features to include in each composite (i.e., cortical or MTL), we mapped vector-based neuroimaging signatures of aging, EOAD, and LOAD to corresponding ROIs from the DKT atlas. we incorporated both cortical thickness and regional volume measures into CoMeT, leading to the next consideration—how to best standardize ROI-level measures to leverage two types of structural measures with very different

units. An established approach in neuroimaging is to adjust thickness measures for age and sex (or age and ICV for volumes), resulting in w-scores which are in standard units. To improve the detection of signal associated with atypicality in symptomatic individuals, we used the amyloid-negative, CU individuals as a reference cohort in the w-scoring algorithm. Another consideration was whether to use ratios, as in CLix (Kouri et al., 2024), or simple subtraction to derive a score that reflected relative atrophy. After evaluating the benefits and drawbacks of both approaches, we settled on subtraction because it would yield a signed value. The sign would indicate which composite exerted greater impact on the final scores, whereas a ratio would mask this effect. In sum, CoMeT is a useful advancement for studying atypical AD phenotypes *in vivo*. In Section 5.2.2, we discuss additional engineering approaches to extend the framework in future work.

The second key engineering contribution related to implementing feature selection in the engineering pipeline designed to develop a novel, "second-generation" epigenetic clock that predicts brain age as well as an episcore of hippocampal volume. we determined that rigorous and meaningful feature selection was necessary for ML models to detect signal from the high-dimensional next-generation methylation arrays, which cover between 850K (EPIC v1) and 950K CpGs (EPIC v2). Established clocks trained without feature selection, such as the multi-tissue clock (Horvath, 2013), were primarily trained on 27K and 450K arrays which are much smaller arrays. In contrast, sample sizes for datasets with both neuroimaging and DNA methylation are a tiny fraction of the feature space ($n = 225$ in ADNI and $n = 215$ in ABC). Without prior feature selection, baseline Elastic Net models either underfit or overfit the methylation data distribution, perhaps due to CpG collinearity and insufficient sample size. To address this challenge, we implemented a rigorous feature selection module coupled with hyperparameter tuning within a nested CV loop. This enabled us to test the performance of multiple filter-based feature selection methods (F-test in a k-best selector, mutual information gain, and Pearson R) and optimize the size of the feature subset input to Elastic Net for each method. Prior feature selection significantly improved the model performance relative to baseline (i.e., naïve Elastic Net). Future generations of epigenetic clocks and EpiScores will benefit from implementing feature selection in training pipelines and potentially leveraging biologically informed feature selection strategies as opposed to univariate filters. These approaches are discussed

in Section 5.2.2.

The empirical findings of this dissertation refine current theories of regional vulnerability in the brain by adding a systemic aging component. Both chapters converge on the central finding that peripheral blood-based epigenetic modifiers involved in age-related processes interact with region-intrinsic factors to shape gradients of atrophy in the brain. However, the ability to detect this signal is limited by various logistical considerations, which we identify in the next subsection 5.1.3. These findings align with emerging multi-system models positing crosstalk between the central nervous system and peripheral aging pathways. Although these findings require replication in larger validation cohorts, they hint at shared molecular substrates linking peripheral epigenetic signals and regional vulnerability to neurodegeneration, providing a framework for studying heterogeneity and atypicality in aging and AD.

5.1.3 Limitations and Challenges

The research presented in this dissertation also has various limitations from both the biological and engineering perspectives, including limited sample size and the tissue-specific nature of DNA methylation. In the next section, we discuss alternative approaches to address these limitations and propose additional research to validate and further advance the findings presented here. Additionally, we offer a roadmap for broader exploration that would significantly expand and enhance this field from both biological and engineering perspectives.

5.2 Future Directions

The findings of this dissertation lay the groundwork for a broad and interdisciplinary research agenda aimed at deepening our understanding of how epigenetic modifiers shape systemic aging processes to influence regional neurodegeneration in AD and related dementias. Building on the empirical advances in Chapters 3 and 4, future directions span two domains: (1) biological insights and translational extensions and (2) innovation in engineering and computational methods.

5.2.1 Biological Insights and Translational Extensions

In this domain, we discuss how to extend the CoMeT framework to more specific *in vivo* biomarkers, expand the scope of epigenetic modeling to atypical AD variants and co-pathologies, and reorient epigenetic clocks and scores towards mechanistic discovery over prediction performance.

Extending the CoMeT Framework

A key next step is to expand the CoMeT framework by incorporating tau PET imaging to define a novel "Tau-CoMeT" index. This approach would align with the corticolimbic index, i.e., CLix (Kouri et al., 2024) described in Section 3.4. *In vivo* tau-PET could help identify pathological subtypes of AD during life, mirroring CLix subtypes observed at autopsy. This approach may also help determine whether cases with atypical, cortex-predominant tau exhibit distinct patterns of epigenetic aging or neurodegenerative progression. Tau-PET offers a clearer window into these relationships given its specificity to AD compared to structural measures of neurodegeneration. However, several challenges must be addressed to implement this approach. In ADNI, tau-PET data were typically collected nearly a decade after DNA methylation sampling. To overcome this limitation, longitudinal measures of methylation could be used to generate a more stable estimate of epigenetic changes, shifting the research focus toward early detection or prediction of tau pathology rather than concurrent measures of neuropathology and neurodegeneration. The subset of ADNI participants with both methylation and tau-PET data is also limited due to participant attrition since collection of DNA samples. Ideally, additional studies with a shorter interval between methylation and PET imaging, such as the ABC or LEADS cohort (Touroutoglou et al., 2023), should be integrated to mitigate this issue.

Given that tau tends to be more regionally localized than amyloid or neurodegeneration, it is important to analyze tau burden in individual regions in addition to relative differences across the cortex and MTL. For example, two individuals with the same "tau-CoMeT" may exhibit markedly different absolute tau levels: one with uniformly high tau across both regions and another with uniformly low levels. These distinct profiles may have different prognostic or mechanistic implications. This approach may also distinguish subgroups with globally elevated pathology, which may indicate

fast-progressors, from those with disproportionate regional involvement suggesting an atypical disease presentation. Furthermore, focusing on neuroanatomical regions instead of composite averages may reveal finer-grained associations. In Chapter 2, we describe the distinct topographical vulnerabilities observed in neuroimaging and neuropathological studies of atypical AD presentations. As such, analyzing tau-PET in specific cortical regions including the frontal, temporal, and occipital lobes could reveal methylation profiles linked to each atypical presentation. Additionally, regional analysis could probe hemispheric asymmetry in tau and atrophy, potentially illuminating why some patients present with lateralized symptoms. This would provide greater insight into selective vulnerability (i.e. within-individual differences in focal neurodegenerative pattern), compared to the intra-individual variability explored in Chapter 3.

From a clinical perspective, prospective studies should aim collect epigenetic in atypical and mixed-pathology cases, which are ideal for probing the epigenetic regulation of neurodegeneration in vulnerable regions beyond the MTL. As detailed in Chapter 2, atypical presentations including PCA, lvPPA, and frontal-variant AD comprise approximately 10 to 15% of all AD cases and often occur at younger ages. Additionally, longitudinal cognitive assessments may help relate baseline epigenetic biomarkers to downstream cognitive changes or rates of cognitive decline across syndromes. For example, an extension of Chapter 3 could be to compute cognitive CoMeT for multiple time points after baseline methylation to assess at what point neurodegeneration translates to the cognitive domain. This approach could improve personalized prognostication and more timely interventions in individuals at risk of accelerated decline.

Broader Applications of Epigenetic Clocks

A recurring insight from our work is that while epigenetic clocks are valuable for summarizing biological age, the next phase of research should repurpose epigenetic data to uncover mechanisms of disease and potential intervention targets. In Chapter 4, we ventured beyond simple age clocks by training models to predict epigenetic surrogate markers (episcores) of hippocampal volume from DNA methylation. While the utility of cross-sectional, blood-based epigenetics as a surrogate for brain-related phenotypes proved to be limited, leveraging cross-tissue methylation datasets and

longitudinal measures could address this issue. Expanding the repertoire of methylation scores for various brain phenotypes including cognition and *in vivo* biomarkers of pathology could transform large epigenomic datasets into a rich source of hypotheses about disease pathways. A recent study made headway in this approach through the development of two episcores characterizing tau spread and tau severity using paired neuropathology and brain methylation data (Goldberg et al., 2024). Unlike clocks that conflate many processes to optimize age prediction, phenotype-focused episcores may help pinpoint underlying biological processes. Akin to the approach employed in a recent study by (Levine et al., 2022), we also envision deconstructing epigenetic models into "modules" that correspond to distinct biological axes such as metabolic aging and inflammatory aging. Instead of monolithic estimates of biological age, a panel of specialized epigenetic biomarkers could collectively explain more variance in neurodegenerative disease processes.

Another intriguing direction, complementary to the study of atypical phenotypes, is to examine how epigenetic "aging" interacts with co-pathologies commonly seen in older adults. For example, LATE is most frequently observed in older individuals, potentially confounding epigenetic measures of biological age (Wolk et al., 2025). We could assess biological age acceleration in the context of co-pathologies like LATE, cerebrovascular disease, or Lewy bodies to see if they additively increase epigenetic aging or have distinct signatures. An improved hippocampal volume episcore could be applied to detect co-pathologies such as TDP-43. For example, individuals meeting A/T criteria for AD but an episcore predicting low hippocampal volume might have underlying TDP-43 pathology. This could enable early identification of LATE during life in individuals who might be misdiagnosed as AD. It may also enable the development of an ML classifier that distinguishes pure AD cases from AD and concomitant LATE cases. Although we determined in Chapter 4 that the hippocampal volume episcore generalized poorly, its performance could improve if the model is trained on larger, perhaps longitudinal datasets.

Addressing Fundamental Biological Complexity

Peripheral blood methylation captures signals from systemic factors including immune response, vascular and metabolic changes, and environmental exposures such as smoking. While these factors

play a role in brain aging and neurodegenerative processes, they are not specific to the brain. In contrast, methylation changes measured in the brain reflect local processes that are region- and cell-specific. This suggests that blood-based epigenetic clocks and episcores must be interpreted with caution in the context of brain aging. Cross-tissue epigenetic studies will help determine the potential of methylation peripheral tissues, such as blood, to serve as surrogates for brain methylation and brain-specific changes. A few databases with methylation data from paired blood-brain samples have already been developed, including BECon (Edgar et al., 2017) and the Blood Brain DNA Methylation Comparison tool from the University of Exeter (Hannon et al., 2015). In Chapter 4, we leverage these databases to show that some CpG sites have significant correlation between blood and region-specific brain methylation, whereas many others do not. We could identify CpG panels based on blood-brain methylation concordance and use this subset of features in episcore development. Integrating multiomic data, including gene expression from RNA-seq and meQTLs will help identify further informative CpGs. This effectively serves as a form of biologically informed feature selection. We expand upon novel feature selection approaches in the next section.

5.2.2 Engineering and Computational Methods Innovation

This second domain addresses the frontiers of modeling high-dimensional epigenetic data using advanced DL architectures that are explainable and biologically aware, bioengineering innovation to develop precision epigenomics assays, and causal inference frameworks.

Feature Engineering and Model Development

The high dimensionality of epigenetic data calls for more sophisticated feature selection and modeling techniques. Our current analyses largely treated individual CpG sites as independent features. As we observed, single CpGs are often noisy and have limited explanatory power. Future work should move beyond univariate filter-based feature selection towards feature engineering methods that consider spatial and functional relationships among CpGs to reduce dimensionality while maintaining biological interpretability. One promising avenue is to leverage graph representations of CpG-gene networks. In Graph Neural Networks (GNNs), each CpG can be a node connected to other CpGs by edges representing a chosen metric, such as association with the target phenotype, genomic prox-

imity, or membership in a common biological pathway (Veličković et al., 2017). GNNs may capture patterns like coordinated methylation changes along signaling pathways or chromatin domains that simpler models miss. Finally, explainable AI techniques such as SHAP can characterize the relative contribution of network clusters or hubs to resulting predictions (Lundberg and Lee, 2017), offering insight into underlying biological pathways.

Model Training and Generalizability

Both EWAS and advanced ML models experience performance limitations with modest sample sizes. In EWAS, the substantial multiple-testing burden severely reduces statistical power, while in ML models tend to overfit and generalize poorly. Consequently, increasing sample size should be a key consideration of future research. Collaborations with large consortia such as ENIGMA can facilitate access to large multimodal datasets and support the pooling of data across studies for both discovery and validation. For instance, the dataset described in (Jia et al., 2021) included over 3,300 individuals with both neuroimaging and DNA methylation data, making it one of the largest resources of its kind. Expanding sample sizes not only improves generalizability but also relaxes the need for restrictive linear assumptions and enables the application of higher-capacity DL models. Larger datasets are particularly important for multi-task networks that simultaneously predict multiple outcomes, such as age and hippocampal volume, by learning shared representations from input features. This dovetails with ongoing efforts to disentangle biological processes from chronological age, as discussed in the previous section. Additionally, future work should prioritize not only larger but more diverse samples, with meaningful representation of racial, ethnic, and gender minorities. This will support the development of ancestry-specific epigenetic clocks that account for variation in genetic background and environmental exposures. The statistical genomics field has already advanced multi-ancestry methods, including SuSiEx for cross-population fine-mapping of GWAS loci to identify putative causal variants (Yuan et al., 2024). Epigenomics research is well positioned to build upon and integrate these developments.

Longitudinal and Multi-Task Modeling

So far, we described analyses based primarily on cross-sectional data. However, aging and neurodegeneration are inherently dynamic processes. Longitudinal neuroimaging and epigenomic data are essential for modeling change in phenotypes of interest. Mixed effects models are popular methods for longitudinal data analysis because they incorporate random slopes and intercepts to account for individual variation, enabling the computation of individualized rates of change. For example, we could estimate slopes in SPARE-BA (Habes et al., 2016)measures over time and develop a "third-generation" epigenetic clock that predicts this "pace of brain aging". Conceptually, this is analogous to the DunedinPACE clock (Belsky et al., 2022). By linking these rates of structural brain change to baseline methylation profiles, we may identify epigenetic signatures associated with AD progression.

However, longitudinal analyses present several challenges, including batch effects and increased measurement noise. Advanced statistical harmonization will be helpful in distinguishing true biological signals from technical variability. Existing methods to stabilize longitudinal epigenetic clock estimates often use dimensionality reduction techniques, such as principal component analysis (PCA), to reduce measurement noise associated with single CpGs (Higgins-Chen et al., 2021). However, this approach can reduce interpretability when components are derived across timepoints. It may also be useful to develop longitudinal clocks that explicitly model temporal dynamics using recurrent neural networks (RNNs), long short-term memory (LSTM) architectures, or transformers may improve forecasting and sequential prediction. Ultimately, leveraging longitudinal data enables the assessment of both biological age and the rate of aging or neurodegenerative change, providing critical insights for the early detection and intervention of AD and related dementias.

Supervised Methods to Disentangle Age from Biological Processes

A recurring theme in our research is the interplay between aging and pathology-specific changes, which are likely entangled in both neuroimaging and epigenetic estimates of biological age. An exciting future direction is to refine the epigenetic signatures in Chapter 4 by disentangling these two components. On the neuroimaging end, Hwang and colleagues developed ML models to disentangle neuroimaging biomarkers of typical brain aging from those of AD using 4,054 participants with

structural MRI (Hwang et al., 2022). By training separate models on clinically and molecularly defined subgroups, the authors generated more specific imaging signatures for aging and AD. The resulting models showed reduced correlation between aging and AD scores, indicating improved separation of the two processes yet the AD-score still tracked with known molecular markers and genetic risk factors such as APOE.

A similar strategy could be employed to develop "disentangled epigenetic clocks". Rather than relying on a single model that captures all age-related changes, two parallel models could be trained: one to optimize the prediction of chronological age in cognitively unimpaired individuals, and another to capture AD-specific methylation changes by distinguishing AD cases from CU individuals, independent of age. Differences between model residuals could then indicate whether an individual's methylation profile reflects accelerated aging due to systemic processes or disease-specific effects. For instance, one clock might be enriched for immune- or inflammation-related CpGs, indicative of systemic aging, while another might focus on CpGs associated with neuronal or synaptic genes, reflecting neurodegenerative processes. This approach would require class-balanced datasets with *in vivo* neuroimaging biomarkers to accurately define groups. As another avenue starting point, we could train models on raw or age-adjusted phenotypes and infer association based on differences in the identified signatures. However, we observed in Chapter 4 that cross-sectional, blood-based estimators perform worse as the target phenotype becomes more "distant" from age. As such, developing an episcore of age-adjusted hippocampal volume would be more difficult in this context, but could be an interesting avenue to pursue in larger datasets.

Biotechnology and Bioengineering Innovation

Leveraging modern biotechnology and engineering tools to design customized epigenetics assays can help boost the signal-to-noise ratio for brain-related phenotypes. Targeted bisulfite sequencing of candidate genomic region, including APOE and other loci identified in Chapter 2, could provide high-resolution methylation data on specific regions hypothesized to be implicated in brain aging and AD-related processes. Generic panels used widely in present-day epigenetics research, as well as this dissertation, include the Illumina Infinium EPIC v1 and EPIC v2 Methylation Bead-

Chip (Pidsley et al., 2016; Kaur et al., 2023). These arrays are designed for general purposes and heavily cover CpG sites relevant to oncology and developmental imprinting. This one-size-fits-all approach is cost effective and efficient for large studies but may omit CpGs critical for neuroscience research. The new Methylation Screening Array (MSA) from Illumina focuses on sites informative for neurodevelopment, neurodegeneration, and systemic processes that have been implicated in brain-related changes such as immune responses and gut microbiome modulation (Goldberg et al., 2025). The MSA can dramatically reduce acquisition costs and enable collection of methylation data in specific AD subtypes, which would enhance our work in Chapter 3. Additionally, given the more focused selection of CpG probes, the MSA could improve the specificity and performance of predictive models. The ABC study has at least 400 methylation samples from the MSA which could be immediately processed and studied within the framework developed in this dissertation.

One of the central challenges in biomarker development for neurodegeneration is that brain phenotypes are clinically informative but difficult to acquire longitudinally, while peripheral molecular profiles such as blood DNA methylation are accessible and scalable. Our work in Chapter 4 was designed to address this by using brain imaging phenotypes (i.e., hippocampal volume) as supervised targets during model training. While our model was only modestly predictive due to discrepancies in data acquisition protocols, more advanced models achieving higher levels of accuracy can be deployed in settings where only blood data are available.

To bridge the gap between peripheral and brain molecular states more directly, future work could consider leveraging MR spectroscopy (MRS), a non-invasive imaging modality that can quantify brain metabolites involved in methylation-related pathways, which serve as *in vivo* proxies for brain methylation potential. By integrating peripheral methylation profiles with MRS-derived neurochemical measures, we could gain a more direct window into methylation-linked processes in the brain. Recently, Lam and colleagues developed an epigenetic MRI (eMRI) method that directly images DNA methylation in the brain, overcoming the challenges of brain-tissue accessibility and enabling ante mortem studies of epigenetic changes in the brain (Lam et al., 2022). eMRI exploits the methionine metabolic pathways for DNA methylation by labeling genomic DNA using ^{13}C en-

riched diets. An MRS sequence is then applied to map the spatial distribution of labeled DNA using radiofrequency detection. eMRI revealed strong regional differences in global DNA methylation in piglets, whose brains bear greater resemblance to human brains compared to rodents.

However, eMRI is currently limited by the sensitivity of isotopic labeling and limited spatial resolution, providing only global or regional methylation signals as opposed to CpG-specific data. In other words, while eMRI reveals regional differences in aggregate methylation, it cannot resolve methylation at individual CpG sites or gene-level resolution. Nontheless, this approach holds promise for eventually monitoring brain epigenetic modifications in humans. If translated to humans, this technology could enable experimental validation of blood-based methylation markers against *in vivo* brain methylation and longitudinal observation of the impact of lifestyle interventions. Even before translation to human research, *in vivo* MRS could be used to detect other metabolites, such as N-acetylaspartate, that are related to epigenetic regulation of oxidative stress in the brain. The development of these tools underscores the broader point that future neuroepigenetics studies will likely integrate both blood- and brain-based readouts to characterize the epigenetic landscape more completely.

Causality and Mechanistic Validation

In this dissertation, we identified associations between peripheral blood methylation and structural brain changes, but the question remains whether these methylation changes are drivers or consequences of age-related neurodegenerative disease processes. Recent approaches in statistical genetics allow for *in silico* testing of causal hypotheses prior to validation *in vitro*. For example, Mendelian Randomization (MR) is an analytical approach that infers causal relationships between risk factors and disease outcomes using genetic variants as natural instruments mediating risk. MR is often considered "nature's clinical trial" because it leverages the random assortment of alleles and genetic variants at birth to minimize confounding and reverse causation, which commonly affect observational studies. In the context of epigenetics, MR can be applied to meQTLs or genes linked to epigenetic signatures derived from ML models, such as TREML2 and BCL11B identified in Chapter 4. Overall, MR provides a powerful tool for uncovering causal mechanisms and prior-

itizing therapeutic targets for aging and AD. Ultimately, the most compelling evidence will come from experimental intervention. Collaborations with wet-lab scientists will enable CRISPR-screens and experimental modification of key epigenetic marks identified in this dissertation. For example, genes identified by epigenomics literature (Chapter 2), or the latest empirical evidence (Chapters 3 and 4) can be targeted in cellular or animal models to test mechanistic hypotheses. Such *in vitro* experiments would allow observation of downstream effects on gene expression and cellular function, providing direct insight into causal relationships and a bridge to translation.

The future directions outlined in this section represent a natural extension of the insights gained through this dissertation. Building on the frameworks established in Chapters 3 and 4, we expanded the definition and application of CoMeT and episcores to better characterize heterogeneity and atypicality in brain aging, neurodegeneration, and cognitive decline in aging and Alzheimer's disease. From this foundation, we proposed a shift from purely predictive modeling toward mechanistically driven and clinically relevant investigations, aiming to uncover how systemic aging processes shape region-specific neurodegenerative patterns across diverse phenotypes. The complexity of high-dimensional epigenomic data further motivates computational and bioengineering innovations, including biologically informed feature selection and model training, as well as precision assays for profiling methylation specific to brain-related genes and pathways. We also highlighted the potential of advanced DL architectures, such as multi-task networks, to capture nonlinear relationships and shared representations between age and age-associated phenotypes. Finally, we emphasized the importance of bioinformatics tools that support causal inference and mechanistic interpretation, laying the groundwork for translational applications.

5.3 Conclusion

In conclusion, this dissertation makes several key contributions to neurobiology and computational engineering in the context of studying the aging brain. Our empirical work advances both the conceptual and methodological frameworks for characterizing heterogeneity. Specifically, we posit epigenetics as a modulator of regional vulnerability to neurodegeneration, proposing a novel framework with which to understand and characterize the biological mechanisms underlying heterogeneity

and atypicality in aging and neurodegenerative diseases such as AD. We introduce the novel CoMeT index and developed epigenetic models of brain age and hippocampal volume using rigorous ML pipelines. Additionally, we demonstrate that peripheral blood-based epigenetic markers can serve as informative, albeit imperfect, surrogates for brain aging. These findings offer compelling evidence that blood-based epigenetic aging is a biologically meaningful and independent contributor to regional neurodegenerative patterns beyond the known effects of chronological age and other established risk factors, such as genetics and life exposures. The broader impact of this work spans both clinical practice and translational research. With advancements in cross-tissue methylation studies, epigenetic signatures in blood may be used as screening tools or biomarkers that are sensitive to atypical variants and mixed pathologies. These biomarkers have the potential to inform patient stratification in clinical and pharmaceutical trials and may help monitor treatment response. Our work also opens avenues for novel drug discovery and therapeutic interventions by proposing future work focused on investigating causality. Ultimately, this dissertation catalyzes a shift towards precision epigenomics, building a foundation for the next generation of personalized diagnostics and interventions in aging and neurodegenerative diseases.

5.4 Acknowledgements

This chapter was shaped by the insights of several mentors and professors, including my advisors and committee members. First and foremost, I am deeply grateful to my advisors, Drs. Corey McMillan and Dave Wolk for their critical analysis of my research and thoughtful perspectives. I also greatly appreciate the guidance on the neuroimaging and ML aspects of my thesis from members of the "HippoGang" and Frontotemporal Degeneration Center, including Drs. Chris Brown, Katie Cousins, Stan Hrybouski, Sandy Das, and Paul Yushkevich. The Methylation Working Group, especially Drs. Rory Boyle, Nadia Dehghani, David Goldberg, Anil Wadhwani, and Wanding Zhou, offered valuable feedback and perspectives on multiomics analyses and interpretation. Dr. Taki Shinohara provided thoughtful input on biostatistical fundamentals and persisting challenges in biological age estimation using neuroimaging and epigenomics. Finally, many meaningful insights came from spontaneous conversations with fellow lab members during pop-in chats at our desks after lab meetings, at the lunch table, and over a latte in one of the many coffee shops in Philly.

BIBLIOGRAPHY

Sanaz Arezoumandan, Sharon X. Xie, Katheryn A. Q. Cousins, Dawn J. Mechanic-Hamilton, Claire S. Peterson, Camille Y. Huang, Daniel T. Ohm, Ranjit Ittyerah, Corey T. McMillan, David A. Wolk, Paul Yushkevich, John Q. Trojanowski, Edward B. Lee, Murray Grossman, Jeffrey S. Phillips, and David J. Irwin. Regional distribution and maturation of tau pathology among phenotypic variants of Alzheimer's disease. *Acta Neuropathologica*, 144(6):1103–1116, December 2022. ISSN 0001-6322, 1432-0533. doi: 10.1007/s00401-022-02472-x.

Nicole M. Armstrong, Yang An, Lori Beason-Held, Jimit Doshi, Guray Erus, Luigi Ferrucci, Christos Davatzikos, and Susan M. Resnick. Predictors of neurodegeneration differ between cognitively normal and subsequently impaired older adults. *Neurobiology of Aging*, 75:178–186, March 2019. ISSN 01974580. doi: 10.1016/j.neurobiolaging.2018.10.024.

Akram Bakkour, John C. Morris, and Bradford C. Dickerson. The cortical signature of prodromal AD: Regional thinning predicts mild AD dementia. *Neurology*, 72(12):1048–1055, March 2009. ISSN 0028-3878, 1526-632X. doi: 10.1212/01.wnl.0000340981.97664.2f.

Daniel W Belsky, Avshalom Caspi, David L Corcoran, Karen Sugden, Richie Poulton, Louise Arseneault, Andrea Baccarelli, Kartik Chamarti, Xu Gao, Eilis Hannon, Hona Lee Harrington, Renate Houts, Meeraj Kothari, Dayoon Kwon, Jonathan Mill, Joel Schwartz, Pantel Vokonas, Cuicui Wang, Benjamin S Williams, and Terrie E Moffitt. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife*, 11, January 2022. ISSN 2050-084X. doi: 10.7554/elife.73420.

D Frank Benson, R Jeffrey Davis, and Bruce D Snyder. Posterior cortical atrophy. *Archives of neurology*, 45(7):789–793, 1988.

H. Braak and E. Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, September 1991. ISSN 0001-6322, 1432-0533. doi: 10.1007/bf00308809.

R Brookmeyer, S Gray, and C Kawas. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *American Journal of Public Health*, 88(9): 1337–1342, September 1998. ISSN 0090-0036, 1541-0048. doi: 10.2105/ajph.88.9.1337.

Tiago C. Silva, Juan I. Young, Lanyu Zhang, Lissette Gomez, Michael A. Schmidt, Achintya Varma, X. Steven Chen, Eden R. Martin, and Lily Wang. Cross-tissue analysis of blood and brain epigenome-wide association studies in Alzheimer's disease. *Nature Communications*, 13(1):4852, August 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32475-x.

Judith Campisi and Fabrizio d'Adda di Fagagna. Cellular senescence: when bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, 8(9):729–740, 2007. doi: 10.1038/nrm2233.

Marianne Chapleau, Renaud La Joie, Keir Yong, Federica Agosta, Isabel Elaine Allen, Liana Apostolova, John Best, Baayla DC Boon, Sebastian Crutch, Massimo Filippi, et al. Demographic,

clinical, biomarker, and neuropathological correlates of posterior cortical atrophy: an international cohort study and individual participant data meta-analysis. *The Lancet Neurology*, 23(2): 168–177, 2024.

Brian H. Chen, Riccardo E. Marioni, Elena Colicino, Marjolein J. Peters, Cavin K. Ward-Caviness, Pei-Chien Tsai, Nicholas S. Roetker, Allan C. Just, Ellen W. Demerath, Weihua Guan, Jan Bressler, Myriam Fornage, Stephanie Studenski, Amy R. Vandiver, Ann Zenobia Moore, Toshiko Tanaka, Douglas P. Kiel, Liming Liang, Pantel Vokonas, Joel Schwartz, Kathryn L. Lunetta, Joanne M. Murabito, Stefania Bandinelli, Dena G. Hernandez, David Melzer, Michael Nalls, Luke C. Pilling, Timothy R. Price, Andrew B. Singleton, Christian Gieger, Rolf Holle, Anja Kretschmer, Florian Kronenberg, Sonja Kunze, Jakob Linseisen, Christine Meisinger, Wolfgang Rathmann, Melanie Waldenberger, Peter M. Visscher, Sonia Shah, Naomi R. Wray, Allan F. McRae, Oscar H. Franco, Albert Hofman, André G. Uitterlinden, Devin Absher, Themistocles Assimes, Morgan E. Levine, Ake T. Lu, Philip S. Tsao, Lifang Hou, JoAnn E. Manson, Cara L. Carty, Andrea Z. LaCroix, Alexander P. Reiner, Tim D. Spector, Andrew P. Feinberg, Daniel Levy, Andrea Baccarelli, Joyce Van Meurs, Jordana T. Bell, Annette Peters, Ian J. Deary, James S. Pankow, Luigi Ferrucci, and Steve Horvath. DNA methylation-based measures of biological age: Meta-analysis predicting time to death. *Aging*, 8(9):1844–1865, September 2016. ISSN 1945-4589. doi: 10.18632/aging.101020.

Jacob Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. ISSN 1939-1455, 0033-2909. doi: 10.1037/0033-2909.112.1.155.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 0 edition, May 2013. ISBN 978-1-134-74270-7. doi: 10.4324/9780203771587.

Alzheimer's Disease Sequencing Project Phenotype Harmonization Consortium. Adsp phenotype harmonization consortium (adsp-phc). https://adsp.niagads.org/funded-programs/phenotype-harmonization/, 2021. U24-AG074855.

Nick Corriveau-Lecavalier, Leland R Barnard, Hugo Botha, Jonathan Graff-Radford, Vijay K Ramanan, Jeyeon Lee, Ellen Dicks, Rosa Rademakers, Bradley F Boeve, Mary M Machulda, Julie A Fields, Dennis W Dickson, Neill Graff-Radford, David S Knopman, Val J Lowe, Ronald C Petersen, Clifford R Jack, and David T Jones. Uncovering the distinct macro-scale anatomy of dysexecutive and behavioural degenerative diseases. *Brain*, 147(4):1483–1496, April 2024. ISSN 0006-8950, 1460-2156. doi: 10.1093/brain/awad356.

Paul K. Crane, Adam Carle, Laura E. Gibbons, Philip Insel, Richard S. Mackin, Alden Gross, Richard N. Jones, Shubhabrata Mukherjee, S. McKay Curtis, Danielle Harvey, Michael Weiner, Dan Mungas, and for the Alzheimer's Disease Neuroimaging Initiative. Development and assessment of a composite score for memory in the alzheimer's disease neuroimaging initiative (adni). *Brain Imaging and Behavior*, 6(4):502–516, 2012. doi: 10.1007/s11682-012-9186-z.

Sebastian J Crutch, Manja Lehmann, Jonathan M Schott, Gil D Rabinovici, Martin N Rossor, and Nick C Fox. Posterior cortical atrophy. *The Lancet Neurology*, 11(2):170–178, 2012.

Lucas Paulo De Lima Camillo, Louis R. Lapierre, and Ritambhara Singh. A pan-tissue DNA-methylation epigenetic clock based on deep learning. *npj Aging*, 8(1):4, April 2022. ISSN 2731-6068. doi: 10.1038/s41514-022-00085-y.

Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, July 2006. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2006.01.021.

Gayatri Devi. A how-to guide for a precision medicine approach to the diagnosis and treatment of alzheimer's disease. *Frontiers in Aging Neuroscience*, 15:1213968, 2023.

Radhika Dhingra, Jamaji C. Nwanaji-Enwerem, Madeline Samet, and Cavin K. Ward-Caviness. DNA Methylation Age—Environmental Influences, Health Impacts, and Its Role in Environmental Epidemiology. *Current Environmental Health Reports*, 5(3):317–327, September 2018. ISSN 2196-5412. doi: 10.1007/s40572-018-0203-2.

Bradford C. Dickerson, Akram Bakkour, David H. Salat, Eric Feczko, Jenni Pacheco, Douglas N. Greve, Fran Grodstein, Christopher I. Wright, Deborah Blacker, H. Diana Rosas, Reisa A. Sperling, Alireza Atri, John H. Growdon, Bradley T. Hyman, John C. Morris, Bruce Fischl, and Randy L. Buckner. The Cortical Signature of Alzheimer's Disease: Regionally Specific Cortical Thinning Relates to Symptom Severity in Very Mild to Mild AD Dementia and is Detectable in Asymptomatic Amyloid-Positive Individuals. *Cerebral Cortex*, 19(3):497–510, March 2009. ISSN 1460-2199, 1047-3211. doi: 10.1093/cercor/bhn113.

Bradford C. Dickerson, Scott M. McGinnis, Chenjie Xia, Bruce H. Price, Alireza Atri, Melissa E. Murray, Mario F. Mendez, and David A. Wolk. Approach to atypical Alzheimer's disease and case studies of the major subtypes. *CNS Spectrums*, 22(6):439–449, December 2017. ISSN 1092-8529, 2165-6509. doi: 10.1017/s109285291600047x.

W. Ding, D. Kaur, S. Horvath, and W. Zhou. Comparative epigenome analysis using infinium dna methylation beadchips. *Briefings in bioinformatics, 24(1)*, 24(1), 2023. ISSN 1477-4054. doi: 10.1093/bib/bbac617,.

Trevor Doherty, Emma Dempster, Eilis Hannon, Jonathan Mill, Richie Poulton, David Corcoran, Karen Sugden, Ben Williams, Avshalom Caspi, Terrie E. Moffitt, Sarah Jane Delany, and Therese M. Murphy. A comparison of feature selection methodologies and learning algorithms in the development of a DNA methylation-based telomere length estimator. *BMC Bioinformatics*, 24(1):178, May 2023. ISSN 1471-2105. doi: 10.1186/s12859-023-05282-4.

R D Edgar, M J Jones, M J Meaney, G Turecki, and M S Kobor. BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *Translational Psychiatry*, 7(8): e1187–e1187, August 2017. ISSN 2158-3188. doi: 10.1038/tp.2017.171.

Daniel Ferreira, Chloë Verhagen, Juan Andrés Hernández-Cabrera, Lena Cavallin, Chun-Jie Guo, Urban Ekman, J-Sebastian Muehlboeck, Andrew Simmons, José Barroso, Lars-Olof Wahlund, and Eric Westman. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: Longitudinal trajectories and clinical applications. *Scientific Reports*, 7(1), April 2017. ISSN 2045-2322. doi: 10.1038/srep46263.

Daniel Ferreira, Agneta Nordberg, and Eric Westman. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology*, 94(10):436–448, March 2020. ISSN 0028-3878, 1526-632X. doi: 10.1212/wnl.0000000000009058.

B. Fischl. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22, January 2004. ISSN 1460-2199. doi: 10.1093/cercor/bhg087.

Jean-Philippe Fortin, Nicholas Cullen, Yvette I. Sheline, Warren D. Taylor, Irem Aselcioglu, Philip A. Cook, Phil Adams, Crystal Cooper, Maurizio Fava, Patrick J. McGrath, Melvin McInnis, Mary L. Phillips, Madhukar H. Trivedi, Myrna M. Weissman, and Russell T. Shinohara. Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167: 104–120, February 2018. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2017.11.024.

Lukas Frontzkowski, Michael Ewers, Matthias Brendel, Davina Biel, Rik Ossenkoppele, Paul Hager, Anna Steward, Anna Dewenter, Sebastian Römer, Anna Rubinski, Katharina Buerger, Daniel Janowitz, Alexa Pichet Binette, Ruben Smith, Olof Strandberg, Niklas Mattsson Carlgren, Martin Dichgans, Oskar Hansson, and Nicolai Franzmeier. Earlier Alzheimer's disease onset is associated with tau pathology in brain hub regions and facilitated tau spreading. *Nature Communications*, 13(1), August 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32592-7.

C. J. Galton. Atypical and typical presentations of Alzheimer's disease: A clinical, neuropsychological, neuroimaging and pathological study of 13 cases. *Brain*, 123(3):484–498, March 2000. ISSN 1460-2156. doi: 10.1093/brain/123.3.484.

X Gao, Y Wang, Z Song, M Jiang, T Huang, and A A Baccarelli. Early-life risk factors, accelerated biological aging and the late-life risk of mortality and morbidity. *QJM: An International Journal of Medicine*, 117(4):257–268, April 2024. ISSN 1460-2725, 1460-2393. doi: 10.1093/qjmed/hcad247.

José María García-Aznar, Sara Alonso Alvarez, and Teresa Bernal del Castillo. Pivotal role of bcl11b in the immune, hematopoietic and nervous systems: a review of the bcl11b-associated phenotypes from the genetic perspective. *Genes & Immunity*, 25(3):232–241, 2024.

Vadim N. Gladyshev. Aging: Progressive decline in fitness due to the rising deleteriome adjusted by genetic, environmental, and stochastic processes. *Aging Cell*, 15(4):594–602, 2016. doi: 10.1111/acel.12480.

David Goldberg, Anil R. Wadhwani, Nadia Dehghani, Lasya P. Sreepada, Hongxiang Fu, Philip L. De Jager, David A. Bennett, David A. Wolk, Edward B. Lee, PART Working Group, Kurt Farrell, John F. Crary, Wanding Zhou, and Corey T. McMillan. Epigenetic signatures of regional tau

pathology and cognition in the aging and pathological brain, November 2024.

David C Goldberg, Cameron Cloud, Sol Moe Lee, Bret Barnes, Steven Gruber, Elliot Kim, Anita Pottekat, Maximillian S Westphal, Luana McAuliffe, Elisa Majounie, et al. Scalable screening of ternary-code dna methylation dynamics associated with human traits. *Cell Genomics*, 2025.

Mitzi M Gonzales, Valentina R Garbarino, Erin Pollet, Juan P Palavicini, Dean L Kellogg, Ellen Kraig, Miranda E Orr, et al. Biological aging processes underlying cognitive decline and neurodegenerative disease. *The Journal of clinical investigation*, 132(10), 2022.

Maria Luisa Gorno-Tempini, Argye E Hillis, Sandra Weintraub, Andrew Kertesz, Mario Mendez, Stefano F Cappa, Jennifer M Ogar, Jonathan D Rohrer, Steven Black, Bradley F Boeve, et al. Classification of primary progressive aphasia and its variants. *Neurology*, 76(11):1006–1014, 2011.

Cheryl L Grady, JV Haxby, B Horwitz, M Sundaram, G Berg, M Schapiro, RP Friedland, and SI Rapoport. Longitudinal study of the early neuropsychological and cerebral metabolic changes in dementia of the alzheimer type. *Journal of clinical and experimental neuropsychology*, 10(5): 576–596, 1988.

Jonathan Graff-Radford, Keir X X Yong, Liana G Apostolova, Femke H Bouwman, Maria Carrillo, Bradford C Dickerson, Gil D Rabinovici, Jonathan M Schott, David T Jones, and Melissa E Murray. New insights into atypical Alzheimer's disease in the era of biomarkers. *The Lancet Neurology*, 20(3):222–234, March 2021. ISSN 1474-4422. doi: 10.1016/s1474-4422(20)30440-3.

Anders Gustavsson, Nicholas Norton, Thomas Fast, Lutz Frölich, Jean Georges, Drew Holzapfel, Tunahan Kirabali, Pierre Krolak-Salmon, Paolo M Rossini, Maria Teresa Ferretti, et al. Global estimates on the number of persons across the alzheimer's disease continuum. *Alzheimer's & Dementia*, 19(2):658–670, 2023.

Bahman Guyuron, David J Rowe, Adam Bryce Weinfeld, Yashar Eshraghi, Amir Fathi, and Seree Iamphongsai. Factors contributing to the facial aging of identical twins. *Plastic and reconstructive surgery*, 123(4):1321–1331, 2009.

Mohamad Habes, Deborah Janowitz, Guray Erus, Jon B Toledo, Susan M Resnick, Jimit Doshi, Sandra Van der Auwera, Katharina Wittfeld, Katrin Hegenscheid, Norbert Hosten, et al. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with alzheimer disease atrophy patterns. *Translational psychiatry*, 6(4):e775–e775, 2016.

Mohamad Habes, Raymond Pomponio, Haochang Shou, Jimit Doshi, Elizabeth Mamourian, Guray Erus, Ilya Nasrallah, Lenore J Launer, Tanweer Rashid, Murat Bilgel, et al. The brain chart of aging: machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the istaging consortium of 10,216 harmonized mr scans. *Alzheimer's & Dementia*, 17(1):89–102, 2021.

Eilis Hannon, Katie Lunnon, Leonard Schalkwyk, and Jonathan Mill. Interindividual methylomic

variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics*, 10(11):1024–1032, November 2015. ISSN 1559-2294, 1559-2308. doi: 10.1080/15592294.2015.1100786.

Gregory Hannum, Justin Guinney, Ling Zhao, Li Zhang, Guy Hughes, SriniVas Sadda, Brandy Klotzle, Marina Bibikova, Jian-Bing Fan, Yuan Gao, Rob Deconde, Menzies Chen, Indika Rajapakse, Stephen Friend, Trey Ideker, and Kang Zhang. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, 49(2):359–367, January 2013. ISSN 1097-2765. doi: 10.1016/j.molcel.2012.10.016.

Oskar Hansson, John Seibyl, Erik Stomrud, Henrik Zetterberg, John Q. Trojanowski, Tobias Bittner, Valeria Lifke, Veronika Corradini, Udo Eichenlaub, Richard Batrla, Katharina Buck, Katharina Zink, Christina Rabe, Kaj Blennow, Leslie M. Shaw, for the Swedish BioFINDER study group, and Alzheimer's Disease Neuroimaging Initiative. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimer's & Dementia*, 14(11):1470–1481, November 2018. ISSN 1552-5260, 1552-5279. doi: 10.1016/j.jalz.2018.01.010.

Maya L Henry and Maria Luisa Gorno-Tempini. The logopenic variant of primary progressive aphasia. *Current opinion in neurology*, 23(6):633–637, 2010.

Albert T. Higgins-Chen, Kyra L. Thrush, and Morgan E. Levine. Aging biomarkers and the brain. *Seminars in Cell & Developmental Biology*, 116:180–193, August 2021. ISSN 10849521. doi: 10.1016/j.semcdb.2021.01.003.

Albert T. Higgins-Chen, Kyra L. Thrush, Yunzhang Wang, Christopher J. Minteer, Pei-Lun Kuo, Meng Wang, Peter Niimi, Gabriel Sturm, Jue Lin, Ann Zenobia Moore, Stefania Bandinelli, Christiaan H. Vinkers, Eric Vermetten, Bart P. F. Rutten, Elbert Geuze, Cynthia Okhuijsen-Pfeifer, Marte Z. Van Der Horst, Stefanie Schreiter, Stefan Gutwinski, Jurjen J. Luykx, Martin Picard, Luigi Ferrucci, Eileen M. Crimmins, Marco P. Boks, Sara Hägg, Tina T. Hu-Seliger, and Morgan E. Levine. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and longitudinal tracking. *Nature Aging*, 2(7):644–661, July 2022. ISSN 2662-8465. doi: 10.1038/s43587-022-00248-2.

Steve Horvath. DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10), December 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-10-r115.

Steve Horvath, Junko Oshima, George M Martin, Ake T Lu, Austin Quach, Howard Cohen, Sarah Felton, Mieko Matsuyama, Donna Lowe, Sylwia Kabacik, et al. Epigenetic clock for skin and blood cells applied to hutchinson gilford progeria syndrome and ex vivo studies. *Aging (Albany NY)*, 10(7):1758, 2018.

Gyujoon Hwang, Ahmed Abdulkadir, Guray Erus, Mohamad Habes, Raymond Pomponio, Haochang Shou, Jimit Doshi, Elizabeth Mamourian, Tanweer Rashid, Murat Bilgel, Yong Fan, Aristeidis Sotiras, Dhivya Srinivasan, John C. Morris, Marilyn S. Albert, Nick R. Bryan, Su-

san M. Resnick, Ilya M. Nasrallah, Christos Davatzikos, David A. Wolk, from the iSTAGING consortium, and for the ADNI. Disentangling Alzheimer's disease neurodegeneration from typical brain ageing using machine learning. *Brain Communications*, 4(3):fcac117, May 2022. ISSN 2632-1297. doi: 10.1093/braincomms/fcac117.

Clifford R. Jack, J. Scott Andrews, Thomas G. Beach, Teresa Buracchio, Billy Dunn, Ana Graf, Oskar Hansson, Carole Ho, William Jagust, Eric McDade, Jose Luis Molinuevo, Ozioma C. Okonkwo, Luca Pani, Michael S. Rafii, Philip Scheltens, Eric Siemers, Heather M. Snyder, Reisa Sperling, Charlotte E. Teunissen, and Maria C. Carrillo. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, page alz.13859, June 2024. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.13859.

Clifford R Jack Jr, David A Bennett, Kaj Blennow, Maria C Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M Holtzman, William Jagust, Frank Jessen, Jason Karlawish, et al. Nia-aa research framework: toward a biological definition of alzheimer's disease. *Alzheimer's & dementia*, 14(4):535–562, 2018.

Purva Jain, Alexandra M. Binder, Brian Chen, Humberto Parada, Linda C. Gallo, John Alcaraz, Steve Horvath, Parveen Bhatti, Eric A. Whitsel, Kristina Jordahl, Andrea A. Baccarelli, Lifang Hou, James D. Stewart, Yun Li, Jamie N. Justice, and Andrea Z. LaCroix. Analysis of Epigenetic Age Acceleration and Healthy Longevity Among Older US Women. *JAMA Network Open*, 5(7): e2223285, July 2022. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2022.23285.

Tianye Jia, Congying Chu, Yun Liu, Jenny Van Dongen, Evangelos Papastergios, Nicola J. Armstrong, Mark E. Bastin, Tania Carrillo-Roa, Anouk Den Braber, Mathew Harris, Rick Jansen, Jingyu Liu, Michelle Luciano, Anil P. S. Ori, Roberto Roiz Santiañez, Barbara Ruggeri, Daniil Sarkisyan, Jean Shin, Kim Sungeun, Diana Tordesillas Gutiérrez, Dennis Van'T Ent, David Ames, Eric Artiges, Georgy Bakalkin, Tobias Banaschewski, Arun L. W. Bokde, Henry Brodaty, Uli Bromberg, Rachel Brouwer, Christian Büchel, Erin Burke Quinlan, Wiepke Cahn, Greig I. De Zubicaray, Stefan Ehrlich, Tomas J. Ekström, Herta Flor, Juliane H. Fröhner, Vincent Frouin, Hugh Garavan, Penny Gowland, Andreas Heinz, Jacqueline Hoare, Bernd Ittermann, Neda Jahanshad, Jiyang Jiang, John B. Kwok, Nicholas G. Martin, Jean-Luc Martinot, Karen A. Mather, Katie L. McMahon, Allan F. McRae, Frauke Nees, Dimitri Papadopoulos Orfanos, Tomáš Paus, Luise Poustka, Philipp G. Sämann, Peter R. Schofield, Michael N. Smolka, Dan J. Stein, Lachlan T. Strike, Jalmar Teeuw, Anbupalam Thalamuthu, Julian Trollor, Henrik Walter, Joanna M. Wardlaw, Wei Wen, Robert Whelan, Liana G. Apostolova, Elisabeth B. Binder, Dorret I. Boomsma, Vince Calhoun, Benedicto Crespo-Facorro, Ian J. Deary, Hilleke Hulshoff Pol, Roel A. Ophoff, Zdenka Pausova, Perminder S. Sachdev, Andrew Saykin, Margaret J. Wright, Paul M. Thompson, Gunter Schumann, and Sylvane Desrivières. Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes: Findings from the ENIGMA Epigenetics Working Group. *Molecular Psychiatry*, 26(8):3884–3895, August 2021. ISSN 1359-4184, 1476-5578. doi: 10.1038/s41380-019-0605-z.

Zbigniew Jost and Sylwester Kujach. Understanding cognitive decline in aging: Mechanisms and mitigation strategies–a narrative review. *Clinical Interventions in Aging*, pages 459–469, 2025.

Juulia Jylhävä, Nancy L. Pedersen, and Sara Hägg. Biological Age Predictors. *EBioMedicine*, 21: 29–36, July 2017. ISSN 2352-3964. doi: 10.1016/j.ebiom.2017.03.046.

Diljeet Kaur, Sol Moe Lee, David Goldberg, Nathan J Spix, Toshinori Hinoue, Hong-Tao Li, Varun B Dwaraka, Ryan Smith, Hui Shen, Gangning Liang, et al. Comprehensive evaluation of the infinium human methylationepic v2 beadchip. *Epigenetics communications*, 3(1):6, 2023.

Esther L.G.E. Koedam, Vivian Lauffer, Annelies E. Van Der Vlies, Wiesje M. Van Der Flier, Philip Scheltens, and Yolande A.L. Pijnenburg. Early-Versus Late-Onset Alzheimer's Disease: More than Age Alone. *Journal of Alzheimer's Disease*, 19(4):1401–1408, March 2010. ISSN 1875-8908, 1387-2877. doi: 10.3233/jad-2010-1337.

Yulia Komleva, Kristina Shpiliukova, Nikolai Bondar, Alla Salmina, Elena Khilazheva, Sergey Illarioshkin, and Michael Piradov. Decoding brain aging trajectory: predictive discrepancies, genetic susceptibilities, and emerging therapeutic strategies. *Frontiers in Aging Neuroscience*, 17:1562453, 2025.

Naomi Kouri, Isabelle Frankenhauser, Zhongwei Peng, Sydney A. Labuzan, Baayla D. C. Boon, Christina M. Moloney, Cyril Pottier, Daniel P. Wickland, Kelsey Caetano-Anolles, Nick Corriveau-Lecavalier, Jessica F. Tranovich, Ashley C. Wood, Kelly M. Hinkle, Sarah J. Lincoln, A. J. Spychalla, Matthew L. Senjem, Scott A. Przybelski, Erica Engelberg-Cook, Christopher G. Schwarz, Rain S. Kwan, Elizabeth R. Lesser, Julia E. Crook, Rickey E. Carter, Owen A. Ross, Christian Lachner, Nilüfer Ertekin-Taner, Tanis J. Ferman, Julie A. Fields, Mary M. Machulda, Vijay K. Ramanan, Aivi T. Nguyen, R. Ross Reichard, David T. Jones, Jonathan Graff-Radford, Bradley F. Boeve, David S. Knopman, Ronald C. Petersen, Clifford R. Jack, Kejal Kantarci, Gregory S. Day, Ranjan Duara, Neill R. Graff-Radford, Dennis W. Dickson, Val J. Lowe, Prashanthi Vemuri, and Melissa E. Murray. Clinicopathologic Heterogeneity and Glial Activation Patterns in Alzheimer Disease. *JAMA Neurology*, 81(6):619, June 2024. ISSN 2168-6149. doi: 10.1001/jamaneurol.2024.0784.

Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016.

Fan Lam, James Chu, Ji Sun Choi, Chang Cao, T. Kevin Hitchens, Scott K. Silverman, Zhi-Pei Liang, Ryan N. Dilger, Gene E. Robinson, and King C. Li. Epigenetic MRI: Noninvasive imaging of DNA methylation in the brain. *Proceedings of the National Academy of Sciences*, 119(10): e2119891119, March 2022. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2119891119.

Susan M. Landau, Christopher Breault, Abhinay D. Joshi, Michael Pontecorvo, Chester A. Mathis, William J. Jagust, and Mark A. Mintun. Amyloid-β Imaging with Pittsburgh Compound B and Florbetapir: Comparing Radiotracers and Quantification Methods. *Journal of Nuclear Medicine*, 54(1):70–77, January 2013. ISSN 0161-5505, 2159-662X. doi: 10.2967/jnumed.112.109009.

S. Lee, C. Loo, R. Prasasya, M. Bartolomei, R. Kohli, and W. Zhou. Low-input and single-cell methods for infinium dna methylation beadchips. *Nucleic acids research*, 2024. ISSN 1362-4962. doi: 10.1093/nar/gkae127.

Jenna M Leser, Anicca Harriot, Heather V Buck, Christopher W Ward, and Joseph P Stains. Aging, osteo-sarcopenia, and musculoskeletal mechano-transduction. *Frontiers in rehabilitation sciences*, 2:782848, 2021.

Morgan E. Levine, Ake T. Lu, Austin Quach, Brian H. Chen, Themistocles L. Assimes, Stefania Bandinelli, Lifang Hou, Andrea A. Baccarelli, James D. Stewart, Yun Li, Eric A. Whitsel, James G Wilson, Alex P Reiner, Abraham Aviv, Kurt Lohman, Yongmei Liu, Luigi Ferrucci, and Steve Horvath. An epigenetic biomarker of aging for lifespan and healthspan. *Aging*, 10(4):573–591, April 2018. ISSN 1945-4589. doi: 10.18632/aging.101414.

Morgan E Levine, Albert Higgins-Chen, Kyra Thrush, Christopher Minteer, and Peter Niimi. Clock Work: Deconstructing the Epigenetic Clock Signals in Aging, Disease, and Reprogramming, February 2022.

Michael T Lin and M Flint Beal. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*, 443(7113):787–795, 2006.

Li Liu, Tsui S Cheung, Gregory W. Charville, and Thomas A. Rando. Epigenetic regulation of stem cell aging and senescence. *Current Opinion in Cell Biology*, 67:43–52, 2020. doi: 10.1016/j.ceb.2020.08.011.

Zaoqu Liu, Qimeng Liang, Yuqing Ren, Chunguang Guo, Xiaoyong Ge, Libo Wang, Quan Cheng, Peng Luo, Yi Zhang, and Xinwei Han. Immunosenescence: molecular mechanisms and diseases. *Signal transduction and targeted therapy*, 8(1):200, 2023.

Ake T. Lu, Austin Quach, James G. Wilson, Alex P. Reiner, Abraham Aviv, Kenneth Raj, Lifang Hou, Andrea A. Baccarelli, Yun Li, James D. Stewart, Eric A. Whitsel, Themistocles L. Assimes, Luigi Ferrucci, and Steve Horvath. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging*, 11(2):303–327, January 2019. ISSN 1945-4589. doi: 10.18632/aging.101684.

Ake T. Lu, Alexandra M. Binder, Joshua Zhang, Qi Yan, Alex P. Reiner, Simon R. Cox, Janie Corley, Sarah E. Harris, Pei-Lun Kuo, Ann Z. Moore, Stefania Bandinelli, James D. Stewart, Cuicui Wang, Elissa J. Hamlat, Elissa S. Epel, Joel D. Schwartz, Eric A. Whitsel, Adolfo Correa, Luigi Ferrucci, Riccardo E. Marioni, and Steve Horvath. DNA methylation GrimAge version 2. *Aging*, December 2022. ISSN 1945-4589. doi: 10.18632/aging.204434.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Xueying Lyu, Michael Tran Duong, Long Xie, Robin De Flores, Hayley Richardson, Gyujoon Hwang, Laura E. M. Wisse, Michael DiCalogero, Corey T. McMillan, John L. Robinson, Sharon X. Xie,

Edward B. Lee, David J. Irwin, Bradford C. Dickerson, Christos Davatzikos, Ilya M. Nasrallah, Paul A. Yushkevich, David A. Wolk, Sandhitsu R. Das, and for the Alzheimer's Disease Neuroimaging Initiative. Tau-neurodegeneration *mismatch* reveals vulnerability and resilience to comorbidities in Alzheimer's continuum. *Alzheimer's & Dementia*, 20(3):1586–1600, March 2024. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.13559.

Carlos López-Otín, Maria A. Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer. The hallmarks of aging. *Cell*, 153(6):1194–1217, 2013. doi: 10.1016/j.cell.2013.05.039.

Jordi A Matías-Guiu, María Nieves Cabrera-Martín, Teresa Moreno-Ramos, María Valles-Salgado, Marta Fernandez-Matarrubia, José Luis Carreras, and Jorge Matías-Guiu. Amyloid and fdg-pet study of logopenic primary progressive aphasia: evidence for the existence of two subtypes. *Journal of neurology*, 262(6):1463–1472, 2015.

Daniel L McCartney, Rosie M Walker, Stewart W Morris, Andrew M McIntosh, David J Porteous, and Kathryn L Evans. Identification of polymorphic and off-target probe binding sites on the illumina infinium methylationepic beadchip. *Genomics data*, 9:22–24, 2016.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020.

Lidija Milicic, Michael Vacher, Tenielle Porter, Vincent Doré, Samantha C. Burnham, Pierrick Bourgeat, Rosita Shishegar, James Doecke, Nicola J. Armstrong, Rick Tankard, Paul Maruff, Colin L. Masters, Christopher C. Rowe, Victor L. Villemagne, Simon M. Laws, Alzheimer's Disease Neuroimaging Initiative (ADNI), Michael Weiner, Paul Aisen, Ronald Petersen, Clifford R. Jack, William Jagust, John Q. Trojanowki, Arthur W. Toga, Laurel Beckett, Robert C. Green, Andrew J. Saykin, John C. Morris, Leslie M. Shaw, Enchi Liu, Tom Montine, Ronald G. Thomas, Michael Donohue, Sarah Walter, Devon Gessert, Tamie Sather, Gus Jiminez, Danielle Harvey, Matthew Bernstein, Nick Fox, Paul Thompson, Norbert Schuff, Charles DeCArli, Bret Borowski, Jeff Gunter, Matt Senjem, Prashanthi Vemuri, David Jones, Kejal Kantarci, Chad Ward, Robert A. Koeppe, Norm Foster, Eric M. Reiman, Kewei Chen, Chet Mathis, Susan Landau, Nigel J. Cairns, Erin Householder, Lisa Taylor Reinwald, Virginia Lee, Magdalena Korecka, Michal Figurski, Karen Crawford, Scott Neu, Tatiana M. Foroud, Steven Potkin, Li Shen, Faber Kelley, Sungeun Kim, Kwangsik Nho, Zaven Kachaturian, Richard Frank, Peter J. Snyder, Susan Molchan, Jeffrey Kaye, Joseph Quinn, Betty Lind, Raina Carter, Sara Dolen, Lon S. Schneider, Sonia Pawluczyk, Mauricio Beccera, Liberty Teodoro, Bryan M. Spann, James Brewer, Helen Vanderswag, Adam Fleisher, Judith L. Heidebrink, Joanne L. Lord, Ronald Petersen, Sara S. Mason, Colleen S. Albers, David Knopman, Kris Johnson, Rachelle S. Doody, Javier Villanueva Meyer, Munir Chowdhury, Susan Rountree, Mimi Dang, Yaakov Stern, Lawrence S. Honig, Karen L. Bell, Beau Ances, Maria Carroll, Sue Leon, Erin Householder, Mark A. Mintun, Stacy Schneider, Angela OliverNG, Randall Griffith, David Clark, David Geldmacher, John Brockington, Erik Roberson, Hillel Grossman, Effie Mitsis, Leyla deToledo-Morrell, Raj C. Shah, Ranjan Duara, Daniel Varon, Maria T. Greig, Peggy Roberts, Marilyn Albert, Chiadi Onyike, Daniel D.' Agostino, Stephanie Kielb, James E. Galvin, Dana M. Pogorelec, Brittany Cerbone, Christina A. Michel, Henry Rusinek, Mony J. de Leon, Lidia Glodzik, Susan De Santi, P. Murali Doraiswamy,

Jeffrey R. Petrella, Terence Z. Wong, Steven E. Arnold, Jason H. Karlawish, David A. Wolk, Charles D. Smith, Greg Jicha, Peter Hardy, Partha Sinha, Elizabeth Oates, Gary Conrad, Oscar L. Lopez, Mary Ann Oakley, Donna M. Simpson, Anton P. Porsteinsson, Bonnie S. Goldstein, Kim Martin, Kelly M. Makino, M. Saleem Ismail, Connie Brand, Ruth A. Mulnard, Gaby Thai, Catherine Mc Adams Ortiz, Kyle Womack, Dana Mathews, Mary Quiceno, Ramon Diaz Arrastia, Richard King, Myron Weiner, Kristen Martin Cook, Michael DeVous, Allan I. Levey, James J. Lah, Janet S. Cellar, Jeffrey M. Burns, Heather S. Anderson, Russell H. Swerdlow, Liana Apostolova, Kathleen Tingus, Ellen Woo, Daniel H. S. Silverman, Po H. Lu, George Bartzokis, Neill R. Graff Radford, Francine ParfittH, Tracy Kendall, Heather Johnson, Martin R. Farlow, Ann Marie Hake, Brandy R. Matthews, Scott Herring, Cynthia Hunt, Christopher H. van Dyck, Richard E. Carson, Martha G. MacAvoy, Howard Chertkow, Howard Bergman, Chris Hosein, Sandra Black, Bojana Stefanovic, Curtis Caldwell, Ging Yuek Robin Hsiung, Howard Feldman, Benita Mudge, Michele Assaly Past, Andrew Kertesz, John Rogers, Dick Trost, Charles Bernick, Donna Munic, Diana Kerwin, Marek Marsel Mesulam, Kristine Lipowski, Chuang Kuo Wu, Nancy Johnson, Carl Sadowsky, Walter Martinez, Teresa Villena, Raymond Scott Turner, Kathleen Johnson, Brigid Reynolds, Reisa A. Sperling, Keith A. Johnson, Gad Marshall, Meghan Frey, Jerome Yesavage, Joy L. Taylor, Barton Lane, Allyson Rosen, Jared Tinklenberg, Marwan N. Sabbagh, Christine M. Belden, Sandra A. Jacobson, Sherye A. Sirrel, Neil Kowall, Ronald Killiany, Andrew E. Budson, Alexander Norbash, Patricia Lynn Johnson, Thomas O. Obisesan, Saba Wolday, Joanne Allard, Alan Lerner, Paula Ogrocki, Leon Hudson, Evan Fletcher, Owen Carmichael, John Olichney, Charles DeCarli, Smita Kittur, Michael Borrie, T. Y. Lee, Rob Bartha, Sterling Johnson, Sanjay Asthana, Cynthia M. Carlsson, Steven G. Potkin, Adrian Preda, Dana Nguyen, Pierre Tariot, Adam Fleisher, Stephanie Reeder, Vernice Bates, Horacio Capote, Michelle Rainka, Douglas W. Scharre, Maria Kataki, Anahita Adeli, Earl A. Zimmerman, Dzintra Celmins, Alice D. Brown, Godfrey D. Pearlson, Karen Blank, Karen Anderson, Robert B. Santulli, Tamar J. Kitzmiller, Eben S. Schwartz, Kaycee M. SinkS, Jeff D. Williamson, Pradeep Garg, Franklin Watkins, Brian R. Ott, Henry Querfurth, Geoffrey Tremont, Stephen Salloway, Paul Malloy, Stephen Correia, Howard J. Rosen, Bruce L. Miller, Jacobo Mintzer, Kenneth Spicer, David Bachman, Elizabether Finger, Stephen Pasternak, Irina Rachinsky, John Rogers, Andrew Kertesz, Dick Drost, Nunzio Pomara, Raymundo Hernando, Antero Sarrael, Susan K. Schultz, Laura L. Boles Ponto, Hyungsub Shim, Karen Elizabeth Smith, Norman Relkin, Gloria Chaing, Lisa Raudin, Amanda Smith, Kristin Fargher, Balebail Ashok Raj, Australian Imaging Biomarkers and Lifestyle (AIBL) Study, Christopher Fowler, Stephanie R. Rainey-Smith, Sabine Bird, Julia Bomke, Pierrick Bourgeat, Belinda M. Brown, Samantha C. Burnham, Ashley I. Bush, Carolyn Chadunow, Steven Collins, James Doecke, Vincent Dore, Kathryn A. Ellis, Lis Evered, Amir Fazlollahi, Jurgen Fripp, Samantha L. Gardener, Simon Gibson, Robert Grenfell, Elise Harrison, Richard Head, Liang Jin, Adrian Kamer, Fiona Lamb, Nicola T. Lautenschlager, Simon M. Laws, Qiao-Xin Li, Lucy Lim, Yen Ying Lim, Andrea Louey, S. Lance Macaulay, Lucy Mackintosh, Ralph N. Martins, Paul Maruff, Colin L. Masters, Simon McBride, Lidija Milicic, Kelly Pertile, Tenielle Porter, Morgan Radler, Joanne Robertson, Mark Rodrigues, Christopher C. Rowe, Rebecca Rumble, Olivier Salvado, Greg Savage, Rosita Shishegar, Brendan Silbert, Magdalene Soh, Hamid R. Sohrabi, Kevin Taddei, Tania Taddei, Christine Thai, Brett Trounson, Regan Tyrrel, Michael Vacher, Shiji Varghese, Victor L. Villemagne, Michael Weinborn, Michael Woodward, Ying Xia, and David Ames. Comprehensive analysis of epigenetic clocks reveals associations between disproportionate biological ageing and hippocampal volume. *GeroScience*, 44

(3):1807–1823, June 2022. ISSN 2509-2715, 2509-2723. doi: 10.1007/s11357-022-00558-8.

Lidija Milicic, Tenielle Porter, Michael Vacher, and Simon M. Laws. Utility of DNA Methylation as a Biomarker in Aging and Alzheimer's Disease. *Journal of Alzheimer's Disease Reports*, 7(1): 475–503, November 2023. ISSN 2542-4823, 2542-4823. doi: 10.3233/ADR-220109.

Christiane Möller, Hugo Vrenken, Lize Jiskoot, Adriaan Versteeg, Frederik Barkhof, Philip Scheltens, and Wiesje M. Van Der Flier. Different patterns of gray matter atrophy in early- and late-onset Alzheimer's disease. *Neurobiology of Aging*, 34(8):2014–2022, August 2013. ISSN 0197-4580. doi: 10.1016/j.neurobiolaging.2013.02.013.

John C. Morris. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11):2412, November 1993. ISSN 0028-3878, 1526-632X. doi: 10.1212/wnl.43.11.2412-a.

Shubhabrata Mukherjee, Seo-Eun Choi, Michael L Lee, Phoebe Scollard, Emily H Trittschuh, Jesse Mez, Andrew J Saykin, Laura E Gibbons, R Elizabeth Sanders, Andrew F Zaman, Merilee A Teylan, Walter A Kukull, Lisa L Barnes, David A Bennett, Andrea Z Lacroix, Eric B Larson, Michael Cuccaro, Shannon Mercado, Logan Dumitrescu, Timothy J Hohman, and Paul K Crane. Cognitive domain harmonization and cocalibration in studies of older adults. *Neuropsychology*, 37(4):409–423, 2022. doi: 10.1037/neu0000835.

Melissa E Murray, Neill R Graff-Radford, Owen A Ross, Ronald C Petersen, Ranjan Duara, and Dennis W Dickson. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. *The Lancet Neurology*, 10(9):785–796, September 2011. ISSN 1474-4422. doi: 10.1016/s1474-4422(11)70156-9.

Sid E. O'Bryant. Staging Dementia Using Clinical Dementia Rating Scale Sum of Boxes Scores: A Texas Alzheimer's Research Consortium Study. *Archives of Neurology*, 65(8):1091, August 2008. ISSN 0003-9942. doi: 10.1001/archneur.65.8.1091.

Rik Ossenkoppele, Daniel R. Schonhaut, Michael Schöll, Samuel N. Lockhart, Nagehan Ayakta, Suzanne L. Baker, James P. O'Neil, Mustafa Janabi, Andreas Lazaris, Averill Cantwell, Jacob Vogel, Miguel Santos, Zachary A. Miller, Brianne M. Bettcher, Keith A. Vossel, Joel H. Kramer, Maria L. Gorno-Tempini, Bruce L. Miller, William J. Jagust, and Gil D. Rabinovici. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*, 139(5): 1551–1567, May 2016. ISSN 0006-8950, 1460-2156. doi: 10.1093/brain/aww027.

Rik Ossenkoppele, Chul Hyoung Lyoo, Jonas Jester-Broms, Carole H. Sudre, Hanna Cho, Young Hoon Ryu, Jae Yong Choi, Ruben Smith, Olof Strandberg, Sebastian Palmqvist, Joel Kramer, Adam L. Boxer, Maria L. Gorno-Tempini, Bruce L. Miller, Renaud La Joie, Gil D. Rabinovici, and Oskar Hansson. Assessment of Demographic, Genetic, and Imaging Variables Associated With Brain Resilience and Cognitive Resilience to Pathological Tau in Patients With Alzheimer Disease. *JAMA Neurology*, 77(5):632, May 2020. ISSN 2168-6149. doi: 10.1001/jamaneurol.2019.5154.

R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, 74(3):201–209, January 2010. ISSN 0028-3878, 1526-632X. doi: 10.1212/wnl.0b013e3181cb3e25.

Ruth Pidsley, Elena Zotenko, Timothy J. Peters, Mitchell G. Lawrence, Gail P. Risbridger, Peter Molloy, Susan Van Djik, Beverly Muhlhausler, Clare Stirzaker, and Susan J. Clark. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, 17(1), December 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-1066-1.

Angelina J. Polsinelli and Liana G. Apostolova. Atypical Alzheimer Disease Variants. *Continuum*, 28(3):676–701, June 2022. ISSN 1080-2371, 1538-6899. doi: 10.1212/con.0000000000001082.

Raymond Pomponio, Guray Erus, Mohamad Habes, Jimit Doshi, Dhivya Srinivasan, Elizabeth Mamourian, Vishnu Bashyam, Ilya M. Nasrallah, Theodore D. Satterthwaite, Yong Fan, Lenore J. Launer, Colin L. Masters, Paul Maruff, Chuanjun Zhuo, Henry Völzke, Sterling C. Johnson, Jurgen Fripp, Nikolaos Koutsouleris, Daniel H. Wolf, Raquel Gur, Ruben Gur, John Morris, Marilyn S. Albert, Hans J. Grabe, Susan M. Resnick, R. Nick Bryan, David A. Wolk, Russell T. Shinohara, Haochang Shou, and Christos Davatzikos. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, 208:116450, March 2020. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2019.116450.

Deepti Putcha, Yuta Katsumi, Alexandra Touroutoglou, Ani Eloyan, Alexander Taurone, Maryanne Thangarajah, Paul Aisen, Jeffrey L. Dage, Tatiana Foroud, Clifford R. Jack, Joel H. Kramer, Kelly N. H. Nudelman, Rema Raman, Prashanthi Vemuri, Alireza Atri, Gregory S. Day, Ranjan Duara, Neill R. Graff-Radford, Ian M. Grant, Lawrence S. Honig, Erik C. B. Johnson, David T. Jones, Joseph C. Masdeu, Mario F. Mendez, Erik Musiek, Chiadi U. Onyike, Meghan Riddle, Emily Rogalski, Stephen Salloway, Sharon Sha, R. Scott Turner, Thomas S. Wingo, David A. Wolk, Kyle Womack, Maria C. Carrillo, Gil D. Rabinovici, Bradford C. Dickerson, Liana G. Apostolova, Dustin B. Hammers, the LEADS Consortium, Laurel Beckett, Bernardino Ghetti, Lea T. Grinberg, Dustin Hammers, Clifford R. Jack, Kala Kirby, Robert Koeppe, Walter A. Kukull, Renaud La Joie, Julien Lagarde, Melissa E. Murray, Kathy Newell, Angelina Polsinelli, Arthur Toga, David Clark, Ian Grant, and Lawrence S. Honig. Heterogeneous clinical phenotypes of sporadic early-onset Alzheimer's disease: A neuropsychological data-driven approach. *Alzheimer's Research & Therapy*, 17(1), February 2025. ISSN 1758-9193. doi: 10.1186/s13195-025-01689-8.

André Rey. *L'examen clinique en psychologie*. Presses Universitaires de France, Paris, 1964.

Jarod Rutledge, Hamilton Oh, and Tony Wyss-Coray. Measuring biological age using omics data. *Nature Reviews Genetics*, 23(12):715–727, December 2022. ISSN 1471-0056, 1471-0064. doi: 10.1038/s41576-022-00511-7.

Lucas A Salas, Devin C Koestler, Rondi A Butler, Helen M Hansen, John K Wiencke, Karl T Kelsey, and Brock C Christensen. An optimized library for reference-based deconvolution of whole-blood

biospecimens assayed using the illumina humanmethylationepic beadarray. *Genome biology*, 19 (1):64, 2018.

Ahmed Salih, Thomas Nichols, Liliana Szabo, Steffen E Petersen, and Zahra Raisi-Estabragh. Conceptual Overview of Biological Age Estimation. *Aging and disease*, 14(3):583, 2023. ISSN 2152-5250. doi: 10.14336/ad.2022.1107.

Nienke ME Scheltens, Betty M Tijms, Teddy Koene, Frederik Barkhof, Charlotte E Teunissen, Steffen Wolfsgruber, Michael Wagner, Johannes Kornhuber, Oliver Peters, Brendan I Cohn-Sheehy, et al. Cognitive subtypes of probable alzheimer's disease robustly identified in four cohorts. *Alzheimer's & Dementia*, 13(11):1226–1236, 2017.

Gemma L Shireby, Jonathan P Davies, Paul T Francis, Joe Burrage, Emma M Walker, Grant W A Neilson, Aisha Dahir, Alan J Thomas, Seth Love, Rebecca G Smith, Katie Lunnon, Meena Kumari, Leonard C Schalkwyk, Kevin Morgan, Keeley Brookes, Eilis Hannon, and Jonathan Mill. Recalibrating the epigenetic clock: Implications for assessing biological age in the human cortex. *Brain*, 143(12):3763–3775, December 2020. ISSN 0006-8950, 1460-2156. doi: 10.1093/brain/awaa334.

Tarun D Singh, Keith A Josephs, Mary M Machulda, Daniel A Drubach, Liana G Apostolova, Val J Lowe, and Jennifer L Whitwell. Clinical, fdg and amyloid pet imaging in posterior cortical atrophy. *Journal of neurology*, 262(6):1483–1492, 2015.

Daniel W. Sirkis, Luke W. Bonham, Taylor P. Johnson, Renaud La Joie, and Jennifer S. Yokoyama. Dissecting the clinical heterogeneity of early-onset Alzheimer's disease. *Molecular Psychiatry*, 27 (6):2674–2688, June 2022. ISSN 1359-4184, 1476-5578. doi: 10.1038/s41380-022-01531-9.

Rebecca G. Smith, Ehsan Pishva, Gemma Shireby, Adam R. Smith, Janou A. Y. Roubroeks, Eilis Hannon, Gregory Wheildon, Diego Mastroeni, Gilles Gasparoni, Matthias Riemenschneider, Armin Giese, Andrew J. Sharp, Leonard Schalkwyk, Vahram Haroutunian, Wolfgang Viechtbauer, Daniel L. A. Van Den Hove, Michael Weedon, Danielle Brokaw, Paul T. Francis, Alan J. Thomas, Seth Love, Kevin Morgan, Jörn Walter, Paul D. Coleman, David A. Bennett, Philip L. De Jager, Jonathan Mill, and Katie Lunnon. A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nature Communications*, 12(1):3517, June 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23243-4.

Lasya P. Sreepada, Christopher A. Brown, Sandhitsu R. Das, Paul A. Yushkevich, David A. Wolk, and Corey T. McMillan. Biological age acceleration in Alzheimer's disease modulates relative cortical to medial temporal lobe neurodegeneration. *Neurobiology of Aging*, 153:21–29, September 2025. ISSN 01974580. doi: 10.1016/j.neurobiolaging.2025.06.003.

Rainulf A. Stelzmann, H. Norman Schnitzlein, and F. Reed Murtagh. An english translation of alzheimer's 1907 paper, "über eine eigenartige erkankung der hirnrinde". *Clinical Anatomy*, 8(6): 429–431, January 1995. ISSN 0897-3806, 1098-2353. doi: 10.1002/ca.980080612.

Anna Steward, Davina Biel, Anna Dewenter, Sebastian Roemer, Fabian Wagner, Amir Dehsarvi, Saima Rathore, Diana Otero Svaldi, Ixavier Higgins, Matthias Brendel, Martin Dichgans, Sergey Shcherbinin, Michael Ewers, and Nicolai Franzmeier. ApoE4 and Connectivity-Mediated Spreading of Tau Pathology at Lower Amyloid Levels. *JAMA Neurology*, 80(12):1295, December 2023. ISSN 2168-6149. doi: 10.1001/jamaneurol.2023.4038.

Russell H. Swerdlow. Pathogenesis of Alzheimer's disease. *Clinical Interventions in Aging*, 2(3): 347–359, 2007. ISSN 1176-9092.

E. M. Taylor. Scoring procedure for the rey auditory verbal learning test. *Archives of Clinical Neuropsychology*, 3:93–97, 1959.

Andrew E. Teschendorff and Steve Horvath. Epigenetic ageing clocks: Statistical methods and emerging computational challenges. *Nature Reviews Genetics*, 26(5):350–368, May 2025. ISSN 1471-0056, 1471-0064. doi: 10.1038/s41576-024-00807-w.

Andrew E Teschendorff, Charles E Breeze, Shijie C Zheng, and Stephan Beck. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC bioinformatics*, 18(1):105, 2017.

Joseph Therriault, Andrea L. Benedet, Tharick A. Pascoal, Sulantha Mathotaarachchi, Mira Chamoun, Melissa Savard, Emilie Thomas, Min Su Kang, Firoza Lussier, Cecile Tissot, Marlee Parsons, Muhammad Naveed Iqbal Qureshi, Paolo Vitali, Gassan Massarweh, Jean-Paul Soucy, Soham Rej, Paramita Saha-Chaudhuri, Serge Gauthier, and Pedro Rosa-Neto. Association of Apolipoprotein E $E4$ With Medial Temporal Tau Independent of Amyloid-β. *JAMA Neurology*, 77(4):470, April 2020. ISSN 2168-6149. doi: 10.1001/jamaneurol.2019.4421.

Huige Tong, Varun B. Dwaraka, Qingwen Chen, Qi Luo, Jessica A. Lasky-Su, Ryan Smith, and Andrew E. Teschendorff. Quantifying the stochastic component of epigenetic aging. *Nature Aging*, 4(6):886–901, May 2024. ISSN 2662-8465. doi: 10.1038/s43587-024-00600-8.

Alexandra Touroutoglou, Yuta Katsumi, Michael Brickhouse, Alexander Zaitsev, Ryan Eckbo, Paul Aisen, Laurel Beckett, Jeffrey L. Dage, Ani Eloyan, Tatiana Foroud, Bernardino Ghetti, Percy Griffin, Dustin Hammers, Clifford R. Jack, Joel H. Kramer, Leonardo Iaccarino, Renaud La Joie, Nidhi S Mundada, Robert Koeppe, Walter A. Kukull, Melissa E. Murray, Kelly Nudelman, Angelina J. Polsinelli, Malia Rumbaugh, David N. Soleimani-Meigooni, Arthur Toga, Prashanthi Vemuri, Alireza Atri, Gregory S. Day, Ranjan Duara, Neill R. Graff-Radford, Lawrence S. Honig, David T. Jones, Joseph C. Masdeu, Mario F. Mendez, Erik Musiek, Chiadi U. Onyike, Meghan Riddle, Emily Rogalski, Stephen Salloway, Sharon Sha, R. Scott Turner, Thomas S. Wingo, David A. Wolk, Kyle Womack, Maria C. Carrillo, Gil D. Rabinovici, Liana G. Apostolova, Bradford C. Dickerson, and the LEADS Consortium. The Sporadic Early-onset Alzheimer's Disease Signature Of Atrophy: Preliminary Findings From The Longitudinal Early-onset Alzheimer's Disease Study (LEADS) Cohort. *Alzheimer's & Dementia*, 19(S9), November 2023. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.13466.

Ryan A Townley, Jonathan Graff-Radford, William G Mantyh, Hugo Botha, Angelina J Polsinelli, Scott A Przybelski, Mary M Machulda, Ahmed T Makhlouf, Matthew L Senjem, Melissa E Murray, Ross R Reichard, Rodolfo Savica, Bradley F Boeve, Daniel A Drubach, Keith A Josephs, David S Knopman, Val J Lowe, Clifford R Jack, Ronald C Petersen, and David T Jones. Progressive dysexecutive syndrome due to Alzheimer's disease: A description of 55 cases and comparison to other phenotypes. *Brain Communications*, 2(1), January 2020. ISSN 2632-1297. doi: 10.1093/braincomms/fcaa068.

Angeliki Vakka, Junco S Warren, and Konstantinos Drosatos. Cardiovascular aging: from cellular and molecular changes to therapeutic interventions. *The journal of cardiovascular aging*, 3(3):23, 2023.

S. N. Vandekar, E. C. Edmonds, L. A. McCollum, B. C. Dickerson, T. J. Ferman, T. J. Hohman, J. K. Johnson, K. Kantarci, K. A. Koenig, M. F. Mendez, G. D. Rabinovici, J. M. Ringman, S. Salloway, D. W. Scharre, G. Tosto, S. Weintraub, S. X. Xie, and L. G. Apostolova. The longitudinal early-onset alzheimer's disease study (leads). *Alzheimer's & Dementia*, 17(S6):e051573, 2021. doi: 10.1002/alz.051573.

S. N. Vandekar, E. C. Edmonds, L. A. McCollum, B. C. Dickerson, T. J. Ferman, T. J. Hohman, J. K. Johnson, K. Kantarci, K. A. Koenig, M. F. Mendez, G. D. Rabinovici, J. M. Ringman, S. Salloway, D. W. Scharre, G. Tosto, S. Weintraub, S. X. Xie, and L. G. Apostolova. Profiling baseline performance on the longitudinal early-onset alzheimer's disease study (leads). *Alzheimer's & Dementia*, 19(11):4280–4293, 2023. doi: 10.1002/alz.13568.

Aparna Vasanthakumar, Justin W. Davis, Kenneth Idler, Jeffrey F. Waring, Elizabeth Asque, Bridget Riley-Gillis, Shaun Grosskurth, Gyan Srivastava, Sungeun Kim, Kwangsik Nho, Kelly N. H. Nudelman, Kelley Faber, Yu Sun, Tatiana M. Foroud, Karol Estrada, Liana G. Apostolova, Qingqin S. Li, Andrew J. Saykin, and for the Alzheimer's Disease Neuroimaging Initiative (ADNI). Harnessing peripheral DNA methylation differences in the Alzheimer's Disease Neuroimaging Initiative (ADNI) to reveal novel biomarkers of disease. *Clinical Epigenetics*, 12(1), December 2020. ISSN 1868-7075, 1868-7083. doi: 10.1186/s13148-020-00864-y.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Shuguang Wang, Zuning Liao, Qiying Zhang, Xinyuan Han, Changqing Liu, and Jin Wang. Mitochondrial dysfunction in alzheimer's disease: a key frontier for future targeted therapies. *Frontiers in immunology*, 15:1484373, 2025.

Si-Yu Wang, Peng-Yu Gong, Yan E, Ying-Dong Zhang, and Teng Jiang. The role of treml2 in alzheimer's disease. *Journal of Alzheimer's Disease*, 76(3):799–806, 2020a.

Xian Wang, Younghee Park, Jae Kyoung Lee, Sunghoon Kim, and Tae-Min Kim. Robust partial reference-free cell composition estimation from tissue expression profiles. *Scientific Reports*, 10 (1):1–12, 2020b. doi: 10.1038/s41598-020-61484-9.

Yucheng Wang, Olivia A. Grant, Xiaojun Zhai, Klaus D. Mcdonald-Maier, and Leonardo C. Schalkwyk. Insights into ageing rates comparison across tissues from recalibrating cerebellum DNA methylation clock. *GeroScience*, August 2023. ISSN 2509-2723. doi: 10.1007/s11357-023-00871-w.

Michael W. Weiner, Dallas P. Veitch, Paul S. Aisen, Laurel A. Beckett, Nigel J. Cairns, Robert C. Green, Danielle Harvey, Clifford R. Jack, William Jagust, John C. Morris, Ronald C. Petersen, Jennifer Salazar, Andrew J. Saykin, Leslie M. Shaw, Arthur W. Toga, John Q. Trojanowski, and Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571, May 2017. ISSN 1552-5260, 1552-5279. doi: 10.1016/j.jalz.2016.10.006.

Jennifer L Whitwell. Atypical clinical variants of alzheimer's disease: are they really atypical? *Frontiers in Neuroscience*, 18:1352822, 2024.

Jennifer L Whitwell, Maria M Shiung, SA Przybelski, Stephen D Weigand, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Clifford R Jack Jr. Mri patterns of atrophy associated with progression to ad in amnestic mild cognitive impairment. *Neurology*, 70(7):512–520, 2008.

Robert S Wilson, Sue E Leurgans, Patricia A Boyle, and David A Bennett. Cognitive decline in prodromal alzheimer disease and mild cognitive impairment. *Archives of neurology*, 68(3): 351–356, 2011.

David A. Wolk, Peter T. Nelson, Liana Apostolova, Konstantinos Arfanakis, Patricia A. Boyle, Cynthia M. Carlsson, Nick Corriveau-Lecavalier, Penny Dacks, Bradford C. Dickerson, Kimiko Domoto-Reilly, Brittany N. Dugger, Rebecca Edelmayer, David W. Fardo, Michel J. Grothe, Timothy J. Hohman, David J. Irwin, Gregory A. Jicha, David T. Jones, Claudia H. Kawas, Edward B. Lee, Karen Lincoln, Gladys E. Maestre, Elizabeth C. Mormino, Chiadi U. Onyike, Ronald C. Petersen, Gil D. Rabinovici, Rosa Rademakers, Rema Raman, Katya Rascovsky, Robert A. Rissman, Emily Rogalski, Philip Scheltens, Reisa A. Sperling, Hyun-Sik Yang, Lei Yu, Henrik Zetterberg, and Julie A. Schneider. Clinical criteria for limbic-predominant age-related TDP-43 encephalopathy. *Alzheimer's & Dementia*, page e14202, January 2025. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.14202.

Kai Yuan, Ryan J Longchamps, Antonio F Pardiñas, Mingrui Yu, Tzu-Ting Chen, Shu-Chin Lin, Yu Chen, Max Lam, Ruize Liu, Yan Xia, et al. Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nature genetics*, 56(9):1841–1850, 2024.

Paul A. Yushkevich, John B. Pluta, Hongzhi Wang, Long Xie, Song-Lin Ding, Eske C. Gertje, Lauren Mancuso, Daria Kliot, Sandhitsu R. Das, and David A. Wolk. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment: Automatic Morphometry of MTL Subfields in MCI. *Human Brain Mapping*, 36(1):258–287, January 2015. ISSN 10659471. doi: 10.1002/hbm.22627.

W. Zhou, T. Hinoue, B. Barnes, O. Mitchell, W. Iqbal, S. Lee, K. Foy, K. Lee, E. Moyer,

A. VanderArk, J. Koeman, W. Ding, M. Kalkat, N. Spix, B. Eagleson, J. Pospisilik, P. Szabó, M. Bartolomei, N. Schaaf, L. Kang, A. Wiseman, P. Jones, C. Krawczyk, M. Adams, R. Porecha, B. Chen, H. Shen, and P. Laird. Dna methylation dynamics and dysregulation delineated by high-throughput profiling in the mouse. *Cell genomics, 2(7)*, 2(7), 2022. doi: 10.1016/j.xgen.2022.100144,.

Wanding Zhou, Timothy J Triche, Peter W Laird, and Hui Shen. SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Research*, July 2018. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gky691.