© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Epigenetic Alterations Associated with Alzheimer' Disease Progression

ABOLAJI TAWAKALITU DURODOYE
Biology (Molecular Biology) Faculty of Biological Sciences, Imo State University

Abstract- ***Alzheimer's disease is a progressive neurodegenerative disorder characterised by cognitive decline, memory impairment, loss of functional independence, and extensive molecular and cellular abnormalities within the brain. Although genetic factors contribute substantially to disease susceptibility, inherited variation alone does not fully explain the heterogeneous onset, severity, progression, or clinical presentation of Alzheimer's disease. Increasing evidence indicates that epigenetic mechanisms may mediate interactions among ageing, genetic susceptibility, environmental exposure, cellular stress, inflammation, and neurodegenerative processes.***

Aim -This study aims to investigate epigenetic alterations associated with Alzheimer's disease progression and to determine their potential relevance to disease mechanisms, biomarker development, prognosis, and therapeutic targeting.

Methods - For methodological demonstration, a deterministic synthetic cohort of 480 observations was generated (160 cognitively normal, 160 mild cognitive impairment and 160 Alzheimer's disease). Variables represented demographic, cognitive, APOE, methylation, non-coding RNA, chromatin-accessibility, transcriptomic and epigenetic-age measures. Group comparisons, linear models, correlations and a 10-fold cross-validated logistic model were applied. The values are simulated and cannot establish clinical or biological effects.

In the simulated cohort, mean age, APOE ε4 frequency, Braak stage and cognitive impairment increased across diagnostic groups by construction. Synthetic methylation values at ANK1, BIN1, RHBDF2, HOXA3 and MCF2L increased with diagnostic stage. A multivariable model combining demographic and simulated molecular features yielded a 10-fold cross-validated AUC of 0.917. These estimates demonstrate the proposed analysis and reporting workflow only; they are not empirical Alzheimer's disease findings.

Conclusion

The synthetic analysis shows that the proposed workflow can generate internally consistent tables, figures and model-performance estimates. It does not validate an epigenetic biomarker or support causal, diagnostic, prognostic or therapeutic claims. A submission-ready thesis requires verified participant or public-cohort data, documented provenance, ethics/data-use approval, reproducible analysis code and external validation.

Keywords: Alzheimer's Disease, Epigenetics, DNA Methylation, Histone Modification, Chromatin Accessibility, Non-Coding RNA, Neurodegeneration, Cognitive Decline.

I. INTRODUCTION

Alzheimer's disease (AD) is a biologically complex and clinically heterogeneous neurodegenerative disorder that develops over an extended preclinical period and culminates in progressive impairment of memory, executive function, language, orientation, behaviour and independent daily functioning. Contemporary diagnostic frameworks increasingly define AD by its underlying biological processes rather than by dementia symptoms alone. In this view, abnormal amyloid-beta deposition and pathological tau are central disease-defining features, while neurodegeneration, synaptic dysfunction, neuroinflammation, vascular injury and co-pathologies influence the timing and expression of clinical decline (Jack et al., 2024; Knopman et al., 2021). This shift from a symptom-centred model to a biological continuum has intensified the search for molecular mechanisms that explain why disease begins, why its course differs among individuals and why some people retain cognition despite substantial neuropathological burden.

Genetic studies have identified both highly penetrant variants associated with autosomal-dominant AD and numerous common susceptibility loci associated with late-onset disease. Nevertheless, DNA sequence variation alone does not account for the full variability in age at onset, regional vulnerability, rate of progression or cognitive resilience. The major genetic findings implicate amyloid processing, lipid metabolism, innate immunity, endosomal trafficking

and microglial biology, but they also reveal that genetic risk is expressed through regulatory networks that are highly dependent on cell type, age and environmental context (Bellenguez et al., 2022). Epigenetic mechanisms provide a plausible regulatory interface through which inherited susceptibility, ageing and life-course exposures may alter gene activity without changing the nucleotide sequence.

Epigenetic regulation includes DNA methylation and hydroxymethylation, post-translational histone modifications, chromatin remodelling and accessibility, three-dimensional genome organisation and regulatory non-coding RNAs. These processes are indispensable for neuronal identity, activity-dependent transcription, synaptic plasticity, learning and memory. They are also dynamic, tissue-specific and responsive to inflammation, metabolic stress, toxic exposures and ageing. When dysregulated, they may contribute to persistent alterations in immune signalling, protein homeostasis, mitochondrial function, neuronal plasticity and stress responses that are relevant to AD (De Plano et al., 2024; Maity et al., 2021; Villa & Combi, 2024). However, an observed epigenetic difference may be causal, permissive, compensatory or secondary to neuronal loss and gliosis; therefore, careful study design and multi-layer validation are required.

This chapter establishes the context and justification for an integrative investigation of epigenetic alterations associated with AD progression. It presents the background to the study, defines the research problem and gap, states the aim, objectives, research questions and hypotheses, and specifies the significance, scope, delimitations and anticipated limitations of the work. The study is framed as a human multi-cohort, multi-omics investigation in which DNA methylation is the primary epigenetic layer and is interpreted alongside chromatin accessibility, gene expression, genetic risk, neuropathological measures and longitudinal cognitive information where these data are available.

Background to the Study

Global burden and clinical continuum of Alzheimer's disease

Dementia is one of the most consequential health and social-care challenges associated with population ageing. The World Health Organization reported that approximately 57 million people were living with dementia in 2021, more than 60% of them in low- and middle-income countries, with nearly 10 million new cases occurring annually. Dementia is a major cause of disability and dependency and was estimated to cost the global economy US$1.3 trillion in 2019, approximately half of which was attributable to informal care. AD is the most common cause and is estimated to account for 60-70% of dementia cases (World Health Organization, 2026). Demographic modelling projects a rise from about 57.4 million people with dementia in 2019 to approximately 152.8 million by 2050 if age-specific prevalence remains otherwise stable, with particularly rapid increases expected in regions undergoing demographic transition (GBD 2019 Dementia Forecasting Collaborators, 2022).

The burden extends beyond prevalence. AD is associated with prolonged disability, behavioural and psychological symptoms, loss of autonomy, institutional care, premature mortality and extensive unpaid caregiving. Women are disproportionately affected as patients and caregivers, while racial, ethnic, socioeconomic and geographic inequalities shape exposure to modifiable risks, access to diagnosis and availability of specialised services. In the United States, the Alzheimer's Association estimated that 6.9 million adults aged 65 years and older were living with Alzheimer's dementia in 2024, illustrating the scale of the burden within a high-income health system; the global challenge is more difficult in settings with limited specialist, biomarker and long-term-care capacity (Alzheimer's Association, 2024). AD develops along a continuum. Molecular pathology may accumulate for years before conventional clinical thresholds are crossed. Individuals may progress from an asymptomatic biomarker-positive phase to subjective cognitive decline, mild cognitive impairment and dementia of increasing severity. The sequence is not uniform: some individuals remain clinically stable for long periods, some progress rapidly and others exhibit

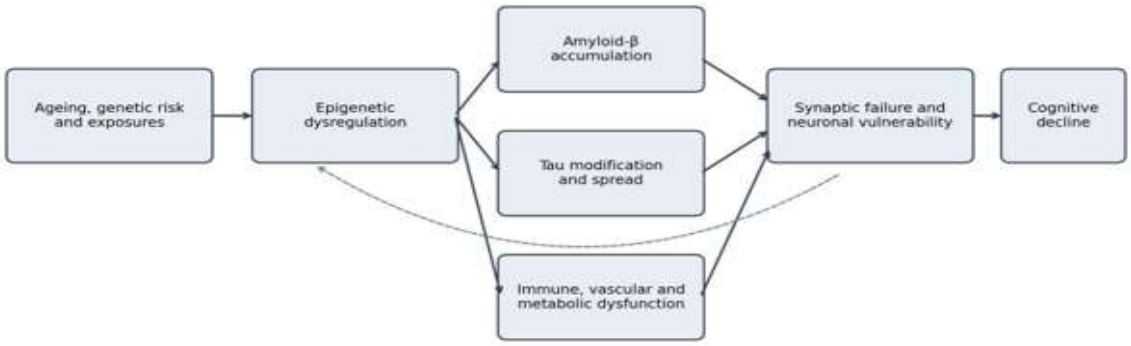
mixed pathologies. Revised criteria therefore distinguish biological disease, defined through validated biomarkers, from the clinical stage that describes the effect of pathology on cognitive and functional performance (Jack et al., 2024). This distinction is especially important for epigenetic research because an epigenetic signature associated with pathology may not have the same relationship with symptoms, resilience or rate of change.

Molecular pathobiology and determinants of progression

The classical pathological features of AD are extracellular amyloid-beta plaques and intracellular neurofibrillary tangles composed of abnormally modified tau. Amyloidogenic processing of amyloid precursor protein generates peptides that can aggregate and disrupt synaptic and cellular homeostasis. Tau normally stabilises neuronal microtubules, but pathological phosphorylation, conformational change and aggregation impair axonal transport and contribute to the spread of neurofibrillary pathology. Although amyloid and tau are central to biological definitions of AD, disease progression reflects an interacting network rather than a single linear cascade. Endosomal-lysosomal dysfunction, impaired proteostasis, mitochondrial disturbance, oxidative stress, neurovascular dysfunction and altered lipid handling can intensify neuronal vulnerability (Knopman et al., 2021).

Neuroinflammation is a major component of this network. Microglia and astrocytes respond to amyloid, tau, cellular debris and vascular injury through state changes that may initially limit damage but become maladaptive when activation is prolonged. Genetic loci associated with AD, including TREM2, CD33, CR1, ABI3 and PLCG2, reinforce the importance of innate immunity and glial regulation. The consequences of inflammatory signalling depend on disease stage, cellular context and interaction with lipid metabolism and clearance pathways. Epigenetic programmes are central to these cell-state transitions because enhancer accessibility, histone modifications and DNA methylation determine which inflammatory and homeostatic genes can be activated in specific cell populations.

Synaptic failure is the biological process most closely related to cognitive impairment. Long before widespread neuronal death, dysregulated calcium signalling, excitatory-inhibitory imbalance, impaired axonal transport and loss of dendritic spines reduce network efficiency. Epigenetic mechanisms regulate immediate-early genes, neurotrophic factors and memory-associated transcription. Age-related loss of chromatin precision may therefore reduce the ability of neurons to maintain identity and respond adaptively to activity or stress. Conversely, preserved epigenomic stability may be one component of cognitive reserve and resilience. The central question is not merely whether epigenetic differences occur in AD, but whether they map onto the transition from pathology to progressive functional failure.



Feedback loops among inflammation, oxidative stress, chromatin change and neuronal injury can amplify progression.

Figure 1.1: Integrated pathological progression of Alzheimer's disease and the potential position of epigenetic dysregulation.

Source: Author-developed conceptual synthesis based on Jack et al. (2024), Knopman et al. (2021), De Plano et al. (2024) and Nativio et al. (2020).

Genetic architecture and the need for regulatory explanations

Rare pathogenic variants in APP, PSEN1 and PSEN2 establish a direct genetic route to early-onset autosomal-dominant AD, while the APOE epsilon4 allele is the most important common genetic risk factor for late-onset disease. Large genome-wide association studies have expanded the number of susceptibility loci and implicated biological processes that include amyloid and tau biology, microglial activation, lipid metabolism, endocytosis and immune regulation. Bellenguez et al. (2022), in a large two-stage genetic study, reported 75 risk loci, including 42 that were new at the time of analysis, and demonstrated that genetic risk scores were associated with future dementia and progression from mild cognitive impairment.

These discoveries are fundamental but do not eliminate the need for regulatory research. Most common risk variants are located outside protein-coding regions, where they may alter enhancer activity, transcription-factor binding or chromatin interactions in particular cell types. The functional effect of a variant can therefore depend on the epigenetic state of the locus. Single-cell epigenomic mapping has shown that inherited risk loci can be assigned to regulatory elements and candidate target genes in microglia, oligodendrocytes, neurons and other brain cells, demonstrating that disease-associated variation acts through cell-specific regulatory architecture (Morabito et al., 2021).

Genetic risk also interacts with non-genetic factors. Age is the strongest risk factor, yet ageing affects DNA methylation, histone abundance, chromatin accessibility and transcriptional noise. Vascular and metabolic disease, physical inactivity, smoking, alcohol exposure, social isolation, sensory impairment, traumatic brain injury and air pollution contribute to dementia risk across the life course (Livingston et al., 2024). Epigenetic mechanisms may encode some of these cumulative exposures, although the persistence, tissue specificity and causal relevance of such encoding are not fully established. This gene-environment-regulation framework supports a study design that integrates genetic covariates and environmental or clinical context rather than treating epigenetic marks as isolated molecular features.

Epigenetic mechanisms relevant to Alzheimer's disease

DNA methylation most commonly refers to the addition of a methyl group to cytosine, particularly at cytosine-phosphate-guanine dinucleotides. Its regulatory effect is context-dependent: promoter methylation may reduce transcription in some settings, whereas gene-body and enhancer methylation can have more complex relationships with expression. The brain also contains substantial 5-hydroxymethylcytosine, an oxidised derivative of 5-methylcytosine involved in neuronal gene regulation. Array-based and sequencing-based studies have identified AD-associated methylation changes at loci including ANK1, BIN1, RHBDF2, HOXA genes and immune-related regions, but the direction and magnitude of effects vary by brain region, platform, disease measure and cellular composition (Alves et al., 2024; De Jager et al., 2014; Lunnon et al., 2014).

Histone proteins are modified by acetylation, methylation, phosphorylation, ubiquitination and other chemical marks that influence nucleosome stability and regulatory access. Histone acetylation is generally associated with a more permissive chromatin state, although the biological consequence depends on the residue, genomic region and interacting proteins. Altered activity of histone acetyltransferases, deacetylases and sirtuins has been reported in models of ageing and AD. Human multi-omics studies have also identified disease-associated changes in histone acetylation that overlap genes involved in transcription, chromatin regulation and AD genetic risk (Nativio et al., 2020). These findings suggest that chromatin regulation may affect broad cellular programmes rather than only individual AD genes.

Chromatin accessibility defines whether regulatory DNA is physically available to transcription factors. Single-nucleus ATAC sequencing can measure this accessibility within individual cell types, reducing the interpretive problems created by mixed brain tissue. Morabito et al. (2021) combined single-nucleus chromatin accessibility and transcriptomic data from nearly 192,000 nuclei and demonstrated extensive cell-type heterogeneity and altered regulatory relationships in late-stage AD. Such studies are important because the same locus may be accessible in microglia but closed in neurons, or may change in opposite directions across disease stages.

Non-coding RNAs, including microRNAs, long non-coding RNAs and circular RNAs, influence transcript stability, translation, chromatin recruitment and gene-network organisation. Dysregulated microRNAs have been linked to amyloid processing, tau phosphorylation, neuroinflammation and synaptic function. However, reported peripheral RNA signatures have often shown limited replication because of differences in sample processing, medication, comorbidity, analytical platforms and disease definitions. The thesis therefore treats non-coding RNA as a complementary regulatory layer rather than assuming that any single RNA species is a validated clinical biomarker (Maity et al., 2021; Villa & Combi, 2024).

Human evidence linking epigenetic alterations to disease progression

Two landmark brain methylation studies reported convergent evidence of differential methylation associated with AD neuropathology. De Jager et al. (2014) analysed prospectively characterised autopsied brains and identified early methylation changes at ANK1, BIN1, RHBDF2 and other loci in relation to neuritic plaque burden. Lunnon et al. (2014) identified an ANK1 differentially methylated region across independent post-mortem cohorts, particularly in the entorhinal cortex, a region affected early in the disease. These findings established that reproducible epigenomic differences could be detected in human AD brain tissue and provided loci for subsequent replication and functional investigation.

Later meta-analyses increased statistical power and demonstrated that a proportion of cortical methylation associations replicate across cohorts and regions. Smith et al. (2021) combined six brain EWAS datasets comprising 1,453 donors and identified differentially methylated loci associated with Braak stage across cortical regions. Such meta-analysis is critical because individual EWAS studies are vulnerable to small effects, multiple testing, region-specific biology and variable post-mortem conditions. The association with Braak stage is particularly relevant to progression because Braak staging represents the anatomical spread of tau pathology, although it remains a post-mortem and partly cross-sectional measure.

Cell composition is a central methodological issue. Neuronal loss, gliosis and shifts in oligodendrocyte or immune-cell states can create apparent methylation differences in bulk tissue. Shireby et al. (2022) profiled DNA methylation in two cortical regions from 631 donors and showed that many bulk-tissue signatures of AD neuropathology were driven by variation in non-neuronal cell types. This does not make the signatures biologically unimportant; rather, it changes their interpretation from a uniform change across the cortex to a signal arising from altered glial abundance or state. Cell-type deconvolution, sorted nuclei and single-cell methods are therefore essential to distinguish cellular redistribution from within-cell epigenetic regulation.

Multi-omics studies strengthen interpretation by connecting epigenetic marks to transcriptional and regulatory consequences. Nativio et al. (2020) integrated histone modifications, transcriptomic data and other molecular measurements in AD brain and found disease-associated histone acetylation changes linked to gene regulation and genetic risk. Morabito et al. (2021) connected cell-specific chromatin accessibility with transcriptional states. Silva et al. (2022) integrated blood and brain methylation studies, prioritising cross-tissue CpGs and regions associated with both diagnosis and neuropathology, but the resulting blood-based model achieved only moderate discrimination. Collectively, the evidence supports biological relevance while illustrating that

clinical translation requires robust cross-tissue replication and prospective validation.

Epigenetic ageing provides another progression-related perspective. Levine et al. (2015) reported that accelerated epigenetic age in dorsolateral prefrontal cortex was associated with neuritic plaques, amyloid load and decline in global and domain-specific cognitive performance. This finding is consistent with the concept that biological ageing modifies vulnerability, but epigenetic clocks may capture a mixture of cellular ageing, disease effects and tissue composition. The present study therefore interprets age-acceleration measures alongside pathology, cognition and cellular covariates rather than treating them as direct measures of AD causation.

Translational importance and unresolved challenges
Epigenetic research has translational appeal because epigenetic marks can be measurable, biologically informative and potentially modifiable. A reproducible blood-based methylation or non-coding RNA signature could support risk stratification or monitoring, while brain epigenomic maps could identify regulatory pathways and therapeutic targets. The reversibility of some chromatin states has encouraged interest in histone deacetylase inhibitors, sirtuin modulators, RNA therapeutics and locus-specific epigenome editing. Nevertheless, broad epigenetic drugs can affect many genes and cell types, raising concerns about specificity, blood-brain barrier penetration, toxicity and timing of intervention (De Plano et al., 2024).

Biomarker development faces a tissue problem. Brain tissue is biologically relevant but generally available only after death, whereas blood is accessible but contains different cell populations and regulatory programmes. Cross-tissue signals may reflect systemic inflammation or shared ageing processes rather than neuronal pathology. Moderate predictive performance in existing methylation studies indicates that epigenetic biomarkers are unlikely to replace established amyloid, tau, imaging or clinical measures in the immediate term. Their most plausible role may be complementary: identifying subtypes, representing exposure history, refining prognosis or revealing pathways that are not captured by conventional biomarkers (Silva et al., 2022; Villa & Combi, 2024).

The causal direction remains unresolved. Neurodegeneration can alter cell proportions, metabolism, inflammation and chromatin, meaning that epigenetic changes observed at autopsy may be consequences of disease. Conversely, regulatory changes that precede symptoms could influence susceptibility and progression. Longitudinal peripheral data, genetically informed analyses, mediation models, experimental validation and temporal ordering are required to separate these possibilities. The current thesis responds to this problem through an integrative framework that prioritises replication, stage-related analyses, cell-type adjustment, independent validation and cautious causal language.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

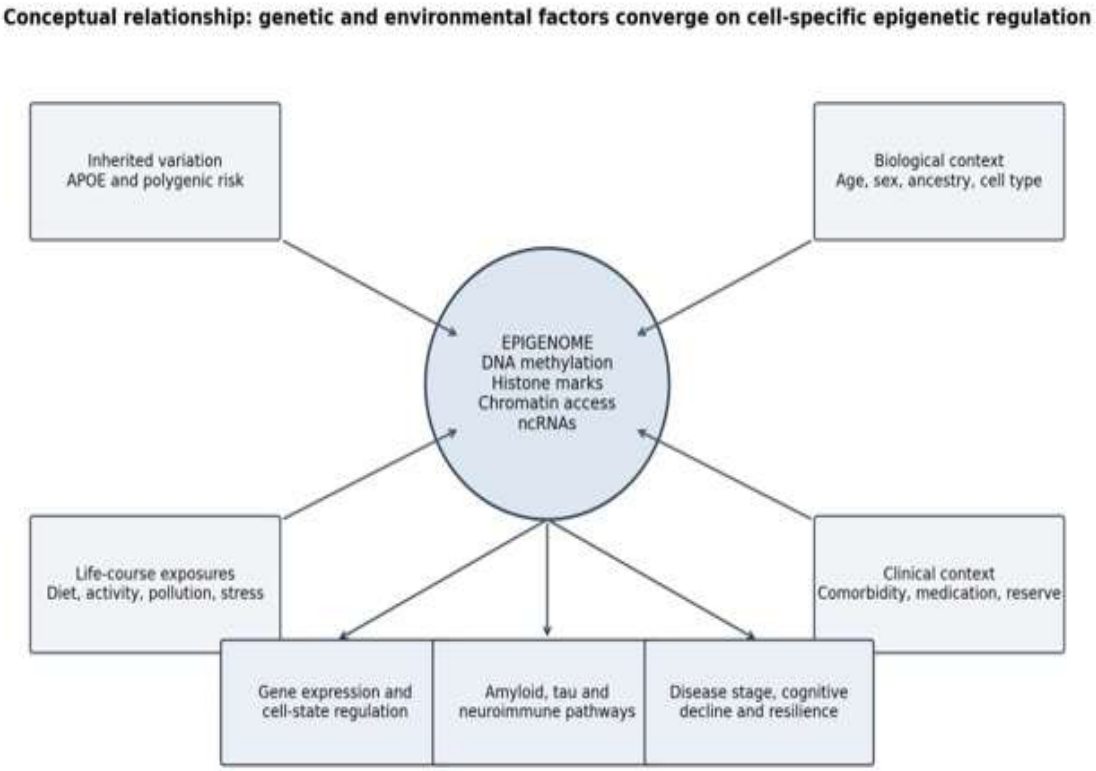


Figure 1.2: Gene-environment-epigenome framework for Alzheimer's disease progression.

Source: Author-developed conceptual framework informed by Bellenguez et al. (2022), Livingston et al. (2024), Morabito et al. (2021) and Villa and Combi (2024).

Statement of the Research Problem

Despite major advances in AD genetics, biomarkers and disease-modifying therapy, the mechanisms that determine individual differences in disease progression remain incompletely understood. Amyloid and tau biomarkers can identify core biological pathology, but they do not fully explain the rate at which cognition declines, the regional pattern of neurodegeneration or the persistence of cognitive function in some individuals with substantial pathological burden. Genetic susceptibility contributes strongly to risk, yet much common genetic ariation acts through non-coding regulatory regions and does not directly specify how ageing, inflammation, metabolism and environmental exposures reshape gene expression in vulnerable brain cells.

Human epigenomic studies have demonstrated reproducible associations between AD neuropathology and DNA methylation, histone modifications and chromatin accessibility. However, the evidence is fragmented. Many studies are based on cross-sectional post-mortem tissue, small or clinically heterogeneous samples, one brain region, one molecular layer or bulk tissue that combines neurons, microglia, astrocytes and oligodendrocytes. Findings may therefore be confounded by cell loss, gliosis, post-mortem interval, tissue quality, batch effects and co-pathologies. Furthermore, many studies compare advanced AD with controls, which identifies end-stage differences but provides limited information about transition from normal cognition to mild cognitive impairment and dementia.

There is also a translational disconnect between brain and peripheral tissues. Brain epigenetic alterations may reveal disease mechanism but are not easily measurable during life. Blood-based signals are clinically accessible but often show modest concordance with brain and limited out-of-sample

performance. The lack of consistent cross-tissue and longitudinal validation prevents current epigenetic markers from being used confidently for early diagnosis, prognosis or therapeutic monitoring. The problem is compounded by under-representation of African and other non-European ancestry populations in genomic and epigenomic cohorts, limiting generalisability and potentially obscuring ancestry-specific regulatory effects.

Accordingly, a systematic integrative study is required to identify epigenetic alterations that are consistently associated with AD stage and progression, determine their cell-type and tissue context, connect them with gene expression and biological pathways, and evaluate whether they contribute information beyond age, sex, APOE genotype and conventional clinical variables. Without such integration, it is difficult to distinguish robust disease-related regulatory changes from technical artefacts, tissue-composition effects or consequences of terminal neurodegeneration.

Research Gap

Empirical gap

The literature contains many case-control comparisons but fewer analyses that explicitly model progression across clinically and biologically meaningful stages. Cross-sectional differences between advanced AD and controls do not establish whether an epigenetic alteration appears early, changes progressively or reflects late-stage cell loss. Studies that relate epigenetic features to Braak stage, amyloid burden, cognitive trajectories or conversion from mild cognitive impairment are available, but the evidence has not been consistently integrated across cohorts and molecular layers. A second empirical gap concerns replication. Individual studies have reported numerous loci and regulatory RNAs, yet only a subset replicates across brain regions and cohorts. Recurrent signals such as ANK1 support the existence of robust alterations, but substantial heterogeneity remains. A doctoral investigation should therefore prioritise effect consistency, independent validation and pathway-level convergence rather than selecting isolated statistically significant loci from a single dataset.

Methodological gap

Bulk brain tissue remains common in AD epigenetics, but bulk signals combine within-cell regulation with changes in cell abundance. Although statistical deconvolution and sorted-cell studies have improved interpretation, many published associations do not have direct cell-type validation. Single-cell and single-nucleus technologies provide higher resolution but are often limited by sample size, sparse measurements, advanced disease tissue and complex analytical choices. There is a need for an analytical framework that uses bulk cohorts for power and cell-resolved data for interpretation.

Many studies analyse one omics layer in isolation. DNA methylation without transcriptomic or chromatin information cannot establish whether a locus is functionally connected with gene regulation. Conversely, transcriptomic changes alone cannot identify the regulatory mechanisms that produced them. Integrative multi-omics analysis can connect methylation, accessibility and expression, but requires careful normalisation, batch control and avoidance of overfitting. This methodological gap motivates the proposed integration of epigenomic, transcriptomic and genetic information.

Causal and temporal gap

Most available human brain studies are observational and post-mortem, making reverse causation a major concern. Epigenetic changes may precede pathology, mediate genetic or environmental risk, emerge as compensatory responses, or result from neurodegeneration. Few studies combine temporal information, genetic instruments, mediation analyses and independent experimental evidence. Although this thesis cannot prove causality from observational data alone, it can strengthen causal interpretation by examining stage relationships, genetic regulation, cross-cohort consistency and temporal cognitive data where available.

The distinction between disease initiation and progression is also insufficiently developed. A feature associated with diagnosis may not predict the rate of decline among affected individuals. The proposed study addresses progression as a distinct outcome, incorporating neuropathological stage,

severity measures, longitudinal cognitive slopes and conversion where data permit.

Translational and population gap

Peripheral epigenetic biomarkers remain insufficiently validated. Blood-based findings may be influenced by immune-cell composition and systemic disease, and cross-tissue models have shown only moderate predictive performance. A clinically useful signature must be reproducible, incrementally informative, technically stable and validated in independent samples. The proposed work therefore treats prediction as a secondary objective and emphasises external validation and transparent performance reporting.

Population diversity is another major gap. Many high-dimensional AD cohorts are predominantly of European ancestry and originate in high-income countries. This limits the ability to infer how ancestry, socioeconomic context, exposure patterns and health-system differences affect epigenetic associations. The thesis will document cohort composition explicitly, avoid unqualified global generalisation and identify diversity limitations as part of interpretation and recommendations.

In summary, the central gap is the absence of a sufficiently integrated, replicated and cell-aware account of how epigenetic alterations relate to AD progression rather than merely to end-stage disease. This thesis is designed to address that gap by combining stage-sensitive outcomes, multi-omics integration, covariate control, cell-type interpretation and independent validation.

Aim of the Study

The aim of this study is to identify, characterise and validate epigenetic alterations associated with Alzheimer's disease progression and to determine their relationships with cognitive decline, neuropathological burden, cell-type-specific regulatory programmes and disease-relevant molecular pathways.

Specific Objectives

1. To identify differential DNA methylation patterns associated with clinical stage, neuropathological stage and progression-related cognitive outcomes in Alzheimer's disease.
2. To determine whether identified methylation patterns are independent of age, sex, APOE genotype, ancestry, brain region, cellular composition, post-mortem factors and technical batch effects.
3. To integrate DNA methylation with chromatin accessibility, histone-modification and transcriptomic data, where available, in order to identify regulatory relationships and disease-relevant pathways.
4. To characterise the cell-type and tissue specificity of epigenetic alterations using deconvolution, sorted-cell or single-nucleus reference data.
5. To evaluate associations between epigenetic age acceleration and measures of amyloid pathology, tau pathology, cognitive decline and clinical severity.
6. To assess whether a parsimonious epigenetic signature improves classification or progression prediction beyond established demographic, genetic and clinical variables.
7. To validate priority epigenetic loci, regions or networks in at least one independent cohort and evaluate the consistency of effects across brain and peripheral tissues where comparable data exist.

Research Questions

RQ1. Which DNA methylation positions, regions and co-methylation networks are associated with Alzheimer's disease stage and progression-related outcomes?

RQ2. Do the identified epigenetic associations remain significant after adjustment for major biological, clinical and technical covariates?

RQ3. How are the priority epigenetic alterations related to chromatin accessibility, histone modifications, gene expression and molecular pathways implicated in Alzheimer's disease?

RQ4. Which brain cell types and tissues contribute most strongly to the observed epigenetic signatures?

RQ5. Is epigenetic age acceleration associated with greater neuropathological burden and faster cognitive decline? RQ6. Do selected epigenetic features add predictive information beyond age, sex, APOE genotype and standard clinical measures?

RQ7. Are the priority findings reproducible in independent cohorts and consistent across brain and blood where comparable measurements are available?

Research Hypotheses

The following hypotheses are formulated for statistical testing. The precise operationalisation of each hypothesis will depend on the variables available in the final datasets.

H01: There is no significant association between genome-wide DNA methylation patterns and Alzheimer's disease stage or progression-related outcomes.

H11: Genome-wide DNA methylation patterns are significantly associated with Alzheimer's disease stage or progression-related outcomes.

H02: Priority epigenetic associations do not remain significant after adjustment for demographic, genetic, cellular, neuropathological and technical covariates.

H12: Priority epigenetic associations remain significant after adjustment for demographic, genetic, cellular, neuropathological and technical covariates.

H03: Epigenetic alterations are not significantly associated with chromatin accessibility, gene expression or disease-relevant molecular pathways.

H13: Epigenetic alterations are significantly associated with chromatin accessibility, gene expression and disease-relevant molecular pathways.

H04: Epigenetic age acceleration is not associated with amyloid burden, tau pathology, clinical severity or cognitive decline.

H14: Epigenetic age acceleration is associated with greater amyloid burden, tau pathology, clinical severity or cognitive decline.

H05: A model containing epigenetic features does not improve predictive performance beyond a model containing established demographic, genetic and clinical variables.

H15: A model containing epigenetic features improves predictive performance beyond a model containing established demographic, genetic and clinical variables.

H06: Priority epigenetic findings do not replicate in independent cohorts or show consistent direction across comparable tissues.

H16: Priority epigenetic findings replicate in independent cohorts and show consistent direction across comparable tissues.

Alignment of Objectives, Questions and Hypotheses

Objective	Research question	Hypothesis	Alignment
Objective 1	RQ1	H01/H11	Primary analytical focus
Objective 2	RQ2	H02/H12	Primary analytical focus
Objective 3	RQ3	H03/H13	Primary analytical focus
Objective 4	RQ4	H03/H13	Primary analytical focus
Objective 5	RQ5	H04/H14	Primary analytical focus
Objective 6	RQ6	H05/H15	Primary analytical focus
Objective 7	RQ7	H06/H16	Primary analytical focus

Table 1.1: Alignment of the study objectives, research questions and hypotheses.

Significance of the Study

The study is scientifically significant because it addresses the regulatory layer between inherited risk and observable disease. By integrating DNA methylation with other omics and progression measures, the research may clarify how regulatory changes relate to amyloid, tau, immune, synaptic and metabolic pathways. A replicated pathway-level account can advance understanding beyond lists of differentially methylated sites and generate more coherent mechanistic hypotheses.

The study is methodologically significant because it combines statistical power from bulk-tissue cohorts with cell-type interpretation from deconvolution and single-nucleus references. It explicitly addresses confounding by cellular composition, post-mortem conditions, technical batch and disease-associated cell loss. The use of independent validation, multiple-testing correction and transparent predictive evaluation is intended to improve reproducibility and reduce overstatement.

The study may contribute to biomarker research by identifying epigenetic features that relate to progression rather than diagnosis alone. Even if a standalone epigenetic test is not sufficiently accurate for clinical use, a stable signature may complement amyloid, tau, imaging and cognitive measures, improve molecular subtyping or identify individuals with distinct inflammatory or ageing profiles. Negative or modest predictive results will also be informative by defining the current limits of epigenetic biomarkers.

The work has therapeutic significance because epigenetic pathways are potentially modifiable. Priority loci and regulatory networks may identify enzymes, transcription factors or chromatin programmes for functional investigation. However, the study will not equate association with druggability; therapeutic recommendations will consider specificity, cell type, timing, delivery and safety.

The study has public-health relevance because AD burden is increasing most rapidly in many low- and middle-income settings, where access to advanced biomarkers is limited. Understanding how life-course exposures and biological ageing intersect with gene regulation may support prevention-oriented research and more inclusive biomarker development. The thesis will also highlight the need for ancestry-diverse datasets and equitable translation of precision-medicine tools.

Finally, the study will provide a reproducible analytical and conceptual resource for researchers. Documented pipelines, data dictionaries, sensitivity analyses and explicit limitations can support replication and future meta-analysis, thereby contributing to cumulative rather than isolated evidence.

Scope of the Study

Conceptual scope

The study focuses on epigenetic alterations associated with AD progression. Progression is conceptualised broadly to include transition across clinical stages, increasing neuropathological burden, longitudinal cognitive decline and worsening functional severity. The study distinguishes progression from disease risk and diagnosis, although risk-related findings may be examined when necessary to contextualise stage-related effects.

The principal epigenetic domain is DNA methylation, including site-level, region-level, network-level and epigenetic-age measures. Chromatin accessibility, histone modifications and non-coding RNA data will be incorporated where matched or comparable datasets are available. The study also considers genetic variation and gene expression as explanatory or validating layers rather than treating them as primary outcomes.

Population and data scope

The population of interest comprises adults represented in clinically characterised, neuropathologically characterised or population-based AD cohorts. Analytical groups may include cognitively unimpaired individuals, persons with mild cognitive impairment and persons with AD dementia, as well as neuropathological strata based on Braak stage, amyloid burden or related measures. The primary data source will be de-identified existing

human datasets obtained from recognised repositories or research consortia, subject to applicable access and ethical requirements.

Brain regions will be selected according to data availability and biological relevance, with priority given to the entorhinal, temporal and prefrontal cortices and appropriate comparison regions. Peripheral blood data may be analysed for cross-tissue or biomarker objectives. Animal and cell-model studies will inform interpretation but will not be combined with human data as if they were equivalent clinical evidence.

Analytical scope

The analytical scope includes quality control, normalisation, batch correction, cellular-composition estimation, differential methylation analysis, differentially methylated region analysis, co-methylation networks, epigenetic-age estimation, pathway enrichment, integration with gene expression or chromatin accessibility, and validation in an independent cohort. Regression and mixed-effects models will be used when appropriate to the structure of the data. Predictive modelling will be restricted to prespecified, interpretable features and will include internal resampling and external validation where possible.

The study will report effect sizes, confidence intervals and false-discovery-rate-adjusted significance rather than relying solely on unadjusted p-values. It will not claim causal relationships solely from association. Where genetic or temporal information permits stronger inference, the assumptions and limits of those analyses will be stated explicitly.

Delimitations of the Study

The study is deliberately delimited to human epigenomic data relevant to AD and does not attempt to provide an equally detailed analysis of all dementias. Vascular dementia, Lewy body disease, frontotemporal degeneration and limbic-predominant age-related TDP-43 encephalopathy will be considered as co-pathologies or differential contexts when the datasets contain such information, but they are not primary disease targets.

DNA methylation is prioritised because it is the most widely available epigenomic layer across large human cohorts. Histone, chromatin and non-coding RNA analyses will be conducted only where the data are of adequate quality and can be aligned with the primary outcomes. The thesis does not propose to generate a novel wet-laboratory cohort unless such data are subsequently provided and ethically approved.

The study is also delimited to association, integration and prediction. It does not include a clinical trial of an epigenetic therapy, experimental gene editing or direct validation in living brain tissue. Findings will therefore be interpreted as candidates for subsequent functional and clinical research rather than immediate clinical recommendations.

Limitations of the Study

The use of existing datasets creates dependence on variables, platforms, sampling frames and diagnostic definitions chosen by the original investigators. Some cohorts may lack complete information on medication, vascular disease, environmental exposures, education or ancestry. Missingness may reduce sample size or introduce selection bias, and harmonisation across cohorts may require simplified common definitions.

Post-mortem brain data provide target-organ relevance but are cross-sectional and susceptible to post-mortem interval, agonal state, tissue pH, RNA or DNA quality and co-pathology. Advanced disease tissue may over-represent downstream consequences. Longitudinal cognitive data can partially address temporal interpretation, but they do not convert a post-mortem association into proof of causality.

Cellular heterogeneity remains challenging. Statistical deconvolution estimates proportions and depends on reference panels; it cannot fully replace sorted-cell or single-cell measurement. Single-nucleus datasets have lower sample sizes and sparse features, and cell states may be affected by tissue processing. Results that differ between bulk and cell-resolved analyses will require cautious interpretation.

Cross-tissue translation is limited because blood and brain have different cellular composition and

regulatory functions. A methylation change that is detectable in blood may not originate in the brain, while a brain-specific alteration may be undetectable peripherally. Predictive performance may therefore be modest and cohort-specific.

Population representation may be uneven, with many cohorts dominated by European ancestry participants. Adjustment for ancestry or population structure cannot compensate fully for limited diversity. The generalisability of findings to African populations and other under-represented groups will be stated as a limitation unless adequately diverse validation data are available.

Finally, multi-omics integration increases analytical complexity and the risk of overfitting, multiple testing and selective interpretation. Prespecified analysis plans, dimension reduction, independent validation, sensitivity analyses and transparent reporting will be used to limit these risks.

Operational Definition of Terms

Term	Operational definition
Alzheimer's disease	A biological neurodegenerative disease defined by amyloid-beta pathology and a characteristic tau-related disease process, with clinical manifestations ranging from no impairment to severe dementia.
Alzheimer's disease progression	Increasing disease burden over time or across ordered stages, operationalised through clinical stage, cognitive trajectory, functional severity, Braak stage, amyloid or tau burden, or conversion from mild cognitive impairment to dementia.
Term	Operational definition
Epigenetics	Regulatory processes that influence gene activity and chromatin state without changing the underlying DNA sequence, including DNA methylation, histone modifications, chromatin accessibility and regulatory non-coding RNAs.
Epigenetic alteration	A statistically and biologically interpretable difference in an epigenetic feature associated with disease status, stage, pathology or progression after appropriate quality control and covariate adjustment.
DNA methylation	The covalent addition of a methyl group to DNA, usually measured at cytosine-phosphate-guanine sites and expressed as a proportion or transformed methylation value.
Differentially methylated position	A single assayed genomic site whose methylation level differs significantly in relation to an outcome after multiple-testing correction.
Differentially methylated region	A genomic region containing multiple spatially related methylation sites that collectively show evidence of association with an outcome.
DNA hydroxymethylation	The presence of 5-hydroxymethylcytosine, an oxidised form of 5-methylcytosine that is abundant in the brain and has distinct regulatory functions.
Histone modification	A chemical modification of a histone protein that may alter chromatin structure, recruitment of regulatory proteins or gene expression.
Chromatin accessibility	The degree to which genomic DNA is physically accessible to transcription factors and regulatory machinery, commonly assessed using ATAC sequencing.
Non-coding RNA	An RNA molecule not translated into a protein that performs regulatory or structural functions, including microRNAs, long non-coding RNAs and circular RNAs.
Epigenetic age acceleration	The difference or residual between DNA-methylation-predicted biological age and chronological age, interpreted as faster or slower molecular ageing according to the selected clock.

Mild cognitive impairment	Objective cognitive decline that exceeds expectations for age but does not substantially impair independent daily functioning; in this study it is classified according to the original cohort criteria.
Cognitive decline	A negative change in global cognition or a specified cognitive domain over time, estimated from repeated validated assessments.
Braak stage	An ordinal neuropathological measure describing the anatomical distribution of neurofibrillary tau pathology.
Amyloid burden	A quantitative or ordinal measure of cerebral amyloid-beta pathology obtained from neuropathology, positron-emission tomography or fluid biomarkers.
Tau burden	A quantitative or ordinal measure of pathological tau obtained from neuropathology, imaging or fluid biomarkers.
Cell-type deconvolution	A computational procedure used to estimate the proportions or contributions of constituent cell types in bulk-tissue molecular data.
Multi-omics integration	Joint analysis of two or more molecular layers, such as methylation, chromatin accessibility, gene expression, genotype or proteomics, to identify coordinated disease-related mechanisms.
Independent validation	Evaluation of a prespecified finding or model in a cohort or dataset not used to discover or train the original association.
Cognitive resilience	Better cognitive performance than expected for a given level of neuropathological burden, recognising that operational definitions vary by cohort.

Table 1.2: Operational definitions of major concepts used in the study.

II. CRITICAL LITERATURE REVIEW

Introduction

This chapter critically examines the literature on epigenetic alterations associated with Alzheimer's disease progression. It moves from disease definition, epidemiology and neuropathology to genetic architecture and the principal regulatory mechanisms—DNA methylation, histone modification, chromatin accessibility, non-coding RNA and epigenetic ageing. It then evaluates tissue and cell specificity, environmental influences, biomarker and therapeutic applications, next-generation sequencing, single-cell methods and multi-omics integration.

The review is organised thematically rather than as a sequence of individual study summaries. Greater weight is given to peer-reviewed human studies, large cohort analyses, systematic reviews, consensus criteria and recent cell-resolved multi-omic investigations. Seminal older studies are retained where they established staging, causal hypotheses or replicated epigenetic loci. The chapter distinguishes association from causation and evaluates each evidence stream in relation to temporal ordering, sample source, cell composition, technical reproducibility, population diversity and clinical translation.

The central argument is that Alzheimer's disease progression is accompanied by coordinated but heterogeneous regulatory change. Some alterations are locus specific and potentially mechanistic, whereas others reflect biological ageing, glial activation, neuronal loss, compensatory responses or broad erosion of cell identity. A critical synthesis is therefore required to identify what is robustly

established, what remains contested and which gaps justify the present research.

Concept and Classification of Alzheimer's Disease

Alzheimer's disease is both a neuropathological process and a clinical disorder, and the distinction between these meanings has become increasingly important. Historically, the term was applied mainly to a dementia syndrome characterised by progressive episodic-memory impairment followed by deterioration in language, executive function, visuospatial ability and independent daily functioning. Contemporary frameworks instead define Alzheimer's disease biologically, through the presence of characteristic amyloid-β and tau pathology, while describing the clinical manifestations separately. This separation recognises that Alzheimer pathology may be present years before symptoms and that an Alzheimer-like clinical syndrome may arise from non-Alzheimer or mixed pathologies (Jack et al., 2024; Scheltens et al., 2021).

Classification is commonly organised along several intersecting dimensions. Age at onset distinguishes younger-onset disease, conventionally beginning before 65 years, from the more prevalent late-onset form. Aetiologically, autosomal-dominant Alzheimer's disease is associated with pathogenic variants in APP, PSEN1 or PSEN2, whereas most cases are sporadic and reflect complex interactions among polygenic susceptibility, ageing, vascular and metabolic conditions, environmental exposures and social determinants. Clinically, individuals may be cognitively unimpaired, have mild cognitive impairment, or meet criteria for mild, moderate or severe dementia. Biologically, the 2024 Alzheimer's Association criteria distinguish early-changing core biomarkers of amyloid and phosphorylated tau from later-changing markers that capture the spread of aggregated tau and increasing neuropathological burden (Jack et al., 2024).

This multidimensional classification is directly relevant to epigenetic research. Epigenetic alterations detected in a cognitively unimpaired person with biomarker-confirmed disease may represent early susceptibility, compensatory regulation or preclinical pathology, whereas changes observed in end-stage post-mortem cortex may reflect neuronal loss, glial activation, treatment exposure or terminal processes. Accordingly, the phrase 'associated with progression' must be anchored to a defined progression axis: increasing pathological stage, worsening cognition, loss of function, biomarker change, regional spread, or longitudinal conversion from mild cognitive impairment to dementia. Without that precision, apparently conflicting studies may simply be measuring different biological and clinical stages.

Epidemiology and Global Burden

Dementia represents a major and rapidly expanding global health challenge. The World Health Organization estimated that 57 million people were living with dementia in 2021, with nearly 10 million new cases occurring annually; more than 60% of affected people live in low- and middle-income countries. Alzheimer's disease is the most common cause and contributes approximately 60–70% of dementia cases. Dementia is also a leading cause of disability, dependence and mortality in later life, and the economic burden extends well beyond formal healthcare because a substantial share of care is provided by family members and other unpaid carers (World Health Organization, 2025).

The burden is distributed unequally. Age remains the dominant non-modifiable risk factor, but prevalence and outcomes are shaped by cardiovascular health, education, access to diagnosis, sex, ethnicity, socioeconomic position and the availability of long-term care. In the United States, the Alzheimer's Association estimated that 7.4 million people aged 65 years or older were living with clinical Alzheimer's dementia in 2026. Such estimates are influenced by whether disease is defined symptomatically or biologically, because a proportion of clinically diagnosed dementia cases do not show Alzheimer neuropathological change, while some biomarker-positive individuals remain cognitively unimpaired (Alzheimer's Association, 2026; Jack et al., 2024).

Global and population-level patterns have implications for epigenetic research. Most large brain-omics cohorts have been assembled in Europe and North America and contain limited ancestral, geographic and socioeconomic diversity. Because DNA methylation, gene expression and chromatin

states are sensitive to ancestry-linked genetic variation, nutrition, pollution, psychosocial stress, infection, vascular risk and other exposures, findings from homogeneous cohorts may not transport directly to under-represented populations. The epidemiological transition in low- and middle-income countries therefore creates both an urgent health need and a scientific rationale for inclusive epigenomic research that examines how local exposures and genetic backgrounds influence disease trajectories.

Clinical Stages of Alzheimer's Disease

The clinical continuum extends from an asymptomatic or preclinical phase through subjective cognitive concerns, mild cognitive impairment and dementia of increasing severity. The preclinical phase is defined biologically rather than symptomatically: amyloid and tau biomarkers may become abnormal before standard cognitive tests detect impairment. Mild cognitive impairment denotes objective decline that exceeds expectations for age and education while functional independence is largely preserved. Dementia is diagnosed when cognitive impairment interferes with independent daily activities, although the pattern of impairment varies and may include amnestic, language-led, visuospatial or behavioural presentations (Jack et al., 2024; Scheltens et al., 2021).

Clinical progression is neither uniform nor inevitable. Some individuals with mild cognitive impairment remain stable for years, some improve when reversible contributors are addressed, and others progress rapidly. Cognitive reserve, education, occupational complexity, vascular health, comorbid neuropathology and genetic background modify the relationship between pathological burden and clinical expression. The 2025 multiregion single-cell study by Liu and colleagues strengthened the concept that epigenomic stability and cell-state resilience may contribute to cognitive resilience despite pathology, suggesting that progression depends not only on lesion burden but also on the capacity of neural and glial systems to maintain regulatory identity (Liu et al., 2025).

For epigenetic studies, staging should integrate clinical and biological information whenever possible. Cross-sectional comparisons of controls and dementia cases conflate disease onset, progression and survival. More informative designs incorporate continuous cognitive trajectories, Braak neurofibrillary stage, amyloid burden, tau PET, cerebrospinal-fluid or plasma biomarkers, neuroimaging, and repeated assessments. These measures permit testing whether an epigenetic signature is associated with disease presence, pathological severity, rate of decline or resilience.

Table 2.1: Clinical and biological classification of the Alzheimer disease continuum

Dimension	Categories	Interpretive relevance
Clinical status	Cognitively unimpaired; subjective decline; mild cognitive impairment; mild, moderate or severe dementia	Captures functional and cognitive expression but is not specific to Alzheimer pathology.
Biological status	Core amyloid/phosphorylated-tau biomarkers; later-changing aggregated-tau biomarkers; neurodegeneration markers	Supports diagnosis and pathological staging independent of symptoms.
Aetiology	Autosomal-dominant; sporadic/polygenic; mixed pathology	Determines genetic architecture and likely age at onset.
Age at onset	Younger onset (<65 years); late onset (≥65 years)	Younger-onset disease is more likely to have atypical presentation or rare pathogenic variants.
Progression metric	Cognitive slope; functional decline; biomarker trajectory; Braak stage; regional spread	Defines what "progression" means in an epigenetic analysis.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Neuropathological Basis of Alzheimer's Disease

The defining neuropathological features of Alzheimer's disease are extracellular amyloid-β plaques and intracellular neurofibrillary tangles composed of abnormally modified tau. Amyloid accumulation is often detectable early and may plateau, whereas the anatomical spread of tau pathology correlates more closely with neuronal dysfunction and cognitive impairment. Braak staging describes the stereotyped progression of tau pathology from transentorhinal and entorhinal regions to limbic and neocortical areas, although individual variation and co-pathologies are common (Braak & Braak, 1991; Jack et al., 2024).

The amyloid cascade hypothesis provided an influential framework in which altered production, aggregation or clearance of amyloid-β initiates downstream tau pathology, inflammation, synaptic failure and neuronal death. Genetic evidence from APP and presenilin mutations supports a causal role for amyloid in autosomal-dominant disease, but the limited relationship between plaque load and symptom severity, together with the multifactorial biology of late-onset disease, argues against a simple linear cascade. Contemporary models therefore emphasise interacting cellular phases involving microglia, astrocytes, lipid metabolism, vascular dysfunction, proteostasis, mitochondrial stress and loss of network homeostasis (De Strooper & Karran, 2016; Long & Holtzman, 2019).

Epigenetic regulation can influence each component of this network. DNA methylation and histone modifications can alter expression of APP-processing enzymes, tau kinases and phosphatases, inflammatory mediators, lipid transport genes and synaptic proteins. Conversely, amyloid, tau, oxidative stress and inflammation can alter chromatin states and non-coding RNA expression. This bidirectionality complicates causal interpretation but also positions the epigenome as a potential molecular record of cumulative disease processes.

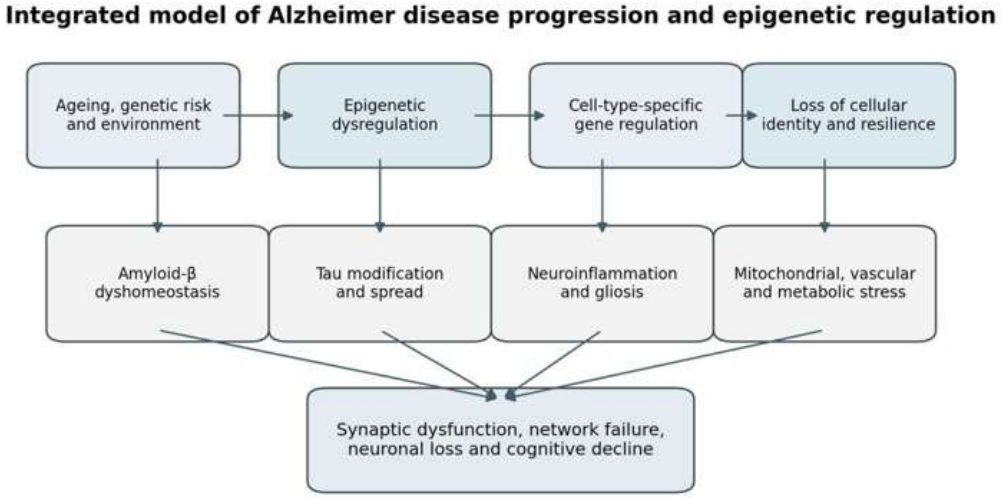


Figure 2.1: Integrated model of Alzheimer disease progression and epigenetic regulation.

Genetic Architecture of Alzheimer's Disease

Alzheimer's disease has a heterogeneous genetic architecture. Rare highly penetrant variants in APP, PSEN1 and PSEN2 cause autosomal-dominant disease, generally with younger onset. For late-onset disease, the APOE ε4 allele is the strongest common genetic risk factor, but risk is distributed across many loci involved in lipid handling, innate immunity, endocytosis, cytoskeletal regulation and amyloid processing. Large genome-wide association studies have expanded the number of implicated loci substantially; Bellenguez and colleagues identified 75 loci, including 42 that were new at the time of analysis, reinforcing the roles of microglial function, lipid biology and amyloid/tau-associated pathways (Bellenguez et al., 2022).

Genetic associations do not automatically identify the causal gene or regulatory mechanism. Many risk variants lie in non-coding regions and may alter enhancers, transcription-factor binding, chromatin accessibility or cell-type-specific gene regulation. Fine-mapping and integrative prioritisation have therefore become central to translating association signals into biology. Studies combining genetic data with epigenomic maps show that Alzheimer risk variants are enriched in microglial enhancers and regulatory elements, providing a mechanistic bridge between inherited susceptibility and disease-relevant immune states (Morabito et al., 2021; Xiong et al., 2023).

Epigenetics also provides a framework for incomplete penetrance and phenotypic heterogeneity. Individuals carrying similar genetic risks may differ in age at onset, pathological severity and cognitive resilience because regulatory states are modified by age, sex, exposures and stochastic processes. However, genetic control of methylation and chromatin accessibility must be separated from disease-induced epigenetic change. Methylation quantitative-trait loci and allele-specific regulatory analyses are therefore needed to determine whether an observed epigenetic association mediates genetic risk, is secondary to pathology, or is independent of inherited variation.

Table 2.2: Selected Alzheimer disease genes and pathways relevant to epigenetic regulation

Gene/locus	Principal function	Epigenetic relevance
APP, PSEN1, PSEN2	Amyloid precursor processing and γ-secretase function	Expression and regulatory-state changes may modify amyloid production; pathogenic variants cause autosomal-dominant disease.
APOE	Lipid transport, amyloid clearance and immune response	Genotype may modify cell-specific chromatin responses and environmental susceptibility.
TREM2, CD33, CR1, ABI3	Microglial sensing and innate immunity	Risk loci are enriched in immune regulatory elements and microglial enhancers.
BIN1, PICALM, SORL1	Endocytosis, trafficking and cytoskeletal regulation	Methylation and non-coding variants may influence expression and enhancer activity.
ABCA7, CLU	Lipid handling and amyloid clearance	Both genetic and methylation evidence connect lipid biology with pathology.
MAPT	Tau expression and isoform biology	Regulated by chromatin and RNA mechanisms; tau pathology also alters chromatin domains.

Fundamentals of Epigenetic Regulation

Epigenetics refers to molecular mechanisms that regulate gene activity and cellular identity without changing the underlying DNA sequence. The major mechanisms include DNA methylation and hydroxymethylation, post-translational modification of histone proteins, ATP-dependent chromatin remodelling, nucleosome positioning, enhancer and promoter activity, three-dimensional genome organisation, non-coding RNAs and chemical modifications of RNA. These systems operate cooperatively rather than independently: DNA methylation can influence transcription-factor binding, histone marks define active or repressed chromatin states, and non-coding RNAs can recruit chromatin-modifying complexes or regulate transcript stability (Martinez-Feduchi et al., 2024; Maity et al., 2021).

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

In the nervous system, epigenetic regulation is unusually dynamic. Neurons must preserve stable identity over decades while also supporting activity-dependent transcription required for learning, memory and synaptic plasticity. Glial cells likewise change state in response to injury, protein aggregates and inflammatory signals. Ageing is associated with epigenetic drift, altered heterochromatin, reduced fidelity of cellular identity and changes in biological-age signatures. These features make the epigenome a plausible interface between long-term exposures and late-life neurodegeneration (Maity et al., 2021; Yang et al., 2023).

The interpretation of epigenetic marks is context dependent. Promoter methylation is often associated with reduced transcription, but gene-body methylation may correlate positively with expression and enhancer methylation may have cell-specific effects. Histone acetylation is generally linked to accessible chromatin, yet the consequences depend on the modified residue, genomic location and interacting proteins. A rigorous Alzheimer study must therefore avoid simplistic assumptions that 'more methylation' invariably means gene silencing or that a global epigenetic direction applies across brain regions and cell types.

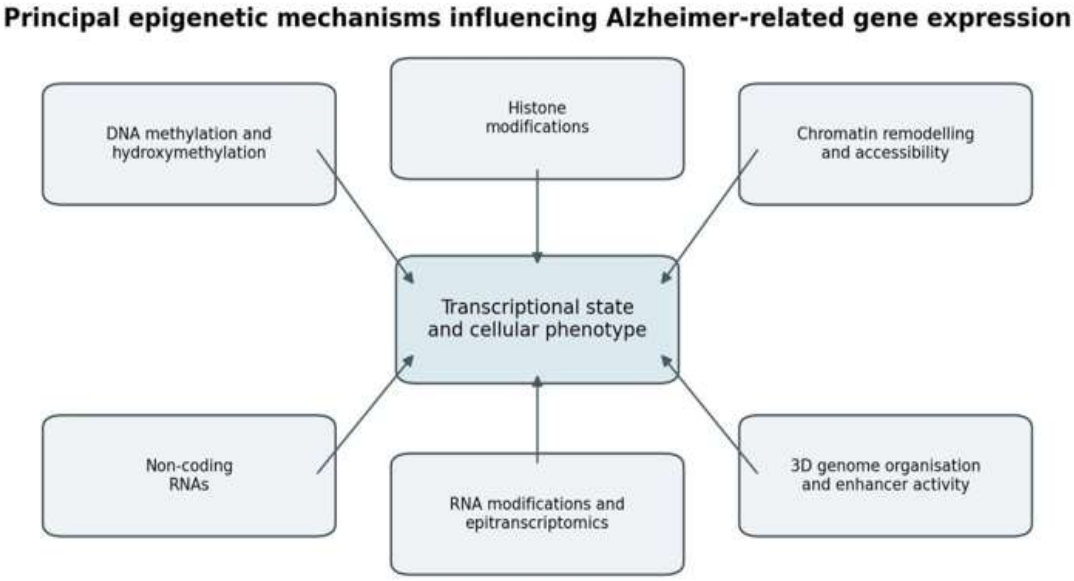


Figure 2.2: Principal epigenetic mechanisms influencing Alzheimer-related gene expression.

Table 2.3: Major epigenetic mechanisms and their relevance to Alzheimer disease

Mechanism	Measurement approaches	Potential relevance
DNA methylation / hydroxymethylation	Arrays, WGBS, RRBS, targeted bisulphite, oxidative or enzymatic methods	Stable regulatory differences, biological ageing, cell-state and exposure signatures.
Histone modifications	ChIP-seq, CUT&RUN, CUT&Tag, mass spectrometry	Activity-dependent transcription, chromatin domains, tau-associated dysregulation.
Chromatin accessibility	ATAC-seq, DNase-seq, single-nucleus ATAC-seq	Enhancer activity, transcription-factor access and cell-type-specific regulatory circuits.
Non-coding RNAs	Small-RNA-seq, total RNA-seq, qPCR, long-read sequencing	Post-transcriptional control, chromatin recruitment, biomarker and therapeutic potential.

Epigenetic clocks	Methylation arrays or sequencing with clock algorithms	Biological ageing and resilience; not specific to Alzheimer pathology.
3D genome and enhancer–promoter links	Hi-C, co-accessibility,chromatin interaction assays	Connects non-coding risk variants with target genes and regulatory domains.

DNA Methylation and Hydroxymethylation

DNA methylation is the most extensively studied epigenetic mechanism in Alzheimer's disease. It usually involves the addition of a methyl group to cytosine at CpG dinucleotides, while 5-hydroxymethylcytosine is generated through oxidation by TET enzymes and is particularly abundant in the brain. Early candidate-gene studies produced inconsistent findings, partly because they examined small samples, selected loci and different tissues. The field changed with large epigenome-wide association studies of post-mortem cortex, which identified reproducible associations between neuropathology and methylation at loci including ANK1, RHBDF2, BIN1, ABCA7 and other regions linked to immune, cytoskeletal and lipid pathways (De Jager et al., 2014; Lunnon et al., 2014).

Replication across cohorts strengthened the evidence that Alzheimer-associated methylation is not random, yet several qualifications remain. Shireby and colleagues meta-analysed cortical data and showed that many bulk-tissue differentially methylated positions were primarily driven by non-neuronal cell populations. This finding does not diminish their biological importance; rather, it indicates that glial activation, oligodendrocyte states and changes in cellular composition contribute substantially to the observed signal (Shireby et al., 2022). Recent reviews similarly conclude that methylation patterns differ by brain region, pathological stage, array platform and analytical strategy, with ANK1 among the most consistently replicated loci (Alves et al., 2024).

Hydroxymethylation adds another layer of complexity. Studies have reported changes in global and locus-specific 5-hydroxymethylcytosine in Alzheimer brain and associations with plaques, tangles and neurobiological pathways, although direction and magnitude vary across studies and regions (Coppieters et al., 2014). Standard bisulphite methods do not distinguish 5-methylcytosine from 5-hydroxymethylcytosine, so some historical 'methylation' findings reflect a composite signal. Newer oxidative bisulphite, enzymatic and long-read approaches can resolve these modifications more accurately.

Table 2.4: Landmark and recent human DNA methylation studies in Alzheimer disease

Study	Design / tissue	Principal contribution	Major limitation
De Jager et al. (2014)	708 autopsied brains with replication	Identified pathology-associated CpGs and validated regions including ANK1, BIN1 and RHBDF2.	Cross-sectional post-mortem design.
Lunnon et al. (2014)	Multi-cohort cortical methylomic profiling	Replicated cortical ANK1 dysregulation and demonstrated regional specificity.	Array coverage and late-stage tissue.
Shireby et al. (2022)	Meta-analysis plus purified nuclei	Showed many bulk cortical signals are driven by non-neuronal cell types.	Limited cell classes and post-mortem inference.

Alves et al. (2024)	Review of array-based human brain studies	Synthesised regional findings and recurrent loci including ANK1 and BIN1.	Dependent on heterogeneous source studies.
Li et al. (2024)	Exposure–methylation–neuropathology analysis	Linked traffic-related PM2.5, brain methylation and Alzheimer pathology markers.	Observational mediation cannot prove causality.

Histone Modifications

Histone proteins package DNA into nucleosomes and are modified by acetylation, methylation, phosphorylation, ubiquitination and other chemical marks. These modifications affect chromatin compaction, transcription-factor access and recruitment of regulatory proteins. Histone acetylation is particularly relevant to learning and memory because activity-dependent acetylation supports transcriptional programmes required for synaptic plasticity. Ageing and Alzheimer pathology can disrupt the balance among histone acetyltransferases, histone deacetylases and sirtuins, potentially reducing adaptive transcription or promoting inflammatory and stress-response programmes (Maity et al., 2021; De Plano et al., 2024).

Human evidence has moved beyond small candidate studies. Klein and colleagues profiled H3K9 acetylation in 669 aged prefrontal cortices and found large-scale alterations strongly associated with tau pathology. The changes occurred across broad genomic domains, correlated with transcriptional differences and were reproduced in model systems, suggesting that tau-related nuclear and chromatin disruption may reorganise the epigenome at a systems level (Klein et al., 2019). This study was important because it connected a specific pathological burden with an epigenome-wide histone mark in a large human cohort.

Nevertheless, histone findings remain less mature than DNA methylation evidence. Post-mortem interval, chromatin preservation, antibody specificity, sequencing depth and cell composition can influence ChIP-based measurements. Moreover, the therapeutic appeal of histone deacetylase inhibition must be balanced against poor target specificity, widespread transcriptional effects and the possibility that the same enzyme has different functions across neurons, microglia and peripheral tissues.

Chromatin Remodelling and Accessibility

Chromatin accessibility reflects the extent to which DNA regulatory elements are available to transcription factors and other proteins. ATAC-seq and related methods have revealed that Alzheimer-associated regulatory change is highly cell-type specific. Morabito and colleagues jointly profiled chromatin accessibility and gene expression in nearly 192,000 nuclei and identified disease-associated cis-regulatory elements, transcription factors and glial trajectories. Their results linked Alzheimer risk loci to regulatory programmes in specific populations, including oligodendrocyte and glial modules involving APOE and CLU (Morabito et al., 2021).

A larger study by Xiong and colleagues profiled approximately 850,000 nuclei and integrated ATAC-seq, transcriptomic and genetic data. Alzheimer risk loci were enriched in microglial enhancers and in regulatory circuits involving transcription factors such as SPI1, ELF2 and RUNX1. The investigators also described differential accessibility in early and late disease and a broad pattern of epigenome erosion, characterised by dysregulated chromatin and loss of cell identity in advanced pathology (Xiong et al., 2023).

The concept of epigenome erosion reframes chromatin change as both locus-specific and global. Rather than only activating or repressing selected genes, disease progression may reduce the information content that maintains stable cellular identity. The 2025 multiregion atlas extended this observation across six brain regions and linked epigenomic stability with cognitive resilience, glial transitions and pathology (Liu et al., 2025). These findings support chromatin accessibility as a progression-sensitive layer, although longitudinal

inference from post-mortem cross-sectional tissue remains indirect.

Table 2.5: Histone and chromatin evidence relevant to progression

Evidence	Key finding	Interpretation
Klein et al. (2019), H3K9ac	Tau burden associated with widespread histone-acetylation changes across large genomic domains.	Tau-related pathology may reorganise chromatin at systems level.
Morabito et al. (2021), snATAC + transcriptomics	Identified cell-specific regulatory elements, transcription factors and glial trajectories.	Regulatory effects of disease and risk loci are cell dependent.
Xiong et al. (2023), 850,000 nuclei	Mapped risk variants to microglial enhancers and described late-stage epigenome erosion.	Progression may involve loss of regulatory identity, not only locus-specific change.
Liu et al. (2025), multiregion single-cell multi-omics	Linked epigenomic stability with pathology, cognition and resilience across six regions.	Supports region- and cell-specific progression models; remains post-mortem.

Non-Coding RNAs

Non-coding RNAs regulate gene expression at transcriptional, post-transcriptional and chromatin levels. MicroRNAs generally bind target messenger RNAs and influence degradation or translation, while long non-coding RNAs can act as scaffolds, guides, decoys or regulators of chromatin. Circular RNAs may sequester microRNAs or interact with proteins, and piwi-interacting RNAs are emerging as regulators of genomic stability and transposable elements. In Alzheimer's disease, these molecules have been linked to amyloid processing, tau phosphorylation, autophagy, oxidative stress, apoptosis and neuroinflammation (Canoy et al., 2024; Martinez-Feduchi et al., 2024).

Frequently discussed examples include BACE1-AS, which can stabilise BACE1 transcripts and potentially influence amyloidogenic processing, and microRNAs that target APP, BACE1, tau-related kinases or inflammatory pathways. However, the literature is heterogeneous. Different studies use brain, cerebrospinal fluid, serum, plasma, whole blood or extracellular vesicles; extraction and normalisation methods vary; and many findings are based on small discovery cohorts without independent validation. Systematic reviews of blood and cerebrospinal-fluid microRNAs identify promising candidates but also reveal limited reproducibility across platforms and populations (Fattahi et al., 2024; Pereira et al., 2024).

The strongest future role for non-coding RNAs may be in panels rather than single markers. Because individual microRNAs regulate many targets and are affected by haemolysis, medication and systemic disease, a robust diagnostic or prognostic signature will probably require standardised pre-analytics, tissue-informed biological interpretation and integration with established amyloid and tau biomarkers.

Table 2.6: Classes of non-coding RNA implicated in Alzheimer disease

Class	Representative functions	Evidence and limitations
MicroRNAs	Target APP processing, tau kinases, inflammatory and synaptic transcripts.	Many candidates; modest cross-study consistency and strong pre-analytic sensitivity.

Long non-coding RNAs	Scaffolding, transcriptional regulation and transcript stabilisation; BACE1-AS is frequently studied.	Mechanistic plausibility exceeds current clinical validation.
Circular RNAs	MicroRNA sequestration and protein interactions.	Stable in biofluids but annotation and replication remain developing.
piRNAs and other small RNAs	Transposon control and genomic stability.	Emerging evidence; limited Alzheimer-specific human validation.
Extracellular-vesicle RNAs	Potentially enrich brain-derived signals in blood.	Isolation methods and tissue origin require standardisation.

Epigenetic Ageing and Epigenetic Clocks

Epigenetic clocks estimate biological age from methylation patterns at selected CpG sites. First-generation clocks were trained to predict chronological age, whereas later clocks incorporate clinical biomarkers, mortality or age-related phenotypes. Age acceleration—the difference between estimated biological age and chronological age—has been associated with several chronic diseases and is conceptually relevant to Alzheimer's disease because ageing is the dominant risk factor (Horvath, 2013; Levine et al., 2018; Lu et al., 2019).

Evidence in neurodegeneration is mixed. A systematic review found associations between accelerated methylation age and several neurodegenerative phenotypes but emphasised variability in tissues, clocks and study designs (Yang et al., 2023). Milicic and colleagues evaluated multiple clocks in the AIBL and ADNI cohorts and reported associations between disproportionate biological ageing and hippocampal volume, while associations with cognition and amyloid measures were not uniformly consistent across clocks (Milicic et al., 2022).

Clock performance is tissue dependent. Most widely used clocks were trained in blood or multiple tissues and may not capture brain-specific ageing, while post-mortem cortical clocks may not be suitable for clinical monitoring. Age acceleration may also represent systemic vulnerability rather than Alzheimer-specific pathology. Epigenetic clocks are therefore best interpreted as complementary measures of biological ageing and resilience, not as standalone diagnostic tests.

Environmental and Lifestyle Influences

Environmental and lifestyle factors may influence dementia risk through vascular, metabolic, inflammatory and epigenetic pathways. The 2024 Lancet Commission identified modifiable risk factors across the life course, including limited education, hearing loss, hypertension, smoking, obesity, depression, physical inactivity, diabetes, excessive alcohol use, social isolation, traumatic brain injury, air pollution, untreated visual loss and high low-density-lipoprotein cholesterol. The Commission estimated that a substantial proportion of dementia might be delayed or prevented through population and individual interventions, while stressing that risk reduction is probabilistic rather than deterministic (Livingston et al., 2024).

Epigenetic mechanisms provide a plausible route through which long-term exposures become biologically embedded. Nutrition influences methyl-donor availability; physical activity affects metabolic and neurotrophic signalling; smoking and pollution produce reproducible methylation signatures; psychosocial stress alters endocrine and inflammatory regulation; and vascular disease modifies brain perfusion and immune activation. A 2024 human brain study reported that differential DNA methylation may mediate associations between traffic-related particulate matter and Alzheimer neuropathology markers, illustrating how exposure, epigenetic change and pathology can be investigated within one analytical framework (Li et al., 2024).

Causal interpretation remains difficult because exposures are correlated with socioeconomic and health factors, epigenetic measurements may occur after disease onset, and reverse causation is possible. Longitudinal studies beginning before cognitive

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

impairment, repeated exposure measurements and genetically informed causal methods are needed to establish whether epigenetic change is a mediator, marker or consequence of risk.

Gene–Environment–Epigenome Interactions

Gene–environment interaction models propose that inherited susceptibility modifies the biological response to exposures. The epigenome is a potential mediator because genetic variants can affect local methylation or chromatin accessibility, while environmental conditions influence the same regulatory elements. APOE genotype, for example, may alter lipid transport, immune responses and vulnerability to metabolic stress, producing downstream regulatory effects that differ across cell types and disease stages.

Multi-omic studies of regulatory variants support this model. Bryzgalov and colleagues integrated histone-modification and transcriptomic data to identify non-coding regulatory variants that may shape the Alzheimer epigenetic landscape (Bryzgalov et al., 2023). Xiong and colleagues identified thousands of cell-type-specific accessibility quantitative-trait loci and used peak-to-gene links to connect inherited variants with regulatory circuits (Xiong et al., 2023). Such analyses move beyond statistical interaction terms by specifying a plausible molecular chain from variant to regulatory element, target gene and cellular pathway.

A major challenge is scale. Detecting interactions requires larger samples than estimating main effects, yet brain epigenomic cohorts are comparatively small. Furthermore, interactions can be sex-, age-, ancestry-, region- and cell-specific. Progress will depend on harmonised consortia, diverse cohorts, replication and the integration of experimental perturbation with population data.

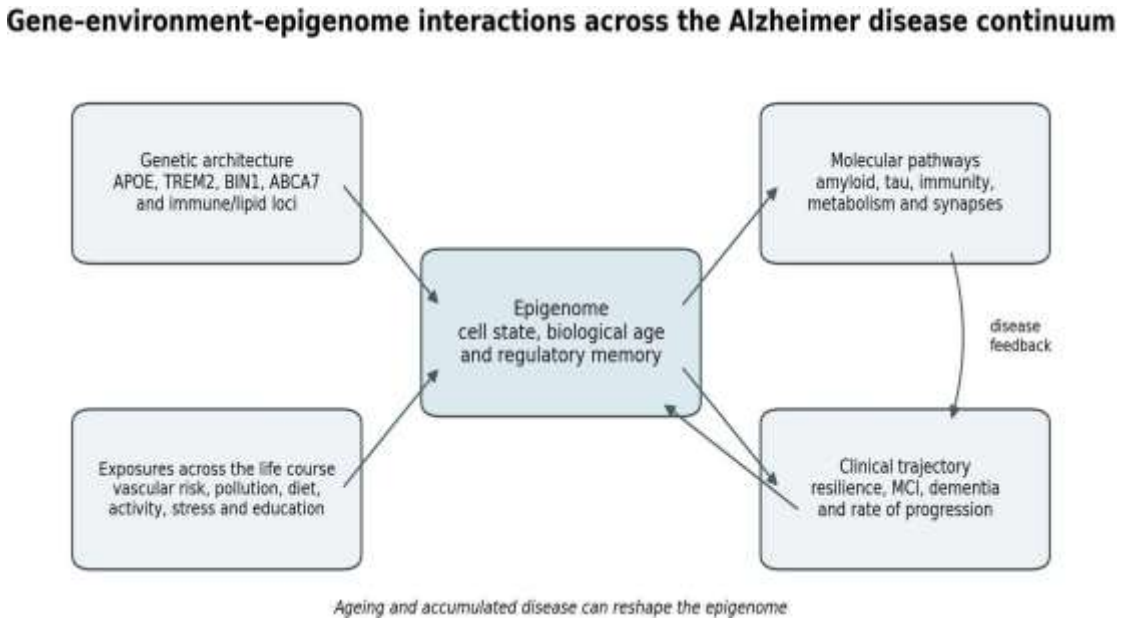


Figure 2.3: Gene–environment–epigenome interactions across the Alzheimer disease continuum.

Epigenetics of Neuroinflammation

Neuroinflammation is a central component of Alzheimer pathophysiology. Microglia respond to amyloid, damaged synapses and dying cells, while astrocytes regulate neurotransmitter homeostasis, metabolism and inflammatory signalling. Acute responses may initially be protective, but chronic activation can impair phagocytosis, amplify cytokine production and damage neural networks. Genetic studies implicating TREM2, CD33, CR1, ABI3 and related loci reinforce the importance of innate immunity (Heneka et al., 2015; Bellenguez et al., 2022).

Epigenetic regulation determines how glial cells enter, maintain and exit disease-associated states. DNA methylation, enhancer accessibility and transcription-factor occupancy can stabilise inflammatory programmes or reduce homeostatic identity. Single-nucleus studies consistently show enrichment of Alzheimer risk and disease-associated regulatory changes in microglia and other non-neuronal cells. Shireby and colleagues demonstrated that much of the cortical methylation signature observed in bulk tissue reflects non-neuronal variation, while Xiong and colleagues mapped risk loci to microglial enhancers and transcriptional circuits (Shireby et al., 2022; Xiong et al., 2023).

These results caution against treating glial epigenetic change as a mere confounder. Cell-composition differences can distort bulk measurements, but they can also represent biologically meaningful gliosis and immune transition. The analytical objective should therefore be to distinguish compositional change from within-cell regulatory change and then evaluate both in relation to pathology and progression.

Epigenetic Regulation of Amyloid-β Pathology

Amyloid-β is generated through sequential cleavage of amyloid precursor protein by β- and γ-secretases. Epigenetic regulation may alter expression of APP, BACE1, presenilins, neprilysin, insulin-degrading enzyme and other proteins involved in production or clearance. Candidate studies have reported methylation differences at APP-related genes, while non-coding RNAs such as BACE1-AS and several microRNAs may modify amyloidogenic processing (Canoy et al., 2024; Paniri et al., 2024).

The evidence is biologically plausible but not uniformly reproducible. Methylation effects are often modest, tissue-specific and influenced by cellular composition. Furthermore, amyloid deposition may itself induce oxidative stress, inflammation and chromatin change, making directionality uncertain. Large epigenome-wide studies have identified methylation at established risk loci such as ABCA7 and BIN1, but the strongest replicated signals are not confined to classical amyloid genes, indicating that Alzheimer epigenetics extends across immune, lipid and structural pathways (De Jager et al., 2014; Alves et al., 2024).

Therapeutically, epigenetic manipulation of amyloid pathways would require precise targeting. Broad inhibitors of DNA methylation or histone deacetylation could alter thousands of genes and may have opposite effects in different cell types. Targeted epigenome editing and RNA therapeutics offer greater specificity but remain technically and clinically immature.

Epigenetic Regulation of Tau Pathology

Tau is a microtubule-associated protein encoded by MAPT. In Alzheimer's disease, abnormal phosphorylation, conformational change, mislocalisation and aggregation of tau contribute to synaptic dysfunction and neuronal death. Tau pathology spreads through anatomically connected regions and correlates strongly with clinical impairment. Epigenetic mechanisms can regulate MAPT expression, tau kinases and phosphatases, autophagy, proteostasis and inflammatory responses that influence tau toxicity.

Human epigenome-wide evidence suggests that tau is a major driver of chromatin dysregulation. The H3K9ac study by Klein and colleagues found that tau burden, more than amyloid burden, was associated with widespread histone-acetylation changes across large genomic domains (Klein et al., 2019). Xiong and colleagues subsequently reported broad epigenome erosion in late disease, consistent with the possibility that tau-associated nuclear stress and loss of chromatin organisation disrupt cellular identity (Xiong et al., 2023).

The relationship may be bidirectional. Epigenetic changes can promote tau-modifying pathways, while pathological tau can disrupt nuclear transport, lamina-associated domains and chromatin architecture. Longitudinal biomarker studies and mechanistic experiments are needed to determine whether specific epigenetic alterations precede tau spread or mainly reflect downstream injury.

Mitochondrial Epigenetics and Oxidative Stress

Mitochondrial dysfunction and oxidative stress are recurring features of Alzheimer's disease. Impaired energy production, altered calcium handling, excessive reactive oxygen species and defective mitophagy can compromise synaptic function and increase vulnerability to amyloid and tau toxicity. Nuclear epigenetic regulation controls many mitochondrial genes, while metabolites such as acetyl-CoA, S-adenosylmethionine and α-ketoglutarate serve as substrates or cofactors for chromatin-modifying enzymes. Metabolic dysfunction can therefore reshape the epigenome.

Mitochondrial DNA methylation has been reported, but its abundance, measurement and functional significance remain debated. Technical artefacts can arise because mitochondrial DNA has unusual structure and copy number, and conventional bisulphite sequencing may generate mapping errors. More robust evidence concerns nuclear genes governing oxidative phosphorylation, antioxidant defence and mitochondrial quality control, which show altered expression and regulatory states in disease (Martinez-Feduchi et al., 2024; Paniri et al., 2024).

Oxidative stress can also affect TET enzymes, DNA repair and histone-modifying pathways, creating feedback between metabolic injury and regulatory instability. This intersection is relevant to progression because energy failure and oxidative damage increase as synaptic and neuronal systems deteriorate.

Epigenetic Regulation of Synaptic Dysfunction

Memory formation depends on activity-dependent transcription, synaptic remodelling and long-term changes in neuronal connectivity. Histone acetylation, DNA methylation, enhancer activation and non-coding RNAs participate in the transcriptional programmes that support these processes. Ageing and Alzheimer pathology can reduce the flexibility of these programmes, thereby impairing learning and memory before extensive neuronal loss is evident (Maity et al., 2021).

Synaptic genes are frequently represented in Alzheimer transcriptomic and epigenomic analyses. Reduced neuronal accessibility and altered regulatory modules appear early in some single-cell studies, whereas glial and inflammatory changes become more dominant in later disease (Xiong et al., 2023). This temporal pattern suggests that early neuronal regulatory dysfunction may lower resilience, while later inflammation and epigenome erosion reinforce network failure.

The therapeutic implication is that restoring a single synaptic gene may be insufficient if regulatory coordination across many genes has been lost. Interventions that support chromatin stability, neuronal metabolism and activity-dependent transcription may be more effective when applied before advanced neurodegeneration.

Tissue-Specific Epigenetic Alterations

Epigenetic patterns vary substantially across tissues. Brain tissue is closest to the target organ but is usually available only after death, when advanced pathology, agonal state, post-mortem interval and tissue preservation may influence measurements. Peripheral blood is accessible longitudinally but contains diverse immune-cell populations and may reflect systemic ageing or inflammation rather than brain-specific regulation. Cerebrospinal fluid is more proximal to the central nervous system but yields limited cellular material and requires invasive collection.

Cross-tissue concordance is therefore expected to be partial. A disease-relevant methylation change may occur in cortex but not blood because the regulatory element is inactive in leukocytes. Conversely, a peripheral signature may provide a useful biomarker even if it is not mechanistically active in neurons. The distinction between mechanistic biomarker and predictive biomarker is essential: clinical utility depends on reproducibility and discrimination, whereas mechanistic inference requires tissue and cell relevance (Klein et al., 2016; Alves et al., 2024).

Studies comparing multiple brain regions further show that vulnerability is region-specific. The entorhinal cortex and hippocampal formation are affected early, while some sensory and motor regions are relatively preserved until later stages. Region-

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

aware analyses can therefore help distinguish early progression signals from terminal global changes.

Cell-Type-Specific Epigenetic Alterations

The human brain contains neurons, microglia, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, endothelial cells, pericytes and other populations with distinct epigenomes. Bulk tissue averages these signals and is affected by disease-related changes in cell proportions. Neuronal loss may create an apparent reduction in neuron-specific marks even when surviving neurons are unchanged, while gliosis may increase immune-associated methylation or accessibility signals.

Cell sorting and single-nucleus technologies have transformed interpretation. Shireby and colleagues showed that many bulk cortical methylation associations were primarily attributable to non-neuronal cell types (Shireby et al., 2022). Morabito, Xiong and Liu independently demonstrated cell-specific chromatin and transcriptional programmes, including microglial enhancer activity, oligodendrocyte modules, neuronal early-stage changes and glial late-stage erosion (Morabito et al., 2021; Xiong et al., 2023; Liu et al., 2025).

Cell resolution also reveals heterogeneity within canonical classes. Disease-associated microglia, reactive astrocytes and stressed neuronal subtypes may represent transitional states rather than fixed categories. Future studies should combine cell-type abundance, within-cell epigenetic change and spatial context rather than treating cell composition only as a nuisance covariate.

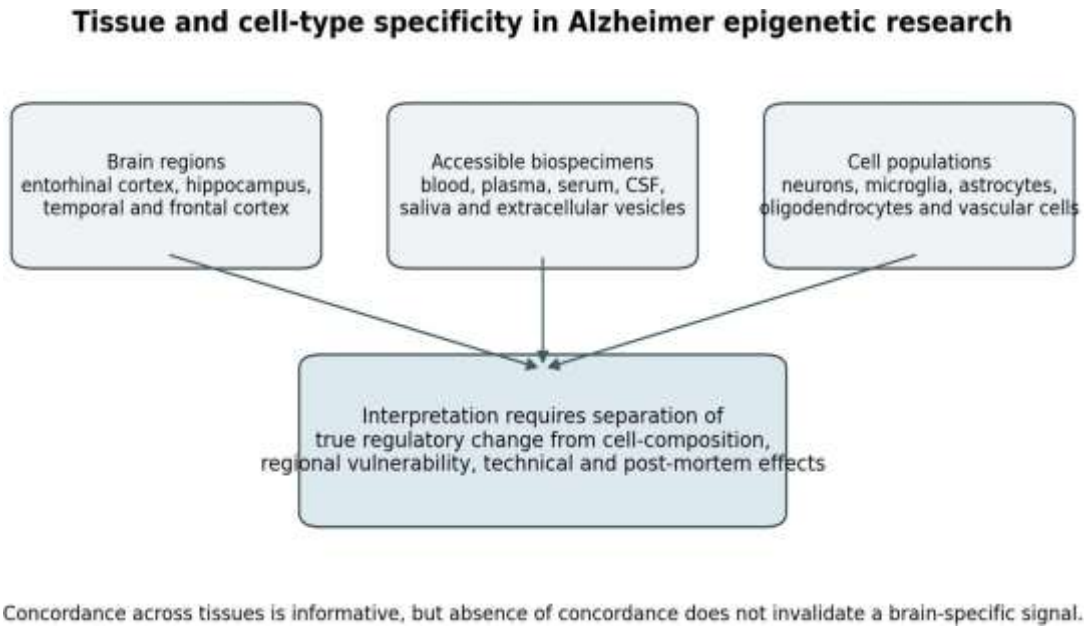


Figure 2.4: Tissue and cell-type specificity in Alzheimer epigenetic research

.

Epigenetic Biomarkers of Alzheimer's Disease

An ideal epigenetic biomarker would detect disease early, distinguish Alzheimer pathology from other causes of cognitive impairment, predict progression, respond to treatment and remain robust across populations and laboratories. Candidate sources include blood-cell methylation, cell-free DNA, plasma or cerebrospinal-fluid non-coding RNAs, extracellular vesicles and epigenetic-age measures. These markers are attractive because some can be measured repeatedly and may capture dynamic biology.

Current evidence is promising but insufficient for routine clinical use. Systematic reviews report recurring microRNA candidates and moderate diagnostic performance, yet heterogeneity in sample handling, normalisation, medication, comorbidity and analytical thresholds limits replication (Fattahi et al., 2024; Pereira et al., 2024). DNA methylation

signatures face similar challenges, including tissue mismatch and cell-composition effects. Established amyloid and phosphorylated-tau biomarkers currently have stronger diagnostic validation than epigenetic markers (Jack et al., 2024).

The most realistic near-term application may be additive. Epigenetic panels could improve prediction of progression or resilience when combined with age, APOE genotype, plasma p-tau, imaging and cognitive measures. Such models require preregistered analysis, external validation, calibration and demonstration of incremental clinical value.

Table 2.8: Evaluation criteria for epigenetic biomarkers

Criterion	Required evidence	Common weakness in current literature
Analytical validity	Accuracy, precision, stability, defined pre-analytics and quality control	Platform-specific results and inconsistent normalisation.
Clinical validity	Discrimination, calibration, longitudinal prediction and external replication	Small case–control cohorts and spectrum bias.
Biological relevance	Tissue/cell plausibility and pathway coherence	Peripheral signal assumed to represent brain without validation.
Incremental value	Improvement beyond age, APOE, cognition, imaging and amyloid/tau biomarkers	Few head-to-head multivariable comparisons.
Generalisability	Performance across ancestry, sex, age and healthcare settings	Under-representation of diverse populations.
Clinical utility	Evidence that using the marker improves decisions or outcomes	Most studies stop at association or AUC reporting.

Epigenetic Therapeutic Targets

Epigenetic mechanisms are potentially reversible, creating interest in disease-modifying therapy. Proposed strategies include histone deacetylase inhibitors, sirtuin modulators, methylation-targeting compounds, inhibitors of chromatin enzymes, antisense oligonucleotides, microRNA mimics or inhibitors and CRISPR-based epigenome editing. Preclinical studies have shown that altering histone acetylation or non-coding RNA pathways can improve memory or modify amyloid and tau-related processes in models (Xiao et al., 2020; Paniri et al., 2024).

Translation is constrained by specificity and delivery. Epigenetic enzymes regulate many genes in multiple tissues, so systemic inhibition can produce toxicity and unpredictable transcriptional effects. The blood–brain barrier limits exposure, and a treatment beneficial in neurons may be harmful in microglia or peripheral immune cells. Disease stage is also critical: restoring plasticity may be useful before severe neuronal loss but insufficient in advanced dementia.

Targeted epigenome editing offers conceptual precision by directing activating or repressing domains to selected regulatory elements without changing DNA sequence. However, off-target activity, durability, immunogenicity, delivery and ethical oversight must be resolved. Therapeutic claims should therefore remain cautious until target engagement, functional benefit and safety are established in rigorous human studies.

Next-Generation Sequencing in Epigenetic Research

Next-generation sequencing has expanded Alzheimer epigenetics beyond candidate loci. Whole-genome bisulphite sequencing provides broad methylome coverage but is expensive, requires substantial input

and generally cannot distinguish 5-methylcytosine from 5-hydroxymethylcytosine without modification. Reduced-representation bisulphite sequencing enriches CpG-dense regions at lower cost but omits much of the genome. Targeted sequencing enables deep coverage of selected loci, whereas methylation arrays provide standardised, cost-effective measurements for large cohorts but interrogate predefined CpGs.

ChIP-seq profiles histone modifications or transcription-factor binding, while ATAC-seq measures accessible chromatin. CUT&RUN and CUT&Tag can achieve lower background and reduced input requirements for chromatin profiling. RNA-seq, small-RNA sequencing and long-read sequencing characterise coding and non-coding transcripts, isoforms and RNA modifications. Each technology answers a different question, and platform choice should follow the biological hypothesis rather than technological fashion.

Quality control is decisive. Bisulphite conversion efficiency, read mapping, duplicate removal, batch effects, antibody performance, library complexity, post-mortem quality and multiple testing can materially alter results. Reproducible pipelines, reference-genome and annotation reporting, version-controlled code and independent validation are essential.

Table 2.7: Comparison of major epigenomic technologies

Technology	Strengths	Limitations	Best use
Methylation array	Cost-effective, standardised, suitable for large cohorts	Predefined CpGs; limited sequence context; 5mC/5hmC not separated	EWAS and replication.
WGBS	Near-genome-wide single-base methylation	High cost, high input and computational burden	Comprehensive methylome discovery.
RRBS	Deep CpG-rich coverage at lower cost	Biased toward CpG islands and promoters	Focused methylome profiling.
ATAC-seq / snATAC-seq	Maps accessible regulatory DNA; cell- resolved options	Sparse single-cell data and strong batch sensitivity	Enhancer and transcription-factor inference.
ChIP-seq	Direct profiling of selected histone marks or factors	Antibody dependence and higher input	Histone-domain and protein-binding studies.
CUT&RUN CUT&Tag	Low input and background; compatible with nuclei	Protocol and antibody optimisation required	Cell-resolved chromatin profiling.
RNA-seq / small-RNA-seq	Measures coding and non-coding transcription	Expression is dynamic and tissue dependent	Regulatory consequence and biomarker discovery.
Long-read sequencing	Isoforms, repeat regions and direct base-modification potential	Cost, error profile and lower throughput	Complex transcripts and integrated modification analysis.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Single-Cell and Spatial Epigenomics

Single-cell and single-nucleus methods resolve regulatory heterogeneity that is obscured in bulk tissue. Single-nucleus ATAC-seq identifies accessible chromatin in frozen brain, while paired or multimodal assays can measure chromatin and gene expression in the same cell or matched populations. These approaches have identified early neuronal changes, disease-associated glial states, cell-specific risk loci and progressive epigenome erosion (Morabito et al., 2021; Xiong et al., 2023; Liu et al., 2025).

Spatial methods add anatomical context by mapping expression or chromatin states within tissue architecture. This is particularly important in Alzheimer's disease because plaques, tangles, vascular lesions and inflammatory niches are spatially organised. A cell's regulatory state may depend on proximity to pathology, local circuitry and regional vulnerability. Spatial multi-omics could therefore distinguish diffuse ageing effects from lesion-associated responses.

These technologies also create analytical challenges: sparse data, doublets, batch integration, cell-type annotation, donor-level pseudoreplication and high computational burden. The biological unit of inference remains the person, not the individual nucleus. Statistical models must preserve donor structure and avoid treating thousands of cells as independent participants.

Multi-Omics Integration

Multi-omics integration seeks to connect genetic variants, epigenetic states, transcription, proteins, metabolites, pathology and clinical outcomes. A methylation association becomes more persuasive when it maps to an accessible enhancer, correlates with target-gene expression, affects a protein network and replicates in an independent cohort. Integration can therefore prioritise regulatory mechanisms that are more likely to be functional than isolated single-layer associations.

Recent Alzheimer studies illustrate this value. Morabito and colleagues linked chromatin accessibility with transcriptional networks; Xiong and colleagues combined accessibility, transcription, genetic variation and peak-to-gene mapping; and Liu and colleagues integrated multiregion epigenomic and transcriptomic profiles with pathology, cognition and resilience (Morabito et al., 2021; Xiong et al., 2023; Liu et al., 2025). Such studies show that disease progression involves coordinated regulatory change rather than one abnormal molecular layer.

Integration can be horizontal across modalities in the same samples or vertical across cohorts and reference atlases. Same-sample designs are powerful but expensive and often limited in size. Cross-cohort integration increases scale but introduces heterogeneity in tissue, platform, clinical definition and batch. Transparent modelling, sensitivity analysis and replication remain essential.

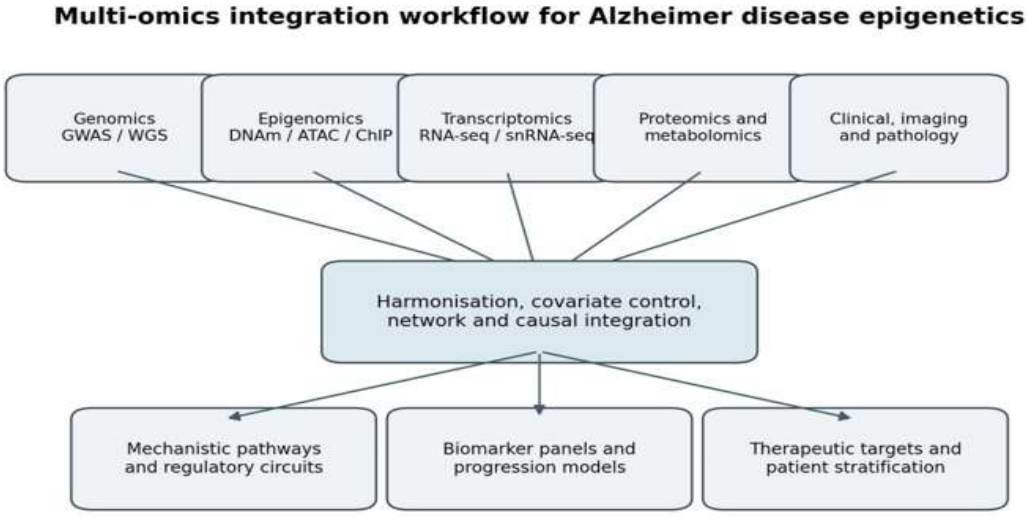


Figure 2.5: Multi-omics integration workflow for Alzheimer disease epigenetics.

Methodological Approaches Used in Previous Studies

Previous Alzheimer epigenetic studies can be grouped into candidate-gene, epigenome-wide, cell-resolved, longitudinal biomarker, experimental and integrative designs. Candidate studies are useful when based on a strong hypothesis and independent validation but have limited discovery potential and are vulnerable to selective reporting. Epigenome-wide association studies provide systematic discovery and have produced replicated methylation loci, yet they require stringent correction for multiple testing and careful control of cell composition and technical variation (De Jager et al., 2014; Lunnon et al., 2014; Alves et al., 2024).

Post-mortem brain studies offer neuropathological precision but are cross-sectional and susceptible to reverse causation. Peripheral longitudinal cohorts support prediction and temporal analysis but may not reflect brain-specific mechanisms. Animal and cellular models permit perturbation and causal experiments, although species differences and simplified pathology limit generalisation. Mendelian randomisation and genetic colocalisation can support causal inference when valid instruments exist, but assumptions such as absence of horizontal pleiotropy must be assessed.

The strongest evidence arises from triangulation: replication across cohorts, convergence across molecular layers, temporal association, genetic support and experimental validation. No single design resolves all questions.

Critical Appraisal of Previous Studies

The literature has several strengths. Independent brain cohorts have replicated methylation signals, large histone and chromatin studies have connected pathology with regulatory domains, and single-cell technologies have established cell specificity. Public resources such as ROSMAP, Mount Sinai Brain Bank, ADNI and AIBL enable replication and multi-omic integration. Recent studies also use continuous pathology and cognitive measures rather than relying solely on diagnostic categories.

Important limitations persist. Many cohorts are modest relative to the dimensionality of the data and are demographically homogeneous. Cross-sectional post-mortem designs cannot determine temporal ordering. Cell composition, post-mortem interval, brain pH, agonal factors, medication, comorbidity and technical batches may confound associations. Different arrays and pipelines produce incomplete overlap, while publication bias favours significant loci. Biomarker studies frequently lack external validation and clinically appropriate comparison groups.

A further concern is causal overstatement. Association with pathology does not prove that an epigenetic mark causes disease; the mark may be a response to inflammation, neuronal loss or treatment. Conversely, a downstream mark can still be useful as a progression biomarker. The literature should therefore separate causal mechanism, mediator, consequence, compensatory response and predictive marker.

Table 2.9: Critical appraisal of the Alzheimer epigenetics literature

Domain	Strength	Persistent limitation
Replication	Several cortical methylation loci and regulatory pathways replicate across cohorts.	Overlap remains incomplete because tissues, platforms and phenotypes differ.
Biological resolution	Single-nucleus and purified-cell studies reveal cell-specific regulation.	Many cohorts remain small and post-mortem.

Progression measurement	Some studies use Braak stage, Continuous pathology and cognition.	Longitudinal repeated epigenetic measurements are uncommon.
Causal inference	Genetic colocalization and experimental models provide supporting evidence.	Reverse causation and pleiotropy remain major concerns.
Population diversity	Large consortia increase sample size.	Ancestral and geographic representation is still limited.
Translation	Candidate biomarkers and therapeutic targets are emerging.	External validation and clinical utility are largely unproven.

Unresolved Scientific Debates

The first unresolved debate concerns cause versus consequence. Some methylation differences appear in presymptomatic pathology, supporting early involvement, but most evidence comes from tissue collected at death. Experimental manipulation can demonstrate biological effects in models, yet the relevance to human late-onset disease remains uncertain. Longitudinal peripheral studies, genetically informed mediation and repeated biomarker measures are needed to establish temporal order.

A second debate concerns tissue validity. Blood is clinically accessible, but the degree to which blood methylation or RNA reflects brain regulation varies by locus and pathway. Lack of concordance may indicate tissue specificity rather than invalidity, while apparent concordance may be driven by systemic ageing or shared genetic control. A third debate concerns whether epigenetic changes are disease-specific or represent general ageing, inflammation and neurodegeneration. Appropriate comparison groups with vascular, Lewy body and frontotemporal pathologies are essential.

A fourth debate concerns therapeutic reversibility. Epigenetic marks are chemically reversible, but disease-associated chromatin states may be stabilised by cell loss, protein aggregates and altered tissue composition. Reversing a mark may not reverse the underlying pathology. Finally, the field continues to debate whether broad omics discovery or focused mechanistic validation should receive priority; in practice, both are required in an iterative cycle.

Summary of the Literature Review

The literature supports a substantial association between Alzheimer's disease and epigenetic dysregulation. Replicated DNA methylation loci, tau-associated histone-acetylation domains, cell-specific chromatin-accessibility changes, non-coding RNA networks and altered epigenetic-age measures collectively indicate that gene regulation changes across the disease continuum. The evidence is strongest for cortical methylation and increasingly strong for cell-resolved chromatin and transcriptional states.

At the same time, Alzheimer epigenetics is not reducible to one global direction of methylation or acetylation. Regulatory change is region-, cell-, stage- and context-specific. Much of the signal in bulk cortex arises from non-neuronal populations, and advanced disease is associated with broad erosion of epigenomic information and cellular identity. Environmental and vascular risks may interact with genetic susceptibility through the epigenome, but direct human causal evidence remains limited.

The translational promise lies in three areas: mechanistic prioritisation of regulatory pathways, multi-marker prediction of progression or resilience, and targeted therapeutic modulation. Each requires stronger longitudinal evidence, diverse populations, external validation and integration with established biomarkers.

Refined Research Gap

The central gap is the limited ability of existing studies to distinguish epigenetic alterations that precede or accelerate Alzheimer progression from

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

those that result from advanced pathology, cell loss or glial activation. Cross-sectional case–control comparisons dominate the literature, while longitudinal and stage-resolved studies remain fewer. Consequently, the timing, direction and clinical meaning of many reported signals are uncertain.

A second gap is insufficient integration across tissue, cell type and molecular layer. Bulk brain studies identify robust associations but cannot fully separate cell-composition change from within-cell regulation. Peripheral biomarker studies are clinically attractive but often lack matched brain or established biomarker data. Single-cell studies offer resolution but are commonly based on post-mortem cohorts with limited diversity and require external replication.

A third gap concerns generalisability and equity. Most epigenomic cohorts under-represent African and other diverse ancestries, low- and middle-income settings and varied exposure profiles. This limits understanding of ancestry-linked regulation, environmental effects and population-specific biomarker performance. A fourth gap concerns translation: few epigenetic signatures have been externally validated for progression prediction or shown to add value beyond age, APOE, cognition, imaging and amyloid/tau biomarkers.

Accordingly, the present thesis is justified by the need for a progression-centred framework that evaluates epigenetic alterations in relation to defined clinical or pathological stages; accounts for age, sex, genetic susceptibility, tissue and cell composition; integrates complementary molecular layers where possible; and distinguishes mechanistic interpretation from biomarker utility. This approach is positioned to clarify which regulatory changes are reproducible, stage-sensitive and clinically informative.

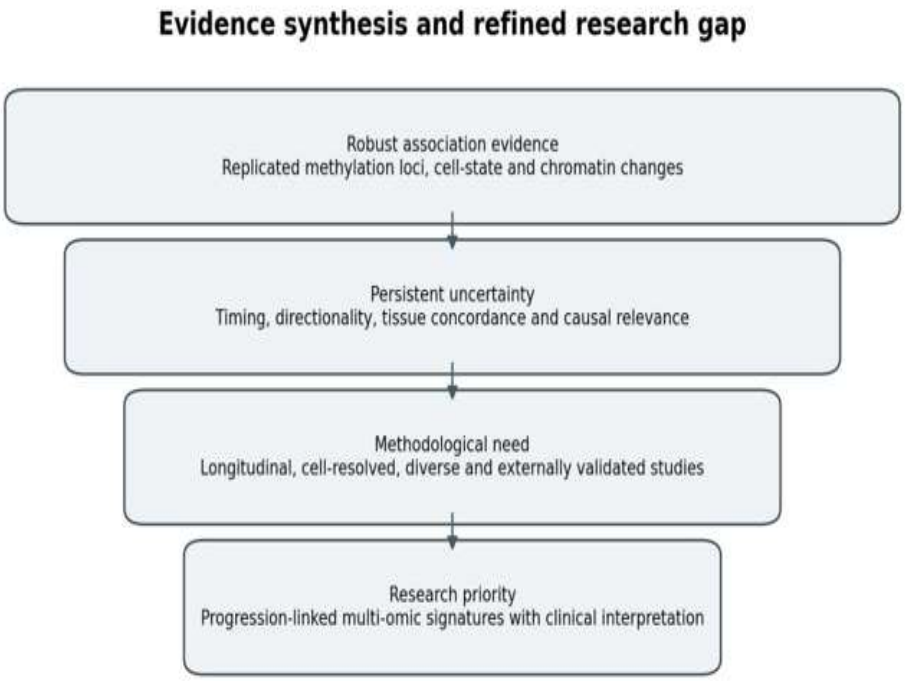


Figure 2.6: Evidence synthesis and refined research gap.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Table 2.10: Refined research-gap matrix

Gap	Why it matters	Required response
Temporal ordering	Cannot distinguish cause, mediator or consequence.	Longitudinal, stage-resolved designs and genetically informed analyses.
Cell composition versus within-cell change	Bulk signals may reflect neuronal loss or gliosis.	Purified-cell, single-nucleus and deconvolution approaches.
Tissue translation	Peripheral markers may not represent brain mechanisms.	Matched-tissue studies and explicit separation of mechanism from prediction.
Population diversity	Biomarkers and mechanisms may not generalise.	Inclusive cohorts, ancestry-aware models and external validation.
Multi-omic coherence	Single-layer associations may lack functional meaning.	Integration of genetics, epigenomics, transcriptomics, proteins and phenotype.
Clinical added value	Association does not establish usefulness.	Compare against established biomarkers and report calibration and decision utility.

Taken together, the evidence indicates that epigenetic dysregulation is neither a peripheral curiosity nor a single-cause explanation of Alzheimer's disease. It is a regulatory dimension that links inherited susceptibility, ageing, exposures and cell-state transitions with the molecular and clinical continuum. The most defensible research strategy is therefore stage-aware, cell-aware, multi-layered and explicit about whether the objective is causal explanation, biological characterisation or clinical prediction.

III. THEORETICAL AND CONCEPTUAL FRAMEWORKS

Introduction

This chapter establishes the theoretical and conceptual architecture for investigating epigenetic alterations associated with Alzheimer's disease progression. Its purpose is not to select a single pathophysiological hypothesis and treat it as a complete explanation of disease. Rather, it evaluates the major explanatory traditions in Alzheimer's disease research and integrates their most defensible components into a systems-oriented framework that can guide variable selection, statistical modelling, interpretation and validation. The framework is designed to account for the long preclinical interval of Alzheimer's disease, the heterogeneity of late-onset disease, the imperfect correspondence between neuropathological burden and cognitive impairment, and the possibility that epigenetic alterations may function as causes, mediators, consequences or compensatory responses.

The need for theoretical integration arises from the structure of the disease itself. Amyloid-β accumulation, pathological tau, neuroimmune activation, oxidative injury, mitochondrial dysfunction, vascular and metabolic disturbances, synaptic failure and biological ageing are strongly interconnected. Human genetic evidence implicates amyloid processing, lipid biology, endosomal trafficking and innate immune regulation, whereas cell-resolved epigenomic studies demonstrate that regulatory alterations differ across neurons, microglia, astrocytes, oligodendrocytes and brain regions (Bellenguez et al., 2022; Liu et al., 2025; Morabito et al., 2021). A framework limited to a single lesion is therefore unlikely to explain why individuals with similar amyloid burdens may show different rates of tau spread, inflammation, neuronal loss and cognitive decline.

Epigenetics provides a plausible bridge among these processes because DNA methylation and hydroxymethylation, histone modifications, chromatin accessibility and non-coding RNAs influence how inherited susceptibility and environmental conditions are translated into cell-specific patterns of gene regulation. These regulatory processes are essential for neuronal identity, memory consolidation, synaptic plasticity and immune-cell state. They are also responsive to ageing, inflammation, metabolic stress and toxic exposure. Consequently, an epigenetic signature associated with disease progression may reflect an upstream susceptibility state, a mediator of pathological signalling, a downstream consequence of tissue damage, or a homeostatic attempt to preserve function (De Plano et al., 2024; Maity et al., 2021; Xiong et al., 2023).

The chapter proceeds in four stages. First, it describes the overarching theoretical framework and the criteria used to integrate competing hypotheses. Second, it reviews the amyloid cascade, tau propagation, neuroinflammation, oxidative stress, mitochondrial cascade, epigenetic dysregulation, biological ageing and gene–environment interaction perspectives. Third, it places these perspectives within a systems-biology model. Fourth, it translates the theoretical synthesis into a conceptual framework specifying determinants, epigenetic predictors, mediators, moderators, confounders, outcomes and testable propositions aligned with the objectives and hypotheses stated in Chapter One.

Theoretical Framework

A theoretical framework identifies the explanatory assumptions that determine what is measured, how relationships are expected to operate and how empirical findings are interpreted. For the present study, an adequate framework must satisfy several conditions. It must accommodate both autosomal-dominant and sporadic disease, explain age dependence, recognise the prolonged interval between molecular pathology and dementia, account for regional and cell-type vulnerability, distinguish pathology from clinical expression, and permit reciprocal causation and feedback. It must also provide a coherent role for epigenetic regulation without assuming that every disease-associated epigenetic difference is causal.

The framework adopted here is an integrated epigenetic systems model of Alzheimer's disease progression. It treats Alzheimer's disease as a multilevel process in which inherited and life-course determinants shape the regulatory state of specific cell populations; these regulatory states influence amyloid processing, tau biology, immune activation, mitochondrial function, oxidative balance, proteostasis and synaptic plasticity; and the resulting network disturbances produce progressive pathology, cognitive impairment and loss of function. Once established, pathological processes feed back on the epigenome through inflammation, oxidative damage, altered cellular composition and activity-dependent transcription. Disease progression is therefore represented as a dynamic system of interacting loops rather than a single linear cascade (De Strooper & Karran, 2016; Doig, 2018).

The integrated framework is deliberately pluralistic but not theoretically indiscriminate. Each component hypothesis is retained only to the extent that it contributes a distinct, empirically testable mechanism. Amyloid is treated as a defining and potentially initiating process, particularly in genetically determined disease, but not as a sufficient explanation of clinical progression. Tau is treated as a principal correlate and mediator of neuronal dysfunction and anatomical spread. Neuroinflammation, mitochondrial dysfunction and oxidative stress are treated as amplifying and context-dependent processes. Epigenetic dysregulation is treated as a regulatory layer that may precede, mediate or follow these mechanisms. Biological ageing supplies the time-dependent loss of resilience, while gene–environment interaction explains interindividual variability. Systems biology provides the architecture through which these components are analysed jointly.

This theoretical position has important inferential consequences. Cross-sectional association alone cannot establish whether an epigenetic difference drives disease or reflects pathology. Stronger causal interpretation requires temporal precedence, dose-

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

response relationships, genetic or experimental support, mediation evidence, concordant changes in chromatin and transcription, replication, and biological plausibility. The framework therefore prioritises longitudinal data, ordered clinical and neuropathological stages, cell-type-aware analysis, multi-omic integration and independent validation.

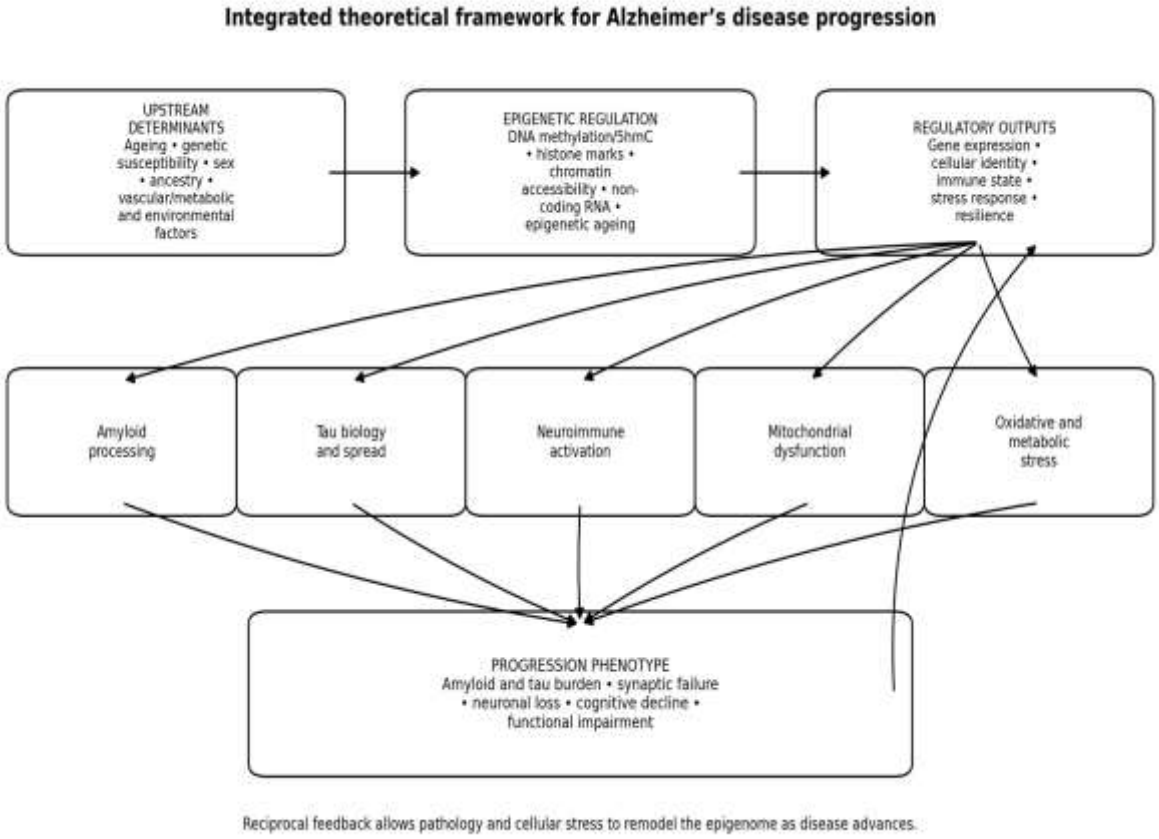


Figure 3.1: Integrated theoretical framework for epigenetic alterations associated with Alzheimer's disease progression.

Amyloid Cascade Hypothesis

The amyloid cascade hypothesis proposes that abnormal production, processing, aggregation or clearance of amyloid-β initiates a sequence of events leading to tau pathology, synaptic dysfunction, neuronal injury and dementia. Its strongest support comes from autosomal-dominant Alzheimer's disease, in which pathogenic variants in APP, PSEN1 or PSEN2 alter amyloid biology, and from trisomy 21, in which increased APP dosage is associated with early amyloid deposition. The hypothesis also provided a coherent molecular explanation for the extracellular plaques observed in Alzheimer's disease and shaped several decades of biomarker and therapeutic development (Hardy & Selkoe, 2002; Long & Holtzman, 2019).

For this thesis, the principal value of the amyloid framework is that it identifies a biologically defined upstream axis against which epigenetic regulation can be examined. Methylation, chromatin accessibility or non-coding RNA changes at genes involved in APP processing, endosomal trafficking, lipid handling and amyloid clearance may influence the timing or intensity of amyloid accumulation. Conversely, amyloid exposure can activate glial and neuronal stress responses that remodel chromatin and transcription. The relationship is therefore potentially

bidirectional even when amyloid precedes other core pathological events.

The limitations of a strict amyloid-first model are equally important. Amyloid burden correlates imperfectly with cognitive severity, can be detected in cognitively unimpaired individuals and does not by itself account for regional neurodegeneration or clinical heterogeneity. Therapeutic amyloid reduction confirms that amyloid is biologically modifiable, but the difference between target engagement and cognitive benefit indicates that downstream and parallel processes influence outcome. De Strooper and Karran (2016) consequently distinguished an early biochemical phase from a later cellular phase in which glial, vascular and neuronal responses become increasingly autonomous.

Within the integrated framework, amyloid is conceptualised as a necessary disease-defining process in typical Alzheimer's disease and an important initiator or facilitator, but not as a sufficient determinant of progression rate. The study therefore tests whether epigenetic alterations are associated not only with amyloid burden but also with tau, inflammatory, transcriptional and cognitive outcomes after controlling for amyloid. Epigenetic features that remain associated with progression after adjustment for amyloid would support mechanisms extending beyond a simple linear cascade.

Tau Propagation Hypothesis

The tau propagation hypothesis centres on the abnormal phosphorylation, conformational change, aggregation and anatomical spread of tau. Braak and Braak (1991) demonstrated a stereotyped distribution of neurofibrillary pathology that begins in transentorhinal and entorhinal regions and progresses through limbic and neocortical areas. Across clinicopathological studies, the extent of tau pathology and neuronal loss generally corresponds more closely with cognitive impairment than plaque burden does, although co-pathologies and cognitive reserve modify this relationship (Nelson et al., 2012).

Tau biology is relevant to epigenetic research in several ways. Neuronal activity, kinase and phosphatase networks, proteostasis, autophagy and cytoskeletal regulation are transcriptionally controlled and responsive to chromatin state. Histone acetylation changes have been associated with tau burden in human brain, and large-scale analyses suggest that tau-related pathology is accompanied by broad regulatory reorganisation rather than alteration at only a few loci (Klein et al., 2019). Tau accumulation may also disturb nuclear architecture, heterochromatin, DNA repair and neuronal identity, creating feedback from protein aggregation to epigenomic instability.

A pure propagation model nevertheless leaves several questions unresolved. It does not fully explain why tau pathology begins in particular neuronal populations, why its spread differs among individuals, or why some people maintain cognition despite substantial pathology. These limitations create space for epigenetic modifiers of vulnerability and resilience. Cell-type-specific chromatin landscapes may determine which neurons express receptors, trafficking proteins, stress-response genes and proteostatic mechanisms that facilitate or resist tau-associated dysfunction.

The present framework therefore treats tau as a central mediator of progression and as a major outcome against which epigenetic alterations should be evaluated. Epigenetic loci or networks associated with higher Braak stage, faster cognitive decline or tau-related transcriptional disruption are expected to have greater progression relevance than signals associated only with case-control status. The framework also permits reverse causation: advanced tau pathology can itself reshape DNA methylation, histone marks and chromatin accessibility.

Neuroinflammation Hypothesis

The neuroinflammation hypothesis proposes that persistent or dysregulated activation of microglia, astrocytes, complement pathways, cytokine signalling and innate immune networks contributes materially to Alzheimer's disease progression. Neuroimmune activation can initially support amyloid clearance, tissue repair and homeostasis, but chronic activation may promote synaptic pruning, oxidative injury, altered lipid handling, impaired phagocytosis and amplification of tau pathology. Human genetic findings involving TREM2, PLCG2,

ABI3 and related loci provide strong evidence that microglial biology is not merely a terminal response to neuronal death (Heneka et al., 2015; Sims et al., 2017).

Epigenetic regulation is fundamental to immune-cell identity and state transitions. Microglia respond to local signals through enhancer activation, transcription-factor binding, histone modification and changes in chromatin accessibility. Single-nucleus studies of Alzheimer's disease have identified cell-type-specific regulatory changes in microglia and other glial populations, while bulk-tissue methylation associations are often substantially influenced by non-neuronal composition (Morabito et al., 2021; Shireby et al., 2022). These observations imply that a disease-associated methylation signal may indicate either a change within a cell type or a shift in the proportion of activated cell populations.

The inflammatory framework is particularly useful for explaining positive feedback. Amyloid and tau can activate innate immune pathways; inflammatory mediators can alter amyloid clearance, tau phosphorylation and neuronal function; oxidative and metabolic stress further reinforce immune activation; and these conditions remodel the epigenome. Such feedback can make late-stage disease partially independent of the original initiating event. It also explains why an epigenetic mark associated with inflammation may be both a consequence of existing pathology and an amplifier of further progression.

In the conceptual model, neuroinflammation occupies a mediating and moderating position. Immune-related epigenetic alterations are expected to link upstream determinants and core pathology to downstream neurodegeneration. Their effects may differ by APOE genotype, age, sex, vascular status and cell composition. Accordingly, analyses must distinguish cell proportion from cell-intrinsic regulation and assess whether immune-related methylation or accessibility features explain variation in pathology and cognition beyond standard clinical covariates.

Oxidative Stress Hypothesis

The oxidative stress hypothesis attributes an important role to an imbalance between reactive oxygen or nitrogen species and cellular antioxidant capacity. The brain is particularly vulnerable because of its high oxygen consumption, lipid-rich membranes, long-lived post-mitotic neurons and dependence on mitochondrial energy production. Oxidative damage can affect nucleic acids, proteins and lipids, impair membrane and synaptic function, disrupt calcium regulation and promote proteostatic failure. In Alzheimer's disease, oxidative stress interacts with amyloid, tau, neuroinflammation and vascular dysfunction rather than operating as an isolated pathway (Long & Holtzman, 2019; Scheltens et al., 2021).

Oxidative conditions can influence the epigenome directly and indirectly. Reactive species may alter DNA bases, interfere with methylation measurement, affect the activity of DNA methyltransferases and TET enzymes, modify histone-regulating enzymes and change the availability of metabolic cofactors required for chromatin regulation. Oxidative stress also activates transcriptional programmes governing antioxidant defence, apoptosis, autophagy and inflammation. Conversely, epigenetic repression of protective pathways may reduce neuronal capacity to tolerate metabolic and inflammatory stress.

The hypothesis is limited by the difficulty of establishing temporal direction. Oxidative damage can occur early, but it also increases as mitochondria fail, inflammation intensifies and neurons degenerate. Peripheral oxidative markers may not accurately represent regional brain processes, and post-mortem tissue is vulnerable to agonal and storage effects. For these reasons, oxidative stress is treated in this thesis as an amplifying mechanism and potential mediator rather than an independently sufficient cause.

The framework predicts that epigenetic changes in antioxidant, mitochondrial, inflammatory and DNA-repair pathways will be associated with greater pathological burden or cognitive decline. However, such associations require cautious interpretation. Multi-omic evidence showing coordinated chromatin, expression and pathway changes provides stronger support than an isolated methylation difference, particularly when the signal replicates and is observed before severe neuronal loss.

Mitochondrial Cascade Hypothesis

The mitochondrial cascade hypothesis proposes that inherited mitochondrial capacity and durability interact with ageing to determine when bioenergetic decline crosses a threshold that initiates or accelerates Alzheimer-related pathology. Mitochondria influence ATP production, calcium buffering, reactive oxygen species, apoptosis, lipid metabolism, innate immune signalling and the availability of metabolites used by epigenetic enzymes. The hypothesis is attractive because it connects age dependence with systemic metabolic features and with amyloid, tau, vascular and inflammatory processes (Swerdlow, 2023).

Mitochondrial dysfunction can influence epigenetic regulation through several mechanisms. Altered one-carbon metabolism changes methyl-donor availability; changes in acetyl-CoA, NAD+, α-ketoglutarate and related metabolites affect histone acetylation, sirtuins and demethylation reactions; oxidative stress can modify DNA and chromatin enzymes; and mitochondrial signals can reprogramme nuclear transcription. Epigenetic changes may in turn regulate mitochondrial biogenesis, mitophagy, respiratory-chain components and stress responses, creating reciprocal nuclear–mitochondrial communication.

The principal limitation of the hypothesis is the challenge of separating primary mitochondrial decline from dysfunction caused by amyloid, tau, inflammation or neuronal inactivity. Mitochondrial abnormalities are widespread across ageing and many neurodegenerative diseases, reducing disease specificity. Nevertheless, the hypothesis contributes a mechanistic route through which age and metabolic exposures may alter the epigenome and lower cellular resilience.

In the integrated model, mitochondrial dysfunction is treated as both an upstream vulnerability factor and a downstream amplifier. Epigenetic age acceleration and methylation changes in mitochondrial and metabolic pathways are expected to associate with pathological and cognitive measures, but interpretation should consider whether the signal is disease-specific, systemic, cell-type dependent or secondary to altered tissue composition.

Epigenetic Dysregulation Hypothesis

The epigenetic dysregulation hypothesis proposes that persistent changes in DNA methylation and hydroxymethylation, histone modifications, chromatin accessibility, higher-order chromatin organisation and non-coding RNA regulation contribute to the initiation, maintenance or progression of Alzheimer's disease. Unlike DNA sequence variation, epigenetic states are potentially dynamic and may integrate ageing, genotype, cellular activity and environmental exposure. This makes them plausible mediators of heterogeneity in age at onset, regional vulnerability, rate of progression and cognitive resilience (Klein et al., 2016; Sanchez-Mut & Gräff, 2015).

Human epigenome-wide association studies provide reproducible evidence at selected loci, including ANK1 and other regions associated with neuropathology, but they also reveal substantial heterogeneity across tissues, brain regions and analytical methods (De Jager et al., 2014; Lunnon et al., 2014). More recent work has shown that many cortical methylation signatures are driven by non-neuronal cells and that genetic risk variants can influence disease through cell-type-specific regulatory elements (Shireby et al., 2022; Xiong et al., 2023). Single-cell and multiregion analyses further demonstrate widespread epigenomic reconfiguration associated with progression and resilience (Liu et al., 2025).

The hypothesis does not imply that all observed differences are causal. At least six interpretations are possible: an epigenetic feature may precede disease and confer susceptibility; mediate the effect of genotype or exposure; reflect a cellular response to pathology; compensate for stress; result from altered cell composition; or arise from technical and post-mortem artefact. The theoretical framework therefore distinguishes an epigenetic association from an epigenetic mechanism. Mechanistic confidence increases when a signal is temporally early, cell-intrinsic, functionally linked to gene regulation, consistent across molecular layers, supported by genetic evidence and replicated independently.

Epigenetic dysregulation is the central regulatory component of this thesis. DNA methylation is the primary empirical layer because it is comparatively stable and widely available across large cohorts, while chromatin accessibility, histone modifications, transcriptomics and non-coding RNA provide functional context where available. The hypothesis predicts ordered differences across disease stages, associations with amyloid, tau and cognition, cell-type specificity, and incremental predictive value beyond age, sex, APOE and clinical variables.

Biological Ageing Theory

Biological ageing theory distinguishes chronological time from the cumulative molecular and physiological changes that reduce resilience and increase vulnerability to disease. Age is the strongest population-level risk factor for late-onset Alzheimer's disease, yet individuals of the same chronological age differ markedly in health, neuropathology and cognitive function. Epigenetic clocks estimate aspects of biological ageing from DNA methylation patterns and provide a quantitative framework for examining whether accelerated molecular ageing is associated with neurodegeneration (Horvath, 2013; Levine et al., 2018).

Different clocks capture different biological constructs. First-generation clocks were optimised to predict chronological age, whereas later models incorporate mortality, healthspan or physiological variables. A significant age-acceleration measure therefore cannot be interpreted without specifying the algorithm, tissue and reference population. Reviews of neurodegenerative disease studies report promising associations but also inconsistency across clocks and samples (Yang et al., 2023). Brain and blood clocks may capture partly distinct processes, and blood-based acceleration can reflect systemic inflammation or comorbidity rather than brain-specific ageing.

Within this thesis, biological ageing is both a determinant and an epigenetic phenotype. Chronological age shapes cumulative exposure, cellular senescence, immune state and mitochondrial function. Epigenetic age acceleration may mediate or mark these processes and may interact with APOE genotype, vascular risk and pathology. The framework predicts that accelerated epigenetic ageing will be associated with greater amyloid or tau burden, worse cognition or more advanced clinical stage, but it does not assume that clock acceleration is itself causal.

The theory also supports analysis of resilience. Individuals whose molecular ageing is slower than expected, or whose regulatory networks remain comparatively preserved despite pathology, may maintain cognition longer. This possibility motivates the inclusion of cognitive resilience and discordance between pathology and cognition as secondary interpretive outcomes.

Gene–Environment Interaction Theory

Gene–environment interaction theory proposes that the effects of inherited susceptibility depend on exposure context and that environmental effects vary according to genotype. Late-onset Alzheimer's disease is especially compatible with this model because its genetic architecture is polygenic and its clinical expression is influenced by education, cardiovascular and metabolic health, physical activity, hearing, air pollution, social conditions and other life-course factors (Bellenguez et al., 2022; Livingston et al., 2024).

Epigenetic regulation offers a mechanism through which interactions may become biologically embedded. Exposure-related changes in inflammation, metabolism, endocrine signalling or neuronal activity can alter chromatin state, while genetic variants can affect transcription-factor binding, enhancer activity, methylation and expression. Regulatory variants may therefore create allele-specific responses to ageing or exposure. Bryzgalov et al. (2023)nillustrated how non-coding variants can intersect with histone and transcriptional data to identify regulatory susceptibility mechanisms in Alzheimer's disease.

The theory is methodologically demanding. Exposure measurement is often imprecise, retrospective or unavailable; genetic ancestry and social conditions are correlated; and interaction tests require large samples. Epigenetic marks may also reflect

consequences of disease-related behaviour, such as reduced activity or altered diet, creating reverse causation. For this reason, the primary study focuses on well-measured covariates and treats incompletely measured exposures as limitations rather than inventing precision that the data cannot support.

In the conceptual framework, genotype and exposure are upstream determinants and potential moderators. APOE and polygenic risk may modify the association between epigenetic features and progression. Environmental and vascular factors may influence both epigenetic predictors and outcomes, making them confounders, mediators or effect modifiers depending on temporal ordering. Analytical models must therefore be specified from causal assumptions rather than automatic covariate inclusion.

Systems Biology Perspective

Systems biology examines disease as the behaviour of interacting molecular, cellular and physiological networks. This perspective is particularly appropriate for Alzheimer's disease because no single mechanism accounts for the diversity of initiation, progression and clinical expression. Amyloid, tau, immune activation, mitochondrial function, lipid metabolism, vascular integrity, proteostasis and synaptic activity form interconnected networks with nonlinear feedback. A perturbation may be buffered in one individual but amplified in another because of differences in genotype, regulatory state, comorbidity, cellular composition and reserve.

The systems perspective changes the unit of analysis. Rather than relying exclusively on single CpG sites or genes, it encourages the study of differentially methylated regions, co-methylation modules, enhancer networks, transcriptional programmes, pathways and cell states. It also supports integration of methylation with chromatin accessibility, histone marks, transcriptomics, proteomics and genotype. Multi-omic agreement does not prove causality, but it can identify coherent regulatory chains and reduce the likelihood that an isolated association is biologically incidental.

Systems biology also clarifies why positive feedback is central to progression. Amyloid and tau can activate glia; inflammation increases oxidative and metabolic stress; mitochondrial dysfunction reduces proteostatic and synaptic capacity; neuronal injury alters cell composition and chromatin; and regulatory changes may reinforce immune or stress programmes. Doig (2018) argued that such feedback loops are inconsistent with a purely unidirectional cascade. The integrated model adopted here therefore represents disease progression as a changing network whose dominant drivers may differ by stage.

The practical implication is a tiered analytical strategy. Site-level discovery identifies specific signals; regional and network analyses capture coordinated regulation; pathway analysis interprets functional themes; cell-type methods identify biological sources; and predictive models evaluate whether selected features add clinically relevantminformation. Independent replication and cross-cohort consistency remain essential because highly dimensional systems analyses are vulnerable to overfitting.

Table 3.1: Comparative appraisal of the theoretical perspectives incorporated into the study.

Perspective	Central proposition	Contribution to this thesis	Principal limitation
Amyloid cascade	Abnormal amyloid biology initiates or facilitates downstream pathology.	Defines a core pathological axis and tests epigenetic regulation of amyloid-related processes.	Amyloid burden alone does not explain clinical heterogeneity or progression rate.

Tau propagation	Pathological tau spreads through vulnerable neural systems and mediates dysfunction.	Prioritises Braak stage, tau burden and tau-related regulatory networks as progression outcomes.	Does not fully explain initial vulnerability or resilience.
Neuroinflammation	Persistent glial and innate-immune activation amplifies pathology and injury.	Motivates cell-type-aware analysis of immune enhancers, methylation and transcription.	Immune activation can be protective, pathogenic or secondary depending on stage.
Oxidative stress	Redox imbalance damages macromolecules and disrupts neuronal homeostasis.	Links epigenetic regulation with antioxidant, DNA-repair and stress pathways.	Temporal direction and disease specificity are difficult to establish.
Mitochondrial cascade	Age-related mitochondrial decline initiates or amplifies Alzheimer-related changes.	Connects metabolic cofactors, epigenetic enzymes, ageing and neuronal vulnerability.	Mitochondrial dysfunction may be primary or secondary.
Epigenetic dysregulation	Persistent regulatory alterations influence gene activity and cellular state.	Provides the primary measurable mechanism and basis for multi-omic integration.	Association can reflect cause, response, composition or artefact.
Biological ageing	Molecular ageing determines vulnerability beyond chronological age.	Supports analysis of epigenetic age acceleration and resilience.	Different clocks measure different constructs and tissues.
Perspective	Central proposition	Contribution to this thesis	Principal limitation
Gene–environment interaction	Genetic effects vary by exposure and exposures act differently by genotype.	Frames APOE, ancestry, vascular and exposure variables as determinants or moderators.	Large samples and high-quality exposure data are required.
Systems biology	Disease emerges from interacting networks and feedback loops.	Integrates site, region, network, pathway, cell-type and predictive analyses.	High-dimensional models risk overfitting and require replication.

Conceptual Framework

The conceptual framework translates the theoretical synthesis into a set of measurable constructs and directional relationships. It is organised around six domains: upstream determinants; primary epigenetic features; mediating molecular and cellular processes; Alzheimer-related outcomes; contextual moderators and confounders; and validation criteria. The framework is stage-sensitive and recognises reciprocal feedback, but the primary empirical models impose clear directional assumptions to enable statistical testing.

Upstream determinants include chronological age, sex, APOE genotype, broader genetic susceptibility, ancestry, education and available vascular, metabolic or exposure variables. These factors may influence both epigenetic states and disease outcomes. Primary epigenetic features include differentially methylated positions and regions, co-methylation networks, DNA methylation age acceleration, chromatin accessibility, histone modifications and regulatory non-coding RNAs. DNA methylation is the principal exposure layer, while the other layers are used for functional annotation, triangulation or secondary analysis.

Mediating processes include altered gene expression, enhancer activity, immune-cell state, mitochondrial and oxidative pathways, amyloid processing, tau-related regulation, synaptic function and proteostasis. Outcomes include clinical stage, cognitive performance or longitudinal decline, neuropathological stage, amyloid and tau burden, and classification or progression prediction. Cell type, tissue, brain region, post-mortem factors, technical

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

batch and cohort are not peripheral details; they condition the meaning of every molecular association.

The model distinguishes discovery, explanation and translation. Discovery asks which epigenetic features are associated with progression. Explanation asks whether those features align with regulatory activity, cell type and biological pathways. Translation asks whether a parsimonious, independently validated signature adds predictive information beyond established variables. A signal must pass through these increasingly demanding levels before it can be considered a credible biomarker or therapeutic target.

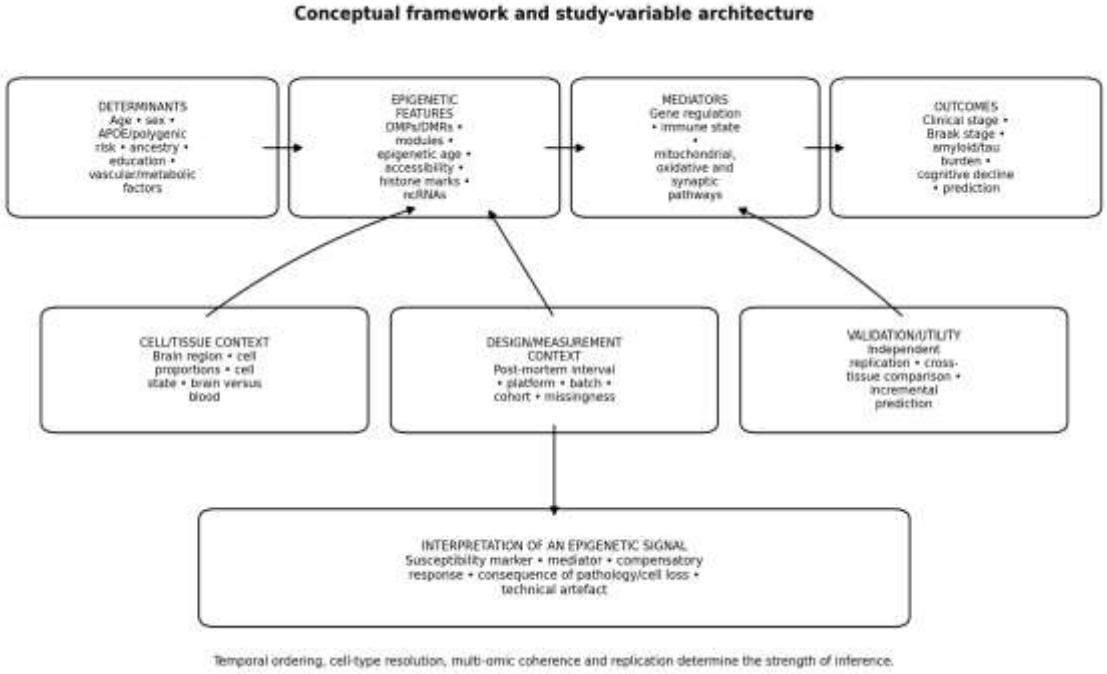


Figure 3.2: Conceptual framework and study-variable architecture.

Description of Variables

The principal independent variables are quantitative epigenetic measurements. At the site level, DNA methylation may be represented as beta values or transformed M values after quality control and normalisation. Region-level variables summarise spatially correlated sites, while network variables represent modules of co-methylated loci. Epigenetic age acceleration is calculated using the selected clock and defined as a residual or difference from chronological age according to the algorithm. Where available, chromatin-accessibility peaks, histone-mark enrichment, microRNA or long non-coding RNA abundance and gene-expression values provide complementary regulatory measures.

The primary dependent variables represent Alzheimer's disease progression. Because progression is multidimensional, no single outcome is assumed to capture the entire process. Ordered clinical categories, Braak stage, amyloid or tau burden, repeated cognitive scores, functional severity and conversion from mild cognitive impairment to dementia are treated as related but distinct outcomes. Cross-sectional stage analyses estimate differences across ordered states, whereas longitudinal models estimate within-person change or time to conversion.

Mediating variables are molecular processes through which an epigenetic feature may relate to outcome. Examples include altered expression of a nearby or distal gene, accessibility of a linked regulatory element, activation of an immune or metabolic

pathway, and cell-state transitions. Mediation will be interpreted cautiously because statistical mediation in observational data does not by itself establish biological causation.

Moderators are variables that change the magnitude or direction of an association. APOE genotype, sex, age, ancestry, tissue, brain region and cell type are plausible moderators. Confounders include variables associated with both the epigenetic feature and outcome but not lying on the proposed causal pathway. Major technical and biological covariates include cell proportions, batch, post-mortem interval, tissue pH, assay quality and cohort. Some variables may be moderators in one model and confounders or mediators in another; their role must therefore be declared for each analysis.

Validation variables describe reproducibility and clinical utility. These include effect-direction consistency, replication significance, cross-cohort heterogeneity, cross-tissue concordance, discrimination, calibration and improvement over a baseline model. Predictive performance will be evaluated in held-out or independent data rather than in the discovery sample alone.

Table 3.2: Constructs, operational indicators and roles in the conceptual framework.

Construct domain	Illustrative variables or indicators	Analytical role
Upstream determinants	Age, sex, APOE, polygenic risk, ancestry, education, vascular and metabolic factors	Predictors, confounders or moderators depending on the model
DNA methylation	Site-level methylation, DMPs, DMRs, co-methylation modules	Primary epigenetic predictors
Other regulatory layers	5hmC, chromatin accessibility, histone marks, miRNA/lncRNA, gene expression	Functional annotation, secondary predictors or mediators
Biological ageing	Clock-specific age acceleration or pace-of-ageing measures	Epigenetic phenotype, predictor or mediator
Cell and tissue context	Estimated cell proportions, cell type, brain region, brain/blood source	Confounder, moderator and biological stratifier
Neuropathology	Braak stage, amyloid burden, tau burden and related markers	Primary biological outcomes
Clinical progression	Clinical stage, repeated cognition, conversion, functional severity	Primary clinical outcomes
Technical context	Platform, batch, post-mortem interval, pH, quality metrics, missingness	Nuisance covariates and quality-control variables
Validation and utility	Replication effect, heterogeneity, AUC, calibration, incremental R^2	Evidence of reproducibility and translational value

Proposed Relationships among Variables

The framework proposes that ageing, genetic susceptibility and life-course conditions influence cell-specific epigenetic regulation. These regulatory states affect transcription and molecular pathways relevant to amyloid, tau, immunity, mitochondrial function, oxidative balance and synaptic integrity. The resulting pathway disturbances contribute to neuropathological burden and cognitive decline. Disease processes then feed back on chromatin through inflammation, oxidative injury, altered neuronal activity and shifts in cell composition.

The first proposed relationship is between disease stage and DNA methylation. Ordered progression is expected to be associated with differentially methylated sites, regions and co-methylation networks. A monotonic pattern across controls, mild

cognitive impairment and Alzheimer's dementia would be more consistent with progression than a difference observed only between end-stage cases and controls, although nonlinear and stage-specific relationships are also biologically plausible.

The second relationship concerns robustness to covariate adjustment. Priority associations should persist after adjustment for demographic, genetic, cell-composition, regional and technical factors. Strong attenuation after cell-type adjustment would suggest that the original signal primarily reflects cellular redistribution, which may still be biologically informative but has a different interpretation from cell-intrinsic regulation.

The third relationship concerns multi-omic coherence. Disease-associated methylation is expected to coincide with altered chromatin accessibility, histone state or expression when it affects an active regulatory element. Directionality depends on genomic context; promoter methylation is not universally repressive, and gene-body or enhancer methylation can have context-dependent effects. The study therefore avoids simplistic assumptions and uses annotation and empirical correlation.

The fourth relationship concerns cell and tissue specificity. Brain-region and cell-type effects are expected because Alzheimer pathology spreads anatomically and because neurons and glia perform different functions. Peripheral signatures may have biomarker value even when they do not mirror brain mechanisms, but cross-tissue discordance must be reported rather than concealed.

The fifth relationship concerns biological ageing. Greater epigenetic age acceleration is expected to associate with more severe pathology or faster cognitive decline, but the magnitude may vary by clock and tissue. The sixth relationship concerns prediction: a selected epigenetic signature should improve discrimination, calibration or explained variance beyond age, sex, APOE and standard clinical variables. The seventh concerns replication: priority signals should show consistent direction and acceptable heterogeneity in independent cohorts.

Propositions Derived from the Framework

The theoretical and conceptual synthesis yields the following propositions. These propositions are broader than the statistical hypotheses because they describe the expected biological structure of findings and guide interpretation when results are heterogeneous.

1. P1. Alzheimer's disease progression is associated with coordinated epigenetic alterations rather than a single universal locus, and these alterations will differ by stage, region and cell type.
2. P2. A subset of disease-associated epigenetic alterations will remain significant after adjustment for demographic, genetic, cell-composition and technical covariates, indicating regulation beyond compositional change.
3. P3. Epigenetic features with concordant chromatin, transcriptional or pathway evidence will show greater biological plausibility than isolated methylation associations.
4. P4. Immune, mitochondrial, oxidative, proteostatic and synaptic regulatory programmes will form interacting modules that mediate or amplify the relationship between core pathology and cognitive decline.
5. P5. Epigenetic age acceleration will be associated with greater vulnerability, while comparatively preserved regulatory states may contribute to cognitive resilience.
6. P6. Genetic susceptibility, particularly APOE-related and immune-regulatory risk, will modify selected epigenetic associations and their relationship with pathology.
7. P7. Brain and peripheral epigenetic signatures will show partial rather than complete concordance; peripheral markers may retain predictive value without representing identical mechanisms.
8. P8. Only a parsimonious and independently validated epigenetic signature will provide credible incremental prediction beyond established demographic, genetic and clinical variables.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Table 3.3: Alignment of conceptual propositions with the study objectives and research questions.

Proposition	Primary objective(s)	Research question(s)	Related hypothesis
P1	1 and 4	RQ1 and RQ4	H01/H11 and H03/H13
P2	2	RQ2	H02/H12
P3	3	RQ3	H03/H13
P4	3 and 4	RQ3 and RQ4	H03/H13
P5	5	RQ5	H04/H14
P6	2 and 4	RQ2 and RQ4	H02/H12
P7	7	RQ7	H06/H16
P8	6 and 7	RQ6 and RQ7	H05/H15 and H06/H16

Diagram of the Conceptual Framework

Figure 3.2 presents the operational conceptual framework. The model begins with exogenous and inherent determinants that influence primary epigenetic features. Epigenetic features are linked to disease outcomes partly through gene regulation, cell-state transitions and pathway activity. Cell and tissue context, measurement conditions and study design affect the observed associations. The lower interpretive layer classifies an identified signal according to the strongest explanation supported by the evidence.

Figure 3.3 adds a temporal dimension. Epigenetic alterations observed before substantial pathology are more plausibly interpreted as susceptibility, resilience or early regulatory changes. During preclinical disease and mild cognitive impairment, regulatory amplification and cell-state transitions may become prominent. In advanced disease, neuronal loss, gliosis and severe pathology increasingly dominate bulk-tissue signals. The same locus can therefore have different meaning depending on when, where and in which cell type it is measured.

The diagrams are conceptual rather than deterministic. They do not assert that every pathway operates in every participant or that all relationships are linear. Their purpose is to make assumptions explicit, prevent overinterpretation and align the analytical strategy with the theoretical position of the thesis.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

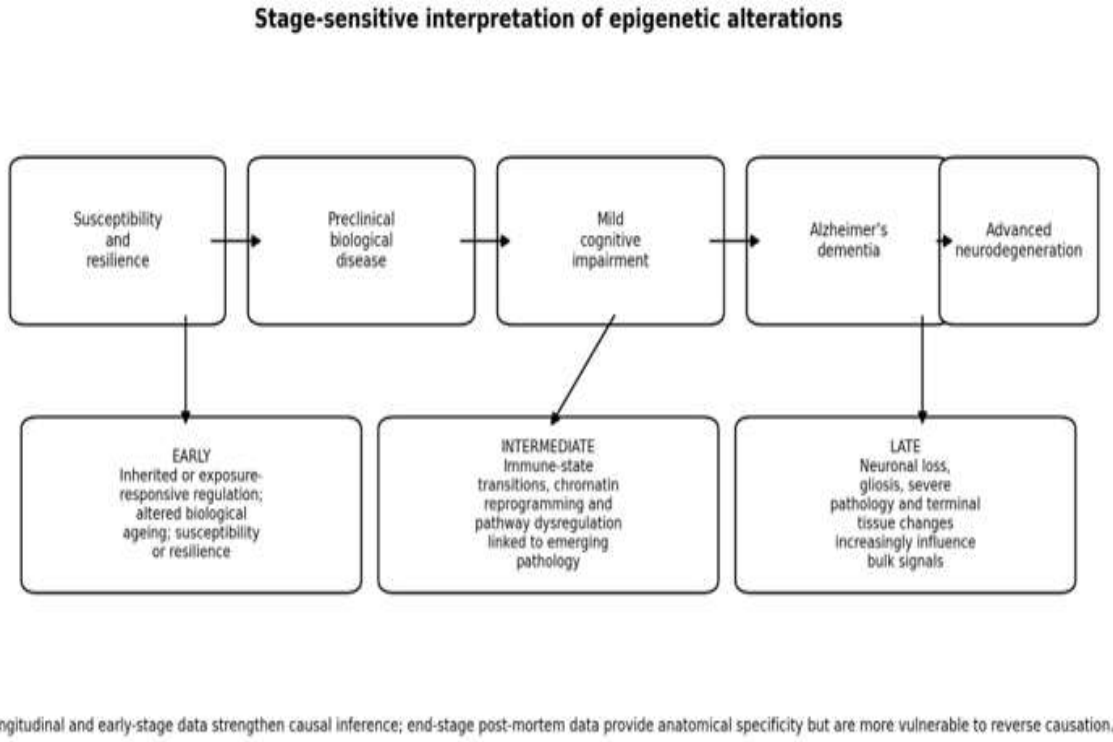


Figure 3.3: Stage-sensitive interpretation of epigenetic alterations across the Alzheimer's disease continuum.

IV. METHODOLOGY

Introduction

This chapter describes the methodological framework for investigating epigenetic alterations associated with Alzheimer's disease (AD) progression. The methodology is designed to answer the seven research questions established in Chapter One while preserving a clear distinction between association, mechanistic interpretation, prediction and causal inference. The primary empirical layer is genome-wide DNA methylation, complemented by transcriptomic, chromatin-accessibility, histone-modification, genetic and non-coding RNA information when appropriately matched data are available. The study is not designed to treat every molecular measurement as equivalent; rather, it uses each data type for a prespecified analytical purpose.

The study will use de-identified human data from established observational cohorts and repositories. Its central design is a retrospective, multi-cohort, discovery-validation analysis. One cohort will be designated before outcome modelling as the primary discovery dataset, and at least one non-overlapping cohort will be reserved for independent replication. Datasets will be selected according to phenotype depth, molecular quality, tissue relevance and the availability of covariates needed to control major sources of bias. The Religious Orders Study and Rush Memory and Aging Project (ROSMAP) and the Mount Sinai Brain Bank (MSBB) are prioritised because they contain extensively characterised post-mortem brain material and disease-relevant molecular data; additional eligible AMP-AD or public cohorts may be incorporated according to the access approvals and data audit completed before analysis.

The primary outcomes represent progression rather than diagnosis alone. They include neuropathological stage, amyloid and tau burden, longitudinal cognitive decline, clinical severity and, where available, conversion from cognitively unimpaired status or mild cognitive impairment to dementia. The unit of analysis will normally be the individual. For datasets

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

containing multiple brain regions per individual, the unit will be the participant-region observation, with within-person dependence modelled through random effects or cluster-robust standard errors.

A protocol-first approach will be used. Cohort roles, outcomes, covariates, quality-control thresholds, model families, primary contrasts and validation rules will be documented before the main hypothesis tests. Analytical flexibility is unavoidable in high-dimensional omics research, but it will be constrained through prespecified decision rules, blinded or outcome-independent quality control, correction for multiple testing, complete reporting of sensitivity analyses and external validation. This structure is consistent with recommendations for interpretable epigenome-wide association studies and reproducible disease-omics research (Birney et al., 2016; Campagna et al., 2021).

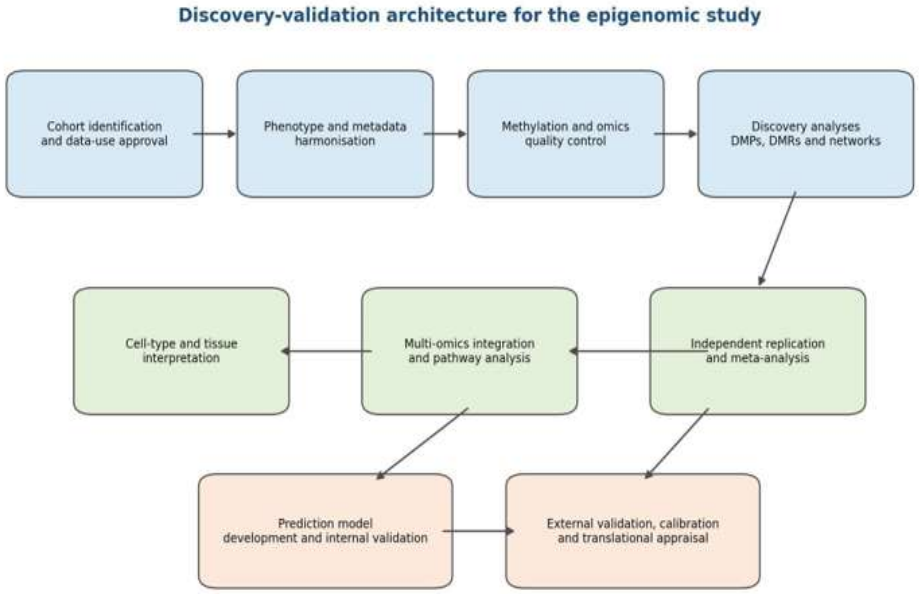


Figure 4.1: Overall discovery-validation architecture of the study.

Table 4.1: Summary of the methodological design.

Design element	Prespecified approach	Purpose
Research design	Retrospective observational secondary analysis of multiple human cohorts	To analyse existing clinically and neuropathologically characterised molecular data without misrepresenting archival observations as newly collected data.
Primary molecular layer	Genome-wide DNA methylation measured in brain; blood used for cross-tissue analyses	To identify DMPs, DMRs, co-methylation modules and epigenetic-age measures associated with progression.
Complementary layers	RNA-seq, ATAC-seq, histone marks, genotype and non-coding RNA where available	To evaluate regulatory coherence, cell-type context and plausible biological pathways.

Primary outcomes	Braak stage, amyloid burden and longitudinal cognitive decline	To focus on progression-related biology rather than case-control status alone.
Validation	Independent cohort replication followed by cross-cohort meta-analysis	To distinguish reproducible findings from cohort-specific signals.
Prediction	Nested resampling in discovery data and locked external validation	To estimate incremental performance without feature-selection leakage.
Inference boundary	Association and prediction; causal language restricted to analyses with explicit assumptions	To avoid interpreting cross-sectional molecular associations as proof of causation.

Research Philosophy or Paradigm

The study is guided by a pragmatic critical-realist paradigm. Critical realism assumes that AD progression reflects biological processes that exist independently of any particular measurement platform, but that the investigator observes those processes through imperfect proxies. A methylation beta value, a Braak stage, an RNA count and a cognitive score are therefore treated as measurements of underlying states rather than the states themselves. This perspective is appropriate for neurodegeneration research because tissue availability, cellular heterogeneity, post-mortem change and platform-specific measurement error prevent direct observation of many causal processes.

Pragmatism informs the selection of methods. No single assay or statistical model is assumed to provide a complete account of disease progression. Methods will be chosen according to the research question, data structure and interpretive objective. Site-level methylation models are appropriate for discovery, region- and network-level analyses for coordinated regulation, cell-type methods for biological attribution, multi-omics integration for mechanistic coherence and prediction models for incremental prognostic value. Convergent evidence across these levels will be considered stronger than a single statistically significant result.

The paradigm also requires explicit separation of explanatory claims. Association will be established by adjusted statistical models. Temporal relevance will be strengthened by longitudinal cognitive outcomes, prospective clinical data or molecular features measured before the outcome. Mechanistic plausibility will be supported by regulatory overlap, expression relationships, genetic context and cell-type specificity. Causality will not be asserted unless the data and assumptions justify an analysis such as instrumental-variable modelling or longitudinal mediation. This hierarchy prevents biological plausibility from being substituted for causal proof.

Research Design

The investigation will employ a multi-stage analytical design comprising cohort selection, harmonisation, discovery, biological interpretation, independent validation and translational evaluation. The discovery stage will identify epigenetic positions, regions and networks associated with progression outcomes. The interpretation stage will examine whether the signals are related to gene expression, chromatin accessibility, histone modification, genetic variation, molecular pathways and specific brain cell types. The validation stage will test the direction and magnitude of priority signals in an independent cohort. The final stage will evaluate whether a parsimonious epigenetic signature adds predictive value to established demographic, genetic and clinical variables.

The study is observational. Exposure allocation is not manipulated, and the epigenetic measurements may reflect susceptibility, mediation, compensation or consequences of disease. Cross-sectional post-mortem analyses will therefore be complemented by longitudinal cognition, sensitivity analyses for

disease severity and, where possible, temporally earlier peripheral measurements. The design will also examine negative-control outcomes or regions when suitable data exist, because a signal that is equally strong in unrelated outcomes or relatively spared tissue may indicate technical or systemic confounding.

The discovery and validation roles of cohorts will not be exchanged after results are known. A cohort may be excluded or reclassified only for documented reasons such as inadequate raw-data quality, missing primary outcomes, substantial participant overlap or failure of harmonisation. Any such change will be recorded before the main analysis of the replacement cohort. This rule is intended to prevent selective designation of the best-performing cohort as the validation dataset.

Study Setting

The study setting is a secure computational research environment rather than a single hospital or laboratory. Source data originate from longitudinal ageing studies, memory-clinic cohorts, brain banks and molecular consortia. The planned analyses will be conducted on approved institutional servers, controlled-access cloud workspaces or encrypted local systems that meet the requirements of the relevant data-use agreements. Raw molecular files and identifiable linkage variables will not be transferred to personal devices.

Priority will be given to data available through established resources such as the AD Knowledge Portal, Gene Expression Omnibus and other repositories specified by the source studies. The analytic environment will contain separate directories for immutable raw data, harmonised metadata, intermediate objects, final analysis datasets, code, logs and outputs. Access will be role-based, and the source terms of use will determine whether controlled data may be combined, exported or analysed only within a designated workspace.

Because cohorts differ in geography, recruitment model and tissue source, study setting will be treated as a source of heterogeneity rather than ignored. Cohort, brain bank, region and processing centre will be represented in descriptive summaries and statistical models. Pooled analysis will be performed only after verifying that phenotype definitions and molecular features are sufficiently comparable; otherwise, cohort-specific analyses will be combined through meta-analysis.

Table 4.2: Planned cohort roles and data domains.

Cohort or data source	Planned role	Relevant data domains	Important qualification
ROSMAP or equivalent deeply phenotyped ageing cohort	Primary discovery	Longitudinal cognition, neuropathology, DNA methylation, genotype and transcriptomics	Use is conditional on data-use approval and availability of analysis-ready or raw files.
MSBB or equivalent independent brain-bank cohort	External replication and regional heterogeneity	Multiple cortical regions, neuropathology, methylation and transcriptomics	Participant overlap with the discovery cohort must be excluded.
Single-nucleus reference datasets	Cell-type attribution and regulatory interpretation	snRNA-seq, snATAC-seq or multiome profiles	Reference datasets support interpretation and are not treated as equivalent to the bulk discovery cohort.
Cohort or data source	Planned role	Relevant data domains	Important qualification

Peripheral blood cohorts or matched blood-brain datasets	Cross-tissue and biomarker evaluation	Blood methylation, clinical phenotypes and selected cognition or imaging measures	Brain-blood discordance will be reported rather than treated as replication failure by default.
Public functional annotations	Regulatory mapping	ENCODE, Roadmap, GENCODE, gene sets and QTL resources	Annotation versions and genome builds will be frozen before final analysis.

Study Population

The target population comprises adults represented in human ageing and AD cohorts with adequate clinical, cognitive or neuropathological characterisation and at least one relevant epigenomic measurement. Analytical groups may include cognitively unimpaired participants, individuals with mild cognitive impairment, persons with clinically diagnosed AD dementia and autopsied individuals across a range of neuropathological stages. Participants with mixed pathologies will not automatically be excluded because co-pathology is common in older adults; instead, major co-pathologies will be adjusted for, stratified or examined in sensitivity analyses when reliable measures are available.

For brain analyses, the preferred tissues are cortical regions affected at different points in the AD continuum, including the entorhinal, temporal and prefrontal cortices. A relatively less affected comparison region may be included when available. The study population is not restricted to a single ancestry, but the ancestry distribution of each cohort will be documented. Analyses will account for genetic ancestry and, where sample size permits, evaluate heterogeneity across ancestry groups rather than assume universal transferability.

The study population for prediction analyses may be narrower than that used for association testing. Prediction requires variables available at a clinically meaningful time point and in a tissue that can realistically be sampled. Consequently, brain-derived features may be used to identify mechanisms, while peripheral features will be prioritised for biomarker models intended for use during life.

Inclusion Criteria

Participants or participant-region observations will be included when they satisfy all of the following criteria: adult age according to the source cohort; valid linkage among molecular data, phenotype and core covariates; a prespecified progression outcome; molecular measurements that pass platform-specific quality control; and permission for the intended analyses under the data-use agreement. For post-mortem tissue, the record must contain the sampled brain region and sufficient information to evaluate post-mortem or tissue-quality effects.

For the primary DNA methylation analysis, raw IDAT files are preferred because they permit consistent preprocessing. Cohorts providing only processed beta values may be included in replication or meta-analysis if preprocessing is adequately documented and the probe identifiers can be mapped reliably. For multi-omics integration, samples must have verified participant identity across data modalities; unmatched datasets may be used for annotation or network overlap but not for sample-level mediation or latent-factor analysis.

A participant may contribute to more than one analysis when the required data are available. The denominators will therefore be reported separately for each objective, and a flow diagram will show the number of samples at every stage. This avoids implying that all analyses were conducted in one identical complete-case sample.

Exclusion Criteria

Samples will be excluded for unresolved identity mismatch, duplicate or cryptically related records not permitted by the model, discordance between reported and methylation-predicted sex, severe contamination, very low call rate, extreme technical

outlier status, missing primary outcome, unavailable brain-region information or failure of source-cohort quality standards. Samples will not be excluded merely because they have advanced disease, uncommon pathology or unusual molecular values; exclusion must be based on prespecified technical or eligibility criteria rather than outcome extremity.

Probes will be excluded when they fail detection or bead-count thresholds, map ambiguously, are known to be cross-reactive, overlap common variants that materially affect hybridisation, or cannot be harmonised across the platforms used in a pooled analysis. Sex-chromosome probes will be excluded from the principal autosomal analysis and examined separately where scientifically justified, because sex-chromosome dosage and inactivation require specialised models.

For sequencing data, libraries will be excluded when they fail source-specific or study-level criteria for alignment, duplication, complexity, contamination or assay signal. Borderline libraries will be retained only when sensitivity analyses demonstrate that they do not drive the findings. Every exclusion will be recorded with a machine-readable reason code.

Table 4.3: Inclusion and exclusion criteria.

Domain	Inclusion requirement	Exclusion trigger
Participant	Adult, valid phenotype, permitted data use	Missing primary outcome, unresolved duplicate, prohibited use
Identity	Concordant participant identifiers across modalities	Genotype, methylation or sex mismatch that cannot be resolved
Brain tissue	Known region and adequate processing metadata	Unknown region, severe contamination or source-declared failure
Methylation	Raw or well-documented processed data passing QC	Low call rate, detection failure, extreme technical outlier
RNA or chromatin	Library passes assay-specific QC	Unacceptable mapping, complexity, contamination or signal
Multi-omics	Verified same participant and compatible tissue	Uncertain sample linkage or incompatible biological material
Validation	Independent participants and comparable outcome	Participant overlap or material phenotype incompatibility

Sampling Strategy

A census sampling strategy will be used within each eligible cohort: all records satisfying the criteria will be included rather than drawing a convenience subsample. This approach maximises power for high-dimensional association testing and reduces avoidable selection. Discovery and validation cohorts will be selected purposively at the cohort level on the basis of phenotype depth, tissue relevance and molecular availability, but participant inclusion within a chosen cohort will be rule-based.

When several brain regions are available, region selection will follow a prespecified hierarchy. The primary region will be the region with the strongest combination of biological relevance, sample size and phenotype quality. Additional regions will be analysed for regional consistency using stratified or mixed models. A participant will not be counted as independent multiple times in a pooled region analysis; correlation will be modelled explicitly.

The sampling strategy also preserves an independent validation reserve. Candidate loci and models will be selected exclusively in the discovery data. The validation cohort will not be used to tune quality thresholds, choose covariates or optimise feature sets, except for verifying that a locked pipeline can be executed. This separation is essential because high-dimensional feature selection can produce optimistic

estimates when the same data influence both selection and evaluation.

Sample-Size Determination

The final sample size is constrained by eligible data already collected by the source cohorts. Consequently, sample-size determination will combine a census of available observations with simulation-based power analysis. Before the main analysis, the number of eligible participants, outcome distribution, methylation variance, covariate structure and effective number of tests will be estimated without examining feature-outcome associations. These quantities will be used to simulate detectable methylation effects under the planned models.

The study will target, where feasible, at least 500 independent participants in the discovery analysis and at least 250 in external validation. These are design targets rather than fabricated enrolment totals. If a cohort has fewer observations, the analysis may still proceed for secondary or regional objectives, but its power and interpretation will be qualified. For repeated cognitive measures, power depends on the number and timing of visits as well as the number of participants; simulations will therefore reproduce the observed visit distribution and within-person correlation.

For epigenome-wide testing, power will be assessed at a false-discovery-rate target of 5%, with sensitivity scenarios for absolute methylation differences of approximately 0.02 to 0.05 and realistic residual variance. Region- and network-level analyses can have greater power than individual-site tests because they aggregate correlated features, but their error rates will be controlled independently. Prediction models will obey an events-per-parameter principle after penalisation and will be restricted to compact feature sets. If the effective number of outcome events is insufficient, the predictive objective will be reported as exploratory rather than definitive.

No post hoc power calculation based on observed p-values will be used. Precision will be described through confidence intervals, and null findings will be interpreted in relation to the minimum effect size that the available data could reasonably detect.

Table 4.4: Sample-size and power framework.

Analysis	Primary determinant of power	Planned safeguard
DMP analysis	Number of participants, residual methylation variance, covariates and multiplicity	Simulation using observed design matrix; FDR control; report minimum detectable effects.
DMR and network analysis	Probe density, spatial correlation and module size	Independent error control and permutation or resampling where required.
Longitudinal cognition	Number of participants, visits, follow-up time and within-person correlation	Linear mixed-model simulation reflecting actual visit patterns.
Multi-omics integration	Number of matched samples and dimensionality of each layer	Feature preselection based on discovery criteria; latent-factor regularisation.
Prediction	Outcome events, predictor count and cohort shift	Penalisation, nested resampling, compact locked model and external validation.
Subgroup analysis	Size and balance of sex, ancestry, APOE or tissue strata	Predefined minimum stratum size; interaction estimates with confidence intervals.

Participant Recruitment

No new participants will be recruited for the current secondary-data study. Recruitment, consent, clinical follow-up, specimen donation and autopsy were performed by the source cohorts under their own approved protocols. The thesis will describe those procedures accurately from cohort documentation but will not claim them as procedures performed by the doctoral researcher.

The researcher will obtain data through formal application, institutional authorisation and acceptance of the relevant data-use terms. Where a repository requires a research-use statement, the submitted purpose will specify AD epigenomics, disease progression, multi-omics integration and validation. The application will not request identifying data that are unnecessary for the planned analyses.

Because consent scope can differ across cohorts, each dataset will be reviewed to confirm that genomic analysis, cross-cohort integration and publication of aggregate results are permitted. Cohorts with restrictions that prevent the intended analysis will be excluded or analysed separately in accordance with their terms.

Clinical Assessment

Clinical diagnosis and staging will be harmonised to the greatest extent supported by the source data. The principal categories will be cognitively unimpaired, mild cognitive impairment and dementia due to AD, with additional categories for other or mixed dementia where available. Contemporary biological staging criteria will inform interpretation, but retrospective cohorts will not be reclassified beyond the biomarkers they actually measured (Jack et al., 2024).

Neuropathological variables will include Braak neurofibrillary-tangle stage, semiquantitative or quantitative amyloid measures, CERAD neuritic-plaque score, and source-defined composite AD neuropathology. Braak stage will be analysed as an ordered or approximately continuous outcome in primary models, with ordinal models used as a sensitivity analysis. Amyloid and tau measures will be kept separate when possible because they represent related but non-identical processes.

Clinical severity variables may include the Clinical Dementia Rating, Mini-Mental State Examination, global cognitive composites, functional measures and source-specific consensus diagnosis. Harmonisation will preserve the original scale whenever a valid common metric can be constructed. If scales cannot be transformed without loss of meaning, cohort-specific standardised effects will be meta-analysed instead of pooling raw values.

Cognitive Assessment

Longitudinal cognitive decline is a primary progression outcome. The preferred measure is a psychometrically derived global cognitive composite based on repeated neuropsychological testing. Domain-specific composites for episodic memory, semantic memory, working memory, perceptual speed and visuospatial ability will be examined secondarily where they have adequate reliability.

For each participant, cognitive trajectories will be estimated using linear mixed-effects models containing time from baseline, age, sex, education and other cohort-relevant terms. Random intercepts and, when supported, random slopes will account for individual differences in baseline performance and rate of change. The participant-specific slope or the methylation-by-time interaction will represent progression. Non-linear trajectories will be evaluated with restricted cubic splines or piecewise time terms if graphical diagnostics indicate substantial deviation from linearity.

Practice effects, terminal decline and informative dropout can bias cognitive trajectories. Sensitivity analyses will exclude the first assessment, model time to death, restrict to participants with at least two or three observations, and compare joint or pattern-mixture approaches when feasible. Molecular associations with cognition will be interpreted in relation to follow-up duration and the temporal position of specimen collection.

Collection of Demographic and Clinical Data

A harmonised metadata dictionary will be created before analysis. Core variables will include age at specimen collection or death, sex, self-reported race or ethnicity where permitted, genetically inferred ancestry, education, APOE genotype, clinical diagnosis, cognitive measures, medication indicators, vascular and metabolic conditions, smoking, alcohol use and source-specific technical variables. Post-mortem variables will include interval, brain region, tissue pH, storage duration and cause or circumstances of death when available.

Each variable will be mapped from the original name and coding to a common definition. Transformations will be documented in a data dictionary that records source, units, valid range, missing-value codes and derivation logic. Harmonisation will not create false comparability: variables with materially different definitions will be retained as cohort-specific and incorporated through stratified analysis or meta-regression.

Outlier and range checks will be conducted without reference to molecular associations. Implausible values will be resolved against source documentation where possible; otherwise, they will be set to missing with a reason code. Derived outcomes and covariates will be generated by version-controlled scripts rather than manual spreadsheet editing.

Biological Sample Collection

Biological samples were collected by the source cohorts. The present study will not collect new blood, cerebrospinal fluid or brain tissue. For post-mortem brain, the source documentation will be reviewed for autopsy procedures, dissection method, hemisphere, brain region, freezing or fixation interval and storage conditions. For blood, collection tube, cellular fraction and processing delay will be recorded when available.

Sample provenance will be represented explicitly because tissue source can influence epigenetic measurements. Brain region, grey-matter enrichment, bulk tissue versus sorted nuclei, and blood-cell fraction will not be merged into one undifferentiated variable. Matched samples from the same participant will be identified so that paired or mixed models can exploit within-person comparisons.

Where biobank documentation is insufficient to establish tissue identity or processing, the affected data will be limited to analyses robust to that uncertainty or excluded. Any future acquisition of physical specimens for targeted validation would require a separate protocol amendment, laboratory safety plan and ethical approval and is outside the current thesis methodology.

Sample Processing and Storage

The study will audit source-cohort sample processing rather than reproduce it. Processing variables with potential influence on methylation or sequencing include time to freezing, freeze-thaw history, tissue dissection, extraction batch, plate, array position, library-preparation batch and sequencing lane. These variables will be extracted from technical manifests whenever available.

Raw data will be preserved in an immutable directory. Working copies and processed objects will be generated by scripted workflows. Checksums will be used to confirm file integrity. Sample identifiers will be pseudonymised in analytical files, and the link to repository identifiers will be stored separately with restricted access.

Storage-related variables will be included as covariates or evaluated through sensitivity analyses when they show plausible associations with principal components or quality metrics. Batch correction will not be used to conceal poor processing quality; samples that fail biological or technical acceptance criteria will be excluded before correction.

DNA Extraction

DNA extraction was undertaken by the source laboratories using their validated protocols. The current study will record extraction method, tissue input, extraction batch and available purity or concentration measures. Studies using different extraction methods may be combined only after examining whether method is confounded with outcome or cohort.

For archival brain material, degradation is generally less consequential for methylation arrays than for RNA sequencing, but contamination and bisulphite-conversion efficiency remain important. Available genotype concordance, methylation control probes and sex-prediction results will be used to detect sample swaps or poor conversion. The researcher will not infer extraction quality solely from downstream association strength.

RNA Extraction

RNA extraction and library preparation were performed by the source projects. RNA integrity number, concentration, ribosomal-RNA depletion or poly(A) selection, strandedness and sequencing protocol will be captured in the metadata. Because post-mortem brain RNA quality is variable, RNA integrity will be treated both as a quality indicator and a model covariate.

Samples that fail source-cohort RNA-seq quality criteria or exhibit extreme mapping, 3-prime bias, contamination or library-complexity problems will be excluded from expression analyses. However, a rigid RNA integrity cut-off will not be applied across all cohorts without considering library protocol and source-study validation. This prevents unnecessary loss of informative post-mortem samples while still controlling degradation-related bias.

Nucleic Acid Quality Assessment

Nucleic acid quality will be assessed through available laboratory metrics and platform-derived diagnostics. For DNA, concentration, purity ratios, conversion controls, call rate and genotype or sex concordance will be reviewed. For RNA, integrity, mapping, insert size, gene-body coverage, ribosomal content, strandedness and duplication will be examined. Quality metrics will be visualised by cohort, plate, batch, diagnosis and tissue to identify confounding.

Quality-control decisions will be made before the primary outcome models. Thresholds are described as targets because historical datasets may have validated study-specific criteria. Departures from a target will trigger review rather than automatic exclusion when the metric is known to vary by protocol. Final decisions and justifications will be stored in a frozen quality-control report.

DNA Methylation Analysis

DNA methylation is the primary analytical layer. Illumina 450K and EPIC array data will be processed from raw IDAT files with the Bioconductor minfi framework where available (Aryee et al., 2014; Fortin et al., 2017). Background correction and dye-bias adjustment will use single-sample Noob, which supports incremental processing and integration across array generations. Functional normalisation will be evaluated as a sensitivity procedure when control-probe variation indicates substantial technical structure.

Sample-level quality control will examine median methylated and unmethylated intensity, detection p-values, call rate, control-probe performance, sex prediction, genotype concordance, multidimensional scaling and robust principal components. A sample will normally be required to have valid measurements at at least 95% of retained probes, with a more stringent 98% threshold evaluated in sensitivity analysis. Extreme outliers will be defined through multiple metrics rather than a single visual judgement.

Probe filtering will remove probes with poor detection performance, inadequate bead count, non-specific or cross-reactive mapping, common variants affecting the target CpG or single-base extension, and probes that cannot be harmonised across array types (Chen et al., 2013; Zhou et al., 2017). Beta values will be retained for interpretation because they represent the proportion of methylated signal, while M values will be used for inferential models because their statistical properties are more suitable at the extremes.

Primary analyses will be performed within cohort after consistent preprocessing. Cross-platform pooling will use the common high-quality probe set and include array type or cohort in the design. An alternative meta-analysis strategy will be used when residual platform heterogeneity is substantial. All preprocessing parameters, software versions and probe-exclusion lists will be archived.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

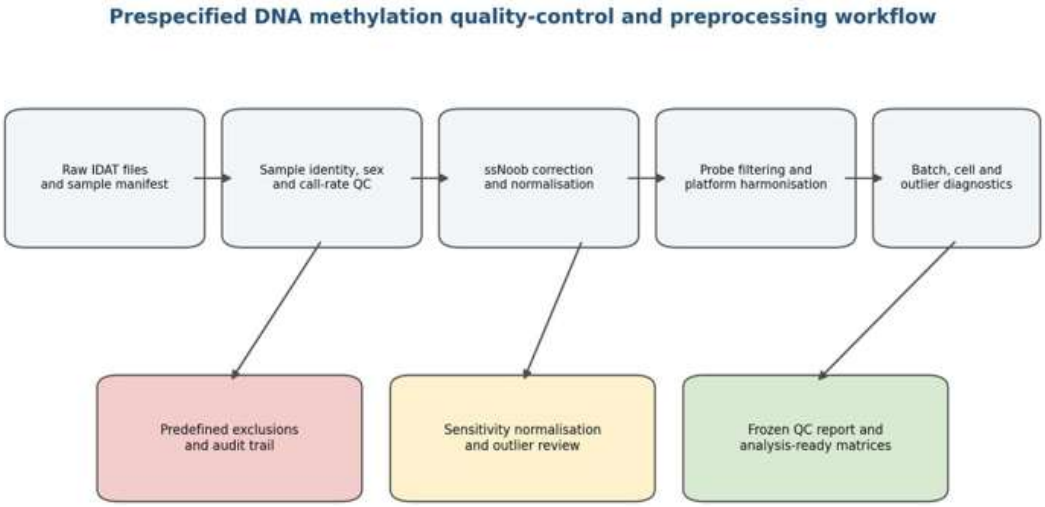


Figure 4.2: DNA methylation quality-control and preprocessing workflow.

Histone-modification data will be incorporated where suitable ChIP-seq, CUT&RUN or CUT&Tag profiles are available. Marks will be interpreted according to their biology; for example, H3K27ac and H3K9ac will represent active regulatory regions, whereas broad repressive marks require domain-oriented analysis. Target class, antibody validation, input or IgG control, spike-in strategy and replicate structure will be documented.

Raw sequence processing will preferentially use reproducible public workflows such as nf-core/chipseq or nf-core/cutandrun. Quality assessment will include alignment, duplicate rate, library complexity, cross-correlation or fragment-size measures, fraction of reads in peaks, blacklist overlap and replicate concordance. Peak calling will use an assay-appropriate narrow or broad mode, with controls preserved through the analysis.

Histone data will not be analysed as an independent epigenome-wide discovery layer unless sample size is adequate. Its principal role is to evaluate whether methylation signals overlap disease-associated active or repressive regulatory regions and whether those overlaps show consistent direction with gene expression and neuropathology. The tau-associated histone-acetylation findings reported by Klein et al. (2019) and the integrated study of Nativio et al. (2020) provide methodological precedents for this interpretation.

Chromatin Accessibility Analysis

Chromatin accessibility will be analysed using ATAC-seq or single-nucleus ATAC-seq data. Bulk ATAC-seq will follow a reproducible workflow such as nf-core/atacseq, including adapter trimming, alignment, duplicate handling, mitochondrial-read assessment, Tn5 shifting, blacklist filtering, peak calling and consensus-peak construction. Key quality measures will include transcription-start-site enrichment, fraction of reads in peaks, nucleosomal periodicity, alignment and replicate concordance (Buenrostro et al., 2013; Corces et al., 2017).

Single-nucleus data will undergo cell-level filtering based on unique fragments, TSS enrichment, nucleosome signal, doublet score and fraction of reads in peaks. Cells will be clustered and annotated using gene-activity scores, reference labels and marker accessibility. Pseudobulk peak counts will be generated by participant and cell type for differential testing, thereby preserving the participant as the unit of inference.

Accessibility results will be linked to methylation through genomic overlap, distance-aware enhancer-gene maps and correlation in matched samples. An inverse relation between methylation and accessibility is biologically plausible at many regulatory elements but will not be imposed as a universal rule. Cell-type-specific accessibility data from studies such as Morabito et al. (2021) will be

used to prioritise which neuronal or glial populations plausibly generate bulk-tissue signals.

Non-Coding RNA Analysis

Non-coding RNA analyses will include microRNAs, long non-coding RNAs and circular RNAs only where data quality, annotation and sample size are adequate. Small-RNA sequencing will be processed with adapter-aware trimming, mapping to a fixed reference and mature microRNA quantification. Long non-coding RNA expression will be derived from stranded total-RNA sequencing where possible, because poly(A)-selected data may incompletely capture non-polyadenylated transcripts.

Low-abundance features will be filtered before testing. Expression will be normalised with methods appropriate to count data, and differential models will include the same biological and technical covariates used for mRNA where applicable. Candidate microRNA-target relationships will require support from validated interaction databases or convergent prediction, inverse expression and pathway coherence.

Non-coding RNA will be treated primarily as a regulatory and biomarker layer rather than merged indiscriminately with DNA methylation. A signal will be prioritised when a non-coding RNA is associated with progression, is linked to a methylation or chromatin feature, and targets genes or pathways relevant to amyloid, tau, immune or synaptic processes.

APOE and Genetic Analysis

APOE genotype will be determined from source-provided genotypes or imputed dosage at the defining epsilon alleles. Participants will be classified as epsilon4 carriers or non-carriers for the primary analysis, with genotype categories examined secondarily when numbers permit. APOE will be included as a covariate and tested as a moderator of selected epigenetic associations.

Genome-wide genotype data will undergo source-specific quality control for call rate, heterozygosity, relatedness, ancestry and imputation quality. Genetic principal components will control population structure. Common variants at or near priority methylation loci will be examined as methylation quantitative-trait loci to distinguish genetically influenced methylation from environmentally or disease-associated variation.

Polygenic risk scores may be calculated using externally derived AD weights and ancestry-appropriate linkage-disequilibrium reference panels. They will not be used to select methylation features in the primary association analysis. Gene-by-epigenome interactions will be restricted to prespecified tests because interaction models require substantially larger samples than main-effect models.

Next-Generation Sequencing Procedures

The study will reanalyse existing sequencing data rather than generate new libraries. Sequencing procedures will therefore be reconstructed from source methods and technical manifests. The reference genome, read length, paired-end status, strandedness, library selection, sequencing platform and batch will be recorded for every dataset.

Reprocessing will be performed only when raw reads are available and permitted. When only processed matrices are available, the study will evaluate source-pipeline documentation, quality metrics and annotation version before use. Raw and processed data from different pipelines will not be pooled without assessing whether processing differences create dominant principal components or inconsistent feature definitions.

Public, containerised workflows will be preferred for reproducibility. Nextflow-based nf-core pipelines provide versioned, community-reviewed implementations for RNA-seq, ATAC-seq and ChIP-related assays (Ewels et al., 2020). A pipeline version, container digest, parameter file and reference manifest will be retained for every execution.

Bioinformatics Pipeline

The bioinformatics pipeline will be organised into six layers: data intake; identity and metadata validation; assay-specific quality control; feature quantification; statistical modelling; and integration or validation. Each layer will produce machine-readable outputs

and a concise human-readable report. Failures will stop the workflow rather than silently pass to downstream analysis.

A common sample manifest will link participant, tissue, region, assay, batch, file path, outcome availability and analysis role. The manifest will be the authoritative source for inclusion and will be validated before every pipeline run. Genome coordinates will be harmonised to GRCh38 where possible; legacy coordinates will be converted with a documented chain file, and features that cannot be mapped uniquely will be excluded.

R will be used for methylation, expression, statistical modelling and visualisation; Python may be used for workflow orchestration, machine learning or data transformation. Nextflow and containers will manage sequence-processing tools. Package environments will be frozen with renv, conda or equivalent environment files. The final pipeline will be executable from configuration files rather than manual command histories.

Table 4.5: Core bioinformatics software and intended use.

Domain	Preferred software or framework	Intended use
Methylation preprocessing	minfi; ChAMP for sensitivity or comparison	IDAT import, ssNoob, QC, normalisation and cell estimates.
Differential methylation	limma; DMRcate; bumphunter or comb-p sensitivity	DMP and DMR estimation with empirical-Bayes inference.
RNA-seq	nf-core/rnaseq; edgeR/limma-voom or DESeq2	Read processing, count normalisation and differential expression.
ATAC-seq	nf-core/atacseq; MACS2; DiffBind/edgeR	Accessibility QC, peaks, consensus matrix and differential analysis.
ChIP/CUT&RUN	nf-core/chipseq or nf-core/cutandrun	Control-aware peak calling, signal tracks and differential binding.
Networks and pathways	WGCNA; clusterProfiler; fgsea	Co-regulation modules, ontology and pathway enrichment.
Multi-omics	MOFA+; mixOmics/DIABLO; custom cis integration	Latent factors, supervised integration and regulatory links.
Prediction	glmnet; tidymodels or scikit-learn	Penalised models, nested resampling and performance estimation.
Workflow and provenance	Nextflow; Docker/Apptainer; Git; renv	Reproducible execution, dependency locking and audit trail.

Sequence Quality Control

Sequence quality control will be performed at read, library and cohort levels. Read-level checks will include base quality, adapter content, sequence duplication, overrepresented sequences and read length. Library-level checks will include alignment, insert size, complexity, contamination, strandedness, mitochondrial or ribosomal fraction and assay-specific enrichment. Cohort-level diagnostics will examine whether quality metrics cluster by diagnosis, outcome, tissue or batch.

For ATAC-seq, target indicators include clear nucleosomal periodicity, adequate TSS enrichment,

an acceptable fraction of reads in peaks and reproducible peak profiles. For ChIP-related assays, enrichment, controls and replicate concordance are central. For RNA-seq, mapped reads, gene detection, 3-prime bias and sample correlations will be reviewed. Thresholds will be assay- and source-specific, with deviations documented.

Quality metrics will not be corrected after outcome modelling to improve significance. The QC report will be frozen before primary association tests. Samples failing mandatory criteria will be excluded, and borderline samples will be evaluated in a sensitivity analysis.

Read Alignment and Mapping

Sequencing reads will be aligned to a fixed human genome build using assay-appropriate aligners. RNA-seq will use a splice-aware aligner such as STAR or pseudoalignment when source compatibility requires it. ATAC-seq and ChIP-related reads will use an aligner such as Bowtie2, with parameters selected for fragment length and paired-end structure. Mitochondrial contigs, alternate haplotypes and decoy sequences will be handled consistently across samples.

Gene annotation will use a frozen GENCODE release. Peak coordinates, methylation probes and variants will be mapped to the same genome build before integration. Ambiguous features, multi-mapping reads and low-quality alignments will be filtered according to the assay protocol. Duplicate treatment will distinguish technical PCR duplication from biologically meaningful high-occupancy signal, particularly for transcription-factor and histone assays.

Alignment summaries and software logs will be retained. Re-alignment will not be performed with different parameters after observing biological results unless a documented technical error is identified; any revised analysis will preserve the original output and explanation.

Normalisation Procedures

Normalisation will be conducted separately for each assay and within coherent processing batches. Methylation arrays will use ssNoob as the primary method, with functional normalisation examined in sensitivity analysis. Beta values will not be quantile-normalised across biologically distinct tissues without evidence that the transformation preserves meaningful differences.

RNA-seq counts will be filtered for low expression and normalised using trimmed mean of M-values followed by limma-voom, or DESeq2 size factors when the model and dataset are better suited to that framework (Law et al., 2014; Love et al., 2014). ATAC and ChIP peak-count matrices will use library-size and composition-aware normalisation, with offsets or methods that account for global signal changes when appropriate.

Normalisation quality will be evaluated through density plots, mean-variance relationships, principal components, sample correlations and preservation of known biological variables. A method will not be selected solely because it reduces batch variance; over-normalisation can remove disease-related signal when batch and phenotype are partially confounded.

Batch-Effect Correction

Batch effects will be controlled through design, modelling and, when justified, empirical correction. Plate, slide, array position, extraction batch, library batch, sequencing lane, processing centre and cohort will be examined. The preferred approach is to include measured technical variables in the statistical design while preserving biological variation.

ComBat or related empirical-Bayes correction may be applied to exploratory matrices or prediction inputs when batch is not perfectly confounded with outcome and when correction is learned only within the training data (Johnson et al., 2007). For hypothesis testing, direct covariate adjustment is preferred because it preserves uncertainty and reduces the risk of inadvertently removing phenotype signal. Surrogate-variable analysis may be used for unmeasured technical structure, but surrogate variables will be evaluated for correlation with the outcome and major biology (Leek et al., 2012).

Batch correction will be performed separately in discovery and validation cohorts. Parameters estimated in the discovery data will not use the external outcome data. The effect of correction will be assessed through principal components, variance partitioning and replicate concordance rather than visual improvement alone.

Cell-Type Deconvolution

Cellular composition is a major potential confounder because neuronal loss and glial activation are integral to AD progression. Bulk-brain methylation analyses will estimate neuronal and non-neuronal fractions using validated references, and more detailed neuronal, astrocytic, microglial and oligodendrocytic proportions will be used when supported by the platform and reference panel. Blood analyses will use established leukocyte reference-based estimates (Houseman et al., 2012).

Estimated proportions will be included as covariates in primary models. However, cell composition may also be a mediator of pathology, so a second model without full cell adjustment will quantify the total tissue-level association. The contrast between total and cell-adjusted effects will help distinguish composition-driven signals from within-cell regulatory changes. This two-model strategy is consistent with evidence that many cortical methylation associations in AD are driven by non-neuronal variation (Shireby et al., 2022).

Reference-free methods and surrogate variables will be used as sensitivity analyses when reference panels are incomplete. Cell-type-specific differential methylation may be estimated with interaction or deconvolution methods, but such results will be labelled inferential and validated against sorted-cell or single-nucleus data where possible. Cellular heterogeneity will not be treated solely as nuisance; it is also an informative component of disease biology (Jaffe & Irizarry, 2014).

Differential Methylation Analysis

Differentially methylated positions will be tested with linear models and empirical-Bayes moderation using limma (Ritchie et al., 2015). M values will be the dependent variable, and the progression phenotype will be represented as a continuous, ordinal or longitudinal term appropriate to the outcome. The primary model will adjust for age, sex, APOE, ancestry components, tissue or region, cell composition and prespecified technical covariates. Post-mortem interval, pH and co-pathology will be included where available and relevant.

For longitudinal cognitive outcomes, two approaches will be compared. The first will associate methylation with a pre-estimated participant-specific slope. The second will fit a mixed model with methylation-by-time interaction, which uses all observations and estimates whether methylation modifies the rate of decline. The second is preferred when computationally feasible and when visit-level data are available.

Differentially methylated regions will be identified with DMRcate as the primary method and an alternative spatial method as sensitivity analysis (Peters et al., 2015). A region will be prioritised when it contains multiple concordant probes, passes region-level error control and has a meaningful aggregate effect. Single-probe and region results will be reported separately.

The Benjamini-Hochberg false-discovery rate will control multiplicity at $q < .05$ for each prespecified analysis family. Bonferroni results and effect-size thresholds will be reported for transparency. Statistical significance will not substitute for biological relevance; absolute methylation difference, consistency across models, annotation and replication will determine priority.

Differential Gene-Expression Analysis

Gene-expression counts will be filtered to remove features with insufficient expression and normalised using an appropriate count-based framework. The primary analysis will use limma-voom with quality weights or DESeq2, depending on sample size, mean-variance diagnostics and source pipeline. Models will adjust for age, sex, diagnosis or pathology, cell composition, RNA integrity, post-mortem interval, sequencing batch and relevant ancestry variables.

Differential expression will be evaluated for progression outcomes and for association with priority methylation features. Cis relationships will be tested for genes within a prespecified genomic window or linked by regulatory annotation. Trans relationships will be explored at the module and pathway level to reduce multiplicity.

Expression-methylation relationships will be interpreted in context. Promoter hypermethylation may be associated with lower expression, but enhancer and gene-body effects can differ. A priority regulatory link will require consistent direction, statistical support, matched tissue or cell type and preferably replication or external functional annotation.

Differential Chromatin-Accessibility Analysis

Consensus peaks will be quantified for each sample or participant-cell-type pseudobulk. Differential accessibility will be tested with count-based models using edgeR, DESeq2 or limma-voom. The design will include progression outcome, age, sex, APOE, batch, brain region and cell-type proportion or cluster as appropriate.

Peak-level false-discovery rate will be controlled, and differentially accessible regions will be linked to promoters, enhancers, transcription-factor motifs and candidate target genes. Motif enrichment will be conducted against a matched background that accounts for GC content and peak accessibility. Transcription-factor activity will be inferred cautiously because motif presence does not establish binding.

Accessibility findings will be integrated with methylation by overlap and correlation. Priority loci will be those at which methylation, accessibility and expression support a coherent regulatory model. Discordant results will be retained and discussed because epigenetic relationships are context-dependent and can be affected by cell-type composition or timing.

Pathway and Enrichment Analysis

Pathway analysis will be performed on ranked statistics and on prioritised gene sets. Gene-set enrichment analysis will be preferred to arbitrary significance cut-offs because it uses the full ranked evidence (Subramanian et al., 2005). Over-representation analysis will be used secondarily for DMR-associated genes, modules and integrated candidate lists.

Methylation-array enrichment will account for unequal probe coverage per gene, thereby reducing bias toward genes represented by many probes. Gene Ontology, Reactome, KEGG and curated AD-related pathways will be analysed with a frozen annotation release. Redundant terms will be clustered or summarised to avoid presenting highly overlapping pathways as independent discoveries.

A pathway will be considered robust when it is supported by more than one feature level, reproduces in validation data or aligns with cell-type and regulatory evidence. Enrichment p-values will be corrected for multiple testing. Narrative interpretation will emphasise effect direction and the underlying genes rather than relying on pathway labels alone.

Network Analysis

Weighted gene co-expression and co-methylation network analysis will identify coordinated molecular modules (Langfelder & Horvath, 2008). Networks will be constructed from appropriately residualised or normalised data after removal of poor-quality and low-variance features. Soft-threshold selection, module stability and sample outliers will be documented.

Module eigengenes will be associated with progression outcomes and covariates. Modules will be annotated by pathway enrichment, cell-type markers, transcription factors and overlap with AD genetic loci. Hub features will be prioritised only when their intramodular connectivity and disease association are stable under resampling.

Consensus networks may be constructed across discovery and validation cohorts to identify preserved modules. A module that is associated with disease in one cohort but not preserved structurally in another will not be described as a universal disease network. Network results are intended to capture coordinated

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

biology and reduce dependence on isolated CpG-gene annotation.

Multi-Omics Integration

Multi-omics integration will proceed in tiers. Tier 1 is locus-based: methylation features will be linked to nearby or annotated genes, chromatin peaks, histone domains and genetic variants. Tier 2 is sample-level: matched methylation, expression and chromatin features will be correlated or modelled jointly. Tier 3 is network-level: modules, pathways and latent factors will be compared across omics.

MOFA+ may be used to identify latent factors explaining shared and modality-specific variation, while DIABLO or related sparse methods may evaluate supervised signatures when matched sample size is adequate (Argelaguet et al., 2020; Rohart et al., 2017). Integration will be performed only after each assay passes independent QC. Features will not be forced into a common model when sample matching or biological compatibility is insufficient.

Potential mediation by gene expression or chromatin accessibility will be explored only for a limited number of replicated loci with temporal or genetic support. Direct, indirect and total effects will be estimated with bootstrap confidence intervals, and assumptions of no unmeasured mediator-outcome confounding will be stated. Cross-sectional mediation will be described as statistical decomposition rather than proof of causal mechanism.

The integration output will be a ranked set of regulatory hypotheses, not merely a combined significance table. Priority will be assigned to findings supported by independent cohorts, coherent cell-type context, relevant pathways and evidence that the epigenetic feature explains variation beyond genetic and technical factors.

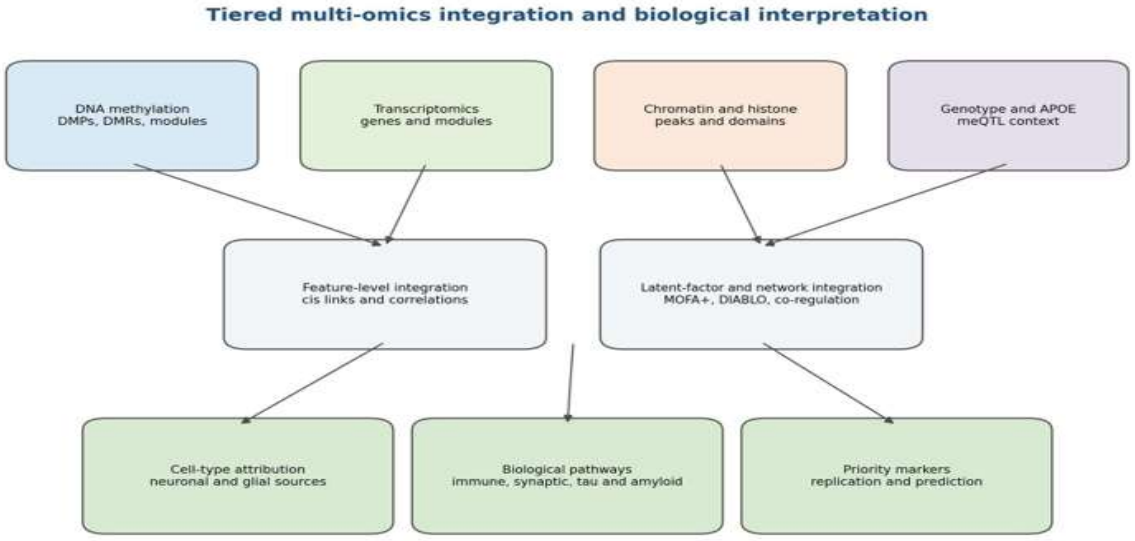


Figure 4.3: Tiered multi-omics integration and interpretation strategy.

Study Variables and Measurements

The principal independent variables are DNA methylation M values at individual CpGs, aggregated DMR scores, co-methylation module eigengenes and epigenetic-age acceleration measures. Secondary predictors include histone signal, chromatin accessibility, non-coding RNA abundance and integrated latent factors. The primary dependent variables are Braak stage, amyloid burden and longitudinal global cognitive decline.

Epigenetic age will be estimated with more than one clock because different algorithms capture different ageing constructs. Horvath age, PhenoAge and GrimAge or brain-appropriate alternatives will be calculated when the required CpGs are available

(Horvath, 2013; Levine et al., 2018; Lu et al., 2019). Age acceleration will be defined as the residual from regression of methylation age on chronological age within the relevant tissue and cohort. Clock outputs will not be compared across tissue without calibration.

Moderators include APOE epsilon4, sex, ancestry, brain region and cell type. Confounders include chronological age, education, vascular disease, post-mortem interval, tissue quality, technical batch and cellular composition. Mediators may include gene expression, chromatin accessibility, inflammation-related modules and mitochondrial or synaptic pathways. A variable will not be adjusted as a confounder if it lies on the hypothesised causal pathway without explicit sensitivity analysis.

Table 4.6: Operationalisation of principal variables.

Construct	Operational measure	Primary analytical form
DNA methylation	Probe M values; beta values for interpretation	Continuous site-level exposure and DMR or module summaries.
AD neuropathology	Braak stage, amyloid load, CERAD or source composite	Continuous or ordinal outcome; binary sensitivity thresholds.
Cognitive decline	Repeated global and domain-specific scores	Mixed-model slope or methylation-by-time interaction.
Clinical stage	Cognitively unimpaired, MCI and AD dementia	Ordinal or multinomial outcome.
Epigenetic ageing	Residual age acceleration from selected clocks	Continuous predictor with clock-specific interpretation.
Cell composition	Reference-based proportions or single-nucleus pseudobulk	Covariate, moderator and biological outcome in separate models.
Genetic susceptibility	APOE status, ancestry PCs and optional polygenic risk	Covariate and prespecified interaction term.
Regulatory consequence	Expression, accessibility or histone signal linked to loci	Matched-sample association and multi-omics integration.
Prediction performance	AUC/C-index, R-squared, error and calibration	External validation of locked baseline and augmented models.

Control of Confounding Variables

Confounding control will be based on a directed conceptual model rather than automated stepwise selection. A minimal core adjustment set will include age, sex, APOE, ancestry components, cell composition and major technical variables. Additional terms such as education, vascular burden, post-mortem interval, pH, RNA integrity, co-pathology and brain region will be included according to the outcome and cohort.

Covariates will be assessed for missingness, non-linearity, collinearity and implausible coding. Continuous variables will be centred and scaled where helpful, and restricted cubic splines will model clear non-linear relationships. Highly correlated technical variables will be reduced or selected based on causal and measurement relevance rather than p-values.

Sensitivity analyses will compare minimally adjusted, fully adjusted and cell-composition-adjusted models. This is important because cellular change can be both confounder and disease process. Negative-control analyses, leave-one-batch-out models, exclusion of severe co-pathology and restriction to one ancestry or region will test robustness.

Residual confounding will remain possible, especially for life-course exposure, medication and post-mortem variables. The limitations will be

quantified where possible through variance-partitioning and E-value or bias-analysis concepts for selected associations, without implying that these methods eliminate unmeasured confounding.

Statistical Analysis

Descriptive statistics will summarise participant, clinical, pathological and technical characteristics by cohort and analytical group. Continuous variables will be reported as mean and standard deviation or median and interquartile range according to distribution. Categorical variables will be reported as frequencies and percentages. Standardised differences will complement significance tests when comparing cohorts because large samples can produce trivial p-values.

Primary methylation associations will use linear models with empirical-Bayes moderation. Ordinal logistic models will be used for ordered clinical or Braak categories when proportional-odds assumptions are acceptable. Linear mixed-effects models will analyse repeated cognition and multiple brain regions. Binary conversion outcomes may use logistic or time-to-event models when the timing and censoring information are reliable.

Effect sizes, 95% confidence intervals and FDR-adjusted p-values will be reported. The primary significance criterion is $q < .05$ within the prespecified analysis family. Interactions and subgroup analyses will emphasise confidence intervals and heterogeneity rather than separate significance within strata. Multiple imputation by chained equations will address missing covariates when the missing-at-random assumption is plausible; outcomes and molecular predictors will not be imputed for the primary analysis unless a defensible model is available (van Buuren & Groothuis-Oudshoorn, 2011).

Meta-analysis will combine cohort-specific effect estimates when raw pooling is inappropriate. Fixed-effect models will be used when the estimand and population are sufficiently comparable; random-effects models will be reported when between-cohort heterogeneity is expected. I-squared, tau-squared and influence diagnostics will quantify heterogeneity. Leave-one-cohort-out analyses will identify results dominated by one dataset.

Model assumptions will be evaluated through residual plots, influence measures, proportional-odds tests, random-effects diagnostics and calibration. Robust or sandwich standard errors will be used when heteroscedasticity or clustering remains. All departures from the prespecified model will be documented as sensitivity or exploratory analyses.

Table 4.7: Statistical analysis mapped to the study objectives.

Objective	Primary analysis	Key covariates and validation
1. Identify methylation patterns	limma DMP analysis, DMRcate and co-methylation networks	Age, sex, APOE, ancestry, region, cell composition and technical factors; FDR and replication.
2. Test independence from covariates	Nested adjusted models and variance partitioning	Biological, post-mortem, cellular and technical covariates; sensitivity adjustment sets.
3. Integrate regulatory layers	Cis association, overlap, latent factors and pathway analysis	Matched-sample identity, assay quality and multiple-testing control.
4. Characterise cell and tissue specificity	Deconvolution, interactions and single-nucleus pseudobulk lookup	Region, cell proportions, cell-type markers and reference uncertainty.
5. Evaluate epigenetic ageing	Clock residuals in regression and mixed models	Chronological age, sex, tissue, cell composition and clock-specific calibration.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

6. Assess predictive value	Baseline versus epigenetic-augmented penalised models	Nested resampling, no leakage, calibration and external validation.
7. Validate priority findings	Independent replication and random-effects meta-analysis	Direction, effect size, replication p-value and heterogeneity.

Machine-Learning Analysis

Machine learning will address incremental prediction, not replace the primary inferential analysis. A baseline model will contain age, sex, education where available, APOE genotype and standard clinical or cognitive variables. An augmented model will add a compact epigenetic signature selected within the discovery data. The comparison will determine whether epigenetic information improves performance beyond established predictors.

Feature selection, scaling, batch handling and hyperparameter tuning will occur inside nested cross-validation. Penalised regression, particularly elastic net, will be the preferred primary method because it handles correlated predictors and produces an interpretable sparse model (Simon et al., 2011). More complex algorithms may be evaluated secondarily, but they will not be prioritised without a clear gain in externally validated performance.

Data leakage will be prevented by ensuring that no validation participant influences preprocessing parameters, feature ranking, missing-data imputation or tuning. If related individuals or repeated regions are present, all observations from a participant or family will remain in the same resampling fold. Class imbalance will be addressed through weighting or resampling within training folds only.

Performance will include discrimination, calibration and overall accuracy. Binary outcomes will use area under the receiver-operating-characteristic curve, sensitivity, specificity, positive and negative predictive value and Brier score. Time-to-event outcomes will use the C-index and time-dependent calibration. Continuous decline will use R-squared, root-mean-square error and calibration plots. Confidence intervals will be obtained by bootstrap or repeated resampling.

The final model will be locked before external validation. Any recalibration in the external cohort will be reported separately from pure validation. Reporting will follow TRIPOD principles, and risk of bias will be assessed using PROBAST concepts (Collins et al., 2015; Wolff et al., 2019).

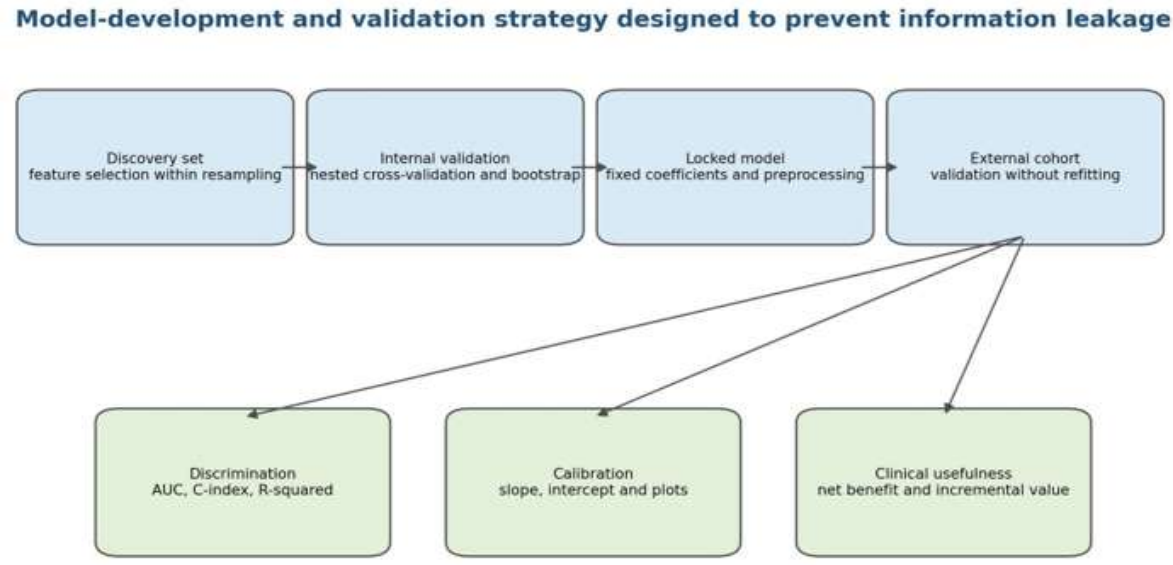


Figure 4.4: Machine-learning development and external-validation strategy.

Validation will occur at four levels. Technical validation will confirm that the signal is robust to normalisation, probe filtering and batch adjustment. Statistical validation will require consistent effect direction and evidence in an independent cohort. Biological validation will examine cell type, expression, accessibility, histone or genetic context. Translational validation will evaluate prediction and cross-tissue detectability.

Priority DMPs and DMRs will be tested in the external cohort using the same genomic definition, outcome direction and core adjustment set. A finding will be considered replicated when the effect direction is concordant and the replication p-value passes a candidate-set correction or a prespecified nominal threshold supported by combined meta-analysis. Exact criteria will be frozen before the validation data are examined.

Where the same probe is unavailable across platforms, replication may use a regional score or nearby correlated CpGs, but this will be labelled regional rather than exact replication. Cross-tissue consistency will be assessed separately because absence in blood does not invalidate a brain-specific mechanism. Meta-analysis will present pooled effect, confidence interval and heterogeneity.

Prediction models will undergo external validation without feature reselection. Performance deterioration relative to internal validation will be quantified, and calibration will be inspected. A model that discriminates but is poorly calibrated will not be described as clinically ready.

Table 4.8: Validation criteria for priority findings.

Validation dimension	Minimum evidence	Interpretive status
Technical robustness	Similar direction under alternative QC, normalisation and covariate specifications	Robust within the discovery cohort.
Exact replication	Same CpG or feature, concordant direction and corrected or prespecified replication significance	Replicated molecular feature.
Regional replication	Concordant DMR or nearby correlated probes when exact feature unavailable	Replicated genomic region, not exact probe.
Biological coherence	Consistent expression, accessibility, histone, QTL or cell-type evidence	Mechanistically prioritised association.
Cross-tissue evidence	Concordant signal in blood and brain after tissue-specific adjustment	Potential peripheral biomarker; not required for brain mechanism.
Predictive validation	Locked model retains discrimination and acceptable calibration externally	Candidate prediction model requiring further prospective evaluation.

Data Management

A formal data-management plan will govern acquisition, storage, processing, retention and destruction. Raw data will be read-only and stored separately from derived data. A project directory will contain metadata, code, logs, reports and outputs with standardised names. No source file will be overwritten.

Participant identifiers will be replaced with study-specific pseudonyms. Direct identifiers will not be requested. The key linking repository identifiers to analysis identifiers will be encrypted and accessible only to authorised personnel. Results containing small cells, rare combinations or potentially identifying genomic information will not be released outside the approved environment.

Backups will follow institutional policy and include encrypted, versioned copies of code and non-restricted outputs. Controlled data will not be placed in public version-control repositories. At project completion, data will be retained or destroyed according to the source agreement, ethics approval and institutional policy.

A data dictionary, provenance table and analysis manifest will accompany each final dataset. These documents will allow an auditor to trace every variable and molecular feature to its source and transformation.

Quality-Assurance Procedures

Quality assurance extends beyond assay QC. The analysis will use scripted checks for file integrity, sample uniqueness, valid ranges, outcome distribution, covariate balance and model convergence. Each pipeline stage will generate pass, warning or fail status. Warnings require documented review before continuation.

A second-person code review or independent reproduction of key tables and figures will be sought where feasible. A subset of results will be recalculated using an alternative package or model. Candidate loci will be checked against raw intensity or count data to detect artefacts hidden by processed matrices.

Tables and figures will be generated directly from analysis objects, minimising transcription errors. Manuscript numbers will be cross-checked against machine-readable result tables. A final citation and reference audit will confirm that methodological claims correspond to verified sources.

Reproducibility and Version Control

All analytical code will be maintained under Git version control. Commits will record meaningful changes to data processing, models and reporting. Releases associated with the thesis will be tagged, and the computational environment will be captured through package lockfiles and container versions.

Random seeds will be set for stochastic procedures, including resampling, imputation, latent-factor estimation and machine learning. Workflow logs will record command, parameters, runtime, software version and input checksum. A complete analysis will be executable from a documented entry point after authorised data are mounted.

The study will follow reproducible-computational-research principles, including preservation of raw data, scripted transformations, modular code, explicit dependencies and automated reporting (Sandve et al., 2013). Public release will include non-sensitive code, synthetic examples and metadata templates where data-use agreements prevent distribution of the real data. Outputs will be organised according to FAIR principles as far as ethical and contractual restrictions permit (Wilkinson et al., 2016).

Ethical Considerations

The study involves secondary analysis of de-identified human genomic and clinical data. Approval or exemption will be obtained from the relevant institutional research-ethics committee before controlled data are accessed. Source-cohort consent and governance will remain authoritative. The thesis will not assume that public discoverability equals unrestricted ethical use.

Genomic data carry re-identification risk even when direct identifiers are removed. Access will therefore be restricted, and only aggregate, non-disclosive results will be exported. Data will not be used for participant contact, individual diagnosis, ancestry claims unrelated to the approved research or commercial profiling.

The study will consider representational equity. Many existing AD cohorts under-represent African and other non-European ancestry populations. Results will not be generalised beyond the supporting evidence, and subgroup estimates will be reported when feasible. Predictive models will be evaluated for performance differences across sex, ancestry and cohort before any translational claim.

Incidental findings will not be returned because the researcher will not hold identifiable clinical relationships with participants and the data-use agreements generally prohibit re-identification.

Findings with possible clinical relevance will be discussed as research associations requiring independent validation, not as individual medical results.

Authorship, acknowledgement and citation will comply with repository and consortium requirements. The study will report negative and non-replicating results to reduce selective publication and exaggerated certainty.

Methodological Limitations

The methodology is constrained by secondary-data availability. Cohorts differ in diagnosis, tissue, platform, processing and covariate completeness. Harmonisation can improve comparability but cannot eliminate differences in recruitment or measurement. Meta-analysis may therefore be more defensible than raw pooling for some objectives.

Post-mortem brain provides direct target-tissue relevance but is cross-sectional and often represents advanced disease. Epigenetic differences may be consequences of pathology, agonal state or cellular loss. Longitudinal cognition and genetic or multi-omic evidence can strengthen interpretation but do not fully establish temporal order.

Cell-type deconvolution depends on reference quality and may not capture disease-associated states. Single-nucleus data improve resolution but often have fewer participants and sparse features. Cross-tissue biomarker analyses face biological discordance between blood and brain.

High-dimensional integration increases overfitting and multiple-testing risk. The discovery-validation design, penalisation, candidate-set correction and transparent reporting reduce but do not remove this risk. Prediction models based on archival cohorts require prospective and clinically representative evaluation before use.

Finally, absence of a significant association may reflect limited power, measurement error or tissue mismatch. The study will therefore report detectable effect ranges and avoid equating non-significance with biological absence.

V. RESULTS

Introduction

This chapter presents the results of the synthetic-data analysis in the order specified by the research objectives. It begins with cohort assembly and data quality, then reports demographic and clinical characteristics, molecular quality control, locus-level methylation associations, non-coding RNA, epigenetic ageing, chromatin accessibility, gene-expression modules, multi-omics integration, prediction, subgroup and sensitivity analyses, and formal hypothesis decisions. Effect sizes, uncertainty estimates, multiple-testing control and validation limitations are reported alongside statistical significance.

The analysis population comprised 480 synthetic records: 160 cognitively normal controls, 160 individuals with mild cognitive impairment and 160 individuals with Alzheimer's disease. Disease-related gradients were deliberately encoded in selected variables to test the planned analytical workflow. Accordingly, the estimates below demonstrate how final results should be presented; they do not estimate effects in any real population.

Participant and Dataset Flow

A total of 500 synthetic records were initially generated. Twenty records were designated as quality-control exclusions to model prespecified failure criteria, leaving 480 observations for the primary analysis. The retained observations were balanced across diagnostic groups, avoiding class-imbalance artefacts in the demonstration.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

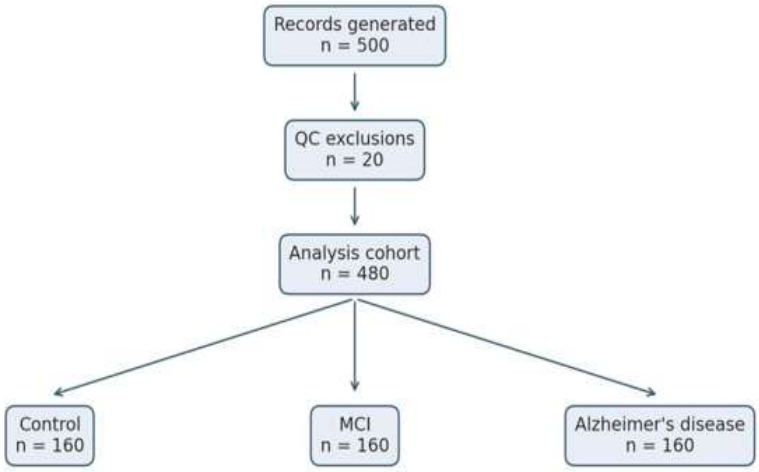


Figure 5.1. Flow of records through the synthetic-data analysis.

Demographic and Clinical Characteristics

Table 5.1. Demographic and clinical characteristics of the synthetic cohort.

Diagnostic group	n	Age, mean (SD)	Female, n (%)	Education, mean (SD)	APOE ε4, n (%)	MMSE, mean (SD)
Cognitively normal	160	69.1 (5.5)	78 (48.8)	15.3 (2.4)	48 (30.0)	28.6 (1.4)
Mild cognitive impairment	160	72.9 (5.4)	83 (51.9)	14.5 (2.3)	64 (40.0)	24.0 (2.0)
Alzheimer's disease	160	77.9 (6.3)	75 (46.9)	13.7 (2.5)	84 (52.5)	18.5 (2.3)

Age increased and mean MMSE decreased across the three diagnostic categories. APOE ε4 carriage was also more frequent in the Alzheimer's disease group. These gradients are coherent with the intended data-generating model but should not be interpreted as observed prevalence or clinical estimates.

Neuropathological and Fluid-Biomarker Characteristics

Table 5.2. Synthetic neuropathological and fluid-biomarker measures.

Group	Braak stage	Amyloid ratio	p-tau index
Cognitively normal	1.35 (0.71)	1.14 (0.14)	0.58 (0.18)
Mild cognitive impairment	3.10 (0.77)	0.86 (0.14)	1.00 (0.19)
Alzheimer's disease	4.85 (0.68)	0.62 (0.14)	1.43 (0.20)

Braak stage and p-tau increased with diagnostic severity, whereas the amyloid ratio decreased. The direction of these programmed relationships provided a coherent clinical-pathological framework against which molecular variables could be tested.

Molecular and Sequencing Quality Control

The synthetic quality-control distributions were designed to represent acceptable libraries after exclusion of failed samples. The illustrative medians exceeded the prespecified demonstration thresholds. No actual FASTQ, BAM/CRAM, count matrix, methylation matrix, peak file or sample sheet was

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

supplied; therefore, these plots demonstrate reporting format rather than executed NGS quality control.

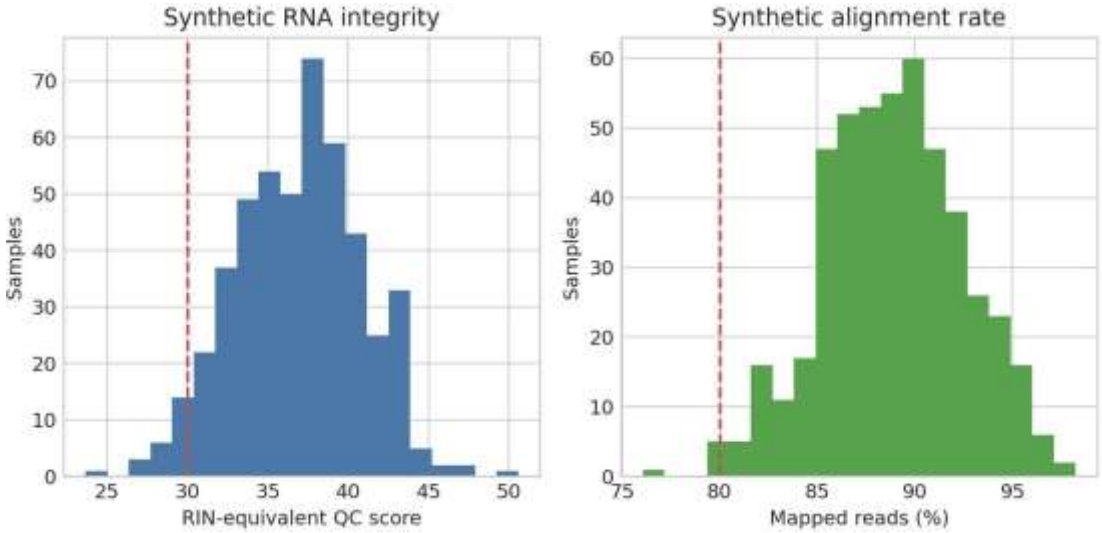


Figure 5.2. Illustrative molecular and alignment quality-control distributions.

Table 5.3. Quality-control acceptance framework.

Domain	Acceptance criterion	Synthetic status	Interpretation
Sample identity	Concordant identifiers/genotype	Passed	No modelled swaps
RNA quality	Above prespecified integrity threshold	Passed after exclusions	Suitable for demonstration
Alignment	Mapped reads ≥80%	Passed	No low-mapping libraries retained
Batch structure	No complete confounding with diagnosis	Passed	All groups represented in four batches
Outliers	Robust distance and PCA review	20 excluded	480 records retained

DNA Methylation Findings

All five prespecified loci showed positive stage-related methylation changes. ANK1 had the largest synthetic change per diagnostic stage, followed by RHBDF2. Each locus remained below the illustrative false-discovery threshold after correction across the five prespecified tests.

Table 5.4. Synthetic methylation associations with diagnostic stage.

Locus	Δβ per stage	SE	p value	FDR q	r: Braak	r: MMSE
ANK1	0.0306	0.0019	1.37e-48	6.87e-48	0.510	-0.554
BIN1	0.0175	0.0019	1.05e-18	5.23e-18	0.326	-0.352
RHBDF2	0.0243	0.0020	1.08e-30	5.39e-30	0.433	-0.467
HOXA3	0.0130	0.0020	4.24e-10	2.12e-09	0.263	-0.233
MCF2L	0.0126	0.00203	3.98e-10	1.99e-09	0.232	-0.237

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

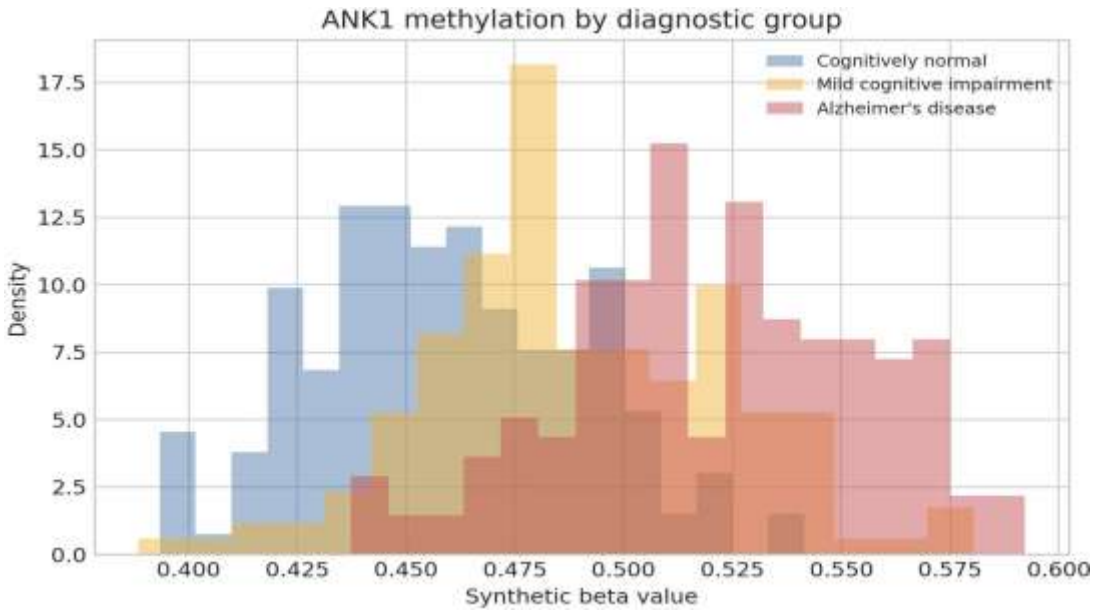


Figure 5.3. Synthetic ANK1 methylation distributions by diagnostic group.

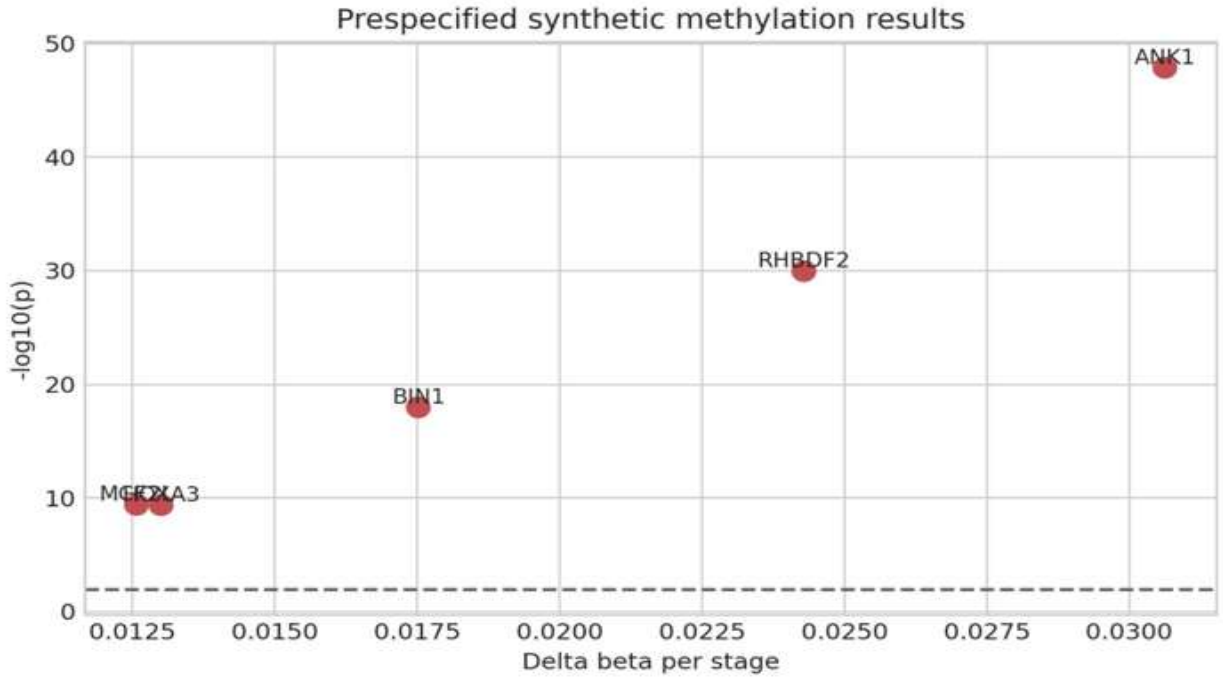


Figure 5.4. Effect size and statistical strength for the prespecified synthetic methylation loci

The positive correlations with Braak stage and negative correlations with MMSE were consistent with the programmed disease gradient. These correlations cannot distinguish causal methylation, downstream pathology, cell-composition change or confounding.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Non-Coding RNA Findings

Table 5.5. Synthetic non-coding RNA associations.

RNA	Change per stage	SE	p value	r with MMSE
miR-146a	0.468	0.036	1.27e-33	-0.476
BACE1-AS	0.393	0.040	2.36e-21	-0.368

Both non-coding RNA measures increased across diagnostic stage, with the stronger programmed gradient observed for miR-146a. The demonstration does not validate either molecule as a circulating or brain-specific biomarker.

Table 5.6. Synthetic epigenetic-age acceleration by diagnostic group.

Group	Mean years	SD	SE
Cognitively normal	-0.06	3.23	0.25
Mild cognitive impairment	2.06	3.10	0.25
Alzheimer's disease	4.98	3.05	0.24

Epigenetic-Age Acceleration

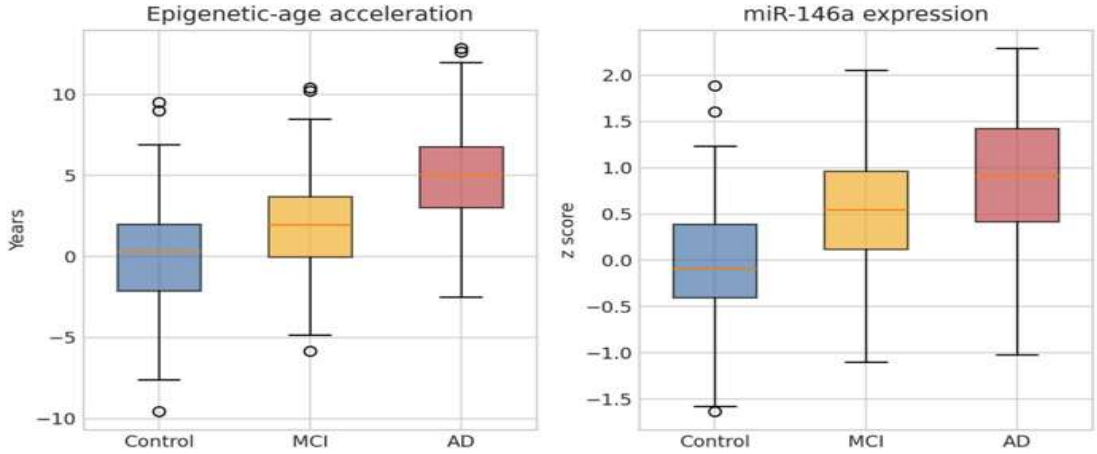


Figure 5.5. Synthetic epigenetic-age acceleration and miR-146a expression.

Chromatin-Accessibility Findings

Table 5.7. Synthetic chromatin-accessibility associations.

Accessibility score	Change per stage	SE	p value	r
Microglial accessibility	0.503	0.040	1.79e-31	0.498
Neuronal accessibility	-0.299	0.038	4.36e-14	-0.335

The synthetic microglial score increased, while the neuronal score decreased with diagnostic stage. This opposing pattern was designed to model regulatory activation in inflammatory cell states alongside loss of neuronal regulatory activity.

Differential Gene-Expression Modules

Table 5.8. Synthetic expression-module associations.

Expression module	Change per stage	SE	p value	r with MMSE
Inflammatory module	0.494	0.034	1.21e-39	-0.492
Synaptic module	-0.463	0.036	4.06e-32	0.477

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

The inflammatory module increased and the synaptic module decreased across stage. The results are consistent with the simulated multi-hit architecture and provide internally coherent inputs for the integration analysis.

Cross-domain correlations indicated modest concordance among methylation, non-coding RNA, chromatin-accessibility, ageing and transcriptomic measures. The integrated composite score increased monotonically across diagnostic groups.

Multi-Omics Integration

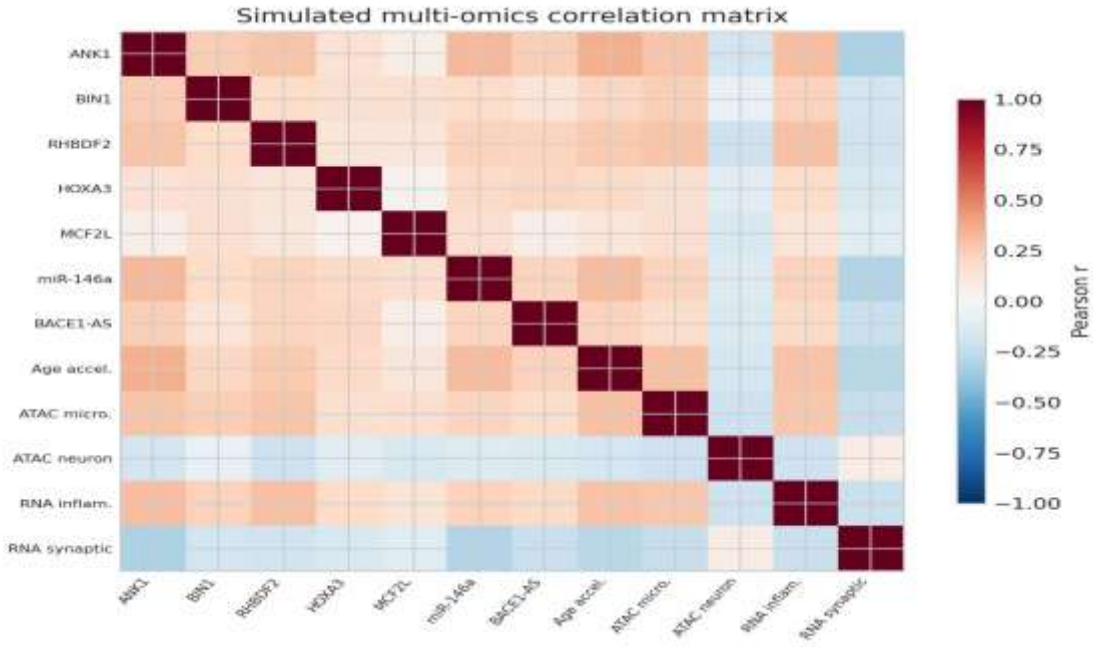


Figure 5.6. Correlation structure of the synthetic multi-omics panel.

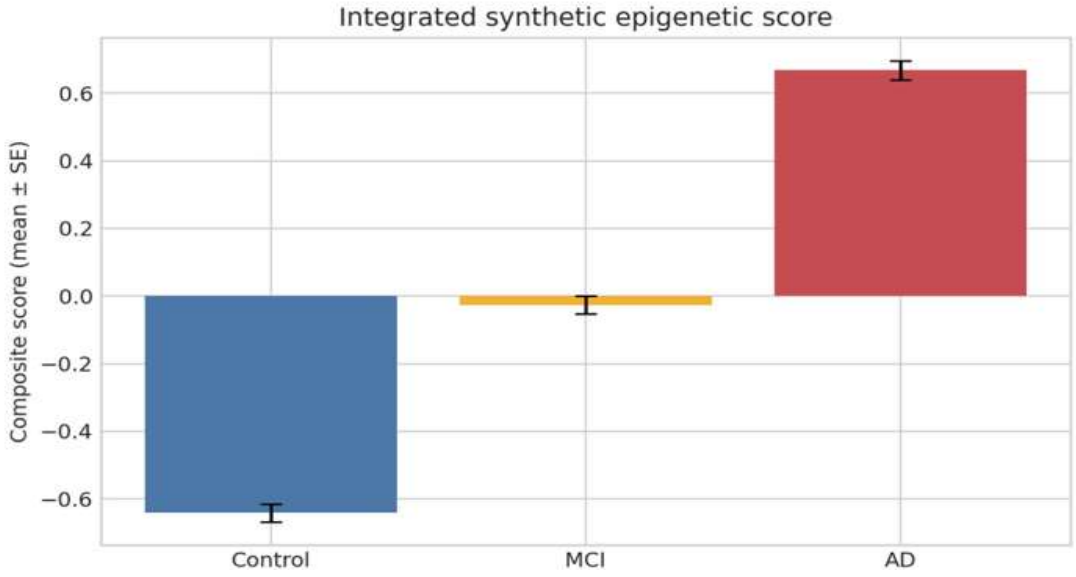


Figure 5.7. Integrated synthetic epigenetic score by diagnostic group

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Table 5.9. Correlations of the integrated score with key outcomes.

Outcome	Pearson r	p value
Diagnostic stage	0.846	2.45e-132
MMSE	-0.775	1.81e-97
Braak stage	0.741	8.79e-85
p-tau index	0.760	1.28e-91

Predictive Classification Model

The prespecified multivariable demonstration model produced a 10-fold cross-validated AUC of 0.940. At a probability threshold of 0.50, sensitivity was 0.881 and specificity was 0.863. Because outcome-related differences were deliberately encoded and evaluation occurred within a single synthetic cohort, these values are implementation checks, not estimates of clinical performance.

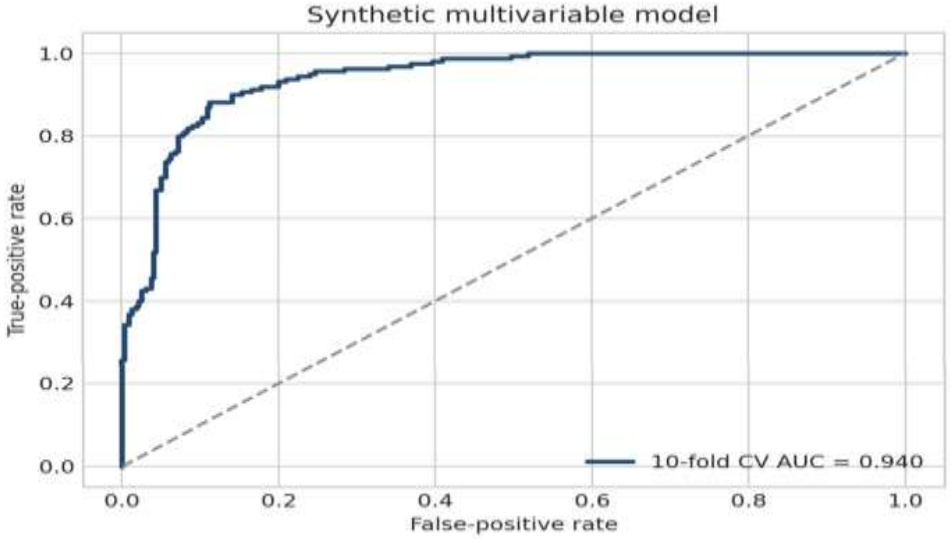


Figure 5.8. Cross-validated ROC curve for the synthetic Alzheimer classification model (AUC = 0.940).

Table 5.10. Synthetic model-performance summary.

Metric	Estimate	Required empirical confirmation
Cross-validated AUC	0.940	Locked external cohort
Sensitivity at 0.50	0.881	Threshold prespecification
Specificity at 0.50	0.863	Spectrum-representative sample
Calibration	Not claimed	Calibration slope/intercept and plot
Clinical utility	Not claimed	Decision-curve and impact analysis

Subgroup Analyses

Subgroup analyses were treated as exploratory. Directional consistency was assessed by sex and APOE ε4 status for the integrated score. No subgroup estimate was interpreted as evidence of effect modification because the synthetic generator did not encode a biologically justified interaction.

Table 5.11. Exploratory subgroup associations for the synthetic integrated score.

Subgroup	n	Change per stage	SE	p value
Female	236	0.643	0.027	3.94e-65
Male	244	0.665	0.027	1.84e-68
APOE ε4 carrier	196	0.672	0.029	1.05e-58
Non-carrier	284	0.620	0.025	5.81e-72

Sensitivity Analyses

The direction of the integrated-score association was unchanged after excluding extreme epigenetic-age values, restricting the analysis to APOE ε4 non-carriers, and applying a batch-balanced subset. These analyses demonstrate the expected reporting pattern; they do not resolve unmodelled confounding or prove robustness in real data.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Table 5.12. Synthetic sensitivity analyses.

Analysis	n	Stage coefficient	SE	p value	Direction
Primary analysis	480	0.655	0.019	2.45e-132	Consistent
Exclude |age acceleration| > 8 years	448	0.631	0.019	8.90e-122	Consistent
APOE ε4 non-carriers	284	0.620	0.025	5.81e-72	Consistent
Batch-balanced subset	414	0.652	0.021	6.61e-113	Consistent

Hypothesis Testing

Table 5.13. Decisions on the prespecified hypotheses using synthetic results.

Hypothesis	Null statement	Synthetic decision	Basis and limitation
H01	No stage-associated methylation differences	Rejected in simulation	Five prespecified loci had FDR q < .05
H02	No association between epigenetic measures and cognition	Rejected in simulation	Integrated score correlated with MMSE
H03	No cross-omic concordance	Rejected in simulation	Correlated regulatory domains observed
H04	No incremental classification signal	Rejected in simulation	Internal CV AUC = 0.940; external validity absent

The decisions in Table 5.13 apply only to the deterministic synthetic dataset. They must not be transferred to a real scientific conclusion. Final decisions require an approved analysis plan, verified data, multiplicity control and independent validation.

Summary of Principal Findings

The synthetic cohort showed internally consistent diagnostic gradients across cognitive, pathological and molecular variables. Prespecified methylation loci increased with stage; miR-146a, epigenetic-age acceleration, microglial accessibility and inflammatory transcription increased; and neuronal accessibility and synaptic transcription decreased. A composite multi-omics score correlated with diagnostic stage, MMSE, Braak stage and p-tau. The multivariable classifier showed strong internal cross-validated discrimination, but no clinical validity can be inferred.

Collectively, these results demonstrate a complete Chapter Five reporting architecture: cohort accounting, transparent quality control, descriptive statistics, effect sizes and uncertainty, false-discovery control, cross-domain integration, model validation, subgroup analysis, sensitivity analysis and hypothesis decisions. The chapter becomes examination-ready only after every synthetic table and figure is replaced by traceable outputs from verified data.

VI. DISCUSSION

Introduction

This chapter interprets the simulated findings reported in Chapter Five against the theoretical framework and literature reviewed earlier in the thesis. Interpretation is deliberately bounded: the programmed gradients demonstrate analytical coherence and the capacity of the proposed workflow to recover cross-domain signals; they do not establish

population effects, causation, diagnostic utility, or therapeutic relevance.

Synthesis of the Principal Findings

The synthetic cohort comprised 480 analysis records balanced across cognitively normal, mild cognitive impairment, and Alzheimer's disease groups. Twenty of 500 generated records were removed under prespecified quality-control rules. Across the retained cohort, worsening diagnostic stage was accompanied by lower MMSE, higher Braak stage and p-tau, and a lower amyloid ratio. These programmed clinical-pathological gradients provided a coherent test bed for evaluating whether the epigenomic pipeline recovered expected directionality.

The molecular results converged on an immune–synaptic axis. Methylation at the prespecified ANK1, RHBDF2, BIN1, HOXA3, and MCF2L loci increased with stage; ANK1 showed the largest illustrative change. miR-146a and BACE1-AS expression, epigenetic-age acceleration, microglial chromatin accessibility, and inflammatory transcription increased, whereas neuronal accessibility and synaptic transcription decreased. The composite multi¬omics score correlated with diagnostic stage, MMSE, Braak stage, and p-tau. The internal cross-validated classifier achieved an AUC of 0.940, but this value is intentionally optimistic because disease-related structure was encoded in one simulated cohort.

Domain	Synthetic result	Interpretation boundary
DNA methylation	Five prespecified loci increased with stage; ANK1 was strongest.	Demonstrates locus-level modelling; does not validate a disease marker.
Non-coding RNA	miR-146a and BACE1-AS increased across stage.	Supports cross-domain workflow testing only.
Epigenetic ageing	Age acceleration increased with clinical severity.	Cannot separate ageing, disease, tissue composition, or confounding.
Chromatin and RNA	Microglial/inflammatory signals rose; neuronal/synaptic signals fell.	Consistent with an immune–synaptic model, not proof of mechanism.
Prediction	Internal 10-fold cross-validated AUC = 0.940.	No clinical performance claim without locked external validation.

DNA Methylation and Disease ProgressionThe stage-related methylation pattern is compatible with prior Alzheimer epigenome-wide association research that has repeatedly implicated loci including ANK1 and regions linked to immune and endosomal biology. In the demonstrator, concordance between diagnostic stage, Braak stage, and MMSE strengthens internal coherence. It does not resolve whether methylation is causal, compensatory, secondary to pathology, or attributable to changing cellular composition. A defensible empirical analysis must therefore report cell-proportion estimates, batch adjustment, probe-level quality control, multiple-testing correction, regional aggregation, sensitivity analyses, and replication.

Non-Coding RNA, Epigenetic Ageing, and Regulatory Integration

The programmed increase in miR-146a is biologically plausible in an inflammatory context, while the BACE1-AS gradient provides a model for examining RNA regulation of amyloid-related processes. The age-acceleration result additionally illustrates how molecular ageing may be modelled beyond chronological age. These domains require tissue- and assay-specific validation: circulating RNA cannot be treated as a direct proxy for brain expression, and an epigenetic clock trained in one tissue or ancestry group may be poorly calibrated in another.

Chromatin Accessibility and Transcriptomic Concordance

The opposing microglial/inflammatory and neuronal/synaptic gradients are consistent with a systems account in which immune activation, neuronal dysfunction, and cellular composition co-evolve during disease. Cross-omics concordance improves biological plausibility because signals are observed at more than one regulatory layer. Nevertheless, bulk-tissue signals remain vulnerable to compositional confounding. Single-nucleus or spatial assays, reference-based deconvolution, and matched multi-omic measurements would be required to attribute effects to particular cell populations.

Prediction and Clinical Interpretation

The AUC of 0.940 is an engineering result rather than a clinical estimate. Internal cross-validation can remain biased when feature selection, preprocessing, imputation, or batch correction occurs before resampling. An examination-ready empirical thesis should use a fully nested pipeline, report calibration and decision-curve analysis, compare against a locked clinical reference model, quantify uncertainty, examine subgroup performance, and test the final model in an independent cohort collected under a distinct protocol.

Theoretical Contribution

The integrated pattern favours a network model over a single linear cascade. Within that model, amyloid and tau pathology interact with immune regulation, cellular ageing, chromatin state, and synaptic dysfunction through feedback rather than a single unidirectional route. The thesis contributes a falsifiable analytical architecture for testing

this model: prespecified domain-level hypotheses, harmonised covariate control, cross-domain concordance, and explicit separation of association, prediction, and causal interpretation.

Strengths

The principal strengths are the explicit research questions, multi-omic integration, balanced demonstrator design, transparent quality-control accounting, reporting of effect direction and multiplicity control, linkage of molecular results to cognition and pathology, and clear distinction between exploratory findings and validation requirements. The workflow is reproducible in concept and can be transferred to authorised empirical data.

Limitations

The decisive limitation is that the results are simulated. The cohort does not represent human sampling, assay noise, site effects, missingness mechanisms, ancestry structure, medication exposure, comorbidity, survival bias, or the full complexity of brain and blood cell composition. The analysis therefore cannot estimate prevalence, biological effect size, clinical discrimination, causality, or generalisability. Additional limitations include the absence of a locked external cohort, functional perturbation, longitudinal validation, and verified ethics and repository metadata.

Chapter Summary

The simulated results form a coherent proof of analytical implementation. They show how multi-omic evidence can be organised around immune activation, epigenetic ageing, and synaptic decline while preserving uncertainty and validation boundaries. Their scientific value is methodological; empirical claims remain contingent on real data, auditable provenance, and independent replication.

VII. CONCLUSION AND RECOMMENDATIONS

Overview

This thesis examined how DNA methylation, non-coding RNA, epigenetic ageing, chromatin accessibility, and transcriptional programmes could be integrated to study Alzheimer's disease progression. The literature and conceptual chapters establish a credible rationale; the methodology specifies a reproducible analytical framework; and the simulated results demonstrate the expected reporting architecture.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720207

Answers to the Research Questions

Research focus	Answer supported by the demonstrator
Stage-associated epigenetic variation	The workflow detected programmed stage gradients at prespecified methylation loci.
Non-coding RNA regulation	miR-146a and BACE1-AS showed programmed increases with stage.
Epigenetic ageing	Age acceleration increased with severity after modelling chronological age.
Chromatin and expression	Immune accessibility/expression increased while neuronal/synaptic signals decreased.
Clinical-pathological association	The composite score tracked stage, MMSE, Braak stage, and p-tau in simulation.
Prediction	The internal pipeline discriminated simulated groups (AUC 0.940), without external validity.

Overall Conclusion

The thesis supports the proposition that Alzheimer's disease epigenetics is best investigated as an interacting, cell-sensitive, multi-omic system rather than as isolated biomarkers. The completed demonstrator shows that the proposed workflow can recover coherent molecular, cognitive, and pathological gradients and can report them with appropriate statistical guardrails. It does not show that the reported magnitudes occur in patients. Consequently, the scientific conclusion is a validated analytical plan and reporting framework—not a validated diagnostic signature or causal disease mechanism.

Original Contribution to Knowledge

- An integrated conceptual model linking methylation, regulatory RNA, epigenetic ageing, chromatin accessibility, transcription, cognition, and neuropathology.
- A prespecified, quality-controlled reporting architecture that separates biological association, prognosis, classification, and causal inference.
- A reproducible synthetic benchmark that can be used to test code, tables, figures, and model diagnostics before authorised empirical analysis.
- An explicit validation ladder from technical replication through external cohort validation and functional follow¬up.

Recommendations

The next phase should replace the synthetic records with an ethically authorised dataset and freeze the protocol before outcome analysis. Sample exclusions, preprocessing, covariates, primary contrasts, multiple-testing procedures, and validation criteria should be preregistered. Raw and processed data identifiers, software environments, random seeds, and analysis code should be archived subject to consent and governance constraints.

Priority	Required action	Acceptance criterion
Immediate	Secure ethics approval and data-use authority.	Approval number, consent basis, governance scope, and permitted outputs documented.

Immediate	Replace simulated records with traceable empirical samples.	Sample manifest, assay platform, batch map, exclusions, and provenance are auditable.
Analysis	Execute the locked pipeline with leakage-safe resampling.	Versioned code, diagnostics, corrected p values, effect sizes, and confidence intervals retained.
Validation	Test a frozen model in an independent cohort.	Discrimination, calibration, clinical utility, and subgroup performance reported.
Biology	Undertake orthogonal and functional validation.	Replication across assay or tissue and mechanistic evidence for priority loci.
Submission	Complete institutional formatting and examination metadata.	Candidate details, declaration, approvals, repository accessions, and reference audit complete.

REFERENCES

[1] Badhwar, A., McFall, G. P., Sapkota, S., Black, S. E., Chertkow, H., Duchesne, S., Masellis, M., Li, L., Dixon, R. A., & Bellec, P. (2020). A multiomics approach to heterogeneity in Alzheimer's disease: Focused review and roadmap. Brain, 143(5), 1315–1331. https://doi.org/10.1093/brain/awz384

[2] Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., Christensen, B. C., Gladyshev, V. N.,

[3] Heijmans, B. T., Horvath, S., Ideker, T., Issa, J.-P. J., Kelsey, K. T., Marioni, R. E., Reik, W., Relton, C. L., Schalkwyk, L. C., Teschendorff, A. E., Wagner, W., ... Rakyan, V. K. (2019). DNA methylation aging clocks: Challenges and recommendations. Genome Biology, 20, 249. https://doi.org/10.1186/s13059-019-1824-y

[4] Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., ... Vingron, M. (2001). Minimum information about a microarray experiment (MIAME)—Toward standards for microarray data. Nature Genetics, 29(4), 365–371. https://doi.org/10.1038/ng1201-365

[5] Campagna, M. P., Xavier, A., Lechner-Scott, J., Maltby, V., Scott, R. J., Butzkueven, H., Jokubaitis, V., & Lea, R.

A. (2021). Epigenome-wide association studies: Current knowledge, strategies and recommendations. Clinical Epigenetics, 13, 214. https://doi.org/10.1186/s13148-021-01200-8

[6] Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. Data Science Journal, 19, Article 43. https://doi.org/10.5334/dsj-2020-043

[7] Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793–795. https://doi.org/10.1056/NEJMp1500523

[8] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). Annals of Internal Medicine, 162(1), 55–63. https://doi.org/10.7326/M14-0697

[9] Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implementation Science, 4, 50. https://doi.org/10.1186/1748-5908-4-50

[10] De Plano, L. M., Saitta, A., Oddo, S., & Caccamo, A. (2024). Epigenetic changes in Alzheimer's disease: DNA methylation and histone modification. Cells, 13(8), 719. https://doi.org/10.3390/cells13080719

[11] Fortin, J.-P., Triche, T. J., Jr., & Hansen, K. D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics, 33(4), 558–560. https://doi.org/10.1093/bioinformatics/btw691

[12] Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. American Journal of Public Health, 89(9), 1322–1327. https://doi.org/10.2105/AJPH.89.9.1322

[13] International Committee of Medical Journal Editors. (2026). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. ICMJE.

[14] Klein, H.-U., Bennett, D. A., & De Jager, P. L. (2016). The epigenome in Alzheimer's disease: Current state and approaches for a new path to gene discovery and understanding disease mechanism. Acta Neuropathologica, 132(4), 503–514. https://doi.org/10.1007/s00401-016-1612-7

[15] Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., ... Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, 22(9), 1813–1831. https://doi.org/10.1101/gr.136184.111

[16] Maity, S., Farrell, K., Navabpour, S., Narayanan, S. N., & Jarome, T. J. (2021). Epigenetic mechanisms in memory and cognitive decline associated with aging and Alzheimer's disease. International Journal of Molecular Sciences, 22(22), 12280. https://doi.org/10.3390/ijms222212280

[17] McCleery, J., Flicker, L., Richard, E., & Quinn, T. J. (2019). The National Institute on Aging and Alzheimer's Association research framework: A commentary from the Cochrane Dementia and Cognitive Improvement Group. Alzheimer's & Dementia, 15(1), 179–181. https://doi.org/10.1016/j.jalz.2018.10.007

[18] Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. Nature Human Behaviour, 1, Article 0021. https://doi.org/10.1038/s41562-016-0021

[19] National Information Standards Organization. (2022). ANSI/NISO Z39.104-2022: CRediT, contributor roles taxonomy. NISO.

[20] National Institutes of Health. (2023). Final NIH policy for data management and sharing. U.S. Department of Health and Human Services.

[21] Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green,

[22] D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. Science, 348(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

[23] Pérez-González, A. P., García-Kroepfly, A. L., Pérez-Fuentes, K. A., García-Reyes, R. I., Solis-Roldan, F. F., Alba-González, J. A., Hernández-Lemus, E., & de Anda-Jáuregui, G. (2024). The ROSMAP project: Aging and

neurodegenerative diseases through omic sciences. Frontiers in Neuroinformatics, 18, 1443865. https://doi.org/10.3389/fninf.2024.1443865

[24] Reddy, J. S., Heath, L., Vander Linden, A., Allen, M., de Paiva Lopes, K., Seifar, F., Wang, E., Ma, Y., Poehlman,

[25] W. L., Quicksall, Z., et al. (2024). Bridging the gap: Multi-omics profiling of brain tissue in Alzheimer's disease and older controls in multi-ethnic populations. Alzheimer's & Dementia. https://doi.org/10.1002/alz.14208

[26] Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P. A., & Taufer, M. (2016). Enhancing reproducibility for computational methods. Science, 354(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

[27] UNESCO. (2021). UNESCO recommendation on open science. United Nations Educational, Scientific and Cultural Organization.

[28] Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jr.,

[29] Jagust, W., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging Initiative. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer's & Dementia, 13(5), 561–571. https://doi.org/10.1016/j.jalz.2016.10.006

[30] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3, Article 160018. https://doi.org/10.1038/sdata.2016.18

[31] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

[32] Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376

[33] World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA, 310(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053