UNIVERSITY OF CALIFORNIA

Los Angeles

# Combining omics approaches to evaluate biomarkers for complex brain disorders

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Neuroscience

by

Marcelo Ricardo Francia

2024

ABSTRACT OF THE DISSERTATION

Combining omics approaches

to evaluate biomarkers for complex brain disorders

by

Marcelo Ricardo Francia

Doctor of Philosophy in Neuroscience

University of California, Los Angeles, 2024

Professor Roel A. Ophoff, Chair

This dissertation is dedicated to enhancing the discovery of biologically relevant information for endophenotypes and biomarkers for complex brain disorders through the integration of data from multiple diverse omics domains, including genomics, transcriptomics, epigenomics, and metabolomics. Chapter 1 focuses on evaluating the functional genomic features captured by skin fibroblasts as an *in vitro* model for circadian studies, particularly within the context of Bipolar Disorder. This investigation utilizes longitudinal gene expression and chromatin accessibility data from six cell lines across thirteen timepoints. In Chapter 2, we delve into how distinct biological layers, including genetics and cerebrospinal fluid metabolites, contribute to understanding various aspects of Alzheimer's Disease pathology as reflected in established cerebrospinal fluid biomarkers, such as amyloid beta 42, total tau, and phosphorylated tau cerebrospinal fluid levels. Our findings underscore the utility of integrating functional genomics platforms to characterize the features that an *in vitro* model captures and assess their relevance for specific endophenotypes and disorders. Additionally, through CSF metabolomics

analysis, we identify novel metabolites associated with phosphorylated tau and total tau CSF levels, such as Anserine and Fucose. This work signifies another step forward in advancing our comprehension of the underlying biology of complex brain disorders. By leveraging current technological advancements and addressing the challenges of integrating omics approaches, we aim to unravel new insights and avenues for diagnosis, treatment, and management of these debilitating conditions.

The dissertation of Marcelo Ricardo Francia is approved.

Christopher S. Colwell

Michael Gandal

Valerie A. Arboleda

Roel A. Ophoff, Committee Chair

University of California, Los Angeles

2024

## DEDICATION

Por los que vendrán

For those that will come

## TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| ABBREVIATION | DEFINITION |
|---|---|
| AD | Alzheimer's Disease |
| ADC | Amsterdam Dementia cohort |
| ADHD | Attention-Deficit/HyperactivityDisorder |
| ALS | Amyotrophic lateral sclerosis |
| ATN | Amyloid, tau, and neurodegeneration |
| BD | Bipolar Disorder |
| CPM | Counts per million |
| CSF | Cerebrospinal fluid |
| DSM | Diagnostic and Statistical Manual of Mental Disorders |
| EHRs | Electronic health records |
| FBS | Fetal bovine serum |
| FDR | False discovery rate |
| FTD | Frontotemporal dementia |
| GO | Gene ontology |
| GR | Glucocorticoid receptor |
| GRE | Glucocorticoid response element |
| GSA | Global Screening Array |
| GWAS | Genome-wide association studies |
| HPA | Hypothalamic-pituitary-adrenal axis |
| MAF | Minor allele frequency |
| MCI | Mild cognitive impairment |
| MDD | Major depression disorder |
| meQTLs | Metabolite quantitative trait loci |
| NC | Normal cognition |

| | |
|---|---|
| NFT | Neurofibrillary tangles |
| NMF | Non-negative matrix factorization |
| NRF | Non-redundant Fraction |
| OCRs | Open chromatin regions |
| P-Tau | Phosphorylated tau |
| PBC | PCR Bottlenecking Coefficients |
| PET | Positron emission tomography |
| PGC | Psychiatric Genomics Consortium |
| PRS | Polygenic risk scores |
| PTSD | Post-traumatic stress disorder |
| SCD | Subjective cognitive decline |
| SCN | Suprachiasmatic nucleus |
| sLDSC | Stratified Linkage Disequilibrium Score Regression |
| SNP | Single nucleotide polymorphism |
| T-Tau | Total tau |
| VUmc | VU University Medical Center |
| WGCNA | Weighted Gene Co-Expression Network Analysis |

## LIST OF FIGURES

## LIST OF TABLES

## ACKNOWLEDGMENTS

"The Universe is made of stories, not atoms", said Muriel Rukeyser. As scientists, we are fundamentally story tellers. We pursue life's most beautiful mysteries, and weave narratives to understand them. Narratives that were built upon the work of those that came before us. Work across multiple generations, lessons learned through time. I would not be here if not for the support of many individuals, whose influence on the road so far cannot be overstated. Firstly, I am truly grateful to Dr. Roel A. Ophoff for his support and mentorship during my graduate experience. The trust he placed upon me allowed me to grow into the independent scientist that I am today. Thank you as well to Dr. Valerie Arboleda, Dr. Chistopher Colwell and Dr. Michael Gandal for their insightful feedback and encouraging advice over the past years on my doctoral committee.

Thank you to all of the Ophoff lab, without whom this work would not have been possible. Merel, for her wet lab work that was used to generate the datasets used in this dissertation, and also for her great encouragement since the beginning. To Naren, who worked on analyzing the datasets presented in chapter two, and whose presence makes the laboratory a brighter space. To Toni, without whom I would not have succeeded in this line of work, and for her great work on all these projects. To Juan, for his guidance on statistics and thinking one step at a time, and for showing me that it can be done. Thank you to Lingyu, Carolinne, Lianne, Kevin and Artemis. It was great great working with you all.

Over the past five years, many incredible friendships have supported and stood beside me across this unknown road. Thank you Leo, for being the first friendship that I made in the West. Ari, for showing me how good people can be. Sarah, for all the wonderful adventures that we had. Thank you Gloria, Nataly, Nathaly, Paul for coming along in such long and rewarding trips. It has been a blessing being able to experience the wonders of nature with you all. Thank

you Gil, for all those tunes and conversations. Thank you Ray, for being a true friend, and for giving me a home on this foreign land. And thank you to all the people that created a space for me, in which we can all be weird. Thank you Ward, Carolinne, Doug, Sara.

Thank you Agosto, for showing me how to be a scientist. And thank you, David, for telling me that I can be scientist. Without you, I would have probably been a physician. Thank you, from the bottom of my heart for being a brother. Thank you Nao, Helix and Rusty, for always being a call away. Thank you Mateo for all these years of friendship.

And lastly, thank you to my family. To my father, for financially supporting me for the past years, and giving me the opportunity of a lifetime. To my mother, for all the challenges that we overcame, teaching me to be strong and to persevere. And to Del, for being a supportive friend. To Eduardo, for teaching me tenacity and to have an inquisitive nature about things. And to all others who have followed my journey so far, I promise to not disappoint your expectations.

For chapter one, this study was supported by funding from the National Institutes of Health (NIH), research grants R01 MH090553, R01 MH115676, the NARSAD Distinguished Investigator Grant (to Roel A. Ophoff), and NS048004 T32 Training grant in Neurobehavioral genetics.

For chapter 2, I am greatly appreciative to those individuals who donated the CSF samples on which this study was based. This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant number 115372) and Stichting Dioraphte. Genotyping of the Dutch case-control samples was performed in the context of EADB (European Alzheimer DNA biobank) funded by the JPco-fuND FP-829- 029 (ZonMW projectnumber 733051061). This project was funded by the NIH National Institute on Aging (NIA) grant RF1AG058484 and the National Institute of Mental Health (NIMH) grant R01MH115676 to Roel A. Ophoff.

## CURRICULUM VITAE

### EDUCATION

| | |
|---|---|
| 2019 - Current | PhD Candidate, Neuroscience<br>University of California, Los Angeles l Los Angeles, CA |
| 2015 - 2019 | BS, Cellular and Molecular Biology<br>University of Puerto Rico, Rio Piedras l San Juan, PR |

### PUBLICATIONS

**Francia, Marcelo**; Bot, Merel; Boltz Toni; De la Hoz, Juan; Boks, Marco; Kahn, Rene; Ophoff, Roel. (2023). Fibroblasts as an *in vitro* model of circadian genetic and genomic studies. bioRxiv : the preprint server for biology. 10.1101/2023.05.19.541494v2. *Submitted and under review for publication*.

**Francia, Marcelo**; Ramesh, Naren; Reus. Lianne M.; Boltz, Toni; van der Flier, Wiesje M.; Visser, Pieter Jelle; van der Lee, Sven; Teunissen, Charlotte E.; Pijnenburg, Yolande A.L.; den Braber, Anouk; Loohuis, Loes Olde; Tjims, Betty M.; Ophoff, Roel. (2024). The integrative analysis of Cerebrospinal Fluid biomarkers of Alzheimer's Disease, Metabolomics, and Genetic risk reveals novel metabolite associations. *Manuscript in preparation*

Reus, Lianne; Boltz Toni; **Francia, Marcelo**; Bot, Merel; Ramesh, Naren; Koromina, Maria; van der Flier, Wiesje F.; Jelle Visser, Pieter; van der Lee, Sven; Tjims, Betty M.; Teunissen, Charlotte E.; Olde Loohuis, Loes; Ophoff, Roel. (2023). Quantitative trait loci mapping of circulating metabolites in cerebrospinal fluid to uncover biological mechanism involved in brain-related phenotypes. bioRxiv : the preprint server for biology. 10.1101/2023.09.26.559021. *Submitted and under review for publication.*

# INTRODUCTION

Complex brain disorders, including Alzheimer's disease (AD), other dementias, and mood disorders such as bipolar disorder (BD), are estimated to be the leading cause of overall disease burden worldwide, contributing significantly to the increase in disability-adjusted life-years [1]. The profound impact of these conditions on global health underscores the pressing need for advancements in diagnosis and treatment. However, their heterogeneous etiology and diverse symptomatology present challenges in both accurate diagnosis and effective treatment. AD, as the foremost cause of dementia [2], is a neurodegenerative disorder characterized by available biomarkers and well-established genetic risk factors. In contrast, BD, a heritable but highly polygenic psychiatric disorder without known biomarkers. These disorders exemplify the broad spectrum of challenges inherent in researching the underlying biology of complex brain disorders.

Clinical diagnosis criteria for AD is based on the disease's progression stages, ranging from mild cognitive impairment to dementia [3]. Dementia manifests as a progressive cognitive decline, impairing an individual's ability to function independently [4]. Along this spectrum, various cognitive deficits emerge, primarily affecting learning and memory recall. Additionally, non-amnestic presentations include deficits in language, spatial cognition, object agnosia, and impaired face recognition [3]. Currently, there are no disease-modifying treatments for AD [5]. In the case of BD, it is categorized along a spectrum based on the severity and duration of mood fluctuations [6]. This spectrum includes bipolar type 2, characterized by depressive and hypomanic episodes, and bipolar type 1, marked by at least one manic episode [7]. Clinical assessment of a bipolar depressive episode employs the diagnostic and statistical manual of mental disorders (DSM) [8]. Symptoms may include hypersomnia, psychosis, catatonia, and psychomotor retardation [9,10]. On the other hand, a manic episode is defined by elevated mood, increased motor activity, and impaired social or occupational functioning, sometimes accompanied by psychotic symptoms. Screening for manic episodes can be conducted using

tools like the mood disorders questionnaire [11]. The primary distinction between manic and hypomanic episodes lies in their duration, with hypomania lasting at least four consecutive days and mania persisting for at least one week [7]. Some individuals may only experience a single mood episode, while others may undergo more than four within a year, a phenomenon known as rapid cycling [12]. Pharmacological treatment for BD depends on the patient's current mood state [13]. For mania, treatment options include lithium, mood stabilizing anticonvulsants, and antipsychotic medications, while classic antidepressants are typically prescribed for depression [14]. The current assessment approaches for complex brain disorders often lead to diagnosis which may change over the course of the disorder, causing delays in proper treatment and worsening health outcomes. This issue is present in AD [15], and particularly in BD. BD frequently faces changes of diagnosis with other mental disorders such as depression, anxiety disorders, schizophrenia, and obsessive-compulsive disorders [16].

The identification of quantifiable biological features, known as biomarkers, holds promise for improving diagnostic criteria, categorizing disease subtypes, and guiding treatment selection. Biomarkers encompass various quantitative features like circulating proteins, gene variants, or combinations thereof, which may correlate with the underlying biology and predict different aspects of an illness [17]. The National Institutes of Health classifies biomarkers based on their representation of disease components or associated treatments [18]. These include diagnostic biomarkers, biomarkers for monitoring disease status, and biomarkers predicting or tracking treatment response, among others. It's important to note that while biomarkers may overlap in the features they capture, a diagnostic biomarker may not necessarily reflect treatment outcomes [19]. For instance, high-density lipoprotein serves as both a prognostic and susceptibility biomarker for atherosclerosis, yet it does not accurately track improvements after pharmacological treatment [20]. Additionally, endophenotypes, a subset of biomarkers, represent heritable, quantifiable biological traits reflecting genetically relevant aspects of a disease's heterogeneous pathophysiology [21]. Initially conceived as state-independent traits,

manifesting regardless of illness presence, endophenotypes have expanded to include developmental and environmental influences as well [22]. Biomarkers have proven particularly valuable in oncology, driving precision medicine approaches. In clinical practice, biomarkers are indispensable for assessing cancer risk, screening, early detection, accurate diagnosis, prognosis, and predicting treatment response [23]. For instance, in lung cancer, epidermal growth factor receptor (EGFR) mutations play a significant role in tumorigenesis. Genetic mutations of EGFR serve as biomarkers, facilitating the identification of patients likely to respond to EGFR inhibitors [24].

The application of biomarkers for complex brain disorders has faced limitations however, due to their diverse etiology and symptom presentations. Progress has been made in identifying biomarkers for AD, such as cerebrospinal fluid values for Amyloid Beta 42, total tau, and phosphorylated tau 181 [25]. These biomarkers can also be evaluated using positron emission tomography (PET) scans, revealing specific deposition patterns in the brain associated with disease progression, particularly with tau [26]. Despite advancements, a standardized biological definition for AD, known as the ATN framework (A stands for Amyloid, T for tau, and N for neurodegeneration) [27], which combines key biomarkers—beta Amyloid deposition, pathologic tau, and neurodegeneration—has not yet achieved successful clinical segregation of AD patients [28]. In contrast, identifying biomarkers for BD has proven more challenging due to the diverse mood episode presentations and high comorbidity rates with other psychiatric disorders [29]. While brain-derived neurotrophic factors [30] and measurements of cortical activity through neuroimaging have shown promise in specific clinical cohorts [31], their broader application in BD research still requires rigorous assessment and validation.

Another valuable avenue for gaining insight into complex brain disorders in the last 15 years has been through the exploration of genetic variants [32]. Genome-wide association studies (GWAS), which analyze associations between traits and hundreds of thousands of genetic variants across the genome [32], offer a means to uncover the underlying biology of a

phenotype and infer potential causal relationships between risk factors and health outcomes. An illustrative example of GWAS's discovery potential is evident in the research by Zhang et al. [34]. Through genome-wide association analysis of more than 50,000 individuals, they identified a novel blood pressure locus. This locus encoded a previously uncharacterized thiamine transporter, SLC35F3, and the associated genetic variants were found to predict gene expression imbalances and disturbances in cardiac function.

AD is highly heritable, with twin studies estimating its heritability between 60% and 80% [35,36]. The latest GWAS on AD, involving 111,326 clinically diagnosed and proxy cases, along with 677,663 controls, identified 75 risk loci encompassing 31 genes. These genes play roles in processes like innate immunity and microglial activation [37]. This study also identified an overlap between the risk loci associated with AD and other neurodegenerative diseases, such as Parkinson's disease, frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS). Despite its polygenic nature, alleles of the APOE gene stand out as the strongest genetic risk factors for AD. This association has remained the most recognized common genetic risk factor for AD susceptibility since it was first identified in 1993 [38], fueling subsequent genetic association studies and highlighting the potential of genome-wide association studies for common disorders. While the main function of the APOE gene has been identified as lipid transport mediation [39], there are currently no FDA-approved therapies targeting the apolipoprotein E for AD treatment [40]. Similarly, BD is a highly heritable disorder, as twin studies have yielded heritability estimates between 60% and 85% [41,42]. The largest GWAS on BD to date, comprising 41,917 BD cases and 371,549 controls, identified 64 risk loci, associated with 161 genes significantly enriched for brain-specific expression and involvement in synaptic signaling pathways [43]. Interestingly, a number of the proteins encoded by these genes are also targets of medications such as antipsychotics and antidepressants. Moreover, this research revealed high genetic correlations between BD and schizophrenia, major depression, as well as moderate correlations with anorexia, attention deficit/hyperactivity disorder, and autism spectrum disorder.

The findings from these GWAS have offered valuable insights into the pathophysiology of AD and BD, emphasizing their complex polygenic nature and genetic overlap with other complex brain disorders. However, while the identified genes and biological pathways provide crucial knowledge, their clinical validation is necessary to fully understand their impact on these conditions. Moreover, the majority of these genetic variant associations represent very small odds ratios, rendering them unsuitable as biomarkers for individual diagnosis. To navigate the genetic complexity and heterogeneity of these disorders, using endophenotypes, which leverage associated measurable phenotypes, proves to be a promising approach.

Various endophenotypes have been proposed for AD, particularly those linked with APOE allele status [44]. These include neuroimaging measures of regional hypometabolism, structural gray and white matter integrity, cerebral blood flow, as well as amyloid and neurofibrillary tangle load [45]. Similarly, for BD, proposed endophenotypes encompass altered neuroimmune states, indicated by pro- and anti-inflammatory cytokines, circadian rhythm instability, and neuroimaging measures of white matter abnormalities and fronto-limbic disconnection [46]. However, these endophenotypes are not unique to BD, as they are also found to be associated with schizophrenia and major depressive disorder [47–49]. Furthermore, GWAS results have revealed that the genetic basis of these endophenotypes can be as complex as the traits themselves [50,51]. While endophenotypes have not led to significant gene discoveries for disease risk, they still offer valuable insights into the development and prognosis of complex brain disorder. Additionally, they can contribute to understanding gene function and shared etiology across different diagnoses [52].

As part of my dissertation, my research focuses on enhancing the discovery of biologically relevant information for endophenotypes and potential biomarkers in both neurological and psychiatric disorders by integrating data from multiple layers of biology. Given the complexity and heterogeneity of these disorders, relying on a single biomarker measurement is not sufficient. Instead, an integrated approach combining various measurements from different biological layers could unveil biomarker patterns, identify

biological targets, and inform the development of new treatments [53,54]. For instance, the detection sensitivity of bladder cancer recurrence was significantly improved by integrating cytological markers and DNA probes [55]. Such integrative approaches have gained momentum in recent years, particularly with the advancement of high-throughput technologies [54]. Across various cancer types, new prognostic methods have emerged by combining mRNA expression data, methylation data, and genetic variation status of cancer risk genes [56]. The central dogma of biology underscores this intricate interplay across layers, wherein DNA is transcribed into RNA, translated into protein products [57], and the resulting biological pathways are reflected by metabolites. Therefore, by amalgamating omics information—such as genomics, transcriptomics, epigenomics, and metabolomics—representing different biological layers, there is a potential to uncover novel biological features and pathways implicated in complex brain disorders.

Chapter 1 delves into the assessment of functional genomic features observed in skin fibroblasts when utilized as an in vitro model for circadian studies. Fibroblasts are a well-established in vitro model for measuring circadian patterns [58]. Despite initial studies on identifying circadian abnormalities associated with BD yielding inconclusive results [59], ongoing research continues to explore this approach [60–62]. Our objective was to investigate the underlying genetic architecture of circadian rhythm in skin fibroblasts, aiming to evaluate its contribution to the polygenic nature of BD risk. To accomplish this, we collected temporal functional genomic features over a 48-hour period from transcriptomic data (RNA-seq) and open chromatin data (ATAC-seq) obtained from primary cell lines of six healthy individuals. Subsequently, we characterized the biological pathways activated in this in vitro circadian model, assessing the relevance of these processes within the context of the genetic architecture of BD and other disorders. Additionally, we highlighted the limitations of this approach and discussed its potential future applications for circadian genomic studies and as a tool for studying circadian abnormalities.

Chapter 2 investigates how distinct biological layers, including genetics and metabolites, contribute to capturing various aspects of AD pathology as reflected in its established cerebrospinal fluid (CSF) biomarkers. CSF serves as a unique source for studying ongoing biological pathways in the brains of patients with neurodegenerative disorders, such as AD. Notably, metabolites in the CSF can reflect intricate biological processes and hold the potential to assess the collective impact of a disease, identify biomarkers, and track the progression of AD. To achieve this objective, we gathered 5,543 CSF metabolite measurements and genotype data from a cohort of N=477 individuals, including subjects across the AD clinical spectrum, from the Amsterdam Dementia cohort (ADC). Additionally, we obtained CSF measurements for phosphorylated tau (P-Tau), total tau (T-Tau), and Amyloid Beta 1-42 levels. Utilizing the latest AD GWAS, we computed polygenic risk scores (PRS) for AD and examined their associations with CSF levels of P-Tau, T-Tau, and Amyloid Beta 1-42. Furthermore, we conducted correlation and elastic regression analyses between our panel of CSF metabolites and the CSF levels of the AD biomarkers. Through this approach, we identified novel metabolites that correlate with and can predict P-Tau and T-Tau levels. Results from pathway enrichment analysis further supported the involvement of these CSF metabolites in established biological pathways affected by AD. This study builds upon previous research in CSF metabolomics by confirming earlier findings and revealing novel associations between metabolites and AD. These findings offer promising avenues for further exploration of metabolic pathways affected by AD pathology.

In the final part of my thesis, I consider how the findings of these projects underscore the value of integrating various ‘omics technologies. Moreover, I explore the potential implications of applying these methodologies to other complex brain disorders with limited biomarker information or ongoing research efforts, the limitations of the current approaches and future directions that can address them.

## Chapter 1: Fibroblasts as an *in vitro* model of circadian genetic and genomic studies: A temporal analysis

**Authors**: Marcelo Francia[1], Merel Bot[2], Toni Boltz[3], Juan F. De la Hoz[4], Marco Boks[5], René Kahn[5], Roel Ophoff[2]

1. Interdepartmental Program for Neuroscience, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

2. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA

3. Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

4. Bioinformatics Interdepartamental Program, University of California Los Angeles, Los Angeles, CA, USA

5. Brain Center University Medical Center Utrecht, Department Psychiatry, University Utrecht, Utrecht, The Netherlands.

### INTRODUCTION

It is estimated that the lifetime worldwide prevalence of bipolar disorder (BD) is 1% [63], with an estimated heritability of 60-85% [42]; [41]. Genome-wide association studies (GWAS) of BD are showing a highly polygenic genetic architecture of disease susceptibility with common genetic variants explaining 20% of the heritability [64]; [43]. BD is primarily characterized by shifts in mood, which result in manic or depressive episodes. Clinical studies have associated abnormalities of the circadian system in Bipolar disorder type 1 (BD1) patients as a hallmark component of its pathophysiology, with disturbed sleep quality being identified as an early symptom of manic episodes [65]. Furthermore, dysregulation of sleep and wake cycles during

manic episodes include sleep abnormalities such as decrease in total sleep time, delta sleep, and REM latency [66]. These abnormalities have also extended to other circadian regulated systems such as cortisol levels. Both differences at morning levels of cortisol within BD subjects when compared to controls [67], as well as higher cortisol levels prior to a manic episode [68] have been reported. Despite these findings, the precise mechanisms of altered circadian rhythms in BD remain unclear.

The circadian rhythms synchronize physiological processes with the environment, creating and maintaining an internal 24 hour cycle. The main controller of the circadian cycle in mammals is the suprachiasmatic nucleus (SCN), a brain region located in the basal hypothalamus. It receives environmental cues, also called zeitgebers, such as light information from the retina which is relayed using synaptic and hormonal signaling [69] to the rest of the central nervous and peripheral systems. At the molecular level, the circadian machinery within every cell [70] consists of multiple transcriptional feedback loops, where core circadian genes BMAL1 and CLOCK induce the expression of their own repressors, PER1, PER2, PER3 and CRY1, CRY2. These genes modulate different layers of gene expression, from modifying the chromatin landscape to make certain regions of the genome more or less accessible [71], to post-transcriptional modifications altering the function of the associated proteins at specific times during the day [72]. Although disruptions in the circadian rhythms have been associated with neuropsychiatric traits, specifically in mood disorders [48], the direct interactions between them, as well as the contributions from genomic loci, are to be elucidated.

The localization of the SCN makes direct interaction and collection in humans impossible, with researchers instead using peripheral fibroblast cells to study the molecular and genetic components of this system [73]. These cells receive cortisol as a circadian signal from the SCN, through the hypothalamic-pituitary-adrenal axis (HPA). In order to study circadian rhythms using cell cultures, the cells need to be synchronized. One approach for this is treating

the cells with dexamethasone, which elicits rhythm synchronization between the cells in a culture [73]. Dexamethasone binds to the glucocorticoid receptor, acting on the same pathways through which cortisol regulates circadian rhythms in vivo. This synchronization method has been used in conjunction with luciferase bioluminescence reporter assays to study the molecular dynamics of selected circadian genes in vitro [74]. Studies using these systems have been applied to both sleep disorders and BD. Although researchers were able to find differences in the period of expression of circadian genes in sleep disorders [75], similar studies using cells derived from BD1 patients were unable to detect significant [59] or replicable[76] differences.

Here we examine the genomic components of circadian related genetic regulation and general biology that this in vitro fibroblast model captures, and assess whether these features relate to the genetic architecture of BD susceptibility. For this, we collected longitudinal temporal sequencing data of both gene expression and accessible chromatin regions. The temporal gene expression was used to identify genes that display circadian oscillations and are under glucocorticoid control, as well as genes with distinct temporal patterns representing other biological pathways. The temporal accessible chromatin data was used to identify regions of the genome and associated transcription factor motifs that are implicated in the temporal regulation of gene expression. Finally, we examined whether the genomic regions showing temporal transcriptomic and epigenomic circadian profiles in primary cultures of fibroblasts were enriched in genetic association signals of BD or other related psychiatric and sleep-related phenotypes.

## RESULTS

### Temporal RNA-seq captures genes with distinct longitudinal expression patterns

Outside of the subset of genes that compose the core circadian transcriptional feedback loop, most rhythmic genes are tissue specific [77]. Within fibroblasts, we aimed to identify the overall longitudinal patterns of all the genes that are temporally regulated and classify them based on

their temporal features. For this purpose, we collected RNA-seq data every 4 hours for a 48-hour period, from cell cultures of 6 human primary fibroblasts that were derived from a skin biopsy of subjects with no psychiatric disorders. To select these subjects, we confirmed that their cell lines displayed measurable circadian oscillations via a bioluminescence assay (Figure S1.2). After quality control, the temporal RNA-seq dataset consisted of n=11,004 genes. We used a cubic spline regression model to identify genes that had a significant effect of time in their expression ([78]; [79]; [80]). This approach identified n=2,767 (~25%) genes with significant evidence (False discovery rate (FDR) < 0.05) for temporal changes of gene expression levels. To cluster these genes according to distinct temporal patterns we applied the Weighted Gene Co-Expression Network Analysis (WGCNA)[81], which identifies genes with highly correlated expression levels. WGCNA produced 11 modules with eigengene values that captured the principal time patterns present in the expression of these genes (i.e: temporal modules; Figure S1.3). Gene ontology (GO) analysis of WGCNA modules with MetaScape [82] highlighted specific cell processes associated with distinct temporal patterns among 11 of these modules. Figure 1.1 depicts the eigengene values of 4 temporal modules with significant enrichment of GO terms (FDR adjusted by Benjamini-Hochberg method). Genes in the turquoise module, which show a linear decrease in expression over time, had GO terms for supramolecular fiber organization (p= 1e-15) and mRNA splicing via spliceosome (p= 2.5e-13). In compasion, genes in the blue module, which show a linear increase in expression, had a GO term for cellular response to hormone stimulus (p= 1.3e-9). The genes in the black module, which show an increase in expression that plateaus by the 16 hour time point (28 hours after dexamethasone treatment), were enriched for chromatin organization(p= 1e-12) and transcription elongation by RNA polymerase II (p= 7.9e-8) GO terms. The genes in the brown module, which show an expression pattern opposite of the black module, had a GO term for intracellular protein transport (p= 1e-18). Lastly, the purple module, which has genes with a peak expression at the

12 hour time point (24 hours after dexamethasone treatment), had GO terms for cell division (p= 1e-67) and mitotic cell cycle (p= 1e-60). The complete results of GO analysis for all the WGCNA modules are available in Table S1.1.

WGCNA results did not yield a module of co-expressed genes with eigengene values representative of oscillating 24 hrs cycles resembling a circadian rhythm, nor were circadian related functional enrichment of GO terms found in any of the modules. Next we focused on closely inspecting known circadian genes for skin fibroblasts, identified by a previous *in vivo* array-based gene-expression circadian study on human skin cells [83]. Out of the 1,439 circadian genes reported in that study, we identified 267 genes in our dataset with significant changes in expression over time (Figure 1.2A). Using the circadian detection tools JTK Cycle [84], LS [85], ARSER [86], Metacycle [87] and RAIN [88], we aimed to detect significant oscillations within these putative circadian genes in the complete temporal RNA-seq dataset. Among these methods, only JTK and ARSER identified significant periodic expression patterns (after Benjamini-Hochberg correction of 0.05) for the circadian gene *NR1D2*, and further only ARSER identified significant oscillations for 73 genes. However, the predicted period differed between the methods. For example, JTK predicted a period of 27.6 hours for *NR1D2*, while ARSER predicted 24.7 hours (Table S1.2 and supplementary files). Therefore instead of using these circadian detection tools, we applied smoothing-splines mixed effect models using the R package "sme" [89] to model the temporal features of these circadian genes (Figure 1.2B and Figure S1.4). These models showed that for some of these circadian genes, such as *CRY2* and *NFIL3*, the circadian expression pattern is only present in some of the cell cultures, whereas for genes such as *NR1D2* and *TEF*, the circadian pattern is ubiquitous across cell cultures from different individuals. The fitted models for these circadian genes were then used in time warping analysis to group genes with known expression dynamics (Figure 1.2C and Figure S1.4). From these expression patterns, we corroborate that *NR1D2* expression follows its inhibition effect on

*ARNTL* and *CRY1* [90]. Similarly, *PER3* expression follows its inhibition effect with *ARNTL*. Despite observing similar expression patterns in *PER2* and *PER3*, these were not consistent across individuals (Figure S1.4). While the expected inhibition relationship between *CRY1* and *ARNTL* was not present (Figure S1.5), this pattern of expression was also reported in the circadian dataset that was used as reference [83].

**Temporal open chromatin levels measured by ATAC-seq highlights potential regulatory regions and transcription factor binding sites**

To identify regions of the genome associated with the regulation and downstream effects of circadian genes, we collected ATAC-seq data following the same temporal design as with the RNA-seq dataset. Quality control metrics such as fraction of reads in peaks and transcription starting site for these samples is available in the supplementary material. After removing a cell line that did not pass quality control, we merged all overlapping regions of open chromatin, also known as peaks, across samples and time points as described previously [91], to define a common set of ATAC-seq signals (n = 126,057). We then used cubic spline regression models to identify peaks that have a significant change in accessibility over time. This approach yielded n=7,568 (6%) time significant peaks, which were functionally annotated using ChipSeeker [92], a software that annotates peaks with the nearest gene and genomic regions (Figure S1.8B). Peaks with significant changes in accessibility over time showed a similar genomic distribution as the full dataset (Figure S1.8A). Following the approach for the RNA-seq data, we applied WGCNA to cluster peaks with similar temporal patterns of accessibility changes (Figure S1.7).

WGCNA identified 4 different modules for the temporal patterns of chromatin accessibility, however the main pattern that characterizes these modules is an overall increase or decrease in accessibility. One module captured all the regions that were decreasing in accessibility (Figure 1.3A), comprising 4,435 peaks. The other 3 modules showed regions increasing in accessibility. Individual motif enrichment analysis conducted with HOMER [93],

showed similar enrichment across these modules, therefore we combined them into a single cluster of regions increasing in accessibility, in total 3,133 peaks. Regions that were decreasing in accessibility over time (Figure 1.3A) had motif sequences for Fos (p= 1e-1047), Fra1 (p= 1e-1041), ATF3 (p= 1e-1034), BATF (p= 1e-1002), Fra2 (p= 1e-996), AP-1 (p= 1e-976), Jun-AP1 (p= 1e-681), Bach2 (p= 1e-330) and JunB (p= 1e-1001). Most of these transcription factors are part of the AP-1 transcription complex. Regions that were increasing in accessibility over time (Figure 1.3B) had motif sequences for BHLHA15 (p= 1e-201), TCF4 (p= 1e-180), NeuroG2 (p= 1e-160), Twist2 (p= 1e-160), Pitx1 (p= 1e-186), Atoh1 (p= 1e-162), Tcf21(p= 1e-147), Olig2 (p= 1e-130), ZBTB18 (p= 1e-139) and NeuroD1 (p= 1e-123). These are dimerizing transcription factors that have the basic helix-loop-helix protein structural motif. In both types of regions HOMER identified the binding sequence of the glucocorticoid response element (GRE), although the rank for the GRE motif in regions that were decreasing in accessibility was higher. For the known circadian transcription factors, HOMER identified significant enrichment of the binding sequences for BMAL1 (p= 1e-29), NPAS2 (p= 1e-9), CLOCK (p= 1e-10), particularly within regions that had increasing accessibility over time. For the regions with decreasing accessibility over time, HOMER identified enrichment of NFIL3 (p= 1e-11).

**Stratified Linkage Disequilibrium Score Regression (sLDSC) analysis**

Functional annotation of the ATAC-seq dataset showed that approximately one third of the peak regions identified are located in distal intergenic regions, with unknown functions. Furthermore, it also showed that these regions displaying transient changes in chromatin state are located across the entire genome. To examine whether these open chromatin regions highlighted in our study are enriched for genetic susceptibility of BD and other neuropsychiatric traits, we used sLDSC (stratified linkage disequilibrium analysis)[94] to calculate the partitioned heritability of these features. For this approach we used published Psychiatric Genomics Consortium (PGC)

summary statistics for BD [43], ADHD (Attention-Deficit/HyperactivityDisorder) [95], schizophrenia [96], PTSD (Post-traumatic stress disorder) [97], MDD (Major depression disorder) [98], insomnia [99], and the circadian trait of morningness [100]. We used the temporally significant ATAC-seq regions with 1 kilobases (kb) and 10 kb genomic windows in both downstream and upstream directions for each region. These ATAC-seq defined annotations were tested jointly with the baseline annotations included with sLDSC [94]. Figure 4 shows the enrichment for the traits tested from the ATAC-seq regions annotations as well as the baseline annotations (Full enrichment results are provided in the Supplementary Material). Among these, only the ATAC-seq regions that were decreasing in accessibility had a nominally significant (p value = 0.00463) less than expected presence for ADHD, and this effect was not present when the regions are extended by either 1kb or 10kb. In comparison, baseline annotations such as conserved regions in mammals showed a significant enrichment for all the traits (ADHD p value = 7.88e-11, schizophrenia p value = 1.67e-23, BD p value = 1.14e-8, MDD p value = 5.92e-22, insomnia p value = 2.65e-15, morningness p value = 3.05e-29), except PTSD (p value = 0.073). We did not identify significant enrichment of ATAC-seq regions in the other psychiatric and behavioral traits tested, indicating that these genomic regions with temporal trends in cromatin accessibility do not play a major role to their genetic architecture.

**DISCUSSION**

Cell cultures of peripheral tissues have been employed as models of *in vitro* circadian clock systems to study their molecular components [101] and the disorders in which they are disrupted [102]. For this particular model that uses primary cultures of fibroblasts derived from skin biopsies, we aimed to characterize the circadian features that are present at gene expression and chromatin accessibility levels, with the goal to identify the circadian genes that are engaged by this system as well as their associated regulatory genomic regions. From the longitudinal RNA-seq data we identified consistent circadian patterns of expression in a limited

amount of genes such as ARNTL, CRY1, PER3, NR1D2 (Rev-erbBeta) and TEF, but observed noticeable differences in the expression patterns between cell cultures. When compared to a recent *in vivo* circadian human skin dataset [83], we identified 267 out of the 1,439 circadian genes previously identified in this tissue to have a significant effect of time in their expression. This limited overlap could indicate that this *in vitro* model for studying circadian rhythms is constrained to the circadian genes that are directly activated by a glucocorticoid-like stimulus. Glucocorticoid response elements have been identified for circadian genes such as PER1, PER2, PER3, CRY1, CRY2, Rev-erbAlpha (NR1D1), Rev-erbBeta (NR1D2), DBP, NPAS2 and BMAL1 [103]. Consistent with these results we identified circadian rhythmicity in most of those genes that had previously been identified to have a glucocorticoid response element (GRE), with the exception of NR1D1. While we do identify expression levels from NR1D1, the lack of a significant circadian oscillation in comparison to the strong results from NR1D2 could be consistent with their known redundant function for circadian rhythms [104].

We found robust circadian expression patterns for NR1D2. However, differences in the lenght of the predicted periods across tools indicates that estimating period duration from longitudinal RNA-seq data is not a straightforward problem. This could be due to the the small number of subjects used, leading to insufficient power. Furthermore, while ARSER identified 73 genes with significant periodic expression patterns, this software is known to have a high false positive rate in high resolution data [87]. Interestingly, while the expression pattern of CRY1 follows the expected inhibition by NR1D2 [105], it does not reflect the expected inhibitory action on ARNTL, nor the similar phase pattern with its heterodimer partner PER3. For the relationship with PER3, CRY1 has been previously reported to have a known phase delay with the PER genes (PER1,2,3)[106], which could be attributed to the multiple binding sites that CRY1 has for different circadian modulators, resulting in stimulus and tissue specific temporal dynamics. Based on the time patterns, changes in the expression of CRY1 appear to precede the

expression of ARNTL by 4 to 8 hours. Similar expression patterns between these genes were reported in a previous *in vivo* study of human skin cells [83], indicating that this model was able to replicate some of the circadian temporal dynamics seen within tissue.

ATAC-seq data can provide an untargeted yet comprehensive view of chromatin accessibility changes over time. Within this fibroblast *in vitro* model, we mainly identified genomic regions with linear increases and decreases of chromatin accessibility. Although we did not found circadian temporal patterns in chromatin accessibility, previous studies conducted *in vivo* have reported such patterns [107]. Specifically, proteins such as CLOCK and BMAL1 have been found to associate and interact with chromatin remodeling and chromatin modifying enzymes [108], as well as act as pioneer factors by directly modifying chromatin accessibility [71]. The absence of anticipated oscillations in chromatin within our dataset, as opposed to our observations in genes, could be due to multiple reasons. One possibility is that the mechanisms governing oscillatory chromatin changes may be exclusive to *in vivo* conditions. Under a physiological setting, cells within a tissue are exposed to multiple stimuli that act as Zeitgebers, such as sunlight (exposure), metabolic signals, temperature, and hormones like cortisol [109]. The exposure to these signals is under a rhythmic control, with levels cycling throughout the day [110]. By using a single exposure to dexamethasone in this model, we are missing the cyclic aspect of cortisol response present *in vivo*, as well as other effects that could be due to the coupling of Zeitgebers. However, our ATAC-seq dataset does replicate previous findings on the broader role of glucocorticoids in the chromatin landscape. Within regions with decreasing accessibility post-dexamethasone, the motif enrichment analyses identified the motifs for the GRE as well as for members of the AP-1 transcriptional complex. These regions may have initially opened due to the dexamethasone treatment (for synchronization of the cells), but are closing without further continued exposure of dexamethasone. Regions with increasing accessibility may have initially closed due to the dexamethasone treatment, and this is

consistent with motif profiles that were unrelated to direct glucocorticoid receptor (GR) targets [111]. The dataset, however, lacks a Chip-seq analysis for GR occupancy, and we are limited to confirm if the identified regions are indeed due to GR activity. The strong glucocorticoid effects observed in our data underscore the need for further exploration of circadian influences on chromatin regulation in fibroblast cell culture models. Other methods for synchronizing these cells and studying the circadian rhythms, such as a switch to serum free media [73], Forskolin treatment [112], or temperature cycles [113], could result in different types of chromatin regulation and gene expression dynamics. This study raises questions about the context-dependent nature of chromatin remodeling events and emphasizes the need to evaluate different synchronization methods to ascertain their implications for circadian rhythms.

The ATAC-seq data showed genome wide transient changes in chromatin conformation, with most of these changes occurring within regions of unknown functions. To evaluate the relevance of these genomic regions for BD and other psychiatric traits, we used partitioned sLDSC regression. This tool identifies genetic susceptibility enrichment for a particular trait across the whole genome and within specific genomic annotations. The partitioned sLDSC analysis mainly showed a significant deflation for the chromatin regions that were decreasing in accessibility over time with ADHD. Although not significant, it mirrored the enrichment for the regions that were increasing in accessibility. When expanding the genomic regions by either 1kb or 10kb both the magnitude and the significance of the enrichment are lost, indicating that this effect could be highly localized for these regions. For the other traits that we examined, the enrichment from the ATAC-seq regions were also attenuated when the genomic region was extended. These results show that these regions with linear changes in chromatin accessibility identified here may not play a relevant role for these neuropsychiatric traits. However the attenuation observed when expanding the genomic window suggests that any relevant signal may be specific to those genomic positions.

The lack of an overlap between the temporal regulatory regions identified in this study and the known genetic architecture of BD could indicate four different interpretations. First, although peripheral tissues capture the genome of an individual, they don't recapitulate brain molecular physiology, the main tissue implicated in the pathophysiology of BD. Second, it could be that the disruptions in the circadian rhythm are not under strong genetic control and are actually influenced by other downstream processes, such as post-translational modifications and differences at the protein level. Third, the specific circadian pathways that are engaged in this *in vitro* model by dexamethasone (i.e: glucocorticoids) are not part of the genetic architecture of BD. This however does not exclude other circadian pathways that could be engaged by a different synchronizing stimulus, such as serum [114], [115]. Furthermore, there are other stimuli that also act in different ways with the circadian system, such as temperature. Whereas dexamethasone acts through binding of the glucocorticoid receptor, temperature affects heat-shock proteins [113]. Lastly, the disrupted circadian phenotype is an episodic state in BD patients, not a constant trait. This could indicate that rather than the regular circadian system being affected by BD, it is the ability to deal with circadian stressors and disruptors that is implicated in BD disease susceptibility.

The circadian analysis of gene expression and chromatin accessibility data faced limitations that could be attributed to variability among cell lines from different subjects. Human skin cell studies, both *in vivo* [83] and *in vitro* [116], have demonstrated that genetic differences contribute to variations in circadian gene expression's phase and amplitude. The variability observed in this dataset was therefore not unexpected, but remains a factor to be considered for this kind of studies. Another potential source of variability was the data collection scheme, involving 13 separate cell cultures for each individual cell line. Distinctions in cell cycle state and growth rates among these cultures might have influenced the data. Previous research has shown that cell cycle and circadian rhythms are coupled processes([117], [118]), and that these

rhythms can be impacted by cell density [119]. Our approach, utilizing a 5% FBS culture that minimizes cell growth, aimed to control for both of these factors. Our lab's prior work also confirmed that the cell density used for our study allows for the production of rhythmic circadian cycles in these cells(Supplementary figure 2). Although various factors known to influence circadian rhythms could have contributed to the variability in this dataset, certain circadian genes appeared resilient, consistently producing rhythmic cycles across cultures and individual cell lines. This could highlight specific genes' resilience to various sources of variation in this kind of studies.

## MATERIALS AND METHODS

### Cell lines and Culture

Fibroblasts were isolated by taking skin biopsies from the nether region from subjects without known psychiatric disorders. Fibroblast cultures were established following standard procedures [120] and stored as frozen aliquots in liquid nitrogen. 6 fibroblast cell lines matched for sex, age and passage number were thawed out and grown to confluence in T75 culture flasks in standard culture media (DMEM containing 10% fetal bovine serum (FBS) and 1x Penicillin-Streptomycin).

Upon reaching confluence, 5x10^4 cells were plated per line into 13 different 6 well plates (1 well per line per plate). All 6 lines were collected in the same experiment for the RNA-seq experiment. Due to the labor-intensive nature of the ATAC protocol and the need to process cells fresh, the 6 lines were split into 2 batches, so 3 lines per batch were processed.

### Assessment of Circadian Expression *in vitro*

In order to collect RNA or cells every 4 hours for 48 hours, cells were split into two batches, which were reset 12 hours apart (see supplementary figure 9). Cells were reset 12 hours before the first collection to exclude the acute effects of dexamethasone and variation in synchronization conditions [116]. 5 days after being plated the cells from batch one were synchronized by treatment with 100 nM Dexamethasone for 30 min. Cells were then washed with PBS and switched to collection media (DMEM containing 5% FBS and 1x Penicillin-Streptomycin). Lower concentration of FBS was used in this media to stop the cells from growing during the experiment, in order to keep all time points at approximately the same culture density. 12 hours later cells from batch 2 were synchronized and switched to collection media and the RNA/cell collection was started (from batch one).

**RNA and Cell collection**

For the collection of RNA, cells were lysed using 350uL RLT lysis buffer from the Qiagen RNeasy mini kit. Lysed cells were then scraped off the plate, transferred to a Qiaschredder (Qiagen 79656) and centrifuged for 2 min at max speed to further homogenize. Cell lysates were kept in -80 until extraction.

For the collection of cells for the ATAC protocol, cells were dissociated using 500uL of prewarmed TrypLE (ThermoFisher 12604013) and left for 5 min at 37°C. TrypLE was inactivated using 500uL of DMEM. Cells were then counted using the Logos Biosystems LUNA-FL automated cell counter, and 50x10^4 cells were used as input for tagmentation. Tagmented DNA for library preparation was collected following the previously described protocol [121].

**RNA extraction**

RNA from cell lysates was extracted using the Qiagen RNeasy mini kit (Qiagen 74106). Cell lysates were extracted in a randomized order to prevent batch effects in downstream analysis. In order to collect total RNA including small RNAs, the standard extraction protocol (Purification of Total RNA from Animal Cells using Spin Technology) was adjusted by making the following changes: (i) adding 1.5 volumes of 100% ethanol, instead of 70%, after the lysis step (step 4 in handbook protocol) and (ii) adding 700 mL of buffer RWT (Qiagen 1067933) instead of the provided RW1 (step 6 in handbook protocol).

**RNA and ATAC sequencing**

For the RNA sequencing, library preps were made using the Lexogen QuantSeq 3' mRNA-Seq Library Prep Kit and sequenced with 65-base single end reads, and sequenced at a targeted depth of 3.8M reads per sample, which is well above the recommended minimum 1M reads per sample read depth for these types of libraries. ATAC seq libraries were generated following the previously described protocol [121] and sequenced with 75-base double end reads, and sequenced at a targeted depth of 61M reads per sample. Library preparation and sequencing was performed at the UCLA Neuroscience Genomics Core (https://www.semel.ucla.edu/ungc). All samples were sequenced on a Illumina HiSeq 4000 sequencer.

**RNA-seq data processing and analysis**

Fastqc [122] software was used to assess the quality of the read files. Low quality reads were trimmed using TrimGalore and Cutadapt.

Alignment of reads was performed with the STAR[122] software and to human gene ensembl version GrCh38. STAR was indexed to the genome using the –runMode genomeGenerate function. For aligning, STAR was run with the parameters –outFilterType BySJout –outFilterMultimapNmax 20 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1 –

outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000.

Samtools was used to index the aligned files from STAR.

Read counts were associated with genes using featureCounts software with the NCBI GRCh38 gene annotation file.

Analysis of the RNA-seq data used the R packages limma, Glimma and edgeR, as previously described [123]. Genes with low read counts were removed and reads were normalized by CPM. GeneIDs were converted to Gene Symbols using the package Homo.sapiens.

WGCNA [81] software was used to classify genes with similar temporal expression patterns. WGCNA was run using a power value of 12 obtained from diagnostic plots and with the "signed" argument. MetaScape was used for Gene Ontology analysis of the resulting gene sets from WGCNA.

Following the method described in [124], MetaCycle [87] was used to run circadian detection tools such as ARSER, JTK [84], LS and metacycle. In order to integrate results from the different individuals, the function meta3d was used from the Metacycle R package. RAIN [87] was run separately.

According to the EdgeR user guide, cubic splines were generated using the splines package in R, with the ns function and 5 degrees of freedom. Resulting p-values were corrected using a false discovery rate of 0.05. Significant genes were then compared to a previously published dataset of circadian human skin gene expression, resulting in 267 genes that were classified according to their time series using the “dtwclust” R package. The resulting clusters are available in the supplementary material. This analysis was also performed with only the known circadian clock genes that had consistent expression patterns across the cell-lines.

**ATAC-seq data processing and analysis**

Sequenced open chromatin data from the ATAC-seq assay followed the standard ENCODE Pipeline for the identification of open chromatin regions (OCRs) of the genome. The steps included using fastqc to evaluate the quality of the sequenced library. Followed by trimming of low quality reads with Trimgalore and Cutadapt. Alignment of the raw reads data to human gene Ensembl version GRCh38 was performed using bowtie2 [125] with a 2kb insert size and allowing up to 4 alignments. Reads within black-listed regions alongside PCR duplicates were removed with samtools. MACS2 [126] software was used to identify OCRs with parameters -g hs -q 0.01 –nomodel –shift -100 –extsize 200 –keep-dup all -B. PCR. Quality control metrics for the ATAC-seq dataset such as peak counts, PCR bottlenecking coefficients, fraction of reads in peaks and enrichment of transcription starting site are provided in the supplementary material.

To compare the ATAC-seq signal across timepoints and subjects, we created a consensus bed file using the bedtools[127] function merge function, combining all the overlapping peak regions across timepoints and subjects into a single file. The Featurecounts software was then used to assign read counts to those regions. Read counts were normalized by RPM.

WGCNA [81] software was used to classify peaks with similar temporal accessibility patterns. WGCNA was run using a power value of 12 obtained from diagnostic plots and with the "signed" argument, identifying 4 modules after merging.

HOMER [93] findMotifsGenome.pllp program was used to identify enriched transcription factor motifs first individually in the peaks that belonged to the largest WGCNA modules, as well as in the resulting set of grouping all the modules that displayed a similar increasing or decreasing pattern of accessibility.

**Stratified linkage disequilibrium score regression analysis**

Following the procedure described in [128], we applied an extension to sLDSR, a statistical method that partitions SNP-based heritability(h2) from GWAS summary statistics [93]. We ran sLDSR (ldsc.py –h2), using an ancestry-match 1000 Genomes Project phase 3 release reference panel, for each annotation of interest while accounting for the full baseline model, as recommended by the developers ([93]; [129]), and an extra annotation of all the ATAC-seq detected in our *in vitro* model (n=3126 for peaks that were decreasing in accessibility, n=4415 for peaks increasing in accessibility), as well as extension of these regions by 1kb and 10kb genomic windows in both directions.

## CHAPTER 1 FIGURES

**Figure 1.1. Eigengene values for RNA-seq modules obtained from WGCNA.**

Eigengene modules from WGCNA of the longitudinal temporal expression patterns of RNA-seq data collected every 4 hours for a 48 hour period. Each color represents a fibroblast cell culture from a different individual. Module names were assigned by WGCNA. The number of genes assigned per module is indicated next to the module name.

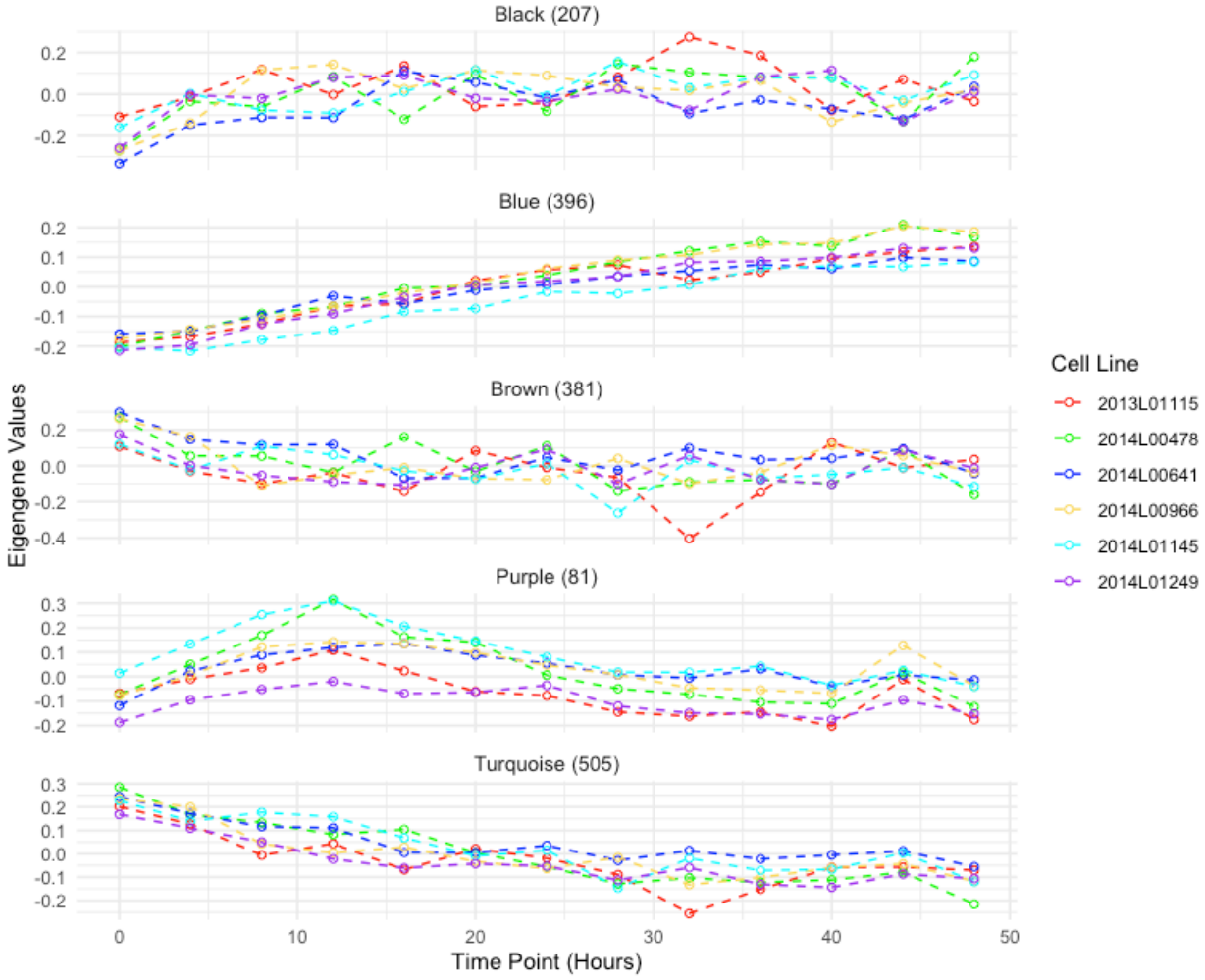



**Figure 1.2. Expression patterns and mixed non-linear modeling of circadian genes.**

(A) Overlap of the genes that were found to have a significant effect of time in their expression as well as being previously identified as circadian within the skin tissue [83]. In bold are those genes that displayed expression patterns consistent with circadian rhythms.

(B) Example of smoothing-splines mixed-effect model of the genes that displayed circadian oscillations and their known redundant gene partners. The red area indicates the 95 percent confidence interval. Gene expression values are presented as lCPM (log of counts per million)

(C) Dynamic time warp clustering results of the 9 circadian clock genes identified in this *in vitro* model. Due to the differences in magnitude of gene expression across circadian genes, expression values were z-scored. Clustering results for all 267 genes can be found in the supplementary File 1.3.

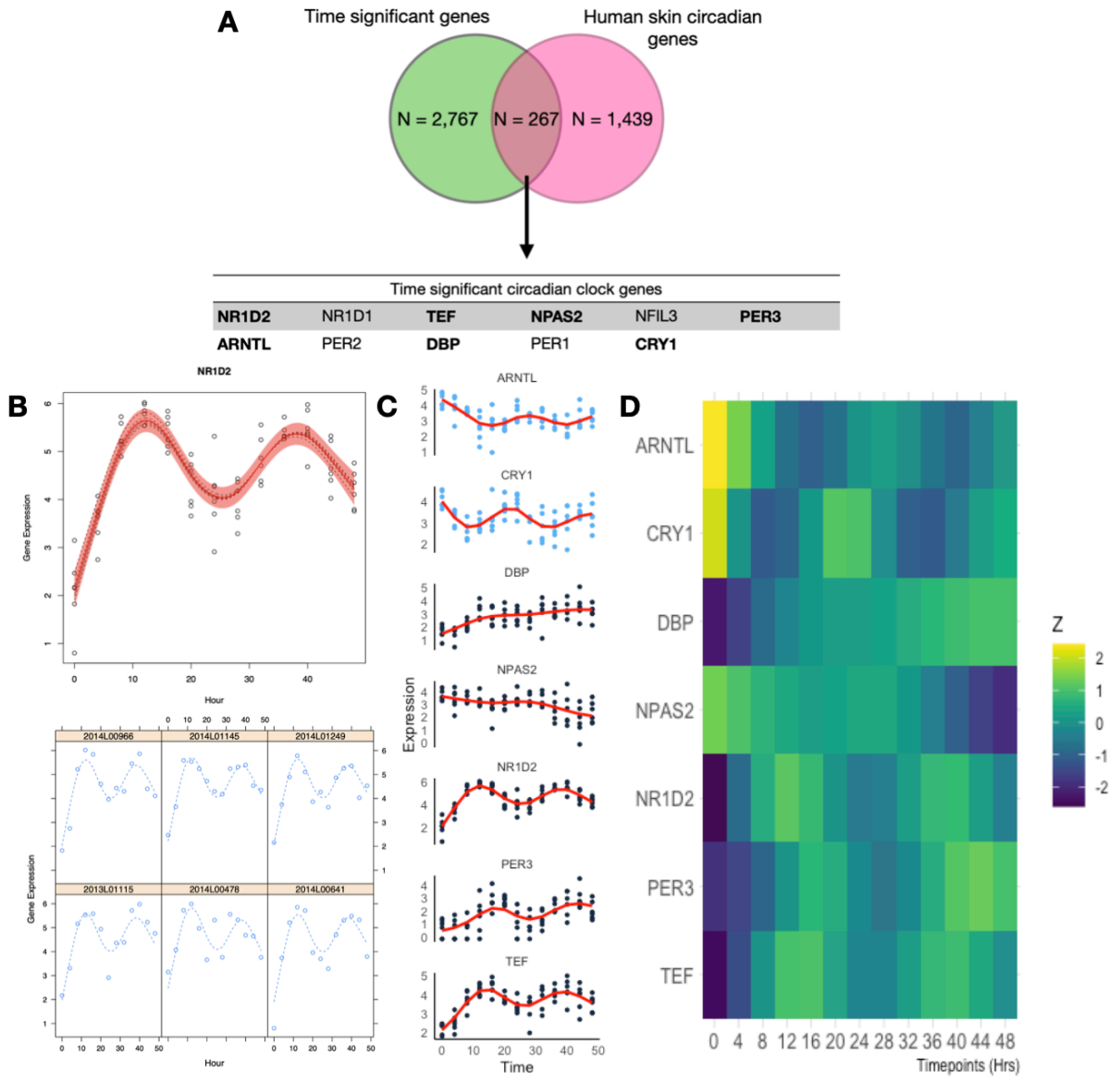
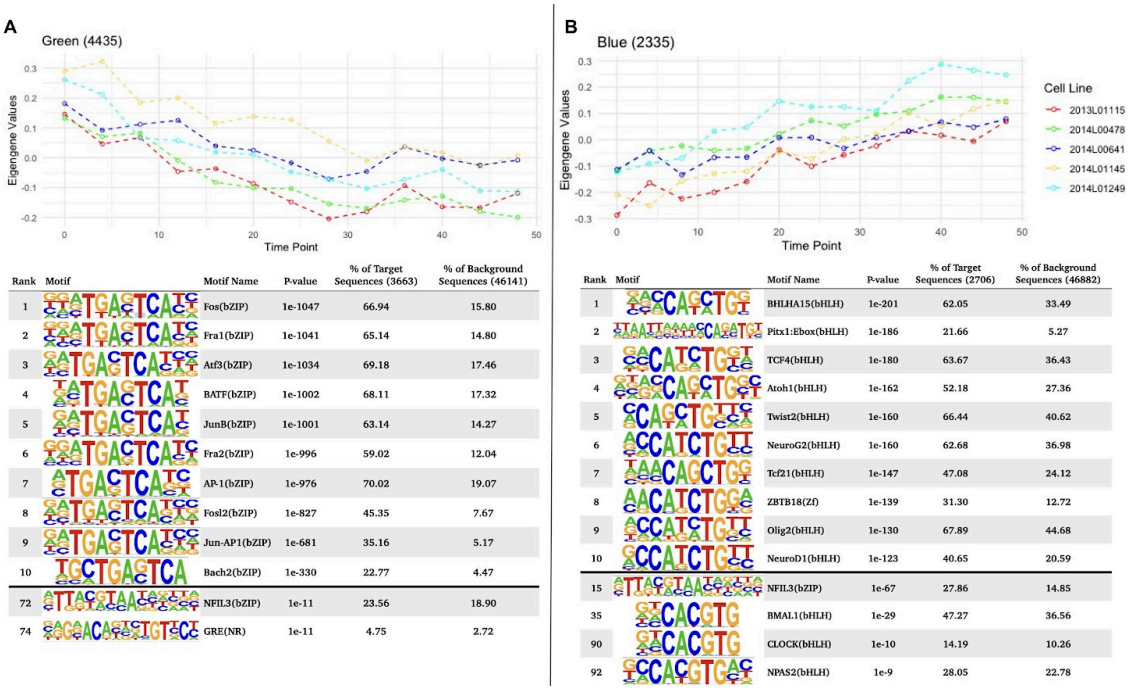



**Figure 1.3 Motif enrichment analysis of time significant peak regions.**

Motif enrichment analysis results after combining all open chromatin regions that followed a similar change in accessibility over time, including the top 10 motifs as well as motifs associated with circadian genes and glucocorticoid response. P-values were confirmed as significant after the Benjamini adjustment cutoff of 1% FDR.

(A) Eigengene values for the Green WGCNA module and motif enrichment analysis of the associated decreasing in accessibility regions.

(B) Eigengene values for the Blue module and motif enrichment analysis of the associated peak regions, including HOMER results for all peak regions decreasing in accessibility.



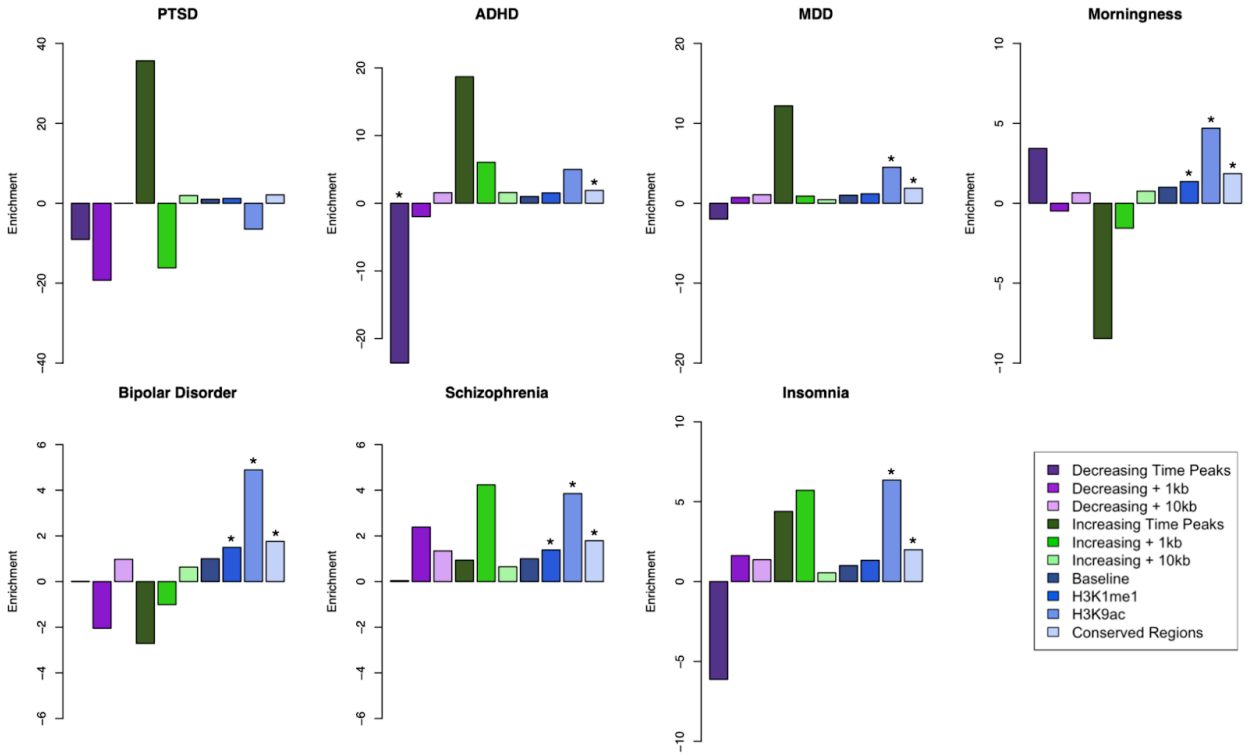

| Rank | Motif | Motif Name | P-value | % of Target Sequences (3663) | % of Background Sequences (46141) |
|---|---|---|---|---|---|
| 1 | | Fos(bZIP) | 1e-1047 | 66.94 | 15.80 |
| 2 | | Fra1(bZIP) | 1e-1041 | 65.14 | 14.80 |
| 3 | | Atf3(bZIP) | 1e-1034 | 69.18 | 17.46 |
| 4 | | BATF(bZIP) | 1e-1002 | 68.11 | 17.32 |
| 5 | | JunB(bZIP) | 1e-1001 | 63.14 | 14.27 |
| 6 | | Fra2(bZIP) | 1e-996 | 59.02 | 12.04 |
| 7 | | AP-1(bZIP) | 1e-976 | 70.02 | 19.07 |
| 8 | | Fosl2(bZIP) | 1e-827 | 45.35 | 7.67 |
| 9 | | Jun-AP1(bZIP) | 1e-681 | 35.16 | 5.17 |
| 10 | | Bach2(bZIP) | 1e-330 | 22.77 | 4.47 |
| 72 | | NFIL3(bZIP) | 1e-11 | 23.56 | 18.90 |
| 74 | | GRE(NR) | 1e-11 | 4.75 | 2.72 |

| Rank | Motif | Motif Name | P-value | % of Target Sequences (2706) | % of Background Sequences (46882) |
|---|---|---|---|---|---|
| 1 | | BHLHA15(bHLH) | 1e-201 | 62.05 | 33.49 |
| 2 | | Pitx1:Ebox(bHLH) | 1e-186 | 21.66 | 5.27 |
| 3 | | TCF4(bHLH) | 1e-180 | 63.67 | 36.43 |
| 4 | | Atoh1(bHLH) | 1e-162 | 52.18 | 27.36 |
| 5 | | Twist2(bHLH) | 1e-160 | 66.44 | 40.62 |
| 6 | | NeuroG2(bHLH) | 1e-160 | 62.68 | 36.98 |
| 7 | | Tcf21(bHLH) | 1e-147 | 47.08 | 24.12 |
| 8 | | ZBTB18(Zf) | 1e-139 | 31.30 | 12.72 |
| 9 | | Olig2(bHLH) | 1e-130 | 67.89 | 44.68 |
| 10 | | NeuroD1(bHLH) | 1e-123 | 40.65 | 20.59 |
| 15 | | NFIL3(bZIP) | 1e-67 | 27.86 | 14.85 |
| 35 | | BMAL1(bHLH) | 1e-29 | 47.27 | 36.56 |
| 90 | | CLOCK(bHLH) | 1e-10 | 14.19 | 10.26 |
| 92 | | NPAS2(bHLH) | 1e-9 | 28.05 | 22.78 |

**Figure 1.4 sLDSC enrichment results for psychiatric disorders and a circadian trait.**

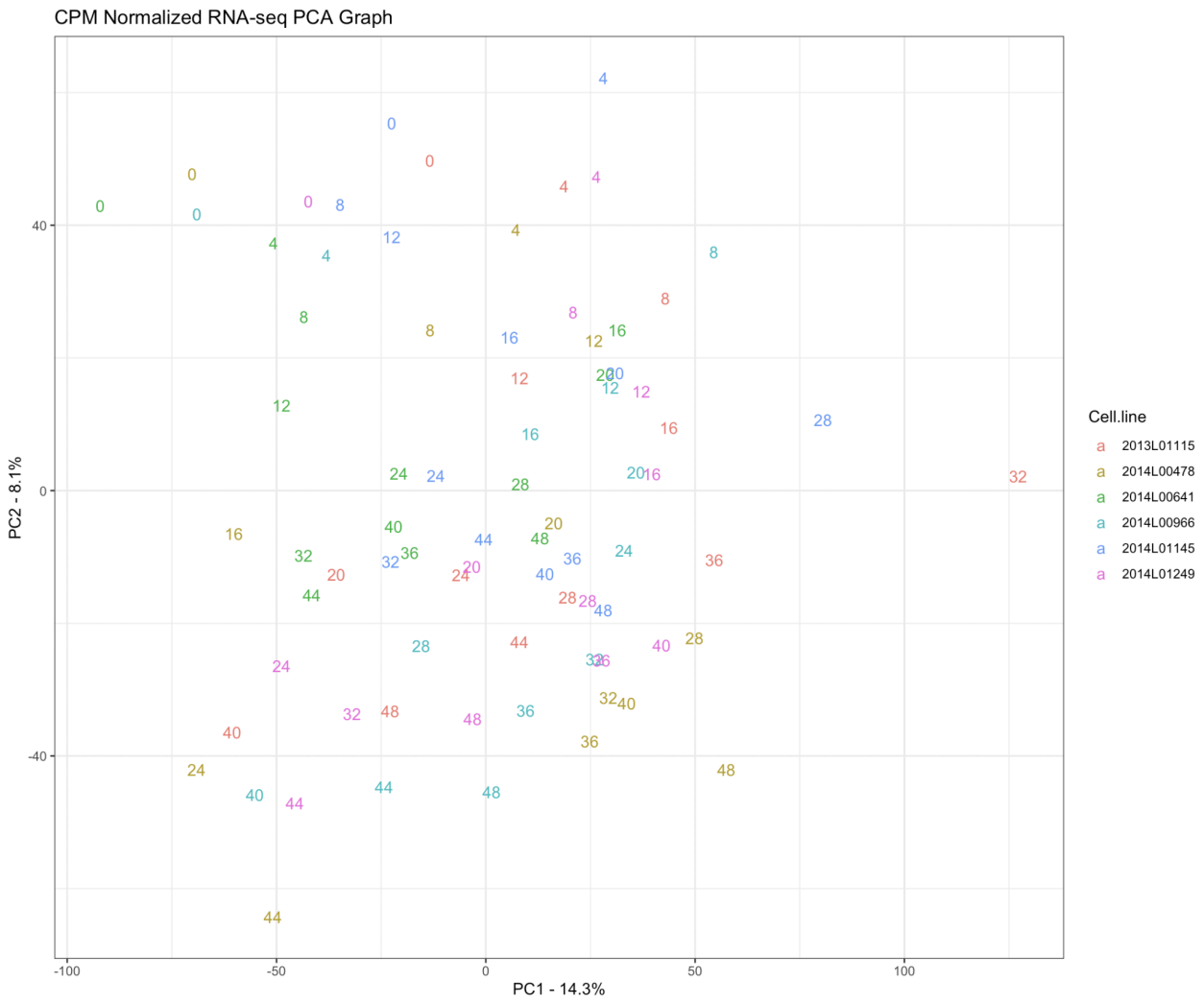
Results of partitioned sLDSC across 3 psychiatric disorders and morningness trait across different genomic annotations. Shown are the enrichment for both the temporal ATAC-seq regions and extended genome windows; as well as annotations part of the baseline model of sLDSC, such as baseline for all the annotations, histone markers H3K9ac, H3K4me1, and conserved regions in mammals. PTSD: Post-traumatic stress disorder; MDD: Major depressive

disorder; ADHD: Attention-Deficit / Hyperactivity Disorder. * indicates enrichment p-value below 0.01.




## CHAPTER 1 SUPPLEMENTARY FIGURES

**Supplementary Figure S1.1 Principal component analysis of RNA-seq temporal dataset.** Principal component component representation of RNA-seq data after CPM(counts per million) normalization. Colors indicate each individual cell line used in this study. Numbers in the plot indicate the corresponding time point for that cell line.

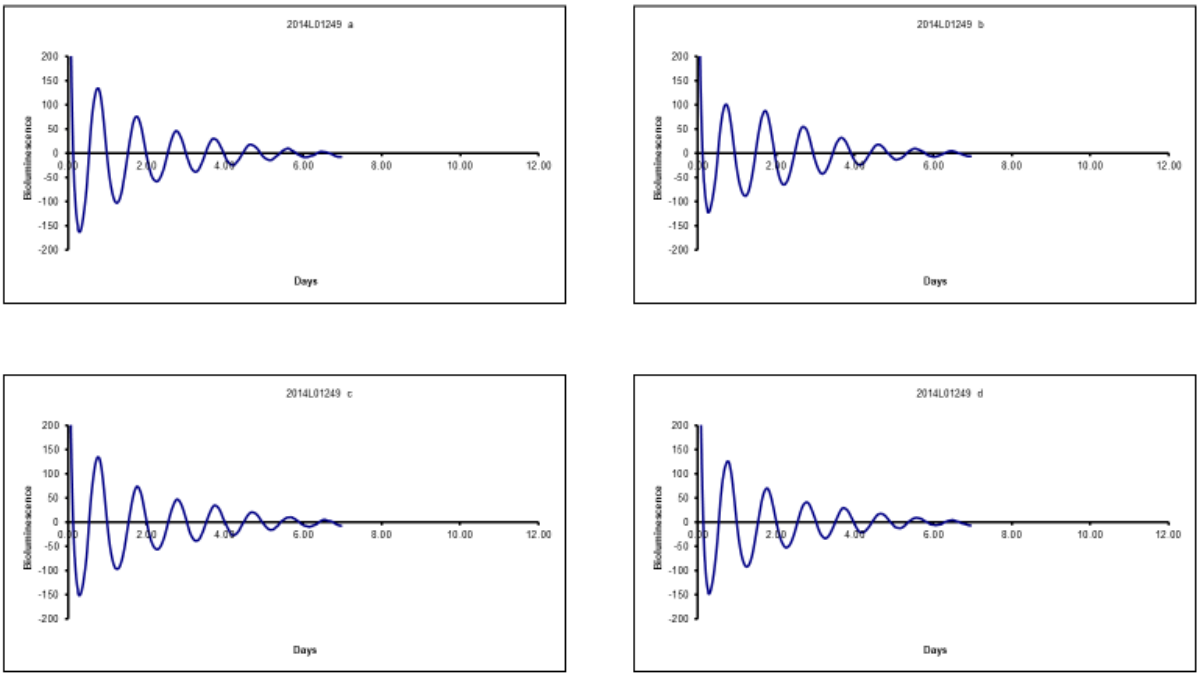



**Supplementary Figure S1.2 Circadian-bioluminescence transduction experiment results.** Example of circadian bioluminescence assay performed using transduced primary fibroblast cell lines with a Bmail1:luc construct, as described previously by Brown, et al. 2005 [116].

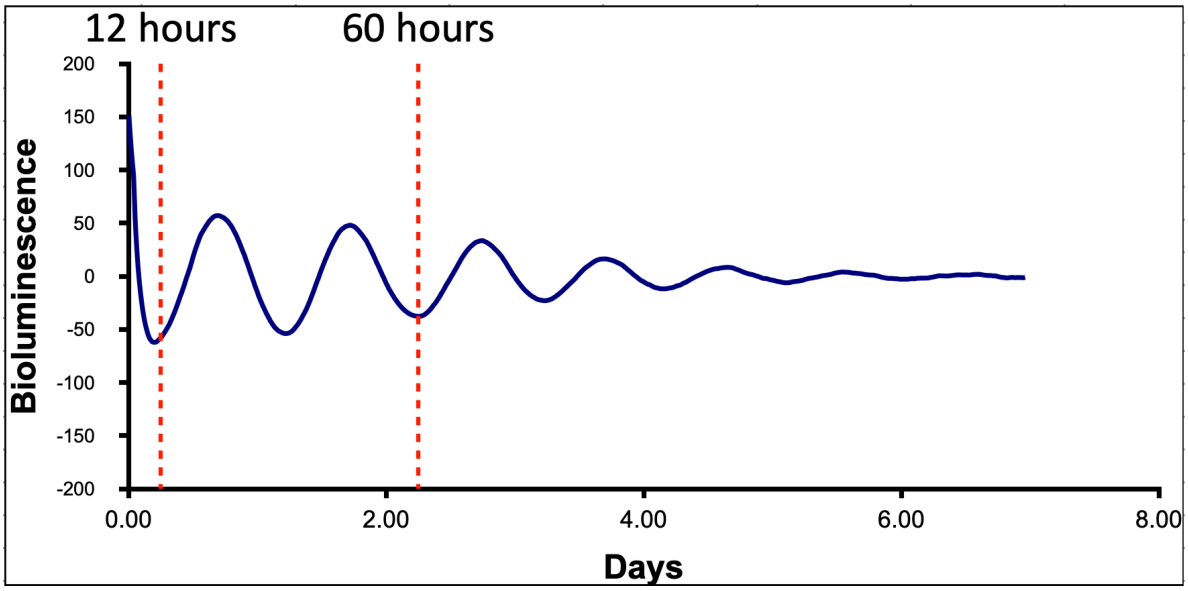
(A) Bioluminescence measurements of cell cultures following synchronization with dexamethasone.

(B) Red lines indicate the time period during which RNA-seq and ATAC-seq data was collected from the cell lines.

**A**

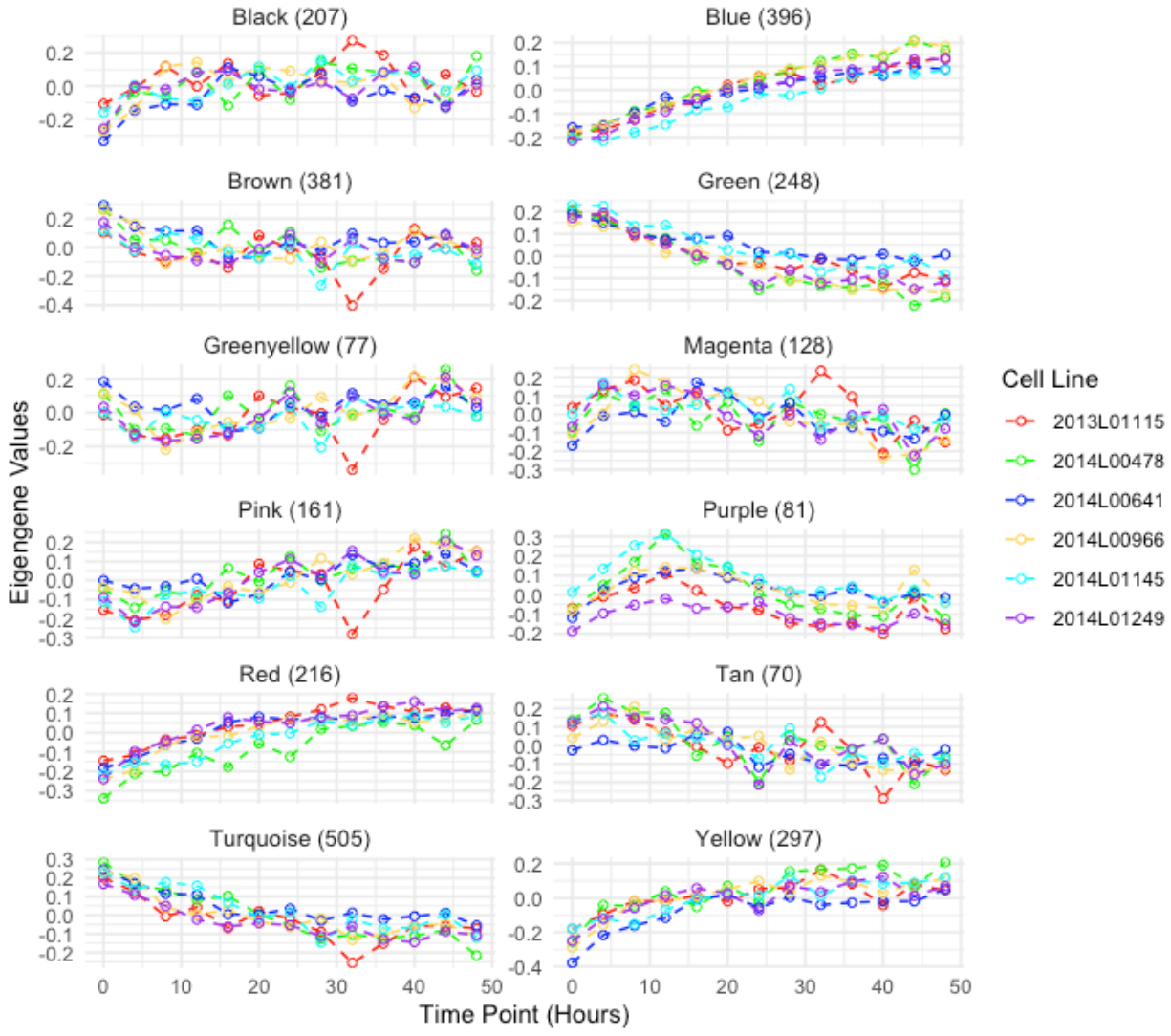

2014L01249 a
2014L01249 b
2014L01249 c
2014L01249 d
Bioluminescence
Days

**B**


12 hours
60 hours
Bioluminescence
Days
200
150
100
50
0
-50
-100
-150
-200
0.00
2.00
4.00
6.00
8.00

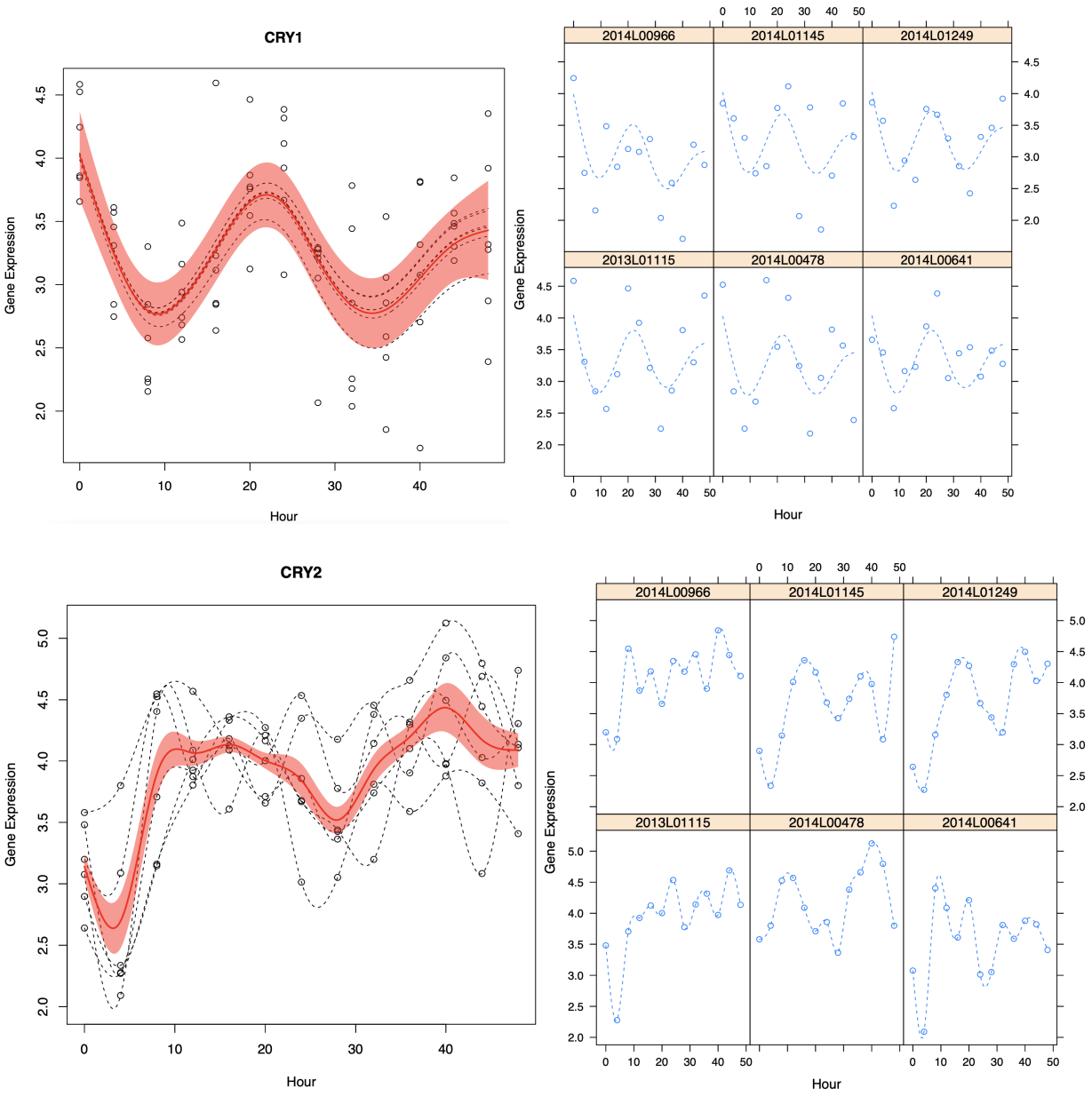
**Supplementary Figure S1.3 WGCNA modules obtained from the RNA-seq temporal dataset.**

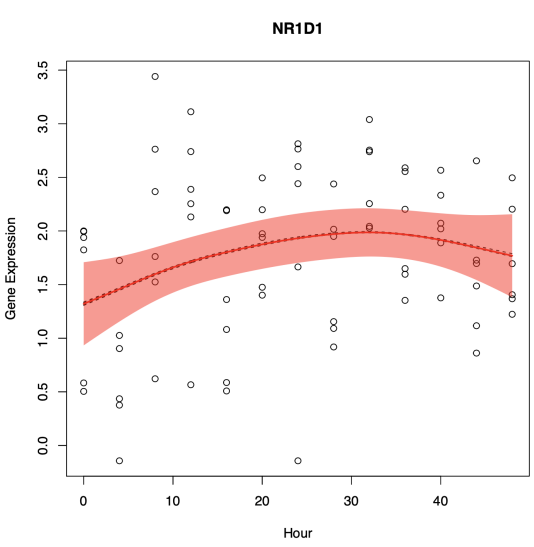
WGCNA modules obtained from n=2,767 genes identified as having a significant effect of time in their expression through cubic splines modeling. Expression values are represented here as eigengene values. Names of the modules were assigned by WGCNA. The number of genes assigned per module is next to the module name.

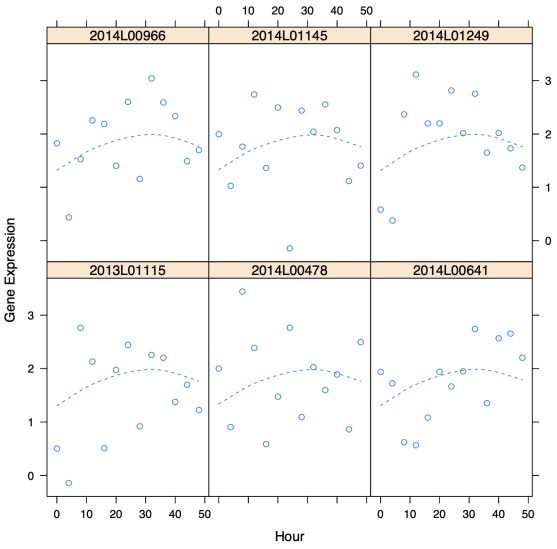


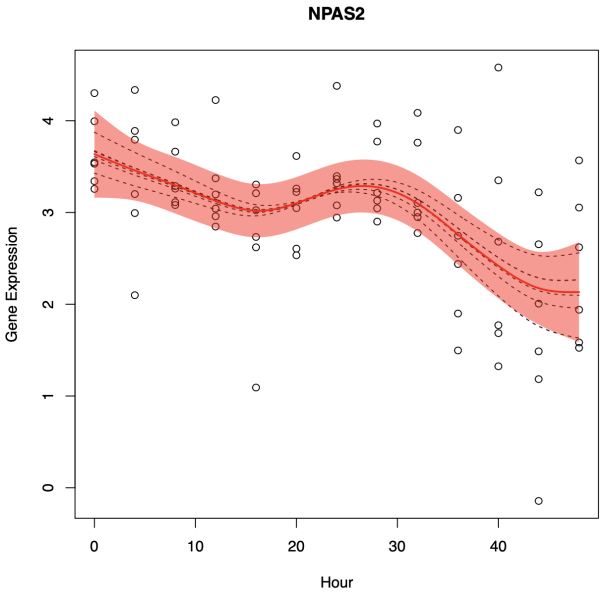

**Supplementary figure S1.4 Mixed non-linear modeling of circadian genes.**

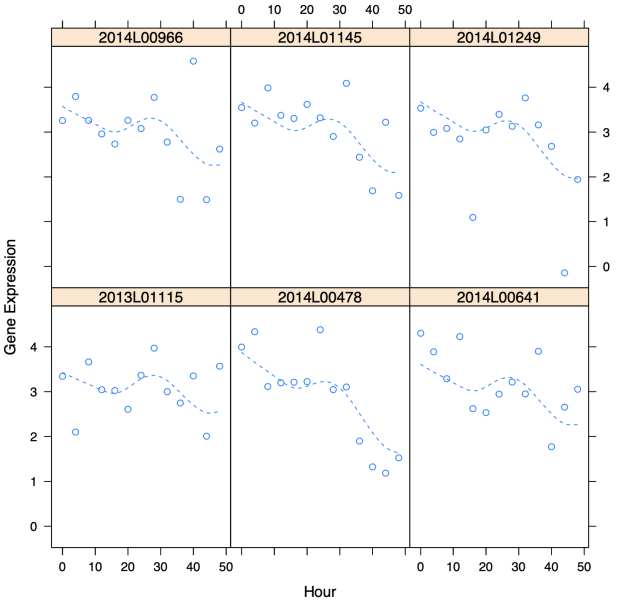
Smoothing-splines mixed-effect models of circadian gene expression across previously reported circadian genes in skin. Red line indicates the average fitted model across cell lines, with the red area representing a 95% confidence interval.

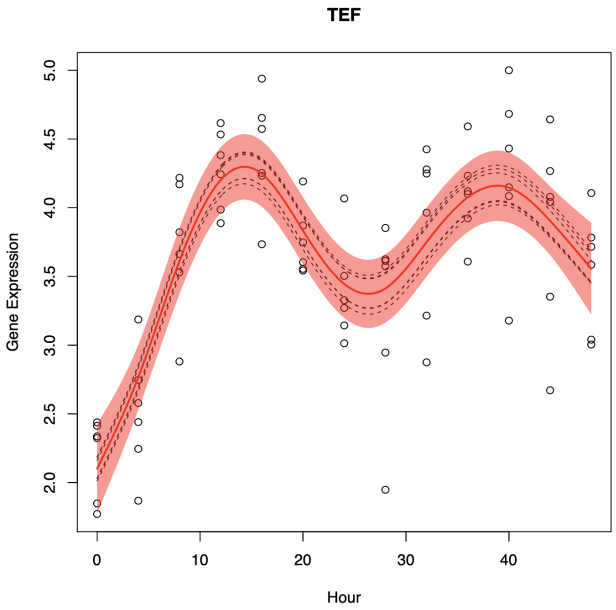


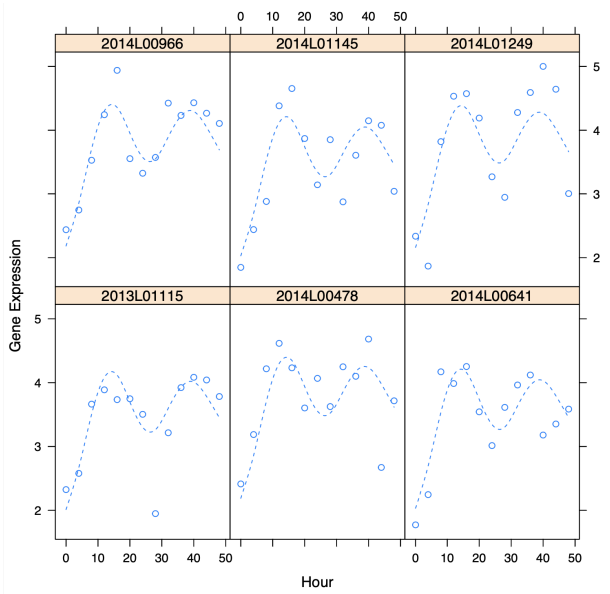

NR1D1
Gene Expression
Hour

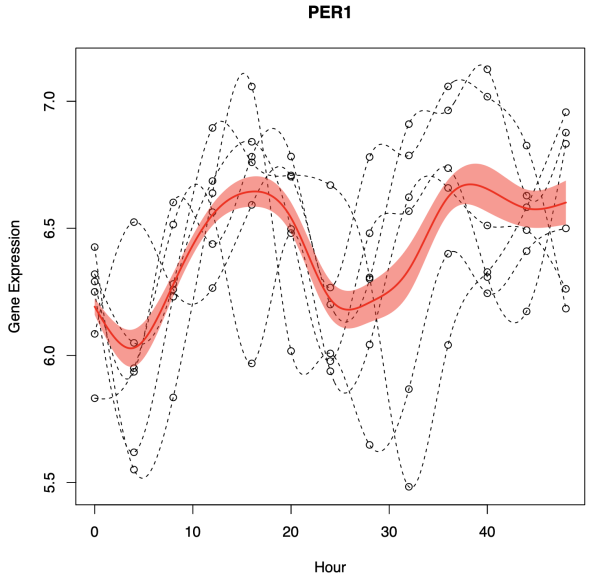

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

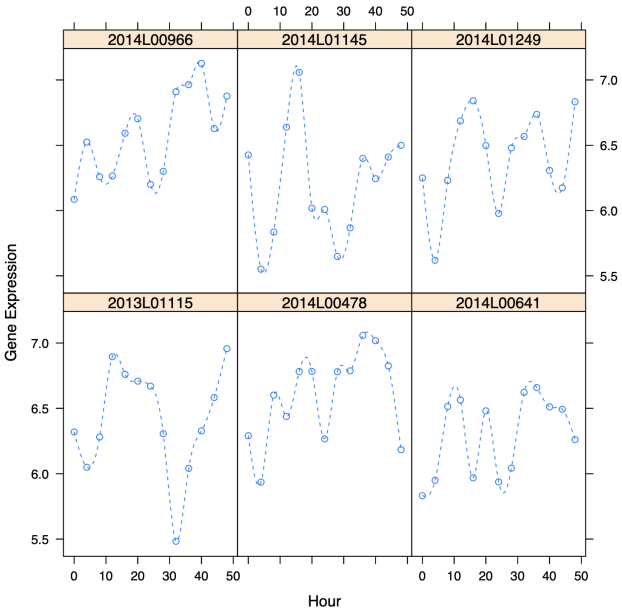

NPAS2
Gene Expression
Hour

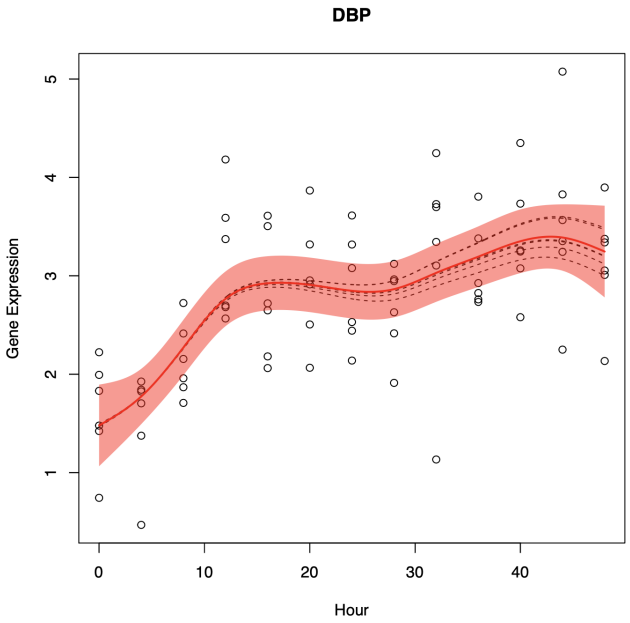

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

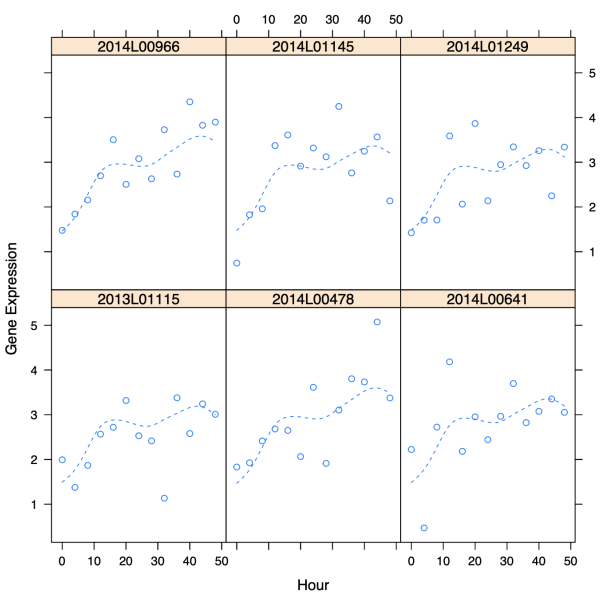

TEF
Gene Expression
Hour

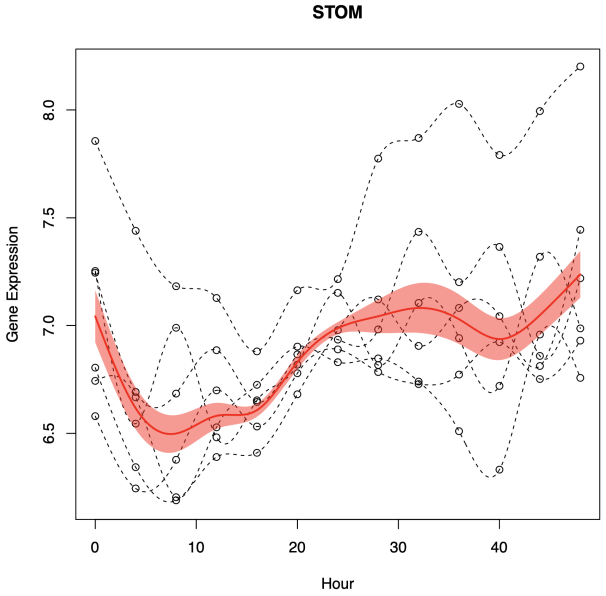

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

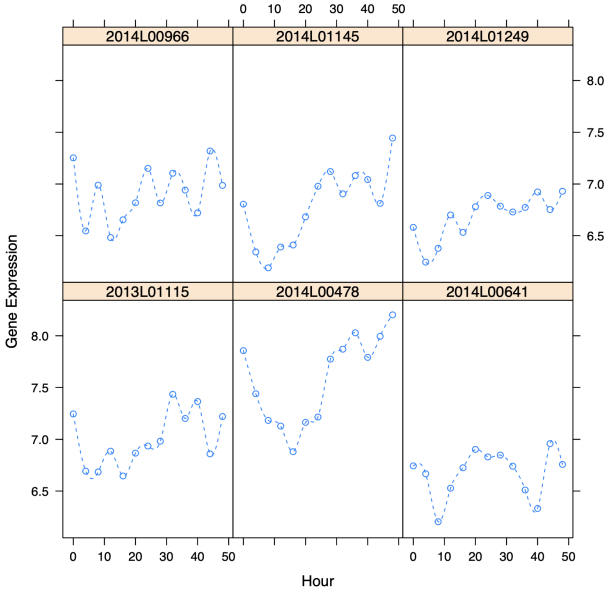

PER1
Gene Expression
Hour

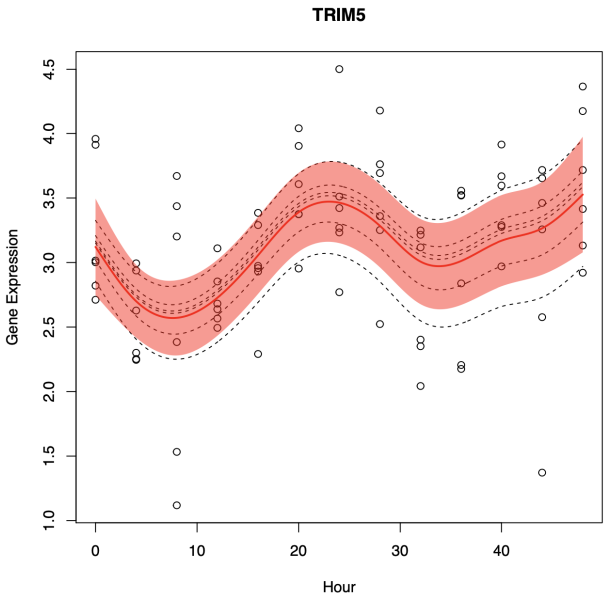

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

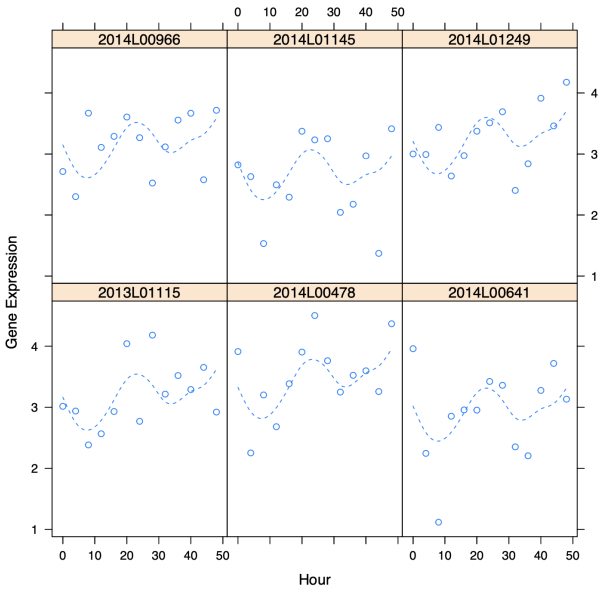

DBP
Gene Expression
Hour

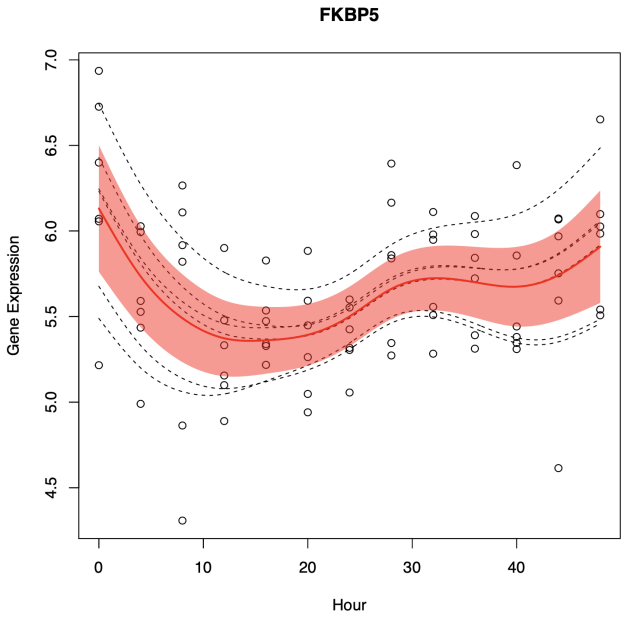

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

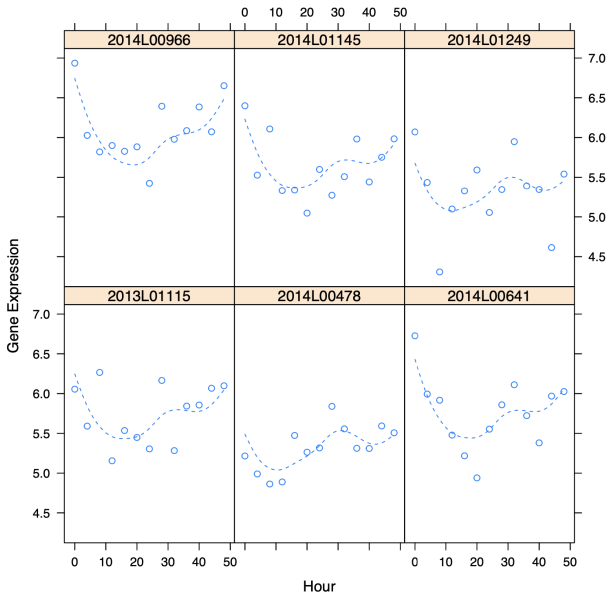

STOM
Gene Expression
Hour

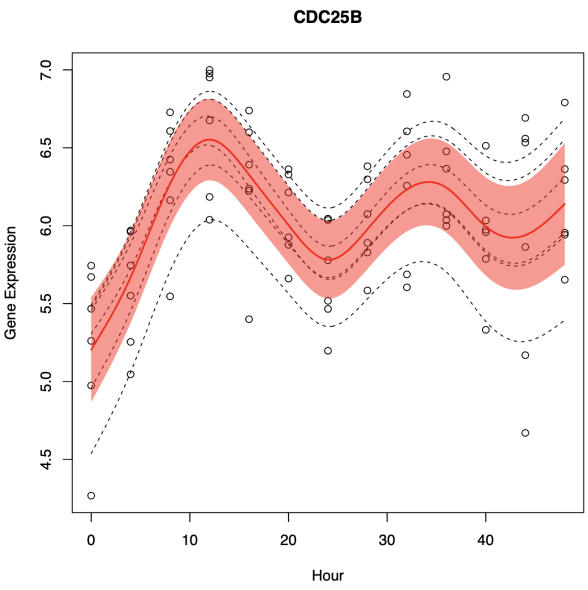

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

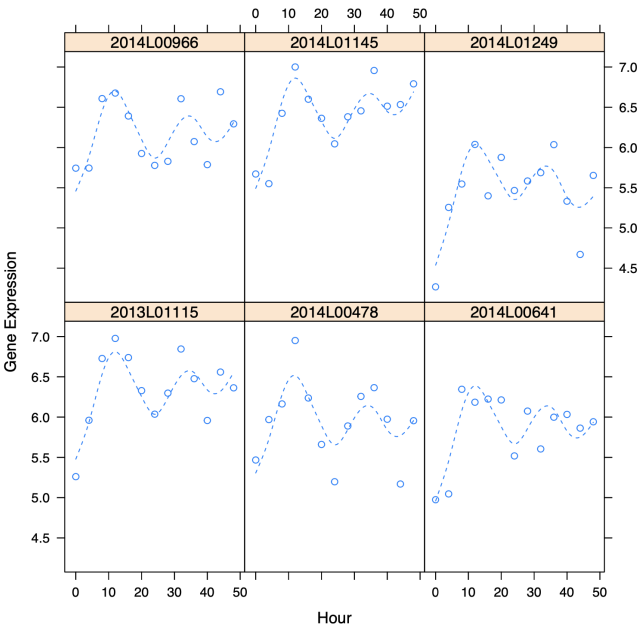

TRIM5
Gene Expression
Hour

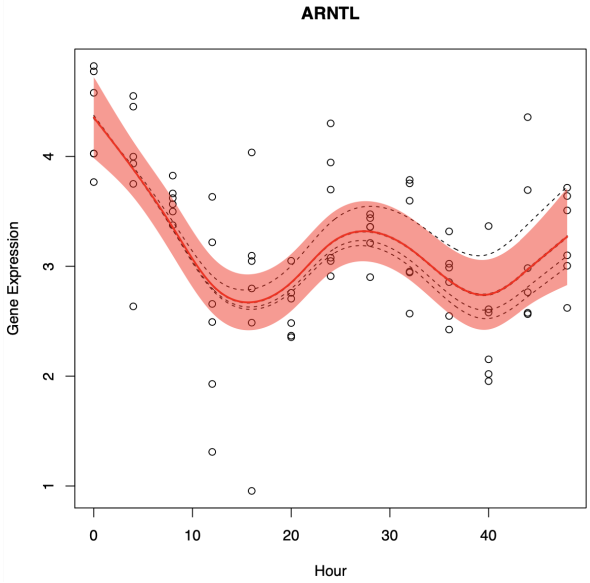

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

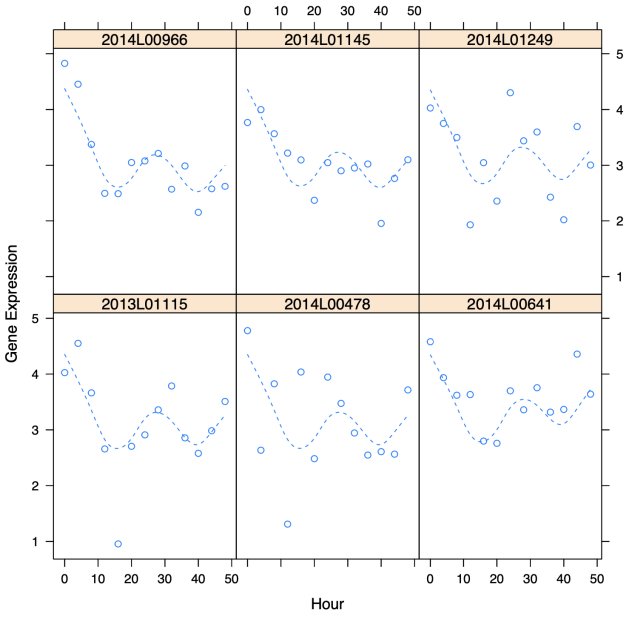

FKBP5
Gene Expression
Hour

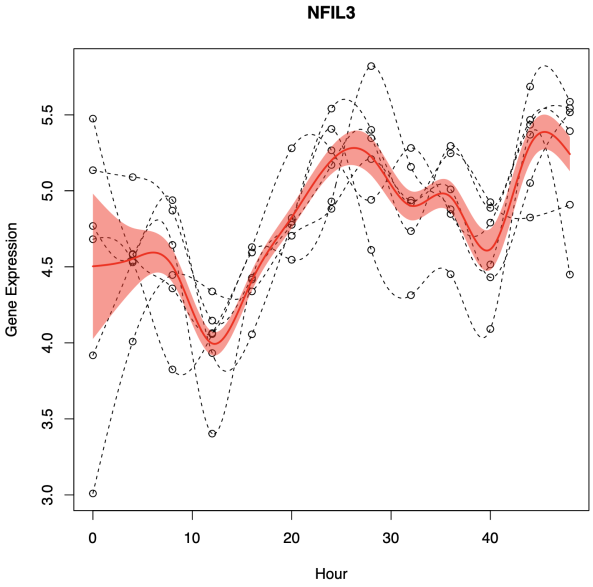

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

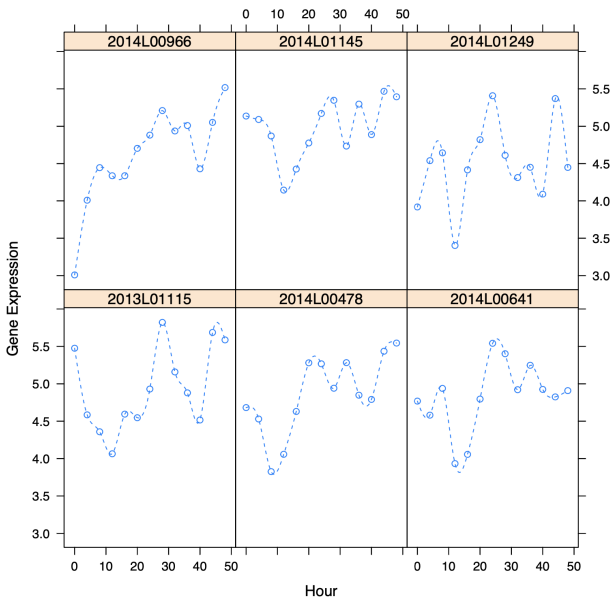

CDC25B
Gene Expression
Hour

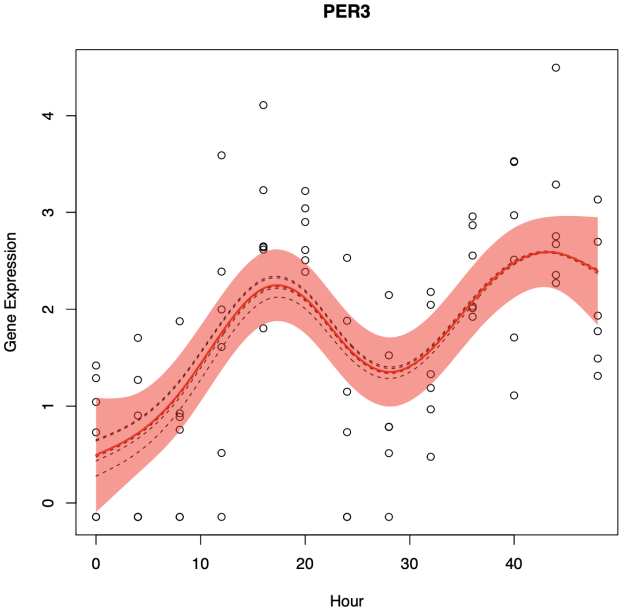

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

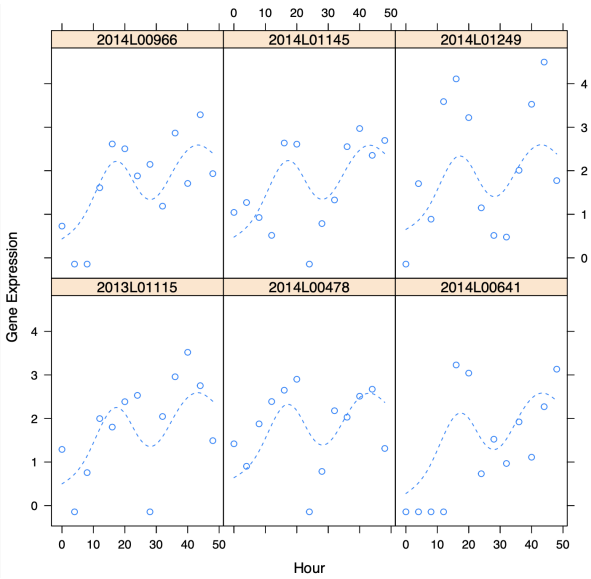

ARNTL
Gene Expression
Hour

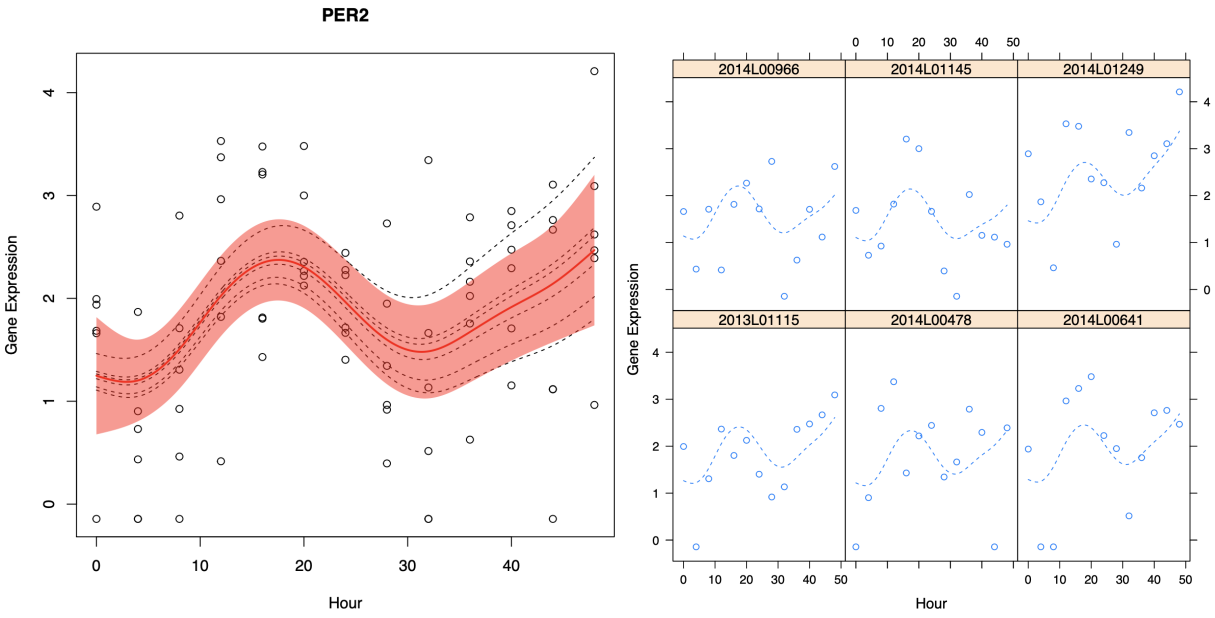

2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour


NFIL3
Gene Expression
Hour


2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour

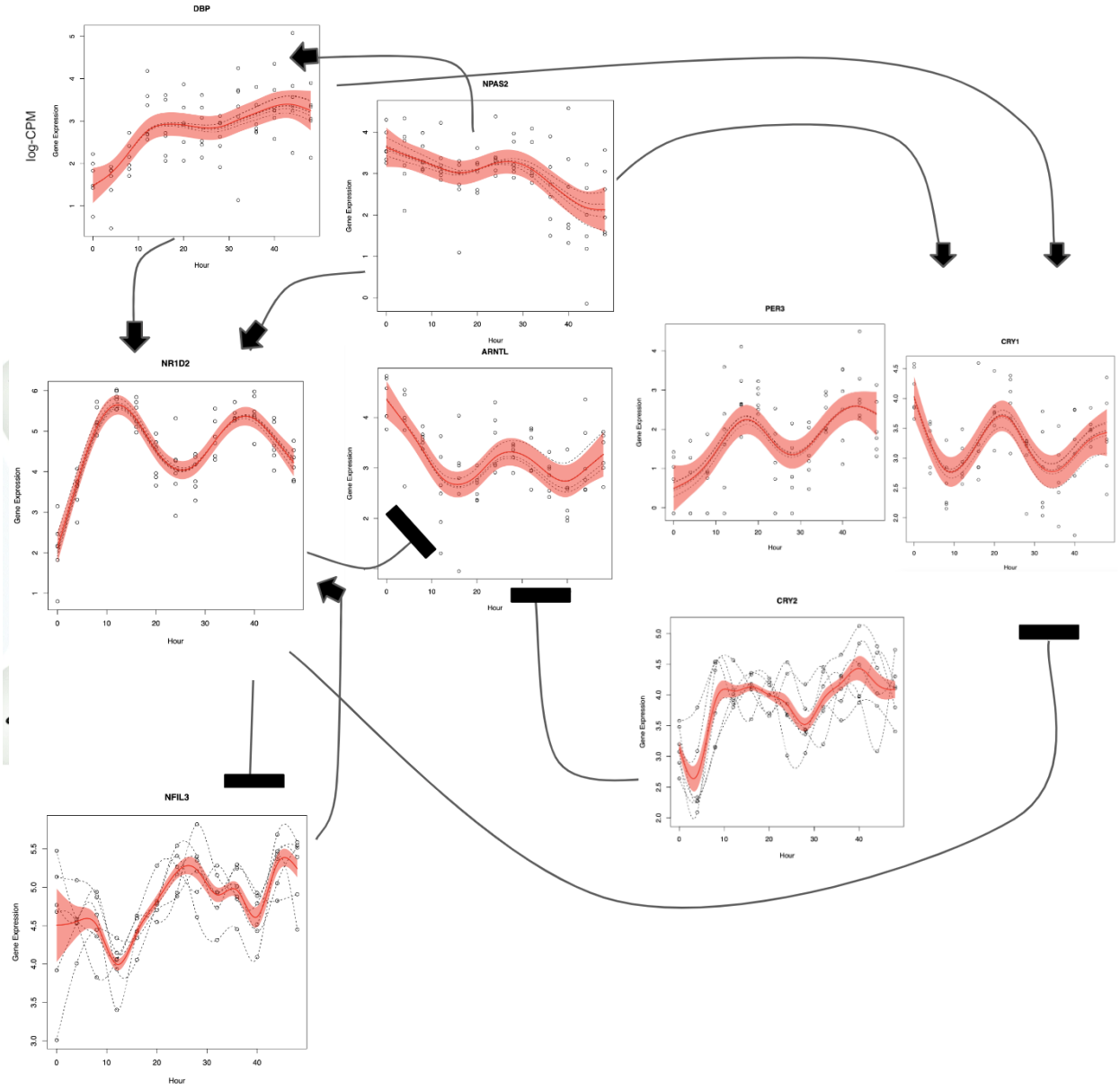

PER3
Gene Expression
Hour


2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
Gene Expression
Hour


PER2
Gene Expression
Hour
0
1
2
3
4
0
10
20
30
40
2014L00966
2014L01145
2014L01249
2013L01115
2014L00478
2014L00641
50

**Supplemental Figure S1.5 Known interactions between circadian genes present identified in this dataset.**

Known gene expression relationships of core circadian genes. Arrows indicate a gene inducing in the expression of another gene. The black bar indicates a gene repressing the expression of another gene.



**Supplemental Figure S1.6 Quality Control for ATAC-seq data**

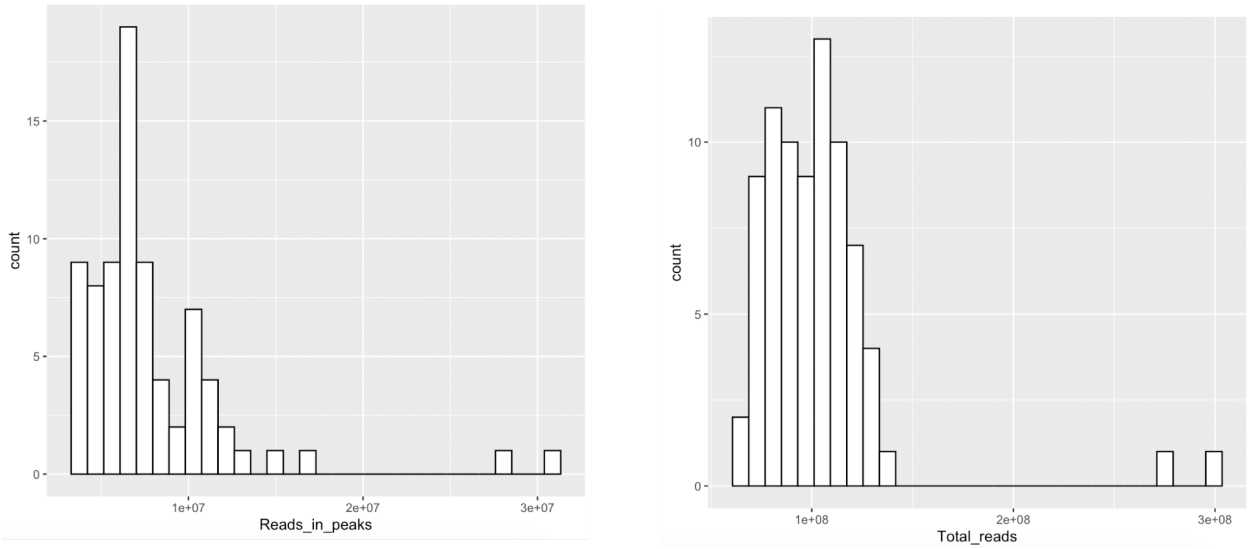
Quality control metrics, using the thresholds provided by the standard ENCODE Pipeline for the identification of open chromatin regions (OCRs) of the genome.

(A) shows distribution of all reads present in the peaks identified in the ATAC-seq dataset, as well as the distribution of all the reads by sample.

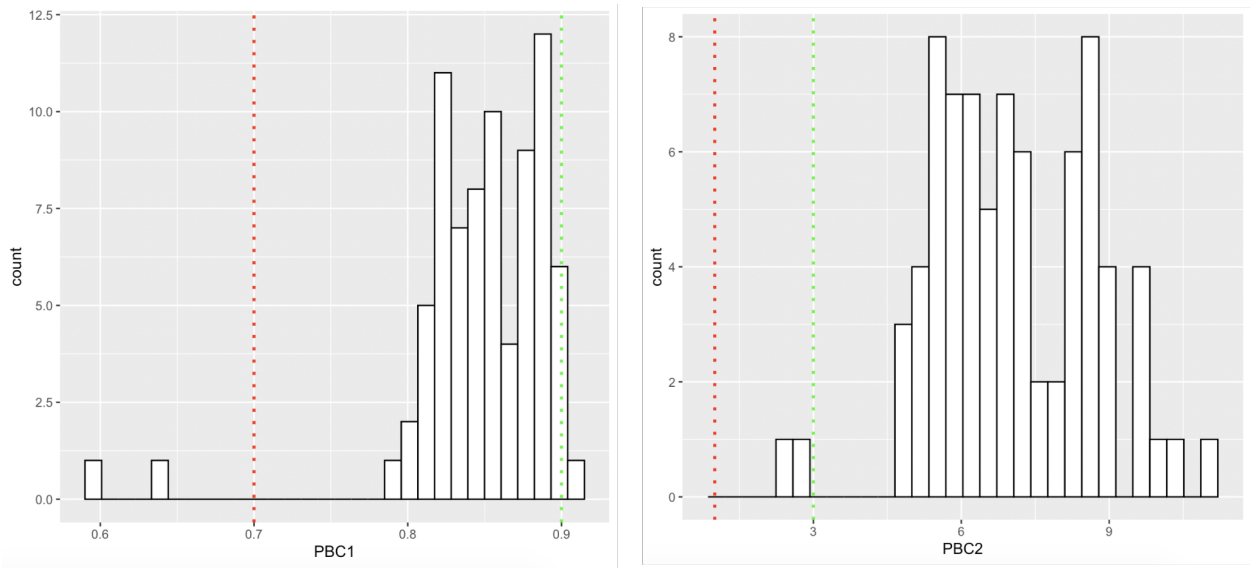
(B) shows library complexity metrics of PCR Bottlenecking Coefficients (PBC), measured as PBC1 = [# of positions with exactly 1 read mapped] / [# of positions with 1 or more reads mapped ] and PBC2 = [# of positions with exactly 1 read mapped] / [# of positions with 2 reads mapped ]. Red cutoff color indicates severe PCR bottlenecking threshold (<0.7 for PBC1, <1 for PBC2), green cutoff corresponds to no PCR bottleneck (>0.9 for PBC1, >3 for PBC2).

(C) shows the distribution of Non-redundant Fraction(NRF) scores. Values within 0.7 and 0.9 are considered acceptable. The outlying samples across QC were from the 2014L00966 cell line, time points 20 and 0.

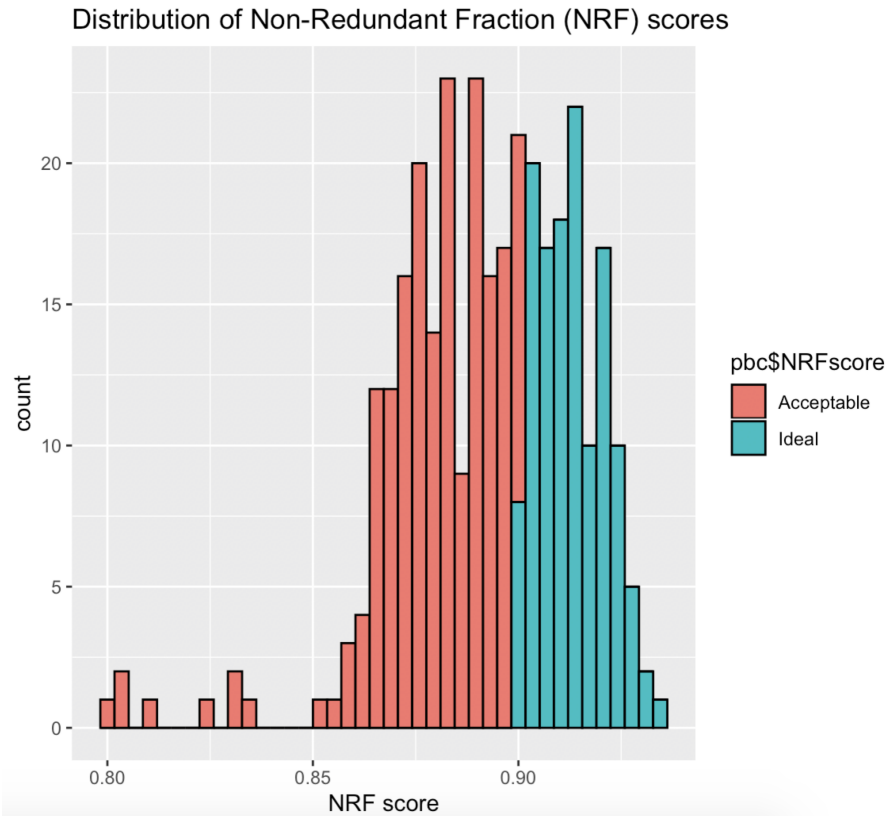
**A**




**B**

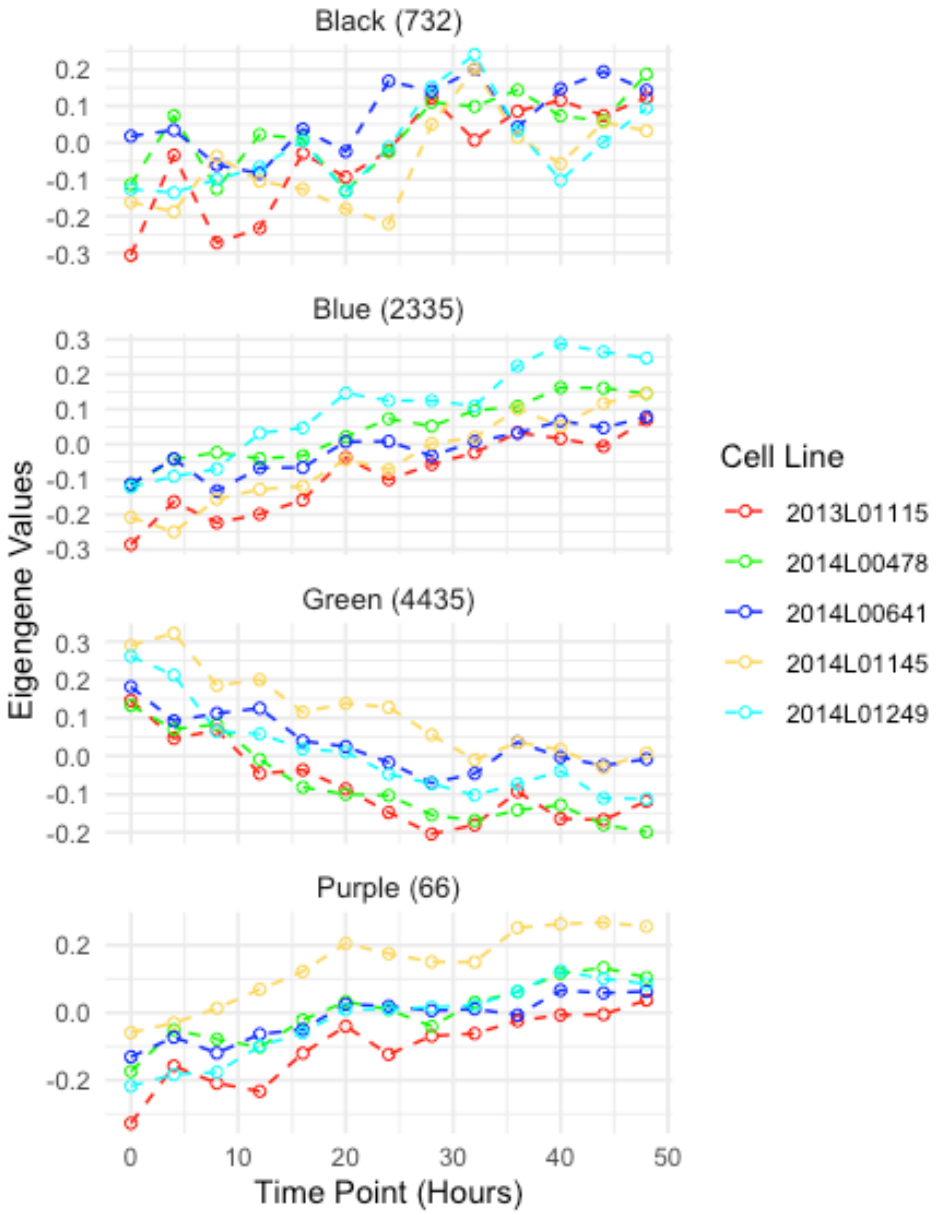



**C**




**Supplemental Figure S1.7 Eigengene values for ATAC-seq modules obtained from WGCNA.**

Eigengene modules from WGCNA of the longitudinal chromatin accessibility patterns of ATAC-seq data collected every 4 hours for a 48 hour period. Each color represents a fibroblast cell culture from a different individual. The number of chromatin accessible regions assigned per module is indicated next to the module name.



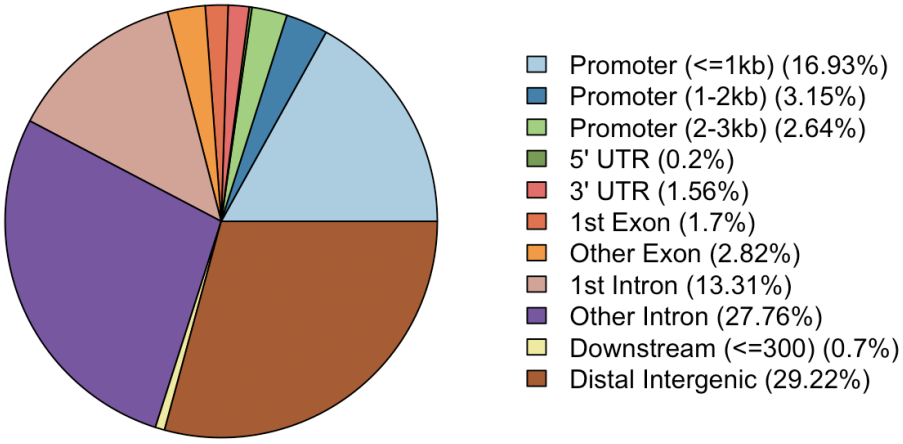
**Supplementary Figure S1.8 Genomic annotation of the consensus peak regions and selected time significant regions.**

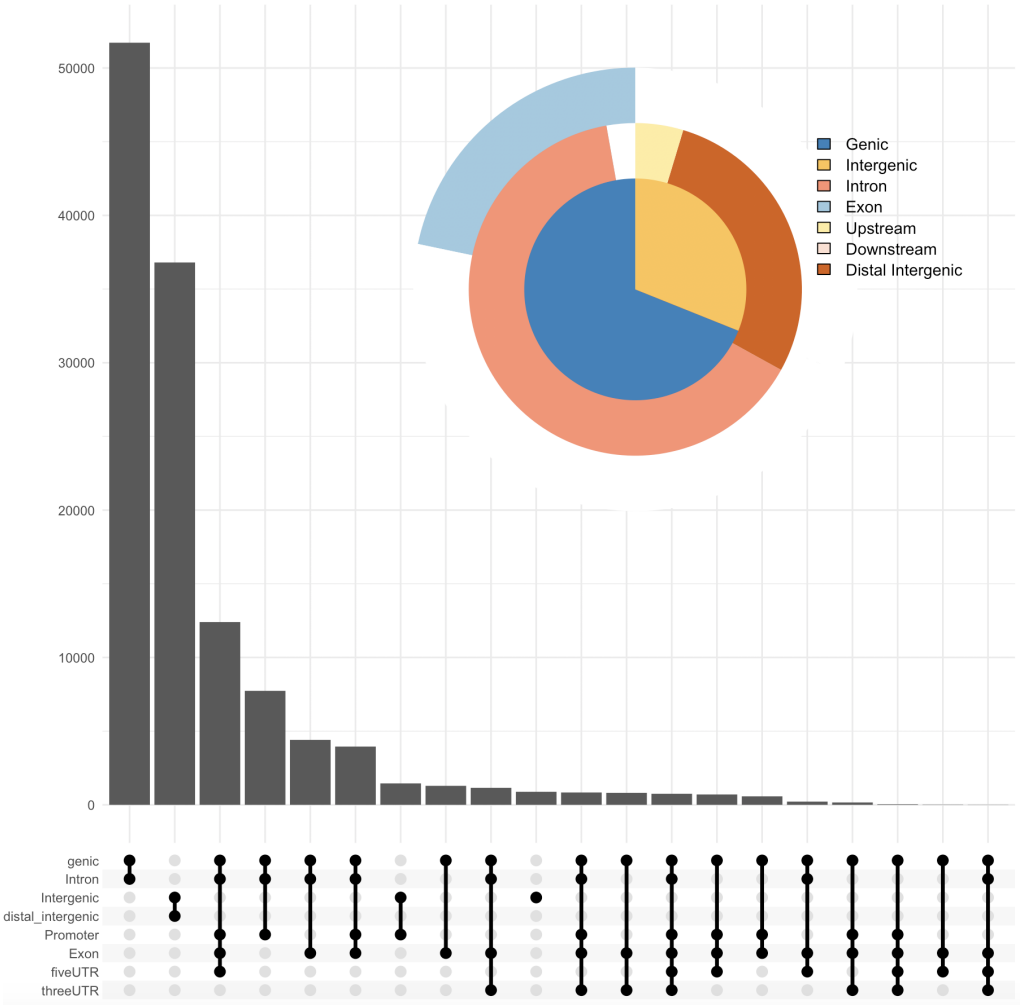
Chipseeker annotations for peak regions.

(A) Genomic annotations for all n=126,057 consensus peak regions.

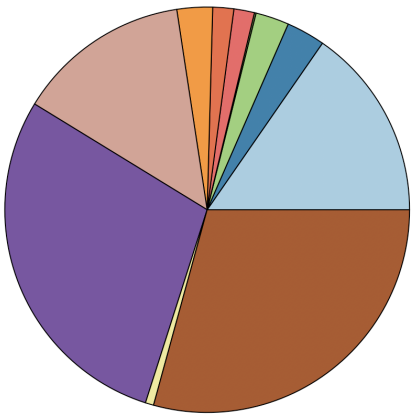
(B) Genomic annotations for the n=7,568 peaks that had a significant change over time in their accessibility.

**A. All consensus regions**






## B. Time significant regions



Promoter (<=1kb) (15.35%)

Promoter (1-2kb) (3.09%)

Promoter (2-3kb) (2.67%)

5' UTR (0.13%)

3' UTR (1.64%)

1st Exon (1.68%)

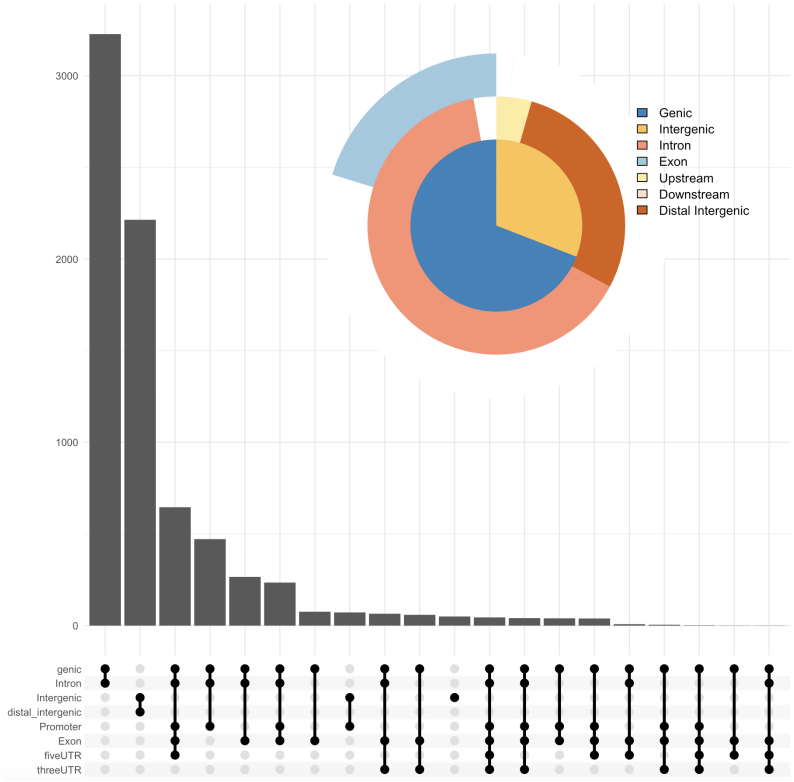
Other Exon (2.84%)

1st Intron (13.86%)

Other Intron (28.81%)

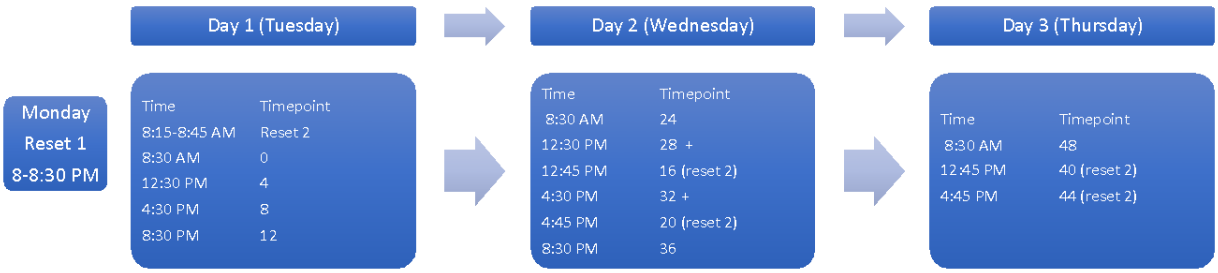
Downstream (<=300) (0.66%)

Distal Intergenic (29.27%)


3000
2000
1000
0
Genic
Intergenic
Intron
Exon
Upstream
Downstream
Distal Intergenic
genic
Intron
Intergenic
distal_intergenic
Promoter
Exon
fiveUTR
threeUTR

**Supplementary figure S1.9 Schematic of synchronization and collection times.**

Collection scheme for both RNA-seq and ATA-seq fibroblast cell culture samples. Cells were reset 12 hours before the first collection. In order to collect RNA or cells every 4 hours for 48 hours, cells were split into two batches, which were reset 12 hours apart.



## Chapter 2: The integrative analysis of Cerebrospinal Fluid biomarkers of Alzheimer's Disease, Metabolomics, and Genetic risk reveals novel metabolite associations

**Authors:** Marcelo Francia[1], Naren Ramesh[1], Toni Boltz[2], Merel Bot[1], Wiesje M. van der Flier[3,4], Pieter Jelle Visser[3,4,5,6], Sven van der Lee[3,4,7], Charlotte E. Teunissen[8], Yolande A.L. Pijnenburg[3,4], Anouk den Braber[3,4,9], Loes Olde Loohuis[1], Lianne M. Reus[1,3,4], Betty M. Tijms[3,4], Roel A. Ophoff[1,2]

1.Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA

2.Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

3.Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands

4.Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

5.Department of Psychiatry, Maastricht University, Maastricht, The Netherlands

6.Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden

7.Section Genomics of Neurodegenerative Diseases and Aging, Department of Clinical Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands

8.Neurochemistry Lab and Biobank, Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam UMC, Amsterdam, The Netherlands

9.Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

### INTRODUCTION

Cerebrospinal fluid (CSF) has emerged as a valuable source to study ongoing neurobiological processes due to its close proximity to the brain. Particularly, biomarkers for Alzheimer's

disease (AD) can be measured in vivo from CSF, with decreased β-amyloid$_{1-42}$(AB) levels and increased levels of phosphorylated tau (P-Tau) and total tau (T-Tau) [130] associated with disease status and being used for diagnostic purposes [131]. These biomarkers are indicators for the presence of extracellular amyloid-beta plaques, intracellular tau tangles, and neurodegeneration in brains of patients with AD [132];[133]. It is therefore considered that CSF metabolomic studies enable the exploration of biological pathways linked to neurological disorders, such as AD. A few smaller and targeted studies have been published reporting an association between CSF metabolites and P-Tau and T-Tau levels [134];[135]. While both studies found metabolic pathways involved in Tau pathology and neurodegeneration [134]; [135], they did not observe a similar pattern for CSF AB levels in AD patients. A more comprehensive approach (with many more metabolites and with an increased sample size) is required to more fully assess the connection between AD CSF biomarkers and in vivo metabolic pathways in the human central nervous system.

AD is a highly heritable disorder, with the APOE-e4 allele contributing the strongest genetic risk factor for AD [136]. Genome-wide association studies (GWAS) on AD have identified more than 80 genetic risk loci, each having a small association with the disease [37]. These genetic loci can be harnessed to construct polygenic risk scores (PRS), collecting the genetic effects of these regions to evaluate individuals' susceptibility to AD [137]. CSF AB, P-Tau, and T-Tau levels are also heritable traits, as underscored by previous GWAS ([138];[133]). While the APOE locus predominantly predicts AB CSF levels, comparatively weaker prediction also exists for P-Tau and T-Tau at this locus, although other genetic loci associate with P-Tau and T-Tau more strongly [133]. Metabolomic profiling introduces an additional layer of biological information which may facilitate the deciphering of the genetic underpinnings of polygenic disorders. Metabolites can reflect intricate biological processes and hold the potential to gauge the collective impact of a disease, pinpoint biomarkers, as well as chart disease progression ([134][139]). Integrating PRS measurements with metabolites can highlight metabolism regulators intertwined with disease mechanisms [140]. We hypothesize that an

integrative analysis of known AD biomarkers, CSF metabolomic profiling and genetic risk (PRS), will reveal metabolic pathways associated with the genetic architecture of AD.

Our study leveraged the different layers of biology captured by genetic risk factors for disease (i.e. polygenic risk scores, PRS) and CSF metabolomics to uncover metabolic pathways involved in AD etiology and pathophysiology. . For this purpose, we collected 5,543 CSF metabolite measurements from AD dementia patients, mild cognitive impairment (MCI) patients, and cognitively unimpaired subjects in three Alzheimer Amsterdam-related cohorts ([141][142][143][144]), along with genome-wide genotype data. Through correlation and elastic net regression analyses, we examined which CSF metabolites are associated with P-Tau, T-Tau, and AB CSF levels. Following an assessment of the contribution of APOE to AD PRS, we employed linear models to gauge the predictive capabilities of AD PRS for P-Tau, T-Tau, and AB CSF levels, and CSF metabolomics. Our study expands on past CSF metabolomics studies, and identifies novel associations that highlight metabolic pathways impacted by AD pathology.

## RESULTS

### Metabolome-wide correlation of AD Biomarkers

Metabolites in the CSF can serve as indicators of a brain's physiological state [145]. To gauge how AD pathology influences physiological processes in vivo, we integrated CSF metabolites with P-Tau and T-Tau CSF levels, both well-established AD biomarkers known to be associated with disease progression. The study design is outlined in Figure 2.1. This analysis was also applied to AB CSF levels within our clinical cohort (Table 2.1). Untargeted metabolomic data were collected from three distinct platforms capturing metabolites across primary metabolism [146], complex lipids [147], and biogenic amines [148] (See Methods section for full details). Out of 5,262 metabolites (n = 678 named, n = 4,587 unnamed), 271 demonstrated significant correlations with T-Tau levels after Bonferroni multiple testing corrections (Figure 2.2A) of which 121 (44.6%) were named metabolites. 262 of these metabolites exhibited positive correlations,

while 9 displayed negative correlations. Similarly, 271 metabolites (119 named) were correlated with P-Tau, with 258 showing positive correlations and 13 showing negative correlations (Figure 2.2B). 254 metabolites were found to be correlated with both P-Tau and T-Tau CSF levels, with 17 unique metabolites associated with each, respectively. (Full list available in supplementary files). Notably, no metabolites were found to be significantly correlated with CSF AB (Figure 2.2C). Table 2.2 summarizes the strongest positive and negative correlations for both CSF Total Tau and CSF Phosphorylated Tau. Furthermore, we stratified the cohort based on case (mild cognitive impairment (MCI) and Alzheimer's disease dementia (AD) and control (normal cognition (NC) and subjective cognitive decline (SCD) status. Upon stratification, the correlations between CSF metabolites and AD biomarkers remained generally consistent irrespective of case-control status for metabolites with statistically significant correlations (T-Tau Pearson R = 0.888, P-Tau Pearson R = 0.915). This indicates that these metabolite associations with CSF T-Tau and P-Tau levels are independent of AD case-control status in this cohort. For non-significant correlations we identified large variations between cases and controls (T-Tau Pearson R = 0.261, P-Tau Pearson R = 0.281) (Supplementary Figure 2.1).

**CSF metabolites can predict AD biomarkers**

In our previous analysis, we identified CSF metabolites correlating with both CSF P-Tau and T-Tau levels. To explore the potential of CSF metabolites as predictors for AD biomarkers, we employed an elastic net regression model. This model utilized CSF metabolites as predictors for P-Tau and T-Tau CSF levels. Base models using only age and sex as predictors were also built for comparison. Upon adjusting for age and sex, elastic net regression models incorporating CSF metabolites demonstrated an average R2 of 0.51 for P-Tau (Figure 2.3A) and 0.46 for T-Tau (Figure 2.3C). In contrast, models using only age and sex as predictors had an R2 of 0.16 for P-Tau and 0.17 for T-Tau. Including CSF metabolite levels in the model provided additional predictive power. To identify consistently predictive CSF metabolites, we

constructed elastic net regression models over a thousand iterations of training and testing sets. In each iteration, 80% of the samples from our cohort were randomly selected to be used as a training set, while the remaining 20% were selected as a test set. Metabolites present in 80% of these models were considered consistent predictors. For models predicting P-Tau, 41 metabolites were identified as consistent predictors, with 13 being named metabolites (Figure 2.3B). For models predicting T-Tau, 33 metabolites were identified as consistent predictors, with 8 being named metabolites (Figure 2.3D) (Full list of metabolites are provided in Supplementary Table 2.1). With the exception of the SM d34:2 metabolite, a phosphosphingolipid, all consistent named metabolites that predicted T-Tau are also highly predictive of P-Tau CSF levels.

**Pathway Analysis of CSF metabolites correlated with P-Tau and T-Tau**

Having identified metabolites that exhibited correlations and consistent predictions for P-Tau and T-Tau CSF levels, our next step was to identify which metabolic pathways are associated with these metabolites. Utilizing 288 unique metabolites that correlated with either P-Tau or T-Tau, we conducted an overrepresentation analysis (See methods section for details). This analysis revealed several significantly enriched pathways (n = 3), including Pentose and Glucuronate Interconversions, Ascorbate and Aldarate Metabolism, and Amino Sugar and Nucleotide Sugar Metabolism (Supplementary Figure 2.4). To corroborate these findings, we employed PaIRKAT, a pathway integrated regression-based kernel association test that associates entire pathways with a phenotype. This alternative tool reaffirmed our results, identifying specific pathways significantly associated with both P-Tau (Figure 2.4A) and T-Tau (Figure 2.4B) levels. Notably, many of the highly enriched pathways exhibited overlap between the two, including Glycerophospholipid Metabolism, ABC Transporters, Linoleic Acid and Arachidonic Acid Metabolism, Retrograde Endocannabinoid Signaling, Alpha-Linoleic Acid Metabolism, and Choline Metabolism in Cancer (Full list of enriched metabolites in the supplementary files).

**Prediction of AD biomarkers and CSF metabolites using polygenic risk scores**

Previous GWAS have identified associations between AD biomarkers (P-Tau, T-Tau, and AB CSF levels) and the genetic architecture of AD [132]. To investigate similar associations in our clinical cohort, we constructed a PRS for AD based on the most recent GWAS [149]. Due to the significant impact of the APOE locus on the PRS (Supplementary figure 2.3), we used an approach that initially generates the AD PRS without the APOE locus and subsequently incorporates the effects of the APOE-e4 and -e2 alleles as a weighted sum [137]. After adjusting for age and sex, linear models revealed that the APOE-weighted AD PRS had a modest association with P-Tau (Adjusted R2 = 0.1368; p = 0.014) and T-Tau (Adjusted R2 = 0.1438; p = 0.007) CSF levels, in comparison with the strong association found for AB CSF levels (Adjusted R2 = 0.2200; p = 9.11E-14). We then compared various models, incorporating different approaches to account for the effects of the APOE alleles. These approaches included using only the APOE-e4 allele counts, the AD PRS without the APOE region, and a weighted risk score for APOE based on autopsy-confirmed AD cases (APOE-npScore) [150]. The summarized results (Table 2.3) revealed that across all three biomarkers, the AD PRS without the APOE region yielded the lowest association (AB R2 = 0.1017, P-Tau R2 = 0.1294, and T-Tau R2 = 0.1340). Although the APOE-weighted AD PRS emerged as the best-performing model, its improvement over solely using the APOE e4 allele counts was marginal.

While metabolomics can offer dynamic insights into an individual's physiological state and environmental exposures, it has also been shown that multiple CSF metabolites are under genetic control [151–153]. To uncover metabolic processes associated with the genetic architecture of AD, we applied linear models to predict the levels of our panel of 5,262 CSF metabolites using the APOE-weighted AD PRS, including age and sex as cofactors. This approach yielded 4 significant associations (FDR < 0.05) between CSF metabolites and the APOE-weighted AD PRS, although these were all among unnamed metabolites (Supplementary Figure 2.4 and supplementary files). Given that the strongest associations with the PRS were

found in unknown compounds, non-negative matrix factorization (NMF) [154] was employed as a dimensional reduction strategy. This approach aimed to capture broader metabolite signals that could potentially be linked to the PRS. In predicting NMF values, the strongest association with an adjusted R2 of 0.09 showed significance for age (adjusted p = 6.70E-10). The second strongest association, with an adjusted R2 of 0.07, had significance for sex (adjusted p = 7.18E-6). Only for the third strongest association, with an adjusted R2 of 0.03, did the APOE-weighted AD PRS show nominal significance (p = 0.008)(Full results in supplementary files).

To further elucidate the associations between the APOE-weighted AD PRS and CSF metabolites, we replicated our analysis on an unrelated, cognitively healthy cohort (n = 449) with the same panel of CSF metabolite measurements [151](Cohort description in Table 2.1). However, linear modeling in this cohort failed to reveal any significant associations between CSF metabolites and the APOE-weighted AD PRS, both nominal and after multiple testing corrections (Supplementary Figure 2.4 and supplementary files). Consistent with the AD cohort findings, metabolites with higher adjusted R2 (>0.2) displayed significance for both age and sex terms. Remarkably, only two named metabolites exhibited significance with age as a significant term: 1,7-Dimethyluric acid (adjusted p = 1.06E-25) and Quinic acid (adjusted p = 1.32E-24). To investigate if the absence of associations was exclusive to the AD PRS, we expanded our analysis to include other brain disorders and traits such as Schizophrenia [96], Attention-deficit/hyperactivity disorder (ADHD) [95], Bipolar Disorder [43], Insomnia [99], Migraine [155], and Alcohol Use Disorder [156]. A summary of these associations can be viewed in supplementary Figure 2.4. None of these polygenic scores yielded significant associations with CSF metabolites in either the AD cohort or the cognitively healthy cohort.

## DISCUSSION

In this study, we leveraged genotype data, an extensive panel of 5,543 CSF metabolites and CSF AD biomarkers in a large cohort, to decipher impacted biological pathways by the disease. Our correlation and elastic net regression analyses revealed 288 unique CSF

metabolites linked with P-Tau and T-Tau CSF levels, yet no metabolites demonstrated associations with AB CSF levels. Among the CSF metabolites predictive of both P-Tau and T-Tau levels found in this study, we identified Anserine and Fucose as novel associations. The pathway enrichment analysis of these CSF metabolites associated with P-Tau and T-Tau levels consistently highlighted glycerophospholipid metabolism and pentose and glucuronate interconversions as significant pathways for both AD biomarkers. By utilizing an AD PRS adjusted for the substantial effect size of APOE alleles, we observed a significant prediction for AB CSF levels in our cohort. This prediction extended to both P-Tau and T-Tau CSF levels, albeit to a moderate extent. Notably, the AD PRS without the APOE locus yielded the lowest performing predictions across all three biomarkers. Linear regression models did not identify consistent associations between CSF metabolites levels and the APOE-weighted AD PRS, nor for additional polygenic scores related to various brain traits and disorders.

Among the CSF metabolite associations with CSF P-Tau and T-Tau levels, we identified two novel CSF metabolites, Anserine and Fucose. Anserine is the methylated analogue of Carnosine, an endogenous dipeptide present in various mammalian tissues, including the brain [157]. Due to its antioxidant and anti-inflammatory functions, Carnosine could be a potential modulator of biological pathways affected by AD [158]. Supplementation with Anserine and Carnosine has demonstrated improvements of AD symptoms within both AD mouse models [159] and elderly healthy human subjects [160], particularly in alleviating memory deficits. The observed strong negative correlation between Anserine and P-Tau/T-Tau aligns with a potential protective mechanism, though further studies are essential to elucidate interactions of Anserine with these AD biomarkers. The other newly identified metabolite, Fucose, is a sugar component of glycolipids and glycoproteins [161]. Fucose's incorporation into glycolipids has been linked to effects on learning, long-term potentiation, and synapse formation in animal models [162,163]. Direct delivery of Fucose into the brain improved retention of learned behavior in rats [164] and mitigated neuroinflammation by inhibiting microglial cell activity in mice [165]. While these functions require validation in humans, a transcriptomic study of brain

samples from an AD cohort identified the upregulation of the FUT8 gene in AD subjects [166]. The protein product of FUT8 catalyzes fucose transfer to glycolipids [167], highlighting a potential role of fucosylation in AD.

The pathway enrichment analysis of the CSF metabolites associated with P-Tau and T-Tau levels consistently highlighted glycerophospholipid metabolism as a significant pathway for both AD biomarkers. Glycerophospholipids, major components of cellular membranes, play a diverse role in altering the functional properties of cells, including signal transduction, vesicle trafficking, and membrane fluidity [168]. In the brain, phospholipases actively catabolize glycerophospholipids through hydrolysis [169]. Abnormal lipid metabolism, particularly changes in glycerophospholipids, has been linked to AD pathogenic features such as amyloidogenesis, oxidative stress, and neuroinflammation [170]. A postmortem lipidomics study on brain tissue identified glycerophospholipids among the lipid subclasses significantly perturbed in AD cases [171]. The CSF metabolites enriched in glycerophospholipid metabolism identified in our study further highlights the importance of this biological pathway in AD. Another enriched pathway was pentose and glucuronate interconversions, previously associated with metabolites linked to P-Tau/T-Tau in a different (much smaller) AD cohort [134]. Although dysregulation of glucose and pentose brain levels has been associated with AD pathology [172,173], the specific interactions of these pathways and associated metabolites with Tau in AD remain to be elucidated. The enrichment of these pathways in metabolites strongly correlated with P-Tau/T-Tau suggests a potential association between elevated levels of these AD biomarkers and processes specifically impacting neuronal activity and viability. A prior study, integrating proteomic, genomic, and imaging data, demonstrated that distinct proteomic profiles and associated pathways were linked to varying levels of P-Tau/T-Tau within individuals with AD (Visser et al., 2022). Further exploration of the data generated in this study offers an opportunity to assess whether different metabolites and pathways are associated with diverse levels of these AD biomarkers.

The absence of associations between CSF metabolites and AB levels may be attributed to the early accumulation of AB in the brain [130]. While AB serves as an AD biomarker, its accumulation has not been shown to mirror or predict disease progression, unlike P-Tau and T-Tau levels [174,175]. Consistent with previous research on CSF metabolites, our study replicates this lack of associations with AB CSF levels [134,135]. Despite significantly expanding the scope of CSF metabolites in this study by ten fold in comparison to previous work in the field, the possibility of overlooking specific metabolites associated with AB remains. While this work stands as the largest CSF metabolites and AD biomarker study to date, it is possible that the impact of AB on CSF metabolites has a small effect size that will require further increases in cohort size to identify it. Future advancements in metabolomics and improvements in metabolite profiling technologies may hold promise for elucidating the intricate relationship between AB levels and CSF metabolites.

In our analysis of the APOE locus and the PRS of AD, we found that the e4 allele counts alone significantly predict these AD CSF biomarkers, corroborating findings from previous studies ([176]; [150]). The APOE gene, encoding a glycoprotein lipid transporter, exhibits isoform-specific impacts on its function [40]. These results align with the proposed relationship between AB and the APOE locus, with AB plaque levels shown to be influenced by different APOE alleles [177]. Regarding AB, it is hypothesized that the e4 allele may influence AB levels by promoting plaque formation [178] and hindering its clearance from the brain [179]. APOE alleles have also been implicated in Tau neurofibrillary tangles (NFT) CSF levels [180], although the precise association between APOE and Tau is still under investigation. Our findings align with these known interactions and support prior reports identifying the APOE locus as a major genetic risk factor for AB, T-Tau, and P-Tau CSF levels [132,133,138]. We did not observe consistent associations between CSF metabolites levels and the APOE-weighted AD PRS, nor for additional polygenic scores related to various brain traits or disorders. Notably, only four CSF metabolites exhibited significant associations with the APOE-weighted AD PRS in the clinical cohort. However, these were among unnamed metabolites, limiting the interpretability

of these findings. It is worth noting that metabolite levels can be heritable traits as demonstrated by prior QTL mapping studies [140,181,182]. Previously, it has been shown, for example for body mass index [183](fang et al), that polygenic scores of trait measures are associated with metabolite levels.

One of the primary limitations of this study is the current inability to validate the strength of our predictors for P-Tau and T-Tau on an independent cohort. While other studies have investigated CSF metabolites in the context of these AD biomarkers, the lack of overlap in the panels of metabolites used presents a challenge for direct comparisons across cohorts. Another limitation is that the cohort utilized in this study comprises exclusively individuals of European ancestry, restricting the generalizability of the results to other ancestry groups. Both the genetic architecture of AD and the performance of AD biomarkers exhibit variations across populations of different ancestries, such as in the case of African American populations. Although the number and size of GWAS studies on African American individuals with AD are limited [184], evidence suggests that, besides the APOE locus, the ABCA7 locus shows strong associations with AD in this population [185]. Similarly, the performance of T-Tau and P-Tau as biomarkers for AD is inconsistent in African American populations [186]. One study demonstrated significantly lower levels of these metabolites in African Americans compared to non-Hispanic white individuals [187]. Future AD studies should aim to broaden their cohorts to include diverse populations, enhancing our understanding of the disease across different ancestry groups.

In this study, our aim was to assess how distinct biological layers, including genetic risk and metabolomic profiles, contribute to capturing various aspects of AD pathology associated with the established clinical biomarkers of disease. We successfully replicated previous findings regarding the impact of the APOE e4 allele on AB, P-Tau, and T-Tau CSF levels. Additionally, we compared the effectiveness of different approaches in modeling the genetic architecture of AD. Notably, we extended prior CSF metabolite investigations by substantially increasing the sample size of a well characterized cohort as well as by

significantly expanding the scope of the metabolomic measurements (with more than five thousand metabolites). This increased effort resulted in the identification of 288 unique CSF metabolites associated with CSF levels of P-Tau and T-Tau, but also highlighted the lack of any significant finding of an association with CSF metabolites and AB,. We identified novel associations between the CSF metabolites Anserine and Fucose, and P-Tau/T-Tau. Pathway analysis of these metabolites further supports their involvement in established biological pathways affected by AD. This makes these metabolites potential targets for studying and gaining a deeper understanding of how AD progression impacts brain physiology.

## METHODS

### Sample information and processing

#### Study Participants

A total of 977 study samples (with an average age of 52.7±16.6 years and 35.9% female) were used in this analysis, drawn from both a memory clinic cohort and a group of cognitively healthy subjects in the Netherlands.

The memory clinic cohort comprised samples from three cohorts affiliated with the Alzheimer Center Amsterdam. These cohorts include the Amsterdam Dementia Cohort (ADC) [141], the 90+ Study [142], and the Twin Study [143,144]. The ADC, initiated in the year 2000, is an ongoing observational follow-up study of patients who visited the memory clinic at Amsterdam UMC, location VU University Medical Center (VUmc), with dementia diagnoses made according to established guidelines for neurodegenerative diseases [3,188–190]. The 90+ Study, part of the Innovative Medicine Initiative European Information Framework for AD (EMIF-AD), focuses on cognitively healthy individuals aged 90 and above, aiming to identify factors associated with resilience to cognitive impairment in the old [142]. For the Twin Study, monozygotic twins (one subject per twin pair) were recruited from the Netherlands Twin Register [143] to participate in the PreclinAD study, a component of the EMIF-AD project (http://www.emif.eu/) [144]. Supplementary materials provide a comprehensive description of each cohort.

Cohort of Cognitively Healthy Subjects: The recruitment of these neurotypical subjects is described before (see Luykx et al. 2014 [151]). In short, inclusion involved patients undergoing spinal anesthesia for minor elective surgical procedures, ages between 18 and 60, and with all four grandparents born in The Netherlands or other Northwestern European countries (Belgium, Germany, UK, France, and Denmark). Each potential participant underwent a telephone interview to screen for self-reported psychotic or major neurological disorders (such as stroke, brain tumors, or neurodegenerative diseases) and to document any use of psychotropic medication.

An overview of the characteristics for the memory clinic cohort (N=487) and the cognitively healthy subjects cohort (N=449, all cognitive healthy controls) can be found in Table 2.1. The Amsterdam Dementia cohort samples comprises n=220 cognitively unimpaired subjects, n=87 subjects with mild cognitive impairment (MCI), and n=180 patients with AD-type dementia. As expected, patient groups in the Amsterdam sample differed from each other, with the AD-type dementia group having more APOE-ε4 carriers, fewer APOE-ε2 carriers, and exhibiting more abnormal AD CSF biomarkers compared to the MCI and cognitively unimpaired subjects. Cognitively healthy subjects, in comparison to memory clinic subjects, were less frequently female, younger, had a lower APOE-ε4 frequency, and a higher APOE-ε2 frequency.

All participating studies received approval from their respective Medical Ethics Committees, and informed consent was obtained from all participants, either directly or from their legal representatives.

**CSF data collection**

Memory clinic cohort: CSF samples were acquired through lumbar puncture using a 25-gauge needle and syringe. Amyloid-beta 42 (Aβ42), total tau (t-tau), and hyperphosphorylated 181 tau (p-tau) levels were assessed as part of the diagnostic work-up, employing enzyme-linked immunosorbent assays (ELISA) (Innotest: Fujirebio, Ghent, Belgium) [141]. In the ADC cohort,

CSF Aβ42 values were adjusted for drift over time, as detailed previously [191]. In this cohort, the biomarker abnormality cut-offs used were CSF Aβ42 < 813 pg/ml, CSF t-tau > 375 pg/ml, and CSF p-tau > 52 pg/ml.

Cognitively healthy subject cohort: Each subject provided a 6ml CSF sample through lumbar puncture, immediately stored in fractions of 0.5 and 1ml at –80°C, as previously described (see Luykx et al. 2014 [151]).

**CSF metabolite processing**

A total of 5,543 CSF metabolites were assessed through three distinct metabolite assays at the West Coast Metabolomics Center at UC Davis, encompassing GC-TOF MS (primary metabolism), CSH-QTOF MS/MS (complex lipids), and HILIC-QTOF MS/MS (biogenic amines) (https://metabolomics.ucdavis.edu/).

To address missingness, metabolite levels were initially examined across each cohort, and metabolites with missing data for over 20% of individuals were excluded. For the remaining metabolites, missing values were imputed to half the median value for the corresponding metabolite across the cohorts. This decision was based on the assumption that these metabolites likely exist in quantities too low to be detected in these individuals, a method consistent with previous approaches [134]. Subsequently, inverse-rank normalization was applied to all metabolites to ensure normality for downstream analysis.

**Genotyping and imputation**

Memory clinic samples underwent genotyping using the Illumina Global Screening Array (GSA), while genotype data for cognitively healthy controls were obtained through the OmniExpress Exome array. Initial filtering of autosomal genotypes involved excluding SNPs with <2% SNP-missingness and >5% minor allele frequency (MAF) using plink [192], performed separately per cohort. Individuals with a call rate below 98% were excluded. Following this, genotype VCF

files were uploaded to the TopMed server for imputation and liftover to hg38. Post-imputation quality was assessed by filtering variants with imputation R2>0.3, resulting in approximately 8 million SNPs per cohort for subsequent analyses. Imputed genotypes were subsequently merged between the two cohorts, and once again filtered for variants with <2% SNP-missingness and >5% MAF.

**Statistical Analyses**

**Polygenic risk scores calculation**

Polygenic risk scores for Alzheimer's Disease were calculated using PRScs software [193] and the latest Alzheimer's GWAS that did not contain the cohort used in this study [149]. We used the European reference panel from 1000 Genomes Phase 3 to model LD in the polygenic risk score computation. To improve prediction accuracy of the AD PRS, we followed the approach suggested by [137], in which the APOE locus(Chr19 43.4M-47.5M) is removed prior to the calculation of the AD PRS, and then the APOE effect is added as a weighted sum to the AD PRS, with effect sizes of -0.47 for the e2 allele and 1.12 for the e4 allele. The region for the APOE locus was chosen after comparing multiple genetic windows and the resulting AD PRS across individuals with different APOE e4 allele counts.

Polygenic scores for other brain disorders and traits such as Schizophrenia [96], Attention-deficit/hyperactivity disorder (ADHD) [95], Bipolar Disorder [43], Insomnia [99], Migraine [155], and Alcohol Use Disorder [156] were also calculated in both cohorts using PRScs and the respective GWAS summary statistics.

**Association of AD PRS with CSF metabolites and AD biomarkers**

Linear regression was used to identify CSF metabolites that could be predicted with the AD PRS. Using the lm function from R we generated linear models predicting CSF metabolite levels using the weighted AD PRS and as cofactors age and sex. We also compared these

models to using the AD PRS calculated without the APOE regions, as well as including the APOE region. This approach was extended to associate Amyloid beta, phosphorylated tau and total tau levels with the AD PRS when available for the sample. AD biomarkers levels were log normalized for linear regression analyses. After correcting for PC and age outliers, we ended up with 461 individuals available for the analysis.

**Correlations between AD biomarkers and CSF metabolites**

In the Amsterdam cohort, individual metabolites were tested for Spearman correlation with CSF phosphorylated tau, total tau, and amyloid beta using the cor.test function in R. Any correlations that passed Bonferroni multiple testing correction were considered significant correlations.

**Elastic Net Regression**

In order to restrict the number of metabolites to those most relevant to predicting Alzheimer's disease biomarkers, elastic net regression was utilized using the train function from the caret package in R. Due to the lack of AD biomarker data in the Utrecht cohort, this analysis was limited to the Amsterdam cohort. This method was implemented to select the most important metabolites out of the (5000+) metabolites for predicting CSF phosphorylated tau, total tau, and amyloid beta. To account for the variability of predictions given the way the data is split between test and training sets, the data was split 80/20 in 1000 random ways. In each iteration, 80% of the data was used to train a model to predict AD biomarkers and 20% of the data was used to test the model's prediction accuracy. In order to determine the added predictive value of CSF metabolites, base models were tested using only age and sex. If any metabolites were included in over 80% of the 1000 models, they were considered to be highly predictive and of importance. To verify the importance of these metabolites, the rest of the metabolites that were not considered highly predictive were corrected for the effects of the

highly predictive metabolites. Elastic net regression was then run using these metabolites to predict phosphorylated tau and total tau  in a similar manner described above.

## CHAPTER 2 TABLES

### Table 2.1: Cohort characteristics

Controls include both individuals with normal cognition and subjective cognitive decline.

| Characteristic | Clinical Cohort (n=610) | Clinical Cohort Samples with CSF and Genotype (n=487) | Cognitively Healthy Cohort (n = 449) |
|---|---|---|---|
| Age, mean(SD) | 65.5 (9.4) | 65.5 (9.7) | 39.7 (11.4) |
| Female, n(%) | 272 (44.6) | 207 (42.5) | 124 (27.6) |
| **Diagnosis** | | | |
| Controls, n(%) | 297 (48.7) | 220 (45.2) | 449 (100) |
| MCI, n(%) | 103 (16.9) | 87 (17.9) | - |
| Dementia, n(%) | 210 (34.4) | 180 (36.9) | - |
| ***APOE* genotype** | | | |
| *ε2/ε2*, n(%) | - | - | 2 (0.4) |
| *ε2/ε3*, n(%) | - | 29 (6.0) | 52 (11.6) |
| *ε2/ε4*, n(%) | - | 8 (1.6) | 17 (3.8) |
| *ε3/ε3*, n(%) | - | 206 (42.3) | 243 (54.1) |
| *ε3/ε4*, n(%) | - | 170 (34.9) | 123 (27.4) |
| *ε4/ε4*, n(%) | - | 74 (15.2) | 12 (2.7) |
| CSF Aβ42 (pg/ml), n(%) | - | 399 (81.9) | - |
| CSF P-Tau (pg/ml), n(%) | - | 407 (83.6) | - |
| CSF T-Tau (pg/ml), n(%) | - | 407 (83.6) | - |
| **Study Cohort** | | | |
| 90+, n(%) | 22 (3.6) | 20 (4.1) | - |
| ADC, n(%) | 477 (78.2) | 412 (84.6) | - |
| Twin Study, n(%) | 61 (10) | 55 (11.3) | - |
| Luykx et al. 2014, n(%) | - | - | 449 (100) |

### Table 2.2 Summary of the strongest CSF metabolite correlations for T-Tau and P-Tau CSF levels

| Metabolites Correlations | | | | |
|---|---|---|---|---|
| | CSF Phosphorylated Tau | | CSF Total Tau | |
| CSF Metabolite | R | Adjusted P-value | R | Adjusted P-value |
| PC 34:1 | 0.508 | 3.530E-28 | 0.498 | 9.190E-27 |
| PC 32:0 | 0.498 | 7.610E-27 | 0.481 | 1.540E-24 |
| PI 40:6 | 0.491 | 8.070E-26 | 0.455 | 1.850E-21 |
| PC 32:0 | 0.482 | 9.580E-25 | 0.462 | 3.260E-22 |
| PE 36:1 | 0.456 | 1.770E-21 | 0.456 | 1.670E-21 |
| Nicotinoylglycine | 0.475 | 8.800E-24 | 0.441 | 7.850E-20 |
| PE O-40:7 | 0.470 | 3.580E-23 | 0.426 | 3.200E-18 |
| isothreonic acid | 0.463 | 2.470E-22 | 0.464 | 1.850E-22 |
| PI 38:5 | 0.459 | 6.200E-22 | 0.438 | 1.650E-19 |
| Glu-Gln | -0.289 | 1.080E-06 | -0.292 | 7.330E-07 |
| Anserine | -0.379 | 1.350E-13 | -0.343 | 1.480E-10 |

**Table 2.3 AD Biomarkers prediction results**

Summary of results from linear regression models, after adjusting for age and sex, using different approaches to account for the APOE allele and genetic architecture of AD. R2 = R-squared; AIC = Akaike information criterion; BIC = Bayesian information criterion.

| Model | CSF Amyloid Beta | | | | | | CSF Phosphorylated Tau | | | | | | CSF Total Tau | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Covariate | n | R2 | AIC | BIC | Beta | P | n | R2 | AIC | BIC | Beta | P | n | R2 | AIC | BIC | Beta | P |
| E4 allele counts | 394 | 0.2198 | 187.04 | 206.8 | -0.3464 | 9.59E-14 | 402 | 0.1351 | 555.23 | 575.1 | 0.1081 | 0.021 | 402 | 0.1419 | 771.94 | 791.8 | 0.1186 | 0.011 |
| *APOE* weighted AD PRS | 394 | 0.2200 | 186.94 | 206.7 | -0.3467 | 9.11E-14 | 402 | 0.1368 | 554.45 | 574.3 | 0.1156 | 0.014 | 402 | 0.1438 | 771.05 | 790.9 | 0.1264 | 0.007 |
| *APOE*-npScore | 394 | 0.2170 | 188.42 | 208.1 | -0.3426 | 1.91E-13 | 402 | 0.1355 | 555.09 | 574.9 | 0.1095 | 0.0197 | 402 | 0.1424 | 771.72 | 791.5 | 0.1206 | 0.00997 |
| AD PRS without *APOE* region | 394 | 0.1017 | 242.55 | 262.3 | -0.0362 | 0.4500 | 402 | 0.1294 | 557.89 | 577.7 | 0.0764 | 0.1023 | 402 | 0.1340 | 775.62 | 795.46 | 0.0778 | 0.0955 |

**CHAPTER 2 SUPPLEMENTARY TABLE**

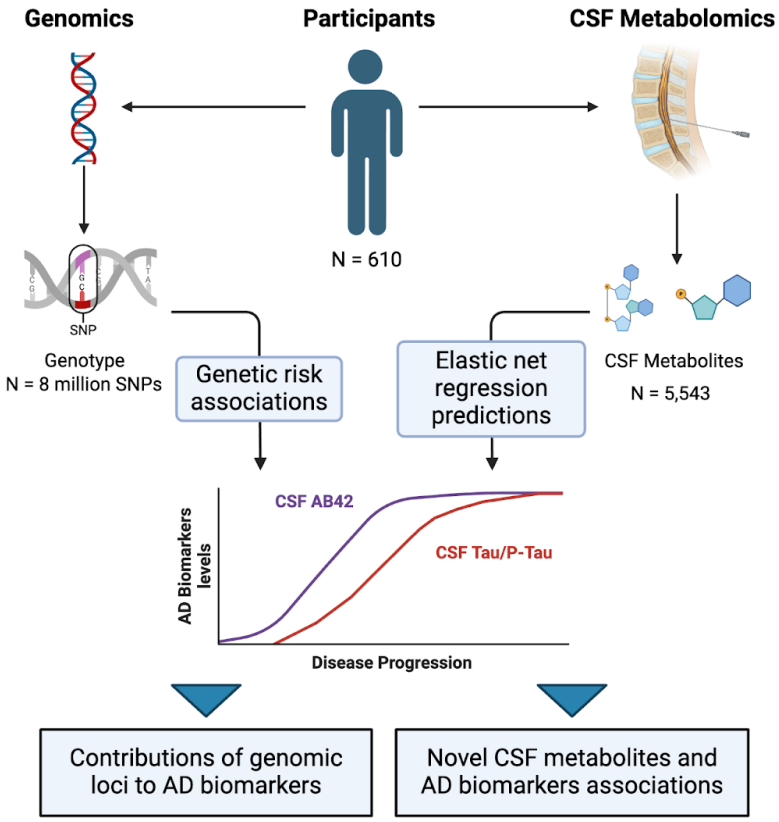
Table S2.1 List of all consistent metabolite predictors and factors (Frequency > 800) for P-Tau and T-Tau CSF levels

| Phosphorylated Tau Consistent Predictors | | Total Tau Consistent Predictors | |
|---|---|---|---|
| **Metabolite** | **Frequency** | **Metabetabolite** | **Frequency** |
| Glu-Gln | 1000 | Glu-Gln | 1000 |
| 9.13_369.25 | 1000 | 7.99_261.03 | 1000 |
| 7.99_261.03 | 1000 | 6.85_124.99 | 1000 |
| 6.85_124.99 | 1000 | 6622 | 1000 |
| 1.53_192.10 | 999 | 1.84_215.07 | 1000 |
| 7.04_275.93 | 998 | Isoxanthopterin | 999 |
| 6622 | 997 | 9.13_369.25 | 992 |
| Age | 994 | 8.02_245.08 | 992 |
| 4.87_233.08 | 990 | 7.04_151.06 | 990 |
| isothreonic acid | 988 | 8.41_406.13 | 986 |
| 12444 | 985 | 11.13_869.63 | 981 |
| 7.04_151.06 | 985 | alpha-Galactosamine-1-phosphate | 975 |
| Isoxanthopterin | 984 | 5.57_151.03 | 970 |
| 8.02_245.08 | 983 | 1.53_192.10 | 968 |
| 1.84_215.07 | 978 | isothreonic acid | 966 |
| gluconic acid | 974 | 3-Methylcrotonylglycine | 965 |
| 11.13_869.63 | 962 | 1690 | 959 |
| 8.47_263.07 | 961 | 7.35_256.06 | 955 |
| 7.35_256.06 | 957 | 8.92_232.13 | 940 |
| Anserine | 954 | 12444 | 924 |
| 8.01_307.07 | 953 | 4.33_233.08 | 924 |
| 5.57_151.03 | 941 | 7.04_275.93 | 915 |
| PI 40:6 | 938 | fucose | 912 |
| 10.89_841.81 | 937 | 1.00_456.34 | 909 |
| 3.67_218.10 | 935 | 6.48_182.03 | 900 |
| 2.10_153.07 | 915 | 7.69_1110.78 | 883 |
| PI 38:5 | 894 | Age | 876 |
| 3-Methylcrotonylglycine | 893 | 4.87_233.08 | 866 |
| 8.41_406.13 | 888 | gluconic acid | 863 |
| fucose | 880 | SM d34:2 | 858 |
| 8.07_215.03 | 880 | 0.98_423.08 | 838 |
| 34135 | 879 | 4.94_1237.83 | 815 |
| 9.31_754.63 | 850 | 0.74_200.24 | 809 |
| alpha-Galactosamine-1-phosphate | 850 | 1.91_202.96 | 807 |
| 1690 | 848 | | |
| 9.03_260.20 | 845 | | |
| N-Methylvaline | 823 | | |
| PI 36:4 | 815 | | |
| Arabitol | 808 | | |
| 7.69_1110.78 | 806 | | |
| 10.84_952.94 | 804 | | |
| 8.92_232.13 | 800 | | |

## CHAPTER 2 FIGURES

### Figure 2.1 Study Design Outline

Study Design Outline. Single nucleotide polymorphisms (SNPs), AD (Alzheimer's Disease), CSF (Cerebrospinal Fluid), β-amyloid1-42 (AB42), Phosphorylated Tau (P-Tau).




### Figure 2.2 CSF metabolites correlations with CSF AD biomarkers

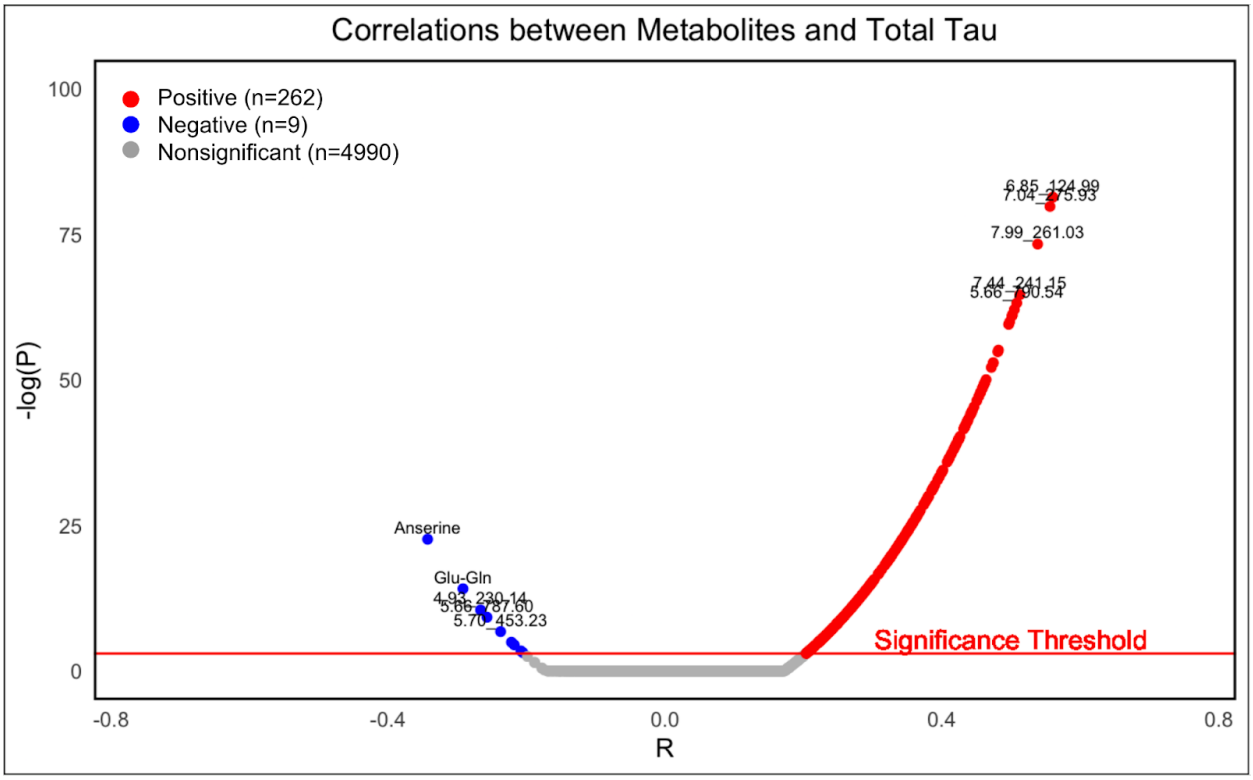
Correlation results between 5,261 CSF metabolites and CSF AD biomarkers. X-axis represents correlation values as an adjusted R. Y-axis represents log transformed p-values. Red dots indicate positive significantly correlated metabolites, blue dots indicate negative significantly correlated metabolites. Red line represents the Bonferroni adjusted significance threshold of -log(0.05).

(A) Shows results for CSF Total Tau

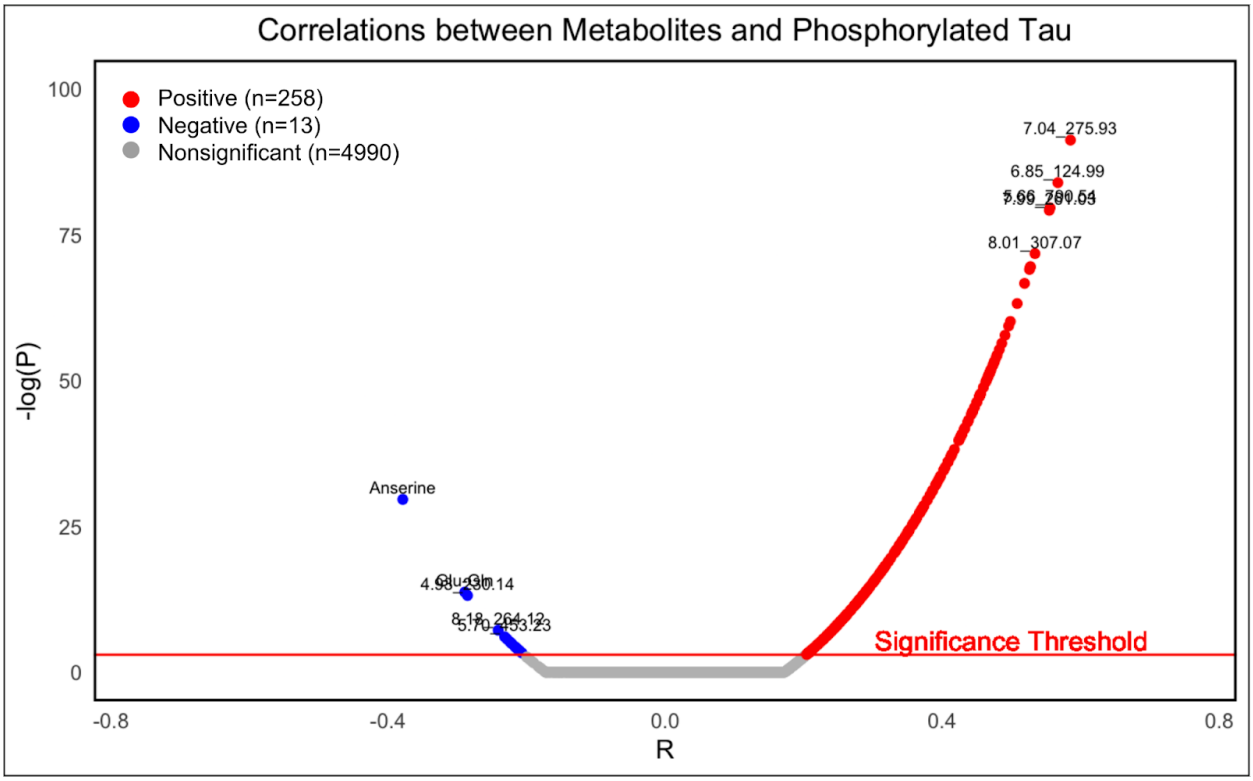
(B) Shows results for CSF Phosphorylated Tau

(C) Shows results for CSF Amyloid-Beta 42 (Aβ42

A




B


Correlations between Metabolites and Phosphorylated Tau
Positive (n=258)
Negative (n=13)
Nonsignificant (n=4990)
7.04_275.93
6.85_124.99
8.01_307.07
Anserine
Significance Threshold
-log(P)
R
100
75
50
25
0
-0.8
-0.4
0.0
0.4
0.8

C

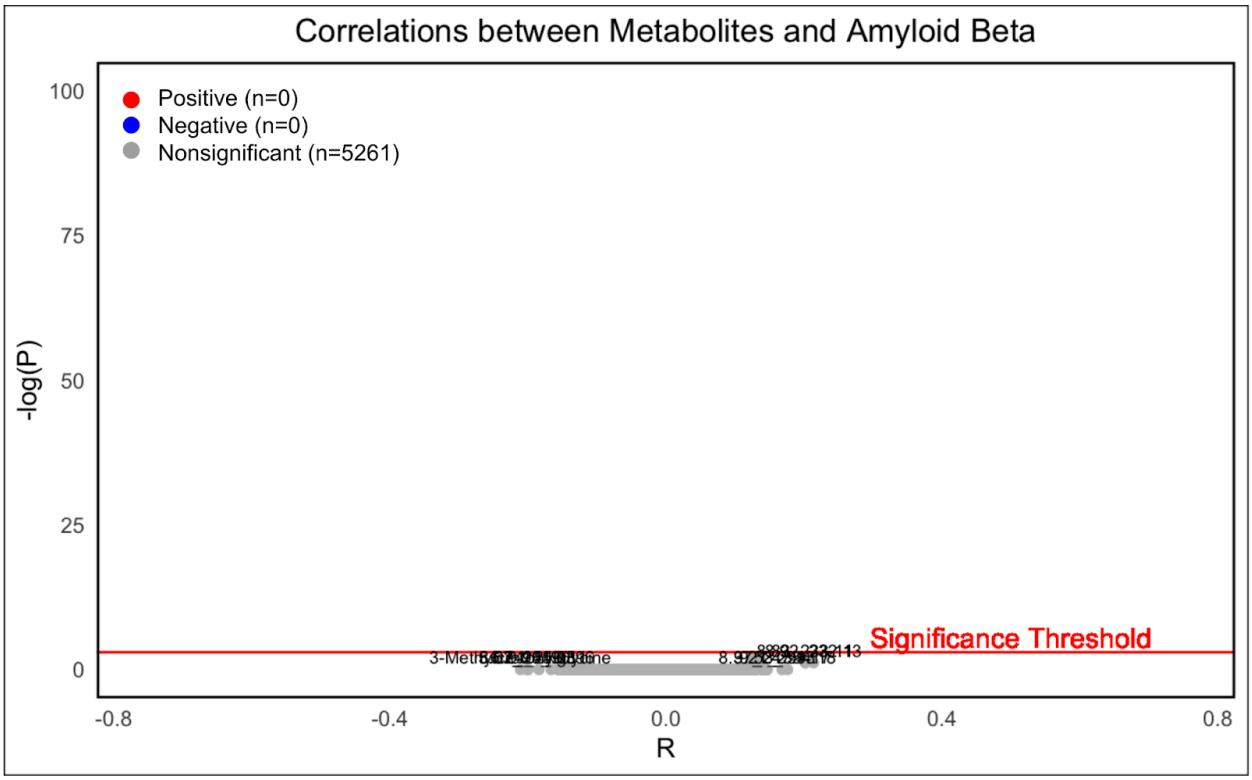
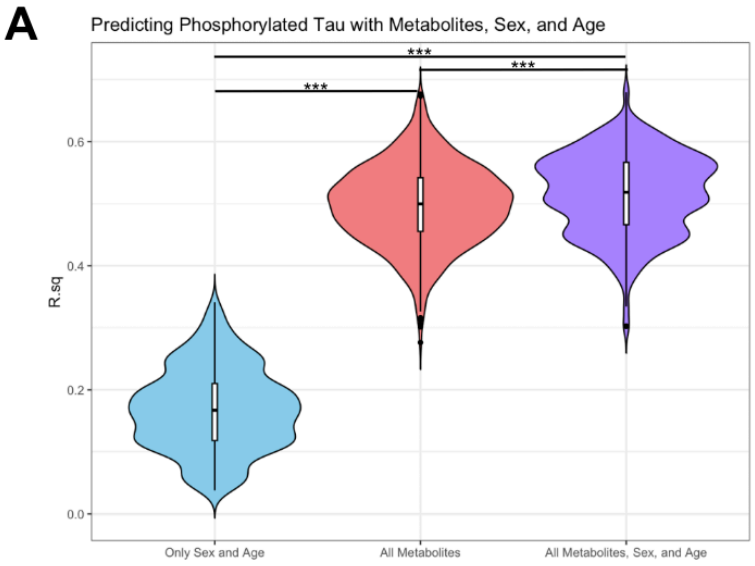
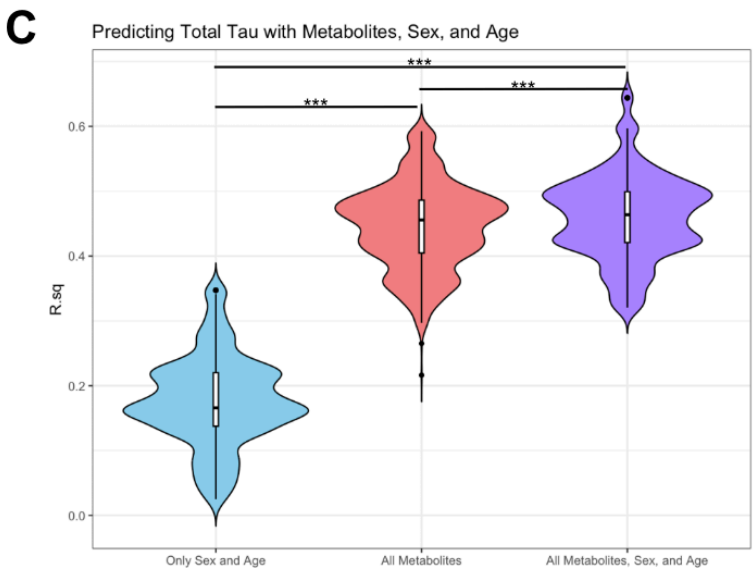

Correlations between Metabolites and Amyloid Beta
Positive (n=0)
Negative (n=0)
Nonsignificant (n=5261)
Significance Threshold
-log(P)
R
100
75
50
25
0
-0.8
-0.4
0.0
0.4
0.8

**Figure 2.3 Prediction of P-Tau and T-Tau CSF levels using CSF metabolites.**

Elastic net prediction results using only age and sex, all the CSF metabolites (n = 5,543), and a combined model of both. X-axis represents R2 values for the models. *** indicates p-value below 0.0001 for ANOVA test.

(A) P-Tau CSF levels prediction results




B

| P-Tau Consistent Named Predictors |
|---|
| Glu-Gln |
| isothreonic acid |
| Isoxanthopterin |
| gluconic acid |
| Anserine |
| PI 40:6 |
| PI 38:5 |
| 3-Methylcrotonylglycine |
| fucose |
| alpha-Galactosamine-1-phosphate |
| N-Methylvaline |
| PI 36:4 |
| Arabitol |

D

| T-Tau Consistent Named Predictors |
|---|
| Glu-Gln |
| Isoxanthopterin |
| alpha-Galactosamine-1-phosphate |
| isothreonic acid |
| 3-Methylcrotonylglycine |
| fucose |
| gluconic acid |
| SM d34:2 |

(B) All named CSF metabolites that consistently came up as predictors for P-Tau (n = 13)

(C) T-Tau CSF levels prediction results CSF levels B and D show all named CSF metabolites

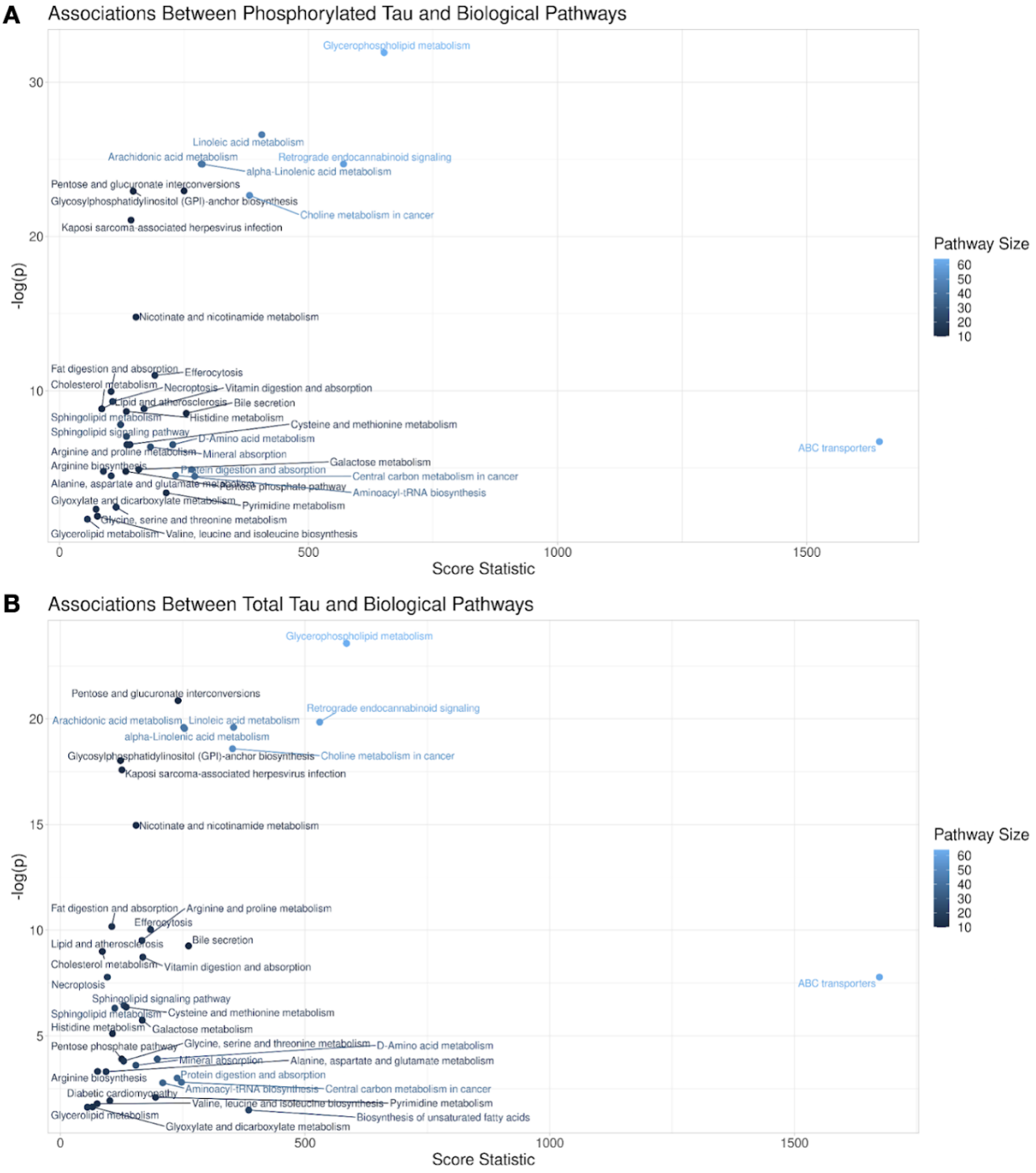
(D) All named CSF metabolites that consistently came up as predictors for T-Tau (n = 8) respectively.

**Figure 2.4 Pathway analysis of metabolites correlated with P-Tau and T-Tau.**

Results of pathway enrichment analysis for 288 metabolites. X-axis represents enrichment score, Y-axis represents log transformed p-values. Color indicates the amount of metabolites that were enriched for that particular pathway.

(A) Pathway enrichment results for P-Tau CSF levels

(B) Pathway enrichment results for T-Tau CSF levels.



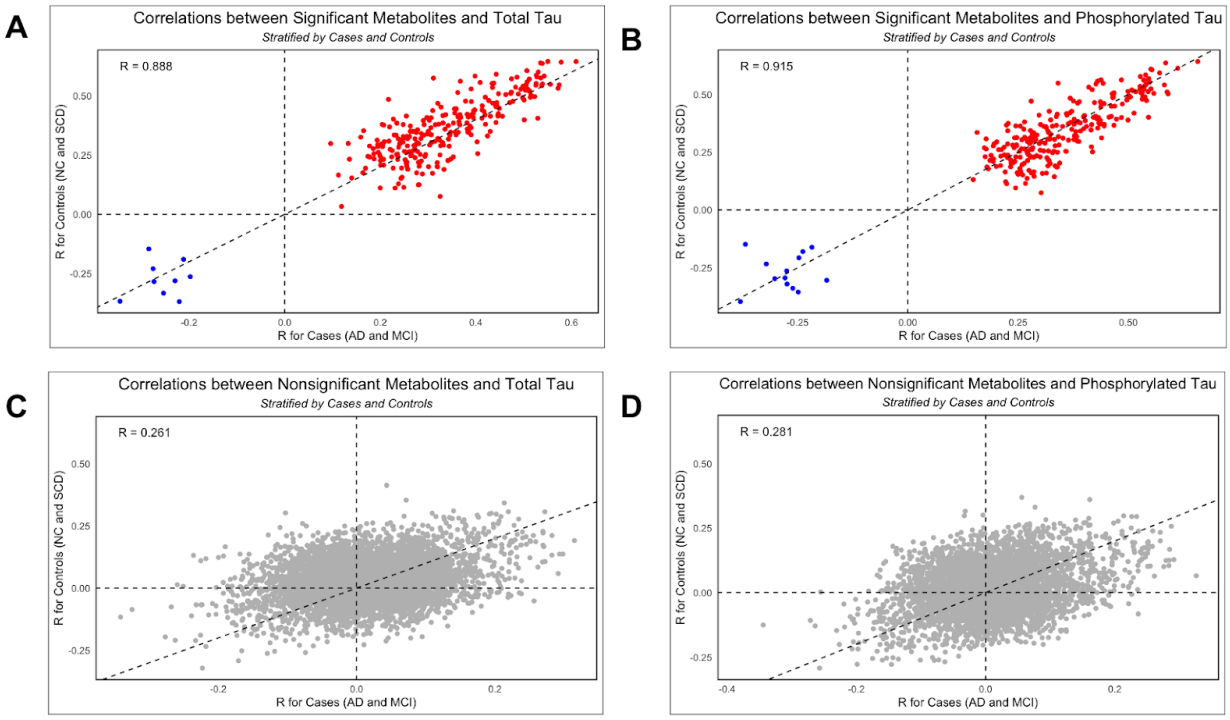

## CHAPTER 2 SUPPLEMENTARY FIGURES

**Supplementary Figure S2.1 CSF metabolites and AD CSF Biomarkers correlations stratified by cofactors**

Stratification of CSF metabolites and AD CSF biomarkers total tau and phosphorylated tau, by Case/Control status. Red and blue colors represent those metabolites found to be significantly positively or negatively correlated in the whole clinical cohort.

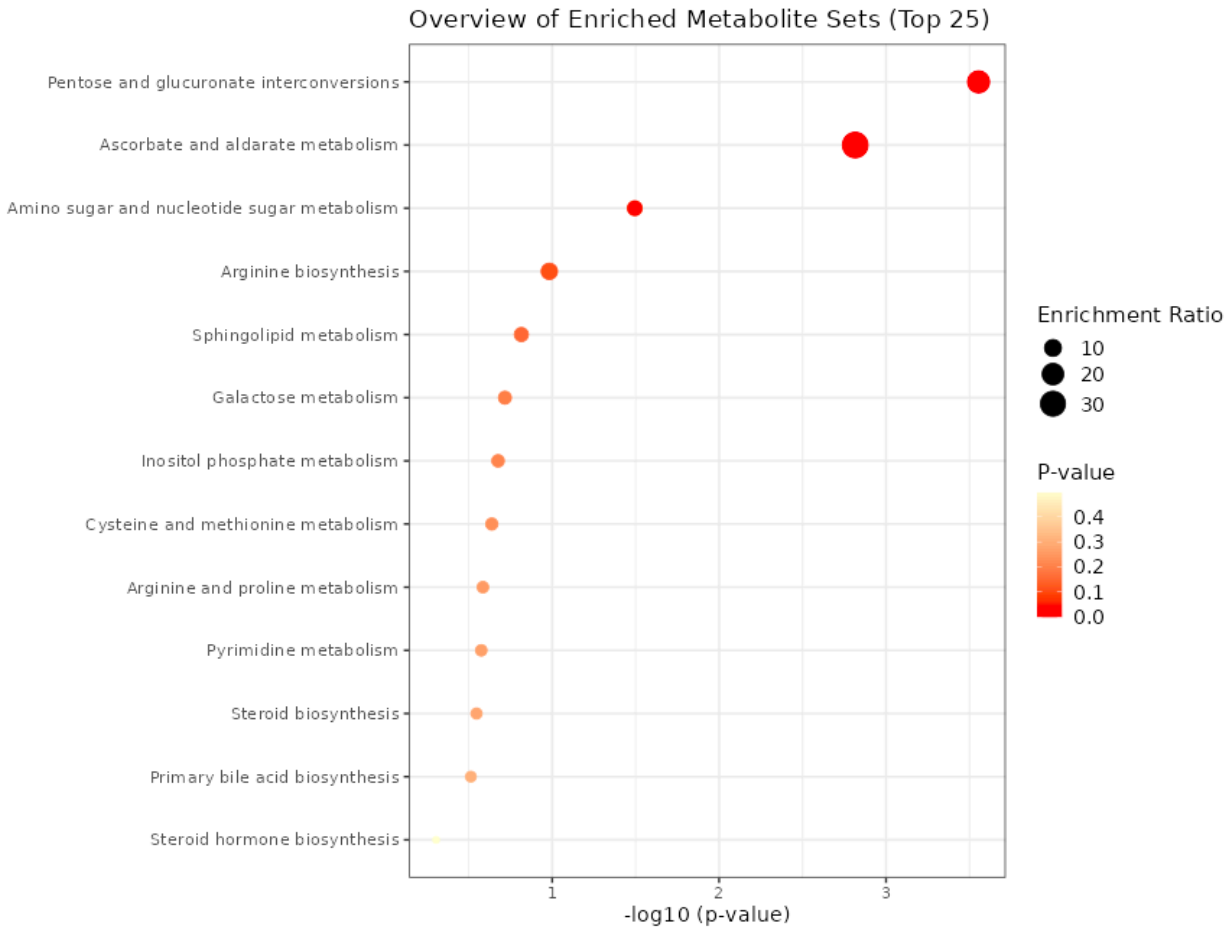
(A) & (B) shows only the significant CSF metabolites correlations for CSF Total Tau and P-Tau respectively

(C) & (D) shows all the non-significant CSF metabolites correlations for CSF Total Tau and P-Tau respectively.




**Supplementary Figure S2.2 Pathway enrichment analysis results by MetaboAnalist**

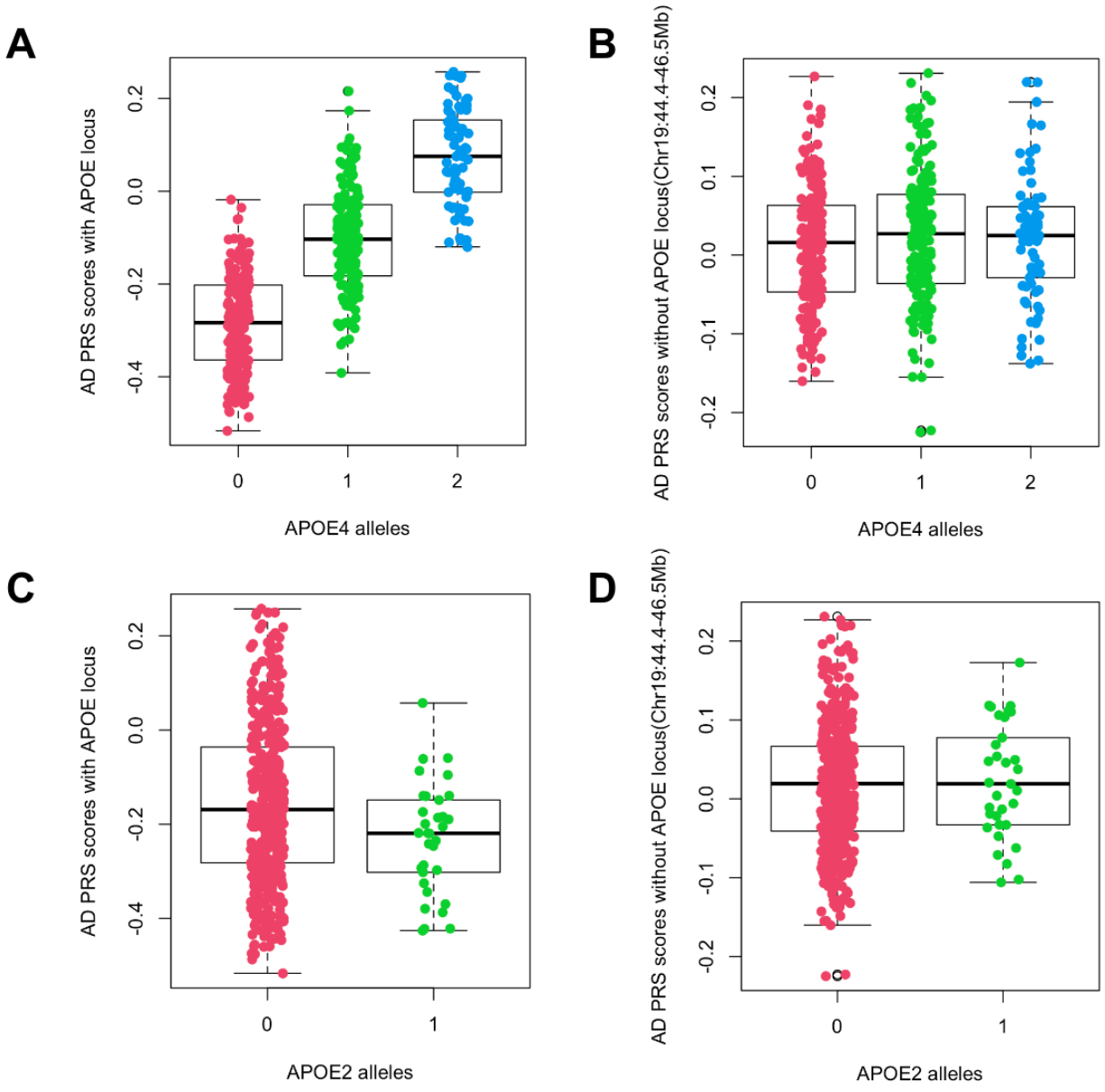
Metabolite pathway enrichment analysis results by MetaboAnalist for CSF metabolites correlated with both CSF P-Tau and T-Tau levels (N = 288). Y-axis shows enriched metabolic pathways, x-axis indicates -log10 of the P-value. Color indicates P-value of the enrichment.




**Supplementary Figure S2.3 Effect of APOE alleles on the AD PRS**

(A) & (B) show the distribution of the AD polygenic risk scores across APOE e4 allele counts, before and after removing the APOE locus region (Chr19: 44.4-46.5 mB).

(C) & (D) show the distribution of the AD scores across APOE e2 allele counts, before and after removing the APOE locus.



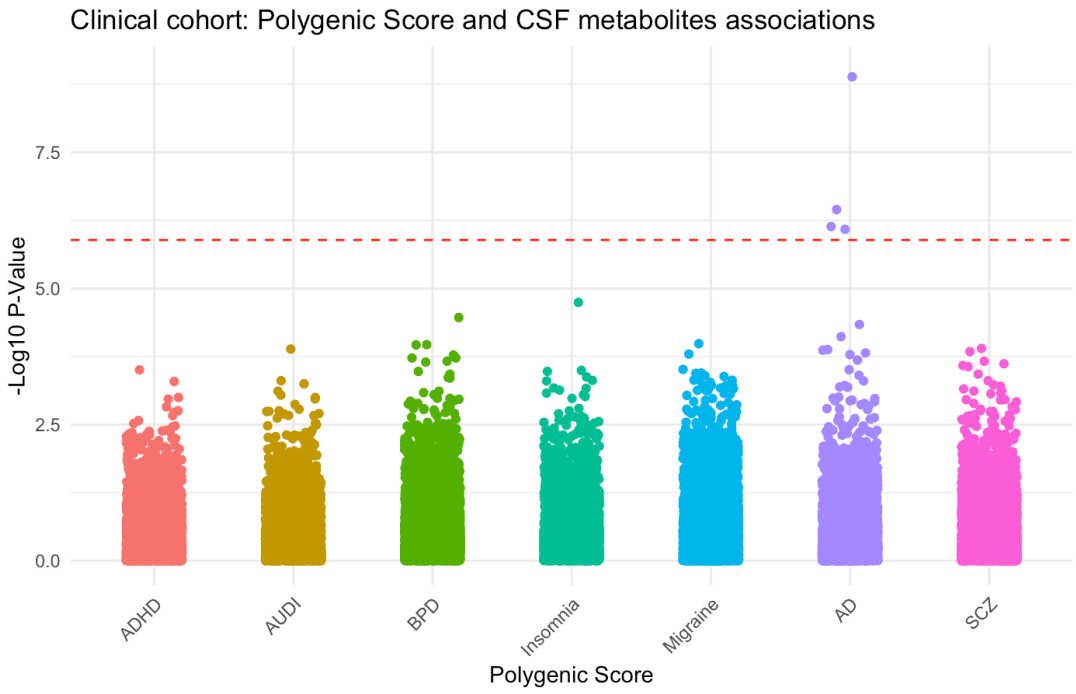

**Supplementary Figure S2.4 Associations between CSF metabolites and Polygenic Scores**

Linear regression analysis results for polygenic scores and CSF metabolites (n = 5,543). Y axis represents -log P value of the CSF metabolites and polygenic score association. Each dot represents a metabolite. Red line indicates Bonferroni adjusted P-value (alpha = 0.05). ADHD = Attention-Deficit / Hyperactivity Disorder; AUDI = Alcohol Use Disorder Identification Test; BPD = Bipolar Disorder; AD = Alzheimer's Disease; SCZ = Schizophrenia

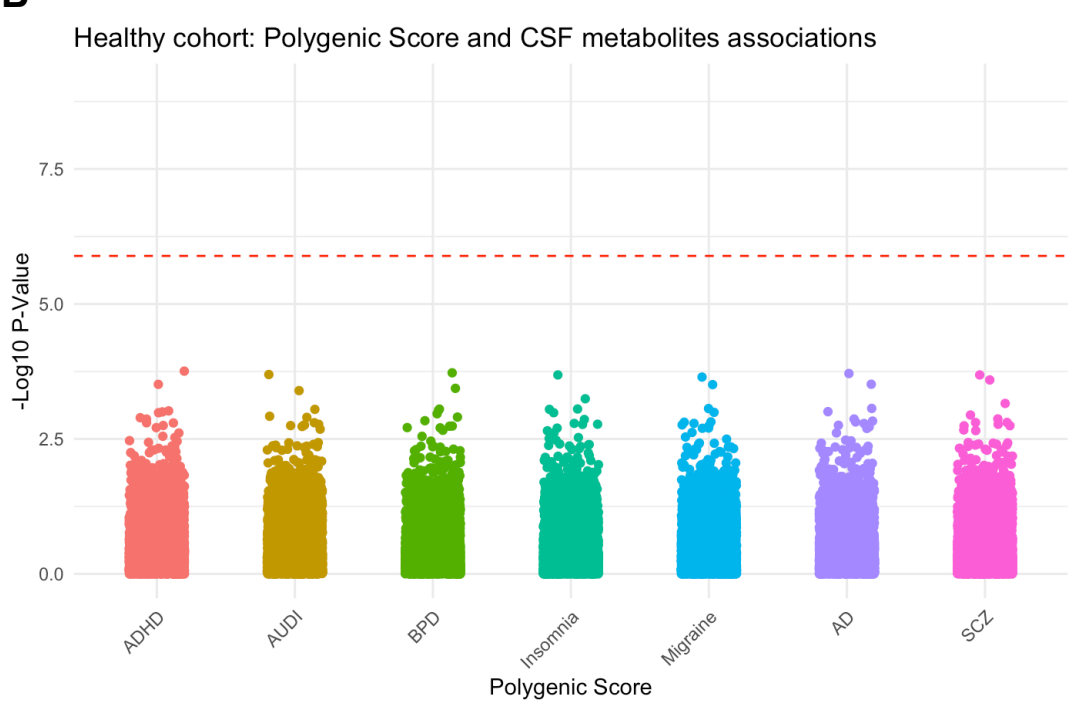
(A) Results for and the clinical cohort

(B) Results for the cognitively healthy cohort

**A**




**B**



## CONCLUSIONS

Current progress in precision medicine has been fueled by advancements in acquiring a broader range of individualized measurements, spanning electronic health records and biological data from various omics platforms [194,195]. While precision medicine approaches are currently being extensively utilized in oncology, particularly for selecting targeted therapies and assessing cancer risk [196], the accurate diagnosis and treatment of complex brain disorders remain challenging due to their heterogeneous and complex etiology. By leveraging their heritable component, genomic technologies, notably through GWAS [33], have identified candidate genes and deepened our understanding of the biological pathways underlying these disorders. Further advancements have stemmed from investigating how these genetic variants impact other biological layers. Transcriptome-wide association studies, for instance, identify trait-associated genes regulated by significant variants [197], while metabolome-wide association studies detect metabolites influenced by significant variants [198].

Another valuable resource for advancing research into the biology of complex brain disorders that has emerged in recent years involves the establishment of large biobanks. These repositories, which collect biological and clinical data from human subjects, offer opportunities to identify biomarkers for diagnosis and prognosis [199]. Notable examples of biobanks include the UK Biobank [200], which has amassed a wealth of biological features across thousands of illnesses from over half a million participants, and the Danish National Patient Registry [201], which has collected comprehensive information at a national scale. Some research work on the UK Biobank has focused on integrating measurements of multiple phenotypes to assess their contributions to mental health [202]. Electronic health records (EHRs) within these Biobanks serve as invaluable sources of clinical information and potential endophenotypes for psychiatric disorders [203]. They consolidate patient disease diagnoses, laboratory assay results, medication statuses, and longitudinal measurements [204]. The longitudinal aspect in particular holds potential to gain insight into the development of a disease and identified

associated risk factors. Ongoing research efforts are dedicated to redefining phenotype definitions currently employed in EHRs and leveraging them to predict diagnostic conversions between psychiatric disorders [205]. While further work is needed to fully integrate these endophenotypes into clinical practice, the linkage of EHRs, biological measures, and genomics underscores the potential of integrating information across clinical datasets and omics technologies.

To study endophenotypes for complex brain disorders, in vitro models utilizing patient-derived cell lines offer a means to capture the genetic architecture of these conditions, serving as accessible and powerful tools for elucidating their biological underpinnings [206]. The results from our study on the in vitro circadian model employing skin cells, in which we examined transcriptomic and open-chromatin data, revealed that the amount of consistent cyclic genes was limited to only 7 core circadian genes. Additionally, accessible chromatin features captured by this model of the circadian system were primarily related to the glucocorticoid response. It was observed that these features are influenced by genetics, cell culture conditions, and, importantly, the method used for circadian cycle synchronization. This modeling approach may lead to a narrowing in the scope of studies of circadian rhythms in the context of complex brain traits. Although the biology captured by this model after circadian synchronization induced by dexamethasone treatment may not be directly involved in the known genetic architecture of BD or other complex brain disorders, it can still be applied to scientific questions that cannot be explored directly in human subjects. For example, this model could be utilized to characterize the specific biological pathways activated during circadian distress. Notably, the dysregulated circadian phenotype in BD patients is characterized by episodic events rather than a constant trait, highlighting the dynamic nature of the disorder [7]. Comparative analysis of responses to circadian distress between fibroblast cell lines derived from BD patients and healthy individuals could provide valuable insights. Furthermore, the accessibility offered by this in vitro model could facilitate investigations into

the effects of lithium, the most commonly prescribed drug treatment for BD [14], during circadian distress scenarios.

Metabolomics has emerged as a potent omics platform capable of complementing information from genomics, by capturing downstream effects of variations in the environment, genome, and other biological layers for an individual [207]. Particularly, metabolites from CSF can offer insights into biological processes related to complex brain traits [152]. In this dissertation, our aim was to assess how genetics and CSF metabolites contribute to capturing various aspects of the pathology of AD that is reflected by the clinical CSF biomarkers. We found that the APOE E4 alleles (in particular) have a major impact on AB CSF levels, and to a moderate degree on P-Tau, and T-Tau CSF levels. By analyzing 5,543 CSF metabolites, we identified 288 unique CSF metabolites associated with CSF levels of P-Tau and T-Tau, but not with AB. We discovered novel associations between the CSF metabolites Anserine and Fucose, and P-Tau/T-Tau CSF levels. Pathway analysis of these metabolites further supports their involvement in established biological pathways affected by AD, such as glycerophospholipid metabolism. These metabolites can be potential targets for studying and gaining a deeper understanding of how AD progression impacts brain physiology.

While this dissertation has uncovered novel associations with AD, further research is needed to fully grasp the relationship between the genetic architecture of AD and its pathophysiology. Notably, our investigation revealed a robust genetic link between APOE E4 alleles and amyloid beta 42 CSF levels, a connection previously documented in multiple studies [44,208–210]. Despite the APOE locus being the most prominent genetic risk factor for AD, it does not encompass the entirety of the genetic risk. The most recent GWAS identified over 75 risk loci with 31 genes associated with AD [37]. The substantial effect size of the APOE locus in our study limited our ability to glean insights from other genetic regions, even with the integration of diverse omics approaches. Future investigations into AD must untangle the significant influence of APOE to evaluate the contribution of other genetic variants to AD pathology. One avenue for further exploration lies in examining cohorts of individuals with AD

who do not carry the APOE E4 allele. Research adopting this approach has uncovered differences in dementia risk factors between non-carriers and carriers of the APOE E4 allele, including variations in multimorbidity risk scores, elevated microalbumin levels in urine, and increased neutrophil counts [211]. Another promising strategy involves studying populations in which other genetic loci exhibit effect sizes comparable to APOE, such as African American populations with the ABCA7 locus [184]. Investigations focusing on non-European ancestry populations underscore the potential benefits of broadening current human genetic studies to encompass individuals from globally diverse ancestries.

In our analysis, we focused solely on common genetic variants, specifically single nucleotide polymorphisms (SNPs), which are single-base variations occurring in at least 1% of European ancestry populations [212]. While SNPs represent the most prevalent genetic variations, research has shown that rare variants, such as structural variants, are also implicated in complex brain disorders. These structural variants, characterized by duplications, insertions, or inversions of DNA segments, can identify specific genes, mutations, and downstream functional effects with high impact on a trait [213]. Preliminary investigations into AD have identified 16 structural variants, including deletions and duplications [214]. The first large-scale studies of these types of rare variants in BD suggest a limited contribution toward disease susceptibility. [215,216]. Despite the challenges posed by their low prevalence in current study populations, studying these rare variants could offer insights into shared disease pathways underlying complex brain disorders.

The solution to biomarker discovery does not solely rest upon a single measurement. When a single strong genetic locus predominantly influences biomarkers, it can obscure the exploration of additional biological mechanisms involving other genetic variants, as evidenced in our study of APOE and AD susceptibility. Conversely, in conditions like BD, where a diverse array of genetic variants each contribute small effects, the interpretability and identification of biologically relevant signals pose continuous challenges. This is compounded by the lack of biological features, such as inflammatory markers or circulating brain protein levels [31], that

we can definitely associate with BD. To advance our understanding, we must develop strategies to investigate the broader genetic landscape of the disorder beyond a single, most significantly associated locus. Additionally, identifying the optimal combination of features that will yield the most informative insights into the disorder is imperative and remains as one of the hardest tasks ahead.

In conclusion, the identification of biomarkers to enhance the diagnosis, risk assessment, and treatment of complex brain disorders remains a significant challenge in the field of neurology and psychiatry. Integration of multiple omics technologies holds promise in addressing these challenges, however, driven by ongoing advancements in refining the integrative analysis between technologies and broadening the populations from which these datasets are collected. In this dissertation, we leveraged two layers of functional genomics information—transcriptomics and epigenomics—to assess the suitability of an in vitro model as a tool for studying endophenotypes in BD. Additionally, we investigated known biomarkers of AD by examining their associations with genomics and CSF metabolomics, identifying novel CSF metabolites and biological pathways linked to the disorder. This work represents another step forward in advancing our understanding of the underlying biology of complex brain disorders, utilizing current technological advancements, and addressing the challenges of combining omics approaches.

## REFERENCES


[1] GBD 2021 Nervous System Disorders Collaborators, Global, Regional, and National Burden of Disorders Affecting the Nervous System, 1990-2021: A Systematic Analysis for the Global Burden of Disease Study 2021, Lancet Neurol. 23, 344 (2024).

[2] C. Zhang, Etiology of Alzheimer's Disease, Discov. Med. 35, 757 (2023).

[3] G. M. McKhann et al., The Diagnosis of Dementia due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease, Alzheimers. Dement. 7, 263 (2011).

[4] S. Duong, T. Patel, and F. Chang, Dementia: What Pharmacists Need to Know, Can. Pharm. J. 150, 118 (2017).

[5] J. M. Long and D. M. Holtzman, Alzheimer Disease: An Update on Pathobiology and Treatment Strategies, Cell 179, 312 (2019).

[6] M. L. Phillips and D. J. Kupfer, Bipolar Disorder Diagnosis: Challenges and Future Directions, Lancet 381, 1663 (2013).

[7] I. Grande, M. Berk, B. Birmaher, and E. Vieta, Bipolar Disorder, Lancet 387, 1561 (2016).

[8] Diagnostic and Statistical Manual of Mental Disorders: DSM-5TM, 5th Ed, 5, 5 (2013).

[9] P. B. Mitchell and G. S. Malhi, Bipolar Depression: Phenomenological Overview and Clinical Characteristics, Bipolar Disord. 6, 530 (2004).

[10] E. G. Hantouche and H. S. Akiskal, Bipolar II vs. Unipolar Depression: Psychopathologic Differentiation by Dimensional Measures, J. Affect. Disord. 84, 127 (2005).

[11] R. M. Hirschfeld et al., Development and Validation of a Screening Instrument for Bipolar Spectrum Disorder: The Mood Disorder Questionnaire, Am. J. Psychiatry 157, 1873 (2000).

[12] M. Valentí, I. Pacchiarotti, J. Undurraga, C. M. Bonnín, D. Popovic, J. M. Goikolea, C. Torrent, D. Hidalgo-Mazzei, F. Colom, and E. Vieta, Risk Factors for Rapid Cycling in Bipolar Disorder, Bipolar Disord. 17, 549 (2015).

[13] F. S. Goes, Diagnosis and Management of Bipolar Disorders, BMJ 381, e073591 (2023).

[14] Y. Lin, R. Mojtabai, F. S. Goes, and P. P. Zandi, Trends in Prescriptions of Lithium and Other Medications for Patients with Bipolar Disorder in Office-Based Practices in the United States: 1996-2015, J. Affect. Disord. 276, 883 (2020).

[15] C. A. Hunter, N. Y. Kirson, U. Desai, A. K. G. Cummings, D. E. Faries, and H. G. Birnbaum, Medical Costs of Alzheimer's Disease Misdiagnosis among US Medicare Beneficiaries, Alzheimers. Dement. 11, 887 (2015).

[16] H. Shen, L. Zhang, C. Xu, J. Zhu, M. Chen, and Y. Fang, Analysis of Misdiagnosis of Bipolar Disorder in An Outpatient Setting, Shanghai Arch Psychiatry 30, 93 (2018).

[17] J. L. Wilson and R. B. Altman, Biomarkers: Delivering on the Expectation of Molecularly Driven, Quantitative Health, Exp. Biol. Med. 243, 313 (2018).

[18] FDA-NIH Biomarker Working Group, BEST (Biomarkers, EndpointS, and Other Tools) Resource (Food and Drug Administration (US), 2016).

[19] R. M. Califf, Biomarker Definitions and Their Applications, Exp. Biol. Med. 243, 213 (2018).

[20] M. Vergeer, A. G. Holleboom, J. J. P. Kastelein, and J. A. Kuivenhoven, The HDL Hypothesis: Does High-Density Lipoprotein Protect from Atherosclerosis?, J. Lipid Res. 51, 2058 (2010).

[21] I. I. Gottesman and T. D. Gould, The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions, Am. J. Psychiatry 160, 636 (2003).

[22] G. Hasler, W. C. Drevets, T. D. Gould, I. I. Gottesman, and H. K. Manji, Toward Constructing an Endophenotype Strategy for Bipolar Disorders, Biol. Psychiatry 60, 93 (2006).

[23] V. K. Sarhadi and G. Armengol, Molecular Biomarkers in Cancer, Biomolecules 12, (2022).

[24] B. Fang, R. J. Mehran, J. V. Heymach, and S. G. Swisher, Predictive Biomarkers in Precision Medicine and Drug Development against Lung Cancer, Chin. J. Cancer 34, 295 (2015).

[25] P. Scheltens, B. De Strooper, M. Kivipelto, H. Holstege, G. Chételat, C. E. Teunissen, J. Cummings, and W. M. van der Flier, Alzheimer's Disease, Lancet 397, 1577 (2021).

[26] J. Lagarde, P. Olivieri, M. Tonietto, C. Tissot, I. Rivals, P. Gervais, F. Caillé, M. Moussion, M. Bottlaender, and M. Sarazin, Tau-PET Imaging Predicts Cognitive Decline and Brain Atrophy Progression in Early Alzheimer's Disease, J. Neurol. Neurosurg. Psychiatry 93, 459 (2022).

[27] C. R. Jack Jr et al., NIA-AA Research Framework: Toward a Biological Definition of Alzheimer's Disease, Alzheimers. Dement. 14, 535 (2018).

[28] D. Altomare et al., Applying the ATN Scheme in a Memory Clinic Population: The ABIDE Project, Neurology 93, e1635 (2019).

[29] M. Kalia and J. Costa E Silva, Biomarkers of Psychiatric Diseases: Current Status and Future Prospects, Metabolism 64, S11 (2015).

[30] B. I. Goldstein and L. T. Young, Toward Clinically Applicable Biomarkers in Bipolar Disorder: Focus on BDNF, Inflammatory Markers, and Endothelial Function, Curr. Psychiatry Rep. 15, 425 (2013).

[31] J. R. Cardoso de Almeida and M. L. Phillips, Distinguishing between Unipolar Depression and Bipolar Depression: Current and Future Clinical and Neuroimaging Perspectives, Biol. Psychiatry 73, 111 (2013).

[32] A. Abdellaoui, L. Yengo, K. J. H. Verweij, and P. M. Visscher, 15 Years of GWAS Discovery: Realizing the Promise, Am. J. Hum. Genet. 110, 179 (2023).

[33] E. Uffelmann, Q. Q. Huang, N. S. Munung, J. de Vries, Y. Okada, A. R. Martin, H. C. Martin, T. Lappalainen, and D. Posthuma, Genome-Wide Association Studies, Nature Reviews Methods Primers 1, 1 (2021).

[34] K. Zhang et al., Genetic Implication of a Novel Thiamine Transporter in Human Hypertension, J. Am. Coll. Cardiol. 63, 1542 (2014).

[35] M. Gatz, C. A. Reynolds, L. Fratiglioni, B. Johansson, J. A. Mortimer, S. Berg, A. Fiske, and N. L. Pedersen, Role of Genes and Environments for Explaining Alzheimer Disease, Arch. Gen. Psychiatry 63, 168 (2006).

[36] I. K. Karlsson, V. Escott-Price, M. Gatz, J. Hardy, N. L. Pedersen, M. Shoai, and C. A. Reynolds, Measuring Heritable Contributions to Alzheimer's Disease: Polygenic Risk Score Analysis with Twins, Brain Commun 4, fcab308 (2022).

[37] C. Bellenguez et al., New Insights into the Genetic Etiology of Alzheimer's Disease and Related Dementias, Nat. Genet. 54, 412 (2022).

[38] E. H. Corder, A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines, and M. A. Pericak-Vance, Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families, Science 261, 921 (1993).

[39] Y. Huang, K. H. Weisgraber, L. Mucke, and R. W. Mahley, Apolipoprotein E: Diversity of Cellular Origins, Structural and Biophysical Properties, and Effects in Alzheimer's Disease, J. Mol. Neurosci. 23, 189 (2004).

[40] A.-C. Raulin, S. V. Doss, Z. A. Trottier, T. C. Ikezu, G. Bu, and C.-C. Liu, ApoE in Alzheimer's Disease: Pathophysiology and Therapeutic Strategies, Mol. Neurodegener. 17, 72 (2022).

[41] O. J. Bienvenu, D. S. Davydow, and K. S. Kendler, Psychiatric "Diseases" versus Behavioral Disorders and Degree of Genetic Influence, Psychol. Med. 41, 33 (2011).

[42] J. Song, S. E. Bergen, R. Kuja-Halkola, H. Larsson, M. Landén, and P. Lichtenstein, Bipolar Disorder and Its Relation to Major Psychiatric Disorders: A Family-Based Study in the Swedish Population, Bipolar Disord. 17, 184 (2015).

[43] N. Mullins et al., Genome-Wide Association Study of More than 40,000 Bipolar Disorder Cases Provides New Insights into the Underlying Biology, Nat. Genet. 53, 817 (2021).

[44] F. Panza, D. Seripa, G. D'Onofrio, V. Frisardi, V. Solfrizzi, P. Mecocci, and A. Pilotto, Neuropsychiatric Symptoms, Endophenotypes, and Syndromes in Late-Onset Alzheimer's Disease: Focus on APOE Gene, Int. J. Alzheimers. Dis. 2011, 721457 (2011).

[45] M. N. Braskie, J. M. Ringman, and P. M. Thompson, Neuroimaging Measures as Endophenotypes in Alzheimer's Disease, Int. J. Alzheimers. Dis. 2011, 490140 (2011).

[46] R. Guglielmo, K. W. Miskowiak, and G. Hasler, Evaluating Endophenotypes for Bipolar Disorder, Int J Bipolar Disord 9, 17 (2021).

[47] Y. Cui et al., White Matter Microstructural Differences across Major Depressive Disorder, Bipolar Disorder and Schizophrenia: A Tract-Based Spatial Statistics Study, J. Affect. Disord. 260, 281 (2020).

[48] W. H. Walker 2nd, J. C. Walton, A. C. DeVries, and R. J. Nelson, Circadian Rhythm Disruption and Mental Health, Transl. Psychiatry 10, 28 (2020).

[49] Y. Chen et al., Neuroimmune Transcriptome Changes in Patient Brains of Psychiatric and Neurological Disorders, Mol. Psychiatry 28, 710 (2023).

[50] S. E. Jones et al., Genome-Wide Association Analyses of Chronotype in 697,828 Individuals Provides Insights into Circadian Rhythms, Nat. Commun. 10, 343 (2019).

[51] B. Zhao et al., Common Genetic Variation Influencing Human White Matter Microstructure, Science 372, (2021).

[52] W. G. Iacono, Endophenotypes in Psychiatric Disease: Prospects and Challenges, Genome Med. 10, 11 (2018).

[53] E. Drucker and K. Krapfenbauer, Pitfalls and Limitations in Translation from Biomarker Discovery to Clinical Utility in Predictive and Personalised Medicine, EPMA J. 4, 7 (2013).

[54] J. E. McDermott, J. Wang, H. Mitchell, B.-J. Webb-Robertson, R. Hafen, J. Ramey, and K. D. Rodland, Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data, Expert Opin. Med. Diagn. 7, 37 (2013).

[55] M. Horstmann, O. Patschan, J. Hennenlotter, E. Senger, G. Feil, and A. Stenzl, Combinations of Urine-Based Tumour Markers in Bladder Cancer Surveillance, Scand. J. Urol. Nephrol. 43, 461 (2009).

[56] E. Demir Karaman and Z. Işık, Multi-Omics Data Analysis Identifies Prognostic Biomarkers across Cancers, Med Sci (Basel) 11, (2023).

[57] F. Crick, Central Dogma of Molecular Biology, Nature 227, 561 (1970).

[58] D. Palm, A. Uzoni, G. Kronenberg, J. Thome, and F. Faltraco, Human Derived Dermal Fibroblasts as in Vitro Research Tool to Study Circadian Rhythmicity in Psychiatric Disorders, Pharmacopsychiatry 56, 87 (2023).

[59] S. Yang, H. P. A. Van Dongen, K. Wang, W. Berrettini, and M. Bućan, Assessment of Circadian Function in Fibroblasts of Patients with Bipolar Disorder, Mol. Psychiatry 14, 143 (2009).

[60] M. J. McCarthy, M. J. Le Roux, H. Wei, S. Beesley, J. R. Kelsoe, and D. K. Welsh, Calcium Channel Genes Associated with Bipolar Disorder Modulate Lithium's Amplification of Circadian Rhythms, Neuropharmacology 101, 439 (2016).

[61] M. J. McCarthy et al., Chronotype and Cellular Circadian Rhythms Predict the Clinical Response to Lithium Maintenance Treatment in Patients with Bipolar Disorder, Neuropsychopharmacology 44, 620 (2019).

[62] H. K. Mishra et al., Contributions of Circadian Clock Genes to Cell Survival in Fibroblast Models of Lithium-Responsive Bipolar Disorder, Eur. Neuropsychopharmacol. 74, 1 (2023).

[63] A. L. R. Moreira, A. Van Meter, J. Genzlinger, and E. A. Youngstrom, Review and Meta-Analysis of Epidemiologic Studies of Adult Bipolar Disorder, J. Clin. Psychiatry 78, e1259 (2017).

[64] E. A. Stahl et al., Genome-Wide Association Study Identifies 30 Loci Associated with Bipolar Disorder, Nat. Genet. 51, 793 (2019).

[65] E. Leibenluft, P. S. Albert, N. E. Rosenthal, and T. A. Wehr, Relationship between Sleep and Mood in Patients with Rapid-Cycling Bipolar Disorder, Psychiatry Res. 63, 161 (1996).

[66] J. Levenson and E. Frank, Sleep and Circadian Rhythm Abnormalities in the Pathophysiology of Bipolar Disorder, Curr. Top. Behav. Neurosci. 5, 247 (2011).

[67] L. Girshkin, S. L. Matheson, A. M. Shepherd, and M. J. Green, Morning Cortisol Levels in Schizophrenia and Bipolar Disorder: A Meta-Analysis, Psychoneuroendocrinology 49, 187 (2014).

[68] M. T. van den Berg, V. L. Wester, A. Vreeker, M. A. Koenders, M. P. Boks, E. F. C. van Rossum, and A. T. Spijker, Higher Cortisol Levels May Proceed a Manic Episode and Are Related to Disease Severity in Patients with Bipolar Disorder, Psychoneuroendocrinology 119, 104658 (2020).

[69] N. Le Minh, F. Damiola, F. Tronche, G. Schütz, and U. Schibler, Glucocorticoid Hormones Inhibit Food-Induced Phase-Shifting of Peripheral Circadian Oscillators, EMBO J. 20, 7128 (2001).

[70] U. Schibler and P. Sassone-Corsi, A Web of Circadian Pacemakers, Cell 111, 919 (2002).

[71] J. S. Menet, S. Pescatore, and M. Rosbash, CLOCK:BMAL1 Is a Pioneer-like Transcription Factor, Genes Dev. 28, 8 (2014).

[72] M. S. Robles, S. J. Humphrey, and M. Mann, Phosphorylation Is a Central Mechanism for Circadian Control of Metabolism and Physiology, Cell Metab. 25, 118 (2017).

[73] S. Yamazaki and J. S. Takahashi, Real-Time Luminescence Reporting of Circadian Gene Expression in Mammals, Methods Enzymol. 393, 288 (2005).

[74] Y. Nakahata, M. Akashi, D. Trcka, A. Yasuda, and T. Takumi, The in Vitro Real-Time Oscillation Monitoring System Identifies Potential Entrainment Factors for Circadian Clocks, BMC Mol. Biol. 7, 5 (2006).

[75] A. Hida et al., Evaluation of Circadian Phenotypes Utilizing Fibroblasts from Patients with Circadian Rhythm Sleep Disorders, Transl. Psychiatry 7, e1106 (2017).

[76] M. J. McCarthy, H. Wei, Z. Marnoy, R. M. Darvish, D. L. McPhie, B. M. Cohen, and D. K. Welsh, Genetic and Clinical Factors Predict Lithium's Effects on PER2 Gene Expression Rhythms in Cells from Bipolar Disorder Patients, Transl. Psychiatry 3, e318 (2013).

[77] R. Zhang, N. F. Lahens, H. I. Ballance, M. E. Hughes, and J. B. Hogenesch, A Circadian Gene Expression Atlas in Mammals: Implications for Biology and Medicine, Proc. Natl. Acad. Sci. U. S. A. 111, 16219 (2014).

[78] Y. Wang, C. Ke, and M. B. Brown, Shape-Invariant Modeling of Circadian Rhythms with Random Effects and Smoothing Spline ANOVA Decompositions, Biometrics 59, 804 (2003).

[79] L. Qin and W. Guo, Functional Mixed-Effects Model for Periodic Data, Biostatistics 7, 225 (2006).

[80] J. M. Madden, X. Li, P. M. Kearney, K. Tilling, and A. P. Fitzgerald, Exploring Diurnal Variation Using Piecewise Linear Splines: An Example Using Blood Pressure, Emerg. Themes Epidemiol. 14, 1 (2017).

[81] P. Langfelder and S. Horvath, WGCNA: An R Package for Weighted Correlation Network Analysis, BMC Bioinformatics 9, 559 (2008).

[82] Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, and S. K. Chanda, Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets, Nat. Commun. 10, 1523 (2019).

[83] M. Del Olmo et al., Inter-Layer and Inter-Subject Variability of Diurnal Gene Expression in Human Skin, NAR Genom Bioinform 4, lqac097 (2022).

[84] M. E. Hughes, J. B. Hogenesch, and K. Kornacker, JTK_CYCLE: An Efficient Nonparametric Algorithm for Detecting Rhythmic Components in Genome-Scale Data Sets, J. Biol. Rhythms 25, 372 (2010).

[85] E. F. Glynn, J. Chen, and A. R. Mushegian, Detecting Periodic Patterns in Unevenly Spaced Gene Expression Time Series Using Lomb-Scargle Periodograms, Bioinformatics 22, 310 (2006).

[86] R. Yang and Z. Su, Analyzing Circadian Expression Data by Harmonic Regression Based on Autoregressive Spectral Estimation, Bioinformatics 26, i168 (2010).

[87] G. Wu, R. C. Anafi, M. E. Hughes, K. Kornacker, and J. B. Hogenesch, MetaCycle: An Integrated R Package to Evaluate Periodicity in Large Scale Data, Bioinformatics 32, 3351 (2016).

[88] P. F. Thaben and P. O. Westermark, Detecting Rhythms in Time Series with RAIN, J. Biol. Rhythms 29, 391 (2014).

[89] M. Berk, Smoothing-Splines Mixed-Effects Models in R Using the Sme Package: A Tutorial, https://rdrr.io/cran/sme/f/inst/doc/Tutorial.pdf.

[90] F. Rijo-Ferreira and J. S. Takahashi, Genomics of Circadian Rhythms in Health and Disease, Genome Med. 11, 82 (2019).

[91] G. R. Keele et al., Integrative QTL Analysis of Gene Expression and Chromatin Accessibility Identifies Multi-Tissue Patterns of Genetic Regulation, PLoS Genet. 16, e1008537 (2020).

[92] G. Yu, L.-G. Wang, and Q.-Y. He, ChIPseeker: An R/Bioconductor Package for ChIP Peak Annotation, Comparison and Visualization, Bioinformatics 31, 2382 (2015).

[93] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass, Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities, Mol. Cell 38, 576 (2010).

[94] H. K. Finucane et al., Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics, Nat. Genet. 47, 1228 (2015).

[95] D. Demontis et al., Discovery of the First Genome-Wide Significant Risk Loci for Attention Deficit/hyperactivity Disorder, Nat. Genet. 51, 63 (2019).

[96] V. Trubetskoy et al., Mapping Genomic Loci Implicates Genes and Synaptic Biology in Schizophrenia, Nature 604, 502 (2022).

[97] C. M. Nievergelt et al., International Meta-Analysis of PTSD Genome-Wide Association Studies Identifies Sex- and Ancestry-Specific Genetic Risk Loci, Nat. Commun. 10, 4558 (2019).

[98] D. M. Howard et al., Genome-Wide Meta-Analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions, Nat. Neurosci. 22, 343 (2019).

[99] K. Watanabe et al., Genome-Wide Meta-Analysis of Insomnia Prioritizes Genes Associated with Metabolic and Psychiatric Pathways, Nat. Genet. 54, 1125 (2022).

[100] S. E. Jones et al., Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci, PLoS Genet. 12, e1006125 (2016).

[101] A. Balsalobre, L. Marcacci, and U. Schibler, Multiple Signaling Pathways Elicit Circadian Gene Expression in Cultured Rat-1 Fibroblasts, Curr. Biol. 10, 1291 (2000).

[102] D. F. Kripke, C. M. Nievergelt, E. Joo, T. Shekhtman, and J. R. Kelsoe, Circadian Polymorphisms Associated with Affective Disorders, J. Circadian Rhythms 7, 2 (2009).

[103] A. Y.-L. So, T. U. Bernal, M. L. Pillsbury, K. R. Yamamoto, and B. J. Feldman, Glucocorticoid Regulation of the Circadian Clock Modulates Glucose Homeostasis, Proc. Natl. Acad. Sci. U. S. A. 106, 17582 (2009).

[104] A. C. Liu, H. G. Tran, E. E. Zhang, A. A. Priest, D. K. Welsh, and S. A. Kay, Redundant Function of REV-ERBalpha and Beta and Non-Essential Role for Bmal1 Cycling in Transcriptional Regulation of Intracellular Circadian Rhythms, PLoS Genet. 4, e1000023 (2008).

[105] Y.-Y. Chiou, Y. Yang, N. Rashid, R. Ye, C. P. Selby, and A. Sancar, Mammalian Period Represses and de-Represses Transcription by Displacing CLOCK-BMAL1 from Promoters in a Cryptochrome-Dependent Manner, Proc. Natl. Acad. Sci. U. S. A. 113, E6072 (2016).

[106] J. M. Fustin, J. S. O'Neill, M. H. Hastings, D. G. Hazlerigg, and H. Dardente, Cry1 Circadian Phase in Vitro: Wrapped up with an E-Box, J. Biol. Rhythms 24, 16 (2009).

[107] N. Koike, S.-H. Yoo, H.-C. Huang, V. Kumar, C. Lee, T.-K. Kim, and J. S. Takahashi, Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals, Science 338, 349 (2012).

[108] Q. Zhu and W. J. Belden, Molecular Regulation of Circadian Chromatin, J. Mol. Biol. 432, 3466 (2020).

[109] T. Roenneberg and M. Merrow, The Circadian Clock and Human Health, Curr. Biol. 26, R432 (2016).

[110] S. Chauhan, R. Norbury, K. C. Faßbender, U. Ettinger, and V. Kumari, Beyond Sleep: A Multidimensional Model of Chronotype, Neurosci. Biobehav. Rev. 148, 105114 (2023).

[111] M. Bothe, R. Buschow, and S. H. Meijsing, Glucocorticoid Signaling Induces Transcriptional Memory and Universally Reversible Chromatin Changes, Life Sci Alliance 4, (2021).

[112] K. Yagita and H. Okamura, Forskolin Induces Circadian Gene Expression of rPer1, rPer2 and Dbp in Mammalian Rat-1 Fibroblasts, FEBS Lett. 465, 79 (2000).

[113] C. Saini, J. Morf, M. Stratmann, P. Gos, and U. Schibler, Simulated Body Temperature Rhythms Reveal the Phase-Shifting Behavior and Plasticity of Mammalian Circadian Oscillators, Genes Dev. 26, 567 (2012).

[114] A. Balsalobre, F. Damiola, and U. Schibler, A Serum Shock Induces Circadian Gene Expression in Mammalian Tissue Culture Cells, Cell 93, 929 (1998).

[115] V. R. Iyer et al., The Transcriptional Program in the Response of Human Fibroblasts to Serum, Science 283, 83 (1999).

[116] S. A. Brown, F. Fleury-Olela, E. Nagoshi, C. Hauser, C. Juge, C. A. Meier, R. Chicheportiche, J.-M. Dayer, U. Albrecht, and U. Schibler, The Period Length of Fibroblast Circadian Gene Expression Varies Widely among Human Individuals, PLoS Biol. 3, e338 (2005).

[117] E. Nagoshi, C. Saini, C. Bauer, T. Laroche, F. Naef, and U. Schibler, Circadian Gene Expression in Individual Fibroblasts: Cell-Autonomous and Self-Sustained Oscillators Pass Time to Daughter Cells, Cell 119, 693 (2004).

[118] E. Farshadi, G. T. J. van der Horst, and I. Chaves, Molecular Links between the Circadian Clock and the Cell Cycle, J. Mol. Biol. 432, 3515 (2020).

[119] T. Noguchi, L. L. Wang, and D. K. Welsh, Fibroblast PER2 Circadian Rhythmicity Depends on Cell Density, J. Biol. Rhythms 28, 183 (2013).

[120] J. Villegas and M. McPhaul, Establishment and Culture of Human Skin Fibroblasts, Curr. Protoc. Mol. Biol. Chapter 28, Unit 28.3 (2005).

[121] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide, Curr. Protoc. Mol. Biol. 109, 21.29.1 (2015).

[122] Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[123] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, STAR: Ultrafast Universal RNA-Seq Aligner, Bioinformatics 29, 15 (2013).

[124] W. Mei, Z. Jiang, Y. Chen, L. Chen, A. Sancar, and Y. Jiang, Genome-Wide Circadian Rhythm Detection Methods: Systematic Evaluations and Practical Guidelines, Brief. Bioinform. 22, (2021).

[125] B. Langmead and S. L. Salzberg, Fast Gapped-Read Alignment with Bowtie 2, Nat. Methods 9, 357 (2012).

[126] Y. Zhang et al., Model-Based Analysis of ChIP-Seq (MACS), Genome Biol. 9, R137 (2008).

[127] A. R. Quinlan and I. M. Hall, BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features, Bioinformatics 26, 841 (2010).

[128] A. P. S. Ori, M. H. M. Bot, R. T. Molenhuis, L. M. Olde Loohuis, and R. A. Ophoff, A Longitudinal Model of Human Neuronal Differentiation for Functional Investigation of Schizophrenia Polygenic Risk, Biol. Psychiatry 85, 544 (2019).

[129] S. Gazal et al., Linkage Disequilibrium-Dependent Architecture of Human Complex Traits Shows Action of Negative Selection, Nat. Genet. 49, 1421 (2017).

[130] C. R. Jack Jr et al., Tracking Pathophysiological Processes in Alzheimer's Disease: An Updated Hypothetical Model of Dynamic Biomarkers, Lancet Neurol. 12, 207 (2013).

[131] F. H. Bouwman et al., Clinical Application of CSF Biomarkers for Alzheimer's Disease: From Rationale to Ratios, Alzheimers. Dement. 14, e12314 (2022).

[132] Y. Deming et al., Genome-Wide Association Study Identifies Four Novel Loci Associated with Alzheimer's Endophenotypes and Disease Modifiers, Acta Neuropathol. 133, 839 (2017).

[133] I. E. Jansen et al., Genome-Wide Meta-Analysis for Alzheimer's Disease Cerebrospinal Fluid Biomarkers, Acta Neuropathol. 144, 821 (2022).

[134] R. Dong et al., CSF Metabolites Associate with CSF Tau and Improve Prediction of Alzheimer's Disease Status, Alzheimers. Dement. 13, e12167 (2021).

[135] B. F. Darst, Q. Lu, S. C. Johnson, and C. D. Engelman, Integrated Analysis of Genomics, Longitudinal Metabolomics, and Alzheimer's Risk Factors among 1,111 Cohort Participants, Genet. Epidemiol. 43, 657 (2019).

[136] C. L. Berkowitz, L. Mosconi, A. Rahman, O. Scheyer, H. Hristov, and R. S. Isaacson, Clinical Application of APOE in Alzheimer's Prevention: A Precision Medicine Approach, J Prev Alzheimers Dis 5, 245 (2018).

[137] G. Leonenko, E. Baker, J. Stevenson-Hoare, A. Sierksma, M. Fiers, J. Williams, B. de Strooper, and V. Escott-Price, Identifying Individuals with High Risk of Alzheimer's Disease Using Polygenic Risk Scores, Nat. Commun. 12, 4506 (2021).

[138] C. Cruchaga et al., GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimer's Disease, Neuron 78, 256 (2013).

[139] T. Kremer et al., Longitudinal Analysis of Multiple Neurotransmitter Metabolites in Cerebrospinal Fluid in Early Parkinson's Disease, Mov. Disord. 36, 1972 (2021).

[140] L. A. Lotta et al., A Cross-Platform Approach Identifies Genetic Regulators of Human Metabolism and Health, Nat. Genet. 53, 54 (2021).

[141] W. M. van der Flier and P. Scheltens, Amsterdam Dementia Cohort: Performing Research to Optimize Care, J. Alzheimers. Dis. 62, 1091 (2018).

[142] N. Legdeur et al., Resilience to Cognitive Impairment in the Oldest-Old: Design of the EMIF-AD 90+ Study, BMC Geriatr. 18, 289 (2018).

[143] D. I. Boomsma et al., Netherlands Twin Register: From Twins to Twin Families, Twin Res. Hum. Genet. 9, 849 (2006).

[144] E. Konijnenberg et al., The EMIF-AD PreclinAD Study: Study Design and Baseline Cohort Overview, Alzheimers. Res. Ther. 10, 75 (2018).

[145] M.-E. Dumas and L. Davidovic, Metabolic Profiling and Phenotyping of Central Nervous System Diseases: Metabolites Bring Insights into Brain Dysfunctions, J. Neuroimmune Pharmacol. 10, 402 (2015).

[146] P. S Haglund, K. Löfstrand, K. Siek, and L. Asplund, Powerful GC-TOF-MS Techniques for Screening, Identification and Quantification of Halogenated Natural Products, Mass Spectrom. 2, S0018 (2013).

[147] S. Dhariwal, K. Maan, R. Baghel, A. Sharma, D. Malakar, and P. Rana, Systematic Untargeted UHPLC-Q-TOF-MS Based Lipidomics Workflow for Improved Detection and Annotation of Lipid Sub-Classes in Serum, Metabolomics 19, 24 (2023).

[148] G. Paglia and G. Astarita, A High-Throughput HILIC-MS-Based Metabolomic Assay for the Analysis of Polar Metabolites, Methods Mol. Biol. 2396, 137 (2022).

[149] D. P. Wightman et al., A Genome-Wide Association Study with 1,126,563 Individuals Identifies New Risk Loci for Alzheimer's Disease, Nat. Genet. 53, 1276 (2021).

[150] Y. Deming et al., Neuropathology-Based APOE Genetic Risk Score Better Quantifies Alzheimer's Risk, Alzheimers. Dement. (2023).

[151] J. J. Luykx et al., Genome-Wide Association Study of Monoamine Metabolite Levels in Human Cerebrospinal Fluid, Mol. Psychiatry 19, 228 (2014).

[152] D. J. Panyard et al., Cerebrospinal Fluid Metabolomics Identifies 19 Brain-Related Phenotype Associations, Commun Biol 4, 63 (2021).

[153] C. Wang et al., Identification of Genetic Regulators of Human Metabolites in CSF, Alzheimers. Dement. 19, (2023).

[154] A. Quintero et al., ShinyButchR: Interactive NMF-Based Decomposition Workflow of Genome-Scale Datasets, Biol Methods Protoc 5, bpaa022 (2020).

[155] H. Hautakangas et al., Genome-Wide Analysis of 102,084 Migraine Cases Identifies 123 Risk Loci and Subtype-Specific Risk Alleles, Nat. Genet. 54, 152 (2022).

[156] S. Sanchez-Roige et al., Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts, Am. J. Psychiatry 176, 107 (2019).

[157] A. R. Hipkiss et al., Pluripotent Protective Effects of Carnosine, a Naturally Occurring Dipeptide, Ann. N. Y. Acad. Sci. 854, 37 (1998).

[158] G. Caruso, F. Caraci, and R. B. Jolivet, Pivotal Role of Carnosine in the Modulation of Brain Cells Activity: Multimodal Mechanism of Action and Therapeutic Potential in Neurodegenerative Disorders, Prog. Neurobiol. 175, 35 (2019).

[159] J. Kaneko, A. Enya, K. Enomoto, Q. Ding, and T. Hisatsune, Anserine (beta-Alanyl-3-Methyl-L-Histidine) Improves Neurovascular-Unit Dysfunction and Spatial Memory in Aged AβPPswe/PSEN1dE9 Alzheimer's-Model Mice, Sci. Rep. 7, 12571 (2017).

[160] Q. Ding, K. Tanigawa, J. Kaneko, M. Totsuka, Y. Katakura, E. Imabayashi, H. Matsuda, and T. Hisatsune, Anserine/Carnosine Supplementation Preserves Blood Flow in the Prefrontal Brain of Elderly People Carrying APOE e4, Aging Dis. 9, 334 (2018).

[161] M. Schneider, E. Al-Shareffi, and R. S. Haltiwanger, Biological Functions of Fucose in Mammals, Glycobiology 27, 601 (2017).

[162] W. Pohle, L. Acosta, H. Rüthrich, M. Krug, and H. Matthies, Incorporation of [3H]fucose in Rat Hippocampal Structures after Conditioning by Perforant Path Stimulation and after LTP-Producing Tetanization, Brain Res. 410, 245 (1987).

[163] S. A. Kalovidouris, C. I. Gama, L. W. Lee, and L. C. Hsieh-Wilson, A Role for Fucose alpha(1-2) Galactose Carbohydrates in Neuronal Growth, J. Am. Chem. Soc. 127, 1340 (2005).

[164] W. Wetzel, N. Popov, B. Lössner, S. Schulzeck, R. Honza, and H. Matthies, Effect of L-Fucose on Brain Protein Metabolism and Retention of a Learned Behavior in Rats, Pharmacol. Biochem. Behav. 13, 765 (1980).

[165] X. Xu et al., Exogenous L-Fucose Attenuates Neuroinflammation Induced by Lipopolysaccharide, J. Biol. Chem. 300, 105513 (2024).

[166] X. Tang, J. Tena, J. Di Lucente, I. Maezawa, D. J. Harvey, L.-W. Jin, C. B. Lebrilla, and A. M. Zivkovic, Transcriptomic and Glycomic Analyses Highlight Pathway-Specific Glycosylation Alterations Unique to Alzheimer's Disease, Sci. Rep. 13, 7816 (2023).

[167] C.-F. Tu, F.-A. Li, L.-H. Li, and R.-B. Yang, Quantitative Glycoproteomics Analysis Identifies Novel FUT8 Targets and Signaling Networks Critical for Breast Cancer Cell Invasiveness, Breast Cancer Res. 24, 21 (2022).

[168] D. Hishikawa, T. Hashidate, T. Shimizu, and H. Shindou, Diversity and Function of Membrane Glycerophospholipids Generated by the Remodeling Pathway in Mammalian Cells, J. Lipid Res. 55, 799 (2014).

[169] A. A. Farooqui, L. A. Horrocks, and T. Farooqui, Glycerophospholipids in Brain: Their Metabolism, Incorporation into Membranes, Functions, and Involvement in Neurological Disorders, Chem. Phys. Lipids 106, 1 (2000).

[170] F. Yin, Lipid Metabolism and Alzheimer's Disease: Clinical Evidence, Mechanistic Link and Therapeutic Promise, FEBS J. 290, 1420 (2023).

[171] S. Akyol et al., Lipid Profiling of Alzheimer's Disease Brain Highlights Enrichment in Glycerol(phospho)lipid, and Sphingolipid Metabolism, Cells 10, (2021).

[172] J. Xu et al., Graded Perturbations of Metabolism in Multiple Regions of Human Brain in Alzheimer's Disease: Snapshot of a Pervasive Metabolic Disorder, Biochim. Biophys. Acta 1862, 1084 (2016).

[173] C. M. Burns, K. Chen, A. W. Kaszniak, W. Lee, G. E. Alexander, D. Bandy, A. S. Fleisher, R. J. Caselli, and E. M. Reiman, Higher Serum Glucose Levels Are Associated with Cerebral Hypometabolism in Alzheimer Regions, Neurology 80, 1557 (2013).

[174] R. van der Kant, L. S. B. Goldstein, and R. Ossenkoppele, Amyloid-β-Independent Regulators of Tau Pathology in Alzheimer Disease, Nat. Rev. Neurosci. 21, 21 (2020).

[175] S. A. Kent, T. L. Spires-Jones, and C. S. Durrant, The Physiological Roles of Tau and Aβ: Implications for Alzheimer's Disease Pathology and Therapeutics, Acta Neuropathol. 140, 417 (2020).

[176] B. F. Darst et al., Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-β Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease, J. Alzheimers. Dis. 55, 473 (2017).

[177] R. Lautner et al., Apolipoprotein E Genotype and the Diagnostic Accuracy of Cerebrospinal Fluid Biomarkers for Alzheimer Disease, JAMA Psychiatry 71, 1183 (2014).

[178] T. E. Mahan, C. Wang, X. Bao, A. Choudhury, J. D. Ulrich, and D. M. Holtzman, Selective Reduction of Astrocyte apoE3 and apoE4 Strongly Reduces Aβ Accumulation and Plaque-Related Pathology in a Mouse Model of Amyloidosis, Mol. Neurodegener. 17, 13 (2022).

[179] K. R. Wildsmith, M. Holley, J. C. Savage, R. Skerrett, and G. E. Landreth, Evidence for Impaired Amyloid β Clearance in Alzheimer's Disease, Alzheimers. Res. Ther. 5, 33 (2013).

[180] J. Therriault et al., Association of Apolipoprotein E ε4 With Medial Temporal Tau Independent of Amyloid-β, JAMA Neurol. 77, 470 (2020).

[181] Y. Chen et al., Genomic Atlas of the Plasma Metabolome Prioritizes Metabolites Implicated in Human Diseases, Nat. Genet. 55, 44 (2023).

[182] L. M. Reus et al., Quantitative Trait Loci Mapping of Circulating Metabolites in Cerebrospinal Fluid to Uncover Biological Mechanisms Involved in Brain-Related Phenotypes, https://doi.org/10.1101/2023.09.26.559021.

[183] S. Fang, M. V. Holmes, T. R. Gaunt, G. Davey Smith, and T. G. Richardson, Constructing an Atlas of Associations between Polygenic Scores from across the Human Phenome and Circulating Metabolic Biomarkers, Elife 11, (2022).

[184] K. E. Stepler, T. R. Gillyard, C. B. Reed, T. M. Avery, J. S. Davis, and R. A. S. Robinson, ABCA7, a Genetic Risk Factor Associated with Alzheimer's Disease Risk in African Americans, J. Alzheimers. Dis. 86, 5 (2022).

[185] C. Reitz et al., Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African Americans, JAMA 309, 1483 (2013).

[186] L. L. Barnes, Alzheimer Disease in African American Individuals: Increased Incidence or Not Enough Data?, Nat. Rev. Neurol. 18, 56 (2022).

[187] J. C. Morris et al., Assessment of Racial Disparities in Biomarkers for Alzheimer Disease, JAMA Neurol. 76, 264 (2019).

[188] M. S. Albert et al., The Diagnosis of Mild Cognitive Impairment due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease, Alzheimers. Dement. 7, 270 (2011).

[189] K. Rascovsky et al., Sensitivity of Revised Diagnostic Criteria for the Behavioural Variant of Frontotemporal Dementia, Brain 134, 2456 (2011).

[190] I. G. McKeith et al., Diagnosis and Management of Dementia with Lewy Bodies: Fourth Consensus Report of the DLB Consortium, Neurology 89, 88 (2017).

[191] B. M. Tijms, E. A. J. Willemse, M. D. Zwan, S. D. Mulder, P. J. Visser, B. N. M. van Berckel, W. M. van der Flier, P. Scheltens, and C. E. Teunissen, Unbiased Approach to Counteract Upward Drift in Cerebrospinal Fluid Amyloid-β 1-42 Analysis Results, Clin. Chem. 64, 576 (2018).

[192] S. Purcell et al., PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses, Am. J. Hum. Genet. 81, 559 (2007).

[193] T. Ge, C.-Y. Chen, Y. Ni, Y.-C. A. Feng, and J. W. Smoller, Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors, Nat. Commun. 10, 1776 (2019).

[194] N. Naithani, S. Sinha, P. Misra, B. Vasudevan, and R. Sahu, Precision Medicine: Concept and Tools, Armed Forces Med. J. India 77, 249 (2021).

[195] M. K. Breitenstein, H. Liu, K. N. Maxwell, J. Pathak, and R. Zhang, Electronic Health Record Phenotypes for Precision Medicine: Perspectives and Caveats From Treatment of Breast Cancer at a Single Institution, Clin. Transl. Sci. 11, 85 (2018).

[196] C. Delpierre and T. Lefèvre, Precision and Personalized Medicine: What Their Current Definition Says and Silences about the Model of Health They Promote. Implication for the Development of Personalized Health, Front Sociol 8, 1112159 (2023).

[197] J. Mai, M. Lu, Q. Gao, J. Zeng, and J. Xiao, Transcriptome-Wide Association Studies: Recent Advances in Methods, Applications and Available Databases, Commun Biol 6, 899 (2023).

[198] A. Dehghan et al., Metabolome-Wide Association Study on ABCA7 Indicates a Role of Ceramide Metabolism in Alzheimer's Disease, Proc. Natl. Acad. Sci. U. S. A. 119, e2206083119 (2022).

[199] L. Coppola, A. Cianflone, A. M. Grimaldi, M. Incoronato, P. Bevilacqua, F. Messina, S. Baselice, A. Soricelli, P. Mirabelli, and M. Salvatore, Biobanking in Health Care: Evolution and Future Directions, J. Transl. Med. 17, 172 (2019).

[200] S. D. Nagar, I. K. Jordan, and L. Mariño-Ramírez, The Landscape of Health Disparities in the UK Biobank, Database 2023, (2023).

[201] M. Schmidt, S. A. J. Schmidt, J. L. Sandegaard, V. Ehrenstein, L. Pedersen, and H. T. Sørensen, The Danish National Patient Registry: A Review of Content, Data Quality, and Research Potential, Clin. Epidemiol. 7, 449 (2015).

[202] K. A. S. Davis et al., Mental Health in UK Biobank - Development, Implementation and Results from an Online Questionnaire Completed by 157 366 Participants: A Reanalysis, BJPsych Open 6, e18 (2020).

[203] J. W. Smoller, The Use of Electronic Health Records for Psychiatric Phenotyping and Genomics, Am. J. Med. Genet. B Neuropsychiatr. Genet. 177, 601 (2018).

[204] S. A. Pendergrass and D. C. Crawford, Using Electronic Health Records To Generate Phenotypes For Research, Curr. Protoc. Hum. Genet. 100, e80 (2019).

[205] S. K. Service et al., Predicting Diagnostic Conversion from Major Depressive Disorder to Bipolar Disorder: An EHR Based Study from Colombia, medRxiv (2023).

[206] I. Pereira, M. J. Lopez-Martinez, and J. Samitier, Advances in Current in Vitro Models on Neurodegenerative Diseases, Front Bioeng Biotechnol 11, 1260397 (2023).

[207] G. J. Patti, O. Yanes, and G. Siuzdak, Innovation: Metabolomics: The Apogee of the Omics Trilogy, Nat. Rev. Mol. Cell Biol. 13, 263 (2012).

[208] J. C. Morris, C. M. Roe, C. Xiong, A. M. Fagan, A. M. Goate, D. M. Holtzman, and M. A. Mintun, APOE Predicts Amyloid-Beta but Not Tau Alzheimer Pathology in Cognitively Normal Aging, Ann. Neurol. 67, 122 (2010).

[209] K. R. Murphy, S. M. Landau, K. R. Choudhury, C. A. Hostage, K. S. Shpanskaya, H. I. Sair, J. R. Petrella, T. Z. Wong, P. M. Doraiswamy, and Alzheimer's Disease Neuroimaging Initiative, Mapping the Effects of ApoE4, Age and Cognitive Status on 18F-Florbetapir PET Measured Regional Cortical Patterns of Beta-Amyloid Density and Growth, Neuroimage 78, 474 (2013).

[210] J. Therriault et al., APOEε4 Potentiates the Relationship between Amyloid-β and Tau Pathologies, Mol. Psychiatry 26, 5977 (2021).

[211] S. Ye et al., Leading Determinants of Incident Dementia among Individuals with and without the Apolipoprotein E ε4 Genotype: A Retrospective Cohort Study, BMC Neurol. 24, 71 (2024).

[212] S. Kim and A. Misra, SNP Genotyping: Technologies and Biomedical Applications, Annu. Rev. Biomed. Eng. 9, 289 (2007).

[213] D. C. Soto, J. M. Uribe-Salazar, C. J. Shew, A. Sekar, S. P. McGinty, and M. Y. Dennis, Genomic Structural Variation: A Complex but Important Driver of Human Evolution, Am J Biol Anthropol 181 Suppl 76, 118 (2023).

[214] H. Wang et al., Structural Variation Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905 Alzheimer's Diseases Sequencing Project Subjects, medRxiv (2023).

[215] J. H. Sul et al., Contribution of Common and Rare Variants to Bipolar Disorder Susceptibility in Extended Pedigrees from Population Isolates, Transl. Psychiatry 10, 74 (2020).

[216] X. Jia et al., Investigating Rare Pathogenic/likely Pathogenic Exonic Variation in Bipolar Disorder, Mol. Psychiatry 26, 5239 (2021).