UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Cell-Type-Specific Epigenomic and Transcriptomic Alterations in the Human Brain in Aging and Neurodegenerative Disease

Permalink

https://escholarship.org/uc/item/69c4j0g8

Author

Chien, Jo-Fan

Publication Date

2024

Supplemental Material

https://escholarship.org/uc/item/69c4j0g8#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Cell-Type-Specific Epigenomic and Transcriptomic Alterations in the Human Brain in Aging and Neurodegenerative Disease

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

in

Physics

by

Jo-Fan Chien

Committee in charge:

Professor Eran A. Mukamel, Chair
Professor David Kleinfeld, Co-Chair
Professor Joseph R. Ecker
Professor Yi-Zhuang You

2024

Copyright

Jo-Fan Chien, 2024

All rights reserved.

The Dissertation of Jo-Fan Chien is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To my father, sister and Blueberry,

who have taught me that

every second counts, every effort matters,

in disease treatments.

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE..... iii

DEDICATION..... iv

TABLE OF CONTENTS..... v

LIST OF FIGURES.....vi

LIST OF SUPPLEMENTAL FILE.....vii

ACKNOWLEDGEMENTS.....viii

VITA..... ix

ABSTRACT OF THE DISSERTATION.....x

INTRODUCTION..... 1

Chapter 1 Cell-type-specific effects of age and sex on human cortical neurons..... 6

Chapter 2 Dynamics of cell-type-specific DNA (hydroxy)methyl-cytosine in human cortical neurons during postnatal development and aging.....48

Chapter 3 Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation..... 80

APPENDIX..... 152

REFERENCES..... 154

Note: The "Dissertation Approval Page" header is in white on page iii, so that the header does not appear on the page itself but appears in the Table of Contents. Do not delete this text.

LIST OF FIGURES

Figure 1.1 | Single nucleus multi-omic analysis of inter-individual transcriptomic-epigenomic diversity in human frontal cortex..... 10

Figure 1.2 | Aging effects on inhibitory and excitatory neuron cell types..... 15

Figure 1.3 | Correlated transcriptome and DNA methylome changes in the aged human brain. 19

Figure 1.4 | Increased expression of subtelomeric genes in aged neurons..... 22

Figure 1.5 | Sex differences in DNA methylation and gene expression in human neurons..... 25

Figure 2.1 | Mapping dynamic changes of genome-wide DNA (hydroxy)methylation and gene expression in two major neuronal subtypes across lifespan..... 52

Figure 2.2 | Correlated transcriptome and DNA methylome changes in the developing human brain..... 57

Figure 2.3 | Dynamics of DNA methylation..... 60

Figure 2.4 | Sex differences in DNA methylation and gene expression in human neurons..... 64

Figure 3.1 | A single cell transcriptomic analysis of human C9-ALS and C9-FTD motor and frontal cortex..... 83

Figure 3.2 | Dysregulation of gene expression in C9-ALS is concentrated in astrocytes and excitatory neurons..... 87

Figure 3.3 | Dysregulation of gene expression in C9-ALS astrocytes suggests activation and structural remodeling..... 91

Figure 3.4 | Excitatory neurons have altered expression of metabolic and protein regulatory pathway genes..... 97

Figure 3.5 | Epigenetic alterations correlate with transcriptome dysregulation in C9-ALS..... 103

Figure 3.6 | Lower proportion of high-quality neurons and alteration of non-neuronal transcription in C9-FTD..... 107

Figure 3.7 | Summary of the major results of our snRNA-seq analysis of postmortem brain samples from C9-ALS donors..... 113

Supplemental Figure 1.1 | Quality control (QC) metrics for snmCT-seq data, related to Figure 1.1..... 39

Supplemental Figure 1.2 | Comparison of RNA expression from snmCT-seq data (this study) with published snRNA-seq data, related to Figure 1.1..... 41

Supplemental Figure 1.3 | Correlation between samples, cell types and relation between DNA

methylation and RNA expression, related to Figure 1.1..... 42

Supplemental Figure 1.4 | Cell type proportions and global DNA methylation changes in aging, related to Figure 1.2..... 43

Supplemental Figure 1.5 | Functional enrichment of age-related transcriptome and DNA methylome changes, related to Figures 1.2, 1.3..... 44

Supplemental Figure 1.6 | Negative correlation between the change of mRNA expression and DNA methylation in aging, related to Figure 1.3..... 46

Supplemental Figure 1.7 | Sex differences in snmCT-seq data (this study) are consistent with published snRNA-seq and WGBS data, related to Figure 1.5..... 47

Supplemental Figure 2.1 | Correlation between samples and genome-wide DNA methylation in each age group, related to Figure 2.1..... 72

Supplemental Figure 2.2 | Consistency of age-DEGs with other published large scale datasets, related to Figure 2.2..... 73

Supplemental Figure 2.3 | DNA methylation features around age-DEGs, related to Figure 2.2. 74

Supplemental Figure 2.4 | DNA Methylation Patterns in Genomic Compartments, related to Figure 2.3..... 75

Supplemental Figure 2.5 | Dynamics of DNA methylation and strand asymmetry, related to Figure 2.3..... 76

Supplemental Figure 2.6 | Sex differences in DNA methylation are consistent across age, related to Figure 2.4..... 78

Supplemental Figure 2.7 | Browser view of DNA methylation levels in imprinting control region 15q11, related to Figure 2.4..... 79

Supplemental Figure 3.1 | Quality control (QC) metrics for snRNA-seq data, related to Figure 3.1..... 134

Supplemental Figure 3.2 | The clusters identified in snRNA-seq were not biased by known covariates and were largely stable, related to Figure 3.1..... 135

Supplemental Figure 3.3 | Annotations of the identified cellular populations in snRNA-seq based on the expression of known markers, related to Figure 3.1..... 136

Supplemental Figure 3.4 | Comparison of the cell types identified in our snRNA-seq dataset with a previously published dataset, related to Figure 3.1..... 137

Supplemental Figure 3.5 | Fluorescence-activated nuclei sorting (FANS) isolation of nuclei of major brain cell types, related to Figure 3.1..... 138

Supplemental Figure 3.6 | Quantification of proteins encoded by genes dysregulated in C9-ALS astrocytes using automated Western blot analysis, related to Figure 3.3..... 139

Supplemental Figure 3.7 | Quantification of proteins encoded by genes dysregulated in C9-ALS

excitatory neurons using automated Western blot analysis, related to Figure 3.3..... 140

Supplemental Figure 3.8 | Functional enrichment analysis of differentially expressed (DE) genes in upper- and deep-layer excitatory neurons, related to Figure 3.4.....142

Supplemental Figure 3.9 | QC metrics used in the snATAC-seq data pre-processing, related to Figure 3.5..... 144

Supplemental Figure 3.10 | The epigenetic landscape of C9-ALS brain cells determined by snATAC-seq, related to Figure 3.5.....145

Supplemental Figure 3.11 | ChromVAR scores in non-redundant transcription factor (TF) motif archetypes, related to Figure 3.5.....147

Supplemental Figure 3.12 | Comparison of the disease effect on gene expression in non-neuronal cells in C9-ALS and C9-FTD, related to Figure 3.6.....148

Supplemental Figure 3.13 | Comparison of C9-ALS vs. control gene expression fold-changes with AD vs. control gene expression fold-changes.....149

Supplemental Figure 3.14 | Principal component analysis of the pseudobulk snRNA-seq counts from each donor in each cell type.....150

Supplemental Figure 3.15 | Separations of disease and control samples in the snRNA-seq dataset.....151

LIST OF SUPPLEMENTAL FILES

Chien_SupplementalFiles.zip

- Chien_SupplementalTable1.1_Donor_metadata.xlsx
- Chien_SupplementalTable1.2_Cell_type_composition.xlsx
- Chien_SupplementalTable1.3_Age_differentially_expressed_genes.xlsx
- Chien_SupplementalTable1.4_Age_differentially_methylated_regions.xlsx
- Chien_SupplementalTable1.5_Sex_differentially_expressed_genes.xlsx
- Chien_SupplementalTable1.6_Sex_differentially_methylated_regions.xlsx
- Chien_SupplementalTable1.7_Cell_metadata.xlsx
- Chien_SupplementalTable2.1_Donor_metadata.xlsx
- Chien_SupplementalTable2.2_Age_differentially_expressed_genes.xlsx
- Chien_SupplementalTable2.3_Sex_differentially_expressed_genes.xlsx
- Chien_SupplementalTable3.1_Donor_metadata.xlsx
- Chien_SupplementalTable3.2_Cell_RNA_metadata.xlsx
- Chien_SupplementalTable3.3_ALS_motor_differentially_expressed_genes.xlsx
- Chien_SupplementalTable3.4_ALS_frontal_differentially_expressed_genes.xlsx
- Chien_SupplementalTable3.5_ALS_gene_ontology.xlsx
- Chien_SupplementalTable3.6_Cell_ATAC_metadata.xlsx
- Chien_SupplementalTable3.7_ALS_H3K27ac_differential_peaks.xlsx
- Chien_SupplementalTable3.8_FTD_motor_differentially_expressed_genes.xlsx
- Chien_SupplementalTable3.9_FTD_frontal_differentially_expressed_genes.xlsx
- Chien_SupplementalTable3.10_Western_blot_signal.xlsx
- Chien_SupplementalTable3.11_correlation_ATAC_H3K27ac.xlsx
- Chien_SupplementalTable3.12_correlation_RNA_ATAC.xlsx

ACKNOWLEDGEMENTS

The path on the Ph.D. road was not always easy, and I am truly grateful for all the people I met and those who have guided and accompanied me throughout my Ph.D. journey. You made this journey possible and fulfilling.

I want to thank my PI, Eran Mukamel, the best advisor ever. He guided me through the world of epigenetics and taught me hands-on coding in Python when I had no prior experience. His critical scientific thinking, coupled with his encouraging words, truly sculpted me into a scientist. His kindness supported me through my most difficult times. He taught me how to accept life's challenges, which shape us, and to always be kind. He is the role model who has guided me on the journey of science and life.

I also want to thank my lab member, Junhao Li, who is incredibly knowledgeable and always willing to teach me the biology and bioinformatics knowledge that I lack, as well as some helpful and fun tips. My gratitude extends to Heng Xu, with whom I spent a lot of time discussing methylation and scientific ideas, as well as weekend ideas. Additionally, I want to thank Fangming Xie, Ethan Armand, Dominic Burrows, Mike Cuoco, Alon Gelber, Abbey Green, and Rohini Gadde for sharing many scientific discussions and fun times together.

I want to express my gratitude to Professor Stella Dracheva, who provided invaluable guidance not only in scientific projects but also as a mentor, caring about my personal well-being. A special thanks to Professor Alexey Kozlenkov for always offering insightful ideas. I feel really lucky to be able to work with you on several projects despite the distance between San Diego and New York.

I am also deeply thankful to Professor Joe Ecker and Professor Margarita Behrens for their advice and comments on the single cell multiomic studies. Their expertise and insights significantly contributed to the depth and quality of my research. Additionally, I appreciate the

fruitful collaborations with Hanqing Liu, Bang-An Wang, and Anna Bartlett. Working with them has been a rewarding experience, and I am grateful for their stimulating discussions.

I would like to thank Professor David Kleinfeld and Professor Yi-Zhuang You for serving as my committee members. Your feedback has always been invaluable. I would also like to thank Professor Terry Hwa for leading the Q-bio and design classes where I truly learned a lot. A special thanks to Professor Doug Smith, who first introduced me to the DNA world in viral packaging. I am also grateful to Sharmila Poddar and Catherine McConney in the Physics department for their help, care, and support.

I am thankful to Professor C.K. Chan at Academia Sinica, who advised my master's thesis on neuron dynamics and encouraged me to pursue a Ph.D. degree. My long-time mentor and friend, Professor Hsiu-Hau Lin from National Tsing Hua University, has always provided genuine support and guidance through all aspects of my life.

I would like to thank my family and friends. Thanks to my mom and dad for supporting my decision to pursue a Ph.D. abroad. Their selfless caring has always supported me. I want to thank my best friend Weiting Lee for always staying by my side, even though we are 10,000 km apart. A special thanks to Olivia Ou and Song-Mao Liao, who took good care of me whenever I was overwhelmed. Your support has been invaluable. My gratitude extends to Huan-Yu Kuo, Claire Huang, and Yifan Xiang, with whom I have shared beers, tears, and many fancy foods. Also, thanks to Yi Guo, Xiang Ji, Da-Chuan Lu, and Zhendi Sun for the numerous nights we spent together in the study lounge, struggling with homework (and our ping-pong skills).

Finally, I want to thank my family here in San Diego: my husband, Wei-Ting Kuo, who has always been the strongest support throughout this journey. Without him, I wouldn’t have experienced the beauty of America and the depth of scientific studies here. His constant encouragement and belief in me has been my anchor. He showed me what true love, caring, and support are. Additionally, my adorable bunnies, Blueberry and Moo, who taught me how to love unconditionally. Their gentle presence brought joy and comfort to my life, especially during

the stressful times of my Ph.D. journey. Lastly, I thank myself for working hard and being persistent, which allowed me to reach heights I never thought possible.

PUBLICATIONS AND CONTRIBUTIONS

Chapter 1, in full, is a reprint of the material as it appears in *Neuron* 2024. Chien, J.-F., Liu, H., Wang, B.-A., Luo, C., Bartlett, A., Castanon, R., Johnson, N. D., Nery, J. R., Osteen, J., Li, J., Altshul, J., Kenworthy, M., Valadon, C., Liem, M., Claffey, N., O'Connor, C., Seeker, L. A., Ecker, J. R., Behrens, M. M. & Mukamel, E. A. Cell-type-specific effects of age and sex on human cortical neurons. *Neuron* (2024). doi:10.1016/j.neuron.2024.05.013. The dissertation author was the first author of this paper.

Chapter 2 is a work done in collaboration with Heng Xu, Alexey Kozlenkov, Ramu Vadukapuram, Junhao Li, Yu Wei, Chunyu Liu, Stella Dracheva and Eran Mukamel.

Chapter 3, in full, is a reprint of the material as it appears in *Nature Communications* 2023. Li, J., Jaiswal, M. K., Chien, J.-F., Kozlenkov, A., Jung, J., Zhou, P., Gardashli, M., Pregent, L. J., Engelberg-Cook, E., Dickson, D. W., Belzil, V. V., Mukamel, E. A. & Dracheva, S. Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation. *Nat. Commun.* **14**, 5714 (2023). The dissertation author was the researcher and author of this paper.

VITA

2013 Bachelor of Science in Physics, National Tsing Hua University

2017 Master of Science in Physics, National Taiwan University

2024 Doctor of Philosophy in Physics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Cell-Type-Specific Epigenomic and Transcriptomic Alterations in the Human Brain in Aging and Neurodegenerative Disease

by

Jo-Fan Chien

Doctor of Philosophy in Physics

University of California San Diego, 2024

Professor Eran M. Mukamel, Chair
Professor David Kleinfeld, Co-Chair

Epigenetic regulation, which modulates gene expression without altering the DNA sequence, is crucial for maintaining cellular function and facilitating responses to various influences such as aging, sex differences, and neurodegenerative diseases. In the brain, epigenetic modifications, including DNA methylation, histone modifications, and chromatin

accessibility, play pivotal roles in establishing and maintaining diverse cell type identities throughout the lifespan. Although DNA methylation typically occurs at CpG dinucleotides, it is also prevalent at non-CpG sites in neurons, underlying a genome-wide reconfiguration of the DNA methylation landscape in early childhood and adolescence. Traditional analysis of bulk tissues obscures cell-type-specific changes, whereas recent single-cell sequencing technologies provide more precise insights into specific excitatory and inhibitory neuron types as well as glia. However, significant gaps remain in our understanding of how human brain epigenetic modifications, such as DNA methylation (mC) and hydroxymethylation (hmC), regulate transcriptomes and influence brain function. This dissertation addresses these gaps through advanced cell-type-specific multi-omics studies of human cortical neurons. We used these technologies to investigate age-related changes, sex differences, and mechanisms underlying neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). The findings offer a comprehensive view of cell-type-specific epigenomic and transcriptomic alterations, providing critical insights for the development of targeted therapies.

Chapter 1 investigates the impact of aging and sex differences on human prefrontal cortical neurons through single-nucleus multi-omics RNA and DNA methylome sequencing. Key findings include synaptic gene downregulation, increased DNA methylation, age-related telomere shortening, and cell-type-specific X-inactivation escape genes.

Chapter 2 examines the dynamics of DNA methylation and hydroxymethylation in the human prefrontal cortex across the lifespan. The study reveals changes in mC and hmC levels, and distinct regulatory roles of mC and hmC in transcriptomic changes with aging.

Chapter 3 focuses on the molecular mechanisms of ALS and FTD associated with the C9orf72 mutation. Single-nucleus RNA and ATAC sequencing identify profound transcriptomic and epigenomic alterations in excitatory neurons and astrocytes in ALS, and in glial cells in FTD.

INTRODUCTION

The human brain is a highly complex organ that undergoes extensive changes throughout the lifespan, influenced by various factors such as development and aging[1–3], sex differences[4–6], and neurodegenerative diseases[7,8]. These changes manifest at multiple levels, including cellular composition[9–13], gene expression[9,14,15], and epigenomic regulation[16–18].

Epigenetics, derived from the Greek prefix “epi-” meaning “above”, refers to physical and covalent chemical modifications of chromatin that occur on top of the DNA sequence and influence gene expression. Conrad Waddington defined epigenetics as the study of the complex interactions between genes and their products that ultimately shape the phenotype[19].

Waddington's epigenetic landscape metaphor vividly illustrates his notion of cellular differentiation: imagine a ball rolling irreversibly down a landscape of hills and valleys. In this metaphor, the ball represents a developing cell, and the landscape's contours symbolize various epigenetic factors that create bifurcations guiding the cell's fate toward an equilibrium representing a mature, differentiated cell type[20,21]. The shape of this landscape is determined by epigenetic modifications such as DNA methylation, histone modifications, and the three-dimensional conformation of chromatin. These epigenetic marks direct the cell along specific developmental pathways, influencing its ultimate function[22,23]. However, it's important to note that Waddington's landscape is a metaphor rather than a physical model. Unlike a static equilibrium, cells are non-equilibrium systems that constantly consume energy and exhibit dynamic behaviors. This means that the cellular “fate” is not fixed but rather a dynamic process shaped by ongoing interactions with its environment.

Epigenetic modifications are dynamically altered in various biological contexts and collectively play pivotal roles in regulating the transcriptome. The groundbreaking work of Yamanaka et al. on induced pluripotent stem cells (iPSCs) demonstrated that the epigenetic landscape is not always irreversible[24]. By introducing a set of transcription factors, they were

able to reprogram differentiated cells back to a pluripotent state, highlighting the reversible nature of epigenetic modifications[25].

DNA methylation

DNA methylation involves the addition of a methyl group to the 5th carbon of the genomic cytosine, resulting in the formation of methylcytosine (mC). This modification is primarily catalyzed by DNA methyltransferases (DNMTs)[26]. DNMT3A and DNMT3B are responsible for de novo methylation, establishing new methylation marks on previously unmethylated DNA[27,28]. In contrast, DNMT1 is principally responsible for maintaining existing methylation patterns during DNA replication[29]. Due to the symmetrical nature of CpG sites on both DNA strands, methylation of CpG sites can be faithfully replicated during cell division, whereas non-CpG sites do not benefit from this symmetry[30]. This results in much higher methylation at CpG sites compared to non-CpG sites in most tissues. However, in neurons, which are post-mitotic and almost entirely produced during early prenatal gestation[31,32], DNA methylation accumulates at non-CpG sites over the course of the first two decades of postnatal life[33]. This non-CpG methylation has been proposed to play critical roles in neuronal function and plasticity[34–36], although the impact of losing mCH in cortical excitatory neurons is limited[37].

Cell-type-specific assay

Traditional studies have focused on bulk tissue analyses, which mask the heterogeneity of cell types and fail to provide insights into cell-type-specific molecular changes[38]. Recent advances in single-cell sequencing technologies have enabled more precise studies[39–44], but significant gaps remain in our understanding of how aging and sex differences affect specific neuronal subtypes in human brains at the epigenomic and transcriptomic levels.

Aging and Hydroxymethylcytosine

Studies have shown that DNA methylation patterns change with age, and the concept of the “methylation clock” has emerged as a biomarker of biological aging[45–47]. However, these studies often do not distinguish between DNA methylation (mC) and hydroxymethylation

(hmC)[48], which can have distinct regulatory roles. Hydroxymethylation, an oxidized form of methylation mediated by TET enzymes, is particularly prevalent in neurons and is implicated in active DNA demethylation and gene regulation[49–52]. The distinction between mC and hmC is crucial because while mC is generally associated with gene repression, hmC is often found to be linked to gene activation[49–52]. To accurately differentiate between mC and hmC, sequence-based techniques such as oxidative bisulfite sequencing (OxBS-seq)[53,54] and TET-assisted bisulfite sequencing (TAB-seq)[55] were developed. These methods selectively convert one form while leaving the other intact, allowing precise mapping of mC and hmC.

Sex differences

Sexual dimorphism in gene expression and epigenetic modifications can lead to different disease susceptibilities. Recent studies have begun to investigate sex-specific transcriptomic and epigenomic signatures to better understand the molecular underpinnings of sexual dimorphism in the human brain[4,5,15,56]. Notably, the non-CG methylation pattern of X-inactivation[16,33,35,57] provides a tool to predict tissue or cell-type-specific X-inactivation genes from methylation data without phasing RNA expression.

Human studies

Moreover, most studies in brain aging and disease come from animal models[3,9,58–61]. Although animal models provide valuable insights into fundamental biological processes, the human brain exhibits unique characteristics and complexities that may not be accurately replicated in animal models[62–64]. Human-specific studies are essential for translating findings into clinical applications and for understanding the precise molecular mechanisms underlying human brain function and disorders.

Dissertation overview

This dissertation addresses these gaps through three interrelated studies that leverage advanced FANS-sorted and single-cell (multi-omics) technologies to provide a comprehensive

view of cell-type-specific changes in the human brain. This dissertation is organized into three chapters:

Chapter 1 Cell-type-specific effects of age and sex on human cortical neurons focuses on the cell-type-specific effects of age and sex on human cortical neurons. Using single-nucleus multi-omics RNA and DNA methylome sequencing (snmCT-seq), we profile neurons from young and aged donors, revealing changes in cell type composition, gene expression, and DNA methylation in aging. We highlight the downregulation of synaptic genes in aged neurons accompanied by the increase of DNA methylation, and age-related telomere shortening. Additionally, we identify cell-type-specific sex differences in gene expression and X-inactivation escape genes. This novel large-scale human brain neuron single-nucleus multi-omics dataset provides valuable insights in cell-type-specific regulation between DNA methylation and transcription in aging.

Chapter 2 Dynamics of cell-type-specific DNA (hydroxy)methyl-cytosine in human cortical neurons during postnatal development and aging explores the dynamics of DNA methylation (mC) and hydroxymethylation (hmC) in the human prefrontal cortex across lifespan. Using whole-genome bisulfite sequencing (BS-seq) and oxidative bisulfite sequencing (OxBS-seq), we measure excitatory glutamatergic (Glu) and medial ganglionic eminence-derived inhibitory GABAergic (MGE-GABA) neurons from donors from 0 to 77 years old. Our findings reveal a genome-wide decrease of mCG and an increase of hmCG, with significant accumulation of non-CG methylation during development. We highlight the different regulatory roles of mC and hmC in transcriptomic changes in aging. We quantify methylation dynamics across the genome and detect regions that lost both mCG and hmCG enriched in enhancer regions. We also examine sex differences in methylation patterns and the regulation of inactivated and escape genes on chromosome X. These findings provide a comprehensive view of the epigenomic landscapes of the human prefrontal cortex, enhancing our

understanding of the dynamics of mC and hmC and their regulation on transcriptome across the lifespan.

Chapter 3 Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation focuses on neurodegenerative diseases, specifically amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) associated with the C9orf72 mutation. Through single-nucleus RNA and ATAC sequencing, we analyze the motor and frontal cortices from postmortem human brains, identifying distinct transcriptomic and epigenomic alterations. Our results show that upper and deep layers of excitatory neurons and astrocytes are profoundly affected in ALS, with upregulation of mitochondrial and proteostasis genes. In contrast, FTD is characterized by extensive transcriptional dysregulation in glial cells. These findings underscore the cell-type-specific mechanisms underlying ALS and FTD and suggest potential therapeutic targets.

Overall, this dissertation advances our understanding of the molecular mechanisms underlying brain aging and neurodegeneration. By leveraging FANS sorting and single-nucleus technologies, we provide a comprehensive view of cell-type-specific changes in the human brain, offering valuable insights for the development of targeted therapies for neurodegenerative diseases.

Chapter 1 Cell-type-specific effects of age and sex on human cortical neurons

1.1 Introduction

Aging is a multifaceted biological process that affects the brain through changes in cell type composition[65], gene expression[9], telomere attrition[66], and multiple levels of epigenomic regulation[47,58,67]. Though many molecular changes have been associated with aging, the impact of age-related changes in the epigenome and transcriptome on specific neuron types remains unclear. Studies of genetic changes such as telomere shortening have focused on dividing cells, without addressing the potential role of telomeres in aging for post-mitotic neurons[68,69]. On the other hand, studies of the brain transcriptome using bulk RNA sequencing in human cohorts[70–73] have implicated age-related downregulation of genes involved in synaptic function and plasticity, potentially contributing to deficits in synaptic plasticity and memory consolidation[71,74,75]. Despite this progress using transcriptomics in bulk tissue samples, we lack cell-type-specific information about the coordinated changes across the transcriptomes and epigenomes of different neuron types.

The gaps in our understanding of cell-type-specific age changes are partly filled by recent single cell or single nucleus RNA sequencing (sc/snRNA-seq) and spatial transcriptomics studies in the fly[59,76], mouse[9,77,78], non-human primate[79,80] and human brain[81–84]. Glial cells, especially oligodendrocytes and their precursors, appear to be particularly affected by advanced age in mouse[9] and human brain[85]. Moreover, recent epigenome studies using snATAC-seq have uncovered associations between cell-type-specific changes in chromatin accessibility and heterochromatin marked by modified histone H3K9me3[58]. Epigenome erosion has also been observed in human neurons in Alzheimer's disease[86]. Studies of chromatin organization have further pointed to a potential role of DNA double strand breaks in disrupting the 3D genome in neurodegenerative disease[87]. Finally, recent studies have highlighted the crucial role of DNA

methylation in aging across multiple tissues[45,88], although the effect of advanced age on the unique DNA methylome of brain cells[33] remains unclear.

Human brain cell types establish a specific DNA methylation landscape over an extended period spanning pre- and postnatal brain development, and alterations to the methylome continue throughout the lifespan[33,89,90]. Array-based studies have shown that methylation at multiple CG sites can reliably predict biological age across multiple species and tissues, including the human brain[46,47,91–93]. Methylation clocks suggest a conserved program of changes in epigenetic regulation in healthy subjects, which may be accelerated by diseases such as cancer or HIV[94–97]. Whole-genome bisulfite sequencing has further revealed a dynamic trajectory of age-related DNA methylation changes in human neurons[33,89,90]. In particular, neurons accumulate considerable methylation at non-CG sites, mainly CA dinucleotides, in a gradual process extending throughout postnatal development and childhood[33,98]. However, the specific changes in DNA methylation between young and aged adults in neuronal and glial cell types have not been characterized with single cell resolution in the human brain.

Here, we use single-nucleus DNA methylome and transcriptome sequencing (snmCT-seq)[99] to profile prefrontal cortex neurons from young adults (23-30 years old) and aged (70-74 yo) male and female donors. Our dataset, with >55,000 high quality single cell profiles, offers a comprehensive view of stable and age- or sex-variable cell regulation in the human brain. We found changes in cell type composition, gene expression, and DNA methylation across brain cell types. Notably, genes implicated in synaptic function were downregulated with advancing age, in parallel with increased DNA methylation within their gene bodies and at nearby candidate cis-regulatory regions. We also used the DNA methylome data to estimate telomere length in individual brain cells, and uncovered an age-related telomere position effect whereby genes located near chromosome ends are broadly upregulated in aged neurons. Finally, we found sex differences in both expression and DNA methylation across autosomes

and chrX, including signatures of genes that escape from X-inactivation in specific neuron types in females.

1.2 Results

1.2.1 A multi-omic atlas of human frontal cortex neurons across age and sex

We measured the transcriptome and DNA methylome[99] in frontal cortex neurons of young (23-30 years old) and aged (70-74 years old) male and female donors (n=11, Fig. 1.1A). We focus on the human dorsolateral prefrontal cortex (Brodmann area 46, BA46; hereafter "frontal cortex"), a critical region involved in cognitive control and executive function which is affected in aging and age-related neurodegenerative disease[9,73,89,100,101]. We obtained 55,447 high-quality nuclei with an average of >1.37 million DNA reads, which were enriched for neurons (NeuN+). Most of these cells (39,830) also passed stringent quality control criteria for RNA-seq data, providing on average 213,000 RNA reads, and 6,800 genes detected per cell (Fig. S1.1, Table S1.1).

To assess donor variability, we annotated cells according to major neuron types rather than fine-grained subtypes. We identified 11 glutamatergic excitatory and 10 GABAergic inhibitory neuron types based on 100 kb bin non-CG methylation (mCH) features[42], which we labeled based on expression of canonical mRNA cell type markers[62] (Fig. 1.1B,C). We also obtained a smaller number of glial cells, including astrocytes, oligodendrocytes and microglia, from some samples. Cell type clusters were similar when identified using mCH, mCG or mRNA features (Fig. S1.1C,D).

The pseudobulk gene expression profiles from each cell type were highly consistent with reference cell types from scRNA-seq of multiple human cortical regions[62] (mean Spearman correlation = 0.84 ± 0.05 for 500 highly variable genes between cell types, Fig. S1.2). As expected, cell type marker gene expression correlated with low gene body DNA methylation in both CG and CH contexts (Fig. 1.1C, D, F, Fig. S1.3B)[99].

RNA expression and DNA methylation features were highly consistent between independent tissue samples from the same donor. For example, samples of L2-4 intra-telencephalic (IT) neurons from the same donor had Pearson correlation r = 0.98 ± 0.01 for RNA, 0.99 ± 0.01 for mCG in gene bodies, and 0.98 ± 0.01 for mCH (Fig. 1.1E and Fig. S1.3C). By contrast, the correlation of these signatures in different individuals is lower, especially between donors with different age or sex (r = 0.90 ± 0.03 for RNA, 0.97 ± 0.01 for gene body mCG, 0.91 ± 0.03 for gene body mCH, Fig. 1.1E and Fig. S1.3C). The correlation between signatures for different cell types was far lower (r = 0.73 ± 0.08 for RNA, 0.80 ± 0.07 for mCG in gene bodies, and 0.48 ± 0.15 for mCH, Fig. S1.3D).

These multi-omic data are displayed in a genome browser that compares base resolution DNA methylation with gene expression for each demographic group (brainome.ucsd.edu/HumanBrainAging). For example, the gene *ADARB2* is highly expressed across CGE-derived inhibitory neuron types (VIP, LAMP5) and shows corresponding patterns of low mCG and mCH across the gene body, while *NXPH4* labels a group of deep layer excitatory neurons (L6b; Fig. 1.1F). Stratifying the genomic data by donor age and sex shows the high reproducibility of DNA methylation and RNA expression tracks for young, aged, male and female donors (Fig. 1.1G). These integrated multi-omic, multi-donor data allowed us to investigate age- and sex-related variability in brain cell transcriptomes and epigenomes.

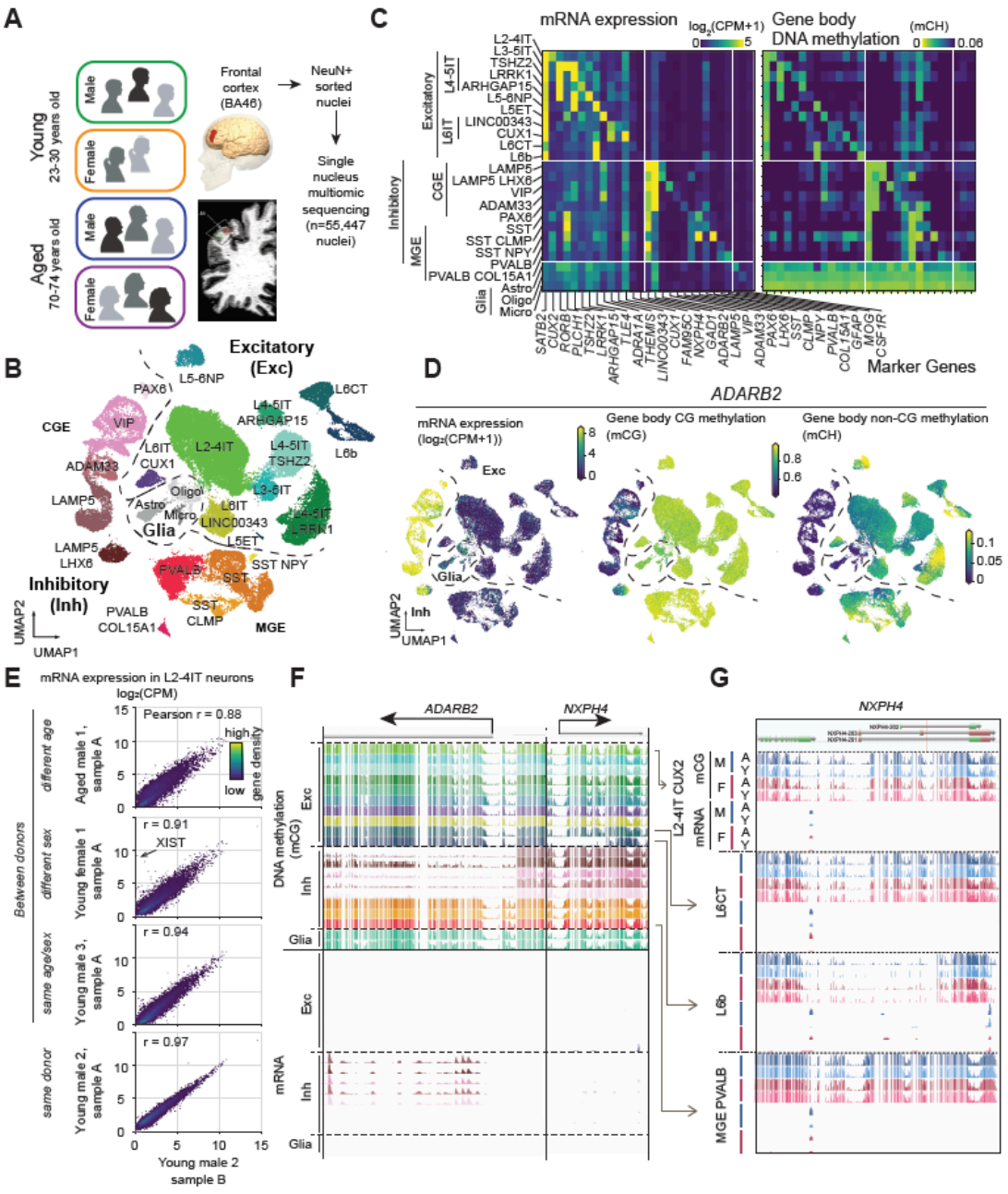


Figure 1.1 | Single nucleus multi-omic analysis of inter-individual transcriptomic-epigenomic diversity in human frontal cortex.

(A) Study design. (B) UMAP embedding of nuclei assigned to 11 excitatory, 10 inhibitory and 3 glial types using DNA methylation in 100 kb bins. (C) Cell type marker RNA expression and gene body mCH. (D) Correspondence of mRNA, mCG and mCH for the CGE interneuron marker, *ADARB2*. (E) Correlation of mRNA expression between independent samples from the same donor, between different donors with the same age/sex and different age or sex. (F) Pseudobulk expression (mRNA) and DNA methylation (mCG) at *ADARB2* and *NXPH4* (L6b excitatory neurons), averaged across subjects. (G) Separate tracks for young, aged, male and female groups showing consistent epigenomic and transcriptomic cell type signatures. MGE/CGE, medial/caudal ganglionic eminence; IT, intra-telencephalic; CT, cortico-thalamic; NP, near-projecting; ET, extra-telencephalic; CPM, counts per million.

1.2.2 SST and VIP expressing GABAergic neurons are reduced in aged frontal cortex

Cortical glutamatergic and GABAergic neurons are produced before birth, and the balance of their excitatory and inhibitory activity throughout the lifespan is critical for healthy cognition and brain function. Although new neurons are not produced in the adult cortex, whether the ratio of excitatory to inhibitory neurons changes with age, e.g. due to selective neuron loss, remains unclear. We found a 68% lower proportion of inhibitory neurons in aged samples (mean 0.28 ± 0.06 s.d.) compared with young samples (0.41 ± 0.06, p=0.01, Wilcoxon rank-sum test, Fig. 1.2A). These differences were not explained by sample quality, as we found no difference in the fraction of high-quality cells or in DNA methylation quality-control metrics in young vs. aged donors (Fig. S1.1).

The age-associated decline of inhibitory neurons was driven by substantial loss of SST- (2.90-fold, false discovery rate (FDR) < 0.1) and VIP-expressing cells (1.71-fold) (Fig. 1.2B; Fig. S1.4A). There was also an increase in the proportion of CGE-derived LAMP5 LHX6 neurons (1.67-fold), and a type of L4-5 IT excitatory neurons (1.50-fold). Notably, the loss of SST or VIP neurons has been observed in the aging human brain using single cell data and deconvolution of bulk brain data from 8 datasets across 6 cohorts, with over 1,100 individuals[65]. Inhibitory neurons were also reported to reduce across multiple brain regions in Alzheimer's disease patients compared with age-matched controls, primarily over 70 years old[12,81,102,103].

1.2.3 Widespread age-related changes in cell-type-specific gene expression and DNA methylation

The deep epigenomic and transcriptomic profiles of our snmCT-seq data allowed us to further examine age-related changes in gene regulation in specific neuron types. The low-dimensional embedding of cells based on CG and CH DNA methylation throughout the genome suggested substantial differences between excitatory neurons from young and aged donors (Fig. 1.2C). By contrast, inhibitory cells from all donors had very similar global

methylation profiles and their UMAP embeddings were largely overlapping. We did not analyze age-related changes in glial cell types because the number of glial cells was not sufficiently balanced across donors.

The genome-wide average level of CG DNA methylation varies across neuron types[104,105], ranging from a low level of methylation in L2-4 IT neurons (mCG=0.82) to the highest levels in SST inhibitory cells (mCG=0.86) (Fig. S1.4B). CH methylation, particularly at CA dinucleotides, varied by more than 2-fold across neuron types, with mCA ranging from 0.07 to 0.15 (Fig. S1.4C, D). In contrast with these large cell type differences, genome-wide average methylation for each cell type was largely stable in aging, with slight increases in mCG in some excitatory neurons (Fig. S1.4B, C, D). This shows that any age-related global changes in mCG and mCH are small relative to the cell-type-specific patterns of DNA methylation that are established in early life and largely persist through the adult lifespan.

To statistically identify robust age differences in gene expression and DNA methylation, we used conservative analyses based on the pseudobulk profile of each cell type per donor (see Methods)[106] with rigorous adjustments for multiple comparisons[107]. We found 1,904 unique differentially expressed genes (age-DEGs) in young vs. aged neurons (FDR <0.05 and fold-change >1.2 in at least one cell type, Fig. 1.2D, Table S1.3). Overall, we found more age-DEGs in excitatory neurons (1,536 unique genes) compared with inhibitory neurons (601). Downsampling analysis confirmed that this pattern was not due to the differential abundance of neuron types (Fig. 1.2E, see below). In particular, there were relatively few age-DEGs in SST- (7 age-DEGs) or VIP-expression (109) inhibitory cells. This indicates that, despite their reduced population in aged donor brains, the surviving SST and VIP cells maintain a similar gene expression profile compared with young adults.

Despite the limited number of donors in our study, the aging changes were highly consistent with mRNA measurements from bulk brain tissue in a large cohort of up to n=255 donors from the gene tissue expression project (GTEx) (r = 0.63, Fig. 1.2F; Fig. S1.4E)[70].

Whereas the GTEx data identified just 28 DEGs from the frontal cortex (Brodmann area 9), our single nucleus data identified more alterations to brain gene expression in specific neuron types.

Changes in gene expression during aging could be driven by alterations to epigenetic modifications, including DNA methylation. Thousands of cytosines throughout the genome gain or lose DNA methylation during childhood brain development[33,89], and the DNA methylome provides a clock-like signature of aging throughout the lifespan in multiple tissues and species[46,108]. However, previous analyses of age-related changes in DNA methylation lacked cell-type-specific resolution, and did not assess the methylation status of all CG and CH sites genome-wide. Focusing on mCG, we called differentially methylated regions (DMRs) between aged and young donors in each neuron type, controlling for sex[109,110]. We found age-DMRs in all cell types, with more regions gaining than losing methylation with age (Fig. 1.2D, Table S1.4). For example, we found 19,715 age-DMRs in L4-5 IT TSHZ2 neurons (covering 8.5 Mbp, ~0.3% of the genome), and 98% of these DMRs increase methylation in aged samples (hypermethylated age-DMRs).

To directly compare age-related changes in expression across cell types without bias due to differences in cell abundance, we repeatedly subsampled 50 cells from each of the major neuron types for each sample followed by DEG and DMR analysis (Fig. 1.2E). We also applied Augur to prioritize cell types that have the most age effect[111] (Fig. S1.4F). Although the number of significant age-DEGs and age-DMRs was greatest in excitatory IT neurons, the effect of aging was similar across neuron types for most age-DEGs even in cell types in which the effect did not reach statistical significance (Fig. 1.2G). Among the 1,904 age-DEGs, only two protein-coding genes had opposite direction of age-related change between excitatory and inhibitory neurons (*ATP2B4*: 3.3-fold increase in L4-5IT TSHZ2 and 2.5-fold decrease in MGE PVALB; *INPP4B*: 3.8-fold increase in L6CT and 2.2-fold decrease in CGE ADAM33). By contrast, age-DMRs are more cell-type-specific and more prominent in excitatory neurons, especially in L2-4 IT and L4-5 IT TSHZ2 (Fig. 1.2G).

1.2.4 Reduced expression and increased methylation of synaptic genes in the aging brain

Genes downregulated with age are enriched in pathways related to synaptic signaling and cell adhesion (Fig. 1.2H). These included neuroligin 1, *NLGN1*, which mediates the formation and maintenance of excitatory synapses[112]; phosphatase and actin regulator 1, *PHACTR1*, encoding a synaptic protein regulating excitatory synapses[113]; and the voltage-gated potassium channel *KCNH7*[114] (Fig. S1.5D). Downregulation of genes involved in synaptic signaling in aged brains has been reported in rodents[115,116], non-human primates[80,117] and humans[71,74,83]. The downregulation of synaptic and cell adhesion genes may be associated with reduced synaptic plasticity[74,118]. Some of the upregulated genes may have neuroprotective functions, such as the activation of histone H4 acetylation (Fig. S1.5C) that has been linked to DNA double-strand break repair[119,120].

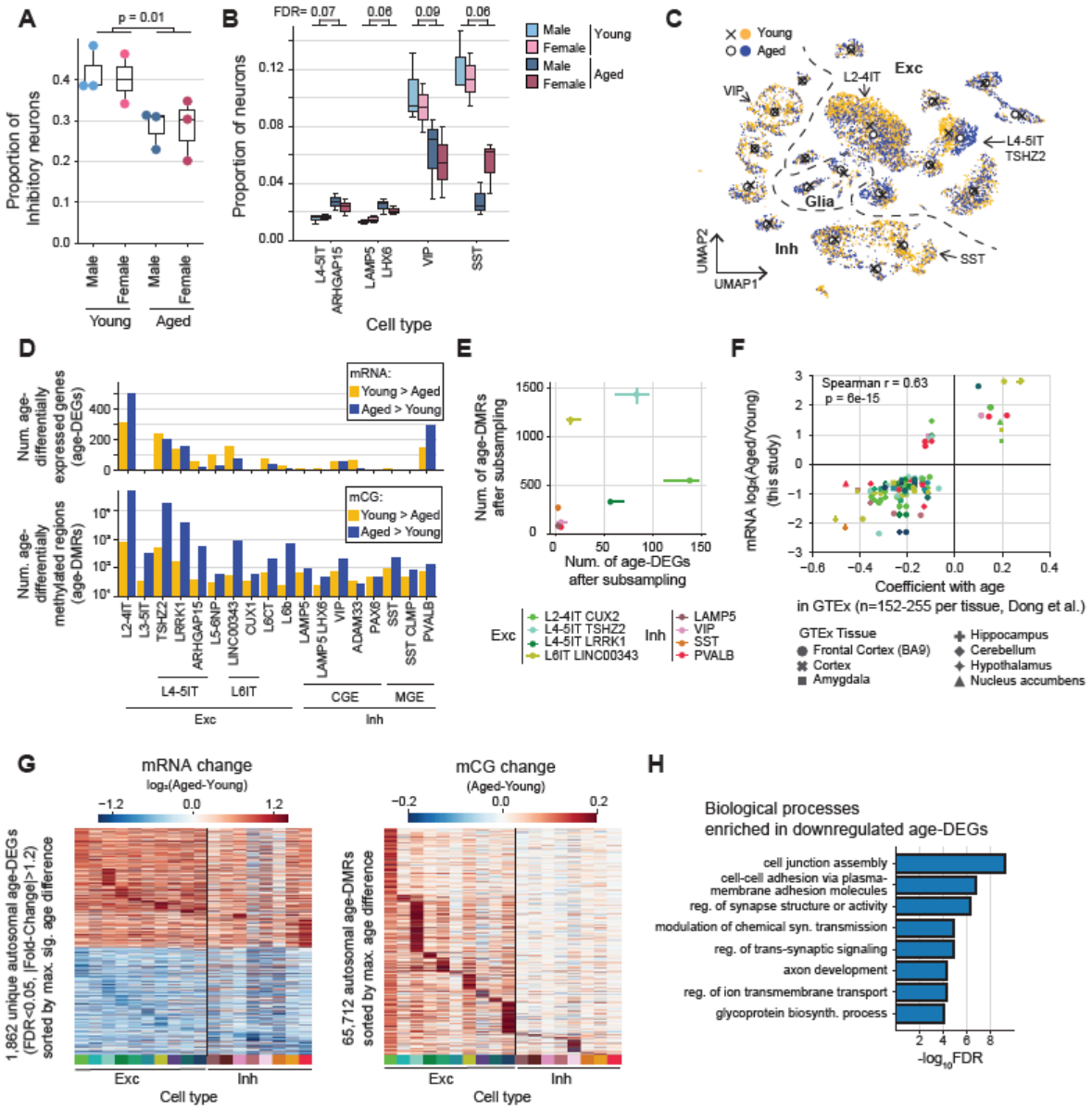


Figure 1.2 | Aging effects on inhibitory and excitatory neuron cell types.
(A) Proportion of all GABAergic inhibitory neurons among all neurons in young and aged donors (Wilcoxon rank-sum test; n=5 young, n=6 aged donors). (B) Proportion of neuron types whose abundance was significantly different between young and aged donors (Wilcoxon rank-sum test, FDR<0.1; n=5 for young donors, n=6 for aged donors). (C) UMAP embedding showing centroids of cells of each type from young (x) and aged (circles) donors. (D) Number of age-differentially expressed genes (age-DEGs) and differentially methylated regions (age-DMRs) in each cell type. (E) Number of age-DEGs and age-DMRs resulting from randomly sampling 50 nuclei per cell type per sample. This subsampling procedure is repeated five times using distinct random seeds. The plotted data points represent the median values, while the error bars indicate the interquartile range (25th to 75th percentile) across the random subsampling. (F) Correlation of age effect on gene expression in our data with GTEx data from bulk brain tissue[70]. (G) Left: Pseudobulk RNA fold change, combined expression alterations from all cells in the same cell type, of the union of age-DEGs from all cell types. Right: Pseudobulk DNA methylation change, the combined alterations in DNA methylation levels from all cells in the same cell type, of the union age-DMRs across all cell types. (H) Biological pathways enriched for age-DEGs that are downregulated in aged neurons (FDR <1e-3).

1.2.5 Increased DNA methylation in aged neurons correlates with reduced gene expression

DNA methylation differences between cell types correlate with cell-type-specific gene expression[104,121], but whether age-related changes to the epigenome correlate with transcriptomic age changes has not been established. To address this, we calculated the Spearman correlation of the fold-change (aged/young) for RNA and DNA methylation features across the union of age-DE genes, i.e. genes which were age-DE in at least one cell type (Fig. 1.3A, Fig. S1.6). Gene body mCH and mCG had a strong negative association with mRNA expression in most neuron types (FDR < 0.01), e.g. r = -0.49 for excitatory L4-5 IT TSHZ2 neurons (Fig. 1.3A, B). This correlation was significant even for cell types in which few genes reached statistical significance (Fig. S1.6B,C). Gene body mCG also correlated with lower expression (Fig. 1.3A), but mCG and mCH at the promoter region (transcription start site ± 1kb) were less strongly correlated (Fig. S1.6B-D). Correlating RNA with DNA methylation features across all expressed genes showed a similar pattern of negative association (Fig. S1.6D). We conclude that changes in neuronal DNA methylation coincide with age-related expression changes in the adult brain, consistent with a potential regulatory role.

To illustrate the age-related changes in gene expression and DNA methylation, we highlight the rho GTPase activating protein 10, *ARHGAP10* (Fig. 1.3C). mRNA for *ARHGAP10* is highly expressed in upper layer excitatory neurons (L2-4 IT) in young donors (>435 transcripts per million mapped reads (TPM)), and subsequently reduced 3.2-fold in older donors. Consistent with this downregulation, we found several DMRs around the *ARHGAP10* locus where DNA methylation increases with age in L2-4 IT neurons. DMRs occurred in the upstream intergenic region, marking putative cis-regulatory elements, as well as within the promoter and gene body. Gene body mC also had a strong negative correlation with *ARHGAP10* mRNA expression across donors (r = –0.68 for mCG, –0.82 for mCH, Fig. 1.3F). Similarly, the highly

expressed potassium voltage-gated channel subfamily H member 7, *KCNH7*, was reduced 1.9-fold in aged donors and showed a corresponding increase of CG methylation (Fig. 1.3D, G).

In total, age-DMRs covered 32.4 Mbp, and occurred mainly in introns (69%) and intergenic regions (25%). A small minority of age-DMRs covered coding regions (exons) (6%). Age-DMRs had a non-random distribution across functional genomic compartments defined by multiple histone modification ChIP-seq datasets from human prefrontal cortex using chromHMM[122] (Fig. S1.5F). DMRs that increase methylation in aged neurons were enriched in enhancers, and in states associated with Polycomb repressive complex 2 (PRC2) (FDR <0.01). DMRs that lose methylation in aged neurons were only enriched in active states.

To explore the potential regulatory significance of DNA methylation changes, we used sequence motif enrichment analysis to identify transcription factors that may bind at age-DMRs[123]. Motifs of early growth response (EGR) family factors, including WT1 and EGR1, were enriched in excitatory neuron age-DMRs that increase methylation with age (FDR <0.001, Fig. 1.3I). Up to 12.5% of hypermethylated (aged > young) age-DMRs overlapped EGR1 motifs (Fig. S1.5G). *EGR1* mRNA expression was also strongly reduced with age (4-30 fold) in both excitatory and inhibitory neurons (Fig. 1.3J), although there was no significant difference in gene body methylation. *EGR1* is an immediate early gene encoding a zinc finger transcription factor that is important for long-term memory consolidation[124,125]. Our findings in the human frontal cortex extend previous observations of reduced *EGR1* expression and increased promoter mCG in the hippocampus of aged mice[126]. By contrast, regions that lost mCG in aged neurons were enriched in transcription factor binding motifs of the AP-1 family of immediate early genes, FOS and JUN, and the transcription elongation factor GRE (Fig. S1.5H).

DNA methylation plays a fundamental role in development, establishing unique sets of hypomethylated regions that mark each mature cell type[42,105]. These cell type DMRs have large differences in mCG across adult neuron types. We hypothesized that age-related changes in DNA methylation may specifically affect these developmentally programmed sites of cell type

regulation. Consistent with this, we found that regions with the greatest difference in mCG or mCH between aged and young cells in any cell type also had large differences in mean methylation between cell types in young adults (Fig. 1.3K).

1.2.6 Low methylated regions gain mCG with age

Studies of the DNA methylation clock have identified systematic changes in mCG throughout the lifespan, particularly at sites with low mCG[127]. Consistent with this, we found increased mCG with age in DNA methylation valleys (DMVs), defined as extended regions (≥5 kb) with low average mCG (<15%)[128] (Fig. 1.3L). DNA methylation also increased at regions associated with Polycomb repressive complex 2 (PRC2) in prefrontal cortex[122], although the expression levels of the four core PRC2 components (*EZH2*, *SUZ12, EED*, and *RBBP4*) were not significantly different between aged and young donors. By contrast, there was no effect of age on mCG at heterochromatin regions marked by the repressive histone modification H3K9me3 (Fig. 1.3L). The lack of increased mCG in heterochromatin regions is notable given the observation of increased chromatin accessibility and reduced H3K9me3 at these sites in excitatory neurons in aged mice[58].

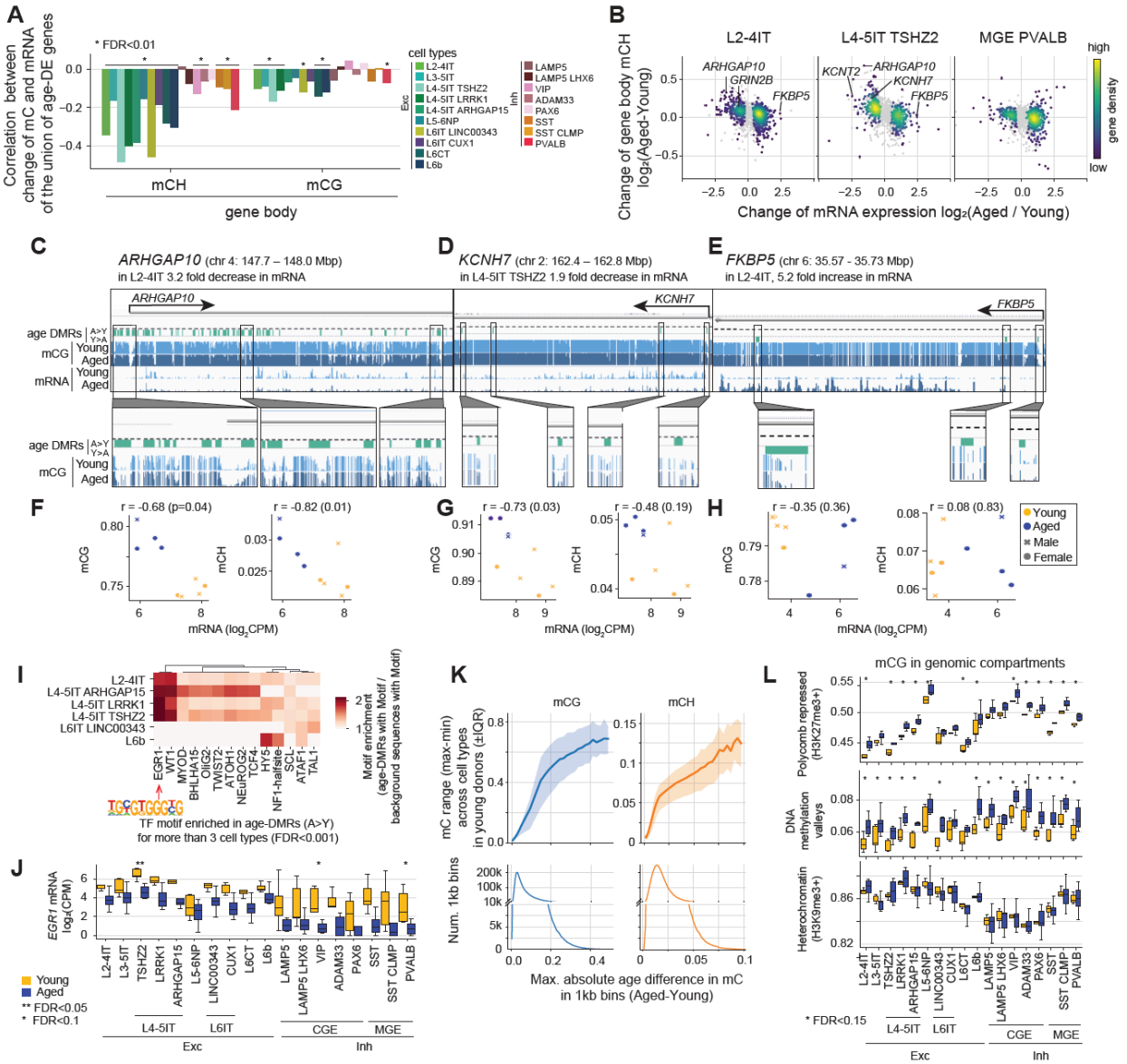


Figure 1.3 | Correlated transcriptome and DNA methylome changes in the aged human brain.
(A) Spearman correlation coefficient between the age change of gene body DNA methylation (aged - young) and the fold-change in gene expression in the union of age-DEGs from all cell types. * FDR <0.01. (B) Example scatter plots showing age-DEGs (colored dots) correlated with the age change in gene body mCH. Gray dots show the union of age-DEGs from all cell types. (C,D) Browser view of the ARHGAP10 (C) and KCNH7 (D) loci, showing reduced gene expression and increased mCG at age-DMRs throughout the gene body. (E) Browser view of the FKBP5 loci, showing reduced mCG at age-DMRs and increased gene expression. (F,G,H) Scatter plots of mRNA expression and CG, CH methylation for each donor in ARHGAP10 (F), KCNH7 (G) and FKBP5 (H). (I) Motif enrichment of hypermethylated age-DMRs. Only motifs significantly enriched (FDR <0.001) in more than two cell types are shown. (J) Expression of EGR1 mRNA. (K) The relation between maximum methylation difference in age and the methylation difference between cell types for all 1kb bins. Bottom row is the histogram of the bins with age methylation difference. (L) mCG age difference in genomic compartments. Polycomb and heterochromatin regions are called by chromHMM whole brain tissue annotation. DNA methylation valleys are called by methylSeekR from our data.

1.2.7 Increased expression of subtelomeric genes in aged brain cells

In contrast with downregulation of synaptic genes, we found other genes with strongly increased expression in aged neurons. *FKBP5*, encoding a co-chaperone of the glucocorticoid receptor complex, increased expression 5.2-fold in aged neurons and harbored hypomethylated (aged < young) age-DMRs in its gene body (Fig. 1.3E). Aging and stress have been associated with increased *FKBP5* expression and decreased methylation at two CG sites in blood in multiple large human cohorts (n >3000), and this was mechanistically linked to NF-k-B driven inflammation[129,130]. Another example of an age-increased gene in L2-4 IT neurons was the mu opioid receptor, *OPRM1* (FDR <0.05, >2-fold increase). Altered opioid receptor signaling has been reported in aged rodents[131].

Despite these examples, we did not identify a clear functional association for age-increased genes as a group, in contrast with the significant enrichment of synaptic function for age-decreased genes (Fig. 1.2H). Moreover, whereas age-decreased genes fit an expected pattern with increased methylation and enriched hypermethylated age-DMRs in their gene body, age-increased genes were not enriched in gene body mCG (Fig. 1.4A, B). Instead, genomic regions that surround age-increased genes had less age-related accumulation of mCH compared with regions around age-decreased genes (Fig. 1.4A). mCG hypo-DMRs (aged < young) were slightly enriched at the promoters of age-increased genes (Fig. 1.4B).

Given the reduced mCH around age-increased genes, we reasoned that these genes may be clustered in genomic regions with a distinct chromatin organization. We found that age-increased genes, but not age-decreased genes, were enriched ~1.5-fold within the subtelomeric regions, i.e. ≤5 Mb from the ends of chromosomes (Fig. 1.4C). We found a similar pattern of enrichment for age-DMRs, including both hypo- and hyper-mCG age-DMRs (Fig. 1.4D). The subtelomeric regions have lower average mCG (~5% lower than the genome average) and mCH (~20% lower, Fig. 1.4E). Subtelomeres also have higher gene density and higher GC content than the rest of the genome (Fig. 1.4F).

Telomere attrition and telomere DNA damage response in aging have been reported in both dividing and post-mitotic cells[132]. We reasoned that the enrichment of age-increasing DEGs in subtelomere regions might be related to the disruption of the chromatin structure at chromosome ends due to telomere damage or shortening. We applied TelomereHunter[133] to estimate telomere length for each donor, using information from snmCT-seq reads that mapped to unique genomic locations as well as unmappable reads with repetitive sequences. Estimated telomere length was significantly shorter in L2-4 IT neurons from aged compared to young donors (Fig. 1.4I, $p < 0.05$, Wilcoxon rank-sum test). Furthermore, shorter telomere length correlated with increased expression of age-DEGs located in the subtelomere regions, consistent with a telomere position effect[134] (Fig. 1.4J).

We found age-DMRs enriched at the ends of many, but not all, chromosomes (Fig. 1.4K). The significant positive correlation between GC content and age-DMRs (Fig. S1.5I) suggests that GC content could contribute to the observed pattern across chromosomes. By contrast, there was no enrichment at centromeres. The enrichment of DMRs was not unique to age-related mC, but was also observed for sex-DMRs and cell type-DMRs (Fig. 1.4G). This enrichment was also not unique to a particular cell type but was observed across multiple excitatory and inhibitory neuron types (Fig. 1.4H, FDR <0.05).

These findings are notable given the observation of age-increased expression of genes in subtelomeric regions in multiple tissues across a large cohort (GTEx)[70]. In that study, six tissues had significant enrichment of age-increased genes in subtelomeres, including whole blood, skeletal muscle, thyroid, and sigmoid colon, all of which have been shown to progressively lose telomere length with advanced age. The GTEx data showed no enrichment of age-increased genes in the subtelomeres for any of the 14 brain tissues tested, including frontal cortex[70]. Despite the lower number of donors in our study, our single-cell data identified a larger number of age-DEGs and recapitulated the enrichment of age-increased DEGs observed in other tissues. These data indicate that post-mitotic neurons experience age-related

alterations of gene expression, including cell-type-specific age-increased expression, in the subtelomeric compartment.

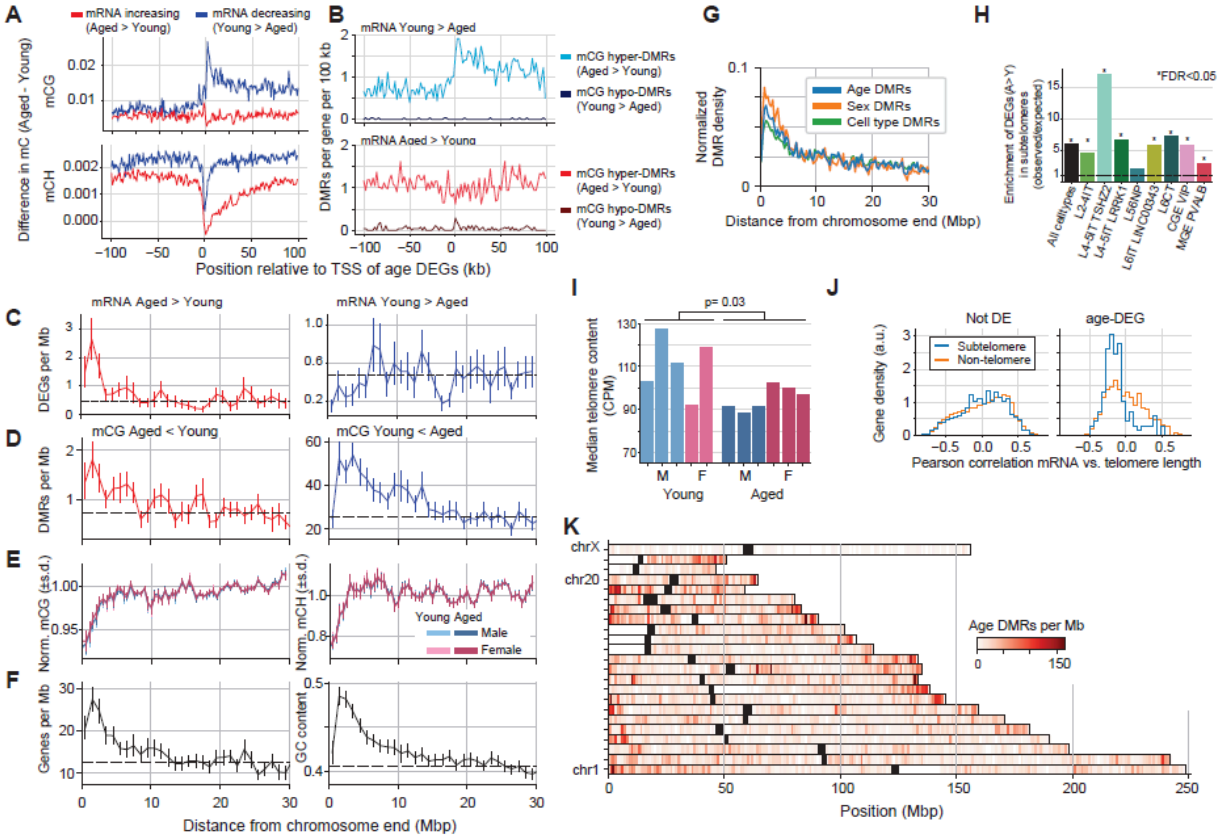


Figure 1.4 | Increased expression of subtelomeric genes in aged neurons.
(A) Average DNA methylation around the promoter of age-DEGs whose expression increases (red) or decreases (blue) with age. Age-decreasing genes have higher gene body mCG, while age-increasing genes have lower mCH both in the gene body and in the surrounding region. (B) Density of age-differentially methylated regions (age-DMRs) around age-DEGs. Hypermethylated age-DMRs are enriched in age-decreasing genes. Hypomethylated age-DMRs are rare, but they are enriched in the promoters of age-increasing genes. (C-F) Distribution of age-DEGs (C), age-DMRs (D), normalized DNA methylation (E), genes and genomic GC content (F) as a function of distance to the nearest chromosome end. Error bars are the standard error across chromosomes, except (E) where error bars represent the standard deviation across cell types. (G) Normalized density of age-, sex- and cell type-DMRs as a function of distance to the nearest chromosome end. (H) Enrichment of age-increasing DEGs in subtelomeres (≤5 Mb from chromosome end). . Enrichment calculated as (observed/expected) subtelomeric A>Y genes, where expected = (# A>Y genes)(# subtelomeric DEGs) / (# DEGs). *FDR <0.05, Fisher exact test (I) Median telomere length in L2-4IT neurons for each donor, estimated by TelomereHunter75. (J) Pearson correlation of pseudobulk mRNA expression with donor telomere length for DE and non-DE genes in subtelomeric or non-telomeric regions in L2-4IT neurons. (K) Map of the density of age-DMRs across the genome.

1.2.8 Transcriptome and DNA methylome differences between male and female neurons

In contrast with the effect of age, sex differences in cell type proportion and global CG or CH DNA methylation were not significant (Fig. S1.4A, Fig. S1.7A,B, FDR >0.1)[82]. We detected a total of 249 autosomal and 13 X-linked differentially expressed genes (sex-DEGs) across all cell types; Y-linked genes were excluded from this analysis (Fig. 1.5A, Table S1.5). We also identified 11,702 unique autosomal sex-DMRs (merged across cell types), covering 3.1 Mb (Table S6). As previously reported, sex-DMRs also occurred throughout chrX, covering 108 Mb (69% of the X chromosome)[35,57].

Sex-specific differential expression in our dataset was consistent with analysis of whole-tissue RNA-seq from human frontal cortex (Brodmann area BA9) in n=175 donors (GTEx, Fig. 1.5B, Fig. S1.7C, D)[5,135]. The strongest correlation was with L2-4 IT neurons (Spearman $r=0.21$, $p < 10^{-10}$), which represent the largest subpopulation of neurons in the frontal cortex (Fig. 1.5B, S7C). A notable example of an autosomal gene with strong sex-biased expression in our data was protein disulfide isomerase family A member 2, *PDIA2* (Fig. 1.5C,D). *PDIA2,* which is reported to modulate estradiol activity through direct binding[136], is expressed 3.5-4.5 fold more in male than female neurons.

Among the thousands of autosomal sex-DMRs, there was a similar number of regions with increased methylation in males and in females. As an example, the long non-coding RNA, *LINC01115,* had a prominent pattern of mCG DMRs (female < male) at the promoter and 5' intronic region in both excitatory and inhibitory neurons (Fig. 1.5F). The gene also had significantly lower gene body mCH in females than males. Consistent with this female hypo-methylation, *LINC01115* was expressed more strongly in females than males (Fig. 1.5E). *LINC01115* was recently reported to have a striking increase in the active chromatin mark H3K4me3 in the prefrontal cortex of schizophrenic patients compared to controls[137].

1.2.9 Cell-type-specific X-inactivation escape genes

X chromosome inactivation balances gene expression on chrX in females and males. This process prevents accumulation of mCH across the majority of the female inactive X chromosome[57]. A subset of X-linked genes escapes X inactivation and accumulates high levels of mCH in the gene body in female neurons[33,35,57]. Previous studies using bulk tissue samples could not identify cell-type-specific X-inactivation escape features. Using our single cell data, we identified 41 chrX genes with female-to-male gene body mCH ratio greater than 1.5 in at least one cell type, which we defined as putative escapee genes (Fig. 1.5F,G). Previously identified escape genes[33,138] such as *UBA1* and *KDM5C* had consistent escape signatures across all cell types. However, some genes previously identified as escape genes exhibited cell-type-specific CH methylation differences between males and females. For instance, *GEMIN8* had higher gene body CH methylation in females compared to males in excitatory neurons and CGE-derived inhibitory neurons, but not in MGE-derived inhibitory neurons. *IQSEC2* lacks escape signatures in VIP and SST neurons. Moreover, genes previously considered to be subject to inactivation exhibited cell-type-specific escape signatures in our data. For example, *BCAP31* and *ATP2B3* had elevated female CH methylation in excitatory neurons but not in inhibitory neurons. By contrast, *USP11* had strong escape signatures in inhibitory neurons but not in excitatory neurons.

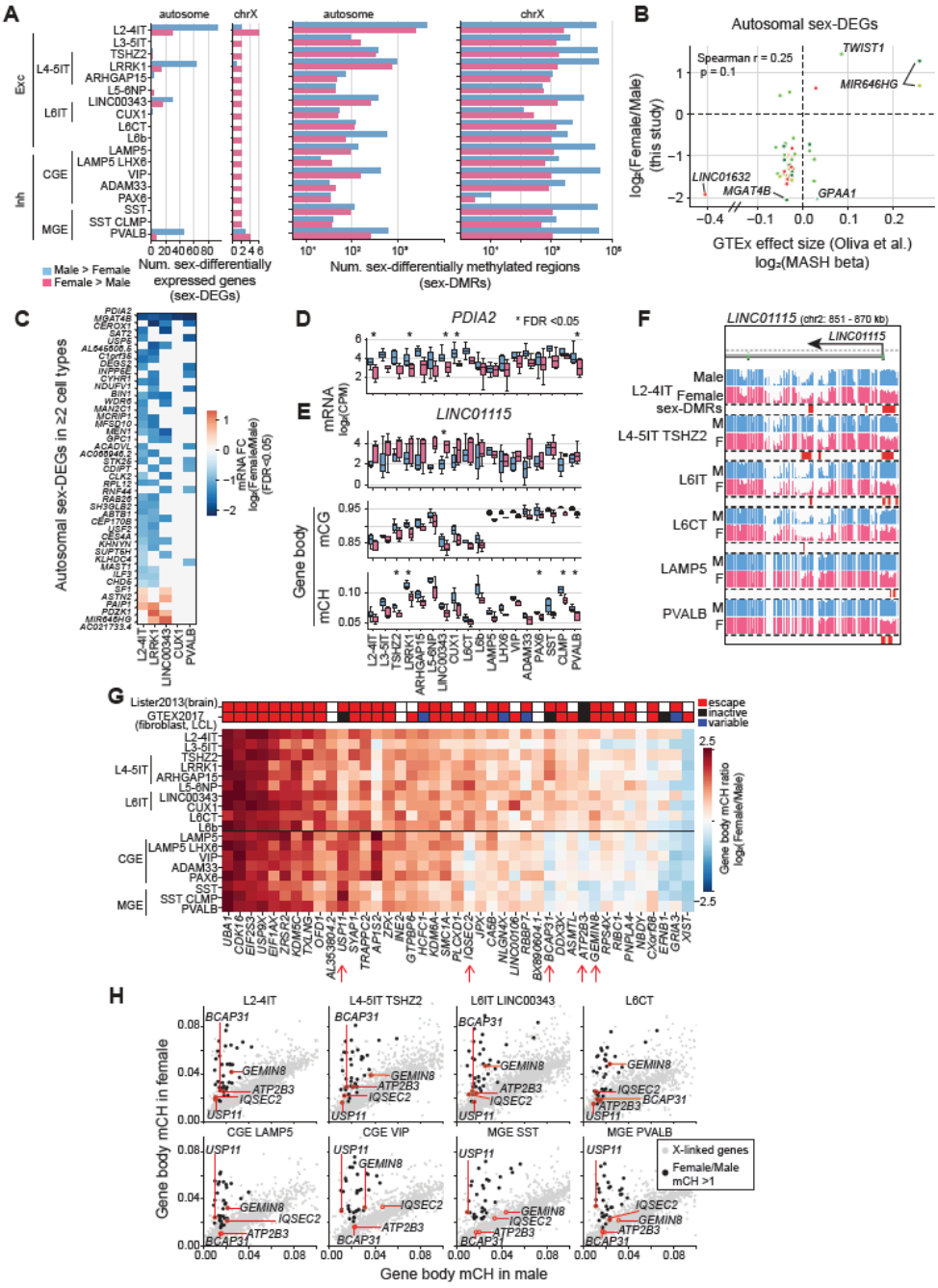


Figure 1.5 | Sex differences in DNA methylation and gene expression in human neurons.
(A) Number of autosomal and X-linked sex differentially expressed genes (sex-DEGs) and differentially methylated regions (sex-DMRs) in each cell type. (B) Comparison of the female/male expression for autosomal sex-DEGs in neural cell types with GTEx data from bulk frontal cortex[5] (colors correspond to cell types shown in Fig. 1.3A). (C) Fold-change of autosomal sex-DEGs that are significant in multiple cell types. (D, E) Sex-specific RNA expression of *PDIA2*, and RNA and gene body DNA methylation of *LINC011115*. (F) CG DNA methylation in the promoter and 5' intronic region of *LINC011115*, with female hypo-mCG sex-DMRs. (G) Gene body CH methylation ratio (female/male) in 41 putative chrX escape genes with female mCH > 0.02 and female/male mCH >1.5 in at least one cell type. (H) Scatter plots of gene body mCH in male and female neurons for X-linked genes (gray dots) and putative escape genes (black). Cell- type-specific X-inactivation escape genes are labeled.

1.3 Discussion

Neurons establish a stable cell identity, encoded in the epigenome and expressed through gene transcription, during prenatal development. Patterns of DNA methylation and gene expression are refined postnatally as neural circuits mature[33,89,90,139]. In contrast with the highly dynamic trajectory of cell regulation in early life, the major epigenetic signatures that define each cell type remain stable throughout adulthood. Despite this stability, changes in DNA methylation at specific loci may occur due to environmental factors experienced through the lifetime or stochastic erosion of the developmentally programmed epigenome[46,86,91,140–142]. Epigenetic changes could impact gene expression and play a role in cognitive alterations associated with aging. Our study provides single-cell multi-omic DNA methylation and RNA expression information about age- and sex-related differences in the epigenome and transcriptome of adult human frontal cortical neurons. By deeply interrogating >55,000 cells across 11 individuals, and validating our findings with public data resources from large populations, we uncovered cell-type-specific patterns of altered epigenetic and transcriptomic regulation in the aged human brain.

Studies of the so-called DNA methylation clock have shown that changes in methylation at several hundred CG sites throughout the genome constitute a broadly shared predictive signature of age with similar features across human tissues and mammalian species[108,127,143]. Our data using whole-genome sequencing of the DNA methylome and transcriptome (snmCT-seq[144]) in the adult human cortex reveal cell-type-specific, age-related changes in each of the specialized excitatory and inhibitory neuron types. Because our data covers the vast majority of cytosines, including both CG and non-CG sites, we uncover widespread age-DMRs spanning over 38 Mbp that extend far beyond the hundreds of sites used in methylation clock models.

We found that intratelencephalic-projecting (IT) excitatory neurons in superficial (L2-4) and middle cortical layers (L4-5) were most affected by aging in terms of both gene expression

and DNA methylation. Age-related changes in DNA methylation correlated with age-related changes in gene expression. Increased DNA methylation at both CG and CH sites accompanied decreased gene expression, particularly for genes implicated in synaptic structure and function[74] such as *EGR1*[126], *NLGN1*[112], *PHACTR1*[113], and *KCNH*[114].

A notable example of coordinated epigenetic and transcriptomic repression of synaptic regulation with age was the neural activity-induced transcription factor, EGR1. Distal enhancers with age-increased CG methylation (hypermethylated age-DMRs) were enriched in the DNA sequence motif associated with binding EGR1. A recent study suggests that EGR1 may regulate gene expression via coordination with TET-mediated demethylation[145]. The transcriptional repression of methylated reporter constructs was reversed by overexpression of EGR1 and TET, but not by expression of either factor alone[145]. Our observation of lower *EGR1* expression in aged donors and the enrichment of the EGR1 motif in hypermethylated age-DMRs is consistent with this link between EGR1 and TET-mediated demethylation. Our data extend the known role of DNA methylation in pre- and postnatal brain development, when cell-type-specific patterns of CG and CH methylation are established as neurons acquire and consolidate their mature identity[33]. Our findings indicate that aging of the adult brain, which broadly impacts cognitive function, coincides with coordinated changes in the DNA methylome and transcriptome of neurons, particularly upper layer excitatory cells.

Whereas age-decreasing genes were associated with synaptic function, we found that age-increasing genes were instead characterized by a distinctive genomic distribution. These genes were enriched in subtelomeric regions, within ~5 Mbp of the ends of chromosomes, consistent with a so-called telomere position effect observed in non-brain tissues in which telomeres shorten with age[70]. Moreover, we found evidence for shorter telomeres in neurons from aged compared to young donors, although telomere length estimates from our snmCT-seq data are difficult to validate[133]. These observations suggest that up-regulation of gene expression in subtelomeres of post-mitotic neurons may not be related to cell division. Instead,

telomere shortening could potentially be related to DNA damage repair in neurons[87,146]. Although there was no pattern of enrichment for specific functional annotations with subtelomeric age-increased genes, these genes included some with notable roles such as the glucocorticoid receptor co-chaperone *FKBP5*[129,130].

By contrast with excitatory neurons, GABAergic inhibitory cells had highly similar gene expression and DNA methylation patterns in young adult and aged humans. GABAergic cells had fewer age-DEGs or age-DMRs compared with excitatory neurons. Recent studies have noted the relative conservation of inhibitory compared to excitatory neuron gene expression across brain regions, mammalian species, and individual human donors[62,63,147]. Instead of altered regulation, we observed a strong reduction in the proportion of two GABAergic neuron types, SST and VIP expressing cells, in aged compared with young adult donors. The reduction of these two neuron types has been previously inferred from deconvolution analyses of bulk human brain tissue and single-cell brain data in aging donors[65] and Alzheimer's disease patients[12,81,103]. A recent snRNA-seq analysis of 75 human brain tissue samples from multiple brain regions obtained from surgeries in epilepsy and cancer patients[82] did not identify age-related differences in neuronal cell type composition. This discrepancy may stem from differences in the study population or the brain regions analyzed. Additionally, our observation could be influenced by the efficiency of sampling nuclei from distinct neuron types using our snmCT-seq procedure, which relies on fluorescence-activated nuclei sorting (FANS) to enrich neurons. To further validate this observation, cell counts using microscopy and/or spatial transcriptome data are warranted.

Differences between males and females in the incidence and symptoms of neuro-cognitive disorders including autism, schizophrenia and depression have spurred interest in sex differences in brain cell regulation[4,148]. To date, sex differences in human brain transcriptomes have mainly been assessed using bulk tissue[5,135]. Although the same cell types are present in the frontal cortex of males and females, we found notable examples of

cell-type-specific sex differences in gene expression and DNA methylation. We found autosomal genes that were more expressed in males, including *PDIA2,* an intracellular estrogen binding protein[136]*.* We also found genes that were more expressed in females, including the non-coding RNA *LINC01115* which had a striking pattern of female hypo-methylated DMRs at its promoter and gene body. Finally, we identified X-linked genes with DNA methylation signatures of escape from X-inactivation, including genes such as *GEMIN8* with cell-type-specific escape signatures[35,57].

Our study has several limitations. First, we performed deep sequencing of >55,000 cells from each of 11 donors, a design that enables analysis of specific brain cell types at the expense of a broader representation of variability across more donors. The consistency of our age- and sex-DEGs and of the pattern of subtelomeric enrichment of age-increasing genes with GTEx data from hundreds of donors[5,70,135] validates the reliability of our sample. As noted above, our observation of the depletion of SST and VIP cells in aged donors could be impacted by artifacts due to nuclei sorting. Finally, our study focused on neurons by using nuclei sorting with the neuronal marker NeuN. Analysis of age-related epigenomic and transcriptomic alterations in human glial cells is important given evidence for their role in aging and age-related disease[9].

Single-cell sequencing has revolutionized biological studies by revealing the landscape of gene regulation across diverse cell types. Our study shows that multi-omics can extend this powerful approach to assess major dimensions of inter-individual diversity, including age and sex. Differences in brain cell regulation between individuals are an important aspect of human brain function and must be addressed as we seek a comprehensive understanding of brain cell diversity[149].

1.4 Methods

1.4.1 Experimental Model and Study Participant Details

Human postmortem brain tissue samples from dorsolateral prefrontal cortex (Brodmann area BA46) were acquired via the NIH Neurobiobank and included samples provided by the Brain Tissue Donation Program at the University of Pittsburgh (IRB Protocol Number: REN14120157/IRB981146), University of Miami Brain Endowment Bank (IRB Protocol Number: 19920358 (CR0001775)), and The Human Brain and Spinal Fluid Resource Center (managed by Sepulveda Research Corporation) (IRB Protocol Number: PCC#: 2015-060672, VA Project #: 0002). All donors had no clinical brain diagnosis. For each donor, two independent samples (biological replicates) were obtained either by using separate tissue punches, or in cases where only a single tissue punch was available by separately dissecting cells from opposite sides of a single piece of tissue. RNA integrity (RIN) values were independently tested using an Agilent TapeStation RIN screentape. RNA for the test was isolated from tissue samples using TRIzol-chloroform extraction and eluted on Qiagen RNeasy silica membrane spin columns after DNase I digestion. Specific characteristics of the individuals are displayed in Table S1.1.

1.4.2 Nuclei isolation, sorting and snmCT sequencing

Nuclei isolation and FANS sorting was performed as previously described[99,150], with added protease inhibitors to the homogenization buffer. A detailed protocol can be found at https://www.protocols.io/view/snmcat-v2-x54v9jby1g3e/v2. snmCT-seq library preparation and sequencing was performed as described in [42,99,151] and protocol is listed at https://www.protocols.io/view/methyl-c-sequencing-of-single-cell-nuclei-snmc-seq-pjvdkn6 Briefly, postmortem brain chunks were ground in liquid nitrogen and stored at - 80 ℃. The ground tissue was then lysed and fractionated to collect nuclei. The purified single-nuclei were stained with nuclei stain Hoechst 33342 and anti-NeuN antibody AlexaFluor 488 followed by fluorescence-activated nuclei (FANS) sorting into 384-well plates, where cDNA was reverse

transcribed and amplified. The samples were bisulfite converted in each well. Finally, the libraries were sequenced by the Illumina NovaSeq 6000 instrument with S4 flowcells and 150-bp paired-end mode.

1.4.3 Mapping snmCT-seq data

We pre-processed snmCT-seq data as previously described[99,105] using pipeline (http://cemba-data.readthedocs.io/). The snmCT-seq library contains both DNA- and RNA-derived (cDNA) fragments. To separate these reads from each modality, we mapped reads in both DNA mode and RNA mode and then separated them by the total methylation level in each read. The main steps were: (1) trimming adapters, (2) DNA mapping, (3) call methylation level and extract DNA reads, (4) RNA mapping, (5) call methylation level per read and extract RNA reads, and (6) RNA expression quantification.

(1) Trimming: Cutadapt (v4.1)[152] in pair-end mode was used to trim reads with R1 adapter (AGATCGGAAGAGCACACGTCTGAAC), R2 adapter (AGATCGGAAGAGCGTCGTGTAGGGA), template switching oligo (TSO) (AAGCAGTGGTATCAACGCAGAGTGAATGG), N6 (AAGCAGTGGTATCAACGCAGAGTAC), ISPCR(AAGCAGTGGTATCAACGCAGAGT), poly T and polyA. Low quality ends (-q 20) and 10 bps from both the 3' and 5' end are further trimmed. Reads shorter than 30 bps were discarded.

(2) DNA mapping: Reads were mapped to the human genome assembly GRCh38 (hg38) using Hisat-3n (v2.2.1-3n-0.0.3)[153] with parameters: --base-change C,T; --no-spliced-alignment; --no-temp-splicesite; --directional-mapping-reverse. Unmapped reads (MAPQ=0) were discarded while unique-mapped reads (MAPQ=60) and multi-mapped reads (MAPQ=1) were kept separately. PCR duplicated reads were removed using Picard MarkDuplicates (v2.27.4).

(3) Call methylation level and extract DNA reads: Reads with global CH methylation level less than 0.5 were collected as DNA reads.

(4) RNA mapping: Reads were mapped to hg38 using Hisat2 (v2.2.1)[154] using default parameters. Multi-mapped reads were removed.

(5) Call methylation level and extract RNA reads: Reads with global CH methylation level greater than 0.9 were collected as RNA reads.

(6) RNA expression quantification: featureCounts (v2.0.1)[155] was used to assign reads to gencode v37 genes including both exon and introns (-t gene) with fraction setting (--fraction) to evenly distributed reads mapped to overlapped features.

1.4.4 Personalized genome sequencing and mapping

The genomic DNA was extracted from ground frozen postmortem brain tissue. The library preparation and sequencing were performed as described in [150].

Reads from personalized genome sequencing were mapped to human genome assembly GRCh38 (hg38) as follows. Reads were first trimmed with Illumina adaptor sequencing using trim_galore (v0.6.7). 10 bp from the 3' end from read1 and read2 were also removed. The trimmed reads were mapped by bwa mem (v0.7.17)[156], deduplicated by Picard MarkDuplicates (v2.27.4), and SNPs were called using the Genome Analysis Toolkit (GATK, v4.2.6.1) as instructed in the GATK tutorial[157]. For each donor, cytosines and adjacent downstream(3') sites which overlap homozygous SNPs called from any donor, were removed for the downstream methylation analysis.

1.4.5 Cell filtering

The criteria for a cell to pass DNA methylome quality control were: (1) DNA unique mapped read count > 300,000, (2) mapping rate > 0.5, (3) global mCG level > 0.5, (4) global mCH level < 0.2, and (5) global mCCC level < 0.03. The criteria for a cell to pass RNA quality control were: (1) number of genes detected > 500, (2) percent of mitochondria reads < 10%, and (3) doublet score (called by scrublet[158]) < 0.3.

1.4.6 Feature filtering

100kb methylation bin features were filtered by coverage and blacklist regions. Bins with mean number of covered cytosine <500 or >2000, or overlapped with ENCODE blacklist[159] regions were removed for the clustering. Genes expressed in fewer than 100 cells were removed for further analysis.

1.4.7 Clustering and cell type annotation

Clustering was performed using non-CG DNA methylation features in 100kb bins throughout the genome, as previously described[42,99]. The top 20,000 most variable principle components of the 100kb bins were used to construct the KNN graph using scanpy.pp.neighbors with k=25. Next, the KNN graph was used to perform clustering using the Leiden community detection algorithm[160] with leiden resolution 1. Consensus clustering was performed by repeating the cluster analysis 500 times with the same resolution (0.7) and different random seed. Clusters were then merged using a balanced random forest classifier (imblearn.ensemble.BalancedRandomForestClassifier[161]) trained using 50% of cells per cluster, while limiting the maximum number of cells to 500 in each cluster to avoid large clusters dominating the training dataset. The merging process stops when it reaches the target accuracy (0.96). Overall, the clustering process is done by ConsensusClustering function in ALLCools (https://lhqing.github.io/ALLCools). Clusters with a high duplication ratio or containing less than 300 cells were excluded from the downstream analysis. Clusters were then annotated based on cell type marker gene expression and gene body methylation features.

1.4.8 Differential expression and gene ontology

We merged nuclei from each sample for each cell type to get peudobulk counts. Protein coding and lncRNA genes from chromosomes 1-22 and X with counts per million (CPM) >5 in at

least 2 samples are defined as “expressed genes”. Differential expression testing was performed separately for each cell type and each comparison (aged vs, young or female vs. male) using Dream[162] (variancePartition v1.28.9, edgeR v3.40.2). Specifically, calcNormFactors was used to control for the raw library size from the raw counts. Then, voomWithDreamWeights and dream were used to prepare for the linear mixed model fit from the normalized count and design matrix. The model included fixed effects of age and sex, and a random intercept by donor: $expression \sim age + sex + (1|donor)$. Low RNA quality samples, including AM2, AM3 and AF2A were excluded from the differential expression testing. Finally, after testing the generalized linear mixed model for each gene, genes with absolute fold-change greater than 1.2 and adjusted p-value < 0.05 are defined as differentially expressed genes (DEGs) in our analysis.

To control for the number of nuclei for each cell type from each condition, we also applied the differential expression test after randomly selecting 50 cells from each sample for each major cell type. The subsampling and DEG testing was repeated five times. The median, 25th and 75th quantile of the number of differentially expressed genes are shown.

Functional gene ontology (GO) enrichment analysis was performed using clusterProfiler (v 3.14.3)[163]. All expressed genes in each cell type were included as the background gene list. Gene ontology with gene size between 25 and 500 were included in the enrichment test. The Benjamini-Hochberg method was used to control the false discovery rate.

1.4.9 Differential methylation analysis

We created pseudobulk counts for nuclei from each donor for each cell type and tested each CpG site between each condition comparison (aged vs. young or female vs. male) to detect differentially methylated regions (DMRs) using DSS[109,110] (DSS v2.34.0, bsseq v1.22.0). DMLfit.multiFactor was applied with model formula: $mCG \sim age + sex$, with a smoothing window size 500 bp. Next, DMLtest.multiFactor was used to call differentially methylated loci. Finally,

callDMR was used to merge significantly differentially methylated loci ($p < 0.01$) with a distance smaller than 50 bp. Merged regions with a length greater than 50 bp and containing at least 3 CpG sites are called differentially methylated regions (DMRs).

To control for the number of nuclei for each cell type from each condition, we also apply the differential methylation test after randomly subsampling 100 cells from each donor for major cell types. The subsampling was done five times. The median, 25th and 75th quantile of the number of differentially methylated regions are shown.

1.4.10 DMR enrichment test

We utilized the chromHMM annotations of 15 chromatin states derived from human frontal cortex tissue[122]. For each chromatin state, we computed the overlap in base pairs (Ntrue) between the DMRs and the designated state regions. Subsequently, this process was repeated 1,000 times, each time with the DMRs randomly shuffled across the genome. The number of base pairs overlapping between shuffled DMRs and the chromatin state region are fitted using a normal distribution. Finally, Ntrue was used to compare with the distribution to derive both the z-score and the associated p-value. The Benjamini-Hochberg method was used to control false discovery rate.

1.4.11 Methylation level in genomic regions

The methylation level in a defined region is calculated by summing over the total coverage of methylcytosine and dividing it by the sum of the total coverage of cytosine. The calculation is performed using allc-to-region-count function in ALLCools (https://lhqing.github.io/ALLCools).

1.4.12 DNA methylation valleys (DMVs)

To call DNA methylation valleys (DMVs), we first identified undermethylated regions (UMRs) and low methylated regions (LMRs) using MethylSeekR[164]. We used methylation cutoff m=0.5, minimum number of CpG, n=10, and FDR < 0.05 as parameters for detecting UMRs and LMRs. DMVs were defined as UMRs with length ≥5 kb and mean methylation level ≤15%. We detected 645-1546 DMVs in each cell type and 35%-63% of them overlapped with the DMVs detected in human embryonic stem cell-derived neural progenitor cells[128].

1.4.13 Estimation of telomere content from snmCT-seq

We estimated telomere content as a proxy of telomere length from snmCT-seq dataset using TelomereHunter (v1.1.0) [133]. In short, TelomereHunter first searched for reads with telomeric repeats from pre-aligned BAM files, and then classified the selected reads into four categories: intratelomeric, junction spanning, subtelomeric and intrachromosomal, using their alignment information.

We first generated per-cell BAM files that contained unmapped reads by re-running Hisat-3n, followed by de-duplication using Picard MarkDuplicates. Next, RNA reads identified from our previous steps described in the “Mapping snmCT-seq data” section were excluded. The resulting BAM files were used as the input of Telomerehunter to search for t-type (TTAGGG) telomeric repeats as well as c-type (TCAGGG), g-type (TGAGGG) and j-type (TTGGGG) telomeric variant repeats. Reads were considered as telomeric reads if they contained at least 6 repeats per 100 bp read length. Given the number of reads of comparable GC content (48-52%) was low per cell, we used the number of uniquely mapped DNA reads in each snmCT-seq library as the normalization factor. The telomere content for each cell was computed as the number of intratelomeric reads divided by the number of uniquely mapped DNA reads, and multiplied by 10^6.

1.4.14 Predicting X-inactivation escape genes

To identify the candidate of X-inactivation escape genes with high signal to noise ratio, we first computed the average mCH level across female donors and all cell types. Genes with mean mCH level smaller than 0.2 were excluded. Additionally, we retained only genes expressed (expression level >0) in all cell types. Finally, X-linked genes with female-to-male gene body mCH ratio greater than 1.5 in at least one cell type were listed as putative escapee genes.

1.4.15 Quantification and Statistical Analysis

Statistical significance was assessed by Wilcoxon rank-sum test (scipy.stats.ranksums) for Fig. 2A and Fig. 4I, and corrected for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) procedure (statsmodels.stats.multitest.fdrcorrection) for Fig. 2B, 3L and Fig. S4A-D, S7A,B. Statistical significance was assessed using Fisher's exact test (scipy.stats.fisher_exact) with Benjamini–Hochberg FDR correction for Fig. 4H. For Fig. S5G, significance was assessed using Fisher's exact test implemented in the HOMER software[123] (q-value (Benjamini)). The values of n (number of donors) and significance were reported in results and figure legends.

1.5 Acknowledgements

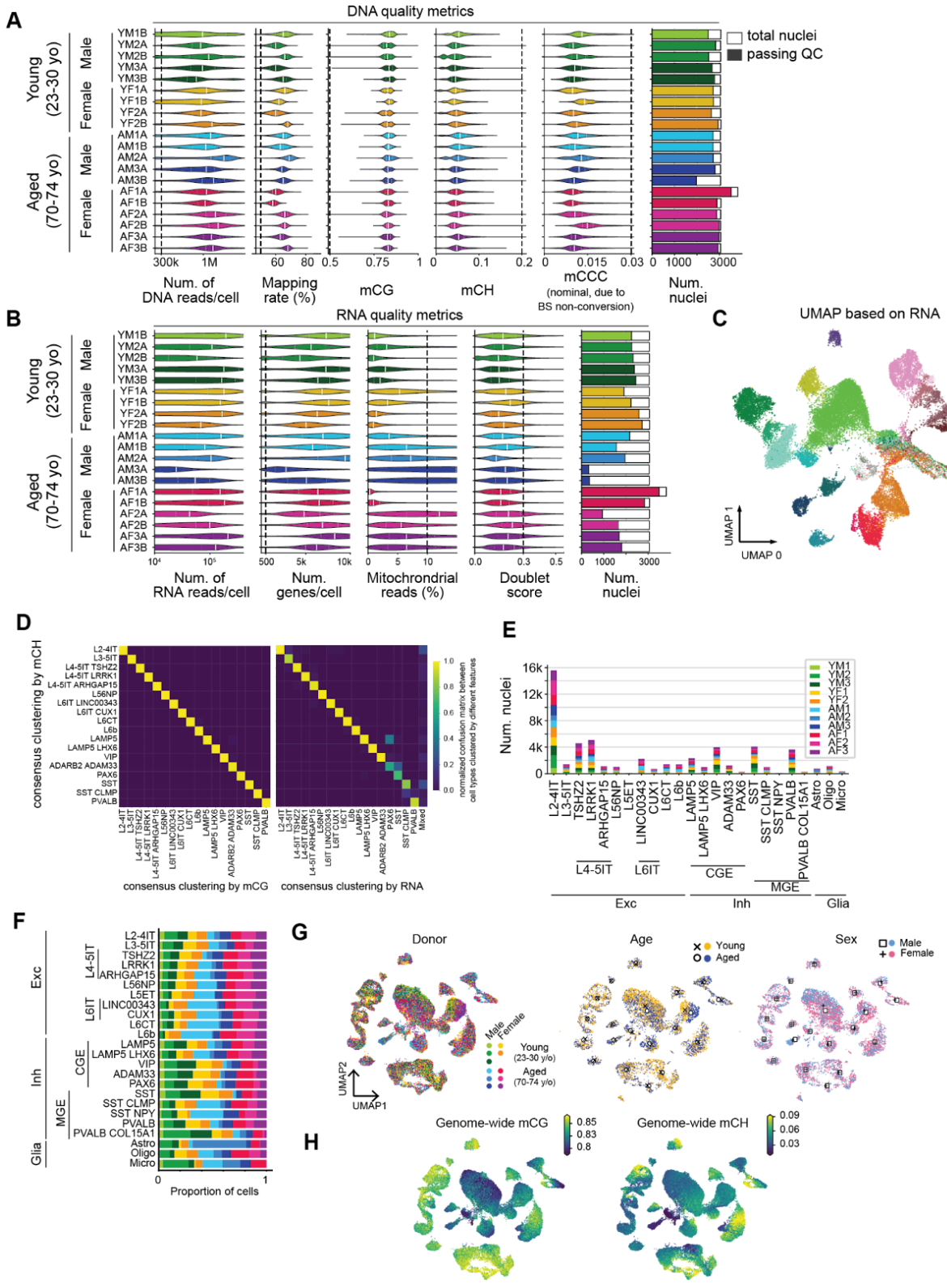
Chapter 1, in full, is a reprint of the material as it appears in *Neuron* 2024. Chien, J.-F., Liu, H., Wang, B.-A., Luo, C., Bartlett, A., Castanon, R., Johnson, N. D., Nery, J. R., Osteen, J., Li, J., Altshul, J., Kenworthy, M., Valadon, C., Liem, M., Claffey, N., O'Connor, C., Seeker, L. A., Ecker, J. R., Behrens, M. M. & Mukamel, E. A. Cell-type-specific effects of age and sex on human cortical neurons. *Neuron* (2024). doi:10.1016/j.neuron.2024.05.013. The dissertation author was the first author of this paper.

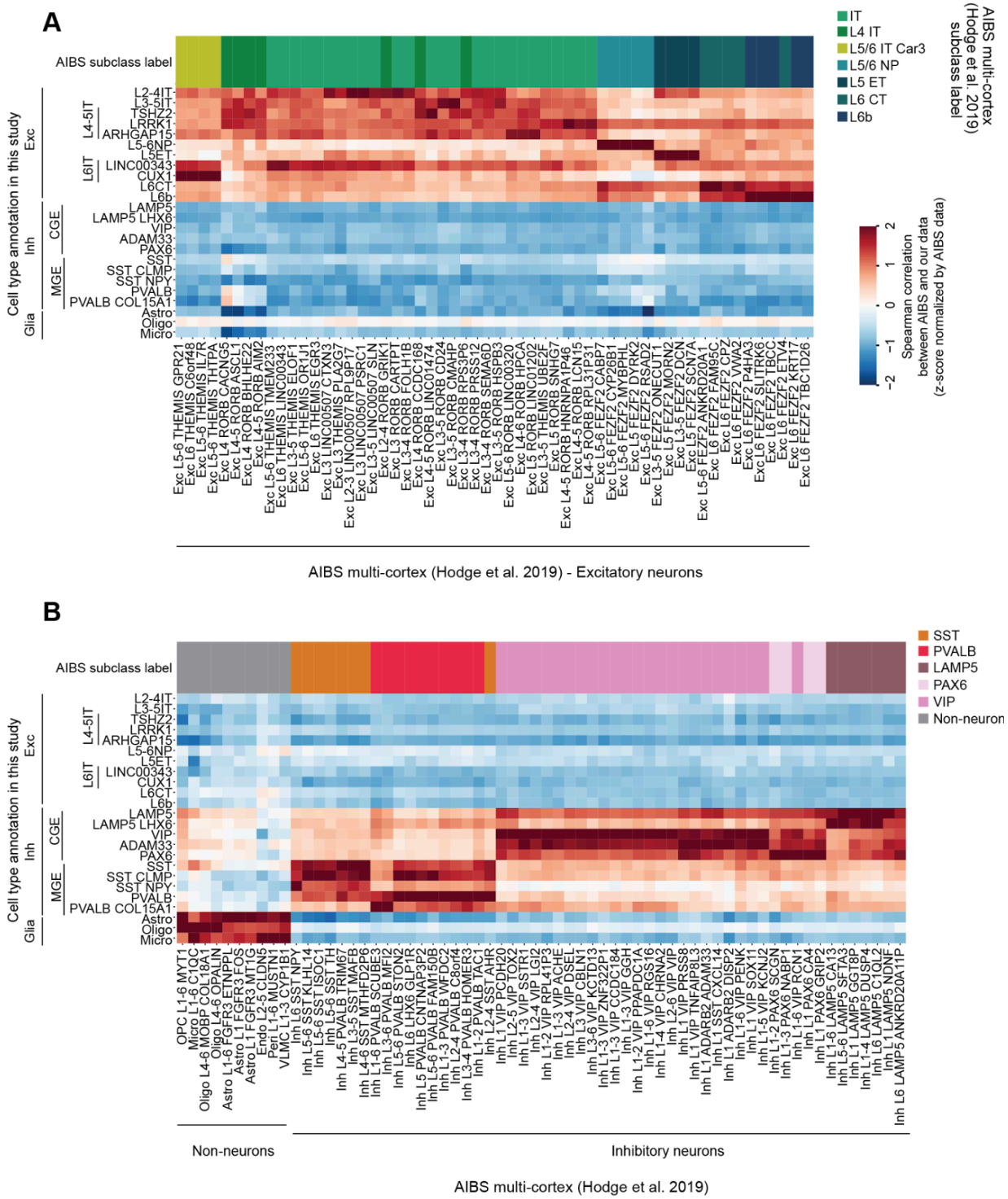
Conceptualization, JC, JRE, MMB, EAM; Methodology, JC, HL, BAW, CL, JRE, MMB, EAM; Sequencing data production, HL, BAW, AB, RC, NDJ, JRN, JO, JA, MK, CV, ML, NC,

COC; Formal analysis, JC, JL, EAM; Writing, JC, EAM; Writing - review and editing, All authors; Supervision, JRE, MMB, EAM.

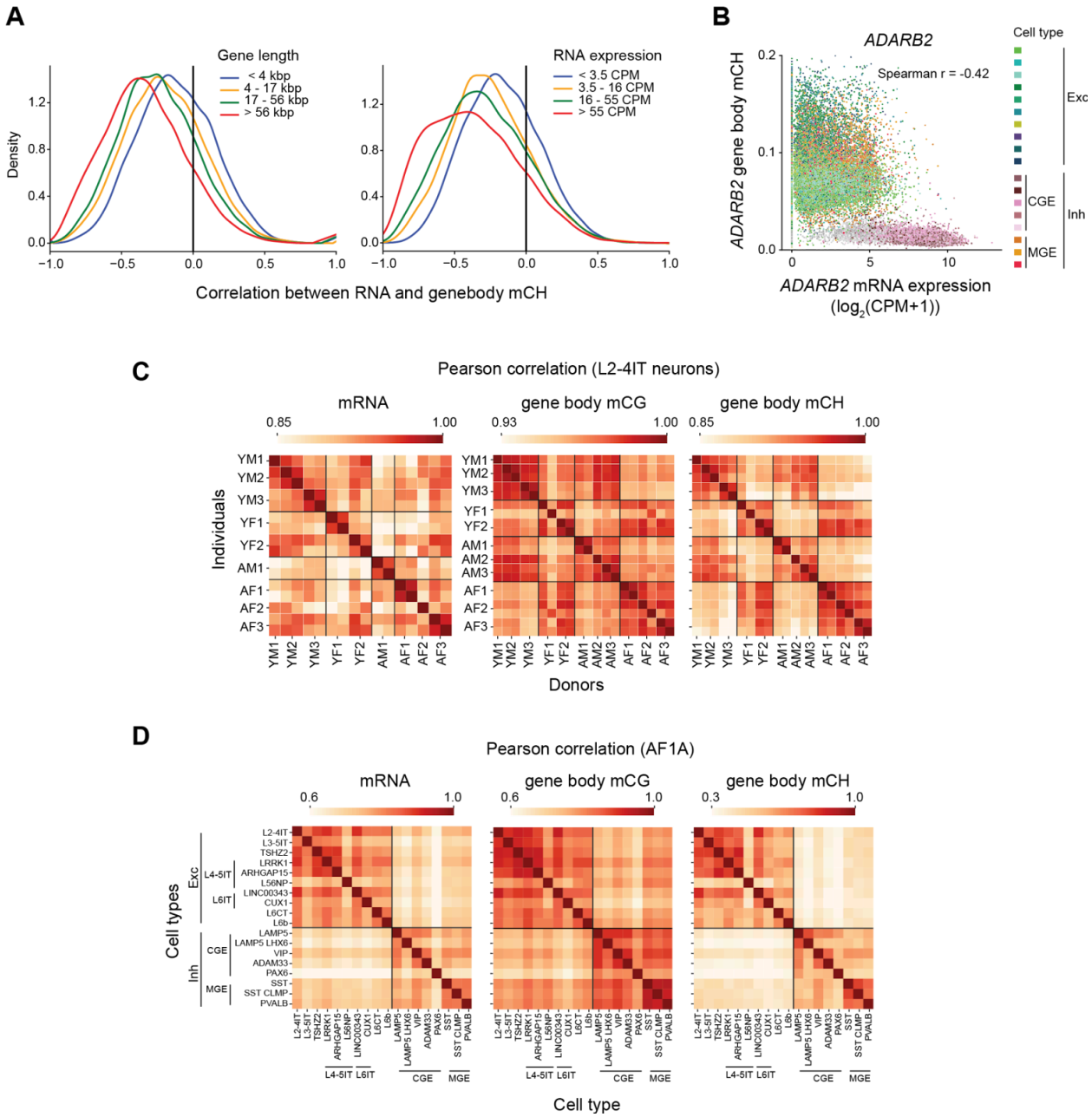
This study was supported by a grant from the Chan Zuckerberg Initiative (CZI) Seed Networks for the Human Cell Atlas (CZF2019-002457). We are grateful to Xin Jin for valuable feedback and discussions.

Supplemental Figure 1.1 | Quality control (QC) metrics for snmCT-seq data, related to Figure 1.1. (A) QC for DNA. Number of DNA reads, mapping rate, genome-wide mCG, mCH, mCCC level for all nuclei grouped by samples. The last column shows the number of total nuclei (open) and the number of nuclei passing DNA QC for each donor (filled, DNA unique mapped read count > 300,000, mapping rate > 0.5, mCG > 0.5, mCH level < 0.2, and mCCC < 0.03). (B) QC for RNA. Number of RNA reads, number of detected genes, percentage of reads mapped to the mitochondrial genome, and doublet score for all nuclei grouped by samples. The last column shows the total number of nuclei (open) and the number of nuclei passing RNA QC for each donor (filled, number of detected genes > 500, percent of mitochondria reads < 10%, and doublet score < 0.3). (C) UMAP embedding of snmCT-seq cells based on mRNA expression. Cells are colored according to their cell type annotation based on mCH features, as in Fig. 1.1B. (D) Confusion matrices showing consistent cell type assignments using mCH, mCG or mRNA features. (E) Number of nuclei per cell type per donor. (F) Proportion of nuclei. (G) UMAP embedding of nuclei colored by donor, donor age, donor sex. Markers show centroids of cells by type from young (x) and aged (circles) or male (square) and female (plus) donors. (H) UMAP embedding of nuclei colored by genome-wide mCG, mCH.


A
DNA quality metrics
Young (23-30 yo)
Aged (70-74 yo)
Male
Female
YM1B
YM2A
YM2B
YM3A
YM3B
YF1A
YF1B
YF2A
YF2B
AM1A
AM1B
AM2A
AM3A
AM3B
AF1A
AF1B
AF2A
AF2B
AF3A
AF3B
total nuclei
passing QC
300k
1M
60
80
0.5
0.75
1
0
0.1
0.2
0
0.015
0.03
0
1000
3000
Num. of DNA reads/cell
Mapping rate (%)
mCG
mCH
mCCC (nominal, due to BS non-conversion)
Num. nuclei
B
RNA quality metrics
10⁴
10⁵
500
5k
10k
0
5
10
0.0
0.3
0.5
0
1000
3000
Num. of RNA reads/cell
Num. genes/cell
Mitochrondrial reads (%)
Doublet score
Num. nuclei
C
UMAP based on RNA
UMAP 1
UMAP 0
D
consensus clustering by mCH
L2-4IT
L3-5IT
L4-5IT TSHZ2
L4-5IT LRRK1
L4-5IT ARHGAP15
L56NP
L6IT LINC00343
L6IT CUX1
L6CT
L6b
LAMP5
LAMP5 LHX6
VIP
ADARB2 ADAM33
PAX6
SST
SST CLMP
PVALB
Mixed
normalized confusion matrix between cell types clustered by different features
1.0
0.8
0.6
0.4
0.2
0.0
consensus clustering by mCG
consensus clustering by RNA
E
Num. nuclei
16k
12k
8k
4k
0
YM1
YM2
YM3
YF1
YF2
AM1
AM2
AM3
AF1
AF2
AF3
L2-4IT
L3-5IT
TSHZ2
LRRK1
ARHGAP15
L56NP
L5ET
LINC00343
CUX1
L6CT
L6b
LAMP5
LAMP5 LHX6
VIP
ADAM33
PAX6
SST
SST CLMP
SST NPY
PVALB
PVALB COL15A1
Astro
Oligo
Micro
L4-5IT
L6IT
CGE
MGE
Exc
Inh
Glia
F
Exc
Inh
Glia
L4-5IT
L6IT
CGE
MGE
L2-4IT
L3-5IT
TSHZ2
LRRK1
ARHGAP15
L56NP
L5ET
LINC00343
CUX1
L6CT
L6b
LAMP5
LAMP5 LHX6
VIP
ADAM33
PAX6
SST
SST CLMP
SST NPY
PVALB
PVALB COL15A1
Astro
Oligo
Micro
0
1
Proportion of cells
G
Donor
Male
Female
Young (23-30 y/o)
Aged (70-74 y/o)
UMAP2
UMAP1
Age
Young
Aged
Sex
Male
Female
H
Genome-wide mCG
0.85
0.83
0.8
Genome-wide mCH
0.09
0.06
0.03

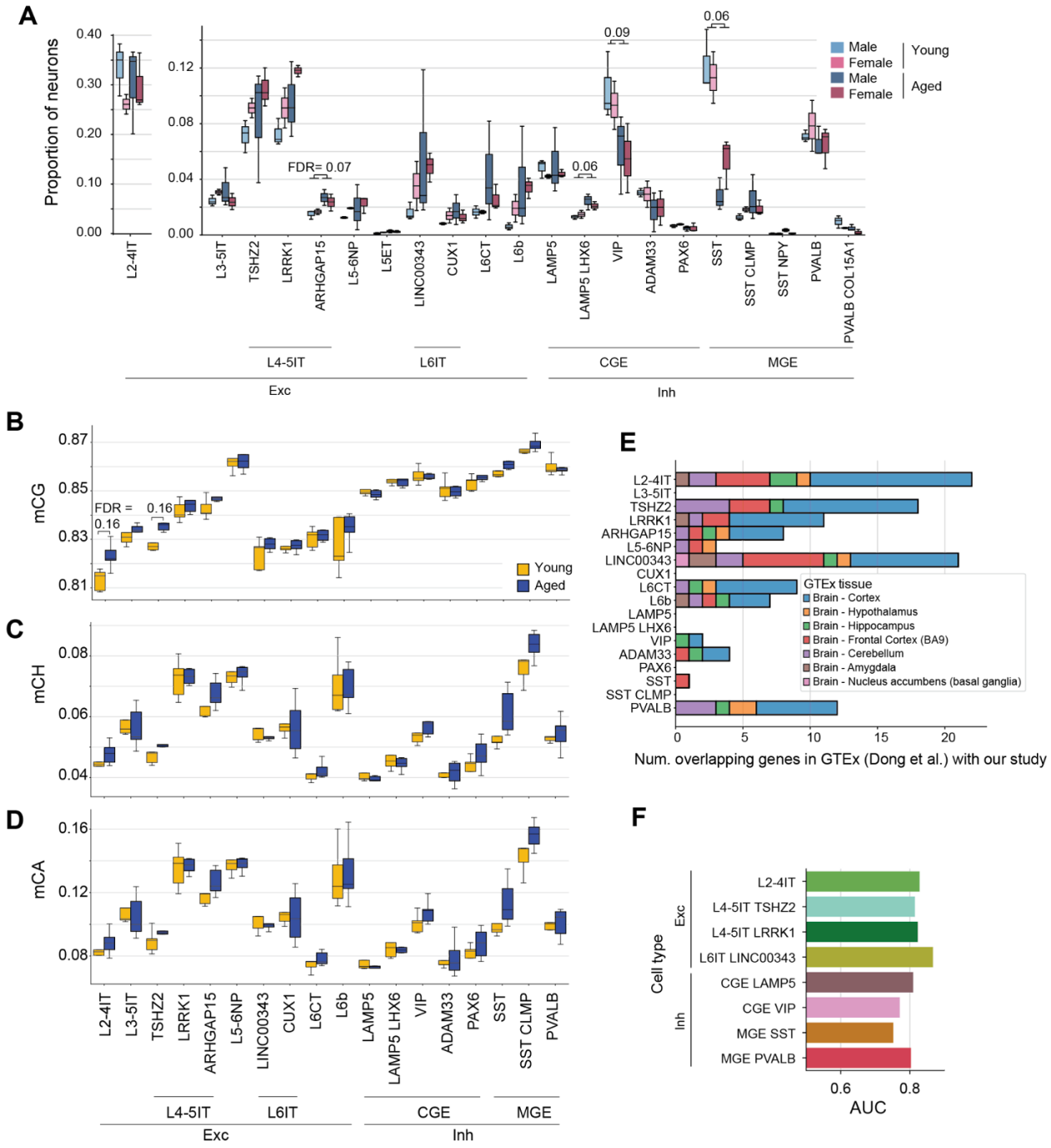


Supplemental Figure 1.2 | Comparison of RNA expression from snmCT-seq data (this study) with published snRNA-seq data, related to Figure 1.1.
Heatmap shows Spearman correlation between pseudobulk RNA expression (log2(CPM+1)) in our data and pseudobulk RNA expression (log2(CPM+1)) from AIBS multi-cortex data[62] for excitatory (A) and inhibitory neurons or non-neurons (B). The correlation is calculated using the top 500 most variable autosomal protein-coding genes.



Supplemental Figure 1.3 | Correlation between samples, cell types and relation between DNA methylation and RNA expression, related to Figure 1.1.

(A) Spearman correlation between RNA expression and gene body CH methylation, stratified by gene length (left) or gene expression (right). (B) The relation between gene body mCH and mRNA expression for each cell colored by cell types. (C) Pearson correlation of mRNA expression (left), gene body CG (middle) and CH (right) methylation between samples in L2-4IT neurons. The correlation was calculated by the top 10,000 variation features. (D) Pearson correlation of mRNA expression (left), gene body CG (middle) and CH (right) methylation between cell types in one donor. The correlation was calculated by the top 10,000 variation features.

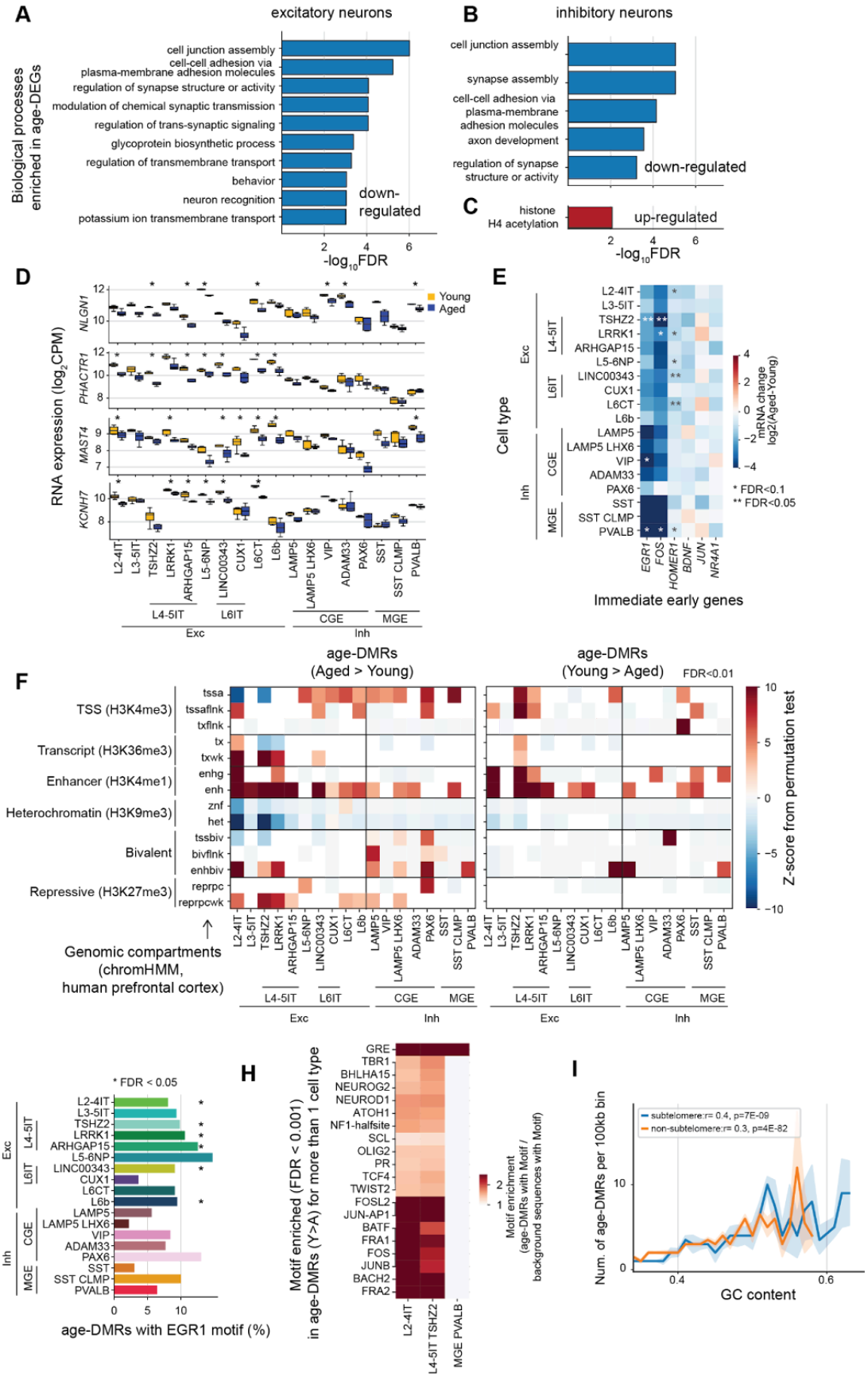


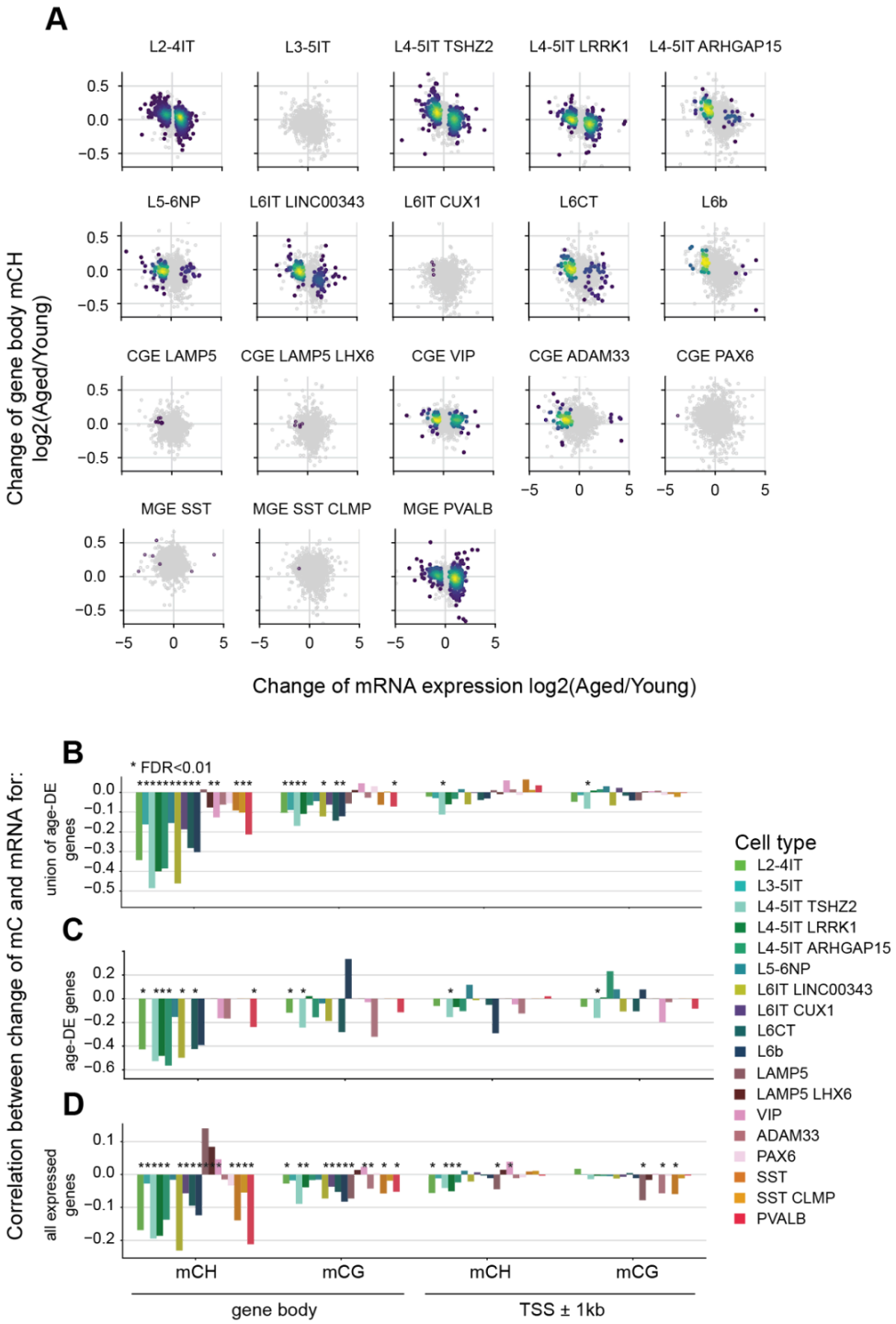
Supplemental Figure 1.4 | Cell type proportions and global DNA methylation changes in aging, related to Figure 1.2.

(A) Proportion of all neuron types in young and aged donors. Significant differences between aged and young donor groups are labeled (FDR < 0.1 from Wilcoxon rank-sum test and Benjamini–Hochberg correction, n=5 for young donors, n=6 for aged donors). (B, C, D) Genome-wide CG (B), CH (C) and CA (D) methylation in young and aged donors. Significant differences between aged and young donor groups are labeled (FDR < 0.2 from Wilcoxon rank-sum test and Benjamini–Hochberg correction, n=5 for young donors, n=6 for aged donors). (E) Number of age-DEGs in our study overlapping with age-DEGs from GTEx[70]. (F) Cell type prioritization in aging using Augur[111].

Supplemental Figure 1.5 | Functional enrichment of age-related transcriptome and DNA methylome changes, related to Figures 1.2, 1.3.

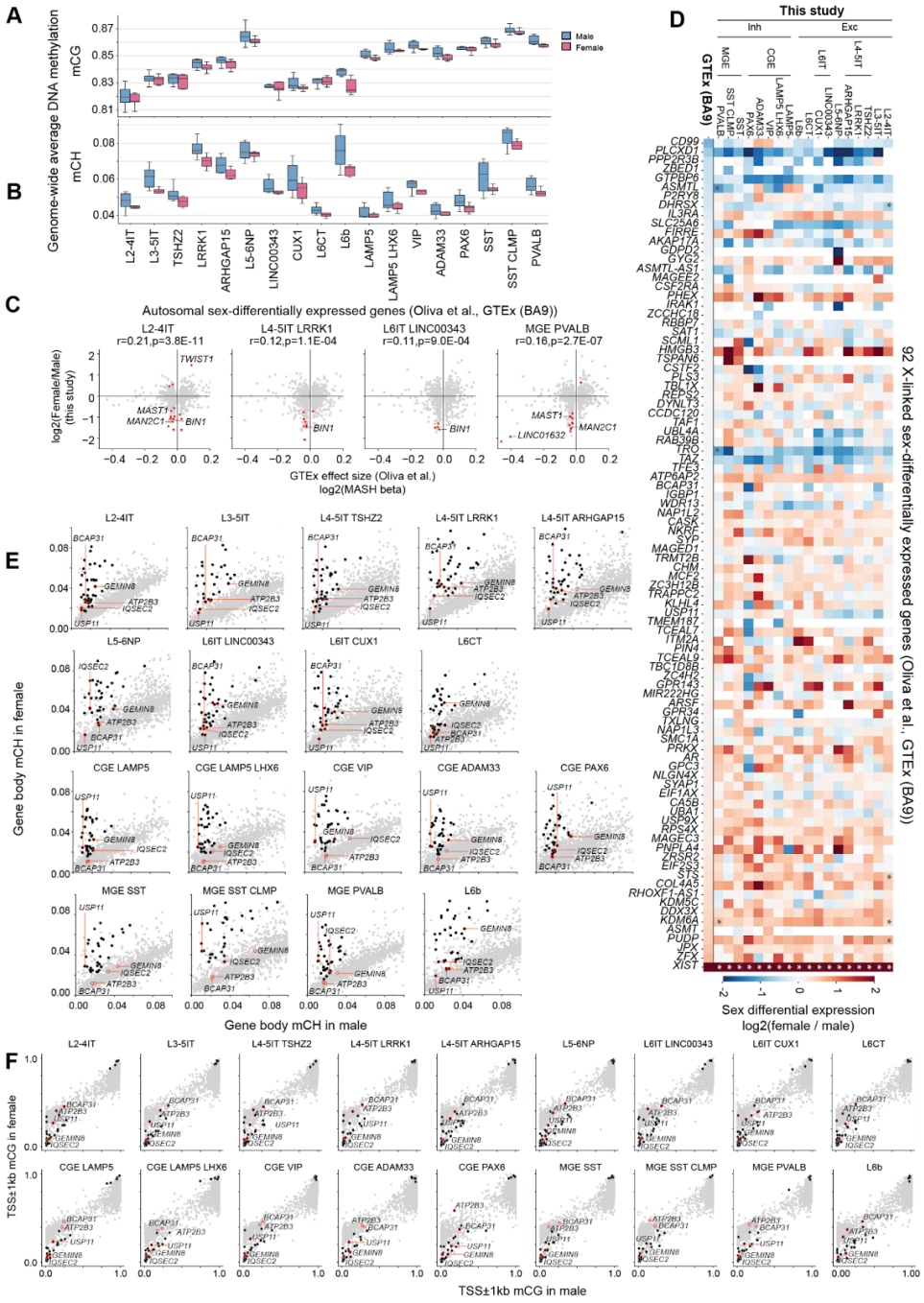
(A,B) Gene ontology (biological process) enriched (FDR <0.001) in all downregulated age-DEGs in (A) excitatory neurons and in (B) inhibitory neurons. (C) Gene ontology (biological process) enriched (FDR <0.01) in all upregulated age-DEGs in inhibitory neurons. (D) Expression of top 4 genes has significant age effects in most number of cell types in aged and young donors. * significantly age-DEGs (FDR < 0.05 and | fold-change | > 1.2). (E) Log2 fold-change of RNA expression in immediate early genes. * FDR <0.1, ** FDR <0.05 (F) Enrichment of age-DMRs in ChromHMM annotated regions. The number of overlapping base pairs was calculated for the set of age-DMRs with each annotation. A baseline was calculated by shuffling the age-DMRs across the genome and overlapping the shuffled DMRs and the annotated regions. The shuffling process was repeated 1,000 times. A normal distribution was fitted for the shuffled results and then compared with the data to get the z-score and p-value. Only significant enrichment/depletion (FDR <0.01) are shown. (G) EGR1 motif enrichment in hypermethylated age-DMRs for all cell types. * FDR <0.05 (H) Motif enrichment of hypomethylated age-DMRs. Only motifs significantly enriched (FDR <0.001) in more than 1 cell type are shown. (I) The relationship between density of age-DMRs and GC content within and outside of subtelomere regions.


A
excitatory neurons
Biological processes enriched in age-DEGs
cell junction assembly
cell-cell adhesion via plasma-membrane adhesion molecules
regulation of synapse structure or activity
modulation of chemical synaptic transmission
regulation of trans-synaptic signaling
glycoprotein biosynthetic process
regulation of transmembrane transport
behavior
neuron recognition
potassium ion transmembrane transport
down-regulated
-log10FDR
B
inhibitory neurons
cell junction assembly
synapse assembly
cell-cell adhesion via plasma-membrane adhesion molecules
axon development
regulation of synapse structure or activity
down-regulated
C
histone H4 acetylation
up-regulated
-log10FDR
D
RNA expression (log2CPM)
NLGN1
PHACTR1
MAST4
KCNH7
Young
Aged
L2-4IT
L3-5IT
TSHZ2
LRRK1
ARHGAP15
L5-6NP
LINC00343
CUX1
L6CT
L6b
LAMP5
LAMP5 LHX6
VIP
ADAM33
PAX6
SST
SST CLMP
PVALB
L4-5IT
L6IT
CGE
MGE
Exc
Inh
E
Cell type
mRNA change log2(Aged-Young)
* FDR<0.1
** FDR<0.05
EGR1
FOS
HOMER1
BDNF
JUN
NR4A1
Immediate early genes
F
age-DMRs (Aged > Young)
age-DMRs (Young > Aged)
FDR<0.01
Z-score from permutation test
TSS (H3K4me3)
tssa
tssaflnk
txflnk
Transcript (H3K36me3)
tx
txwk
Enhancer (H3K4me1)
enhg
enh
Heterochromatin (H3K9me3)
znf
het
Bivalent
tssbiv
bivflnk
enhbiv
Repressive (H3K27me3)
reprpc
reprpcwk
Genomic compartments (chromHMM, human prefrontal cortex)
G
* FDR < 0.05
Cell type
age-DMRs with EGR1 motif (%)
H
Motif enriched (FDR < 0.001) in age-DMRs (Y>A) for more than 1 cell type
GRE
TBR1
BHLHA15
NEUROG2
NEUROD1
ATOH1
NF1-halfsite
SCL
OLIG2
PR
TCF4
TWIST2
FOSL2
JUN-AP1
BATF
FRA1
FOS
JUNB
BACH2
FRA2
Motif enrichment (age-DMRs with Motif / background sequences with Motif)
L2-4IT
L4-5IT TSHZ2
MGE PVALB
I
Num. of age-DMRs per 100kb bin
subtelomere:r= 0.4, p=7E-09
non-subtelomere:r= 0.3, p=4E-82
GC content



Supplemental Figure 1.6 | Negative correlation between the change of mRNA expression and DNA methylation in aging, related to Figure 1.3.

(A) Scatter plots showing significant age-DEGs (colored points) correlated with the age change in gene body mCH for each cell type. Gray points show the union of age-DEGs from all cell types. (B-D) Spearman correlation coefficient between the age change of DNA methylation (aged - young) and the fold-change in gene expression in the union of age-DEGs from all cell types (B), age-DEGs (C), and all expressed genes (D). * FDR <0.01, Benjamini–Hochberg.



Supplemental Figure 1.7 | Sex differences in snmCT-seq data (this study) are consistent with published snRNA-seq and WGBS data, related to Figure 1.5.

(A,B) Genome-wide CG (A) and CH (B) methylation in male and female donors. No significant differences between male and female donor groups. (FDR > 0.2 from Wilcoxon rank-sum test and Benjamini–Hochberg correction, n=6 for male donors, n=5 for female donors). (C) Comparison of the regulation of autosomal RNA expression between females and male in our data (y-axis) and significant sex-DEGs in GTEx (x-axis). Gray dots: significant sex-DEGs in GTEx data. Red dots: overlapped sex-DEGs in GTEx data and our data. (D) Comparison of the regulation of X-linked RNA expression between females and males in our data and significant sex-DEGs in GTEx data[5]. * significantly sex-DEGs in our data. (E, F) Scatter plots of X-linked gene body mCH (E) or promoter mCG (F) in male and female (gray dots). Black dots are the 41 genes shown in (Fig. 1.5G). Putative cell type-specific X-inactivation escape genes are labeled.

Chapter 2 Dynamics of cell-type-specific DNA (hydroxy)methyl-cytosine in human cortical neurons during postnatal development and aging

2.1 Introduction

DNA methylation is a key epigenetic process that has been recognized for its crucial role in regulating gene expression and maintaining cellular identity. DNA methylation typically occurs in two main contexts: CG dinucleotides (CG) and non-CG contexts (CH, where H can be A, T, or C). CG methylation is common across most tissues and cell types. In contrast, CH methylation is prevalent in neurons[34,35] and accumulates during postnatal development[33,90].

A subset of methylcytosine (mC) undergoes further conversion to hydroxymethylcytosine (hmC) via TET enzymes[52]. However, the functional role and lifelong dynamics of hmC in human neurons remain largely unknown. In the brain, the dynamics of mC and hmC are particularly important due to their abundant presence in neurons[34,51,165] and significant accumulation during early development[166–168] in parallel with the genome-wide increase of CH methylation[33,169].

The further oxidation of hmC has been proposed to mediate active demethylation processes, contributing to a closed cycle converting C to mC to hmC and returning to C via one of several potential pathways[170,171]. The dynamics of (h)mC could be critical for cognitive development and neural function[168,172]. Specifically, the interaction between hmC and the methyl-CpG-binding protein MECP2 is critical for understanding the molecular pathology of neurodevelopmental disorders such as Rett syndrome[28,50,173,174]. Additionally, alterations in hmC levels have been associated with the pathogenesis of neurodegenerative diseases, including Alzheimer's disease[140,175,176].

Despite the recognition of DNA methylation and hydroxymethylation as pivotal epigenetic mechanisms in neuronal function, comprehensive studies detailing the dynamics of these modifications in specific neuron subtypes across the lifespan are limited. Although single cell bisulfite sequencing has revealed a highly cell type specific landscape of total

methylation[151,177–179], profiling hmC with single cell resolution has so far been limited to small scale samples[180–183].

To address these scientific gaps, we used whole-genome bisulfite sequencing (BS-seq) and oxidative bisulfite sequencing (OxBS-seq) to distinguish mC and hmC in nuclei of glutamatergic (Glu) and of medial ganglionic-eminence derived GABAergic (MGE-GABA) neurons from the human prefrontal cortex. We collected a large-scale dataset with postmortem samples from 103 donors representing the entire postnatal lifespan in both males and females. Our findings reveal distinct patterns of age-related changes in DNA methylation and gene expression, providing insights into the regulatory mechanisms that influence neuronal function and plasticity throughout the human lifespan.

2.2 Results

2.2.1 Genome-wide conversion of mC to hmC during human postnatal brain development

We collected human prefrontal cortex (Brodmann area BA9) samples from 102 donors ranging from 38 days to 77 years old (Fig. 2.1A). We categorized the donors into six age groups including Infancy (0-1 year old (yo)), Early Childhood (1-5 yo), Late Childhood (5-12 yo), Adolescence (12-20 yo), Early Adulthood (20-50 yo) and Adulthood (50-80 yo) (Fig. 2.1A).

We applied multicolor fluorescence-activated nuclei sorting (FANS) to separate two major neuronal classes using antibodies to label two neuronal markers, NeuN and Sox6. We defined a population consisting mainly of excitatory glutamatergic neurons (Glu, NeuN+/Sox6–) and a medial ganglionic eminence-derived inhibitory GABAergic interneurons (MGE-GABA, NeuN+/Sox6+) population. The MGE-GABA cells include the parvalbumin (PV) and somatostatin (SST) expressing populations[104,184,185]. Note that the Glu population also includes a small proportion of non-MGE-derived GABAergic neurons, such as VIP and LAMP5 expressing cells. We profiled the sorted nuclei using RNA sequencing (RNA-seq) to assess gene expression, whole genome bisulfite sequencing (BS-seq) to quantify total modified cytosine

(total mC = mC + hmC), and oxidative bisulfite sequencing (OxBS-seq) to quantify methylcytosine (mC)[53]. The level of hydroxymethylcytosine (hmC) was calculated as the difference between the total mC and mC. In total, we generated 116 transcriptomes and 338 methylome samples, with more than 80 billion total sequenced reads. On average, we obtained 15-fold genome coverage per sample in BS-seq, and 12-fold coverage per sample in OxBS (Fig. S2.1A).

Overall, the deeply sequenced transcriptomes and DNA methylomes were highly consistent across subjects. There was a high degree of correlation of transcriptome (RNA), gene body total CG methylation level (total mCG), CG methylation level (mCG) and non-CG methylation level (mCH) across the samples (Fig. S2.1B). The Spearman correlation of samples from donors within the same age group (RNA: 0.96 ± 0.00; total mCG in gene bodies: 0.98 ± 0.00; mCG: 0.98 ± 0.00; mCH: 0.96 ± 0.03) was higher than across age groups (RNA: 0.93 ± 0.04; total mCG: 0.97 ± 0.01; mCG: 0.96 ± 0.02; mCH: 0.93 ± 0.04). The correlation between different neuronal classes (Glu vs. MGE-GABA) within each age group was much lower (RNA: 0.88 ± 0.02; total mCG: 0.93 ± 0.01; mCG: 0.92 ± 0.01; mCH: 0.87 ± 0.04). Notably, samples from the Infancy group were less correlated with samples from other age groups and demonstrated less distinction between the neuronal classes (Fig. S2.1C).

Principal component analysis (PCA) of transcriptomes and DNA methylomes showed a clear separation between neuronal classes and ages by the first two PCs. While the first two PCs accounted for more than 50% of the variance in DNA methylation, they explained considerably less variance in RNA expression. Interestingly, the separation between the neuronal classes became more pronounced with increasing age, especially for the methylation marks, suggesting that the distinct molecular identities of excitatory and inhibitory neurons are refined and increasingly differentiated during early childhood (Fig. 2.1B).

Previous studies have reported that hmCG is enriched in neurons, with a higher concentration in GABAergic neurons than Glu cells[104]. Consistent with this, we observed higher

levels of mCG in MGE-GABA (~54%) compared with Glu (~44%) and higher levels of hmCG in Glu (~37%) compared with MGE-GABA (~30%) in early adult samples[104] (Fig. S2.1D). Our data from samples across the lifespan enabled us to analyze the dynamic changes of mC and hmC throughout the genome. We observed a genome-wide decrease of mCG and concomitant increase of hmCG levels across the lifespan, consistent with a gradual conversion of methylcytosine to hydroxymethylcytosine. A larger fraction of mCG was converted to hmCG in Glu (25%) compared to MGE-GABA neurons (21%) (Fig. 2.1C, S2.1D). In contrast, the total mCG level (mCG + hmCG) changed by less than 3% in both cell types, suggesting that loss of mCG is largely balanced by gain of hmCG.

Neurons accumulate substantial non-CG methylation during childhood[33,9033,98]. We observed an accumulation of non-CG methylation (mCH) levels during development from 1% to 6%. By contrast, non-CG hydroxymethylation (hmCH)remained low at less than 2% (Fig. 2.1C, S2.1D). The genome-wide methylation level changes most rapidly in the first 20 years of life. The change in genome-wide methylation levels can be fitted well by an exponential model (R-squared = 0.84~0.89 in mCG and hmCG for both neuronal classes), with the fitted time constants representing the time scale of methylation conversion (mCG: 4.9 ± 0.7 years Glu, 11.9 ± 1.4 years MGE-GABA; hmCG: 4.4 ± 0.5 years Glu, 6.4 ± 0.9 years MGE-GABA).

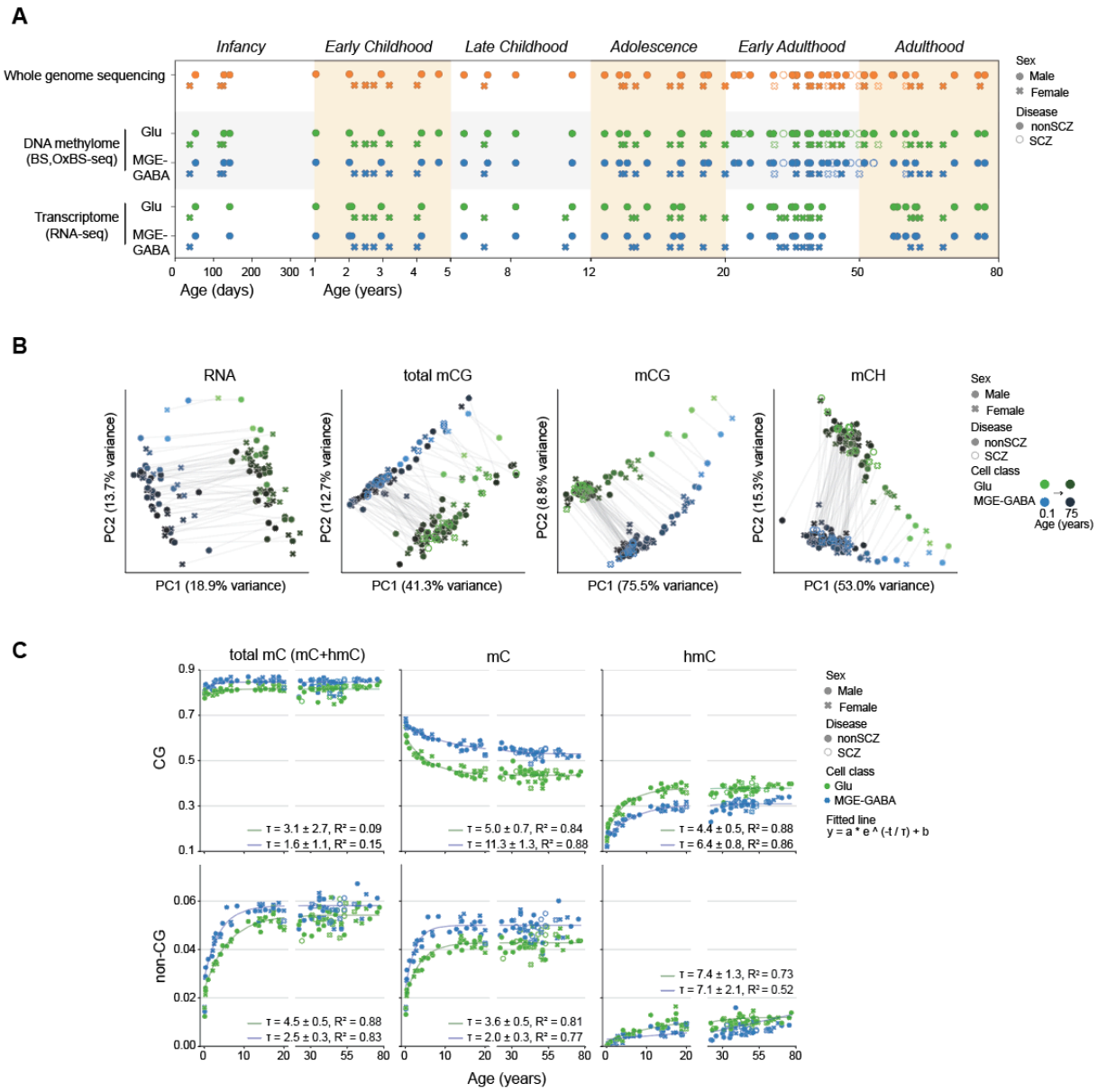


Figure 2.1 | Mapping dynamic changes of genome-wide DNA (hydroxy)methylation and gene expression in two major neuronal subtypes across lifespan.
(A) Number of samples in this dataset. (B) Principal component analysis (PCA) of gene expression, gene body total mCG, mCG, and mCH levels across age groups and neuronal classes. (C) Genome-wide methylation levels across age. τ is the time constant fitted from the data. The time constant fitted from the data and the standard deviation are labeled on each panel. SCZ: schizophrenia.

2.2.2 Age-related changes in gene expression across human lifespan

We examined the transcriptomic consequences of postnatal development and aging by analyzing differentially expressed genes between age groups (age-DEGs). We found 8,109 age-DEGs (FDR < 0.05, Table S2.2) in at least one pairwise comparison between the age groups in at least one neuronal class (Fig. S2.2A).

The aging changes were highly consistent with single nuclei human brain mRNA measurements from the PsychENCODE consortium and other studies with up to n=388 donors (Pearson correlation of log-fold change, r = 0.62 between Glu and L2/3 IT with more than 3300 age-DEGs genes overlapped; r=0.69 between GABA and SST (>1600 genes), Fig. S2.2B, C)[14,84]. The age-DEGs represent 58% of all expressed genes in both neuronal classes, suggesting that childhood brain development and aging have a substantial impact across most of the neuronal transcriptome.

We found more age-DEGs in Glu neurons (6,808 unique genes) compared with MGE-GABA neurons (4,929 unique genes). Of these, 1,981 (24%) were differentially expressed between early childhood and infancy (Fig. S2.2A). We observed a median of 1.3-fold more genes being downregulated compared to upregulated with age across all pairwise comparisons, consistent with a progressive refinement of the neuronal gene expression program.

We clustered the age-DEGs into five groups based on their pattern of dynamic regulation: early downregulated (Early-Down, Glu/MGE-GABA: 782/537 genes), late downregulated (Late-Down, 2,692/2,065), early upregulated (Early-Up, 509/289), late upregulated (Late-Up, 2,597/1,985), and non-monotonic (Non-Mono, 228/53) (Fig. 2.2A). Early downregulated genes, whose expression declines starting in infancy (1 year old), were enriched in pathways crucial for neuronal development, such as protein synthesis, cell adhesion, and axon guidance[118,186]. For example, *DRAXIN*, an axon guidance gene that is highly expressed in the developing brain[187,188], was downregulated by up to 4-fold in early childhood compared to infancy in both neuronal classes (Fig. 2.2A). In contrast, late downregulated genes, whose

expression decreases starting in childhood or later, were enriched in more specific pathways such as regulation of trans-synaptic signaling, synapse organization, and telencephalon development, which are essential for advanced neural communication and establishing neural circuits (Fig. 2.2B). For example, *KCNG1,* which encodes voltage-gated potassium channels, was 1.9 / 4.7 -fold decreased in late childhood compared to infancy in Glu / MGE-GABA neurons (Fig. 2.2A). This suggests a shift towards more specialized neural activities and cognitive functions[83,189]. Notably, a group of activity-dependent genes, known as the late response genes, such as *IGF1* and *PNOC*, were significantly enriched in the late downregulated group (Fig. S2.2D, E). This finding highlights the critical role of late downregulated genes in synaptic plasticity[190].

In contrast with the downregulated genes, we found fewer upregulated genes during postnatal development. The upregulated genes did not have a clear pattern of functional enrichment (Fig. 2.2B). This suggests that as neurons age, the expression of genes tightly regulated during early development collectively decreases, while genes related to more mature, cell-type-specific functions may increase[179].

The vast majority of DE genes had monotonic patterns of increased or decreased expression across the lifespan. A small but notable group of non-monotonic genes (228 genes in Glu and 53 genes in MGE-GABA) exhibited significant upregulation in some age group comparisons and downregulation in others. For example, the gene *MICAL2*, which is involved in cytoskeletal dynamics and organ development[191], increased 0.48-fold from infancy to adolescence, but then decreased 0.42-fold from adolescence to adulthood in MGE-GABA. Another example is *ARHGAP10*, which is involved in small GTPase signaling[192], increased 1.1-fold from infancy to early childhood, but then decreased 3.1-fold from early childhood to adulthood in Glu (Fig. 2.2A, C).

The majority (92%) of the age-DEGs present in both neuronal classes were regulated in the same direction in Glu and MGE-GABA cells. We observed a subset (58 genes) that had

age-dependent monotonic regulation in opposite directions between Glu and MGE-GABA neurons. For instance, genes related to synaptic signaling, such as *SLC6A7*[193], *CADPS2*[194],*OLFM3*[195] , *KCNS2*[196] and *KCNIP4*[197], had similar expression levels in both Glu and MGE-GABA neurons during infancy, but with age, they were upregulated in Glu neurons and downregulated in MGE-GABA neurons . Conversely, the G protein-coupled receptor gene *GPRC5B*[198] shows an opposite trend: it was upregulated with age in MGE-GABA neurons and downregulated in Glu neurons. *CPNE6,* a brain-specific gene associated with synaptic plasticity[199]*,* was more strongly expressed in Glu neurons during early development, but after adolescence it showed a stronger expression in MGE-GABA neurons (Fig. 2.2C).

2.2.3 Expression of age-DEGs correlates with reduced gene body mCG and increased hmCG

Although a negative association between DNA methylation and gene expression is widely observed, it is not clear to what extent changes in mC during development and aging relate to corresponding changes in gene expression. We found that genes whose expression increases from early childhood to adolescence exhibit a more pronounced conversion of mCG to hmCG compared with genes that are downregulated in this period (Fig. S2.3A), consistent with previous findings that hmCG is enriched in highly expressed genes[50]. Genes whose expression decreases with age gain substantially more mCH in their gene bodies. Whereas changes in mCG and mCH were observed not only within gene bodies but also upstream of the transcription start sites (TSS), the differences in hmCG were primarily observed within the gene bodies (Fig. S2.3A).

We directly correlated the RNA expression fold change with the DNA methylation fold change in 1kb bins within the flanking regions of TSS of age-DEGs across all pairwise comparisons. We observed a strong negative correlation for mCG and a strong positive

correlation for hmCG in gene bodies, while there was a strong negative correlation for mCH upstream of TSS regions (Fig. 2.2D, S2.3B, C, D).

Next, for each gene, we developed a regularized regression model using LASSO to predict gene expression levels based on the methylation features (mCG, hmCG, and mCH) across different genomic regions: promoter (TSS±1k), gene body, and genome-wide methylation. Features were standardized first, and then 5-fold cross-validation was used to conduct a grid search for the best penalty parameter (10^{-4} to 1). The average cross-validated (test set) correlation between predicted and actual gene expression for each donor is 0.66, indicating that while methylation and hydroxymethylation dynamics are crucial in regulating age-related gene expression changes, other biological mechanisms, such as activating and repressive histone modifications, may also be involved[37,104].

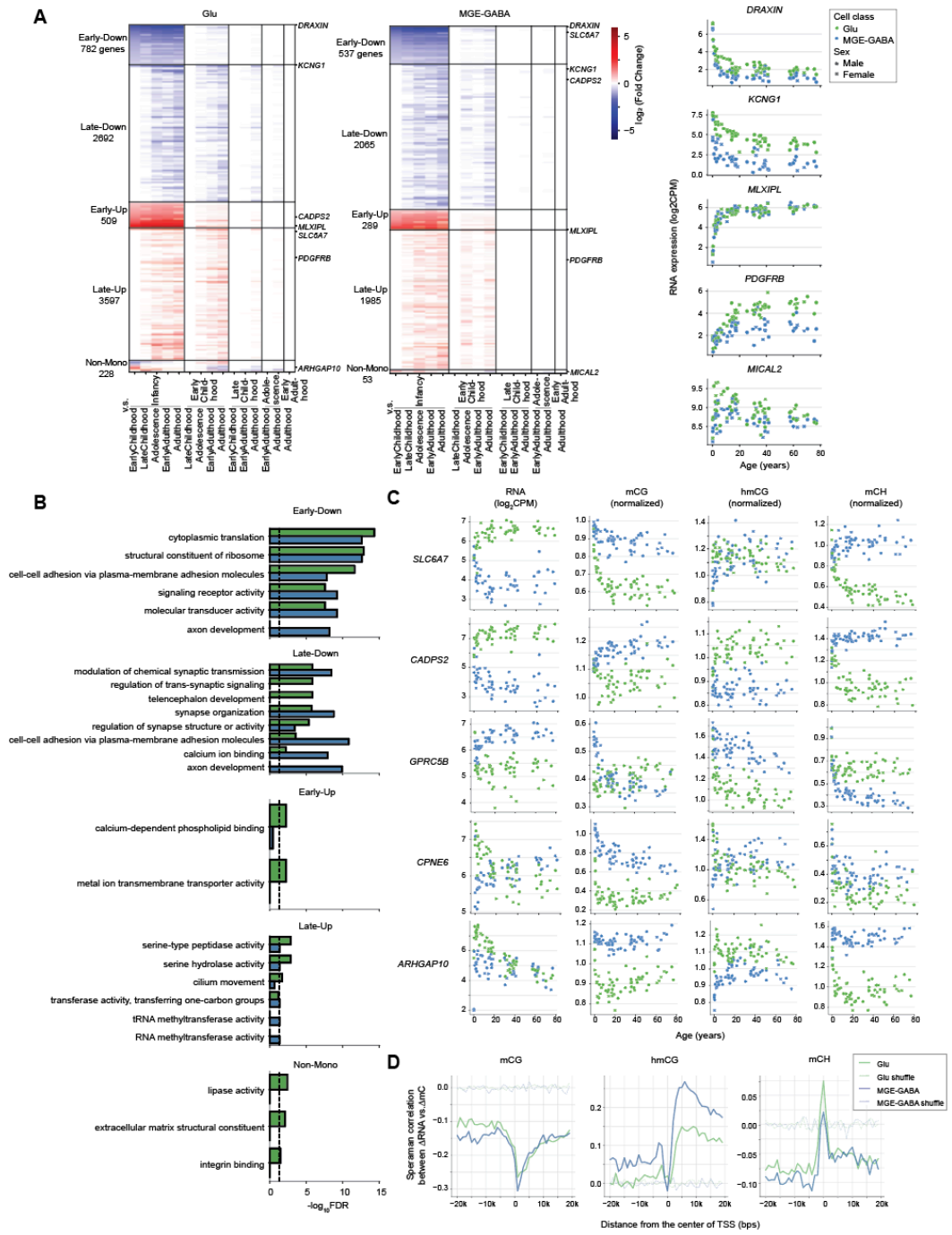


Figure 2.2 | Correlated transcriptome and DNA methylome changes in the developing human brain.

(A) Five groups of age-DEGs based on their pattern of dynamic regulation: early downregulated (Early-Down), late downregulated (Late-Down), early upregulated (Early-Up), late upregulated (Late-Up), and non-monotonic (Non-Mono). (B) Gene ontology of age-DEGs from the five groups. Top five significant (FDR<0.05) terms in biological pathway and molecular function were shown. (C) Example genes with opposite expression trends with age in the two neuronal classes. Normalized methylation was calculated as the gene body methylation level divided by the genome-wide methylation level of the sample. (D) Spearman correlation between RNA expression fold change and DNA methylation fold change within the flanking regions of TSS of age-DEGs across all pairwise comparisons (solid line). Dashed lines represent the correlation between shuffled fold changes of RNA and DNA methylation from the age-DEGs.

2.2.4 Dynamics of DNA methylation and age-related differentially methylated regions

Different genomic regions regulate various functions and thus exhibit distinct methylation dynamics. Generally, methylation levels are inversely related to genomic activity. For example, transcription start sites (TSS) typically have low levels of mCG, hmCG, and mCH, whereas gene bodies show intermediate levels of these methylation marks that correlate with the gene's expression level. Repressive regions, such as L1HS retrotransposon-derived sequences, are characterized by high mCG, low hmCG, and low mCH (Fig. S2.4A).

We examined DNA methylation valleys (DMVs), defined as extended regions (≥5 kb) with low average mCG (<15%), which play a crucial role in development[49,128,200]. These regions are typically marked by the histone modification H3K27me3, which is associated with polycomb repressive complex 2 (PRC2). Using total mCG data, we identified 500-700 DMVs which were largely conserved across all age groups for each neuronal class (Fig. S2.4B). Both the number and length of DMVs decreased with age, consistent with previous findings in human and mice neurons[90,121]. DMVs exhibited low levels of mCG, hmCG, and mCH, with minimal changes over the lifespan (2% of mCG, 3% of hmCG, and 0.5% of mCH, Fig. S2.4A). Given the role of DMVs in development, we hypothesized that the expression of genes within DMVs would change with age. Indeed, we detected an enrichment of downregulated age-DEGs in DMVs ($p<0.001$, Fisher's exact test, Fig. S2.4C), with the mean expression of these genes decreasing with age across all levels of gene expression (Fig. S2.4D).

To detect genome-wide DNA methylation variations with age, we fit linear regression models to 10kb bins across the genome. We detected significant (FDR<0.01) changes in total mCG, mCG, or hmCG in more than 99% of the genome. Among these bins, 90% showed a loss of mCG and a gain of hmCG with age, while total mCG either increased or remained constant. Additionally, we identified a small number of bins that lost total mCG, mCG, and hmCG (n=102 in Glu, 24 in MGE-GABA) (Fig. S2.4E). Bins losing total mCG were enriched in active genome regions, such as transcription start sites (TSS) and enhancer regions (Fig. S2.4F).

We further detected age-related differentially methylated regions (age-DMRs) by pairwise age group comparisons on autosomes. In total, we detected more than 280,000 age-DMRs where total mCG changes with age,covering 8.5% of the genome in Glu and 4.7% in MGE-GABA. Using our OxBS-seq data, we found 720,000 mCG age-DMRs covering 26% of the genome in Glu and 18% in MGE-GABA (Fig. S2.5A). These data show that changes in mCG are more widespread than the total methylation changes observed with BS-seq. We clustered age-DMRs called from total mCG into five groups based on mCG and hmCG dynamics: late hypomethylation (Late-Hypo), early hypomethylation (Early-Hypo), late hypermethylation (Late-Hyper), early hypermethylation (Early-Hyper), and de-methylation (De-M) (Fig. 2.3A, S2.5B).

Most DMRs exhibited an increase in total mCG and hmCG with age. However, a subset of DMRs (8.5% in Glu and 4.5% in MGE-GABA) showed a loss of both mCG and hmCG (Fig. 2.3C). Those de-methylation DMRs had higher hmCG and lower mCH in infancy compared with other age-DMRs (Fig. 2.3B, S2.5C) and were enriched in enhancers mapped by chromHMM using multiple ChIP-seq histone modification data obtained in the human prefrontal cortex[122] (Fig. 2.3D, S2.5D), as well as neuronal activity-regulated transcription factor motifs identified by HOMER[123] (Fig. 2.3E, S2.5E). Examples of de-methylation DMRs in both neuronal classes, Glu-specific and MGE-GABA specific are shown in Fig. 2.3F.

To quantify the dynamics of DNA methylation conversion in each DMR, we developed a dynamical model to estimate the time scales of DNMT3A-dependent methylation of cytosine (C) to methylcytosine (mC), TET-dependent oxidation of mC to hydroxymethylcytosine (hmC), and the subsequent demethylation (hmC to C) (see Methods). Our model reveals that the de-methylation group of DMRs exhibits a longer time scale (~10 years) for converting C to mC compared to the time scales for mC to hmC and hmC to C, as well as compared to other DMR groups. (Fig. S2.5F).

Interestingly, we identified age-DMRs that were oppositely regulated with age between Glu and MGE-GABA neurons. Specifically, we found 268 DMRs where total mCG decreased in Glu but increased in MGE-GABA, and 51 DMRs where total mCG increased in Glu but decreased in MGE-GABA. Genes near these DMRs had similar RNA expression levels between the neuronal classes in infancy but showed opposite expression trends later in life (Fig. S2.5G).

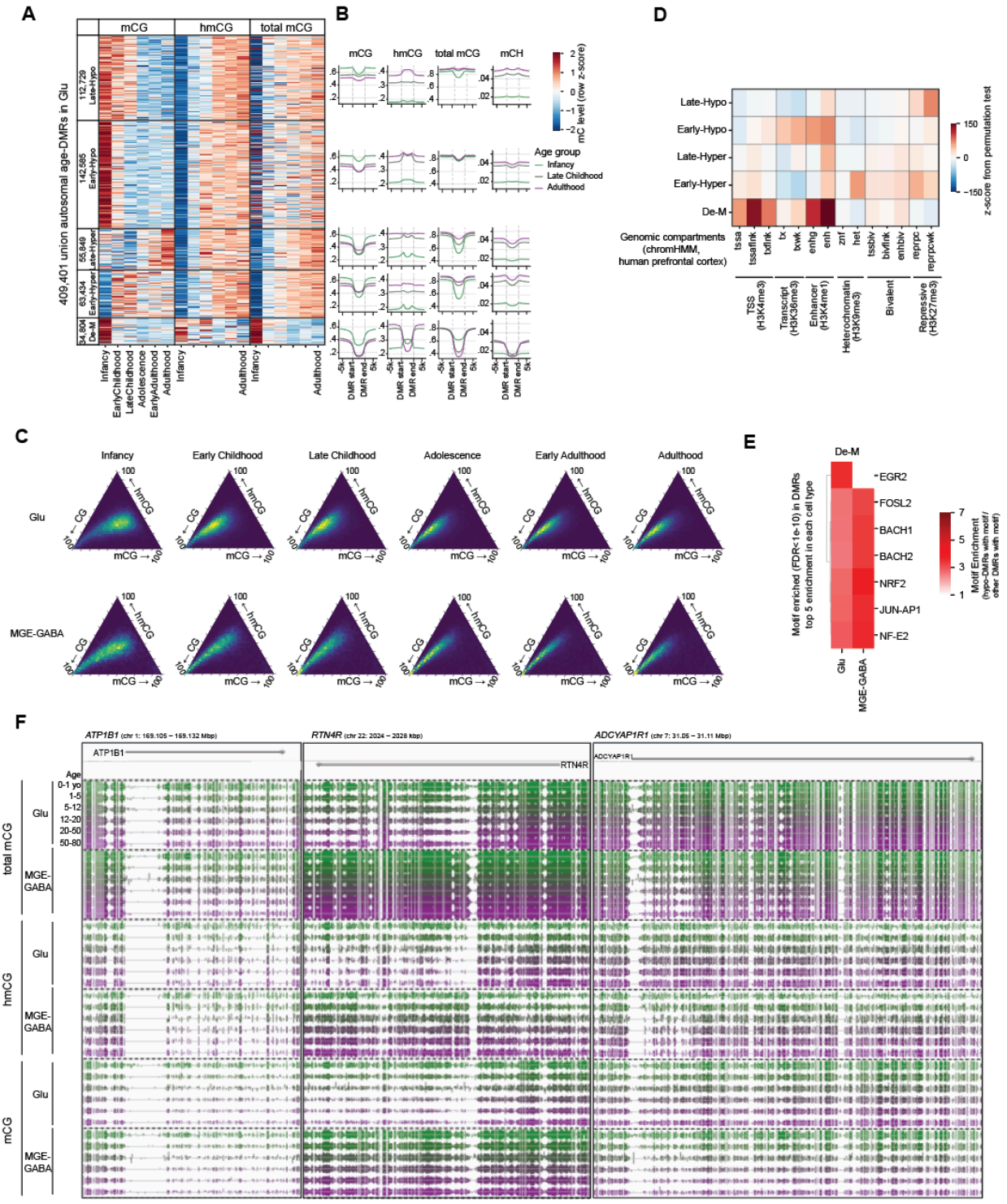


Figure 2.3 | Dynamics of DNA methylation.
(A) Five clusters of age-DMRs based on mCG and hmCG dynamics in Glu. (B) Average methylation levels of different age-DMR clusters in infancy, late childhood and adulthood in Glu. (C) Ternary plot illustrating the swift of DMRs in the de-methylation group. (D) Enrichment of Glu age-DMRs in ChromHMM-defined chromatin states[122]. (E) Transcription factors motif enrichment in the de-methylation DMRs by HOMER[123]. (F) Examples of common and cell-type-specific de-methylation DMRs.

2.2.5 Strand asymmetry of hmCG increases with age for expressed genes

Previous studies reported hmCG enrichment on the sense strand and mCG enrichment on the antisense strand within gene bodies of highly expressed genes in the adult brain[165]. We categorized all expressed genes into 10 groups according to their average expression levels across all samples. Consistent with previous findings, we observed strand-specific asymmetry in adult samples of both neuronal classes in highly expressed genes, with this asymmetry confined to the gene bodies and not extending into the flanking regions (Fig. S2.5H).

Furthermore, for many genes we observed age-dependent changes in hmCG asymmetry between the sense and antisense strands. Specifically, a considerable number of genes (1,235 in Glu and 821 in MGE-GABA) exhibit a significant (FDR <0.05) correlation between hmCG strand differences and age (Fig. S2.5I). Notably, the majority of these genes (1,173 in Glu and 760 in MGE-GABA) demonstrated a positive correlation, indicating an increasing enrichment of hmCG on the sense strand with age. Furthermore, 89% of these genes displayed high expression levels in both neuronal subtypes, implying that strand asymmetry is likely influenced by transcriptional activity. Notably, we did not observe any age-dependent changes in total mCG. These findings extend recent observations of clock-like changes in DNA methylation across the lifespan. Our data show that hmCG strand asymmetry provides a remarkably linear predictor of age with no fitting required.

2.2.6 Sex differential gene expression with age

We detected 13 X-linked and one autosomal sex-specific differential expression genes (sex-DEGs, Fig. 2.4A, Table S2.3). Three of the male-biased genes, *ASMTL*, *GTPBP6,* and *PLCXD1*, are located in pseudoautosomal regions, consistent with findings by other studies[201,202]. Additionally, *GTPBP6* shows increased expression with age in both males and females.

Notably, we did not detect significant expression differences between males and females in sex hormone receptor genes (androgen receptor *AR*, estrogen receptor *ESR1* and *ESR2*)[203,204]. However, we observed cell-type-specific age-related changes for some of these genes. *AR* was upregulated with age in MGE-GABA neurons, but downregulated in Glu neurons in both sexes. *ESR1* was lowly expressed in both neuronal classes during infancy but increases in Glu neurons during early childhood and then gradually decreases with age (Fig. S2.6A).

2.2.7 Sex-specific DNA methylation patterns and X-inactivation escape genes

It has been established that females exhibit lower total mCG levels on the X chromosome compared to males due to X-inactivation in one of the X chromosome in females, while autosome methylation levels are similar between sexes[57,205]. Our studies further extend this knowledge by separately examining mCG and hmCG.

We found that the higher total methylation (mCG+hmCG) levels in males on the X chromosome are primarily due to increased hmCG, whereas mCG levels remain similar between male and female neurons (Fig. 2.4B). Additionally, the X chromosome mCH levels were higher in males compared to females.

We further compared the methylation features between inactivated and escape genes on the X chromosome. Our data showed that genes which escape X-inactivation exhibited higher gene body mCH levels, whereas inactivated genes displayed higher promoter total mCG levels, which is consistent with previous studies[33,57] (Fig. 2.4C). By distinguishing between mCG and hmCG, we discovered that the higher promoter total mCG in inactivated genes was driven by mCG rather than hmCG (Fig. 2.4C). Furthermore, we observed that less hmCG accumulates with age near inactivated genes, indicating that hmCG fails to accumulate in heterochromatic regions. This finding suggests that the TET enzyme, which converts mC to hmC, might be less effective or excluded in these densely packed chromatin areas, reinforcing the silenced state of inactivated genes. The X chromosome DNA methylation differences between males and

females, as well as the differences between inactivated and escape genes, begin in infancy, persist throughout the lifespan, and are consistent in both neuronal classes (Fig. S2.6B, C).

2.2.8 Hydroxymethylation with age in imprinting control regions

Hydroxymethylation has been shown to play an important role in imprinting during fetal development[206–208]. Non-CG methylation was also detected in mouse brain imprinting regions. Loss of imprinting is often observed in various human tumors[209–211]. However, there is still an ongoing debate about whether imprints are erased in disease and aging[212,213].

We analyzed hydroxymethylation within a well-recognized imprinting control region, the 15q11 segment[214]. We observed a lower level of methylation in both neuronal classes in the imprinting region (Fig. 2.4D, S2.7A). Notably, the change of DNA methylation level with age is smaller in imprinting regions than non-imprinting regions (two-sample Kolmogorov-Smirnov test, $p<0.01$). We detected age-DEGs of the known imprinting genes in those regions: paternally expressed genes such as *MKRN3* and *NDN*, along with the maternally expressed gene *UBE3A*, showed a decrease in expression as age increased. Conversely, the expression of the paternally expressed gene *SNURF-SNRPN* was found to increase with age.

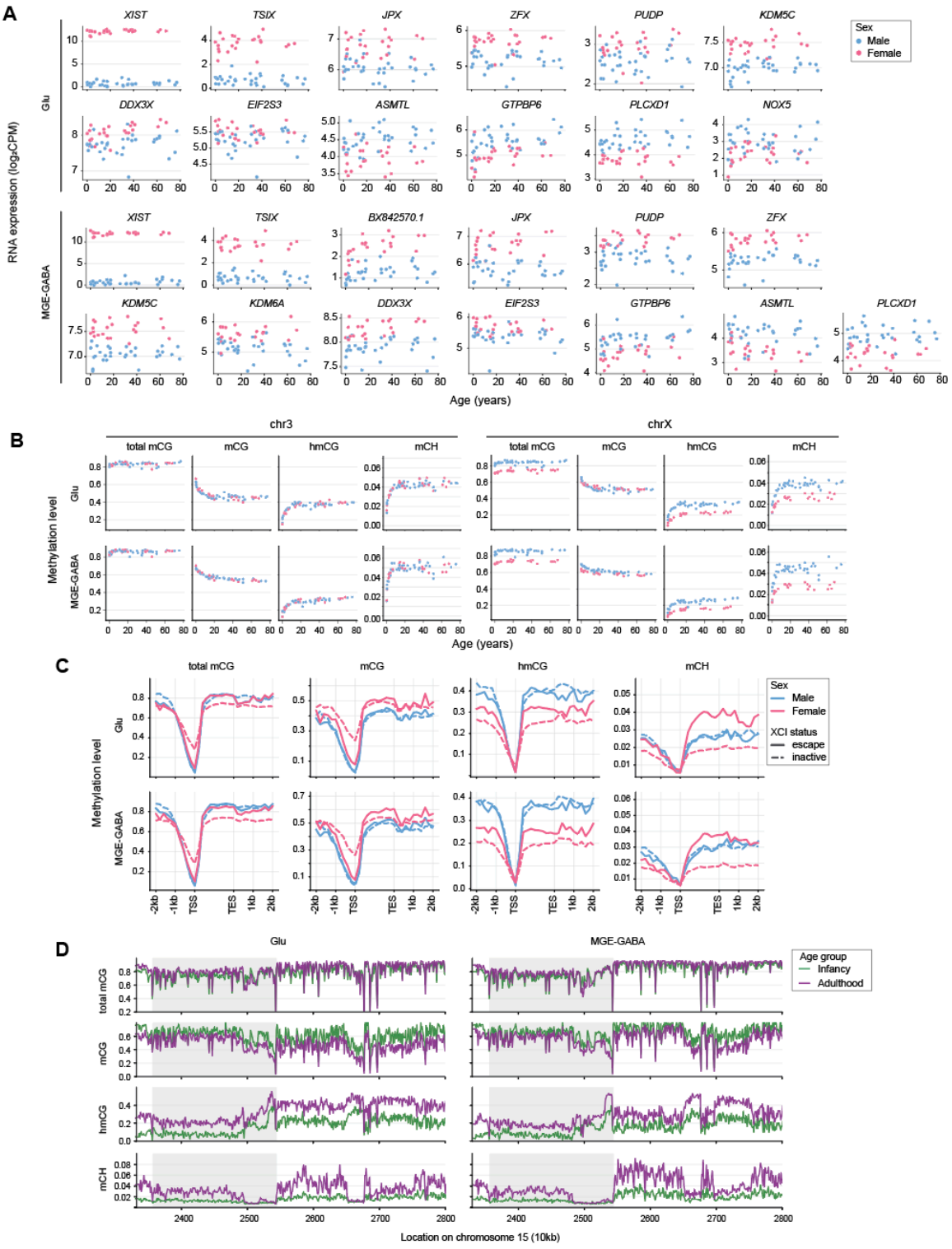


Figure 2.4 | Sex differences in DNA methylation and gene expression in human neurons.
(A) Expression changes of sex-specific differential expression genes (sex-DEGs) with age. (B) Comparison of mCG and hmCG levels on the X chromosome between males and females. (C) Methylation features of inactivated and escaped genes on the X chromosome. Data shown in early adulthood. XCI: X chromosome inactivation. (D) DNA methylation in known imprinting control region 15q11 (shaded region).

2.3 Discussion

In this study, we employed neuronal cell type specific purification (FANS) and large-scale transcriptome and DNA (hydroxy)methylome sequencing to profile the dynamics of genome-wide DNA methylation (mC) and hydroxymethylation (hmC) in glutamatergic (Glu) and MGE-derived GABAergic (MGE-GABA) neurons across the human lifespan. Our comprehensive analysis reveals distinct patterns of age-related changes in DNA methylation marks and gene expression, with implications for neuronal development and synaptic plasticity.

The observed genome-wide decrease in mCG levels coupled with an increase in hmCG levels with age, while total mCG levels remain constant. This indicates an ongoing conversion of mCG to hmCG throughout the period of early to late childhood. This transformation primarily occurs before the age of 20, with a prolonged conversion timescale of 5-10 years, suggesting a significant regulatory role for DNA methylation conversion in human brain development. We developed a simple 3-state model for the cyclic dynamic of C, mC and hmC, which captures the complex trajectory of methylation across the lifespan. These changes are more pronounced in Glu neurons compared to MGE-GABA neurons, highlighting the importance of cell-type specificity in brain development.

Consistently, we identified a substantial number of age-related differentially expressed genes (age-DEGs), with more age-DEGs found in Glu neurons compared to MGE-GABA neurons. This disparity may reflect the higher synaptic activity and plasticity associated with excitatory neurons. Clustering of age-DEGs into distinct expression patterns revealed a shift from genes involved in early neural development to those associated with advanced synaptic signaling and brain function as neurons age. The observation of genes exhibiting age-dependent regulation in opposite directions between Glu and MGE-GABA neurons again underscores the functional diversity of neural circuits during aging.

Our study found a significant inverse correlation between gene expression changes and mCG levels[215], and a positive correlation with hmCG levels within gene bodies. These findings

are consistent with the role of hydroxymethylation in promoting gene expression. The stronger correlation observed in gene body regions compared to promoter regions suggests that gene body methylation may play a more prominent role in regulating age-related gene expression changes.

We detected age-related differentially methylated regions (age-DMRs) across the genome, exhibiting distinct patterns of hypo- and hyper- methylation. Hypomethylated regions, characterized by a decrease in total mCG over time, were enriched in enhancer regions and neuronal activity-regulated transcription factors motifs. Additionally, some age-DMRs exhibited opposing methylation trends between Glu and MGE-GABA neurons, emphasizing the complex interplay between different neuronal subtypes and their unique epigenetic landscapes. Furthermore, by fitting our developed dynamical model, we can more precisely and quantitatively understand the time scales of the conversion between C, mC, and hmC, as well as the variation in these time scales across the genome.

The identification of differences in mCG and hmCG patterns between males and females provides novel insights into the sex-dependent regulation of the neuronal epigenome. We observed lower levels of total mCG on the X chromosome in females compared to males, primarily due to decreased hmCG but not mCG in females. Conversely, the higher promoter total mCG levels in inactivated genes in females were attributed to mCG, not hmCG. These findings support the notion that TET enzymes may be less effective in heterochromatic regions, thereby limiting the conversion of mCG to hmCG and reinforcing the silenced state of these genes.

In conclusion, our study provides valuable insights into the dynamics of DNA methylation and hydroxymethylation in human neurons across the lifespan. The distinct patterns of age-related changes in Glu and MGE-GABA neurons underscore the importance of cell-type-specific epigenetic regulation in maintaining neuronal development and aging. Our findings of the regulation role of mC and hmC have important implications for future research on

neurodevelopmental and neurodegenerative diseases for developing potential therapeutic strategies targeting epigenetic modifications.

2.4 Methods

2.4.1 Human brain samples

Human postmortem brain tissues were sourced from the dorsolateral prefrontal cortex (Broadman area 9). The majority of the samples were obtained from the NIH NeuroBioBank. Twenty nine adult samples were obtained from the New York State Psychiatric Institute Macedonian Brain Collection. Specific characteristics of the individuals are displayed in Table S2.1.

2.4.2 Nuclei isolation

The protocol and reagents for brain nuclei isolation before flow cytometry were previously described[104,184]. Briefly, tissue was homogenized in lysis buffer, underlaid with high sucrose buffer, and centrifuged for 1 hour. The nuclear pellets were resuspended in antibody-incubation buffer, incubated with primary antibodies for 1.5 hours, followed by a second centrifugation and a 1-hour incubation with secondary antibodies. We used FANS, employing NeuN (RBFOX3) as a marker for neuronal nuclei and SOX6, a transcription factor in MGE-derived GABA neurons, to separate nuclei of MGE-GABA and Glu neurons. We used mouse monoclonal anti-NeuN phycoerythrin (PE)–conjugated antibodies (Millipore, FCMAB317PE, 1:1000), donkey anti-guinea pig AX647-conjugated secondary antibodies (Jackson ImmunoResearch, 1:1500) to detect SOX6-positive nuclei.

2.4.3 RNA isolation and RNA-seq library construction

Nuclei of cell populations were prepared using the FANS-based method as previously described[104,184], with modifications: proteinase inhibitors were omitted, and RNase inhibitors (1:40 during lysis, 1:100 during other steps) were added. Tissue (100-200 mg) was homogenized, and ~40,000 nuclei per sample were used for RNA extraction with the PicoPure RNA Isolation Kit (Thermo Fisher). RNA-seq libraries were prepared from 10 ng of RNA using

the SMARTer Stranded Total RNA-Seq Kit (Clontech) and sequenced on a HiSeq 2500 to an average of ~50 million read pairs per sample.

2.4.4 DNA methylation assay

Between 350,000 and 700,000 nuclei of Glu and MGE-GABA neurons were isolated from each tissue sample by flow cytometry. High-quality genomic DNA (average fragment length >20,000 bp) was isolated using the DNeasy Blood and Tissue Kit (Qiagen) and concentrated with a Genomic DNA Cleanup Kit (Zymo). Genomic DNA (500 ng) was used for BS/OxBS processing and library preparation with the TrueMethyl Whole Genome kit (CEGX). Libraries were sequenced on an Illumina HiSeq 2500 system using a paired-end 100-cycle protocol. Conversion performance in oxidation and bisulfite treatments was assessed using CEGX spike-in control oligonucleotides.

2.4.5 Mapping RNA data

RNA reads were trimmed using Trim-Glalore (0.6.4_dev) for quality trimming and adaptor trimming. The first 3 bps from the 5' end of read1 is further removed. Trimmed reads were then mapped to hg38 and GENCODE gene annotation (v37) using STAR (v2.7.3a) with the following parameters consistent with ENCODE standard (--outFilterType BySJout --outFilterMultimapNmax 20 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --sjdbScore 1). Gene count was estimated by RSEM(v1.3.1) with the following parameters (--estimate-rspd --paired-end --forward-prob 1).

2.4.6 Mapping DNA methylation data

Whole genome (oxidative)bisulfite sequence was mapped to hg38, C to T converted genome index in paired-end by Hisat-3N (v2.2.1-3n-0.0.3) after adaptor trimming by trim_galore

(v0.6.4_dev). Only unique mapped reads were kept for further analysis. Duplicated reads were removed by Picard (v2.27.1). Number of methylated and unmethylated reads covered for each cytosine site was called by methylpy (v1.4.4). For each donor, cytosines and adjacent downstream (3') sites which overlap homozygous SNPs called from any donor, were removed for the downstream methylation analysis.

2.4.7 Differential expression and gene ontology

Protein coding and lncRNA genes from chromosomes 1-22 and X with counts per million (CPM) >7 in at least 3 samples are defined as "expressed genes". Age differential expression testing was performed separately for each cell type and each comparison using Dream[162] (variancePartition v1.28.9, edgeR v3.40.2). Specifically, calcNormFactors was used to control for the raw library size from the raw counts. Then, voomWithDreamWeights and dream were used to prepare for the linear mixed model fit from the normalized count and design matrix. The model included fixed effects of age group, batch, PMI hours, RIN,race and sex: $expression \sim Age\ Group\ +\ batch\ +\ PMI\ hours\ +\ RIN\ +\ Race\ +\ Sex$. Finally, after testing the generalized linear fixed model for each gene, genes with an adjusted p-value < 0.05 in any pairwise age group comparison are defined as age-related differentially expressed genes (age-DEGs) in our analysis.

For genes exhibiting monotonic changes, we compared the fold change between early childhood and infancy to the fold change between adulthood and early childhood. Those genes are resulted in four categories: early downregulated genes, late downregulated genes, early upregulated genes, and late upregulated genes. Additionally, we identified genes showing non-monotonic change patterns.

Functional gene ontology (GO) enrichment analysis was performed using clusterProfiler (v 3.14.3)[163]. All expressed genes in each cell type were included as the background gene list. Gene ontology with gene size between 25 and 500 were included in the enrichment test. The

Benjamini-Hochberg method was used to control the false discovery rate. Results from biological pathway (BP) and molecular function (MF) are shown.

2.4.8 Differential methylation analysis

We detect differentially methylated regions (DMRs) using DSS[109,110] (DSS v2.34.0, bsseq v1.22.0). DMLfit.multiFactor was applied with model formula: $mCG \sim age\ group + sex$, with a smoothing window size 500 bp. Next, DMLtest.multiFactor was used to call differentially methylated loci. Finally, callDMR was used to merge significantly differentially methylated loci ($p < 0.05$) with a distance smaller than 50 bp. Merged regions with a length greater than 50 bp and containing at least 3 CpG sites are called differentially methylated regions (DMRs). Only control donors were included in the DMRs analysis.

2.4.9 DNA methylation valleys (DMVs)

To call DNA methylation valleys (DMVs), we first identified undermethylated regions (UMRs) and low methylated regions (LMRs) using MethylSeekR[164]. We used methylation cutoff m=0.5, minimum number of CpG, n=10, and FDR < 0.05 as parameters for detecting UMRs and LMRs. DMVs were defined as UMRs with length ≥5 kb and mean methylation level ≤15%.

2.4.10 DNA methylation conversion dynamical model

To estimate the time scale of the conversion between C, mC, and hmC, we used coupled ordinary differential equations (ODEs), assuming the time scale is constant across the lifespan. The time constants for C, mC, and hmC, as well as the initial values of mC and hmC, were fitted to the methylation levels within each DMR from all donors using *optimize.curve_fit*. The ODEs used are as follows:

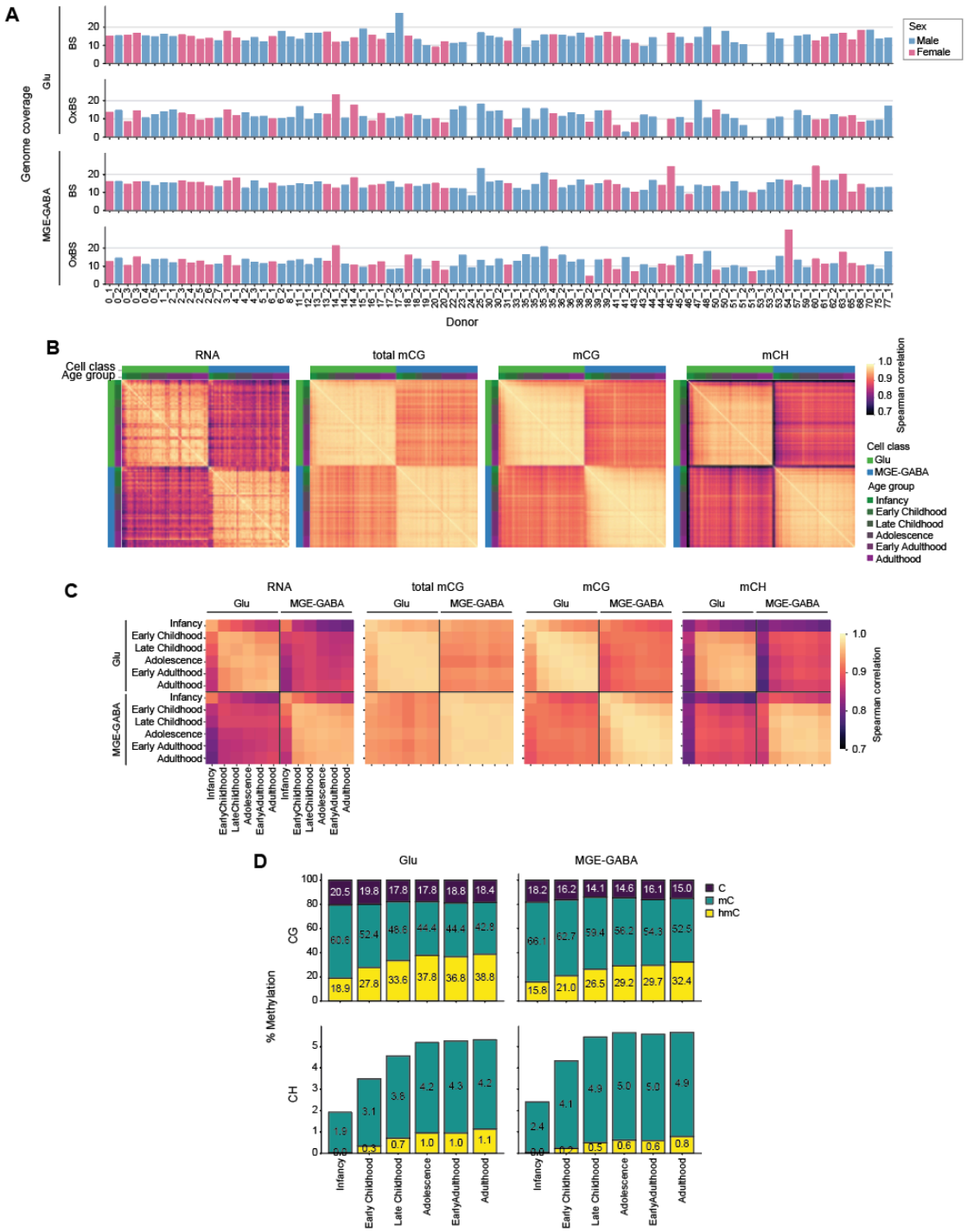
$$\frac{d\,C}{dt} = -\frac{C}{\tau_c} + \frac{hmC}{\tau_{hmC}}$$

$$\frac{d\,mC}{dt} = -\frac{mC}{\tau_{mC}} + \frac{C}{\tau_{c}}$$

$$\frac{d\,hmC}{dt} = -\frac{hmC}{\tau_{hmC}} + \frac{mC}{\tau_{mc}}$$

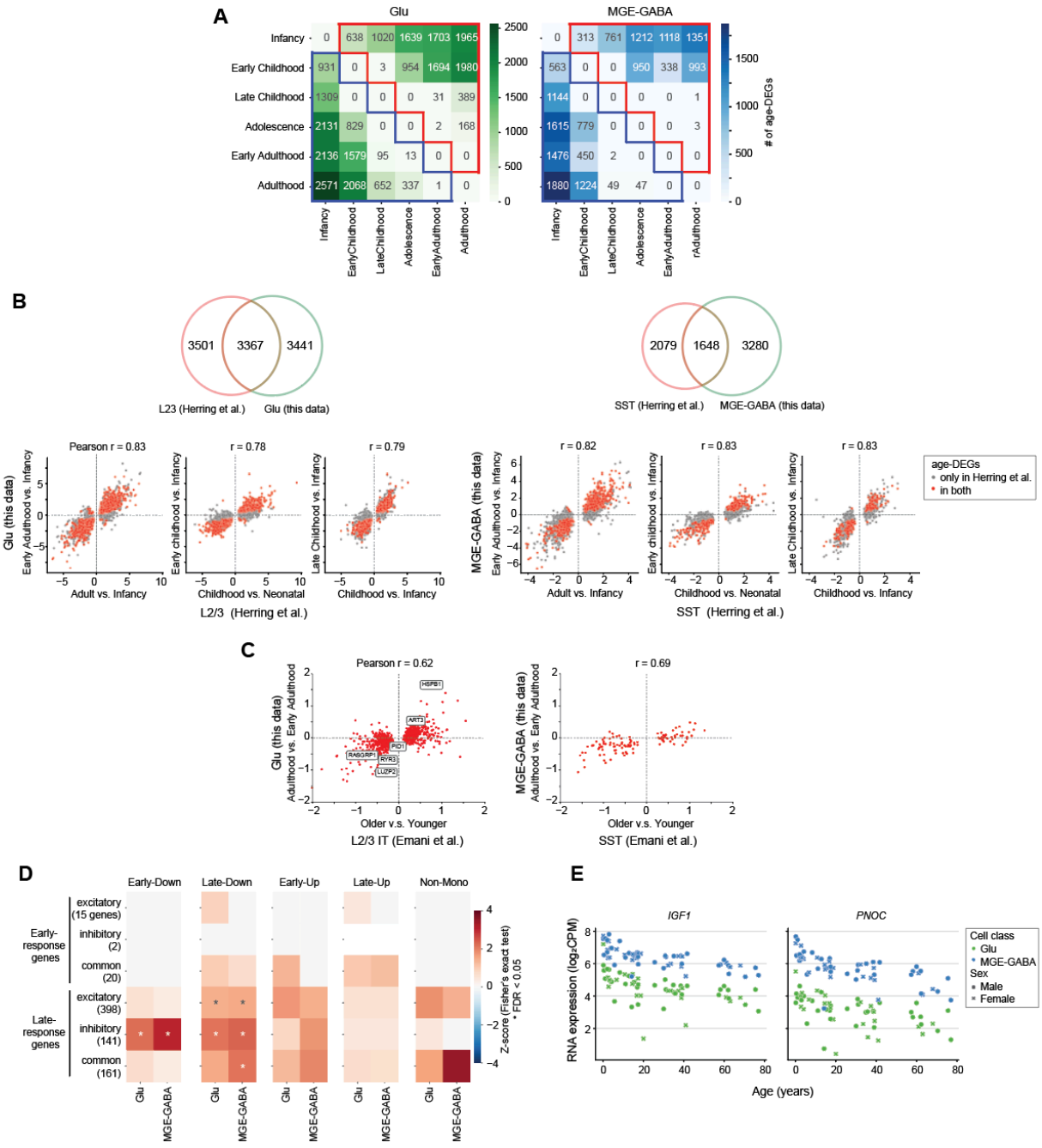
2.5 Acknowledgements

Chapter 2 is a work done in collaboration with Heng Xu, Alexey Kozlenkov, Ramu Vadukapuram, Junhao Li, Yu Wei, Chunyu Liu, Stella Dracheva and Eran Mukamel.



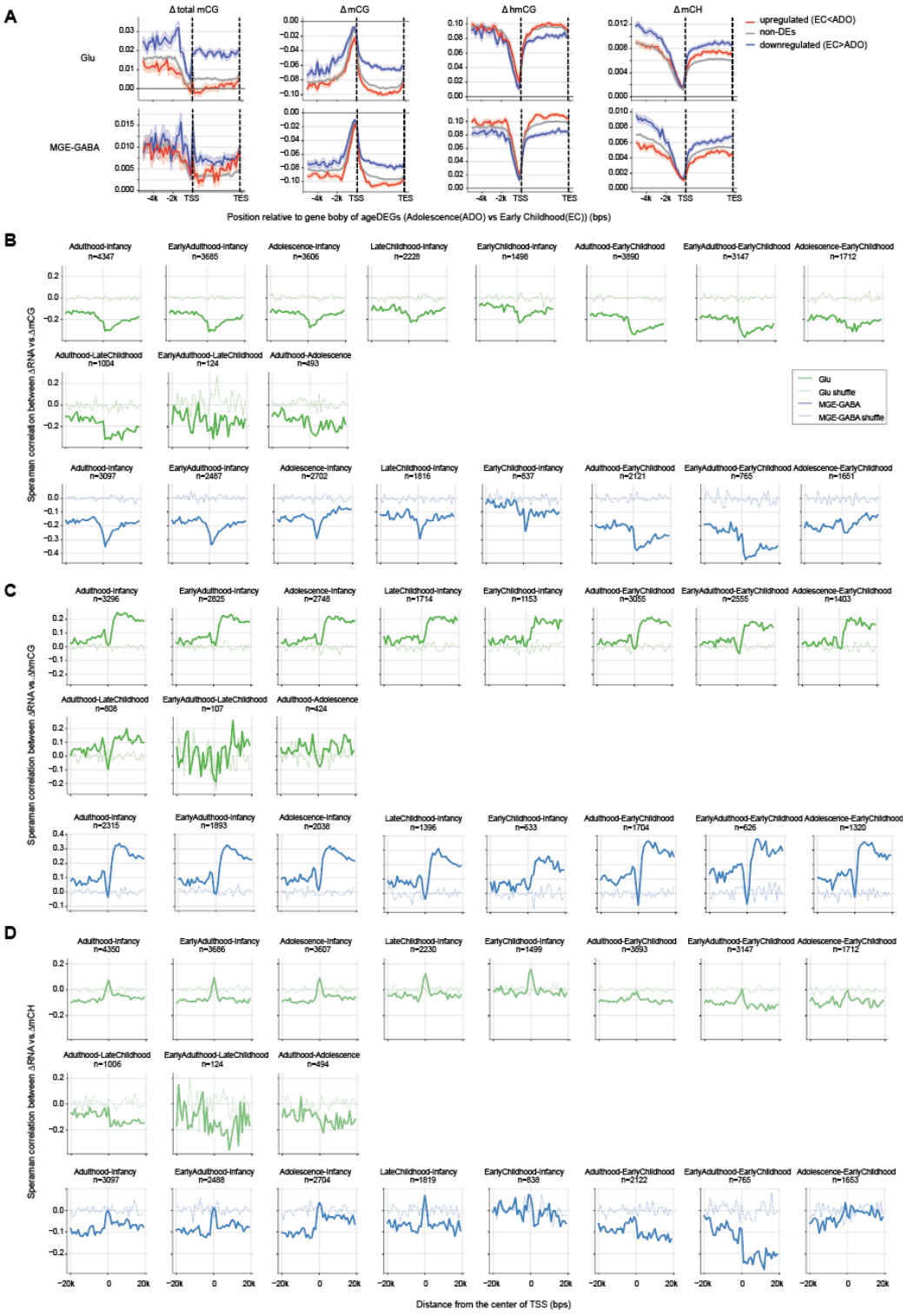
Supplemental Figure 2.1 | Correlation between samples and genome-wide DNA methylation in each age group, related to Figure 2.1.

(A) Genome-wide coverage of individual DNA samples. (B) Spearman correlation between donors in RNA expression and gene body total mCG, mCG and mCH. Expressed autosomal protein coding genes with length >1 kb were included (11,160 genes). The correlation between transcriptomes was calculated using normalized gene counts (log_2CPM). The correlation of DNA methylation was calculated using gene body methylation level. (C) Spearman correlation between age groups in RNA expression and gene body total mCG, mCG and mCH. (D) Mean genome-wide CG and non-CG (CH) methylation levels in each age group for Glu and MGE-GABA.

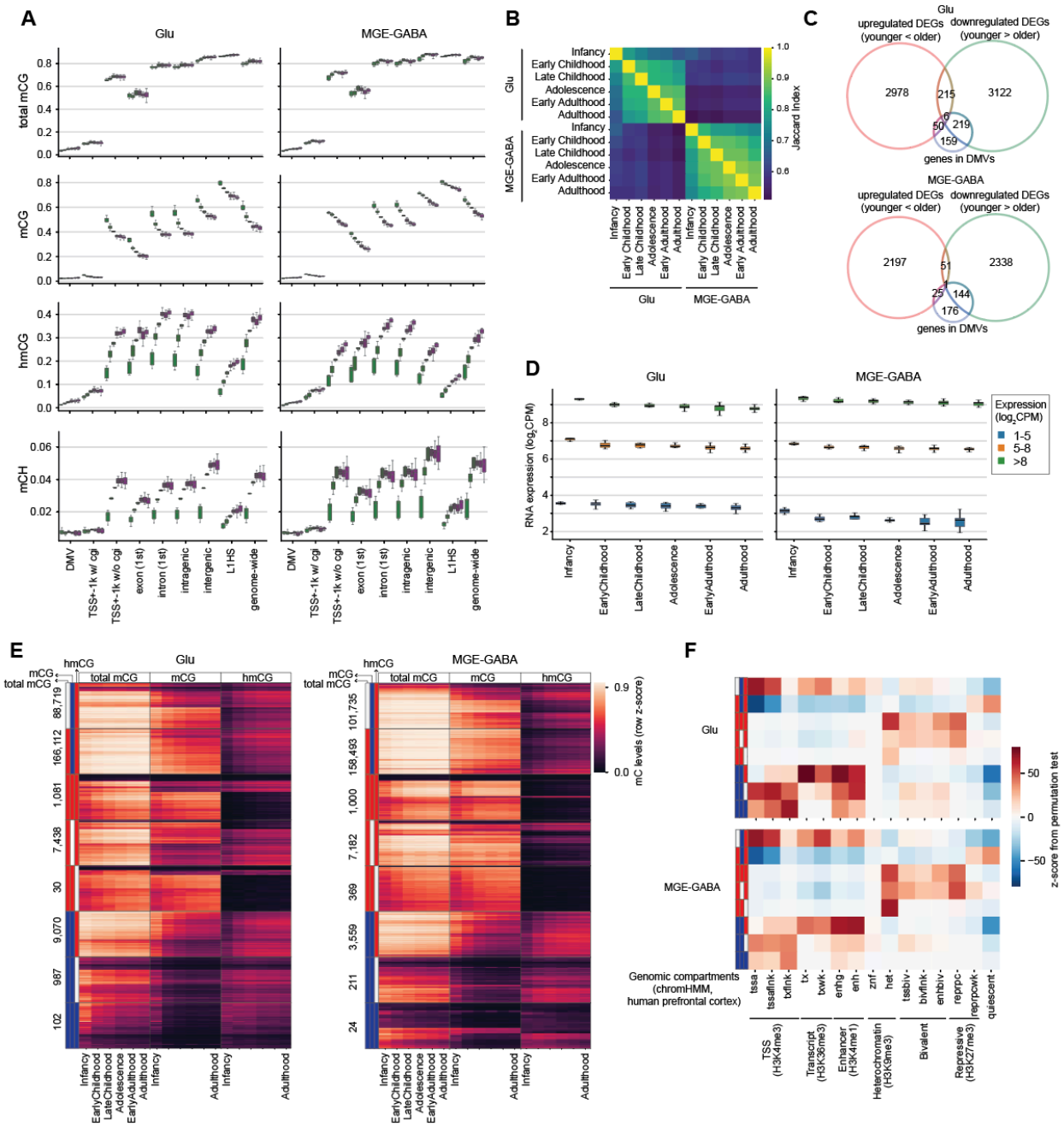


Supplemental Figure 2.2 | Consistency of age-DEGs with other published large scale datasets, related to Figure 2.2.

(A) Number of age-related differentially expressed genes (age-DEGs, FDR < 0.05) in pairwise age group comparisons. The upper triangle shows the number of upregulated genes (younger < older); the lower triangle shows the number of downregulated genes (younger > older). (B, C) age-DEGs compared with published dataset. Herring et al.[84](B). Emani et al.[14](C). (D) Enrichment of annotated activity-dependent genes[190] in our age-DEG groups. (E) Gene expression profiles of two late response genes, *IGF1* and *PNOC*.



Supplemental Figure 2.3 | DNA methylation features around age-DEGs, related to Figure 2.2.
(A) DNA methylation levels around transcription start sites (TSS) of age-DEGs separated by up- and down- regulated genes. (B-D) Spearman correlation between RNA expression fold change and (B) mCG (C) hmCG (D) mCH fold change within the flanking regions of TSS of age-DEGs separated by pairwise age group comparisons. n represents the number of age-DEGs in the comparison.

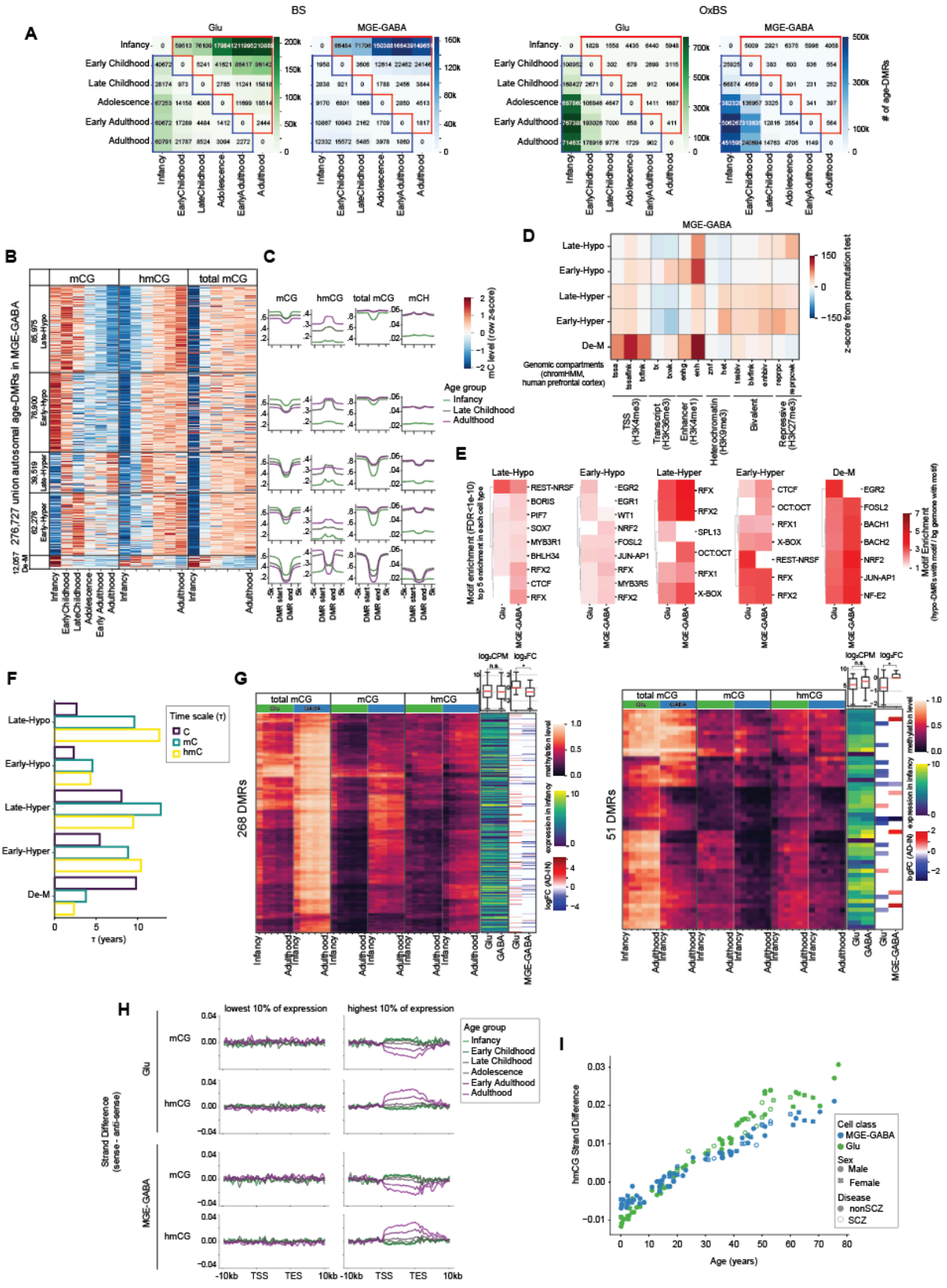


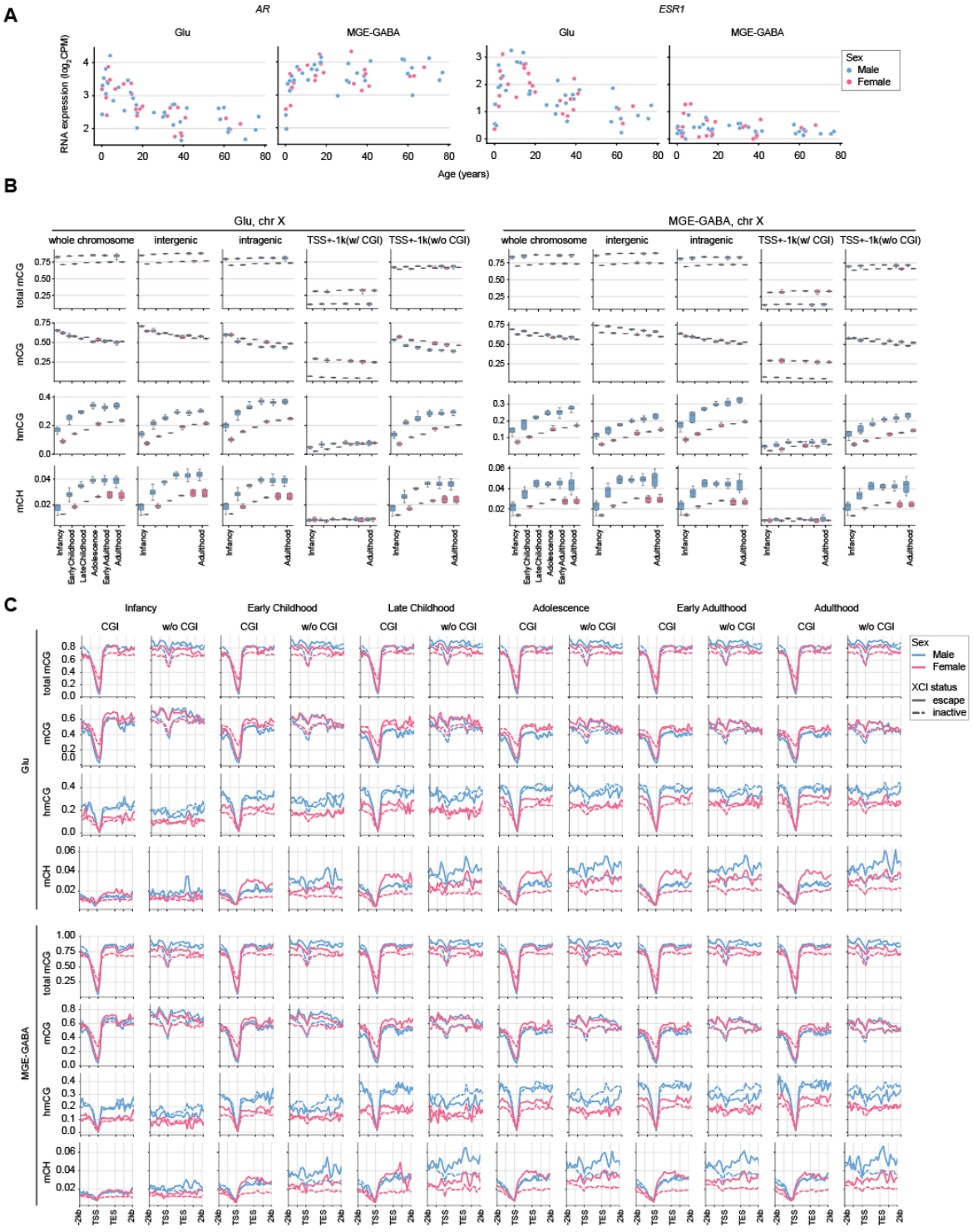
Supplemental Figure 2.4 | DNA Methylation Patterns in Genomic Compartments, related to Figure 2.3.

(A) DNA methylation in different genomic compartments. (B) Jaccard index between DMVs called in each age group and neuronal class. (C) Venn diagrams comparing age-DEGs and genes in DMV regions.. (D) RNA expression of genes in DMV regions stratified by expression level. (E) Categorization of 10kb bins by linear regression models. Colors in total mCG, mCG and hmCG columns represent the regulations of the 10kb bins: red for hypermethylation with age, blue for hypomethylation with age, and white for no significant change. (F) Enrichment of clusters of 10kb bins in ChromHMM-defined chromatin states[122]. DMV: DNA methylation valley. TSS: transcription start site. CGI: CpG island.

Supplemental Figure 2.5 | Dynamics of DNA methylation and strand asymmetry, related to Figure 2.3.

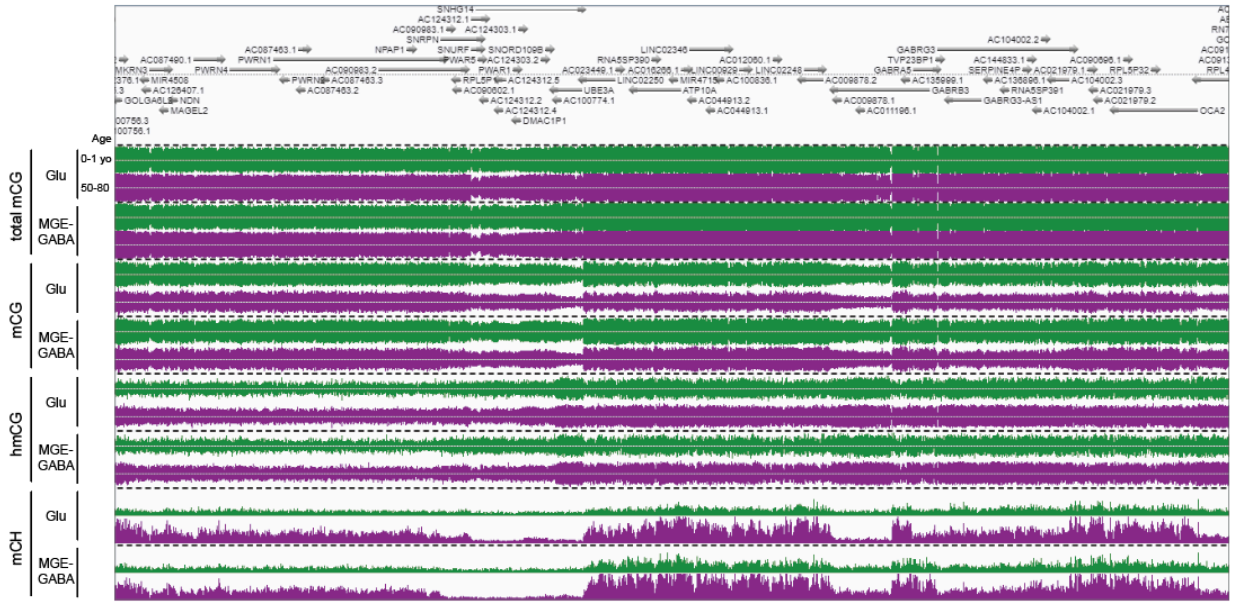
(A) Number of age-related differentially methylated regions (age-DMRs) in pairwise age group comparisons from bisulfite sequencing (BS) and oxidative bisulfite sequencing (OxBS). The upper triangle shows the number of hypermethylated DMRs (younger < older); the lower triangle shows the number of hypomethylated DMRs (younger > older). (B) Five clusters of age-DMRs based on mCG and hmCG dynamics in MGE-GABA. (C) Average methylation levels of different age-DMR clusters in infancy, late childhood and adulthood in MGE-GABA. (D) Enrichment of MGE-GABA age-DMRs in ChromHMM-defined chromatin states[122]. (E) Transcription factors motif enrichment in the age-DMRs from all clusters by HOMER[123]. (F) Mean estimated time scale from our dynamic model for all age-DMR in each age-DMR group. (G) Methylation levels in age-DMRs that show opposite trends of total mCG in Glu and MGE-GABA. The last two columns display the expression of genes near these DMRs in infancy and log fold change of the genes between infancy and adulthood. (H) Metagene plots showing the strand difference of mCG and hmCG for low expressed genes and high expressed genes across different age groups in both neuronal classes (I) Average hmCG strand difference for genes showing significant correlation between age and gene body strand asymmetry (1,173 genes in Glu and 760 genes in MGE-GABA neurons).


A
BS
OxBS
Glu
MGE-GABA
Infancy
Early Childhood
Late Childhood
Adolescence
Early Adulthood
Adulthood
# of age-DMRs
B
mCG
hmCG
total mCG
276,727 union autosomal age-DMRs in MGE-GABA
C
mCH
mC level (row z-score)
Age group
Infancy
Late Childhood
Adulthood
D
z-score from permutation test
Genomic compartments (chromHMM, human prefrontal cortex)
TSS (H3K4me3)
Transcript (H3K36me3)
Enhancer (H3K4me1)
Heterochromatin (H3K9me3)
Bivalent
Repressive (H3K27me3)
E
Motif enrichment (FDR<1e-10)
top 5 enrichment in each cell type
Motif Enrichment (hypo-DMRs with motif / bg genome with motif)
Late-Hypo
Early-Hypo
Late-Hyper
Early-Hyper
De-M
REST-NRSF
BORIS
PIF7
SOX7
MYB3R1
BHLH34
RFX2
CTCF
RFX
EGR2
EGR1
WT1
NRF2
FOSL2
JUN-AP1
RFX
MYB3R5
RFX2
RFX
RFX2
SPL13
OCT:OCT
RFX1
X-BOX
CTCF
OCT:OCT
RFX1
X-BOX
REST-NRSF
RFX
RFX2
EGR2
FOSL2
BACH1
BACH2
NRF2
JUN-AP1
NF-E2
F
Time scale (τ)
C
mC
hmC
τ (years)
G
log-CPM
logFC
268 DMRs
51 DMRs
methylation level
expression in infancy
logFC (AD-N)
H
lowest 10% of expression
highest 10% of expression
Strand Difference (sense - anti-sense)
Glu
MGE-GABA
-10kb
TSS
TES
10kb
I
hmCG Strand Difference
Age (years)
Cell class
MGE-GABA
Glu
Sex
Male
Female
Disease
nonSCZ
SCZ



Supplemental Figure 2.6 | Sex differences in DNA methylation are consistent across age, related to Figure 2.4.

(A) RNA expression of sex hormone receptor genes, androgen receptor (*AR*) and estrogen receptor (*ESR1*), with age (B) Methylation difference between males and females in different genomic regions on chromosome X. TSS: transcription start site. CGI: CpG island. (C) Methylation features of inactivated and escaped genes on the X chromosome in each age group.



Supplemental Figure 2.7 | Browser view of DNA methylation levels in imprinting control region 15q11, related to Figure 2.4.

DNA methylation in the imprinting control region 15q11 is lower in both infancy (labeled in green) and adulthood (labeled in purple). Each tick represents a cytosine site, with the height indicating the DNA methylation level at that site.

Chapter 3 Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation

3.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that manifests in progressive loss of cortical, brainstem and spinal motoneurons resulting in paralysis, and eventually death, typically, within 3-5 years of symptoms onset[216]. In ~10% of ALS patients, there is a clear family history of the disease, and pathogenic mutations are found in more than half of these cases[217]. By contrast, apparently sporadic ALS is considered a complex trait with an estimated heritability of 40–50%[218]. The pathology and genetics of ALS overlap with another neurodegenerative disease, frontotemporal dementia (FTD), which is the second most common cause of dementia in patients under 65 years[219]. Aggregates of the RNA-binding transactive response DNA binding 43 protein (TDP-43) accumulate in neurons of almost all ALS cases and approximately half of FTD cases[220], and ALS and FTD can occur within the same family and even in the same person[221].

Mutations in several genes are causative of ALS, FTD, or ALS/FTD[222]. The most common genetic cause is a dynamic hexanucleotide (GGGGCC; G_4C_2) repeat expansion in the first intron of the *C9orf72* (C9) gene, which is found in ~40% of familial and ~12% of all ALS and FTD cases[223–225]. The non-coding G_4C_2 expansion appears to have two direct consequences: a loss-of-function that causes *C9orf72* haploinsufficiency, and a gain-of-function associated with the expression of abnormal, bidirectionally transcribed RNAs containing the repeat[226]. These RNAs accumulate as RNA foci and/or are translated into toxic dipeptide repeat proteins (DPRs)[227–229]. ALS- and FTD-associated genes (including *C9orf72)* are expressed in multiple neuronal and non-neuronal cell types in the brain and spinal cord[230], raising the question of what impact the mutations have in each individual cell type.

ALS is caused by the degeneration of upper and lower motor neurons, affecting the giant pyramidal cells (Betz cells) in layer 5 of primary motor cortex and the large multipolar alpha motor neurons of the brainstem and spinal cord, whereas FTD results mainly from the degeneration of the large bipolar Von Economo neurons located in layer 5 of several specific cortical regions[231–233]. Animal and *in vitro* models associated multiple motoneuron-intrinsic (cell-autonomous) pathways with ALS pathogenesis, including glutamate excitotoxicity, mitochondrial dysfunction, RNA metabolism, nucleocytoplasmic transport, and alterations in protein homeostasis[234]. Although these pathways responded to therapeutic modulation in experimental models, clinical translation has been slow. The only approved therapies for ALS (riluzole, edaravone, and relyvrio) prolong survival by only several months[235], suggesting limitations of current disease models, especially those uniquely focusing on motoneurons. Other neuron types[236], as well as astrocytes and microglia, have been implicated in the pathogenesis of ALS and FTD[237,238].

We used single-nucleus transcriptomic (snRNA-seq) and epigenomic (snATAC-seq) analysis of human brain samples to investigate the cell-type-specific molecular alterations in C9-ALS or C9-FTD donors. In both diseases, we studied two key cortical regions, the motor and frontal cortices, in an attempt to uncover regional differences. In C9-ALS, we found the most pronounced transcriptional disruption in upper layer excitatory neurons and astrocytes, and parallel disruption of the epigenome, with concordant changes in chromatin accessibility, histone modifications, and gene expression in specific cell types. Our comparative analysis of C9-ALS and C9-FTD highlights distinct molecular pathologies in these two neurodegenerative diseases.

3.2 Results

3.2.1 Single nucleus transcriptomes in C9-ALS and C9-FTD specimens identify fine-grained cortical cell types

We performed snRNA-seq on nuclei isolated from the motor cortex (Brodmann area 4) and dorsolateral prefrontal cortex (Brodmann area 9; hereafter “frontal cortex”) of autopsied C9-ALS (n = 6), C9-FTD (n = 5), and pathologically normal (hereafter “controls”, n = 6) brains (Table S3.1). We obtained 105,120 high-quality nuclear gene expression profiles (45,376 from C9-ALS, 9,445 from C9-FTD, and 50,299 from controls), detecting a median of 6,351 unique mRNA molecules and 2,665 genes per nucleus, indicating the high quality of the data. Far fewer nuclei passed quality control in C9-FTD (median 24.3%) compared with controls (66.2%) or C9-ALS (68.4%) (two-sided Welch’s t-test, $p < 0.05$; Fig. S3.1a-c**;** see also “A distinct signature of glial dysregulation in C9-FTD” below).

We used iterative clustering to identify 49 distinct subpopulations of excitatory neurons, inhibitory neurons and non-neuronal cells (Fig. 3.1a-b; Table S3.2). The clustering was not biased by brain region, donor sex, sequencing batch, individual samples, or diagnosis (Fig. 3.1c, Fig. S3.2a-e), and the resulting clusters were stable with respect to the resolution parameter for community detection (Fig. S3.2f). The subpopulations, which we annotated based on the expression of known marker genes (Fig. S3.3), were highly consistent with a recent large-scale study of the human motor cortex (Fig. S3.4)[63]. These results demonstrated the fine-grained cell subtype resolution of our dataset. However, to leverage the information shared across multiple donors when estimating the molecular dysregulation in C9-ALS and C9-FTD, a statistically meaningful number of cells from each donor is required. Therefore, we focused our analysis on 14 major cell types (8 neuronal, 6 glial), defined by grouping subpopulations sharing well-defined markers (Supplementary Fig. 3.3). At this level of cell type resolution, we observed consistent expression patterns across donors in each major cell type and brain region of the control donors (Fig. 3.1d).

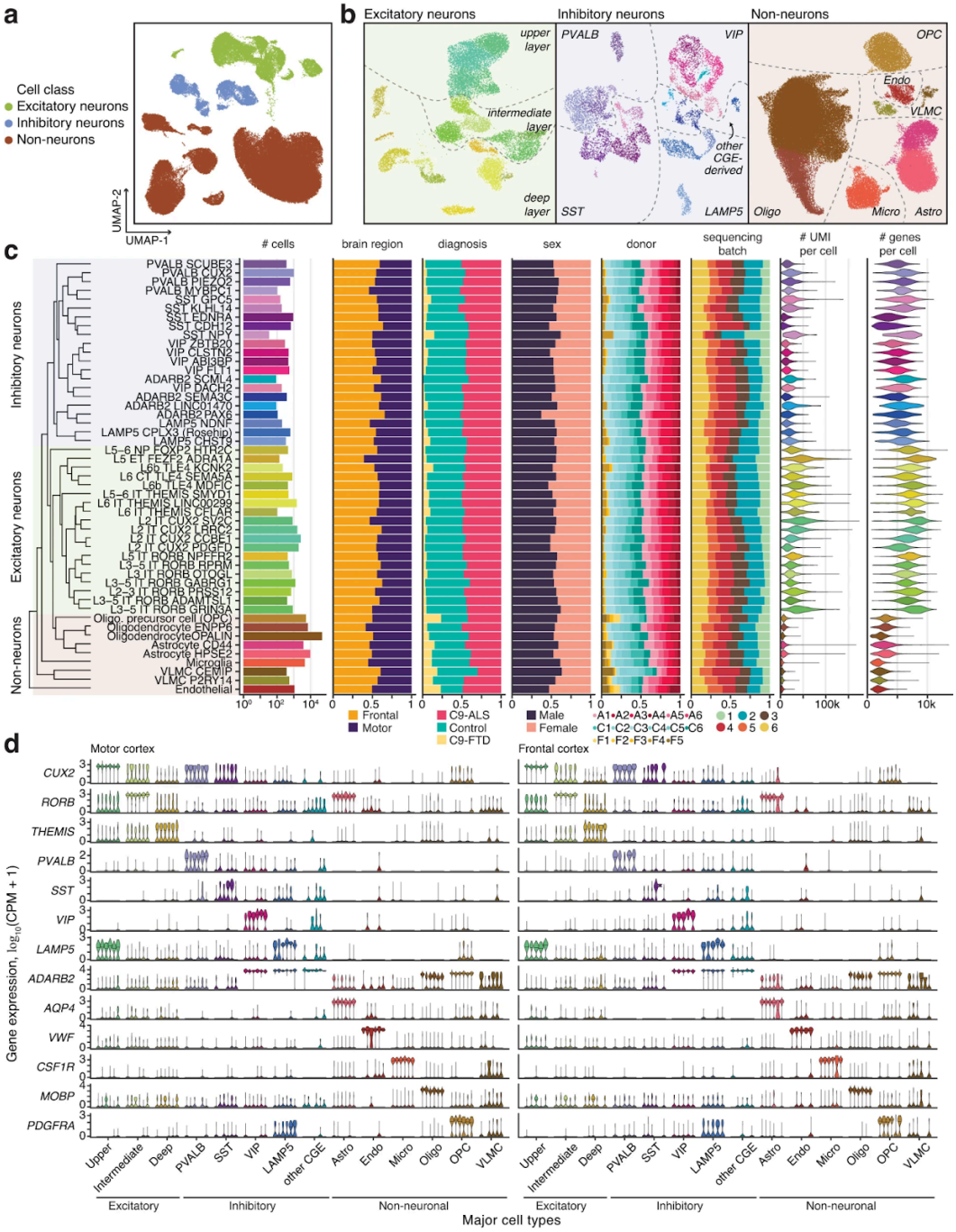


Figure 3.1 | A single cell transcriptomic analysis of human C9-ALS and C9-FTD motor and frontal cortex.

(a, b) Identification of transcriptomic cell types from snRNA-seq (n = 105,120 cells from 17 donors). Nuclei were first classified into three cell classes in (a), then subpopulations in each class were identified in (b). CGE caudal ganglionic eminence, Astro astrocytes, Endo endothelial cells, Micro microglia, Oligo oligodendrocytes, OPC oligodendrocyte precursor cells, VLMC vascular leptomeningeal cells, IT intratelencephalic, CT corticothalamic, NP near-projecting, ET extratelencephalic. (c) Neuronal and non-neuronal cell types were distributed across brain regions, donor diagnosis groups, sex, and sequencing batches. UMI unique molecular identifier. d Violin plots show marker gene expression in major cell types in each of the 6 control donors. CPM counts per million.

3.2.2 Disruption of gene expression in C9-ALS astrocytes and excitatory neurons

The rare giant pyramidal Betz cells in layer 5b of the human primary motor cortex are considered especially vulnerable in ALS[239,240]. Yet, other excitatory and inhibitory neurons, as well as other cortical areas such as the frontal regions, have also been reported to be affected in the disease[236,241]. Global analyses using principal components, Augur[111], and linear discriminant analysis demonstrate significant differences between disease and control samples (Fig. S3.14-3.15). To identify genes dysregulated in C9-ALS in a cell type- and region-specific manner, we compared nuclear transcriptional profiles from C9-ALS and control subjects using a mixed effects model[242], while adjusting for demographic and technical covariates (Methods).

Excitatory neurons in all cortical layers were profoundly affected in C9-ALS, with hundreds of differentially expressed (DE) genes in each cell type (FDR < 0.05) (Table S3, S4). There were more DE genes in upper layer excitatory neurons (L2/3) compared with excitatory neurons in other layers (Fig. 3.2a). In particular, there were 2.48-fold more DE genes in upper- compared with deep-layer (L5/6) excitatory neurons in motor cortex; in frontal cortex, the enrichment was 2.69-fold. This difference was evident even after subsampling each cell type to ensure equal statistical power for detecting DE genes (Fig. 3.2b). We also found that deep layer excitatory neurons were more affected in motor than in frontal cortex, with a larger number of DE genes after controlling for sample size (Fig. 3.2b). Inhibitory neurons had significantly fewer DE genes. Among non-neuronal cells, astrocytes were the most strongly affected in both brain regions (Fig. 3.2a-b).

Overall, transcriptional changes in C9-ALS were highly consistent in motor and frontal cortices (Pearson r = 0.49-0.74, Fig. 3.2c). Cluster analysis of the most strongly affected DE genes (fold-change > 2; Fig. 3.2d**)** showed a unique set of upregulated (cluster C3) or downregulated (C6) genes in astrocytes. In contrast, transcriptional changes in neurons were either shared between multiple neuronal subtypes (Clusters C2 and C5) or specific for excitatory neurons (Cluster C4).

We validated our snRNA-seq findings using bulk RNA-seq measurements in purified nuclei from major cortical cell types (Table S3.1, Methods). Using fluorescence-activated nuclear sorting (FANS), we purified neurons, oligodendrocyte lineage cells (mature oligodendrocytes and oligodendrocyte precursor cells, OPCs), and non-oligodendrocyte glial cells. The third group mainly consisted of astrocytes and microglia (Fig. S3.5a, b; Methods). Differential gene expression fold-changes in the single nucleus were strongly correlated with FANS RNA-seq data for corresponding cell types in both cortical regions (Fig. 3.2e), especially for astrocytes in motor cortex (Spearman r = 0.517, Fig. 3.2f).

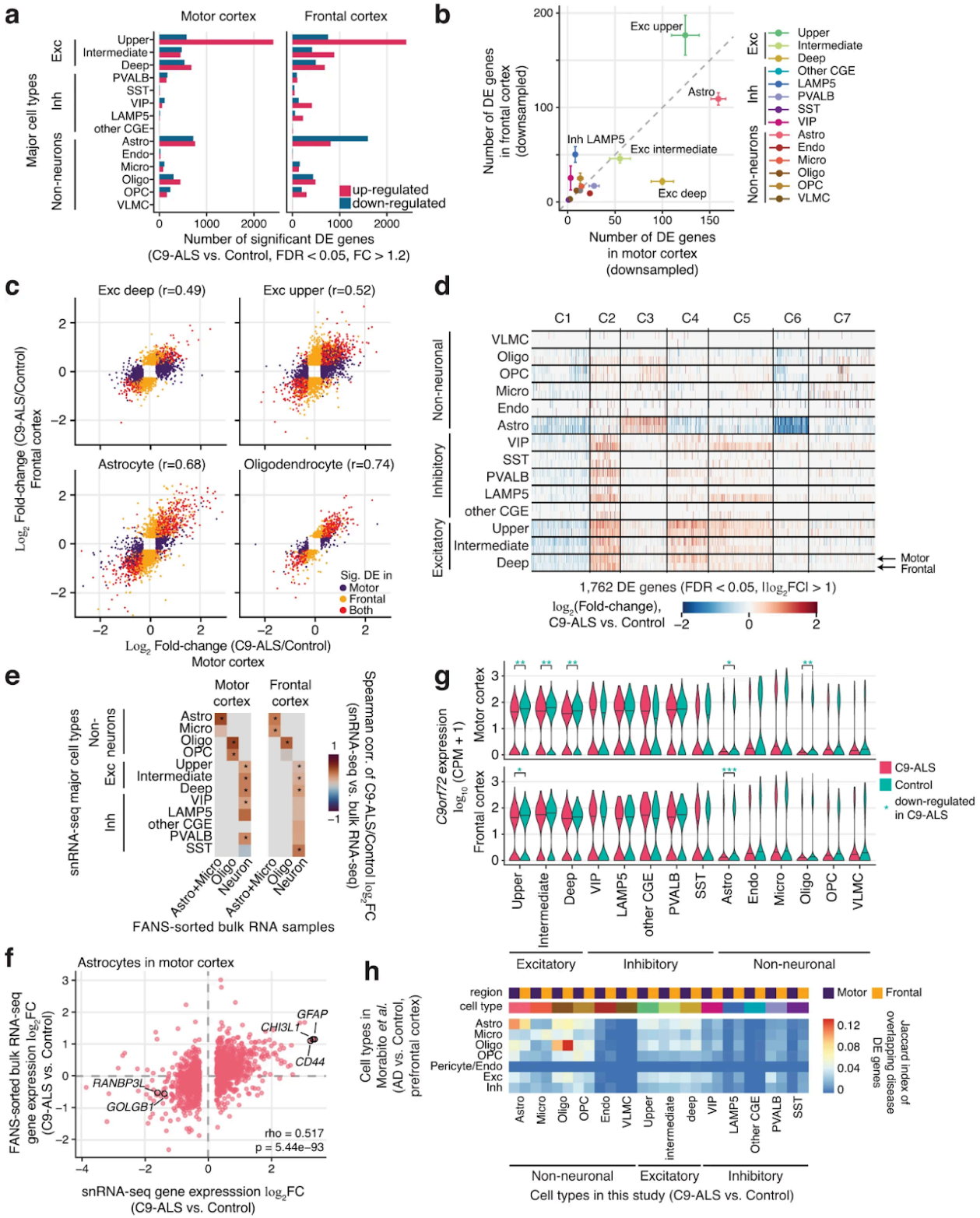
Studies in bulk brain tissues showed that the hexanucleotide repeat expansion in *C9orf72* lowers expression of this gene[224,243]. Using our single cell data, we found the highest expression of the *C9orf72* gene in neurons, with comparable expression levels in all neuronal subtypes (Fig. 3.2g). *C9orf72* was downregulated in C9-ALS excitatory, but not inhibitory neurons (FDR < 0.05). Among glial cells, only astrocytes and oligodendrocytes showed lower *C9orf72* expression in C9-ALS. However, because C9-ALS and control samples differ by both genotype and disease status, our study cannot determine whether downregulation of *C9orf72* expression in specific cell types is directly caused by the repeat expansion mutation, or is also influenced by ALS disease processes.

To connect transcriptional alterations in C9-ALS to protein expression, we used automated Western blotting in bulk motor cortex tissue (Methods). We tested multiple antibodies for proteins that were encoded by DE genes. Antibodies against 19 DE gene products consistently detected a protein band of the expected size in control samples. Using these antibodies, we confirmed the predicted changes in protein levels for 6 of the 19 DE genes, with no discordant changes observed for any of the tested proteins (Figs. 3.3-3.4; Figs. S3.6-S3.7; Table S3.10).

To examine whether our findings were specific to C9-ALS or were general responses to a degenerative process, we compared the DE genes in our C9-ALS vs. control analysis with the

reported DE genes in Alzheimer's disease (AD) in prefrontal cortex[244]. We found little overlap between the C9-ALS DE genes and the AD DE genes (the Jaccard index < 0.14 for all comparisons) (Fig. 3.2h). The overlap was highest for glial cell types (astrocytes and oligodendrocytes), suggesting a shared component of transcriptional disruption in glia across these two neurodegenerative diseases. The differences in neurons in C9-ALS or AD were more distinct. For genes that were significantly DE in both diseases, the fold-changes were in general correlated between the two diseases ($r > 0.58$, $p < 1E\text{-}14$ for astrocytes and oligodendrocytes, and $r > 0.23$, $p < 0.006$ for neuronal subtypes in prefrontal cortex) (Fig. S3.13a). We also examined the significant DE genes that had concordant or discordant effects in ALS (our study) vs. AD[244]. For this comparison, we focused on the cell types that were most affected in C9-ALS, such as astrocytes and excitatory upper/deep neurons, as well as oligodendrocytes that had the highest Jaccard index (Fig. S3.13b-e). We found that *CD44* and *MAOB* were upregulated in astrocytes in both C9-ALS and AD, whereas *DNMT3A* was downregulated and *HSPH1* was upregulated in both neuronal subtypes. In contrast, *KNDC1* was upregulated in C9-ALS but downregulated in AD in prefrontal cortex astrocytes, whereas *HTR1E* was downregulated in C9-ALS but upregulated in AD in the prefrontal cortex deep excitatory neurons.

Figure 3.2 | Dysregulation of gene expression in C9-ALS is concentrated in astrocytes and excitatory neurons.
(a) The number of differentially expressed (DE) genes (fold-change>1.2, FDR < 0.05). Exc, excitatory neurons; Inh, inhibitory neurons. (b) After downsampling to 30 nuclei per donor in each cell type to ensure equivalent statistical power, the majority of DE genes were detected in astrocytes and excitatory neurons. Astrocytes and deep layer excitatory neurons were more affected in motor cortex compared with frontal cortex. Random downsampling was performed 10 times, and the dots and error bars represent mean ± SEM. (c) C9-ALS transcriptional differences were consistent in two cortical regions. r, Pearson correlation coefficient. (d) K-mean clustering analysis of strongly DE genes (>2 fold-change) showed distinct groups of genes are affected in neurons and astrocytes. (e) Validation of snRNA-seq differential expression by bulk RNA-seq in FANS-purified cells. Spearman correlations of the C9-ALS vs. control FC between the snRNA-seq and bulk-RNA-seq were computed using significant DE genes found in snRNA-seq shown in (a). *Spearman correlation FDR < 0.05. (f) C9-ALS astrocytes exhibit consistent differential expression in snRNA-seq and in bulk RNA-seq. Dots represent significant DE genes (FC > 1.2, FDR < 0.05) found in snRNA-seq. Selected genes of interest with concordant FC are highlighted. rho, Spearman correlation coefficient. (g) *C9orf72* expression was downregulated in C9-ALS in excitatory neurons, astrocytes and oligodendrocytes. Asterisks indicate significant differential expression identified with MAST (*FDR < 0.05; **FDR < 0.01; ***FDR < 0.001). CPM, counts per million. (h) Overlaps of significant C9-ALS vs. control DE genes (this study) and AD vs. control DE genes (Morabito et al.) for each cell type, measured in Jaccard index.


a
Motor cortex
Frontal cortex
Major cell types
Exc
Inh
Non-neurons
Upper
Intermediate
Deep
PVALB
SST
VIP
LAMP5
other CGE
Astro
Endo
Micro
Oligo
OPC
VLMC
up-regulated
down-regulated
Number of significant DE genes
(C9-ALS vs. Control, FDR < 0.05, FC > 1.2)
b
Exc upper
Astro
Inh LAMP5
Exc intermediate
Exc deep
Number of DE genes in frontal cortex (downsampled)
Number of DE genes in motor cortex (downsampled)
c
Exc deep (r=0.49)
Exc upper (r=0.52)
Astrocyte (r=0.68)
Oligodendrocyte (r=0.74)
Log2 Fold-change (C9-ALS/Control) Frontal cortex
Log2 Fold-change (C9-ALS/Control) Motor cortex
Sig. DE in
Motor
Frontal
Both
d
C1
C2
C3
C4
C5
C6
C7
Non-neuronal
Inhibitory
Excitatory
1,762 DE genes (FDR < 0.05, log2FCl > 1)
log2(Fold-change), C9-ALS vs. Control
Motor
Frontal
e
Motor cortex
Frontal cortex
snRNA-seq major cell types
Non-neurons
Exc neurons
Inh
Astro+Micro
Oligo
Neuron
FANS-sorted bulk RNA samples
Spearman corr. of C9-ALS/Control log2FC (snRNA-seq vs. bulk RNA-seq)
f
Astrocytes in motor cortex
FANS-sorted bulk RNA-seq gene expression log2FC (C9-ALS vs. Control)
GFAP
CHI3L1
CD44
RANBP3L
GOLGB1
rho = 0.517
p = 5.44e−93
snRNA-seq gene expresssion log2FC (C9-ALS vs. Control)
g
C9orf72 expression log10 (CPM + 1)
Motor cortex
Frontal cortex
C9-ALS
Control
down-regulated in C9-ALS
Excitatory
Inhibitory
Non-neuronal
h
Cell types in Morabito et al. (AD vs. Control, prefrontal cortex)
region
cell type
Motor
Frontal
Pericyte/Endo
Exc
Inh
Jaccard index of overlapping disease DE genes
Non-neuronal
Excitatory
Inhibitory
Cell types in this study (C9-ALS vs. Control)

3.2.3 Astrocyte transcriptional dysregulation in C9-ALS

We observed a distinct transcriptional signature of astrocyte activation in C9-ALS brains[245]. mRNA for *GFAP,* a classic marker of astrocyte activation that encodes an intermediate filament protein, increased 10.3-fold in astrocytes of the motor cortex, and 2.4-fold in frontal cortex (Fig. 3.3a-b). Likewise, we found higher expression of the reactive astrocyte-associated genes *CD44* (9.4-fold in motor cortex, 3.7-fold in frontal) and *CHI3L1* (10-fold in both regions). CD44-positive astrocytes were found in a variety of neurological diseases[246], and elevated levels of chitinases (including CHI3L1) have been proposed as a biomarker of ALS severity and progression[247,248]. We confirmed that GFAP and CD44 proteins were upregulated in C9-ALS bulk motor cortex tissue ($p < 0.05$; Fig. 3.3c; Fig. S3.6**).** Furthermore, immunofluorescence confirmed that GFAP protein is highly enriched throughout the cell body and processes of astrocytes in C9-ALS motor cortex (Fig. 3.3d).

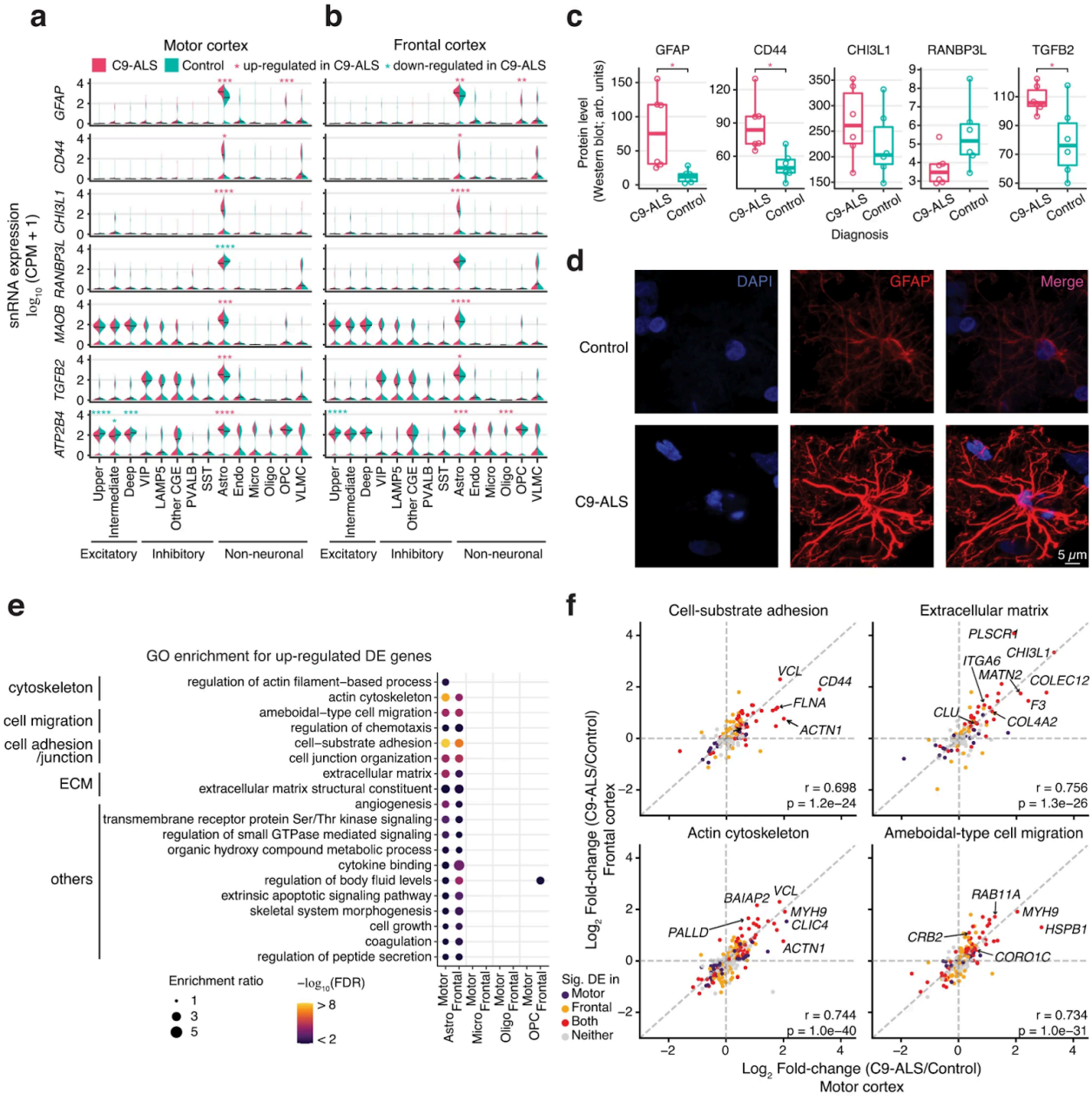
To assess the functional relevance of the altered gene expression in astrocytes, we performed functional enrichment analysis of DE genes[249]. In both cortices, upregulated genes were enriched for multiple functions including cell adhesion and extracellular matrix (exemplified by *ITGA6-7, CLU, COL4A2* genes), actin cytoskeleton (e.g., *ACTN1*, *PALLD*), and cell migration (e.g., *CRB2, CORO1C*), suggesting substantial cytoskeleton and cell-surface protein remodeling of the C9-ALS astrocytes (GO enrichment FDR < 0.05, Fig. 3.3e; Table S3.5). DE genes in these functional categories were affected in a consistent manner in both motor and frontal cortices ($r > 0.7$, Fig. 3.3f, Methods). Remodeling of astrocyte morphology is a hallmark of neurodegeneration[250], and our findings highlight the genes involved in this process in C9-ALS. We did not find any functional categories enriched for genes that were downregulated in astrocytes.

We detected changes in the expression of genes whose protein products or functional pathways were previously associated with ALS. For example, several studies reported alterations of transforming growth factor-β2 (TGF-β) signaling in ALS models[251,252]. Consistent

with these reports, we found strong upregulation of the *TGFB2* gene (1.56-1.95 fold in the two cortices), as well as an increased TGF-β protein level in C9-ALS astrocytes **(**Fig. 3.3a-c, Fig. S3.6). We found that the expression of *RANBP3L*, a gene encoding a Ran-binding protein involved in nucleocytoplasmic transport[253], was downregulated 3-fold in motor cortex of C9-ALS astrocytes (Fig. 3.3a-b). We also detected a trend to a reduction of RANBP3L protein in the C9-ALS motor cortex ($p = 0.063$) (Fig. 3.3c, Fig. S3.6). This finding is consistent with reports of reduced expression of neuronal nucleocytoplasmic transport proteins in C9-ALS[254–257].

Our data enabled a granular cell-type-specific analysis of transcriptional alterations in C9-ALS, in some cases, identifying genes with opposite patterns of altered expression in different cell types. For example, we found that *ATP2B4,* which encodes plasma membrane calcium ATPase PMCA4 that catalyzes the hydrolysis of ATP coupled with transport of Ca^{2+} from the cytoplasm to the extracellular milieu, showed ~2.0-fold higher expression in C9-ALS *vs.* control astrocytes in both cortical regions (Fig. 3.3a-b). In contrast, *ATP2B4* was downregulated in upper and deep layer excitatory neurons. Astrocytes exhibit dynamic cell state changes mediated by cytosolic Ca^{2+} [258]. Thus, increased expression of *ATP2B4* in astrocytes likely represents a protective mechanism to counterbalance upregulation of Ca^{2+} levels in C9-ALS.

Figure 3.3 | Dysregulation of gene expression in C9-ALS astrocytes suggests activation and structural remodeling.
(a, b) Examples of genes that were prominently dysregulated in C9-ALS astrocytes. Asterisks marked the significant upregulated (pink) and downregulated (cyan) DE genes in C9-ALS from the MAST analysis. *FDR < 0.05; **FDR < 0.01; ***FDR < 0.001; ****FDR < 0.0001. (c) Examples of DE genes in astrocytes with corresponding changes in their protein products. Protein levels were measured by automated Western blot analysis. Arb. units, arbitrary units. N = 6 C9-ALS and 6 control biologically independent samples, and each sample was measured in two technical replicates. In each box plot, dots are the average from the two technical replicates for each sample; the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. *Two-sided Welch's *t*-test p = 0.033, 0.012, and 0.033 for GFAP, CD44 and TGFB2, respectively. (d) Representative immunofluorescence images of the GFAP-positive astrocytes in motor cortex (red) obtained using confocal microscopy. DAPI was used to stain nuclei (blue). Immunofluorescence showed strong upregulation of GFAP immunoreactivity in astrocytes in C9-ALS donors. The images were acquired in 2 control and 2 C9-ALS subjects, 3 sections per subject and 3 fields in each section. (e) Top Gene Ontology (GO) terms enriched for upregulated DE genes in astrocytes. Enrichment of the same terms for upregulated DE genes in other glia cell types are shown as a comparison. Enrichment ratio is the number of observed genes divided by the number of expected genes from each GO term. ECM extracellular matrix. See Table S3.5 for the full list of GO enrichment results. (f) Genes in four functional categories had consistent patterns of differential expression in astrocytes from both brain regions. r and p, two-sided Pearson correlation test coefficient and p-value.


a
b
c
d
e
f
Motor cortex
Frontal cortex
C9-ALS
Control
up-regulated in C9-ALS
down-regulated in C9-ALS
snRNA expression
log10 (CPM + 1)
GFAP
CD44
CHI3L1
RANBP3L
MAOB
TGFB2
ATP2B4
Upper
Intermediate
Deep
VIP
LAMP5
Other CGE
PVALB
SST
Astro
Endo
Micro
Oligo
OPC
VLMC
Excitatory
Inhibitory
Non-neuronal
Protein level
(Western blot; arb. units)
Diagnosis
DAPI
GFAP
Merge
Control
C9-ALS
5 μm
GO enrichment for up-regulated DE genes
cytoskeleton
cell migration
cell adhesion /junction
ECM
others
regulation of actin filament–based process
actin cytoskeleton
ameboidal–type cell migration
regulation of chemotaxis
cell–substrate adhesion
cell junction organization
extracellular matrix
extracellular matrix structural constituent
angiogenesis
transmembrane receptor protein Ser/Thr kinase signaling
regulation of small GTPase mediated signaling
organic hydroxy compound metabolic process
cytokine binding
regulation of body fluid levels
extrinsic apoptotic signaling pathway
skeletal system morphogenesis
cell growth
coagulation
regulation of peptide secretion
Enrichment ratio
−log10(FDR)
Cell-substrate adhesion
Extracellular matrix
Actin cytoskeleton
Ameboidal-type cell migration
VCL
CD44
FLNA
ACTN1
PLSCR1
CHI3L1
ITGA6
MATN2
COLEC12
F3
CLU
COL4A2
BAIAP2
MYH9
CLIC4
PALLD
RAB11A
HSPB1
CRB2
CORO1C
r = 0.698
p = 1.2e−24
r = 0.756
p = 1.3e−26
r = 0.744
p = 1.0e−40
r = 0.734
p = 1.0e−31
Log2 Fold-change (C9-ALS/Control)
Frontal cortex
Log2 Fold-change (C9-ALS/Control)
Motor cortex
Sig. DE in
Motor
Frontal
Both
Neither

3.2.4 C9-ALS-associated transcriptional alterations in excitatory neurons

Our analysis showed that upper layer (L2/3) and deep layer (L5/6) excitatory neurons were strongly affected in C9-ALS (Fig. 3.2b), and we found a large number of DE genes in both cortical regions (examples in Fig. 3.4a; Table S3.3, S3.4**).** Functional analysis of genes upregulated in motor cortex showed the most significant enrichments for gene ontology (GO) categories associated with mitochondrial function, protein synthesis, and cellular proteostasis, the process by which the health of the cell's proteins is monitored and maintained[259] (FDR < 0.05) (Fig. 3.4c; Table S3.5). There were also notable enrichments for categories associated with nucleocytoplasmic transport and DNA damage.

Although C9-ALS upper and deep excitatory neurons showed a largely consistent effect on gene expression in motor and frontal cortices (Fig. 3.2c), only a small number of GO categories were enriched for genes upregulated in frontal cortex (Table S3.5**)**. To understand the regional differences of the disrupted pathways, we compared genes from the enriched GO categories across the two regions (Fig. 3.4c-e; Methods**)**. We found a similar pattern of differential expression for genes associated with cellular proteostasis (protein folding) in both regions. However, genes associated with mitochondrial function (respiratory chain) were more affected in motor compared to frontal cortex in deep layer excitatory neurons. Genes associated with protein synthesis (protein localization to endoplasmic reticulum) were more affected in motor cortex in both neuronal subtypes. This observation was consistent with our downsampling analysis (Fig. 3.2b), showing that certain genes in deep layer excitatory neurons were more strongly affected in C9-ALS in motor cortex compared to frontal cortex. The majority of the DE genes showed stronger upregulation in upper *vs.* deep layer excitatory neurons (Fig. 3.4f).

We detected widespread upregulation of mitochondrial genes in C9-ALS (Fig. 3.4c-d**)**. Among the DE genes, were those coding for subunits of all four mitochondrial respiratory chain complexes (Complexes I-IV), such as NADH ubiquinone oxidoreductase (*NDUF* family genes), succinate dehydrogenase (*SDHA, SDHB, SDHC*), cytochrome bc1 complex (*UQCR* family and

CYC1), and cytochrome c oxidase (*COX* family and *CYCS*), as well as ATP synthase complex (*ATP5* family) (Fig. 3.4a; Table S3.3, S3.4). Genes that encode proteins involved in mitochondrial membrane transport (*SLC25A* family, including *SLC25A4*) and chaperones that assist the import of proteins from the cytoplasm into the mitochondrial inner membrane (*TIMM* family, e.g., *TIMMDC1*) were also upregulated.

Among upregulated genes associated with proteostasis, there was a large group of genes that encode members of the heat shock protein 70 (HSP70) molecular chaperone family (e.g., *HSPA4, HSPA8, HSPA9*) (Fig. 3.4e; Table S3.3, S3.4)[260]. HSP70 proteins are required for aggregation prevention, folding of newly synthesized proteins, conformational maintenance, as well as degradation of misfolded proteins and protein aggregates in autophagic-lysosomal or ubiquitin-proteasome pathways[261]. HSP70s carry out their functions together with members of large families of co-chaperones. Genes encoding multiple co-chaperones of HSP70 were also upregulated in C9-ALS excitatory neurons, such as members of *DNAJ* and *BAG* families (Table S3.3-S3.4), as well as several other heat shock proteins (e.g., *HSPB1*, *HSPB11*) that facilitate the function of HSP70s in different protein quality control systems, including the degradation of protein aggregates in chaperone assisted selective autophagy[262–264]. Members of the HSP90 family of chaperons were also upregulated in C9-ALS excitatory neurons (Fig. 3.4a, e). Similar to HSP70, HSP90 chaperones bind to misfolded proteins, including TDP-43, and assist in their transition to native conformation or prevent their accumulation in misfolded state[265]. Of particular interest is the 1.8-fold increase in expression of *CLU* in upper layer excitatory neurons (Fig. 3.4a). Notably, upregulation of *CLU* was also detected in astrocytes (Fig. 3.3f). *CLU*, which is a risk gene for AD[266], encodes a normally secreted chaperone protein, clusterin, that is redirected to the cytosol during ER stress. Clusterin directly interacts with TDP-43 *in vitro* and potently inhibits its aggregation. Thus, increased expression of clusterin could provide an important defense against intracellular proteotoxicity associated with ALS[267]. We confirmed the upregulation of HSP90 and clusterin proteins in C9-ALS motor cortex (Fig. 3.4b; Fig. S3.7).

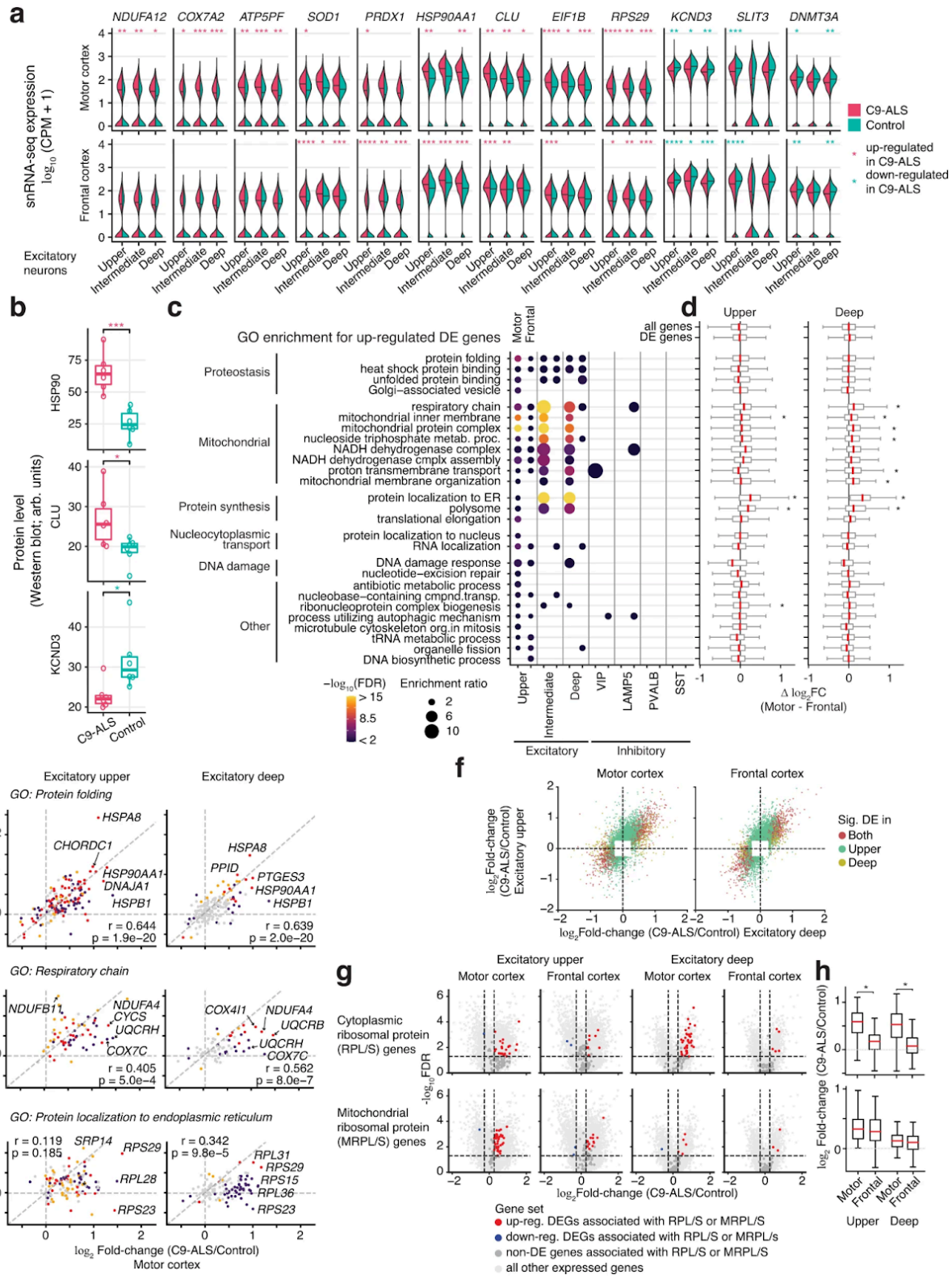
Nearly half of the nucleus-encoded genes for cytoplasmic (*RPL/S* families) and mitochondrial (*MRPL/S* families) ribosomal proteins were upregulated in C9-ALS (Table S3.3-S3.4). In both upper and deep excitatory neurons, we detected a larger number of upregulated cytoplasmic ribosomal genes in motor cortex compared to frontal cortex (Fig. 3.4g-h). In contrast, this difference was not observed for the mitochondrial ribosomal genes. Other C9-ALS upregulated genes that are associated with protein synthesis were those encoding translation initiation factors (e.g., *EIF1B*) (Fig. 3.4a). These intricate differences could result from region and/or cell specificity of the pathological changes in C9-ALS brain and will require further study.

Genes upregulated in C9-ALS excitatory neurons were linked to processes that are essential for all cell types (e.g., mitochondria function, proteostasis, protein synthesis), whereas genes downregulated in motor cortex were enriched for categories that are specific to neuronal cells (*e.g.,* neuronal cell body category that includes *C9orf72,* and genes that encode several potassium channels) (Fig. 3.4a; Fig. S3.8; Table S3.3-3.4). Downregulated DE genes were also enriched for categories related to neuronal projections, and cell adhesion. Similar to upregulated genes, in both cortical regions, the C9-ALS-associated downregulated genes were more strongly affected in upper *vs.* deep layer excitatory neurons (Fig. 3.4f). Likewise, we did not find significant differences in differential expression between motor and frontal cortices when we compared genes from the GO categories that were enriched in either cortical region (Fig. S3.8a-b**)**. We also examined the distribution of DE genes across fine-grained excitatory neuron cell types, and found broad enrichment across upper-layer neuron types as well as L6 intra-telencephalic (IT) neurons (Fig. S3.8c).

The downregulated genes include genes that encode brain derived neurotrophic factor (*BDNF*), slit guidance ligands and their receptor (*SLIT1, SLIT3, ROBO2)*, semaphorin proteins and their receptors (*SEMA4B, SEMA4C, SEMA5B, PLXNA1, PLXNA2, PLXNB2*), and RAP1 GTPase activating protein (*RAP1GAP)*. All these proteins are linked to neuronal plasticity in the

developing and/or adult brain, and *BDNF* has been implicated in the pathophysiology of several psychiatric and neurodegenerative diseases including ALS[268–273]. Also notable is downregulation of genes that encode several potassium channels (e.g., *KCNN3, KCND3*)[274]. Unexpectedly, we observed ~45% decrease in expression of *de novo* DNA methyltransferase *DNMT3A* in the upper- and deep-layer excitatory neurons in motor cortex, suggesting a link to epigenetic dysregulation in C9-ALS. C9-ALS-associated changes in *DNMT3A* levels were previously reported in several studies, yielding conflicting results[275–277]. Consistent with our gene expression data, we detected downregulation of KCND3 protein, as well as a trend for RAP1GAP and DNMT3A proteins in C9-ALS motor cortex (Fig. 3.4b; Fig. S3.7).

Figure 3.4 | Excitatory neurons have altered expression of metabolic and protein regulatory pathway genes.
(a) Examples of genes that were dysregulated in C9-ALS excitatory neurons. Asterisks marked the upregulated and downregulated DE genes in C9-ALS from the MAST analysis. *FDR < 0.05; **FDR < 0.01; ***FDR < 0.001; ****FDR < 0.0001. CPM counts per million. (b) Examples of DE genes for which corresponding changes in protein abundance were confirmed by automated Western blot. Arb. units, arbitrary units. $N = 6$ C9-ALS and 6 control biologically independent samples, and each sample was measured in two technical replicates. Dots are the average from the two technical replicates for each sample. Asterisks denote two-sided Welch's *t*-test $p = 6.35e\text{-}4$, 0.041, and 0.036 for HSP90, CLU and KCND3, respectively. (c) Top GO terms enriched for genes upregulated in upper and/or deep layer excitatory neurons. Enrichments of the same terms for upregulated DE genes in other neuronal cell types are shown for comparison. Enriched GO categories (FDR < 0.01) were selected by affinity propagation. Enrichment ratio is the number of observed genes divided by the number of expected genes from each GO term. (d) Differences in the C9-ALS vs. control fold-changes between motor and frontal cortex (Δ log2FC = log2FC in motor cortex - log2FC in frontal cortex). The boxes denote the distribution of these differences (Δ log2FC) of all expressed genes in each GO category. As background comparisons, the distributions of these differences for all expressed genes (labeled as "all genes") and for all C9-ALS vs. control DE genes (labeled as "DE gene") are shown on top. Two-sided Welch's *t*-tests were used to test whether the Δ log2FC in each group of genes were significantly different from the Δ log2FC of the "all genes" control set, and * mark the significant differences (FDR < 0.005). Exact N numbers and *p*-values for this analysis are provided in Table S3.5. (e) Comparison of effects in motor and frontal cortices for GO categories exemplifying three major cellular processes enriched for upregulated genes. r, Pearson correlation coefficient. (f) Comparison of the C9-ALS vs. control expression fold-changes between DE genes in upper- and deep-layer excitatory neurons. (g) Volcano plots demonstrating the dysregulation of genes associated with cytoplasmic ribosomal proteins (RPL/S, top row) and mitochondrial ribosomal proteins (MRPL/S, bottom row) in upper- and deep-layer excitatory neurons. Dots represent significant upregulated genes (red), downregulated genes (blue), and non-DE genes (dark gray) that are associated with cytoplasmic/mitochondrial ribosomal proteins, and all other expressed genes (light gray). (h) Comparison of the C9-ALS vs. control expression fold-changes of cytoplasmic ribosomal protein genes ($n = 87$, top row) and mitochondrial ribosomal protein genes ($n = 72$, bottom row) across the neuronal subtypes and cortical regions. *Two-sided Welch's *t*-test *p*-value 2.21e-15 and 1.38e-13 from left to right, respectively. In each box plot in panels (b), (d) and (h), the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median.


a
NDUFA12
COX7A2
ATP5PF
SOD1
PRDX1
HSP90AA1
CLU
EIF1B
RPS29
KCND3
SLIT3
DNMT3A
snRNA-seq expression
log10 (CPM + 1)
Motor cortex
Frontal cortex
Excitatory neurons
Upper
Intermediate
Deep
C9-ALS
Control
up-regulated in C9-ALS
down-regulated in C9-ALS
b
Protein level (Western blot; arb. units)
HSP90
CLU
KCND3
C9-ALS
Control
c
GO enrichment for up-regulated DE genes
Motor
Frontal
Proteostasis
protein folding
heat shock protein binding
unfolded protein binding
Golgi−associated vesicle
Mitochondrial
respiratory chain
mitochondrial inner membrane
mitochondrial protein complex
nucleoside triphosphate metab. proc.
NADH dehydrogenase complex
NADH dehydrogenase cmplx assembly
proton transmembrane transport
mitochondrial membrane organization
Protein synthesis
protein localization to ER
polysome
translational elongation
Nucleocytoplasmic transport
protein localization to nucleus
RNA localization
DNA damage
DNA damage response
nucleotide−excision repair
Other
antibiotic metabolic process
nucleobase−containing cmpnd.transp.
ribonucleoprotein complex biogenesis
process utilizing autophagic mechanism
microtubule cytoskeleton org.in mitosis
tRNA metabolic process
organelle fission
DNA biosynthetic process
−log10(FDR)
> 15
8.5
< 2
Enrichment ratio
2
6
10
Upper
Intermediate
Deep
VIP
LAMP5
PVALB
SST
Excitatory
Inhibitory
d
Upper
Deep
all genes
DE genes
Δ log2FC
(Motor - Frontal)
e
Excitatory upper
Excitatory deep
GO: Protein folding
HSPA8
CHORDC1
HSP90AA1
DNAJA1
HSPB1
PPID
PTGES3
r = 0.644
p = 1.9e−20
r = 0.639
p = 2.0e−20
GO: Respiratory chain
NDUFB11
NDUFA4
CYCS
UQCRH
COX7C
COX4I1
UQCRB
r = 0.405
p = 5.0e−4
r = 0.562
p = 8.0e−7
GO: Protein localization to endoplasmic reticulum
SRP14
RPS29
RPL28
RPS23
RPL31
RPS15
RPL36
r = 0.119
p = 0.185
r = 0.342
p = 9.8e−5
log2 Fold-change (ALS/Control) Frontal cortex
log2 Fold-change (C9-ALS/Control) Motor cortex
Sig. DE in
Motor
Frontal
Both
Neither
f
Motor cortex
Frontal cortex
log2Fold-change (C9-ALS/Control) Excitatory upper
log2Fold-change (C9-ALS/Control) Excitatory deep
Sig. DE in
Both
Upper
Deep
g
Excitatory upper
Excitatory deep
Motor cortex
Frontal cortex
Cytoplasmic ribosomal protein (RPL/S) genes
Mitochondrial ribosomal protein (MRPL/S) genes
-log10 FDR
log2Fold-change (C9-ALS/Control)
Gene set
up-reg. DEGs associated with RPL/S or MRPL/S
down-reg. DEGs associated with RPL/S or MRPL/s
non-DE genes associated with RPL/S or MRPL/S
all other expressed genes
h
log2 Fold-change (C9-ALS/Control)
Motor
Frontal
Upper
Deep

3.2.5 Disruption of epigenetic regulation in C9-ALS neurons and glial cells

The widespread alterations of gene expression in C9-ALS excitatory neurons, astrocytes, and other cell types raises the question of how such transcriptional changes are established and maintained. We used two complementary assays to determine the epigenetic landscape of C9-ALS brain cells. Single nucleus ATAC-Seq (snATAC-seq) identifies regions of accessible chromatin in single cells[278,279]. We generated 109,198 high-quality snATAC-seq profiles (TSS enrichment ≥ 4, unique fragments ≥ 1,000 per cell, Fig. S3.9), and clustered these to generate pseudo-bulk accessibility profiles for 11 major brain cell types (Fig. S3.10a). The snATAC clusters were annotated by transferring labels from the snRNA-seq data (Table S3.6; Methods). The clusters were not driven by brain region, donor sex, individual samples, or diagnosis (Fig. S3.10b), and they had consistent patterns of accessibility at cell type marker genes (Fig. S3.10c). Similar to our snRNA-seq data, we observed fewer nuclei that passed QC in C9-FTD, and C9-FTD nuclei that passed QC also generally had lower TSS enrichment score and number of unique fragments in all major cell types (Fig. S3.10d-e).

We found 39,830 cell-type-specific snATAC peaks marking regions of accessible chromatin (FDR < 0.05, $\log_2$FC > 0.5 in one cell type vs. the rest; Fig. 3.5a). However, we identified no significant (FDR < 0.05) differential peaks between C9-ALS and control in each major cell type. We also estimated the chromatin accessibility within peaks sharing the same transcription factor (TF) motif with ChromVAR[280] (Fig. S3.11). Again, we observed that the variability of ChromVAR scores across nuclei for TF motifs were mainly contributed by the cell type differences. For example, SOX, SPI and TCF were among the topmost variable TF motifs in terms of ChromVAR scores (Fig. S3.11a), and their accessibility were enriched in oligodendrocytes, microglia, and neurons, respectively (Fig. S3.11b-d).

Although snATAC-seq can in principle provide fine-grained resolution of cell types, the accuracy of clustering and thus the reliability of the pseudo-bulk profiles is limited by the sparse data from individual cells. Indeed, we found that the distinction between neuronal cell types was

not as clear using snATAC-seq data (Fig. S3.10a, c) as it was using snRNA-seq (Fig. 3.1a). We therefore used a complementary strategy, isolating bulk samples of nuclei from four major populations by FANS. Using antibodies against three cell-type-specific nuclear markers, we purified astrocytes, oligodendrocyte lineage cells, microglia, and neurons (Fig. S3.5b). We then profiled active enhancers and promoters marked by histone 3 lysine 27 acetylation (H3K27ac) using ChIP-seq (Fig. S3.5c)[281]. The ChIP-seq data had very high signal/noise ratio, with strong correlation between replicates for the same cell type (r=0.94±0.029, mean ± s.d.) and lower correlation for different cell types (r=0.41±0.084; Fig. 3.5b; Table S3.11). The correlation was equally strong between C9-ALS and control subjects as it was between control subjects. snATAC data had lower inter-replicate correlation for the same cell type (r=0.76±0.16) and higher correlation between cell types (r=0.52±0.11; Fig 5b; Table S3.11).

We observed that the ChIP-seq data correlated strongly with the snATAC-seq tracks from corresponding cell types (r=0.62±0.086), while the correlation for different cell types was lower (r=0.36±0.087; Fig. 3.5b; Table S3.11). Both assays marked cell-type-specific regions of accessible and active regulatory chromatin. The correspondence between H3K27ac and snATAC data at cell-type-specific enhancers was evident at the locus of the ALS risk gene *ERGIC1*[282], where multiple snATAC peaks coincide with H3K27ac peaks (Fig. 3.5c). The ChIP-seq tracks generated from bulk nuclei had deeper coverage than the snATAC-seq data (on average 2.25-fold more unique mapped sequence fragments), and we observed a corresponding higher signal/noise ratio in the H3K27ac tracks (Fig. 3.5c).

We found multiple differentially acetylated (DA) chromatin regions in C9-ALS subjects compared with controls (FDR < 0.05; Table S3.7). There was a large difference in numbers of DA regions between the cell types, with 16 DA regions detected in neurons vs. 2945 in astrocytes. For example, we observed more active chromatin (increased H3K27ac), along with increased gene expression, at the *CD44* locus in astrocytes from C9-ALS compared to control subjects (Fig. 3.5i). This disparity is likely explained by heterogeneity of neuronal subtypes

which precludes the detection of the DA regions in the bulk population. Notably, as stated above,differential accessibility analysis of the snATAC-seq datasets did not reveal significant differences between ALS and control subjects even in astrocytes. We attributed the lack of differential snATAC-seq peaks to the lower signal/noise ratio in these datasets, as well as the statistical burden of multiple comparisons in analysis of the thousands of genome-wide open chromatin regions.

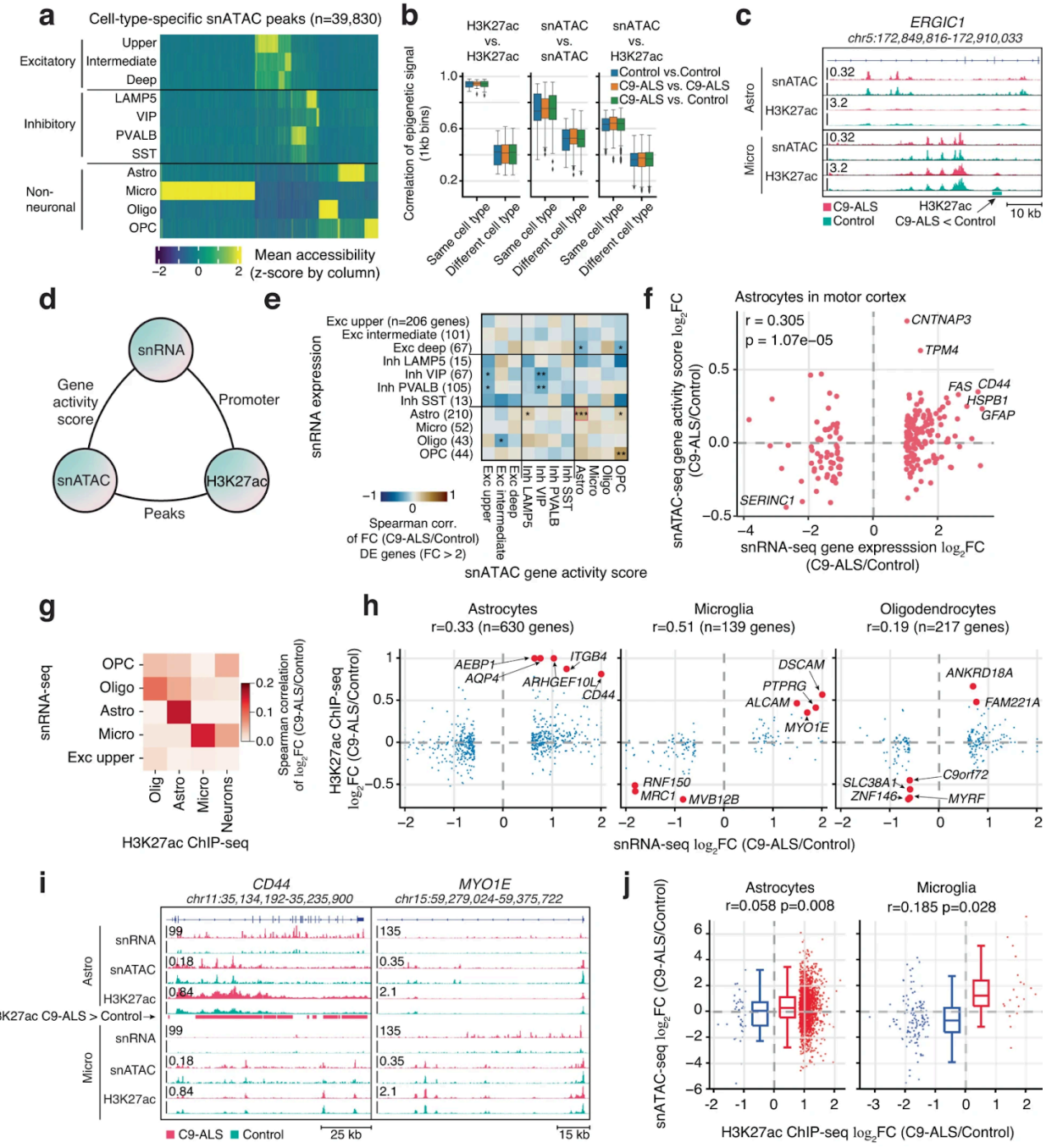
To further characterize patterns of disease-related differences in epigenetic regulation and connect them with transcriptional alterations, we investigated genome-wide correlations among all three data types: H3K27ac ChIP-Seq, snATAC-seq, and snRNA-seq. In particular, we examined the relationship between cell-type-specific signatures of ALS in each pair of datasets using gene expression, gene activity score, and H3K27ac peaks (Fig. 3.5d). To link gene expression with open chromatin (snRNA vs. snATAC), we calculated the gene activity score for strongly DE genes (FC > 2) (Methods). Disease-associated changes in gene expression correlated with open chromatin for astrocytes and OPCs in motor cortex, and in astrocytes and microglia in frontal cortex (FDR < 0.05, Fig. 3.5e, Fig. S3.10f; Table S3.12). Indeed, for astrocytes we found that key upregulated markers including *GFAP, CD44, HSPB1, TPM4,* and *CNTNAP3,* contained more open chromatin in C9-ALS subjects (Fig. 3.5f). By contrast, significantly downregulated genes such as *SERINC1* contained less open chromatin. Notably, we did not observe a significant correlation between snATAC and snRNA signatures of C9-ALS neurons, oligodendrocytes, or microglia, potentially indicating that chromatin accessibility is less affected in those cell types.

Next, we compared differential expression with chromatin activation at the gene promoter (snRNA vs. H3K27ac). This analysis revealed a strong positive correlation for three glial cell types (astrocytes, oligodendrocytes, and microglia; $p < 10^{-5}$, Fig. 3.5g-h). The correlation was especially strong for microglia, even though fewer genes were DE in microglia compared with astrocytes (r=0.51 for microglia and r=0.33 for astrocytes). The strong

consistency of C9-ALS-related differential snRNA and H3K27ac signals in microglia supports the robustness of our observations in these cells. For example, we found that the gene *MYO1E* was transcriptionally upregulated in microglia and had a corresponding enrichment of active chromatin at its promoter (Fig. 3.5h-i). Whereas ALS-related differences in expression were correlated with H3K27ac in a cell-type-specific manner for the non-neuronal cells (Fig. 3.5g), the neuronal ChIP-seq data was less consistent with snRNA, most likely because it represents a mixture of diverse neuronal subtypes.

Finally, we directly compared the two epigenetic data types for the glial cells (Fig. 3.5j), focusing on H3K27ac differentially acetylated peaks. We found that C9-ALS-related changes in chromatin accessibility were positively correlated with H3K27ac in both astrocytes and microglia ($p < 0.05$, Fig. 3.5j).

Figure 3.5 | Epigenetic alterations correlate with transcriptome dysregulation in C9-ALS.
(a) Normalized chromatin accessibility at cell-type-specific snATAC peaks in 11 major cell types identified in snATAC-seq. (b) Pearson correlation of snATAC and H3K27ac ChIP-seq signal in 1 kb genomic bins. Box-and-whisker plots show the distribution of correlations between replicates for the same cell type, or across different cell types. (c) Genome browser view of snATAC-seq and H3K27ac ChIP-seq signals from astrocytes and microglia at the *ERGIC1* locus. Track height represents pseudo-bulk counts normalized by reads in TSS for snATAC-seq, or average signal (counts normalized to one million reads) across donors for ChIP-seq. Bottom track highlights a H3K27ac peak that is significantly reduced in C9-ALS. (d) Schematic of pairwise comparisons (shown in subsequent panels) of C9-ALS effects on the transcriptome (snRNA-seq), chromatin accessibility (snATAC-seq), and histone modification H3K27ac (ChIP-Seq). For each pair of data modalities, we correlated the fold-change of gene- or H3K27ac peak-associated signals. (e, f) Comparison of snRNA with snATAC. (e) Spearman correlation of the C9-ALS vs. control fold-change (FC) for snRNA expression vs. snATAC gene activity score in motor cortex. The analysis was limited to strongly DE genes (FC > 2) in each major cell type in motor cortex. Two-sided Spearman's rank correlation test: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. (f) Scatter plot illustrating the significant correlation between differential gene expression (snRNA-seq) and snATAC-seq changes in astrocytes in motor cortex. Selected genes with high concordant FC are labeled. *r* and *p*, two-sided Spearman's rank correlation test coefficient and *p*-value. (g, h) Comparison of snRNA with H3K27ac ChIP-Seq, showing Spearman correlation (g) and scatter plots (h) of DE gene expression (snRNA-seq) vs. promoter H3K27ac signal (ChIP-seq). Genes with concordant and biggest H3K27ac signal fold-changes were highlighted in red. (i) Browser view of the *CD44* and *MYO1E* loci, showing the correspondence of epigenomic and transcriptomic signals. Track height represents average RPKM across donors for snRNA-seq, pseudo-bulk counts normalized by reads in TSS for snATAC-seq, or average signal (counts normalized to one million reads) across donors for ChIP-seq. Pink rectangles highlight significant H3K27ac differential peaks that were increased in astrocytes from C9-ALS samples. (j) Correlation of snATAC vs. H3K27ac ChIP-seq signal at differential H3K27ac peaks. Box plots show the distribution for upregulated ($N = 2054$ for astrocytes and 20 for microglia) and downregulated peaks ($N = 40$ for astrocytes and 122 for microglia). *r* and *p*, two-sided Spearman's rank correlation test coefficient and *p*-value. In each box plot in panels (b) and (j), the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median.


a
Cell-type-specific snATAC peaks (n=39,830)
Excitatory
Upper
Intermediate
Deep
Inhibitory
LAMP5
VIP
PVALB
SST
Non-neuronal
Astro
Micro
Oligo
OPC
Mean accessibility (z-score by column)
b
H3K27ac vs. H3K27ac
snATAC vs. snATAC
snATAC vs. H3K27ac
Control vs.Control
C9-ALS vs. C9-ALS
C9-ALS vs. Control
Correlation of epigenetic signal (1kb bins)
Same cell type
Different cell type
c
ERGIC1
chr5:172,849,816-172,910,033
Astro
Micro
snATAC
H3K27ac
C9-ALS
Control
H3K27ac C9-ALS < Control
10 kb
d
snRNA
Gene activity score
Promoter
snATAC
H3K27ac
Peaks
e
snRNA expression
Exc upper (n=206 genes)
Exc intermediate (101)
Exc deep (67)
Inh LAMP5 (15)
Inh VIP (67)
Inh PVALB (105)
Inh SST (13)
Astro (210)
Micro (52)
Oligo (43)
OPC (44)
Spearman corr. of FC (C9-ALS/Control) DE genes (FC > 2)
snATAC gene activity score
f
Astrocytes in motor cortex
r = 0.305
p = 1.07e−05
CNTNAP3
TPM4
FAS
CD44
HSPB1
GFAP
SERINC1
snATAC-seq gene activity score log2FC (C9-ALS/Control)
snRNA-seq gene expresssion log2FC (C9-ALS/Control)
g
snRNA-seq
OPC
Oligo
Astro
Micro
Exc upper
Spearman correlation of log2FC (C9-ALS/Control)
Olig
Astro
Micro
Neurons
H3K27ac ChIP-seq
h
Astrocytes
r=0.33 (n=630 genes)
Microglia
r=0.51 (n=139 genes)
Oligodendrocytes
r=0.19 (n=217 genes)
AEBP1
AQP4
ITGB4
ARHGEF10L
CD44
DSCAM
PTPRG
ALCAM
MYO1E
ANKRD18A
FAM221A
RNF150
MRC1
MVB12B
SLC38A1
ZNF146
C9orf72
MYRF
H3K27ac ChIP-seq log2FC (C9-ALS/Control)
snRNA-seq log2FC (C9-ALS/Control)
i
CD44
chr11:35,134,192-35,235,900
MYO1E
chr15:59,279,024-59,375,722
Astro
Micro
snRNA
snATAC
H3K27ac
H3K27ac C9-ALS > Control
C9-ALS
Control
25 kb
15 kb
j
Astrocytes
r=0.058 p=0.008
Microglia
r=0.185 p=0.028
snATAC-seq log2FC (C9-ALS/Control)
H3K27ac ChIP-seq log2FC (C9-ALS/Control)

3.2.6 A distinct signature of glial dysregulation in C9-FTD

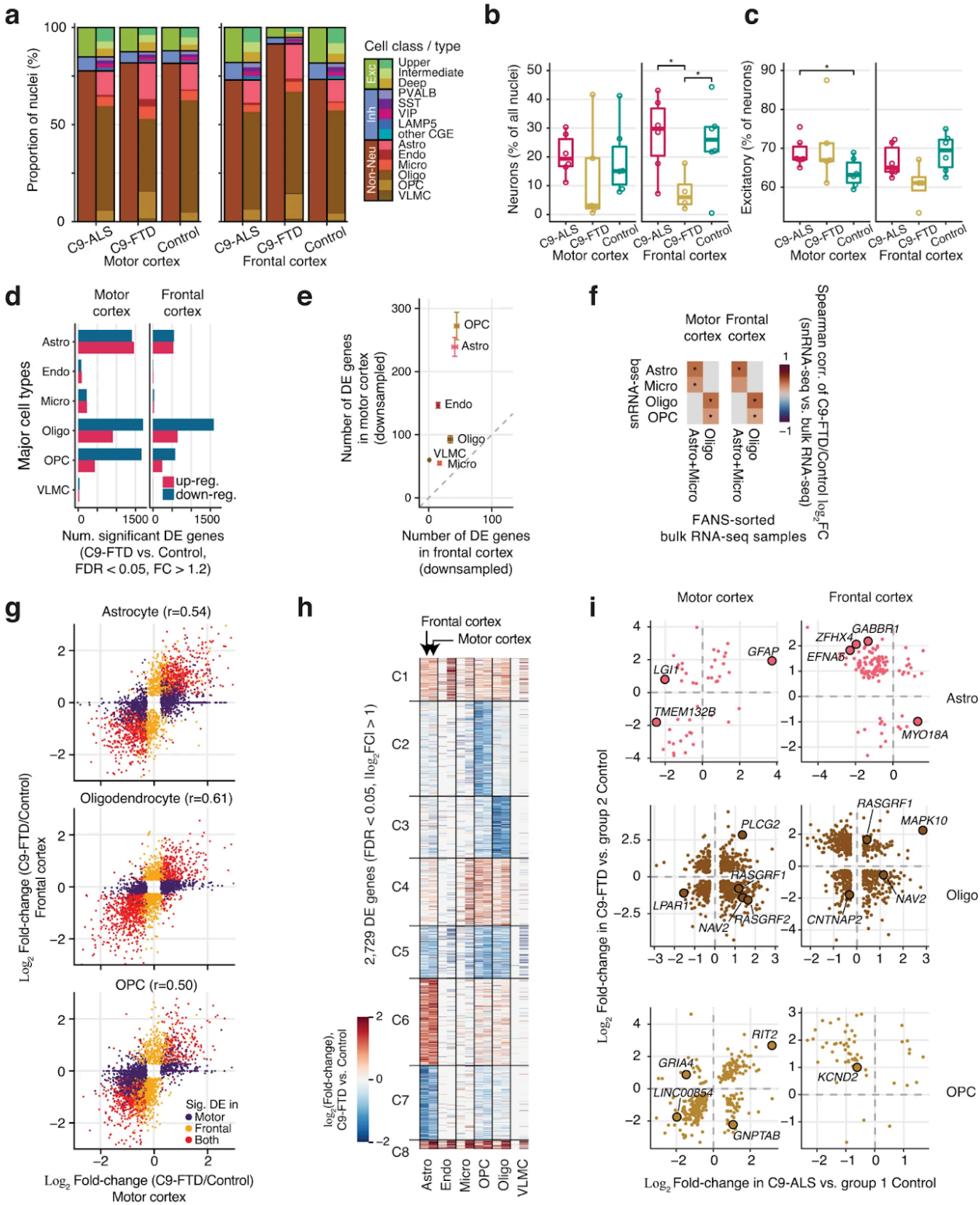
Our initial analyses showed striking differences in snRNA-seq data from C9-FTD compared with C9-ALS, despite the shared genetic risk factor. Notably, the proportion of high-quality excitatory and inhibitory neurons was 2.1-fold lower in frontal cortex in C9-FTD compared with control (Fig. 3.6a-b). No such reduction in the proportion of neurons was observed in C9-ALS. The reduction in high-quality neurons in C9-FTD samples was mainly due to a high proportion of low-quality nuclei with a smaller number of detected genes and a high proportion of mitochondrial RNA (Fig. S3.1a-c; Fig. S3.12a**).** Notably, flow cytometry showed no difference in the overall proportion of neurons (NeuN+ SOX10- cells) in any of the donor groups (Fig. S3.5d), suggesting that the depletion of the frontal cortex C9-FTD neurons in snRNA-seq was due to the lower quality of neuronal nuclear RNA. Neurons were also depleted among high-quality snRNA-seq nuclei in many of the motor cortical samples, but the proportion was more variable and not significantly different from controls on average (Fig. 3.6b). Among neurons, the ratio of excitatory to inhibitory neurons was not significantly different (Fig. 3.6c; Fig. S3.12b). Although our study cannot distinguish between lower tissue quality due to technical factors (e.g., tissue handling) vs. disease effects, the lower proportion of high-quality nuclei that we observed was limited to a specific cell type (neurons), donor population (C9-FTD), and brain region (frontal cortex), which is consistent with the expected pattern of tissue-level pathology in C9-FTD.

The number of high-quality neurons sampled from C9-FTD patients was too low for statistically meaningful analysis of the transcriptome. However, our data enabled high-resolution analysis of the molecular dysregulation in C9-FTD glial cells. We found thousands of DE genes in astrocytes, oligodendrocytes and OPCs, as well as a smaller but significant number of DE genes in microglia (Fig. 3.6d; Table S3.8, S3.9). After downsampling to equalize the statistical power across non-neuronal cell types and brain regions, we found that astrocytes and OPCs were the most strongly affected cell types (Fig. 3.6e). We were surprised to find more DE genes

in glial cells in motor cortex than in frontal cortex, in contrast to the more pronounced depletion of high quality neurons in frontal cortex (Fig. 3.6e). This observation could indicate that glial responses to neuronal degeneration in C9-FTD are distributed through multiple cortical regions. In both motor and frontal cortices, we found strongly correlated differential gene expression fold-changes in the single nucleus and FANS RNA-seq data for corresponding glial cell types (Fig. 3.6f). As in C9-ALS, we found that glial DE genes in C9-FTD had consistent effect sizes in both cortical regions (Fig. 3.6g). We clustered the glial DE genes in C9-FTD, identifying groups of genes specifically affected in oligodendrocytes and OPCs (C2-C5) or in astrocytes (C6, C7) (Fig. 3.6h). These genes exhibited generally consistent effects in both motor and frontal cortices.

We next directly compared the disease effects in glia cells between C9-ALS and C9-FTD. To avoid double dipping, we split the 6 control donors into two groups and ran a second differential test for each glial cell type (Methods). We detected only a small number of shared DE genes between the two diseases (Fig. S3.12c). Also, the most strongly affected DE genes (fold-change > 2) in one disease showed inconsistent disease-associated changes in the other disease (Fig. S3.12d-e). To identify the most pronounced differences, we looked at genes that were DE in both C9-ALS and C9-FTD, and highlighted several genes that were altered in the same or in opposite directions (Fig. 3.6i). We found that *GFAP* was strongly upregulated in motor cortex in both diseases, consistent with activation of astrocytes. On the other hand, the voltage-gated potassium channel regulator protein, *LGI1*, was upregulated in C9-FTD but downregulated in C9-ALS in astrocytes from motor cortex[283]. Notably, clearance of extracellular potassium by astrocytes is essential for controlling neuronal excitability[284]. *NAV2* that encodes Neuron Navigator 2, showed an opposite pattern of expression in oligodendrocytes from frontal cortex. Neuron Navigator 2 was suggested to be involved in CNS development, and the *NAV2* gene has been implicated in Alzheimer's Disease[285,286]. These results suggest a distinct pattern of transcriptome dysregulation in C9-ALS and C9-FTD.

Figure 3.6 | Lower proportion of high-quality neurons and alteration of non-neuronal transcription in C9-FTD.
(a) Distribution of the abundance of three cell classes and 14 major cell types across brain regions and diagnoses. (b, c) Relative abundance of neurons (b), and percent of excitatory neurons among all neurons (c). Circles represent biologically independent individual donors; N = 6 C9-ALS and 6 control samples for both brain regions, and N = 5 and 4 C9-FTD samples from motor cortex and frontal cortex, respectively. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. *Two-sided Welch's *t*-test p = 0.016, 0.041, and 0.039 for the comparisons from left to right, respectively. (d) The number of DE genes in FTD vs. control (FDR < 0.05, FC > 1.2) in glia. (e) Comparison of the number of FTD vs. control DE genes between the motor and frontal cortex after downsampling to 30 nuclei per donor in each cell type. Random downsampling was performed 10 times. Dots and error bars represent mean ± SEM. (f) Spearman correlation between gene expression FC from snRNA-seq vs. FANS-sorted bulk RNA-seq. Correlation was performed using significant DE genes in C9-FTD vs. control identified in snRNA (FDR < 0.05, FC > 1.2). *Spearman correlation test, FDR < 0.05. (g) Transcriptional changes in C9-FTD were consistent in two cortical regions. r, Pearson correlation. (h) K-mean clustering analysis of strongly DE genes in C9-FTD vs. control (>2 fold-change). (i) Comparison of the disease fold-changes in C9-ALS and C9-FTD for astrocytes (Astro), oligodendrocytes (Oligo), and OPC. To avoid double-dipping, the control donors were split into two groups and used to compute the fold-changes for C9-ALS and C9-FTD respectively (see Methods). Only genes that were significantly DE in both comparisons with these split controls are shown in the scatter plot. Highlighted are examples of genes that are also significantly DE in our full model reported in Fig. 3.2a and Fig. 3.6d.


a
Proportion of nuclei (%)
Cell class / type
Exc
Upper
Intermediate
Deep
Inh
PVALB
SST
VIP
LAMP5
other CGE
Non-Neu
Astro
Endo
Micro
Oligo
OPC
VLMC
C9-ALS
C9-FTD
Control
Motor cortex
Frontal cortex
b
Neurons (% of all nuclei)
c
Excitatory (% of neurons)
d
Major cell types
up-reg.
down-reg.
Num. significant DE genes
(C9-FTD vs. Control,
FDR < 0.05, FC > 1.2)
e
Number of DE genes
in motor cortex
(downsampled)
Number of DE genes
in frontal cortex
(downsampled)
f
snRNA-seq
Astro+Micro
FANS-sorted
bulk RNA-seq samples
Spearman corr. of C9-FTD/Control log2FC
(snRNA-seq vs. bulk RNA-seq)
g
Astrocyte (r=0.54)
Oligodendrocyte (r=0.61)
OPC (r=0.50)
Log2 Fold-change (C9-FTD/Control)
Frontal cortex
Sig. DE in
Motor
Frontal
Both
Log2 Fold-change (C9-FTD/Control)
Motor cortex
h
2,729 DE genes (FDR < 0.05, |log2FCl > 1)
log2(Fold-change),
C9-FTD vs. Control
C1
C2
C3
C4
C5
C6
C7
C8
i
Log2 Fold-change in C9-FTD vs. group 2 Control
Log2 Fold-change in C9-ALS vs. group 1 Control
GFAP
LGI1
TMEM132B
ZFHX4
GABBR1
EFNA5
MYO18A
PLCG2
RASGRF1
LPAR1
NAV2
RASGRF2
MAPK10
CNTNAP2
RIT2
GRIA4
LINC00854
GNPTAB
KCND2

3.3 Discussion

Both sporadic and C9-associated ALS and FTD are widely considered to be initiated by degenerative processes in motor or von Economo neurons, respectively[231–233]. Yet, mounting evidence indicates that the broader population of neuronal and non-neuronal cells play key roles in the pathology of these diseases. Here, we used single nucleus transcriptome and epigenome sequencing from autopsied human motor and frontal cortices of cases and controls to identify the specific cell types and the biological pathways that are altered in C9-ALS or C9-FTD. These data, together with bulk RNA-seq and epigenetic (H3K27ac ChIP-seq) assays in FANS-separated nuclei, revealed distinct molecular pathologies in C9-ALS and C9-FTD across cell types and brain regions.

Previous studies have implicated all major classes of glial cells and blood-borne immune cells in the pathology of ALS[237,287–290]. Yet, it remains unclear how these findings, largely based on animal or *in vitro* models of (C9)-ALS, may translate to human ALS patients. Our analysis showed that among non-neuronal cells, astrocytes were the most strongly affected in C9-ALS. We found comparatively fewer gene expression changes in microglia and oligodendrocytes. This finding may be partly explained by the fact that our data reflect the end-of-life transcriptional disruption in C9-ALS, rather than gene expression changes at earlier stages of the disease. We detected transcriptional upregulation of markers of astrocyte activation, and of cytoskeleton and cell-surface protein remodeling in C9-ALS astrocytes. This is consistent with the hypothesis of a common astrocytic response to many forms of injury and neurodegenerative diseases, termed ‘astrogliosis’ or ‘astrocyte reactivity’, which are often accompanied by profound structural changes, including changes in different filament systems and the actin cytoskeleton[245]. Similar to C9-ALS, C9-FTD astrocytes also exhibited upregulation of activation markers (*GFAP, CD44*).

Anatomical studies showed that ALS causes degeneration of the giant upper motor neurons (Betz cells) which are rare cells located in layer 5b of the primary motor cortex and

account for only ~10% of the pyramidal cells in this layer[239]. A recent large-scale study of the human motor cortex identified cells consistent with the size and shape of Betz cells in two Layer 5 extratelencephalic clusters (Exc L3-5 *FEZF2 ASGR2* and Exc L5 *FEZF2 CSN1S1*), but these clusters also included other neurons with pyramidal morphologies[291]. In the absence of specific markers, we were not able to reliably separate a cluster corresponding to Betz cells in our snRNA-seq data. However, all deep layer (L5/6) excitatory neurons are generally considered vulnerable in ALS[236,292], and we found transcriptional dysregulation of L5/6 excitatory neurons which was more extensive in motor than in frontal cortex. Surprisingly, we found even more extensive gene dysregulation in C9-ALS in upper layer (L2/3) excitatory neurons, which were equally afflicted in both cortical regions. This finding is intriguing given that L2/3 neurons, which mainly project intratelencephalically to connect to distant cortical regions and striatum, are greatly expanded as a proportion of the cortical neuron population in human compared with marmoset or mouse[63,293]. The extensive pathological alterations in L2/3 neurons in both motor and frontal cortices could thus contribute to the cognitive and behavioral symptoms in C9-ALS[294].

Whereas genes downregulated in C9-ALS excitatory neurons showed the most significant enrichments for categories associated with neuron-specific functions, probably reflecting neuronal degeneration associated with the disease, the upregulated genes mostly affected mitochondrial function, proteostasis, and protein synthesis pathways. Impaired mitochondrial energy production and functions (e.g., maintenance of Ca^{2+} homeostasis) have been described in many animal and *in vitro* models of ALS, albeit with conflicting results[295]. Our data showed upregulation of the nucleus-encoded mitochondrial genes in deep and upper layer excitatory neurons in both cortical regions. By contrast, recent electron microscopy observations in autopsied motor cortex show decreased mitochondrial density and length in ALS patients[296]. Mitochondria were also decreased in animal and cell models of ALS[297]. We detected no changes for the three mitochondrial proteins that were encoded by the C9-ALS vs. controls DE

genes and that were assessed in our Western blot experiments. Of note, mitochondrial genes are highly expressed in all cell types and our protein assay was performed in bulk brain tissue, precluding the detection of the neuron-specific signal. Therefore, we cannot reliably evaluate the relationship between these transcriptional and protein alterations in the C9-ALS neurons. Still, our findings suggest that upregulation of mitochondrial genes in C9-ALS could reflect adaptation in the surviving mitochondria, at least at the transcriptional level, potentially to satisfy the energy demands of the neuronal cells. It has been suggested that energy demand might be substantially higher in C9-ALS than healthy brains, as the accumulation of DPRs and TDP43 protein aggregates or their pathogenic oligomers in C9-ALS neurons triggers upregulation of multiple systems that maintain cellular homeostasis[298]. In turn, the high energy cost of homeostasis might require an unsustainable increase in energy production. Consistent with this connection, in addition to upregulation of mitochondrial genes, we detected upregulation of proteostasis genes that encode heat shock proteins and their co-chaperones, and we confirmed upregulation of several proteostasis proteins encoded by the DE genes in C9-ALS motor cortex. Our major snRNA-seq findings in C9-ALS are summarized in Fig. 3.7. Our global analyses of gene expression differences across donors showed particularly pronounced disruption of transcriptome in excitatory neurons in C9-ALS and astrocytes in both C9-ALS and C9-FTD (Supplementary Figs. 14-15).

Our epigenomic data, including single nucleus open chromatin (snATAC-seq) and histone modification H3K27ac chromatin immunoprecipitation (ChIP-seq), complement our cell-type-specific transcriptome findings. The high quality of our epigenomic data from multiple donors allowed us to identify significant C9-ALS disease-associated changes of H3K27ac at promoters and distal regulatory elements, which correlated with differential gene expression and chromatin accessibility in specific glial cell types. We further found coordinated effects in C9-ALS on a large number of distal regions marked by open chromatin and H3K27ac. Notably, epigenomic alterations were more pronounced in glial cells compared with neurons, although

this could partly reflect the difficulty of distinguishing cells from diverse neuronal types in sparse snATAC-seq data.

Our study provides a high-resolution view of the transcriptional and epigenetic alterations in individual cell types from the brains of C9-ALS and C9-FTD donors and offers a valuable resource for the scientific community. A limitation of this approach is that we cannot distinguish cell-type-specific alterations that represent direct effects of C9 mutation from end-of-life consequences of neurodegeneration and the resulting disease which are present in the C9-ALS or C9-FTD cases. Also, our methods do not allow for a spatial resolution of the ALS-associated transcriptome dysregulation, which provides valuable information of how this dysregulation participates in the spread of ALS or FTD pathology[299]. Although we confirmed that differential expression of some genes results in the corresponding changes in protein expression, it is possible that some disease-associated mRNA effects are counteracted by homeostatic regulation of translation, justifying future large-scale proteomic studies of ALS and FTD brain. Lastly, a large number of low-quality neurons in C9-FTD patients precluded statistically meaningful analysis of the neuronal transcriptome. Overall, this work demonstrates the value of large-scale single cell/nucleus studies in patients to establish the cell-type-specific molecular pathology of C9-ALS and C9-FTD, which will be essential for developing targeted disease-modifying therapies.

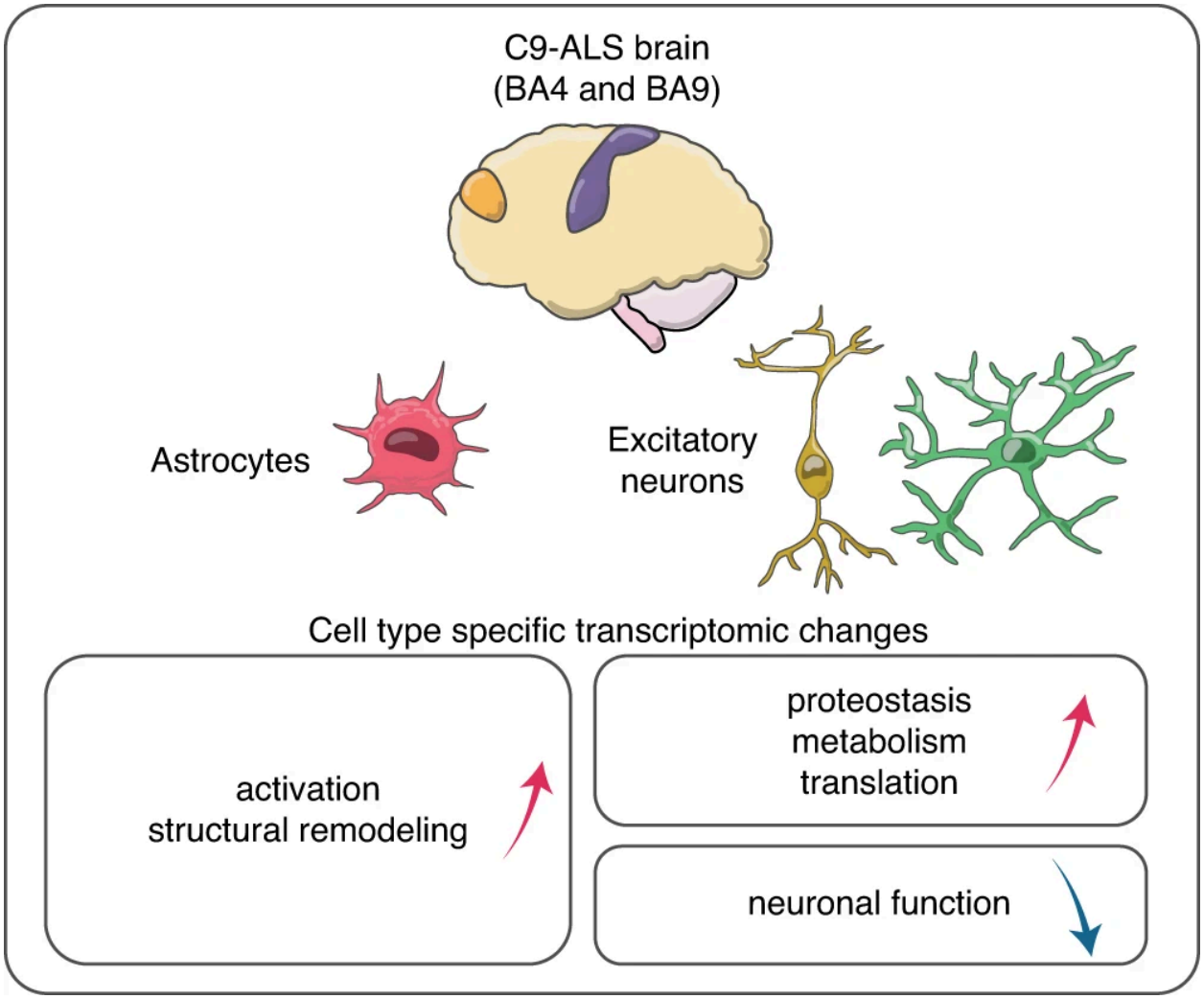


Figure 3.7 | Summary of the major results of our snRNA-seq analysis of postmortem brain samples from C9-ALS donors.
We detected that the most pronounced transcriptional disruption in C9-ALS concentrated in excitatory neurons and astrocytes. These changes were highly consistent in motor and frontal cortices. C9-ALS astrocytes showed increased expression of genes associated with activation and structural remodeling. C9-ALS upper-layer (L2/3) and deep-layer (L5/6) excitatory neurons had increased expression of genes related to proteostasis, metabolism, whereas genes related to neuronal function were downregulated. In both cortical regions, there were more extensive gene dysregulation in upper layer vs. deep-layer excitatory neurons.

3.4 Methods

3.4.1 Human samples

Human frozen postmortem tissues were obtained from the Mayo Clinic Brain Bank (Jacksonville, FL USA). Notably, 3 of the 12 control brains were obtained via the Einstein Aging Study. All brains were acquired with informed consent, and procedures were conducted according to the approved Institutional Review Board protocol (IRB# 15.009452). Data analysis was conducted as exempt human research, considering that these were postmortem brain samples, and they were not specifically collected for this study. The samples were dissected from the dorsolateral prefrontal cortex (Brodmann area 9; BA9; hereafter “frontal cortex”) and/or motor cortex (Brodmann area 4; BA4), which are of particular relevance for ALS and FTD pathological diagnosis, respectively. A total of 17 donors with C9-ALS (N=6), C9-FTD (N=5), or controls determined to have no remarkable pathology (N=6) were used for snRNA-seq, snATAC-seq, bulk RNA-seq in FANS-separated nuclei from major brain cell type, and automated Western blotting studies (Table S3.1). A separate group of C9-ALS (N=6) and normal (no remarkable pathology, N=9) donors was used in H3K27ac ChIP-seq studies (Table S3.1).

All subjects were confirmed negative for protein-coding mutations in *TARDBP, FUS, NEK1, GRN, MAPT,* or *TBK1*. *C9orf72* repeat expansions were confirmed via repeat-primed PCR and Southern blotting. ALS/FTD neuropathologic changes were assessed in both BA4 and BA9 using immunohistochemistry for TAR DNA binding protein 43 (TDP-43). C9-FTD cases presented TDP pathology in PFC with almost no inclusions in motor cortex, along with no Betz cells depletion. In C9-ALS cases, TDP pathology was concentrated in motor cortex with a severe depletion of Betz cells. All C9-ALS cases had clinical features of muscle weakness and atrophy with fasciculations, with at most mild cognitive and memory impairment (3 of 10 cases) or mild dementia (2 of 10 cases) (Supplementary Dataset 1). None of the C9-ALS cases had parkinsonism or significant apathy. Only 1 case had mild depression and 2 cases had some personality changes. In contrast, all 7 FTD cases had dementia with behavioral abnormalities (most often frontal type), but only 2 had mild muscle weakness.

3.4.2 Statistics and Reproducibility

This study was designed to determine transcriptional and epigenetic alterations in the brains of ALS and FTD donors with the repeat expansion in *C9orf72* (C9) gene in single cell resolution. For this purpose, we performed snRNA-seq and snATAC-seq in postmortem samples from C9-ALS (N=6), C9-FTD (N=5) and pathologically normal donors (N=6). To compare the impact across different cortical regions, these studies were performed in motor cortex and frontal cortex, which have been considered to be afflicted the most in ALS and FTD, respectively. For validation, the single nucleus studies were complemented by RNA-seq and epigenetic (H3K27ac ChIP-seq) assays in FANS-separated nuclei from major brain cell types, as well as by automated Western blotting in bulk motor cortex tissue. No statistical methods were used to predetermine sample size because effect sizes were unknown before experiments. The investigators were not blinded to the disease status of the donors. Samples were randomly assigned to batches for nuclear isolation and library preparation. All statistics tests were performed in R (v4.1.2) and Python (v3.7.12), and the details of the tests used in each analysis are provided in the main text, figure legends and the following paragraphs of the Method section.

3.4.3 Nuclei isolation and snRNA/snATAC-seq using the 10X Genomics platform

Nuclei were isolated from 50-70 mg frozen brain tissue, using the previously published protocol[300]. For each donor, brain tissues from the motor cortex and frontal cortex were dissected with a scalpel blade. The tissues were then homogenized on ice in 5mL lysis buffer (0.32 M sucrose, 3 mM $Mg(Ac)_2$, 3 mM $CaCl_2$, 0.1 mM EDTA, 1 mM DTT, 0.1% Triton-X, and 10 mM Tris-HCl pH 8.0 in DEPC-treated water), containing 0.4 U/µl freshly added Recombinant RNase Inhibitor (RRI; Takara, Cat. # 2313A) until no chunks of tissue were visible in the tissue suspension (~20 strokes) using a glass Dounce homogenizer (Thomas Scientific; Cat. # 3431D76; size A). The homogenized tissues were filtered through a 40 µm strainer (Thermofisher; Cat. # 22-363-547) and transferred to a 30 mL thick polycarbonate

ultracentrifuge tube (Beckman Coulter; Cat. # 355631). Nine mL of the sucrose buffer (1.8 M Sucrose, 3 mM $Mg(Ac)_2$, 1 mM DTT, and 10 mM Tris-HCl pH 8.0 in DEPC-treated water) were added to the bottom of the tubes with the tissue homogenate, and the tubes were centrifuged at 112,500 g (25,000 rpm using SW28 rotor) for 2.5 hours at 4°C. The supernatant was aspirated, and the nuclear pellet was submerged on ice for 20 min in 250 uL of DEPC-treated water-based PBS, containing 1% BSA and 0.4 U/μL RRI. The nuclear pellet was then resuspended and filtered twice through a 30 μm cell strainer (Miltenyi Pre-separation Filters; Cat. # 130-041-407). Nuclei were counted using a hemocytometer and diluted to ~2,000 nuclei/μL before performing single-nucleus capture on the 10X Genomics Single-Cell 3' system. The 10X capture and library preparation protocols were used without modification. Matched control, C9-ALS and C9-FTD samples were loaded on the same 10X chip to minimize potential batch effects. Single-nucleus libraries from individual samples were pooled and sequenced on the NovaSeq 6000 instrument (Illumina) with an average depth of ~10,870 unique molecular identifiers (UMIs) per nucleus for snRNA-seq and ~8,223 unique fragments per nucleus for snATAC-seq.

3.4.4 Fluorescence-activated nuclei sorting (FANS) and nuclear RNA-seq of the three populations of the human brain cells (neurons, oligodendrocyte lineage cells, and other glia)

The isolation of brain nuclei prior to the flow cytometry separation was based on the previously published protocol[104]. Tissue (~300 mg) was homogenized in ice-cold lysis buffer (320 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 0.1 mM EDTA, 0.1% Triton X-100, 1 mM DTT, 1 U/μl RRI, and 10mM Tris-HCl, pH 8.0), underlaid with the sucrose buffer (1.8 M Sucrose, 3 mM $Mg(Ac)_2$, 1 mM DTT, 0.4 U/μl RRI, and 10 mM Tris-HCl pH 8.0), and centrifuged for 1h at 98,600 g (24,000 rpm using SW41Ti rotor). The nuclear pellets were then resuspended in the antibody-incubation buffer (0.5% BSA, 3mM $MgCl_2$, 1 U/μl RRI, and 10 mM Tris-HCl, pH 8.0) and incubated with antibodies against NeuN and SOX10 for 1 hr at 4°C. FANS method was then

used to isolate neurons (NeuN+SOX10- population), oligodendrocyte lineage cells (NeuN-SOX10+ population), consisting of mature oligodendrocytes and a smaller population of OPCs, and a third population (NeuN-SOX10-; hereafter “other glia”) that mostly consisted of astrocytes and microglia.

NeuN (also known as RNA-Binding Protein RBFOX3) is a well-established marker of neuronal nuclei[301]. SOX10 is a transcription factor specifically expressed in oligodendrocyte lineage cells. The application of anti-SOX10 antibodies to isolate oligodendrocyte lineage nuclei was described by Frisen and colleagues[302]. We used Alexa488-conjugated anti-NeuN antibodies (1:1000 dilution, Millipore, Cat. # MAB377x) and anti-SOX10 antibodies (R&D Systems, Cat. # AF2864) that were custom conjugated to Alexa647 (1:150 dilution, Cat# FCMAB317PE, Millipore). DNA stain DAPI was used to label intact nuclei (see Fig. S3.5 for details). From each sample, we collected 250-300 thousand (K) nuclei of neurons, 250-300K of oligodendrocyte lineage cells, and 150-200K nuclei of other glia. Nuclei were collected directly to the lysis buffer from PicoPure RNA Isolation Kit (ThermoFisher Scientific, Cat. # KIT024), which was then used for RNA isolation. RNA-seq libraries were prepared with the SMARTer Stranded Total RNA-seq Kit, Pico-Input v2 (Takara, Cat. # 634414) from 10ng of the RNA. Libraries were sequenced on a NovaSeq 6000 instrument (Illumina), using paired-end 100 cycles protocol to an average of 120 million read pairs per sample.

3.4.5 FANS and H3K27ac ChIP-seq of the four populations of the brain cells (neurons, oligodendrocyte lineage, astrocytes, and microglia)

FANS protocol to isolate neurons, oligodendrocyte lineage, astrocytes, and microglia was performed as described in the previous section, except removing RRI from all buffers, adding 0.1 mM benzamidine, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) to the lysis buffer, and adding antibodies against interferon response factor 5 (IRF5). IRF5 is highly enriched in cells of myeloid origin, including microglia[303]. The IRF5 antibodies allowed the separation of

astrocytes and microglia within the other glia (NeuN-OLIG10-) population (see Fig. S3.5 for details). In this FANS protocol, we used PE-conjugated anti-NeuN antibodies (1:1,000 dilution, Millipore, Cat. # FCMAB317PE), anti-SOX10 antibodies (R&D Systems, Cat. # AF2864), which were custom-conjugated to Alexa647 (1:150 dilution), and Alexa488-conjugated anti-IRF5 antibodies (1:200 dilution, R&D Systems, Cat. # IC4508G).

We employed the native ChIP protocol (N-ChIP) in which chromatin fragmentation is performed using micrococcal nuclease (MNase) without crosslinking proteins to DNA[304]. 100-150K of each cell type were collected and used for each ChIP reaction. We used anti-H3K27ac antibodies from Active Motif (Cat# 39133; rabbit polyclonal, 3 µg per sample). ChIP-Seq libraries were prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Cat. # E7645). The resulting libraries were sequenced on a NovaSeq 6000 instrument (Illumina), using paired-end 100 cycles protocol, to an average of 60 million read pairs per sample. For each diagnostic group, three input control samples obtained from MNAse-digested DNA were prepared and sequenced.

3.4.6 Jess/Wes automated, multiplex Western blot assay

Capillary Western analyses were performed using the ProteinSimple Jess-Wes System (San Jose, CA, USA). The following reagents were used: Jess/Wes-Separation Module [2-40kDa Kit (Cat. # SM-W012) 12-230 kDa Kit (Cat. # SM-W004) and 66-440 kDa kit (Cat. # SM-W007)]; Jess/Wes- anti-rabbit, anti-mouse, and anti-goat Detection Module kit (Cat. # DM-001, DM-002 and DM-006, respectively); anti-mouse, anti-rabbit and anti-goat secondary antibodies (Cat. # 042-205, 042-206 and 043-522-2, respectively); antibody diluent (Cat # 042-203); EZ standard Pack 1, 3 and 5, containing biotinylated ladder (molecular weight 12-230 kDa, 66-440 kDa, and 2-40 kDa, respectively); fluorescent (FL) standards, containing 29 kDa (Cat. # PS-ST01EZ), 90 kDa (PS-ST03EZ), and 26 kDa (PS-ST05EZ) system controls; dithiothreitol (DTT); streptavidin-HRP; luminol-S, peroxide; sample buffer (Cat. # 042-195); and

wash buffer (Cat. # 042-202). The Separation Module kit included capillary cartridge and pre-filled microplates. Primary Antibodies used were specific for HSP90 (Mouse, R&D Systems Cat, # MAB3286; Clone # 341320), GFAP (Mouse, Sigma Cat, # G3893), CD44 (Mouse, R&D Systems, Cat. # BBA10; Clone # 2C5), CHI3L1 (Goat, R&D Systems, Cat. # AF2599), HSP70 (Rabbit, R&D Systems, Cat. # AF1663), HSP27 (Rabbit, R&D Systems, Cat # AF1580), RANBP3L (Rabbit, Novus Cat # NBP2-38347), KCND3/Kv4.3 (Rabbit, Novus Cat. # NBP2-76945), DNMT3A (Mouse, Novus Cat # 64B1446), NEFL (Mouse, R&D Cat # MAB22163; Clone #1002615), Clusterin (Mouse, Novus Cat. # MAB2937, clone 3500227), RAP1GAP (Rabbit, Novus Cat. # NBP1-53072), UbiquitinB (Mouse, Novus Cat. # NBP3-07163, clone UBB/1748), TGFB2 (Goat, Novus Cat # AB-112-NA), MAOB (Rabbit, Novus Cat # NBP3-15411, clone 5M8A5), ATP5A (Rabbit, Novus Cat # NBP2-92928), Cytochrome C (CYCS) (Rabbit, Novus Cat # NBP2-21569), and COX5A (Rabbit, Novus Cat # NBP1-32550).

The assay was performed in the motor cortex samples from C9-ALS and control donors (Table S3.1). For each subject, ~ 80-100 mg of tissue were homogenized in lysis buffer (3µl/mg of tissue) [0.32M Sucrose, 5mM CaCl2, 3mM Mg(Ac)2, 150 mM NaCl, 0.1 mM EDTA, 0.1% Triton X100, 3% Igepal.NP-40 (v/v), 1% sodium deoxycholate (w/v), 0.1% SDS (w/v), 1mM PMSF, Protease Inhibitor Cocktail (Sigma, St. Louis, MO; 1:100 v/v), and 10mM Tris (pH 8.0)] using a ME220 focused ultrasonicator (Covaris, Cat # 500506). Burst setting of 20 Sec, power setting of 180 watts were used; the suspension was chilled at 4 °C between ultrasonic bursts. Following homogenization, an additional 0.2mM of PMSF was added to each sample, and the samples were incubated on ice for 30 min. The homogenates were then centrifuged at 10,000 g for 20 min at 4 °C, and the supernatant (whole-tissue extract) was transferred to a new tube. Total protein concentrations of the whole-tissue extracts were measured using Qubit Protein Assay (ThermoFisher Scientific, Cat # Q33211). Samples were stored at −80 °C.

On the day of an experiment, samples (whole-tissue extracts) were defrosted and diluted with 0.1X Sample Buffer (from a 10X stock). The diluted samples were then combined with 5X

Fluorescent (5X FL) Master Mix (containing 5X sample buffer, 5X fluorescent standard, and 200 mM DTT) in a 4:1 ratio and denatured by heating at 95 °C for 5 min using TruTemp DNA Microheating System (Robbins Scientific, Cat # 1057-30-0). The Fluorescent Master Mix contains three fluorescent proteins that serve to normalize the separation distance within each capillary, as the molecular weight ladder is loaded only to the first capillary. Following the denaturation step, the samples, blocking reagent, chemiluminescent substrate (mixture of luminol-S and peroxide in a 1:1 ratio), washing buffer, HRP-conjugated secondary antibodies, and a primary antibody were dispensed into designated wells in the assay plate. Following plate-loading, the samples were subjected to a fully automated capillary separation electrophoresis and immunodetection. One target protein was tested in each assay, and the corresponding primary antibodies were diluted to ensure the linear dynamic range of detection for each protein. The following dilutions were used: 1:5 dilutions for clusterin (CLU), 1:10 dilution for CHI3L1, CD44, HSP90, KCND3/Kv4.3, DNMT3A, RAP1GAP, Ubiquitin B, COX5A and 1:20 dilution for GFAP, HSP27, NRFL, RANBP3L, MAOB, ATP5A, CYCS.

For the statistical analysis, ProteinSimple Compass for SW software was used to extract areas under peak values, which were used to analyze the immunoreactive signals. A two-sided Welch's t-test was used to compare the signals between C9-ALS and control, where the averaged signals between two replicates for each donor were used as observations (6 donors each for C9-ALS and control), and the comparisons with a p-value less than 0.05 were deemed significant.

3.4.7 GFAP immunofluorescence

Frozen human postmortem motor cortex blocks from control and C9-ALS samples were embedded in OCT (TissueTek Sakura) and cryosectioned at 10μm at −20 °C (Thermo Cryostar). Sections were placed onto Superfrost Plus glass slides (Fisher Scientific), dried for 20 min at −20 °C, sealed, and stored at −80 °C until use. Immunofluorescence staining was performed in

accordance with the previously published protocol[305]. Sections were washed in 1x PBS for 60 min, treated with 0.5% sodium borohydride to remove aldehydes, rinsed in PBS, incubated for 10 min in 3% hydrogen peroxide to inhibit endogenous peroxidase, and then incubated in blocking solution (1xPBS, 0.3% Triton-X 100, 10% Normal Goat Serum, 1% Bovine Serum Albumin) for 2 hours at room temperature (RT). This was followed by overnight incubation at 4 °C in blocking solution with rabbit anti-GFAP primary antibodies (1:1,000; Millipore Sigma, Cat. #AB5804). Sections were then washed 3 times in 1x PBS (10 min each) and incubated with goat anti-rabbit IgG Alexa Fluor 594 secondary antibodies (1:500, Invitrogen). Subsequently, sections were incubated in 0.1% Sudan Black in 70% ethanol for 5 min at RT to suppress autofluorescence, washed in Tris buffer saline (pH 7.6) 3 times (15 min each), cover-slipped with ProLong Gold antifade reagent with DAPI (Invitrogen), and seal with nail polish. Fluorescent labeled sections were imaged using 63x oil-immersion objectives. The images were captured on an inverted Zeiss LSM 700 confocal microscope (Zeiss, Germany). Double labeling was confirmed by acquiring z-stacks at 1 to 0.5 µm intervals through cells of interest and by maximum intensity projection (MIP) views. The images were acquired in 2 control and 2 C9-ALS subjects, 3 sections per subject and 3 fields in each section.

3.4.8 snRNA-seq data preprocessing, clustering and cell type annotation

snRNA-seq reads were mapped to the human GRCh38 genome with a custom pre-mRNA annotation (modified RefSef gene annotation where each gene transcript locus was listed as an exon) using 10x Genomics Cell Ranger (v3.0.2)[306] with the default parameters. The raw feature-by-droplet matrices generated from the “cellranger count” module were then fed to CellBender (v0.2.0)[307] to distinguish cell-containing from cell-free droplets and retrieve noise-free cell-by-gene quantification tables using the following parameters: “--expected-cells 1500 --total-droplets-included 18000 --model full --epochs 150 --cuda --low-count-threshold 15 --fpr 0.01 --learning-rate 0.0001 --posterior-batch-size 5 --cells-posterior-reg-calc 50”.

Scrublet (v0.2.3)[158] was then used in the CellBender-generated h5 file for each sample to compute the doublet score for each nucleus with an expected doublet rate of 0.06. Based on the simulated doublet histogram of doublet scores, we set the doublet score cutoff to 0.2 and this threshold worked well for all samples.

Next, Seurat (v4.0.4)[308] was used to create a merged object containing all the samples, and to perform the following downstream processing. Nuclei with fewer than 500 detected genes, more than 1% reads that mapped to the mitochondrial genome, or a doublet score greater than 0.2 were excluded from further analysis. Sctransform[309] normalization was performed on the remaining nuclei, with the mitochondrial mapping percentage and the sequencing batches set as the confounding sources of variation to be removed. Principal component analysis (PCA) was run with the top 3000 highly variable genes in the normalized nuclei-by-gene matrix, and the first 30 principal components were used to run Harmony (v0.1.0)[310] integration with sample IDs, donor IDs and sequencing batches set as covariants. Then, a k-nearest-neighbor graph was constructed with k set as 30, and the nuclei were clustered using Leiden clustering with a resolution of 1. This gave rise to 29 clusters, and UMAP (Uniform Manifold Approximation and Projection)[311] was used for visualization (Fig. 3.1a, the leftmost panel).

The nuclei were then classified into three classes (excitatory neurons, inhibitory neurons, and non-neurons) using several well-established marker genes (Fig. S3.3a). For each of these three classes, a subset of nuclei was created from the full dataset, and the aforementioned Sctransform normalization, PCA, Harmony integration and Leiden clustering steps were run on each of these subsets. This iterative clustering process reselects the highly variable genes in the context of each cell class and hence may give better resolution in finding distinct subclusters. We used SC3 stability index[312] and Clustree (v0.5.0)[313] to evaluate and visualize the stability of the clusters across a range of resolutions for the Leiden community detection analysis (Fig. S3.2f). With the Leiden resolution set as 1.5, 0.5 and 1, we identified 24, 23, and

23 subclusters for excitatory neurons, inhibitory neurons and non-neurons, respectively. These subclusters were annotated using the expression of selected marked genes (Fig. S3.3b-d). Note that six subclusters seemed to express marker genes of multiple cell classes (for example an excitatory subcluster with expressions of oligodendrocyte marker genes), which were potential doublets that failed to be captured by the Scrublet pipeline. We labeled these subclusters as "ambiguous" and removed them from further consideration. We also noticed one excitatory neuron subcluster that highly expressed the *TUBB2A* gene but had a low number of detected genes, and one non-neuron subcluster that expressed marker genes of immune cells (but no expression of microglia marker genes). We also excluded these two subclusters from the downstream analysis. After annotation, we identified 49 distinct subpopulations of excitatory neurons, inhibitory neurons and non-neuronal cells. To estimate how well our annotation matched with the cell types annotated in the recent human motor cortex single nuclei study by the Allen Institute for Brain Science[63], we used Seurat's "FindTransferAnchors" and "TransferData" functions to transfer the annotation labels from the AIBS reference dataset to our dataset. A confusion matrix was constructed with the prediction labels (label with the highest prediction score) from the data transfer and the labels from our annotation, and the adjusted rand index was calculated to measure the similarity between the two (Fig. S3.4). Cell type nomenclature mostly followed the common cell type nomenclature (CCN) convention presented by the Allen Institute for Brain Science[63,314], with slight modifications to reflect the specific marker genes found in our dataset.

3.4.9 Differential expression analysis in snRNA-seq

We first performed unsupervised analysis to characterize the similarity of samples from each patient in an unbiased manner. We created pseudobulk counts with nuclei originating from each donor for each cell type, computed normalized counts with DEseq2's median of ratios method (v1.36.0)[315], and performed PCA separately for each cell type and each brain region

using the log normalized counts (Fig. S3.14). Although the number of subjects is limited, we did observe a separation of the disease and control samples, especially in astrocytes and excitatory neurons. The inter-individual variability of gene expression was greater in some cell types than others. Second, we used Augur (v1.0.3)[111] to quantify the separability of nuclei from different diagnoses. Augur employs a machine-learning framework to quantify the separability within a high-dimensional space. For each cell type, Augur withholds a proportion of sample labels and trains a random forest classifier on the labeled subset. The classifier predictions are compared with the experimental labels, and cell types are prioritized on the basis of the area under the receiver operating characteristic (AUC) of these predictions in cross-validation.This analysis showed that we can distinguish nuclei from disease or control samples using the gene expression profile, with the highest AUC of ~0.78 achieved in astrocytes for C9-ALS vs. control (Fig. S3.15a). Furthermore, for the C9-ALS vs. control analysis, we generated random shuffles to check how much better these observed AUCs would perform than the shuffles. We permuted the diagnosis labels of the individual donors (so that all nuclei from each donor will still have the same diagnosis labels), and required that the two shuffled groups must each contain three C9-ALS donors and three Control donors, of which three are male and three are female. This resulted in 82 possible permutations, and we ran Augur to get the AUCs for distinguishing the two groups. We found that in motor cortex the AUCs from the real diagnosis labels are higher than the 95% upper confidence intervals in astrocytes, excitatory neurons and VIP inhibitory neurons. Astrocytes and deep-layer excitatory neurons also have higher AUCs in the observed than those in the shuffles in the frontal cortex samples (Fig. S3.15b). Last, we applied linear discriminant analysis (Fig. S3.15c) to quantify the discriminability of C9-ALS cells vs. control cells. We separated the data into training and test sets based on the donor, using 10 donors for training and 2 for testing (one each from C9-ALS and control). We trained a LDA classifier using the top 1-10 principal components for each cell type and brain region. The results showed that

C9-ALS cells could be meaningfully distinguished from control using the top few principal components in deep and superficial excitatory neurons, and in astrocytes.

Next, we performed the traditional differential expression analysis with the labels of diagnosis. To ensure a statistically meaningful number of cells from each donor for the differential expression analysis, we grouped the subpopulations into 14 major cell types with shared marker genes. In these major cell types, there were at least 15 high-quality nuclei in at least 4 out of the 6 donors in each group (the combination of diagnosis and brain region). For the comparison between C9-FTD and control samples, we focused on the non-neuronal cell types as the C9-FTD samples showed depletion in the numbers of neuronal nuclei (Fig. 3.5a). To identify the differentially expressed genes (DEGs) between disease (C9-ALS or C9-FTD) and control across these major cell types in each brain region, we used the model-based analysis of single-cell transcriptomes (MAST, v1.18.0)[242], where a mixed-effect hurdle model was employed to model the snRNA expression data as a mixture of a binomial and normal distribution while systematically accounting for predefined covariates. For each comparison, the raw counts of chosen nuclei in question were extracted from the "RNA" slot of the Seurat R object and then normalized to $\log_2$CPM (counts per million). Genes that were expressed in at least 10% of the selected nuclei were kept. The following linear mixed model was then fit with MAST:

$$y \sim D + G + U + M + A + S + B + (1|I) \qquad (1)$$

Here, y is the $\log_2$-normalized count of the gene; D is the diagnosis (ALS, FTD or control); G is the number of genes detected, U is the number of UMIs; M is the percentage of reads mapped to mitochondrial genes; A is the age of the donor; S is the sex of the donor; B is the sequencing batch of the sample; and I is the ID of the donor. G, U, M and A were centered and scaled across the selected nuclei in comparison. The donor ID I was modeled as a random-effect term while the rest of the covariates were modeled as fixed-effect terms.

Next, a likelihood ratio test (LRT) was performed to identify DEGs by comparing the model with and without the diagnosis term. Hurdle p-values were reported by MAST, and

Benjamini & Hochberg's false discovery rate (FDR) method was used to adjust p-values for multiple comparisons. The MAST model also reported fold-changes (FC) due to the disease effect (hereafter referred to as model FC), which were contributed by both the continuous component (nonzero expression) and the discrete component (expressed or not) of the hurdle model. We also computed more straightforward fold changes between the groups (hereafter referred to as average FC) by subtracting the mean $\log_2$CPM of selected nuclei in the control sample from the mean $\log_2$CPM of nuclei in the disease sample. In most cases these two fold-changes were concordant, and we excluded genes with extremely high model FC that deviated from the average FC. Genes were defined as significantly differentially expressed if they passed all of the following criteria: the FDR was less than 0.05; the absolute model FC was greater than 20%; both the continuous and the discrete fits of the hurdle model were convergent; the upper and lower bounds of the 95% confidence model had the same sign; the difference between the $\log_2$(model FC) and $\log_2$(average FC) was less than 2.

To control for the number of nuclei when comparing the disease effect across the major cell types, we randomly downsampled the dataset to 30 nuclei per donor in each cell type, and performed the same MAST test as described above. The downsample analysis was done 10 times, and the number of significant DEGs for each cell type was reported as mean ± SEM (standard error of the mean).

To avoid double dipping when comparing the disease effect in non-neuronal cells between C9-ALS and C9-FTD, we split the 6 control donors into two groups (group 1: C1, C2 and C3, two males and one female; group 2: C4, C5 and C6, one male and two females). The same MAST test was run as described above, except that the group 1 control donors were used in the C9-ALS vs. control comparison while group 2 control donors were used in the C9-FTD vs. control comparison. The overlaps of significant DEGs between the two comparisons were reported, and the correlations of model fold-changes were estimated using the Pearson correlation coefficient.

3.4.10 Gene Ontology (GO) enrichment test

GO enrichment analysis was done using WebGestalt (v2019)[249]. Up- and Down-regulated DEGs identified in each cell type and each comparison were used separately as input, and all the expressed genes in the corresponding group were set as the reference gene set (background). GO terms in the functional databases “Biological Process noRedundant”, “Cellular Component noRedundant” and “Molecular Function noRedundant” were used, and only terms with 25 to 500 genes were included in the over-representation analysis. Significantly enriched GO terms (FDR < 0.05) were then clustered by Affinity Propagation to further remove reductant terms.

To compare the effect size of differential expression between motor cortex and frontal cortex, we calculated the fold-change difference ($\Delta \log_2$FC) between the two brain regions for genes associated with each GO term of interest. The $\Delta \log_2$FC for all expressed genes (labeled as “all genes” in Fig. 3.4d and Fig. S3.8a) and all DE genes (labeled as “DE genes”) were also computed and used as baseline control. The two-sided Welch’s t-test was used to test whether the $\Delta \log_2$FC in each group of genes were significantly different from the $\Delta \log_2$FC of the “all genes” control set.

3.4.11 Bulk FANS-sorted nuclear RNA-seq data processing

FastQC (v0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to examine the quality of the RNA-seq reads. Reads were then trimmed to remove sequencing adapters and low-quality sequences (minimum Phred score 20) using Trim Galore (v0.5.0, a wrapper tool powered by Cutadapt[152]) in paired-end mode. The first 3 bp from the 5′ end of read 1 were also removed. Trimmed reads were then mapped to the human hg38 genome and the GENCODE annotated transcriptome (release V35) with STAR (Spliced Transcripts Alignment to a Reference, v 2.7.1a)[316]. Gene expression was estimated using RSEM (RNA-Seq by

Expectation Maximization, v1.2.30)[317]. Gene-level 'expected count' from the RSEM results were rounded and fed into edgeR (v3.34.1)[318] to perform differential expression tests separately for each sorted population in each brain region. Only genes that were expressed (with CPM > 2) in at least six samples were kept. The read counts of the remaining genes were then normalized using the TMM method[319], and DE genes were called in the quasi-likelihood F-test mode with a cutoff of FDR < 0.05. No DE genes were detected at this FDR cutoff, and we only used the fold-changes in the bulk RNA-seq data to perform the correlation test with the fold-changes in snRNA-seq.

3.4.12 snATAC-seq data preprocessing, clustering and cell type annotation

snATAC-seq reads were mapped to the human GRCh38 genome using 10x Genomics Cell Ranger ATAC (v1.1.0)[320] with the default parameters. Barcode multiplets[321] were removed using the "clean_barcode_multiplets_1.1.py" script provided by 10x Genomics. The output fragments files were then imported into ArchR (v1.0.2)[322] to create Arrow files and an "ArchRProject" object for the downstream analysis. Fragment size distribution in each sample was inspected for nucleosomal periodicity, and nuclei with TSS (Transcription Start Site) enrichment score less than 4 and a number of unique fragments less than 1000 were removed. Doublet inference and removal were performed using the "addDoubletScores" and "filterDoublets" functions in ArchR. A cell-by-tile matrix containing insertion counts across genome-wide 500-bp bins was created, and dimensionality reduction was applied to the matrix using the iterative Latent Semantic Indexing (LSI) implemented in ArchR (two iterations with the top 25,000 features). The top 30 dimensions after LSI were used to perform Louvain clustering (with resolution set as 2) and UMAP visualization.

Next, a cell-by-gene matrix was computed with the gene activity score model implemented by ArchR. The model accounted for both the accessibility within the entire gene body and the activity of putative distal regulatory elements using an exponential weighting

function that depends on the distance between the insertions and the TSS of the gene of interest, and gene boundaries were imposed to minimize the contribution of unrelated regulatory elements to the gene activity score. Gene activity scores of well-defined cell class marker genes were used to classify the clusters into excitatory neurons, inhibitory neurons and non-neurons. A subset of nuclei was created for each of these three classes, and a similar LSI and clustering pipeline was run to identify subclusters in each category. For visualization, the "addImputeWeights" function was used to impute gene activity scores by smoothing signals across nearby cells based on a MAGIC[323] diffusion matrix.

To annotate these subclusters, we integrated the gene activity score matrix in the snATAC dataset with the gene expression matrix in our snRNA dataset using the "addGeneIntegrationMatrix" function. The integration was done in a constrained way, where subclusters from the three cell classes in the snATAC dataset were only allowed to align to the subclusters with matching cell classes in the snRNA dataset. The annotation labels from the 14 major snRNA cell types were transferred to each snATAC nuclei with a prediction score to represent the assignment accuracy. The final annotations of the snATAC subclusters were determined by first filtering out nuclei with a prediction score less than 0.7 and then taking the transferred label of 70% supermajority in remaining nuclei for each cluster. Subclusters that failed to reach the supermajority were labeled as "Mixed" and removed from further consideration. This resulted in 109,198 high-quality snATAC nuclei of 11 major brain cell types.

3.4.13 snATAC-seq peak calling

To call peaks in each group of interest (the combination of major cell types, brain regions and diagnosis), we first created pseudo-bulk replicates using the "addGroupCoverages" function in ArchR with the following settings: the minimum and maximum numbers of replicates were set to 6 and 28; the minimum and maximum numbers of nuclei per replicate were set to 50 and 500; the sampling ratio to use if a particular group lacks sufficient cells to make the desired replicates

was set to 0.8. Next, the “addReproduciblePeakSet” in ArchR was run with MACS2 (v2.2.7.1)[324] set as the peak calling method, where the iterative overlap peak merging procedure was performed to generate a single merged peak set of fixed-width (501 bp), reproducible peaks that can be called in at least two samples. A cell-by-peak count matrix was then computed with the “addPeakMatrix” function. The counts were normalized by the number of reads in TSS across cells prior to performing marker feature identification, and the cell-type-specific peaks were identified using the Wilcoxon test in the “getMarkerFeatures” function for each major cell type, adjusting for the potential bias introduced by the number of unique fragments and TSS enrichment.

3.4.14 Differential chromatin accessibility between disease and control

Differential chromatin accessibility between disease and control was accessed in two ways. First, we ran gene-based pair-wise differential tests with the gene activity scores for each major cell type using the Wilcoxon test in the “getMarkerFeatures” function, accounting for bias of the number of unique fragments and TSS enrichment. Genes with FDR less than 0.05 were called significant. In addition, peak-based differential tests were performed to identify differentially accessible regions (DARs) between disease and control for each major cell type. The cell-by-peak matrix was normalized by the number of reads in TSS across cells and DARs were identified using the Wilcoxon test in the “getMarkerFeatures” function, adjusting for the potential bias introduced by the number of unique fragments and TSS enrichment. No significant DARs were found with an FDR threshold of 0.05.

3.4.15 Accessibility of transcription factor motif

ChromVAR[280] was used to estimate chromatin accessibility within peaks sharing the same transcription factor (TF) motif while controlling for technical biases. First, the non-redundant transcription factor (TF) motif archetypes (v2.0-beta,

https://github.com/jvierstra/motif-clustering)[325] were used to scan and annotate the peaks with the “addMotifAnnotations” function in ArchR. Background peaks were selected based on similarity in GC content and the number of fragments across all samples using the “addBgdPeaks” function. The chromVAR deviations and z scores per cell were then computed for each TF motif using the “addDevationsMatrix” function. To find TF motifs with differential chromVAR scores between disease and control in each major cell type, t-test and Brown-Forsythe test were performed on the deviations and z scores, respectively. Motifs with FDR < 0.05 were called significant. Motif sequence logos were drawn using the “ceqlogo” module in the MEME Suite (v5.4.1)[326].

3.4.16 H3K27ac ChIP-seq data processing

Raw sequencing data were pre-processed to remove adapters and low-quality sequences with the HTStream tool (https://github.com/s4hts/HTStream). Reads were mapped to the human hg38 genome build with BWA-MEM2[327] and filtered to remove multi-mapping reads and low-quality alignments using Samtools[328]. Reads mapping to ENCODE-blacklisted genomic regions were excluded using BEDTools[329]. H3K27ac-enriched peaks were detected using MACS2, including input controls for each cell type and condition, as previously described[330]. Promoter (2 kb upstream and 1 kb downstream from TSS) H3K27ac signal was computed using the “multiBigwigSummary” module from deepTools (v.3.3.1)[331]. Differential peaks between C9-ALS and control for each FANS-sort population were identified using DiffBind[332] in edgeR[318] mode. Peaks with FDR < 0.05 were called significant.

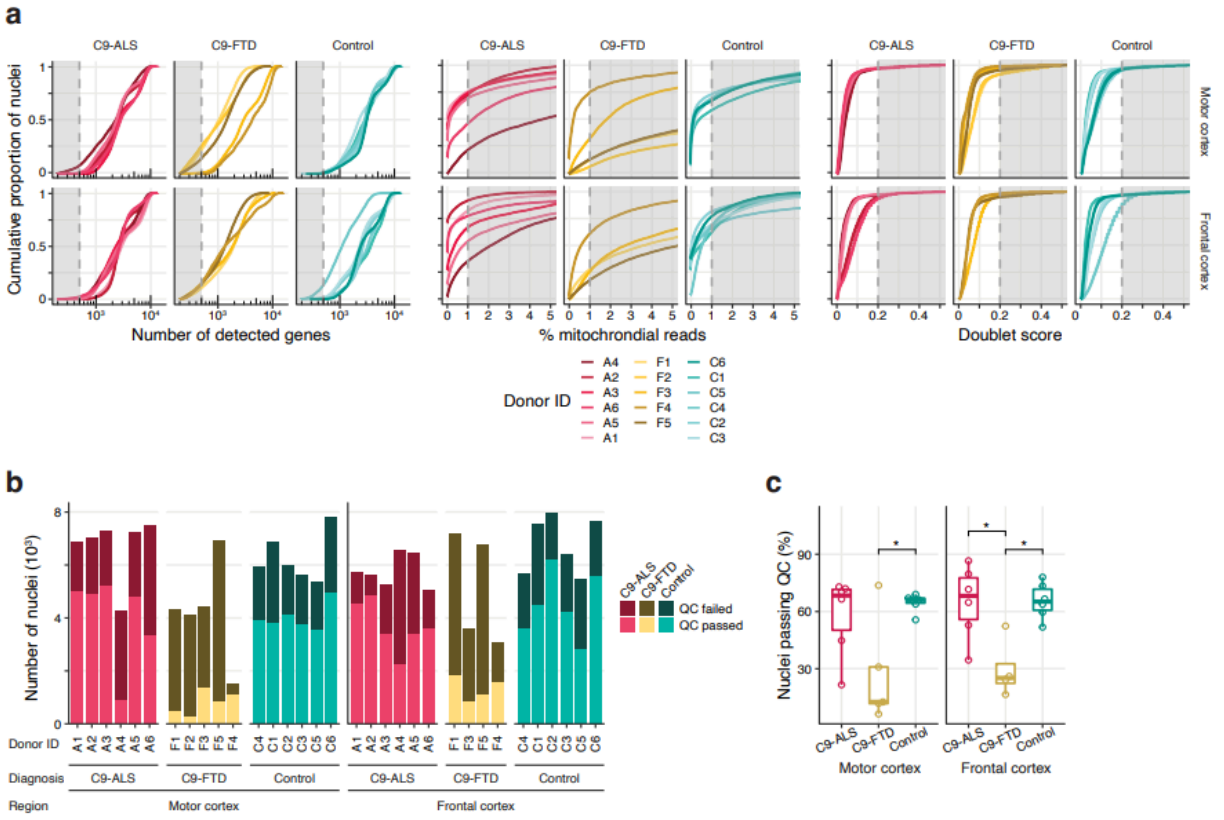
3.5 Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in *Nature Communications* 2023. Li, J., Jaiswal, M. K., Chien, J.-F., Kozlenkov, A., Jung, J., Zhou, P., Gardashli, M., Pregent, L. J., Engelberg-Cook, E., Dickson, D. W., Belzil, V. V., Mukamel, E. A. & Dracheva, S.

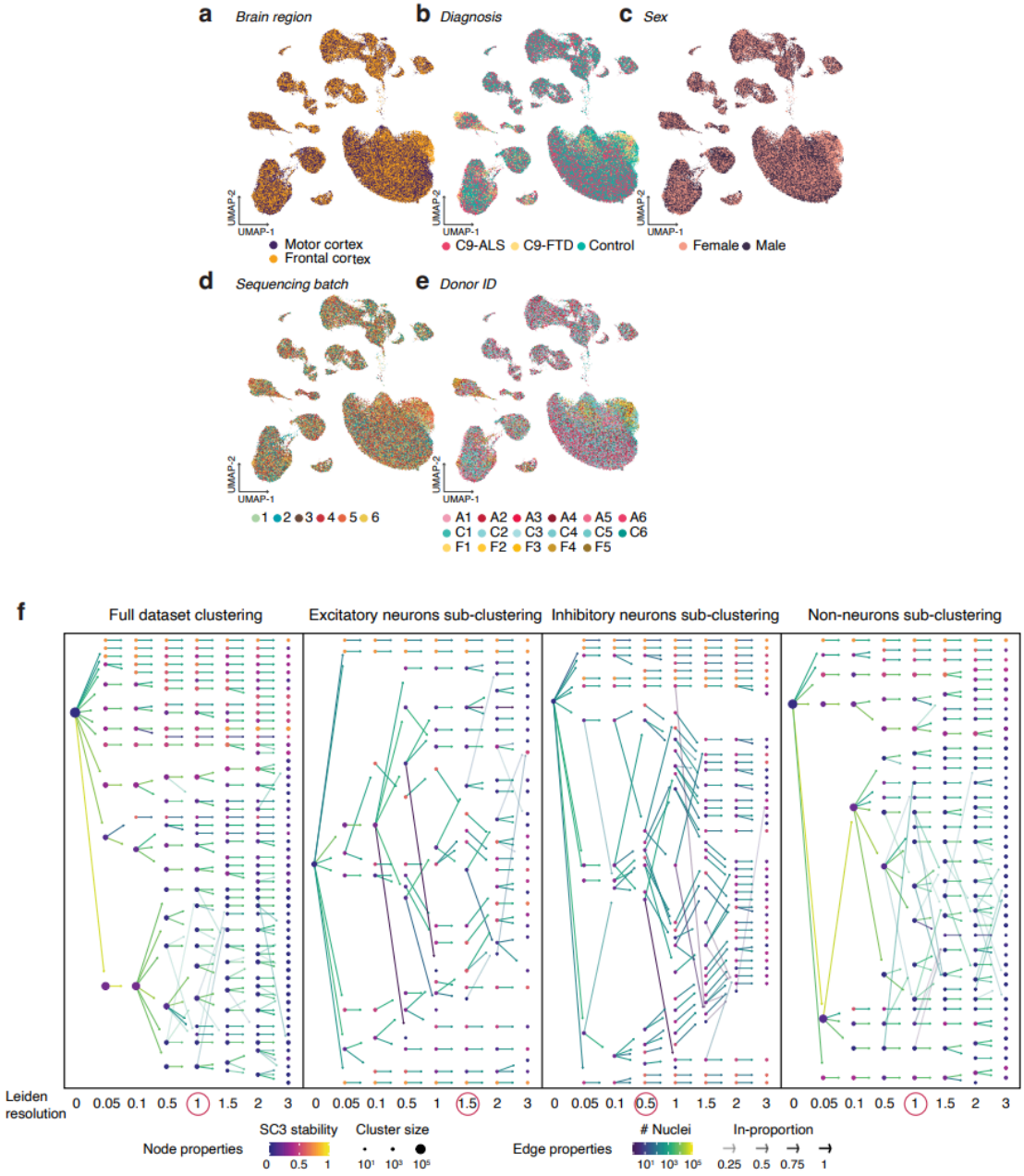
Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with C9orf72 mutation. *Nat. Commun.* **14**, 5714 (2023). The dissertation author was the researcher and author of this paper.

The study was conceived and designed by JL, VVB, EAM and SD. JL performed computational analyses of all sequencing data, and generated the majority of figures and tables. MKJ performed snRNA-seq, snATAC-seq, FANS-sorted bulk RNA-seq, ChIP-seq, and GFAP immunofluorescence experiments, and generated the GFAP immunofluorescence figure. JC participated in computational analysis, performed GO analyses and generated GO-related figures. AK designed and performed FANS-sorted ChIP-seq experiments, and processed the ChIP-seq data. MKJ and JJ performed and analyzed digital Western Blot validation experiments. PZ assisted in single cell and FANS-sorted sequencing experiments. MG and LJP performed dissection of the motor and prefrontal cortices. DWD and EE-C performed cases chart review to provide clinical information. VVB and DWD provided biospecimens. JL, JC and EAM designed and built the data browser. JL, SD, and EAM interpreted the results and wrote the original draft of the manuscript. SD and EAM supervised and provided funding for the study. All authors read, edited and approved the final manuscript.

We gratefully acknowledge Eugene Koonin for helpful comments, and Ramu Vadukapuram for performing nuclear sorting for ChIP-seq studies. This work was supported by VA Merit Awards BX003625 and BX005585 (S.D.); PsychEncode consortium NIH/National Institute of Mental Health R01MH122590 (former U01MH122590) (S.D.) and R01MH122592 (former U01MH122592) (E.A.M.); R01AG067151 (V.V.B). We gratefully acknowledge Mayo Clinic for its financial support of “The Mayo Clinic Brain Bank”. We also gratefully acknowledge Drs. Boylan and Oskarsson, who are clinicians who were essential in obtaining C9-ALS autopsies and the Einstein Aging Study (P01AG003949) for providing tissue of 3 of the 11 control samples in our study.

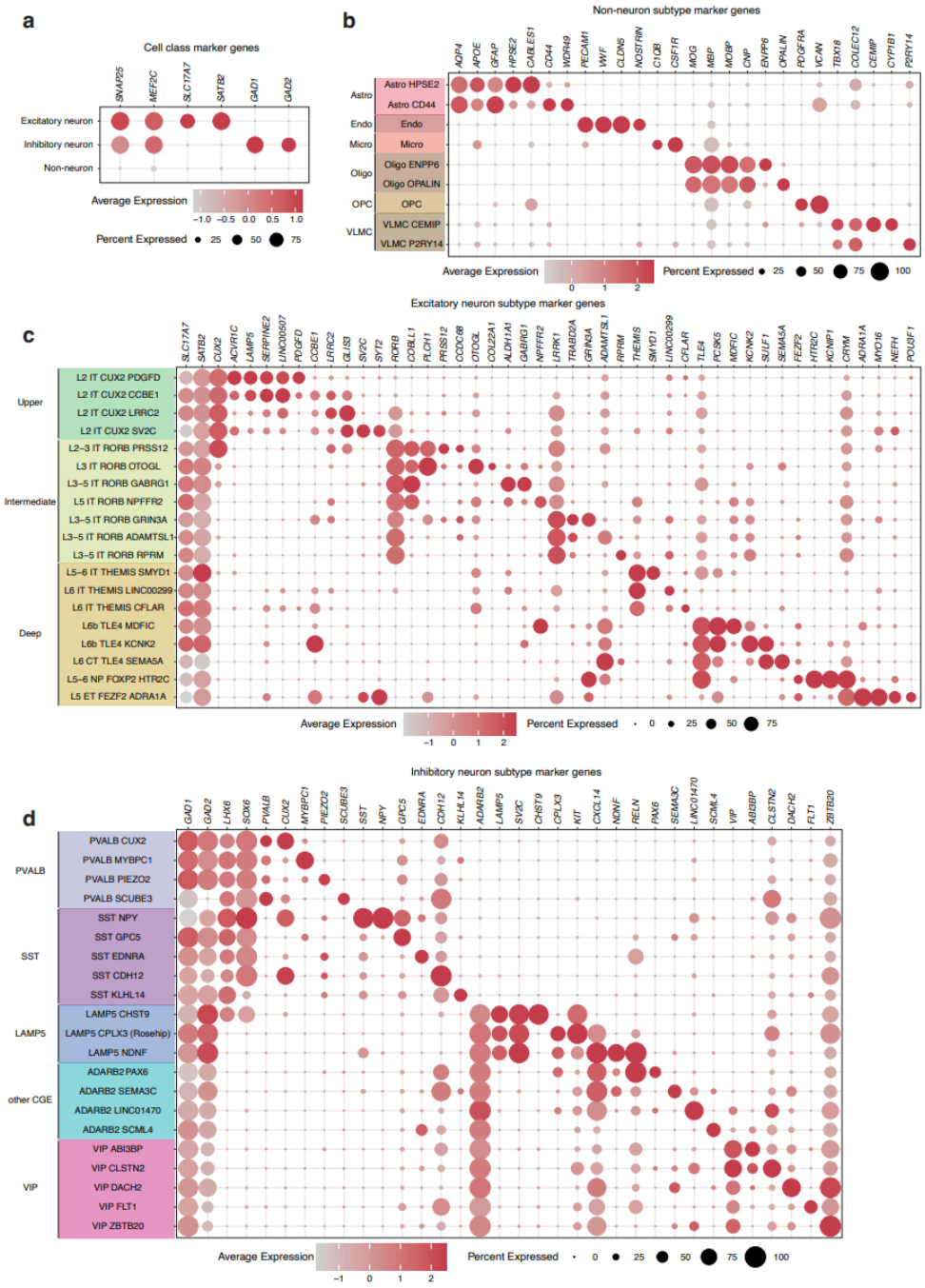


Supplemental Figure 3.1 | Quality control (QC) metrics for snRNA-seq data, related to Figure 3.1. (a) Cumulative proportion of the number of detected genes, percentage of reads mapped to the mitochondrial genome, and doublet score for nuclei in each snRNA-seq sample, grouped by donor ID, diagnosis and brain region. Shaded areas in gray mark the nuclei that failed the QC filtering (number of detected genes > 500, % mitochondrial reads < 1, or doublet score < 0.2). Nuclei in the C9-FTD samples failed QC mainly due to high percentage of mitochondrial reads. (b) Summary of the total number of collected nuclei, and the number that passed QC, in each sample. (c) Proportion of nuclei that passed QC. Circles represent biologically independent individual donors; N = 6 C9-ALS and 6 control samples for both brain regions, and N = 5 and 4 C9-FTD samples in motor cortex and frontal cortex, respectively. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. *, two-sided Welch's t-test p = 0.038, 0.014, and 0.012 for the comparisons from left to right, respectively.

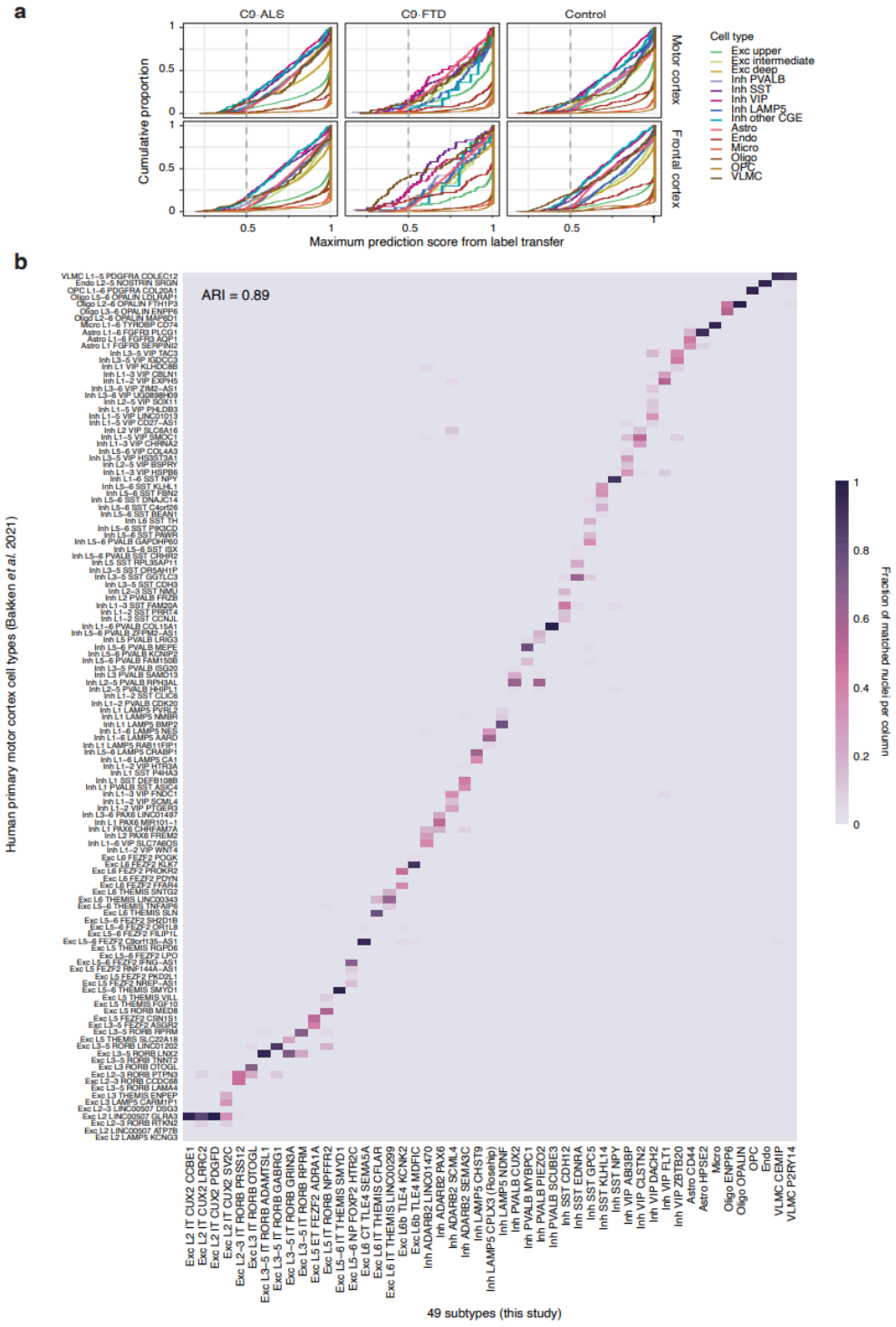


Supplemental Figure 3.2 | The clusters identified in snRNA-seq were not biased by known covariates and were largely stable, related to Figure 3.1.

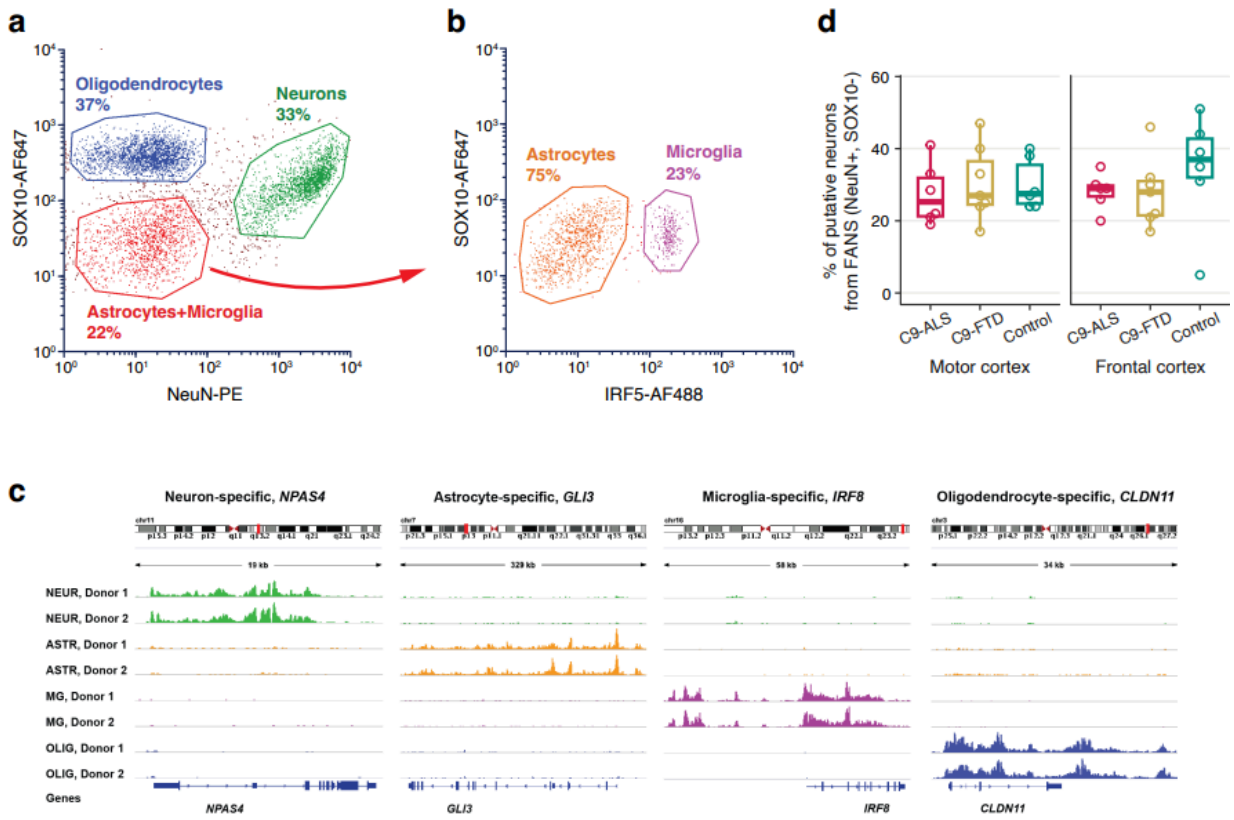
(a-e) Uniform manifold approximation and projection (UMAP2) embedding of snRNA-seq profiles (see Fig. 3.1a) were colored by brain region (a), diagnosis (b), donor sex (c), sequencing batch (d), or individual donor (e). (f) Clustree analysis to evaluate cluster stability with increasing Leiden resolutions from 0 to 3 in our full dataset clustering and the sub-clusterings for the three major cell classes. Clusters found at each resolution value are represented by nodes in the corresponding column. The nodes are colored by the SC3 stability index4 and sized proportional to the number of nuclei in the cluster. The transparency of the edges is adjusted according to the in-proportion, a metric defined as the ratio between the number of nuclei on the edge and the number of nuclei in the cluster it goes toward. Low in-proportion edges tend to arise as clusters become unstable. The selected resolutions we used before cell type annotation are highlighted in red circles.



Supplemental Figure 3.3 | Annotations of the identified cellular populations in snRNA-seq based on the expression of known markers, related to Figure 3.1.
Nuclei were first categorized into excitatory and inhibitory neurons, and nonneuronal cells using cell class marker genes (a). Fine-grained cell types were then annotated using marker genes for non-neuronal cells (b), excitatory neurons (c), and inhibitory neurons (d). Major cell types (left colored bars) were defined using shared marker genes. Normalized expression is defined as the z-score of log(CPM) for each gene across all cells. Dots were colored by the average normalized expression in each cell type, and dot size represents the percentage of nuclei expressing the marker genes (with non-zero counts) in the cell type. CGE, caudal ganglionic eminence; Astro, astrocytes; Endo, endothelial cells; Micro, microglia; Oligo, oligodendrocytes; OPC, oligodendrocyte precursor cells; VLMC, vascular leptomeningeal cell; IT, intratelencephalic; CT, corticothalamic; NP, near-projecting; ET, extratelencephalic.

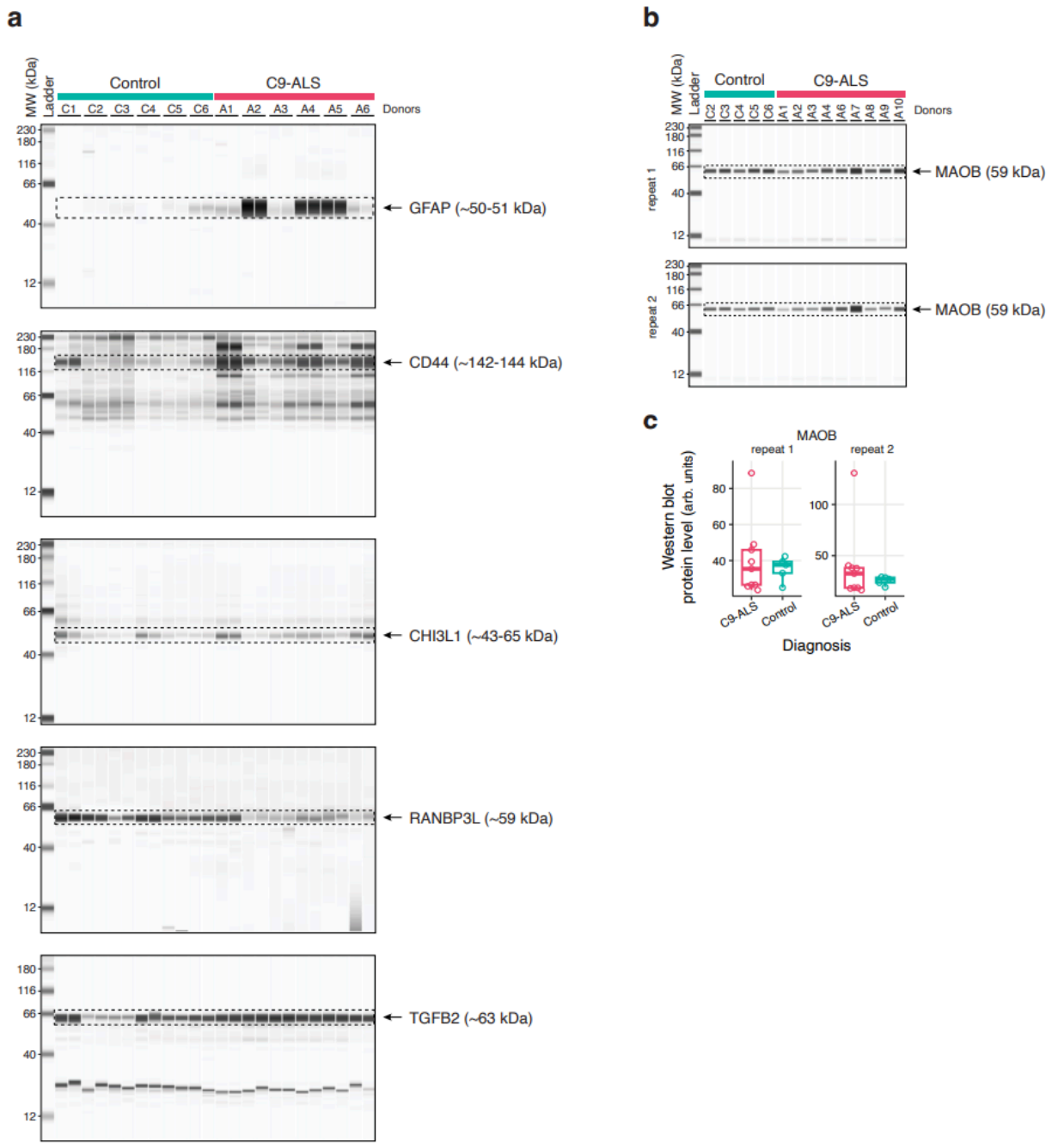


Supplemental Figure 3.4 | Comparison of the cell types identified in our snRNA-seq dataset with a previously published dataset, related to Figure 3.1.
(a) Cumulative proportion of the maximum prediction score across nuclei from the label transfer analysis using the published human primary motor cortex cell types as the reference dataset (see Methods). Curves are grouped and colored by diagnosis, brain region, and the major cell types identified in our study. (b) The annotations in Fig. 3.1 and Fig. S3.3 were consistent with cell types in the study of the human motor cortex5 . ARI, adjusted Rand index.



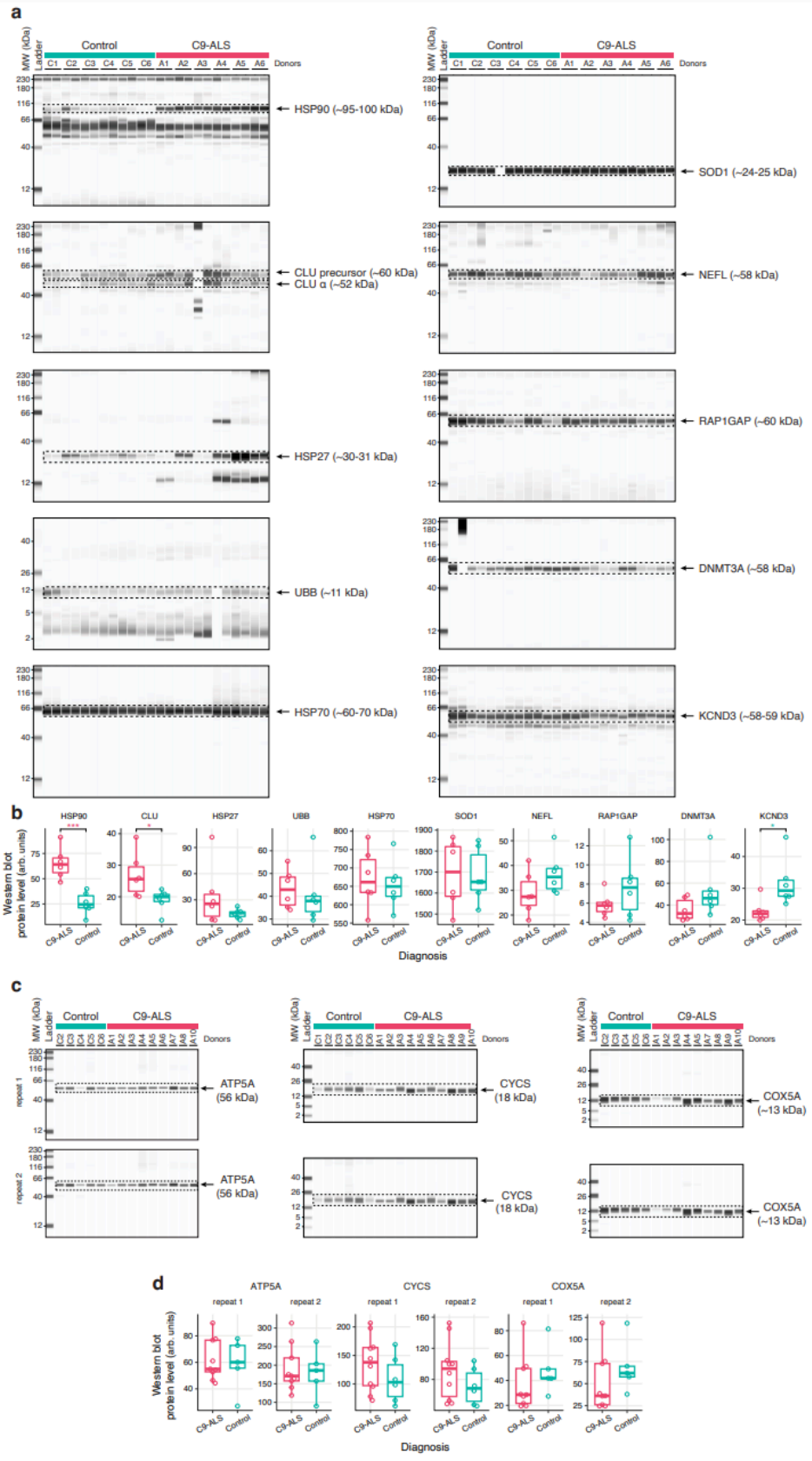
Supplemental Figure 3.5 | Fluorescence-activated nuclei sorting (FANS) isolation of nuclei of major brain cell types, related to Figure 3.1.

(a,b) Sequential FANS gating procedure to isolate nuclei from cell populations. (a) Three cell populations were isolated for bulk RNA-seq (Figs. 2e-f and 6f): neurons (NeuN+), oligodendrocyte lineage cells (NeuN–/Sox10+, consisting mainly of mature oligodendrocytes and a smaller population of OPCs), and other glia (NeuN–/Sox10–, mostly consisting of microglia and astrocytes). Anti-NeuN and anti-SOX10 antibodies were used to separate these three populations. (b) For H3K27ac ChIP-seq (Fig. 3.5), the NeuN–/Sox10– population was further split to isolate microglia (IRF5+), and astrocytes (IRF5–). (c) Validation of FANSseparated neurons, oligodendrocyte lineage cells, microglia and astrocytes using H3K27ac ChIP-seq. Signals from two donors are shown. In each population, H3K27ac signal enrichment was detected for known cell-type-specific genes. NEUR, neurons; ASTR, astrocytes; OLIG, oligodendrocyte lineage; MG, microglia. (d) Percentage of putative neurons (NeuN+SOX10- nuclei) from FANS. Circles represent biologically independent individual donors; N = 6 C9-ALS, 6 control and 7 C9-FTD samples, respectively. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. No significant differences were found in any pairwise comparison between diagnosis groups (two-sided Welch's t-test, p=0.41-0.97).

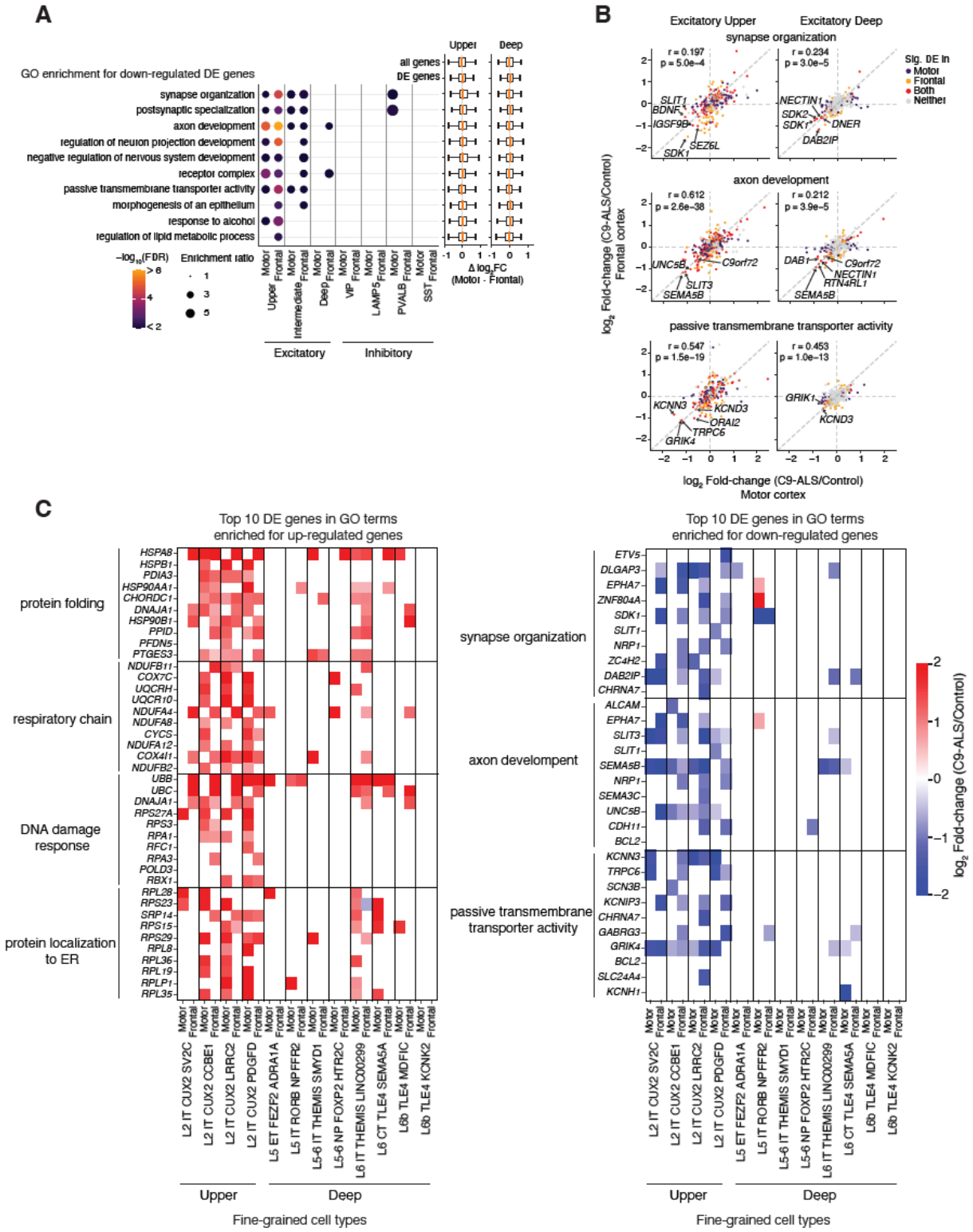


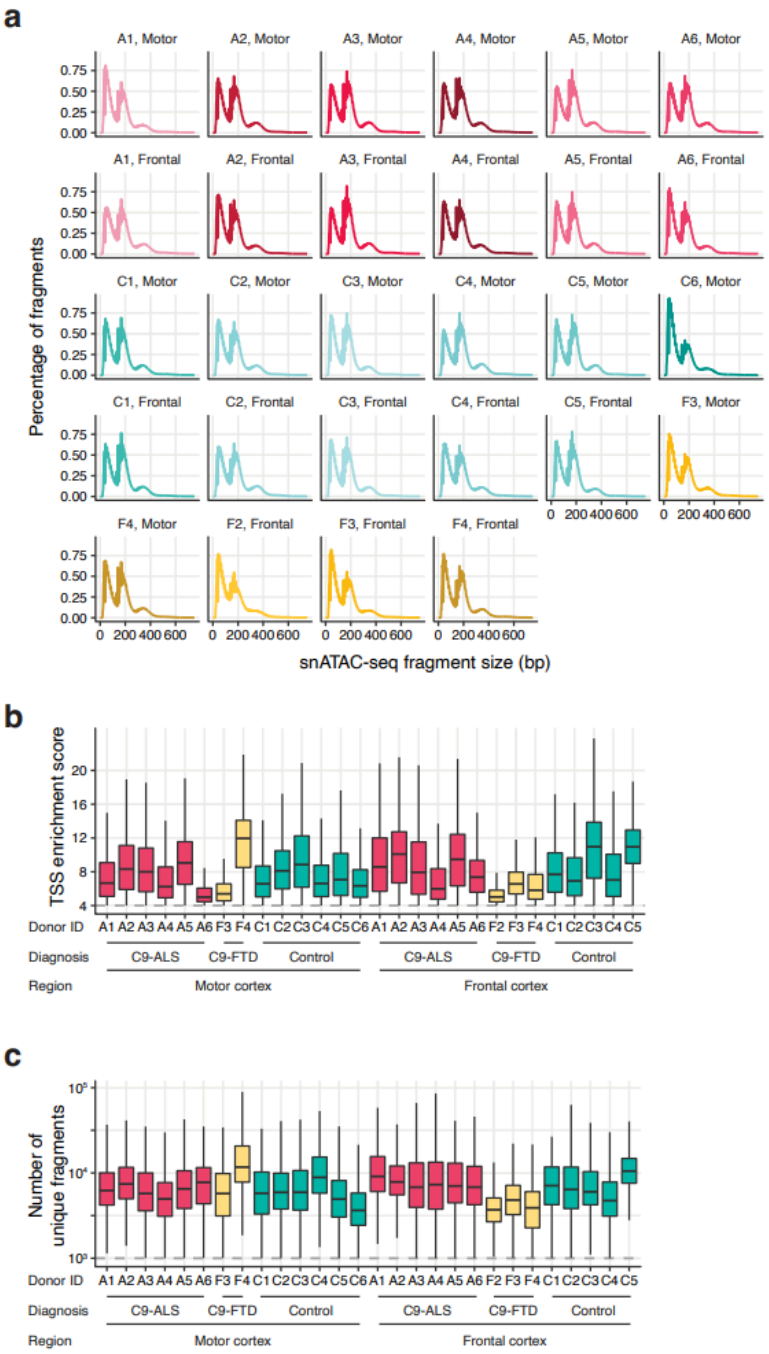
Supplemental Figure 3.6 | Quantification of proteins encoded by genes dysregulated in C9-ALS astrocytes using automated Western blot analysis, related to Figure 3.3.
(a) Protein levels were determined in selected DE genes identified in astrocytes in our snRNA-seq analysis. (b) MAOB protein levels from nine C9-ALS and five control donors. Each sample was assessed in two repeated experiments. (c) Quantifications of protein levels in (b). Arb. units, arbitrary units. Circles represent the signal for each biologically independent individual donor; N = 9 C9-ALS and 5 control samples. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. For each repeat experiment, a two-sided Welch's t-test was used to compare the immunoreactive signals between C9-ALS and control. No significant changes of MAOB protein levels were detected (p=0.549 and 0.297).

Supplemental Figure 3.7 | Quantification of proteins encoded by genes dysregulated in C9-ALS excitatory neurons using automated Western blot analysis, related to Figure 3.3.
(a) Same as Fig. S3.6a but for selected DE genes identified in excitatory neurons with our snRNA-seq dataset. (b) Quantification of protein levels in (a) with automated Western blot analysis. Arb. units, arbitrary units. Circles represent the average signal across duplicates for each donor; N = 6 C9-ALS and 6 control biologically independent samples. *, two-sided Welch's t-test $p < 0.05$; ***, $p < 0.001$. HSP90 and CLU proteins were upregulated ($p = 6.35e\text{-}4$ and 0.041, respectively), KCND3 was downregulated ($p = 0.036$), and HSP27, RAP1GAP and DNMT3A proteins were not significant ($p = 0.226$, 0.238 and 0.163, respectively). None of the tested genes had a change in protein abundance in the opposite direction from the mRNA expression change. (c) ATP5A, CYCS and COX5A protein levels (n=10 C9-ALS and 6 control donors for CYCS; n=9 C9-ALS and 5 control donors for ATP5A and COX5A). Each sample was assessed in two repeated experiments. (d) Quantifications of protein levels in (c). Arb. units, arbitrary units. Circles represent the signal for each donor; N = 10 C9-ALS and 6 control donors for CYCS, and N = 9 C9-ALS and 5 control donors for ATP5A and COX5A. For each repeat experiment, a two sided Welch's t-test was used to compare the immunoreactive signals between C9-ALS and control. No significant changes of ATP5A, CYCS or COX5A protein levels were detected (p=0.206-0.757). In each box plot in panels (b) and (d), the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median.


a
Control
C9-ALS
Donors
C1 C2 C3 C4 C5 C6 A1 A2 A3 A4 A5 A6
MW (kDa)
Ladder
HSP90 (~95-100 kDa)
SOD1 (~24-25 kDa)
CLU precursor (~60 kDa)
CLU α (~52 kDa)
NEFL (~58 kDa)
HSP27 (~30-31 kDa)
RAP1GAP (~60 kDa)
UBB (~11 kDa)
DNMT3A (~58 kDa)
HSP70 (~60-70 kDa)
KCND3 (~58-59 kDa)
b
HSP90
CLU
HSP27
UBB
HSP70
SOD1
NEFL
RAP1GAP
DNMT3A
KCND3
Western blot protein level (arb. units)
C9-ALS
Control
Diagnosis
c
repeat 1
repeat 2
ATP5A (56 kDa)
CYCS (18 kDa)
COX5A (~13 kDa)
d
ATP5A
CYCS
COX5A
repeat 1
repeat 2
Western blot protein level (arb. units)
Diagnosis

Supplemental Figure 3.8 | Functional enrichment analysis of differentially expressed (DE) genes in upper- and deep-layer excitatory neurons, related to Figure 3.4.
(a) Left panel: Top gene ontology (GO) terms enriched for genes downregulated in upper and/or deep layer excitatory neurons. Enrichments of the same terms for downregulated DE genes in other neuronal cell types are shown for comparison. Enriched GO categories (FDR<0.01) were selected by affinity propagation. Right panel: Differences in the C9-ALS vs. control fold changes between motor and frontal cortex (Δ log2FC = log2FC in motor cortex - log2FC in frontal cortex). The boxes denote the distribution of these differences (Δ log2FC) for all expressed genes in each GO category. As background comparisons, the distributions of these differences for all expressed genes (labeled as “all genes”) and for all ALS vs. control DE genes (labeled as “DE gene”) are shown on top. Two-sided Welch’s t-tests were used to test whether the Δ log2FC in each group of genes were significantly different from the Δ log2FC of the “all genes” control set. No significant differences were found. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. Exact N numbers and p-values for this analysis are provided in Table S3.5. (b) Comparison of effects in motor vs. frontal cortices for GO categories exemplifying processes/structures that are specific for neurons (synapse organization, axon development), or important for neuronal function (passive transmembrane transport activity). r, Pearson correlation coefficient. (c) mRNA expression fold-change (C9-ALS vs. control) of top 10 DE genes for GO terms enriched for upregulated DE genes (left panel) and downregulated genes (right panel) in fine-grained upper- and deep-layer excitatory neurons.


A
GO enrichment for down-regulated DE genes
synapse organization
postsynaptic specialization
axon development
regulation of neuron projection development
negative regulation of nervous system development
receptor complex
passive transmembrane transporter activity
morphogenesis of an epithelium
response to alcohol
regulation of lipid metabolic process
Upper
Deep
all genes
DE genes
Excitatory
Inhibitory
B
Excitatory Upper
Excitatory Deep
synapse organization
axon development
passive transmembrane transporter activity
Sig. DE in
Motor
Frontal
Both
Neither
log2 Fold-change (C9-ALS/Control)
Frontal cortex
log2 Fold-change (C9-ALS/Control)
Motor cortex
C
Top 10 DE genes in GO terms
enriched for up-regulated genes
protein folding
respiratory chain
DNA damage response
protein localization to ER
Top 10 DE genes in GO terms
enriched for down-regulated genes
synapse organization
axon developent
passive transmembrane transporter activity
log2 Fold-change (C9-ALS/Control)
Upper
Deep
Fine-grained cell types

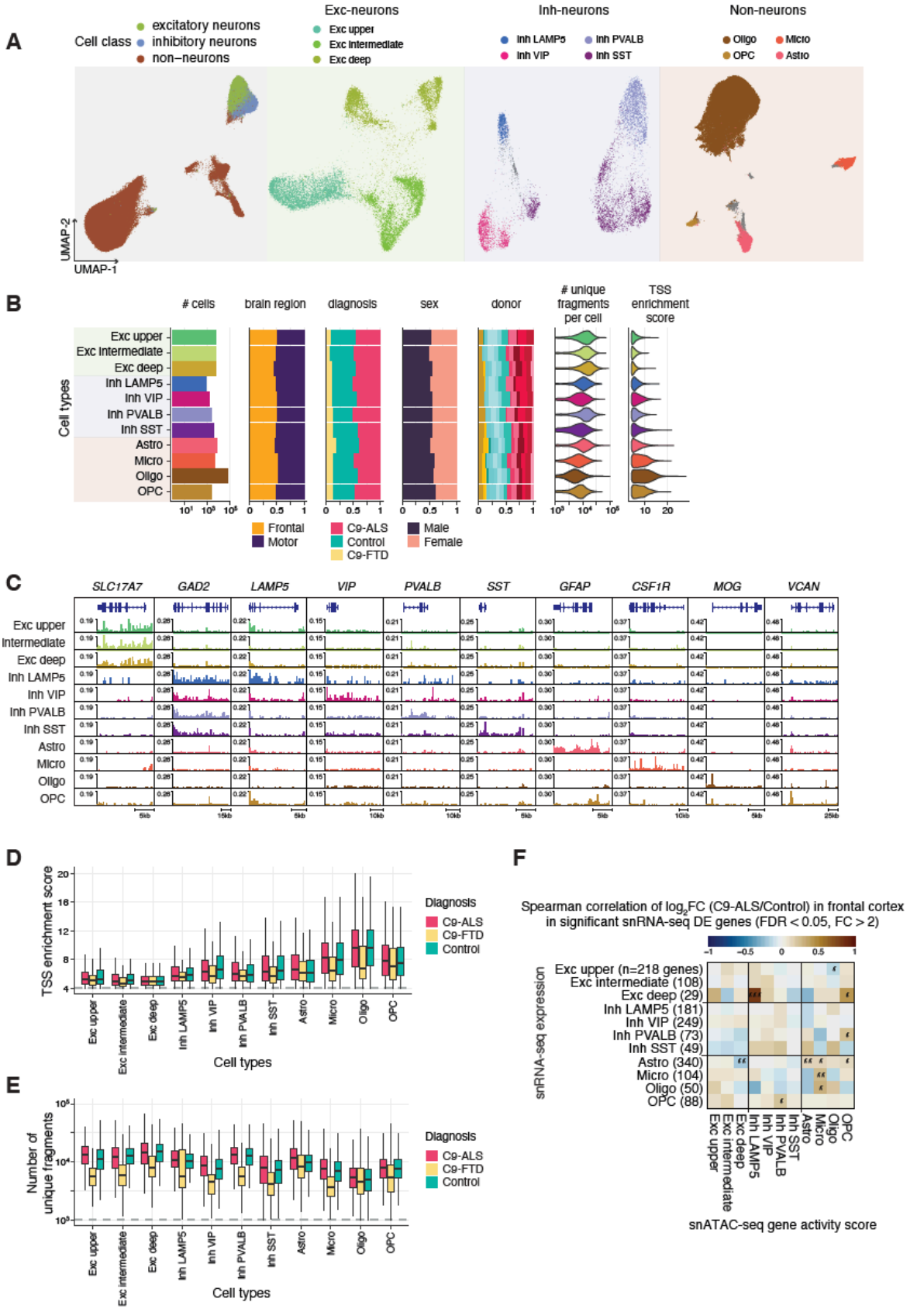


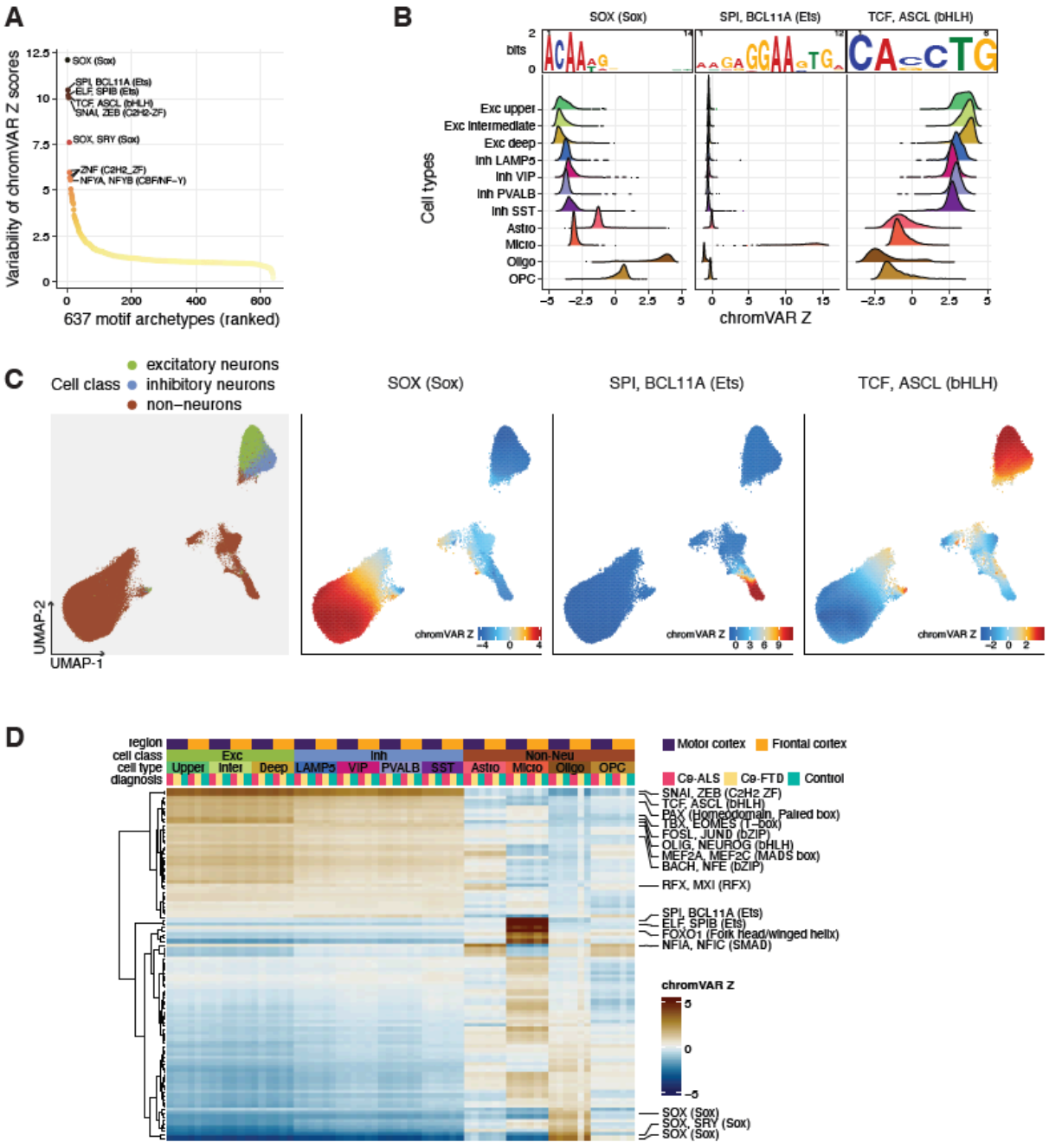
Supplemental Figure 3.9 | QC metrics used in the snATAC-seq data pre-processing, related to Figure 3.5.

(a) Distribution of snATAC-seq fragment sizes in each sample that passed QC, showing periodicity related to nucleosome spacing. (b-c) Box-and-whisker plots to show the distribution of transcription start site (TSS) enrichment score (b) and number of unique fragments (c) across nuclei grouped by samples. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median.

Supplemental Figure 3.10 | The epigenetic landscape of C9-ALS brain cells determined by snATAC-seq, related to Figure 3.5.

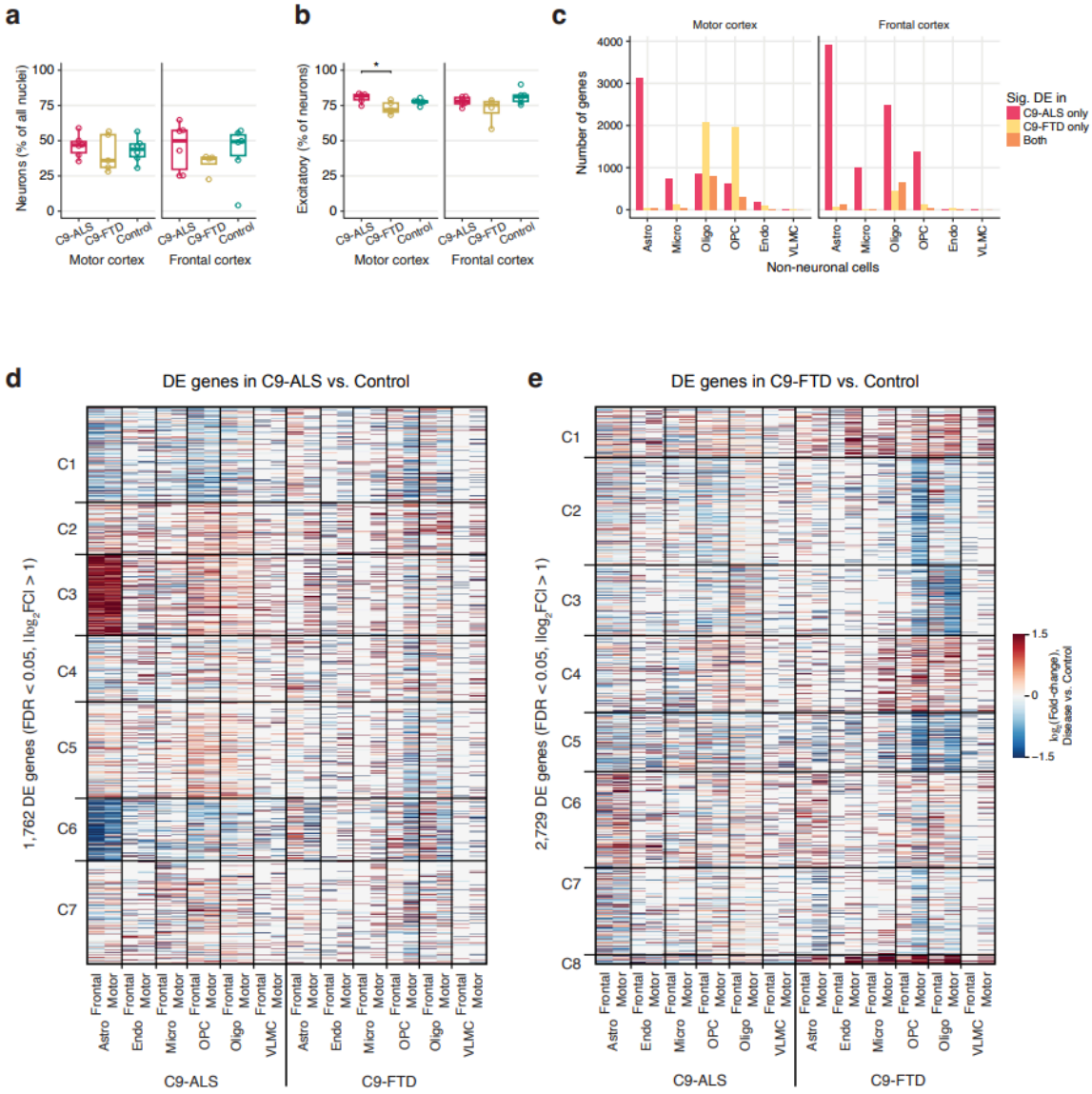
(a) Clustering n=109,198 high-quality snATAC-seq profiles (TSS enrichment ≥ 4, unique fragments ≥ 1,000 per cell) identified 11 major brain cell types. The clusters were annotated by transferring labels from the snRNAseq data (see Methods). (b) Major cell types from snATAC-seq were distributed across brain regions, diagnosis groups, sex, and donors. TSS, transcription start site. (c) snATAC-seq signal at cell-type-specific marker genes. Track height represents pseudo-bulk counts normalized by reads in TSS. (d-e) Box-and-whisker plots to show the distribution of TSS enrichment score (d) and number of unique fragments (e) across nuclei grouped by major cell types. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. N numbers for each box plot in panels (d) and (e) can be derived from the full metadata information of each nucleus from the snATAC-seq data provided in Table S3.6. (f) Spearman correlation of the C9-ALS vs. control fold-change (FC) for snRNA expression vs. snATAC gene activity score in frontal cortex; the corresponding analysis for motor cortex is shown in Fig. 3.5e. The analysis was performed for strongly DE genes (FC > 2) in each major cell type in frontal cortex. Two-sided Spearman's rank correlation test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.


A
Cell class
excitatory neurons
inhibitory neurons
non−neurons
Exc-neurons
Exc upper
Exc intermediate
Exc deep
Inh-neurons
Inh LAMP5
Inh PVALB
Inh VIP
Inh SST
Non-neurons
Oligo
Micro
OPC
Astro
UMAP-2
UMAP-1
B
cells
brain region
diagnosis
sex
donor
unique fragments per cell
TSS enrichment score
Cell types
Exc upper
Exc intermediate
Exc deep
Inh LAMP5
Inh VIP
Inh PVALB
Inh SST
Astro
Micro
Oligo
OPC
Frontal
Motor
C9-ALS
Control
C9-FTD
Male
Female
C
SLC17A7
GAD2
LAMP5
VIP
PVALB
SST
GFAP
CSF1R
MOG
VCAN
D
TSS enrichment score
Diagnosis
C9-ALS
C9-FTD
Control
Cell types
E
Number of unique fragments
F
Spearman correlation of log2FC (C9-ALS/Control) in frontal cortex in significant snRNA-seq DE genes (FDR < 0.05, FC > 2)
Exc upper (n=218 genes)
Exc intermediate (108)
Exc deep (29)
Inh LAMP5 (181)
Inh VIP (249)
Inh PVALB (73)
Inh SST (49)
Astro (340)
Micro (104)
Oligo (50)
OPC (88)
snRNA-seq expression
snATAC-seq gene activity score



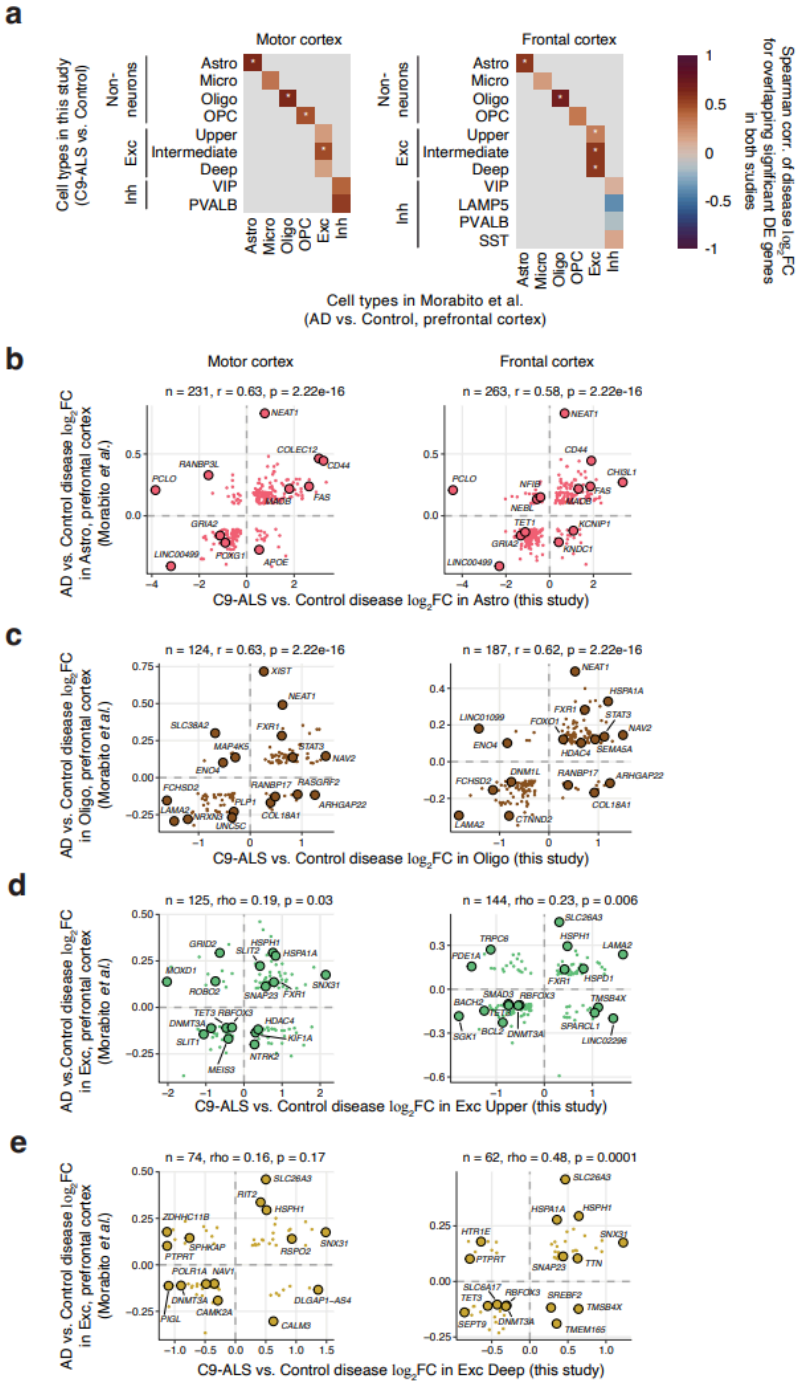
Supplemental Figure 3.11 | ChromVAR scores in non-redundant transcription factor (TF) motif archetypes, related to Figure 3.5.

(a) Variability (standard deviation) of the ChromVAR6 Z scores across all nuclei for TF motif archetypes (see Methods). (b) Distribution of ChromVAR Z score across nuclei from each major cell type for three representative TF motif archetypes with high variability. Sequence logos of the motif archetypes are shown on top. (c) ChromVAR Z scores of motif archetypes for each cell are displayed on the UMAP embedding. (d) Heatmap of average ChromVAR Z scores across groups of nuclei (columns) by major cell types, brain region and diagnosis, for the top 100 motif archetypes (rows) with highest variability across all nuclei. Top 3 cell-type-specific motif archetypes for each cell type are labeled.

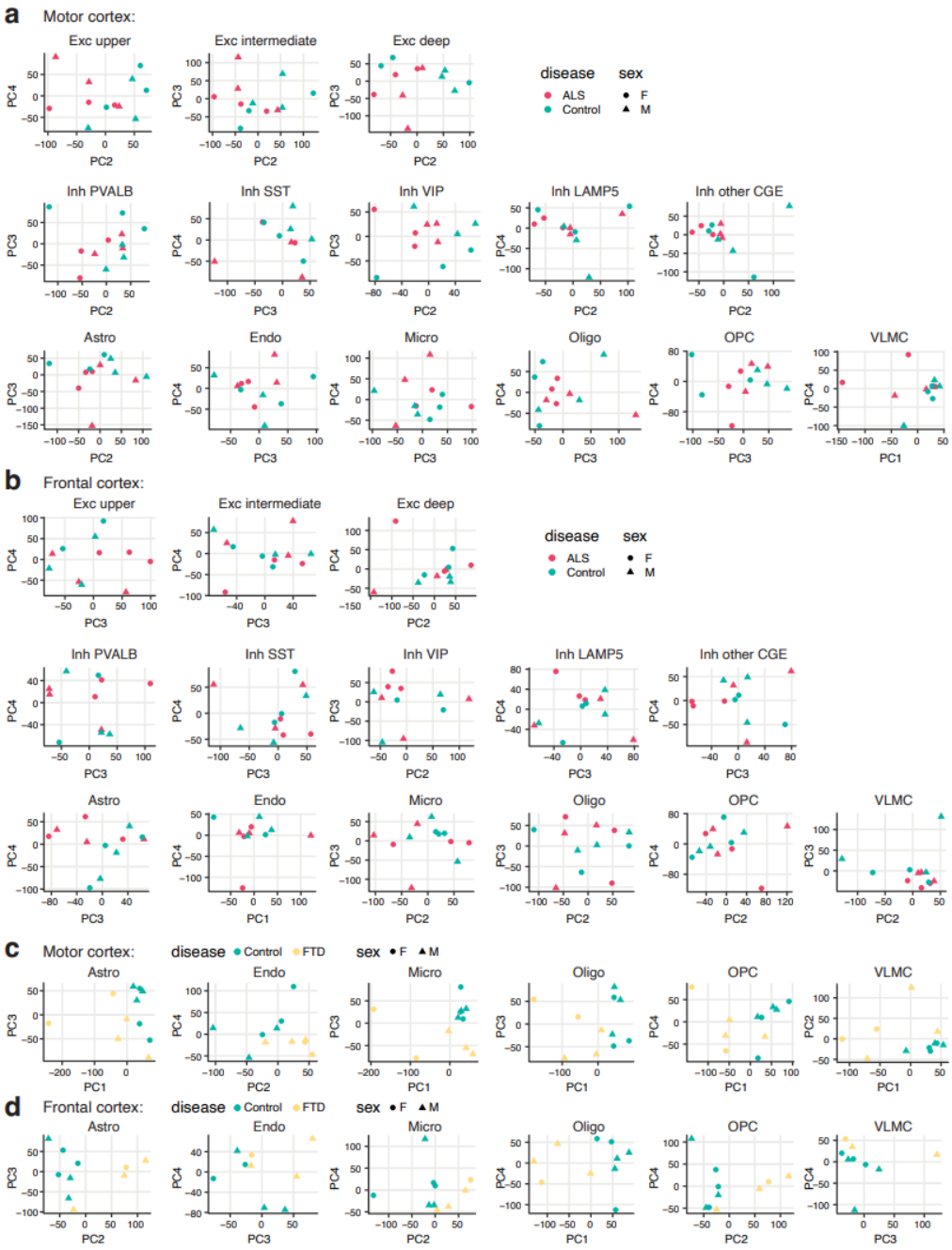


Supplemental Figure 3.12 | Comparison of the disease effect on gene expression in non-neuronal cells in C9-ALS and C9-FTD, related to Figure 3.6.

(a-b) Related to Fig. 6b-c. Relative abundance of neurons (a), and percent of excitatory neurons among all neurons (b) using relaxed QC criteria (without thresholds on the number of detected genes, the percentage of mitochondrial reads, and doublet scores). This analysis included all 195,144 non-empty droplets determined by CellBender. No significant difference in the percentage of neurons was detected between diagnostic groups in (a) (two-sided Welch's t-test, p=0.20-0.76). Circles represent individual donors; N = 6 C9-ALS and 6 control samples for both brain regions, and N = 5 and 4 C9-FTD samples from motor cortex and frontal cortex, respectively. *, two-sided Welch's t-test p = 0.021. In each box plot, the lower and upper hinges correspond to the first and third quartiles; the whiskers extend 1.5 * IQR (interquartile range) away from the hinges; and the center denotes the median. (c) Number of significant DE genes identified in the two diseases. For this analysis, control samples were split into two independent groups (from different donors) for comparison with C9-ALS and with C9-FTD to avoid double dipping (see Methods). (d-e) Heatmap to compare the gene expression fold-changes of C9-ALS vs. control and C9-FTD vs. control. Strongly DE genes (fold-change > 2) identified from the full dataset (same genes shown in Fig. 3.2d for C9-ALS and Fig. 3.6h for C9-FTD) were used, and the fold-changes from the split control analysis were shown to avoid double dipping.

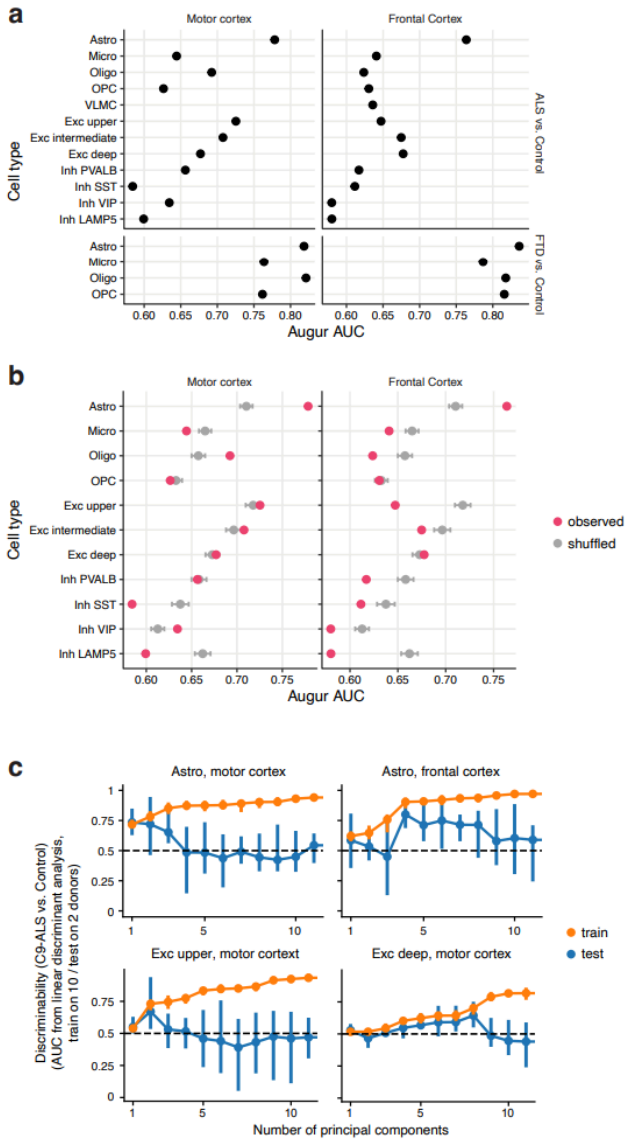


Supplemental Figure 3.13 | Comparison of C9-ALS vs. control gene expression fold-changes with AD vs. control gene expression fold-changes.
(a) Spearman correlations of disease fold-changes between C9-ALS (this study) and AD for matching cell types. Only genes that were significantly DE in both diseases were included in this analysis. Asterisk denotes significant correlations (FDR < 0.05). (b-e) Scatter plots showing the correlation of the disease fold-changes between the two diseases for astrocytes, oligodendrocytes, and excitatory neurons in the two studies. Examples of genes with concordant or discordant changes in the two diseases are highlighted. n, number of significant DE genes in both studies in the corresponding cell types; rho and p, two-sided Spearman's rank correlation test coefficient and p-value.



Supplemental Figure 3.14 | Principal component analysis of the pseudobulk snRNA-seq counts from each donor in each cell type.

(a-b) Separation of the pseudobulk gene expression profiles in each major cell type for C9-ALS and control donors in motor cortex (a) and frontal cortex (b). (c-d) Separation of the pseudobulk gene expression profiles in each non-neuronal nuclei for C9- FTD and control donors in motor cortex (c) and frontal cortex (d). Counts were normalized with DEseq2's median of ratios method , and the natural logarithm of the normalized count plus one was used in the principal component analysis. CGE, caudal ganglionic eminence; Astro, astrocytes; Endo, endothelial cells; Micro, microglia; Oligo, oligodendrocytes; OPC, oligodendrocyte precursor cells; VLMC, vascular leptomeningeal cell; Exc, excitatory neurons; Inh, inhibitory neurons; PC, principal component.



Supplemental Figure 3.15 | Separations of disease and control samples in the snRNA-seq dataset. (a-b) Cell type prioritization using Augur. (a) The area under the receiver operating characteristic curve (AUC) was reported by Augur as a metric of the separability between the disease (C9-ALS or C9-FTD) and control samples. (b) For C9-ALS vs. control, n = 82 random shuffles with balanced diagnosis and sex were generated by permutation of the diagnosis labels of the donors. Augur was run on these shuffles and the mean AUCs are shown in grey dots. Data are presented as mean values +/- 95% confidence intervals of the mean. Observed AUCs with real diagnosis labels are shown in red dots. Cell types that were not detected in all 12 donors were omitted in this analysis. (c) Discriminability of C9-ALS vs. control cells is shown as the area under the receiver operating characteristic (ROC) for four sets of cells. Donors were split using k-fold cross-validation, with 10 subjects used for training a linear discriminant classifier and 2 subjects used for testing (one control and one C9-ALS subject). The AUC for training subjects (orange) increases with the number of PCs used, whereas the test-set AUC (blue) shows that the C9-ALS cells can be significantly discriminated from the control cells using the top few PCs. Data are presented as mean values +/- 95% confidence intervals of the mean; n = 6 groups of split samples. Astro, astrocytes; Endo, endothelial cells; Micro, microglia; Oligo, oligodendrocytes; OPC, oligodendrocyte precursor cells; VLMC, vascular leptomeningeal cell; Exc, excitatory neurons; Inh, inhibitory neurons; AUC, area under the ROC curve.

APPENDIX

Supplemental tables below are provided separately in Excel format in a zip file.

Chien_SupplementalTable1.1_Donor_metadata.xlsx: Donor age, sex and total nuclei passing quality control metrics in Chapter 1.

Chien_SupplementalTable1.2_Cell_type_composition.xlsx: Number of cells per donor and cell type in Chapter 1.

Chien_SupplementalTable1.3_Age_differentially_expressed_genes.xlsx: List of age-DEGs with adjusted p-value (FDR) < 0.05 and their statistics analyzed by Dream in Chapter 1.

Chien_SupplementalTable1.4_Age_differentially_methylated_regions.xlsx: List of age-DMRs called using DSS in Chapter 1.

Chien_SupplementalTable1.5_Sex_differentially_expressed_genes.xlsx: List of sex-DEGs with adjusted p-value (FDR) < 0.05 and their statistics analyzed by Dream in Chapter 1.

Chien_SupplementalTable1.6_Sex_differentially_methylated_regions.xlsx: List of sex-DMRs called using DSS in Chapter 1.

Chien_SupplementalTable1.7_Cell_metadata.xlsx: Quality control information and cell type annotation for each cell in Chapter 1.

Chien_SupplementalTable2.1_Donor_metadata.xlsx: Metadata of the donors in Chapter 2.

Chien_SupplementalTable2.2_Age_differentially_expressed_genes.xlsx: List of age-DEGs with adjusted p-value (FDR) < 0.05 and their statistics analyzed by Dream in Chapter 2.

Chien_SupplementalTable2.3_Sex_differentially_expressed_genes.xlsx List of sex-DEGs with adjusted p-value (FDR) < 0.05 and their statistics analyzed by Dream in Chapter 2.

Chien_SupplementalTable3.1_Donor_metadata.xlsx: Metadata of the donors in Chapter 3.

Chien_SupplementalTable3.2_Cell_RNA_metadata.xlsx: Metadata and cell type annotations of all high-quality nuclei from snRNA-seq in Chapter 3.

Chien_SupplementalTable3.3_ALS_motor_differentially_expressed_genes.xlsx: List of C9-ALS vs. control differentially expressed genes identified in each major cell type in motor cortex in Chapter 3.

Chien_SupplementalTable3.4_ALS_frontal_differentially_expressed_genes.xlsx: List of C9-ALS vs. control differentially expressed genes identified in each major cell type in frontal cortex in Chapter 3.

Chien_SupplementalTable3.5_ALS_gene_ontology.xlsx: Gene Ontology enrichment results for C9-ALS vs. control differentially expressed genes in Chapter 3.

Chien_SupplementalTable3.6_Cell_ATAC_metadata.xlsx: Metadata and cell type annotations of all high-quality nuclei from snATAC-seq in Chapter 3.

Chien_SupplementalTable3.7_ALS_H3K27ac_differential_peaks.xlsx: H3K27ac differential peaks between C9-ALS and control identified in ChIP-seq in Chapter 3.

Chien_SupplementalTable3.8_FTD_motor_differentially_expressed_genes.xlsx: List of C9-FTD vs. control differentially expressed genes identified in each non-neuronal cell type in motor cortex in Chapter 3.

Chien_SupplementalTable3.9_FTD_frontal_differentially_expressed_genes.xlsx: List of C9-FTD vs. control differentially expressed genes identified in each non-neuronal cell type in frontal cortex in Chapter 3.

Chien_SupplementalTable3.10_Western_blot_signal.xlsx: Digital Western Blot signal for tested genes in C9-ALS and control samples in Chapter 3.

Chien_SupplementalTable3.11_correlation_ATAC_H3K27ac.xlsx: Pearson correlations of snATAC-seq and H3K27ac ChIP-seq signal in 1kb genomic bins in C9-ALS and control samples in Chapter 3.

Chien_SupplementalTable3.12_correlation_RNA_ATAC.xlsx: Spearman correlation of the C9-ALS vs. control fold-changes for snRNA expression vs. snATAC gene activity score in Chapter 3.

REFERENCES

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186,** 243–278 (2023).

2. Miller, J. A., Guillozet-Bongaarts, A., Gibbons, L. E., Postupna, N., Renz, A., Beller, A. E., Sunkin, S. M., Ng, L., Rose, S. E., Smith, K. A., Szafer, A., Barber, C., Bertagnolli, D., Bickley, K., Brouner, K., Caldejon, S., Chapin, M., Chua, M. L., Coleman, N. M., Cudaback, E., Cuhaciyan, C., Dalley, R. A., Dee, N., Desta, T., Dolbeare, T. A., Dotson, N. I., Fisher, M., Gaudreault, N., Gee, G., Gilbert, T. L., Goldy, J., Griffin, F., Habel, C., Haradon, Z., Hejazinia, N., Hellstern, L. L., Horvath, S., Howard, K., Howard, R., Johal, J., Jorstad, N. L., Josephsen, S. R., Kuan, C. L., Lai, F., Lee, E., Lee, F., Lemon, T., Li, X., Marshall, D. A., Melchor, J., Mukherjee, S., Nyhus, J., Pendergraft, J., Potekhina, L., Rha, E. Y., Rice, S., Rosen, D., Sapru, A., Schantz, A., Shen, E., Sherfield, E., Shi, S., Sodt, A. J., Thatra, N., Tieu, M., Wilson, A. M., Montine, T. J., Larson, E. B., Bernard, A., Crane, P. K., Ellenbogen, R. G., Keene, C. D. & Lein, E. Neuropathological and transcriptomic characteristics of the aged brain. *Elife* **6,** (2017).

3. Tan, L., Ma, W., Wu, H., Zheng, Y., Xing, D., Chen, R., Li, X., Daley, N., Deisseroth, K. & Xie, X. S. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* (2021). doi:10.1016/j.cell.2020.12.032

4. Hoffman, G. E., Ma, Y., Montgomery, K. S., Bendl, J., Jaiswal, M. K., Kozlenkov, A., Peters, M. A., Dracheva, S., Fullard, J. F., Chess, A., Devlin, B., Sieberts, S. K. & Roussos, P. Sex Differences in the Human Brain Transcriptome of Cases With Schizophrenia. *Biol. Psychiatry* **91,** 92–101 (2022).

5. Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S. E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E. A., Skol, A. D., Garrido-Martín, D., Reverter, F., Brown, A., Evans, P., Gamazon, E. R., Payne, A., Bonazzola, R., Barbeira, A. N., Hamel, A. R., Martinez-Perez, A., Soria, J. M., GTEx Consortium, Pierce, B. L., Stephens, M., Eskin, E., Dermitzakis, E. T., Segrè, A. V., Im, H. K., Engelhardt, B. E., Ardlie, K. G., Montgomery, S. B., Battle, A. J., Lappalainen, T., Guigó, R. & Stranger, B. E. The impact of sex on gene expression across human tissues. *Science* **369,** (2020).

6. Govender, P., Ghai, M. & Okpeku, M. Sex-specific DNA methylation: impact on human health and development. *Mol. Genet. Genomics* **297,** 1451–1466 (2022).

7. Geevasinga, N., Menon, P., Özdinler, P. H., Kiernan, M. C. & Vucic, S. Pathophysiological and diagnostic implications of cortical dysfunction in ALS. *Nat. Rev. Neurol.* **12,** 651–661 (2016).

8. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M. & Tsai, L.-H. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* (2019). doi:10.1038/s41586-019-1195-2

9. Allen, W. E., Blosser, T. R., Sullivan, Z. A., Dulac, C. & Zhuang, X. Molecular and spatial

signatures of mouse brain aging at single-cell resolution. *Cell* **186,** 194–208.e18 (2023).

10. Wen, X., Luo, Z., Zhao, W., Calandrelli, R., Nguyen, T. C., Wan, X., Charles Richard, J. L. & Zhong, S. Single-cell multiplex chromatin and RNA interactions in ageing human brain. *Nature* **628,** 648–656 (2024).

11. Wamsley, B., Bicks, L., Cheng, Y., Kawaguchi, R., Quintero, D., Margolis, M., Grundman, J., Liu, J., Xiao, S., Hawken, N., Mazariegos, S. & Geschwind, D. H. Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. *Science* **384,** eadh2602 (2024).

12. Mathys, H., Peng, Z., Boix, C. A., Victor, M. B., Leary, N., Babu, S., Abdelhady, G., Jiang, X., Ng, A. P., Ghafari, K., Kunisky, A. K., Mantero, J., Galani, K., Lohia, V. N., Fortier, G. E., Lotfi, Y., Ivey, J., Brown, H. P., Patel, P. R., Chakraborty, N., Beaudway, J. I., Imhoff, E. J., Keeler, C. F., McChesney, M. M., Patel, H. H., Patel, S. P., Thai, M. T., Bennett, D. A., Kellis, M. & Tsai, L.-H. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186,** 4365–4385.e27 (2023).

13. Yap, C. X., Vo, D. D., Heffel, M. G., Bhattacharya, A., Wen, C., Yang, Y., Kemper, K. E., Zeng, J., Zheng, Z., Zhu, Z., Hannon, E., Vellame, D. S., Franklin, A., Caggiano, C., Wamsley, B., Geschwind, D. H., Zaitlen, N., Gusev, A., Pasaniuc, B., Mill, J., Luo, C. & Gandal, M. J. Brain cell-type shifts in Alzheimer's disease, autism, and schizophrenia interrogated using methylomics and genetics. *Sci. Adv.* **10,** eadn7655 (2024).

14. Emani, P. S., Liu, J. J., Clarke, D., Jensen, M., Warrell, J., Gupta, C., Meng, R., Lee, C. Y., Xu, S., Dursun, C., Lou, S., Chen, Y., Chu, Z., Galeev, T., Hwang, A., Li, Y., Ni, P., Zhou, X., PsychENCODE Consortium‡, Bakken, T. E., Bendl, J., Bicks, L., Chatterjee, T., Cheng, L., Cheng, Y., Dai, Y., Duan, Z., Flaherty, M., Fullard, J. F., Gancz, M., Garrido-Martín, D., Gaynor-Gillett, S., Grundman, J., Hawken, N., Henry, E., Hoffman, G. E., Huang, A., Jiang, Y., Jin, T., Jorstad, N. L., Kawaguchi, R., Khullar, S., Liu, J., Liu, J., Liu, S., Ma, S., Margolis, M., Mazariegos, S., Moore, J., Moran, J. R., Nguyen, E., Phalke, N., Pjanic, M., Pratt, H., Quintero, D., Rajagopalan, A. S., Riesenmy, T. R., Shedd, N., Shi, M., Spector, M., Terwilliger, R., Travaglini, K. J., Wamsley, B., Wang, G., Xia, Y., Xiao, S., Yang, A. C., Zheng, S., Gandal, M. J., Lee, D., Lein, E. S., Roussos, P., Sestan, N., Weng, Z., White, K. P., Won, H., Girgenti, M. J., Zhang, J., Wang, D., Geschwind, D., Gerstein, M. & PsychENCODE Consortium. Single-cell genomics and regulatory networks for 388 human brains. *Science* **384,** eadi5199 (2024).

15. Xia, Y., Xia, C., Jiang, Y., Chen, Y., Zhou, J., Dai, R., Han, C., Mao, Z., PsychENCODE Consortium, Liu, C. & Chen, C. Transcriptomic sex differences in postmortem brain samples from patients with psychiatric disorders. *Sci. Transl. Med.* **16,** eadh9974 (2024).

16. Kundakovic, M. & Tickerhoof, M. Epigenetic mechanisms underlying sex differences in the brain and behavior. *Trends Neurosci.* (2023). doi:10.1016/j.tins.2023.09.007

17. Tarkhov, A. E., Lindstrom-Vautrin, T., Zhang, S., Ying, K., Moqri, M., Zhang, B., Tyshkovskiy, A., Levy, O. & Gladyshev, V. N. Nature of epigenetic aging from a single-cell perspective. *Nat. Aging* **4,** 854–870 (2024).

18. Zeng, B., Bendl, J., Deng, C., Lee, D., Misir, R., Reach, S. M., Kleopoulos, S. P., Auluck, P., Marenco, S., Lewis, D. A., Haroutunian, V., Ahituv, N., Fullard, J. F., Hoffman, G. E. &

Roussos, P. Genetic regulation of cell type-specific chromatin accessibility shapes brain disease etiology. *Science* **384,** eadh4265 (2024).

19. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150,** 563–565 (1942).

20. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *CABI Databases* at <https://www.cabidigitallibrary.org/doi/full/10.5555/19580101706>

21. Waddington, C. H. Genetic assimilation. *Adv. Genet.* **10,** 257–293 (1961).

22. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).

23. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128,** 635–638 (2007).

24. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).

25. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17,** 183–193 (2016).

26. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38,** 23–38 (2013).

27. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).

28. Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., Hemberg, M., Ebert, D. H. & Greenberg, M. E. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522,** 89–93 (2015).

29. Feng, J., Zhou, Y., Campbell, S. L., Le, T., Li, E., Sweatt, J. D., Silva, A. J. & Fan, G. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13,** 423–430 (2010).

30. Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T., Gnirke, A. & Meissner, A. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7,** e1002389 (2011).

31. Telley, L., Agirman, G., Prados, J., Amberg, N., Fièvre, S., Oberst, P., Bartolini, G., Vitali, I., Cadilhac, C., Hippenmeyer, S., Nguyen, L., Dayer, A. & Jabaudon, D. Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364,** (2019).

32. Subramanian, L., Calcagnotto, M. E. & Paredes, M. F. Cortical Malformations: Lessons in Human Brain Development. *Front. Cell. Neurosci.* **13,** 576 (2019).

33. Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu,

J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., Sejnowski, T. J., Behrens, M. M. & Ecker, J. R. Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 1237905 (2013).

34. Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M. & Carell, T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5,** e15367 (2010).

35. Schultz, M. D., He, Y., Whitaker, J. W., Hariharan, M., Mukamel, E. A., Leung, D., Rajagopal, N., Nery, J. R., Urich, M. A., Chen, H., Lin, S., Lin, Y., Jung, I., Schmitt, A. D., Selvaraj, S., Ren, B., Sejnowski, T. J., Wang, W. & Ecker, J. R. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523,** 212–216 (2015).

36. Lavery, L. A., Ure, K., Wan, Y.-W., Luo, C., Trostle, A. J., Wang, W., Jin, H., Lopez, J., Lucero, J., Durham, M. A., Castanon, R., Nery, J. R., Liu, Z., Goodell, M., Ecker, J. R., Behrens, M. M. & Zoghbi, H. Y. Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. *Elife* **9,** (2020).

37. Li, J., Pinto-Duarte, A., Zander, M., Cuoco, M. S., Lai, C.-Y., Osteen, J., Fang, L., Luo, C., Lucero, J. D., Gomez-Castanon, R., Nery, J. R., Silva-Garcia, I., Pang, Y., Sejnowski, T. J., Powell, S. B., Ecker, J. R., Mukamel, E. A. & Behrens, M. M. Dnmt3a knockout in excitatory neurons impairs postnatal synapse maturation and increases the repressive histone modification H3K27me3. *Elife* **11,** (2022).

38. Ham, S. & Lee, S.-J. V. Advances in transcriptome analysis of human brain aging. *Exp. Mol. Med.* **52,** 1787–1797 (2020).

39. Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U. & Linnarsson, S. Molecular architecture of the mouse nervous system. *Cell* **174,** 999–1014.e22 (2018).

40. Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A. & McCarroll, S. A. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).

41. Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., Hawrylycz, M., Koch, C. & Zeng, H. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** 335–346 (2016).

42. Luo, C., Keown, C. L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J. R., Sandoval, J. P., Bui, B., Sejnowski, T. J., Harkins, T. T., Mukamel, E. A., Behrens, M. M. & Ecker, J. R. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357,** 600–604 (2017).

43. Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., Zhang, Y., Sos, B. C., Afzal, V., Dickel, D. E., Kuan, S., Visel, A., Pennacchio, L. A., Zhang, K. & Ren, B. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21,** 432–439 (2018).

44. Lake, B. B., Chen, S., Sos, B. C., Fan, J., Kaeser, G. E., Yung, Y. C., Duong, T. E., Gao, D., Chun, J., Kharchenko, P. V. & Zhang, K. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36,** 70–80 (2018).

45. Trapp, A., Kerepesi, C. & Gladyshev, V. N. Profiling epigenetic age in single cells. *Nat Aging* **1,** 1189–1201 (2021).

46. Johnstone, S. E., Gladyshev, V. N., Aryee, M. J. & Bernstein, B. E. Epigenetic clocks, aging, and cancer. *Science* **378,** 1276–1277 (2022).

47. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14,** R115 (2013).

48. Seale, K., Horvath, S., Teschendorff, A., Eynon, N. & Voisin, S. Making sense of the ageing methylome. *Nat. Rev. Genet.* 1–21 (2022).

49. Stoyanova, E., Riad, M., Rao, A. & Heintz, N. 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. *Elife* **10,** (2021).

50. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151,** 1417–1430 (2012).

51. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

52. Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. & Rao, A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

53. Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W. & Balasubramanian, S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336,** 934–937 (2012).

54. Booth, M. J., Ost, T. W. B., Beraldi, D., Bell, N. M., Branco, M. R., Reik, W. & Balasubramanian, S. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* **8,** 1841–1851 (2013).

55. Yu, M., Hon, G. C., Szulwach, K. E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.-H., Jin, P., Ren, B. & He, C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149,** 1368–1380 (2012).

56. Gershoni, M. & Pietrokovski, S. The landscape of sex-differential transcriptome and its consequent selection in human adults. *BMC Biol.* **15,** 7 (2017).

57. Keown, C. L., Berletch, J. B., Castanon, R., Nery, J. R., Disteche, C. M., Ecker, J. R. & Mukamel, E. A. Allele-specific non-CG DNA methylation marks domains of active chromatin in female mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **114,** E2882–E2890 (2017).

58. Zhang, Y., Amaral, M. L., Zhu, C., Grieco, S. F., Hou, X., Lin, L., Buchanan, J., Tong, L., Preissl, S., Xu, X. & Ren, B. Single-cell epigenome analysis reveals age-associated decay

of heterochromatin domains in excitatory neurons in the mouse brain. *Cell Res.* (2022). doi:10.1038/s41422-022-00719-6

59. Lu, T.-C., Brbić, M., Park, Y.-J., Jackson, T., Chen, J., Kolluru, S. S., Qi, Y., Katheder, N. S., Cai, X. T., Lee, S., Chen, Y.-C., Auld, N., Liang, C.-Y., Ding, S. H., Welsch, D., D'Souza, S., Pisco, A. O., Jones, R. C., Leskovec, J., Lai, E. C., Bellen, H. J., Luo, L., Jasper, H., Quake, S. R. & Li, H. Aging Fly Cell Atlas identifies exhaustive aging features at cellular resolution. *Science* **380,** eadg0934 (2023).

60. Zeng, H., Huang, J., Zhou, H., Meilandt, W. J., Dejanovic, B., Zhou, Y., Bohlen, C. J., Lee, S.-H., Ren, J., Liu, A., Tang, Z., Sheng, H., Liu, J., Sheng, M. & Wang, X. Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26,** 430–446 (2023).

61. Saunders, L. M., Srivatsan, S. R., Duran, M., Dorrity, M. W., Ewing, B., Linbo, T. H., Shendure, J., Raible, D. W., Moens, C. B., Kimelman, D. & Trapnell, C. Embryo-scale reverse genetics at single-cell resolution. *Nature* **623,** 782–791 (2023).

62. Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T., Close, J. L., Long, B., Johansen, N., Penn, O., Yao, Z., Eggermont, J., Höllt, T., Levi, B. P., Shehata, S. I., Aevermann, B., Beller, A., Bertagnolli, D., Brouner, K., Casper, T., Cobbs, C., Dalley, R., Dee, N., Ding, S.-L., Ellenbogen, R. G., Fong, O., Garren, E., Goldy, J., Gwinn, R. P., Hirschstein, D., Keene, C. D., Keshk, M., Ko, A. L., Lathia, K., Mahfouz, A., Maltzer, Z., McGraw, M., Nguyen, T. N., Nyhus, J., Ojemann, J. G., Oldre, A., Parry, S., Reynolds, S., Rimorin, C., Shapovalova, N. V., Somasundaram, S., Szafer, A., Thomsen, E. R., Tieu, M., Quon, G., Scheuermann, R. H., Yuste, R., Sunkin, S. M., Lelieveldt, B., Feng, D., Ng, L., Bernard, A., Hawrylycz, M., Phillips, J. W., Tasic, B., Zeng, H., Jones, A. R., Koch, C. & Lein, E. S. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573,** 61–68 (2019).

63. Bakken, T. E., Jorstad, N. L., Hu, Q., Lake, B. B., Tian, W., Kalmbach, B. E., Crow, M., Hodge, R. D., Krienen, F. M., Sorensen, S. A., Eggermont, J., Yao, Z., Aevermann, B. D., Aldridge, A. I., Bartlett, A., Bertagnolli, D., Casper, T., Castanon, R. G., Crichton, K., Daigle, T. L., Dalley, R., Dee, N., Dembrow, N., Diep, D., Ding, S.-L., Dong, W., Fang, R., Fischer, S., Goldman, M., Goldy, J., Graybuck, L. T., Herb, B. R., Hou, X., Kancherla, J., Kroll, M., Lathia, K., van Lew, B., Li, Y. E., Liu, C. S., Liu, H., Lucero, J. D., Mahurkar, A., McMillen, D., Miller, J. A., Moussa, M., Nery, J. R., Nicovich, P. R., Niu, S.-Y., Orvis, J., Osteen, J. K., Owen, S., Palmer, C. R., Pham, T., Plongthongkum, N., Poirion, O., Reed, N. M., Rimorin, C., Rivkin, A., Romanow, W. J., Sedeño-Cortés, A. E., Siletti, K., Somasundaram, S., Sulc, J., Tieu, M., Torkelson, A., Tung, H., Wang, X., Xie, F., Yanny, A. M., Zhang, R., Ament, S. A., Behrens, M. M., Bravo, H. C., Chun, J., Dobin, A., Gillis, J., Hertzano, R., Hof, P. R., Höllt, T., Horwitz, G. D., Keene, C. D., Kharchenko, P. V., Ko, A. L., Lelieveldt, B. P., Luo, C., Mukamel, E. A., Pinto-Duarte, A., Preissl, S., Regev, A., Ren, B., Scheuermann, R. H., Smith, K., Spain, W. J., White, O. R., Koch, C., Hawrylycz, M., Tasic, B., Macosko, E. Z., McCarroll, S. A., Ting, J. T., Zeng, H., Zhang, K., Feng, G., Ecker, J. R., Linnarsson, S. & Lein, E. S. Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598,** 111–119 (2021).

64. Sousa, A. M. M., Zhu, Y., Raghanti, M. A., Kitchen, R. R., Onorati, M., Tebbenkamp, A. T. N., Stutz, B., Meyer, K. A., Li, M., Kawasawa, Y. I., Liu, F., Perez, R. G., Mele, M., Carvalho, T., Skarica, M., Gulden, F. O., Pletikos, M., Shibata, A., Stephenson, A. R., Edler, M. K., Ely, J. J., Elsworth, J. D., Horvath, T. L., Hof, P. R., Hyde, T. M., Kleinman, J. E., Weinberger, D.

R., Reimers, M., Lifton, R. P., Mane, S. M., Noonan, J. P., State, M. W., Lein, E. S., Knowles, J. A., Marques-Bonet, T., Sherwood, C. C., Gerstein, M. B. & Sestan, N. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358,** 1027–1032 (2017).

65. Chen, Y., Hunter, E., Arbabi, K., Guet-McCreight, A., Consens, M., Felsky, D., Sibille, E. & Tripathy, S. J. Robust differences in cortical cell type proportions across healthy human aging inferred through cross-dataset transcriptome analyses. *Neurobiol. Aging* **125,** 49–61 (2023).

66. Demanelis, K., Jasmine, F., Chen, L. S., Chernoff, M., Tong, L., Delgado, D., Zhang, C., Shinkle, J., Sabarinathan, M., Lin, H., Ramirez, E., Oliva, M., Kim-Hellmuth, S., Stranger, B. E., Lai, T.-P., Aviv, A., Ardlie, K. G., Aguet, F., Ahsan, H., GTEx Consortium, Doherty, J. A., Kibriya, M. G. & Pierce, B. L. Determinants of telomere length across human tissues. *Science* **369,** (2020).

67. De Jager, P. L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., Klein, H.-U., White, C. C., Peters, M. A., Lodgson, B., Nejad, P., Tang, A., Mangravite, L. M., Yu, L., Gaiteri, C., Mostafavi, S., Schneider, J. A. & Bennett, D. A. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5,** 180142 (2018).

68. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345,** 458–460 (1990).

69. Daniali, L., Benetos, A., Susser, E., Kark, J. D., Labat, C., Kimura, M., Desai, K., Granick, M. & Aviv, A. Telomeres shorten at equivalent rates in somatic tissues of adults. *Nat. Commun.* **4,** 1597 (2013).

70. Dong, X., Sun, S., Zhang, L., Kim, S., Tu, Z., Montagna, C., Maslov, A. Y., Suh, Y., Wang, T., Campisi, J. & Vijg, J. Age-related telomere attrition causes aberrant gene expression in sub-telomeric regions. *Aging Cell* **20,** e13357 (2021).

71. Dillman, A. A., Majounie, E., Ding, J., Gibbs, J. R., Hernandez, D., Arepalli, S., Traynor, B. J., Singleton, A. B., Galter, D. & Cookson, M. R. Transcriptomic profiling of the human brain reveals that altered synaptic gene expression is associated with chronological aging. *Sci. Rep.* **7,** 16890 (2017).

72. Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A. M. M., Pletikos, M., Meyer, K. A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M. B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S. N., Vortmeyer, A., Weinberger, D. R., Mane, S., Hyde, T. M., Huttner, A., Reimers, M., Kleinman, J. E. & Sestan, N. Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).

73. Mostafavi, S., Gaiteri, C., Sullivan, S. E., White, C. C., Tasaki, S., Xu, J., Taga, M., Klein, H.-U., Patrick, E., Komashko, V., McCabe, C., Smith, R., Bradshaw, E. M., Root, D. E., Regev, A., Yu, L., Chibnik, L. B., Schneider, J. A., Young-Pearse, T. L., Bennett, D. A. & De Jager, P. L. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21,** 811–819 (2018).

74. Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J. & Yankner, B. A. Gene regulation and DNA damage in the ageing human brain. *Nature* **429,** 883–891 (2004).

75. Yap, E.-L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100,** 330–348 (2018).

76. Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., Poovathingal, S., Hulselmans, G., Spanier, K. I., Moerman, T., Vanspauwen, B., Geurs, S., Voet, T., Lammertyn, J., Thienpont, B., Liu, S., Konstantinides, N., Fiers, M., Verstreken, P. & Aerts, S. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174,** 982–998.e20 (2018).

77. Ximerakis, M., Lipnick, S. L., Innes, B. T., Simmons, S. K., Adiconis, X., Dionne, D., Mayweather, B. A., Nguyen, L., Niziolek, Z., Ozek, C., Butty, V. L., Isserlin, R., Buchanan, S. M., Levine, S. S., Regev, A., Bader, G. D., Levin, J. Z. & Rubin, L. L. Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22,** 1696–1708 (2019).

78. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583,** 590–595 (2020).

79. Chiou, K. L., Huang, X., Bohlen, M. O., Tremblay, S., DeCasien, A. R., O'Day, D. R., Spurrell, C. H., Gogate, A. A., Zintel, T. M., Cayo Biobank Research Unit, Andrews, M. G., Martínez, M. I., Starita, L. M., Montague, M. J., Platt, M. L., Shendure, J. & Snyder-Mackler, N. A single-cell multi-omic atlas spanning the adult rhesus macaque brain. *Sci Adv* **9,** eadh1914 (2023).

80. Zhang, H., Li, J., Ren, J., Sun, S., Ma, S., Zhang, W., Yu, Y., Cai, Y., Yan, K., Li, W., Hu, B., Chan, P., Zhao, G.-G., Belmonte, J. C. I., Zhou, Q., Qu, J., Wang, S. & Liu, G.-H. Single-nucleus transcriptomic landscape of primate hippocampal aging. *Protein Cell* **12,** 695–716 (2021).

81. Gazestani, V., Kamath, T., Nadaf, N. M., Dougalis, A., Burris, S. J., Rooney, B., Junkkari, A., Vanderburg, C., Pelkonen, A., Gomez-Budia, M., Välimäki, N.-N., Rauramaa, T., Therrien, M., Koivisto, A. M., Tegtmeyer, M., Herukka, S.-K., Abdulraouf, A., Marsh, S. E., Hiltunen, M., Nehme, R., Malm, T., Stevens, B., Leinonen, V. & Macosko, E. Z. Early Alzheimer's disease pathology in human cortex involves transient cell states. *Cell* **186,** 4438–4453.e23 (2023).

82. Johansen, N., Somasundaram, S., Travaglini, K. J., Yanny, A. M., Shumyatcher, M., Casper, T., Cobbs, C., Dee, N., Ellenbogen, R., Ferreira, M., Goldy, J., Guzman, J., Gwinn, R., Hirschstein, D., Jorstad, N. L., Keene, C. D., Ko, A., Levi, B. P., Ojemann, J. G., Pham, T., Shapovalova, N., Silbergeld, D., Sulc, J., Torkelson, A., Tung, H., Smith, K., Lein, E. S., Bakken, T. E., Hodge, R. D. & Miller, J. A. Interindividual variation in human cortical cell type abundance and expression. *Science* **382,** eadf2359 (2023).

83. Ling, E., Nemesh, J., Goldman, M., Kamitaki, N., Reed, N., Handsaker, R. E., Genovese, G., Vogelgsang, J. S., Gerges, S., Kashin, S., Ghosh, S., Esposito, J. M., Morris, K., Meyer, D., Lutservitz, A., Mullally, C. D., Wysoker, A., Spina, L., Neumann, A., Hogan, M., Ichihara, K., Berretta, S. & McCarroll, S. A. A concerted neuron-astrocyte program declines in ageing and schizophrenia. *Nature* (2024). doi:10.1038/s41586-024-07109-5

84. Herring, C. A., Simmons, R. K., Freytag, S., Poppe, D., Moffet, J. J. D., Pflueger, J., Buckberry, S., Vargas-Landin, D. B., Clément, O., Echeverría, E. G., Sutton, G. J., Alvarez-Franco, A., Hou, R., Pflueger, C., McDonald, K., Polo, J. M., Forrest, A. R. R., Nowak, A. K., Voineagu, I., Martelotto, L. & Lister, R. Human prefrontal cortex gene

regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185,** 4428–4447.e28 (2022).

85. Soreq, L., UK Brain Expression Consortium, North American Brain Expression Consortium, Rose, J., Soreq, E., Hardy, J., Trabzuni, D., Cookson, M. R., Smith, C., Ryten, M., Patani, R. & Ule, J. Major Shifts in Glial Regional Identity Are a Transcriptional Hallmark of Human Brain Aging. *Cell Rep.* **18,** 557–570 (2017).

86. Wang, B.-A., Jones, J. R., Zhou, J., Tian, W., Wu, Y., Wang, W., Berube, P., Bartlett, A., Castanon, R., Nery, J. R., Chen, H., Kenworthy, M., Altshul, J., Valadon, C., Wang, Y., Kang, A., Goodman, R., Liem, M., Claffey, N., O'Connor, C., Metcalf, J., Luo, C., Gage, F. H. & Ecker, J. R. Epigenome erosion in Alzheimer's disease brain cells and induced neurons. *bioRxiv* 2023.10.15.562394 (2023). doi:10.1101/2023.10.15.562394

87. Dileep, V., Boix, C. A., Mathys, H., Marco, A., Welch, G. M., Meharena, H. S., Loon, A., Jeloka, R., Peng, Z., Bennett, D. A., Kellis, M. & Tsai, L.-H. Neuronal DNA double-strand breaks lead to genome structural variations and 3D genome disruption in neurodegeneration. *Cell* **186,** 4404–4421.e20 (2023).

88. Hernando-Herraez, I., Evano, B., Stubbs, T., Commere, P.-H., Jan Bonder, M., Clark, S., Andrews, S., Tajbakhsh, S. & Reik, W. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat. Commun.* **10,** 4361 (2019).

89. Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R. & Kleinman, J. E. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19,** 40–47 (2016).

90. Price, A. J., Collado-Torres, L., Ivanov, N. A., Xia, W., Burke, E. E., Shin, J. H., Tao, R., Ma, L., Jia, Y., Hyde, T. M., Kleinman, J. E., Weinberger, D. R. & Jaffe, A. E. Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation. *Genome Biol.* **20,** 196 (2019).

91. Lu, A. T., Fei, Z., Haghani, A., Robeck, T. R., Zoller, J. A., Li, C. Z., Lowe, R., Yan, Q., Zhang, J., Vu, H., Ablaeva, J., Acosta-Rodriguez, V. A., Adams, D. M., Almunia, J., Aloysius, A., Ardehali, R., Arneson, A., Baker, C. S., Banks, G., Belov, K., Bennett, N. C., Black, P., Blumstein, D. T., Bors, E. K., Breeze, C. E., Brooke, R. T., Brown, J. L., Carter, G. G., Caulton, A., Cavin, J. M., Chakrabarti, L., Chatzistamou, I., Chen, H., Cheng, K., Chiavellini, P., Choi, O. W., Clarke, S. M., Cooper, L. N., Cossette, M. L., Day, J., DeYoung, J., DiRocco, S., Dold, C., Ehmke, E. E., Emmons, C. K., Emmrich, S., Erbay, E., Erlacher-Reid, C., Faulkes, C. G., Ferguson, S. H., Finno, C. J., Flower, J. E., Gaillard, J. M., Garde, E., Gerber, L., Gladyshev, V. N., Gorbunova, V., Goya, R. G., Grant, M. J., Green, C. B., Hales, E. N., Hanson, M. B., Hart, D. W., Haulena, M., Herrick, K., Hogan, A. N., Hogg, C. J., Hore, T. A., Huang, T., Izpisua Belmonte, J. C., Jasinska, A. J., Jones, G., Jourdain, E., Kashpur, O., Katcher, H., Katsumata, E., Kaza, V., Kiaris, H., Kobor, M. S., Kordowitzki, P., Koski, W. R., Krützen, M., Kwon, S. B., Larison, B., Lee, S. G., Lehmann, M., Lemaitre, J. F., Levine, A. J., Li, C., Li, X., Lim, A. R., Lin, D. T. S., Lindemann, D. M., Little, T. J., Macoretta, N., Maddox, D., Matkin, C. O., Mattison, J. A., McClure, M., Mergl, J., Meudt, J. J., Montano, G. A., Mozhui, K., Munshi-South, J., Naderi, A., Nagy, M., Narayan, P., Nathanielsz, P. W., Nguyen, N. B., Niehrs, C., O'Brien, J. K., O'Tierney Ginn, P., Odom, D. T., Ophir, A. G., Osborn, S., Ostrander, E. A., Parsons, K. M., Paul, K. C., Pellegrini, M., Peters, K. J., Pedersen, A. B., Petersen, J. L., Pietersen, D. W., Pinho, G. M.,

Plassais, J., Poganik, J. R., Prado, N. A., Reddy, P., Rey, B., Ritz, B. R., Robbins, J., Rodriguez, M., Russell, J., Rydkina, E., Sailer, L. L., Salmon, A. B., Sanghavi, A., Schachtschneider, K. M., Schmitt, D., Schmitt, T., Schomacher, L., Schook, L. B., Sears, K. E., Seifert, A. W., Seluanov, A., Shafer, A. B. A., Shanmuganayagam, D., Shindyapina, A. V., Simmons, M., Singh, K., Sinha, I., Slone, J., Snell, R. G., Soltanmaohammadi, E., Spangler, M. L., Spriggs, M. C., Staggs, L., Stedman, N., Steinman, K. J., Stewart, D. T., Sugrue, V. J., Szladovits, B., Takahashi, J. S., Takasugi, M., Teeling, E. C., Thompson, M. J., Van Bonn, B., Vernes, S. C., Villar, D., Vinters, H. V., Wallingford, M. C., Wang, N., Wayne, R. K., Wilkinson, G. S., Williams, C. K., Williams, R. W., Yang, X. W., Yao, M., Young, B. G., Zhang, B., Zhang, Z., Zhao, P., Zhao, Y., Zhou, W., Zimmermann, J., Ernst, J., Raj, K. & Horvath, S. Universal DNA methylation age across mammalian tissues. *Nat Aging* **3,** 1144–1166 (2023).

92. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T. & Zhang, K. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49,** 359–367 (2013).

93. Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T. & Adams, P. D. DNA Methylation Clocks in Aging: Categories, Causes, and Consequences. *Mol. Cell* **71,** 882–895 (2018).

94. Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., Gibson, J., Henders, A. K., Redmond, P., Cox, S. R., Pattie, A., Corley, J., Murphy, L., Martin, N. G., Montgomery, G. W., Feinberg, A. P., Fallin, M. D., Multhaup, M. L., Jaffe, A. E., Joehanes, R., Schwartz, J., Just, A. C., Lunetta, K. L., Murabito, J. M., Starr, J. M., Horvath, S., Baccarelli, A. A., Levy, D., Visscher, P. M., Wray, N. R. & Deary, I. J. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **16,** 25 (2015).

95. Gross, A. M., Jaeger, P. A., Kreisberg, J. F., Licon, K., Jepsen, K. L., Khosroheidari, M., Morsey, B. M., Swindells, S., Shen, H., Ng, C. T., Flagg, K., Chen, D., Zhang, K., Fox, H. S. & Ideker, T. Methylome-wide Analysis of Chronic HIV Infection Reveals Five-Year Increase in Biological Age and Epigenetic Targeting of HLA. *Mol. Cell* **62,** 157–168 (2016).

96. Horvath, S. & Levine, A. J. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J. Infect. Dis.* **212,** 1563–1573 (2015).

97. Yang, Z., Wong, A., Kuh, D., Paul, D. S., Rakyan, V. K., Leslie, R. D., Zheng, S. C., Widschwendter, M., Beck, S. & Teschendorff, A. E. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* **17,** 205 (2016).

98. Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G.-L. & Song, H. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17,** 215–222 (2014).

99. Luo, C., Liu, H., Xie, F., Armand, E. J., Siletti, K. & Bakken, T. E. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics* (2022). at <https://www.sciencedirect.com/science/article/pii/S2666979X22000271>

100. Li, J., Jaiswal, M. K., Chien, J.-F., Kozlenkov, A., Jung, J., Zhou, P., Gardashli, M., Pregent, L. J., Engelberg-Cook, E., Dickson, D. W., Belzil, V. V., Mukamel, E. A. & Dracheva, S. Divergent single cell transcriptome and epigenome alterations in ALS and FTD patients with

C9orf72 mutation. *Nat. Commun.* **14,** 1–22 (2023).

101. Cain, A., Taga, M., McCabe, C., Green, G. S., Hekselman, I., White, C. C., Lee, D. I., Gaur, P., Rozenblatt-Rosen, O., Zhang, F., Yeger-Lotem, E., Bennett, D. A., Yang, H.-S., Regev, A., Menon, V., Habib, N. & De Jager, P. L. Multicellular communities are perturbed in the aging human brain and Alzheimer's disease. *Nat. Neurosci.* **26,** 1267–1280 (2023).

102. Lauterborn, J. C., Scaduto, P., Cox, C. D., Schulmann, A., Lynch, G., Gall, C. M., Keene, C. D. & Limon, A. Increased excitatory to inhibitory synaptic ratio in parietal cortex samples from individuals with Alzheimer's disease. *Nat. Commun.* **12,** 2603 (2021).

103. Gabitto, M., Travaglini, K., Ariza, J., Kaplan, E., Long, B., Rachleff, V., Ding, Y., Mahoney, J., Dee, N., Goldy, J., Melief, E., Brouner, K., Compos, J., Campos, J., Carr, A., Casper, T., Chakrabarty, R., Clark, M., Cool, J., Dalley, R., Darvas, M., Dolbeare, T., Ding, S.-L., Egdorf, T., Esposito, L., Ferrer, R., Gala, R., Gary, A., Gloe, J., Guilford, N., Guzman, J., Ho, W., Jarsky, T., Johansen, N., Kalmbach, B., Keene, L., Khawand, S., Kilgore, M., Kirkland, A., Kunst, M., Lee, B., Donald, C. M., Malone, J., Maltzer, Z., Martin, N., McCue, R., McMillen, D., Meyerdierks, E., Meyers, K., Mollenkopf, T., Montine, M., Nolan, A., Nyhus, J., Olsen, P., Pacleb, M., Pham, T., Pom, C., Postupna, N., Ruiz, A., Schantz, A., Sorensen, S., Staats, B., Sullivan, M., Sunkin, S., Thompson, C., Tieu, M., Ting, J., Torkelson, A., Tran, T., Cuevas, N. V., Wang, M.-Q., Waters, J., Wilson, A., Haynor, D., Gatto, N., Jayadev, S., Mufti, S., Ng, L., Mukherjee, S., Crane, P., Latimer, C., Levi, B., Smith, K., Close, J., Miller, J., Hodge, R., Larson, E., Grabowski, T., Hawrylycz, M., Keene, C. & Lein, E. Integrated multimodal cell atlas of Alzheimer's disease. *Res Sq* (2023). doi:10.21203/rs.3.rs-2921860/v1

104. Kozlenkov, A., Li, J., Apontes, P., Hurd, Y. L., Byne, W. M., Koonin, E. V., Wegner, M., Mukamel, E. A. & Dracheva, S. A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* **4,** eaau6190 (2018).

105. Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J. K., Nery, J. R., Chen, H., Rivkin, A., Castanon, R. G., Clock, B., Li, Y. E., Hou, X., Poirion, O. B., Preissl, S., Pinto-Duarte, A., O'Connor, C., Boggeman, L., Fitzpatrick, C., Nunn, M., Mukamel, E. A., Zhang, Z., Callaway, E. M., Ren, B., Dixon, J. R., Behrens, M. M. & Ecker, J. R. DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* **598,** 120–128 (2021).

106. Junttila, S., Smolander, J. & Elo, L. L. Benchmarking methods for detecting differential states between conditions from multi-subject single-cell RNA-seq data. *Brief. Bioinform.* **23,** (2022).

107. Mukamel, E. A. Multiple Comparisons and Inappropriate Statistical Testing Lead to Spurious Sex Differences in Gene Expression. *Biol. Psychiatry* **91,** e1–e2 (2022).

108. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14,** R115 (2013).

109. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42,** e69 (2014).

110. Park, Y. & Wu, H. Differential methylation analysis for BS-seq data under general

experimental design. *Bioinformatics* **32,** 1446–1453 (2016).

111. Skinnider, M. A., Squair, J. W., Kathe, C., Anderson, M. A., Gautier, M., Matson, K. J. E., Milano, M., Hutson, T. H., Barraud, Q., Phillips, A. A., Foster, L. J., La Manno, G., Levine, A. J. & Courtine, G. Cell type prioritization in single-cell data. *Nat. Biotechnol.* (2020). doi:10.1038/s41587-020-0605-1

112. Kwon, H.-B., Kozorovitskiy, Y., Oh, W.-J., Peixoto, R. T., Akhtar, N., Saulnier, J. L., Gu, C. & Sabatini, B. L. Neuroligin-1-dependent competition regulates cortical synaptogenesis and synapse number. *Nat. Neurosci.* **15,** 1667–1674 (2012).

113. Hamada, N., Ogaya, S., Nakashima, M., Nishijo, T., Sugawara, Y., Iwamoto, I., Ito, H., Maki, Y., Shirai, K., Baba, S., Maruyama, K., Saitsu, H., Kato, M., Matsumoto, N., Momiyama, T. & Nagata, K.-I. De novo PHACTR1 mutations in West syndrome and their pathophysiological effects. *Brain* **141,** 3098–3114 (2018).

114. Shi, W., Wymore, R. S., Wang, H. S., Pan, Z., Cohen, I. S., McKinnon, D. & Dixon, J. E. Identification of two nervous system-specific members of the erg potassium channel gene family. *J. Neurosci.* **17,** 9423–9432 (1997).

115. Wood, S. H., van Dam, S., Craig, T., Tacutu, R., O'Toole, A., Merry, B. J. & de Magalhães, J. P. Transcriptome analysis in calorie-restricted rats implicates epigenetic and post-translational mechanisms in neuroprotection and aging. *Genome Biol.* **16,** 285 (2015).

116. Jin, K., Yao, Z., van Velthoven, C. T. J., Kaplan, E. S., Glattfelder, K., Barlow, S. T., Boyer, G., Carey, D., Casper, T., Chakka, A. B., Chakrabarty, R., Clark, M., Departee, M., Desierto, M., Gary, A., Gloe, J., Goldy, J., Guilford, N., Guzman, J., Hirschstein, D., Lee, C., Liang, E., Pham, T., Reding, M., Ronellenfitch, K., Ruiz, A., Sevigny, J., Shapovalova, N., Shulga, L., Sulc, J., Torkelson, A., Tung, H., Levi, B., Sunkin, S. M., Dee, N., Esposito, L., Smith, K., Tasic, B. & Zeng, H. Cell-type specific molecular signatures of aging revealed in a brain-wide transcriptomic cell-type atlas. *bioRxiv* 2023.07.26.550355 (2023). doi:10.1101/2023.07.26.550355

117. Graham, L. C., Naldrett, M. J., Kohama, S. G., Smith, C., Lamont, D. J., McColl, B. W., Gillingwater, T. H., Skehel, P., Urbanski, H. F. & Wishart, T. M. Regional Molecular Mapping of Primate Synapses during Normal Healthy Aging. *Cell Rep.* **27,** 1018–1026.e4 (2019).

118. Kozlova, I., Sah, S., Keable, R., Leshchyns'ka, I., Janitz, M. & Sytnyk, V. Cell Adhesion Molecules and Protein Synthesis Regulation in Neurons. *Front. Mol. Neurosci.* **13,** 592126 (2020).

119. Bird, A. W., Yu, D. Y., Pray-Grant, M. G., Qiu, Q., Harmon, K. E., Megee, P. C., Grant, P. A., Smith, M. M. & Christman, M. F. Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419,** 411–415 (2002).

120. Nativio, R., Donahue, G., Berson, A., Lan, Y., Amlie-Wolf, A., Tuzer, F., Toledo, J. B., Gosai, S. J., Gregory, B. D., Torres, C., Trojanowski, J. Q., Wang, L.-S., Johnson, F. B., Bonini, N. M. & Berger, S. L. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat. Neurosci.* **21,** 497–505 (2018).

121. Mo, A., Mukamel, E. A., Davis, F. P., Luo, C., Henry, G. L., Picard, S., Urich, M. A., Nery, J. R., Sejnowski, T. J., Lister, R., Eddy, S. R., Ecker, J. R. & Nathans, J. Epigenomic

Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86,** 1369–1384 (2015).

122. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y.-C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J. M., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L.-H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T. & Kellis, M. Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).

123. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).

124. Duclot, F. & Kabbaj, M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity and Neuropsychiatric Disorders. *Front. Behav. Neurosci.* **11,** 35 (2017).

125. Jones, M. W., Errington, M. L., French, P. J., Fine, A., Bliss, T. V., Garel, S., Charnay, P., Bozon, B., Laroche, S. & Davis, S. A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. *Nat. Neurosci.* **4,** 289–296 (2001).

126. Penner, M. R., Parrish, R. R., Hoang, L. T., Roth, T. L., Lubin, F. D. & Barnes, C. A. Age-related changes in Egr1 transcription and DNA methylation within the hippocampus. *Hippocampus* **26,** 1008–1020 (2016).

127. Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J.-B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T. & Zhang, K. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49,** 359–367 (2013).

128. Xie, W., Schultz, M. D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J. W., Tian, S., Hawkins, R. D., Leung, D., Yang, H., Wang, T., Lee, A. Y., Swanson, S. A., Zhang, J., Zhu, Y., Kim, A., Nery, J. R., Urich, M. A., Kuan, S., Yen, C.-A., Klugman, S., Yu, P., Suknuntha, K., Propson, N. E., Chen, H., Edsall, L. E., Wagner, U., Li, Y., Ye, Z., Kulkarni, A., Xuan, Z., Chung, W.-Y., Chi, N. C., Antosiewicz-Bourget, J. E., Slukvin, I., Stewart, R., Zhang, M. Q., Wang, W., Thomson, J. A., Ecker, J. R. & Ren, B. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).

129. Zannas, A. S., Jia, M., Hafner, K., Baumert, J., Wiechmann, T., Pape, J. C., Arloth, J., Ködel, M., Martinelli, S., Roitman, M., Röh, S., Haehle, A., Emeny, R. T., Iurato, S., Carrillo-Roa, T., Lahti, J., Räikkönen, K., Eriksson, J. G., Drake, A. J., Waldenberger, M., Wahl, S., Kunze, S., Lucae, S., Bradley, B., Gieger, C., Hausch, F., Smith, A. K., Ressler, K.

J., Müller-Myhsok, B., Ladwig, K.-H., Rein, T., Gassen, N. C. & Binder, E. B. Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB-driven inflammation and cardiovascular risk. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 11370–11379 (2019).

130. Blair, L. J., Nordhues, B. A., Hill, S. E., Scaglione, K. M., O'Leary, J. C., 3rd, Fontaine, S. N., Breydo, L., Zhang, B., Li, P., Wang, L., Cotman, C., Paulson, H. L., Muschol, M., Uversky, V. N., Klengel, T., Binder, E. B., Kayed, R., Golde, T. E., Berchtold, N. & Dickey, C. A. Accelerated neurodegeneration through chaperone-mediated oligomerization of tau. *J. Clin. Invest.* **123,** 4158–4169 (2013).

131. Fullerton, E. F., Karom, M. C., Streicher, J. M., Young, L. J. & Murphy, A. Z. Age-Induced Changes in μ-Opioid Receptor Signaling in the Midbrain Periaqueductal Gray of Male and Female Rats. *J. Neurosci.* **42,** 6232–6242 (2022).

132. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153,** 1194–1217 (2013).

133. Feuerbach, L., Sieverling, L., Deeg, K. I., Ginsbach, P., Hutter, B., Buchhalter, I., Northcott, P. A., Mughal, S. S., Chudasama, P., Glimm, H., Scholl, C., Lichter, P., Fröhling, S., Pfister, S. M., Jones, D. T. W., Rippe, K. & Brors, B. TelomereHunter - in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics* **20,** 272 (2019).

134. Robin, J. D., Ludlow, A. T., Batten, K., Magdinier, F., Stadler, G., Wagner, K. R., Shay, J. W. & Wright, W. E. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* **28,** 2464–2476 (2014).

135. Lopes-Ramos, C. M., Chen, C.-Y., Kuijjer, M. L., Paulson, J. N., Sonawane, A. R., Fagny, M., Platig, J., Glass, K., Quackenbush, J. & DeMeo, D. L. Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep.* **31,** 107795 (2020).

136. Fu, X.-M. & Zhu, B. T. Human pancreas-specific protein disulfide isomerase homolog (PDIp) is an intracellular estrogen-binding protein that modulates estrogen levels and actions in target cells. *J. Steroid Biochem. Mol. Biol.* **115,** 20–29 (2009).

137. Gusev, F. E., Reshetov, D. A., Mitchell, A. C., Andreeva, T. V., Dincer, A., Grigorenko, A. P., Fedonin, G., Halene, T., Aliseychik, M., Filippova, E., Weng, Z., Akbarian, S. & Rogaev, E. I. Chromatin profiling of cortical neurons identifies individual epigenetic signatures in schizophrenia. *Transl. Psychiatry* **9,** 256 (2019).

138. Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B. B., Castel, S. E., Karczewski, K. J., Aguet, F., Byrnes, A., GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lappalainen, T., Regev, A., Ardlie, K. G., Hacohen, N. & MacArthur, D. G. Landscape of X

chromosome inactivation across human tissues. *Nature* **550,** 244–248 (2017).

139. Velmeshev, D., Perez, Y., Yan, Z., Valencia, J. E., Castaneda-Castellanos, D. R., Wang, L., Schirmer, L., Mayer, S., Wick, B., Wang, S., Nowakowski, T. J., Paredes, M., Huang, E. J. & Kriegstein, A. R. Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382,** eadf0834 (2023).

140. De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., Eaton, M. L., Keenan, B. T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H. S., Younkin, C., Younkin, S. G., Zou, F., Szyf, M., Epstein, C. B., Schneider, J. A., Bernstein, B. E., Meissner, A., Ertekin-Taner, N., Chibnik, L. B., Kellis, M., Mill, J. & Bennett, D. A. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17,** 1156–1163 (2014).

141. Oliva, M., Demanelis, K., Lu, Y., Chernoff, M., Jasmine, F., Ahsan, H., Kibriya, M. G., Chen, L. S. & Pierce, B. L. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nat. Genet.* **55,** 112–122 (2023).

142. Rizzardi, L. F., Hickey, P. F., Idrizi, A., Tryggvadóttir, R., Callahan, C. M., Stephens, K. E., Taverna, S. D., Zhang, H., Ramazanoglu, S., GTEx Consortium, Hansen, K. D. & Feinberg, A. P. Human brain region-specific variably methylated regions are enriched for heritability of distinct neuropsychiatric traits. *Genome Biol.* **22,** 116 (2021).

143. Haghani, A., Li, C. Z., Robeck, T. R., Zhang, J., Lu, A. T., Ablaeva, J., Acosta-Rodríguez, V. A., Adams, D. M., Alagaili, A. N., Almunia, J., Aloysius, A., Amor, N. M. S., Ardehali, R., Arneson, A., Baker, C. S., Banks, G., Belov, K., Bennett, N. C., Black, P., Blumstein, D. T., Bors, E. K., Breeze, C. E., Brooke, R. T., Brown, J. L., Carter, G., Caulton, A., Cavin, J. M., Chakrabarti, L., Chatzistamou, I., Chavez, A. S., Chen, H., Cheng, K., Chiavellini, P., Choi, O.-W., Clarke, S., Cook, J. A., Cooper, L. N., Cossette, M.-L., Day, J., DeYoung, J., Dirocco, S., Dold, C., Dunnum, J. L., Ehmke, E. E., Emmons, C. K., Emmrich, S., Erbay, E., Erlacher-Reid, C., Faulkes, C. G., Fei, Z., Ferguson, S. H., Finno, C. J., Flower, J. E., Gaillard, J.-M., Garde, E., Gerber, L., Gladyshev, V. N., Goya, R. G., Grant, M. J., Green, C. B., Hanson, M. B., Hart, D. W., Haulena, M., Herrick, K., Hogan, A. N., Hogg, C. J., Hore, T. A., Huang, T., Izpisua Belmonte, J. C., Jasinska, A. J., Jones, G., Jourdain, E., Kashpur, O., Katcher, H., Katsumata, E., Kaza, V., Kiaris, H., Kobor, M. S., Kordowitzki, P., Koski, W. R., Krützen, M., Kwon, S. B., Larison, B., Lee, S.-G., Lehmann, M., Lemaître, J.-F., Levine, A. J., Li, X., Li, C., Lim, A. R., Lin, D. T. S., Lindemann, D. M., Liphardt, S. W., Little, T. J., Macoretta, N., Maddox, D., Matkin, C. O., Mattison, J. A., McClure, M., Mergl, J., Meudt, J. J., Montano, G. A., Mozhui, K., Munshi-South, J., Murphy, W. J., Naderi, A., Nagy, M., Narayan, P., Nathanielsz, P. W., Nguyen, N. B., Niehrs, C., Nyamsuren, B., O'Brien, J. K., Ginn, P. O., Odom, D. T., Ophir, A. G., Osborn, S., Ostrander, E. A., Parsons, K. M., Paul, K. C., Pedersen, A. B., Pellegrini, M., Peters, K. J., Petersen, J. L., Pietersen, D. W., Pinho, G. M., Plassais, J., Poganik, J. R., Prado, N. A., Reddy, P., Rey, B., Ritz, B. R., Robbins, J., Rodriguez, M., Russell, J., Rydkina, E., Sailer, L. L., Salmon, A. B., Sanghavi, A., Schachtschneider, K. M., Schmitt, D., Schmitt, T., Schomacher, L., Schook, L. B., Sears, K. E., Seifert, A. W., Shafer, A. B. A., Shindyapina, A. V., Simmons, M., Singh, K., Sinha, I., Slone, J., Snell, R. G., Soltanmohammadi, E., Spangler, M. L., Spriggs, M., Staggs, L., Stedman, N., Steinman, K. J., Stewart, D. T., Sugrue, V. J., Szladovits, B., Takahashi, J. S., Takasugi, M., Teeling, E. C., Thompson, M. J., Van Bonn, B., Vernes, S. C., Villar, D., Vinters, H. V., Vu, H., Wallingford, M. C., Wang, N., Wilkinson, G. S., Williams, R. W., Yan,

Q., Yao, M., Young, B. G., Zhang, B., Zhang, Z., Zhao, Y., Zhao, P., Zhou, W., Zoller, J. A., Ernst, J., Seluanov, A., Gorbunova, V., Yang, X. W., Raj, K. & Horvath, S. DNA methylation networks underlying mammalian traits. *Science* **381,** eabq5693 (2023).

144. Luo, C., Liu, H., Xie, F., Armand, E. J., Siletti, K., Bakken, T. E., Fang, R., Doyle, W. I., Stuart, T., Hodge, R. D., Hu, L., Wang, B.-A., Zhang, Z., Preissl, S., Lee, D.-S., Zhou, J., Niu, S.-Y., Castanon, R., Bartlett, A., Rivkin, A., Wang, X., Lucero, J., Nery, J. R., Davis, D. A., Mash, D. C., Satija, R., Dixon, J. R., Linnarsson, S., Lein, E., Behrens, M. M., Ren, B., Mukamel, E. A. & Ecker, J. R. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom* **2,** (2022).

145. Sun, Z., Xu, X., He, J., Murray, A., Sun, M.-A., Wei, X., Wang, X., McCoig, E., Xie, E., Jiang, X., Li, L., Zhu, J., Chen, J., Morozov, A., Pickrell, A. M., Theus, M. H. & Xie, H. EGR1 recruits TET1 to shape the brain methylome during development and upon neuronal activity. *Nat. Commun.* **10,** 3892 (2019).

146. Ain, Q., Schmeer, C., Penndorf, D., Fischer, M., Bondeva, T., Förster, M., Haenold, R., Witte, O. W. & Kretz, A. Cell cycle-dependent and -independent telomere shortening accompanies murine brain aging. *Aging* **10,** 3397–3420 (2018).

147. Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., Larsen, R., Casper, T., Barkan, E., Kroll, M., Parry, S., Shapovalova, N. V., Hirschstein, D., Pendergraft, J., Sullivan, H. A., Kim, T. K., Szafer, A., Dee, N., Groblewski, P., Wickersham, I., Cetin, A., Harris, J. A., Levi, B. P., Sunkin, S. M., Madisen, L., Daigle, T. L., Looger, L., Bernard, A., Phillips, J., Lein, E., Hawrylycz, M., Svoboda, K., Jones, A. R., Koch, C. & Zeng, H. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563,** 72–78 (2018).

148. Blokland, G. A. M., Grove, J., Chen, C.-Y., Cotsapas, C., Tobet, S., Handa, R., Schizophrenia Working Group of the Psychiatric Genomics Consortium, St Clair, D., Lencz, T., Mowry, B. J., Periyasamy, S., Cairns, M. J., Tooney, P. A., Wu, J. Q., Kelly, B., Kirov, G., Sullivan, P. F., Corvin, A., Riley, B. P., Esko, T., Milani, L., Jönsson, E. G., Palotie, A., Ehrenreich, H., Begemann, M., Steixner-Kumar, A., Sham, P. C., Iwata, N., Weinberger, D. R., Gejman, P. V., Sanders, A. R., Buxbaum, J. D., Rujescu, D., Giegling, I., Konte, B., Hartmann, A. M., Bramon, E., Murray, R. M., Pato, M. T., Lee, J., Melle, I., Molden, E., Ophoff, R. A., McQuillin, A., Bass, N. J., Adolfsson, R., Malhotra, A. K., Bipolar Disorder Working Group of the Psychiatric Genomics Consortium, Martin, N. G., Fullerton, J. M., Mitchell, P. B., Schofield, P. R., Forstner, A. J., Degenhardt, F., Schaupp, S., Comes, A. L., Kogevinas, M., Guzman-Parra, J., Reif, A., Streit, F., Sirignano, L., Cichon, S., Grigoroiu-Serbanescu, M., Hauser, J., Lissowska, J., Mayoral, F., Müller-Myhsok, B., Świątkowska, B., Schulze, T. G., Nöthen, M. M., Rietschel, M., Kelsoe, J., Leboyer, M., Jamain, S., Etain, B., Bellivier, F., Vincent, J. B., Alda, M., O'Donovan, C., Cervantes, P., Biernacka, J. M., Frye, M., McElroy, S. L., Scott, L. J., Stahl, E. A., Landén, M., Hamshere, M. L., Smeland, O. B., Djurovic, S., Vaaler, A. E., Andreassen, O. A., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Baune, B. T., Air, T., Preisig, M., Uher, R., Levinson, D. F., Weissman, M. M., Potash, J. B., Shi, J., Knowles, J. A., Perlis, R. H., Lucae, S., Boomsma, D. I., Penninx, B. W. J. H., Hottenga, J.-J., de Geus, E. J. C., Willemsen, G., Milaneschi, Y., Tiemeier, H., Grabe, H. J., Teumer, A., Van der Auwera, S., Völker, U., Hamilton, S. P., Magnusson, P. K. E., Viktorin, A., Mehta, D., Mullins, N., Adams, M. J., Breen, G., McIntosh, A. M., Lewis, C. M., Sex Differences

Cross-Disorder Analysis Group of the Psychiatric Genomics Consortium, iPSYCH, Hougaard, D. M., Nordentoft, M., Mors, O., Mortensen, P. B., Werge, T., Als, T. D., Børglum, A. D., Petryshen, T. L., Smoller, J. W. & Goldstein, J. M. Sex-Dependent Shared and Nonshared Genetic Architecture Across Mood and Psychotic Disorders. *Biol. Psychiatry* **91,** 102–117 (2022).

149. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598,** 86–102 (2021).

150. Tian, W., Zhou, J., Bartlett, A., Zeng, Q., Liu, H., Castanon, R. G., Kenworthy, M., Altshul, J., Valadon, C., Aldridge, A., Nery, J. R., Chen, H., Xu, J., Johnson, N. D., Lucero, J., Osteen, J. K., Emerson, N., Rink, J., Lee, J., Li, Y. E., Siletti, K., Liem, M., Claffey, N., O'Connor, C., Yanny, A. M., Nyhus, J., Dee, N., Casper, T., Shapovalova, N., Hirschstein, D., Ding, S.-L., Hodge, R., Levi, B. P., Keene, C. D., Linnarsson, S., Lein, E., Ren, B., Behrens, M. M. & Ecker, J. R. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382,** eadf5357 (2023).

151. Luo, C., Rivkin, A., Zhou, J., Sandoval, J. P., Kurihara, L., Lucero, J., Castanon, R., Nery, J. R., Pinto-Duarte, A., Bui, B., Fitzpatrick, C., O'Connor, C., Ruga, S., Van Eden, M. E., Davis, D. A., Mash, D. C., Behrens, M. M. & Ecker, J. R. Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.* **9,** 3824 (2018).

152. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

153. Zhang, Y., Park, C., Bennett, C., Thornton, M. & Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* (2021). doi:10.1101/gr.275193.120

154. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37,** 907–915 (2019).

155. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).

156. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

157. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. ('O'Reilly Media, Inc.', 2020).

158. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8,** 281–291.e9 (2019).

159. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9,** 9354 (2019).

160. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9,** 5233 (2019).

161. Lemaître, G., Nogueira, F. & Aridas, C. K. Imbalanced-learn: A Python Toolbox to Tackle the

Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **18,** 1–5 (2017).

162. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37,** 192–201 (2021).

163. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16,** 284–287 (2012).

164. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41,** e155 (2013).

165. Wen, L., Li, X., Yan, L., Tan, Y., Li, R., Zhao, Y., Wang, Y., Xie, J., Zhang, Y., Song, C., Yu, M., Liu, X., Zhu, P., Li, X., Hou, Y., Guo, H., Wu, X., He, C., Li, R., Tang, F. & Qiao, J. Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15,** R49 (2014).

166. Szulwach, K. E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A. K., Gearing, M., Levey, A. I., Vasanthakumar, A., Godley, L. A., Chang, Q., Cheng, X., He, C. & Jin, P. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14,** 1607–1616 (2011).

167. Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., Chandran, S., Hay, D. C., Bradley, M., Wilmut, I. & De Sousa, P. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res.* **21,** 1332–1342 (2011).

168. Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P. & He, C. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29,** 68–72 (2011).

169. Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L. & Ren, B. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148,** 816–831 (2012).

170. Guo, J. U., Su, Y., Zhong, C., Ming, G.-L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434 (2011).

171. Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. & Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).

172. Rudenko, A., Dawlaty, M. M., Seo, J., Cheng, A. W., Meng, J., Le, T., Faull, K. F., Jaenisch, R. & Tsai, L.-H. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79,** 1109–1122 (2013).

173. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikov, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T. & Vermeulen, M. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).

174. Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U. & Zoghbi, H. Y. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23,** 185–188 (1999).

175. Coppieters, N., Dieriks, B. V., Lill, C., Faull, R. L. M., Curtis, M. A. & Dragunow, M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35,** 1334–1344 (2014).

176. Zhao, J., Zhu, Y., Yang, J., Li, L., Wu, H., De Jager, P. L., Jin, P. & Bennett, D. A. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers. Dement.* **13,** 674–688 (2017).

177. Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R. S., Aldridge, A. I., Ament, S. A., Bartlett, A., Behrens, M. M., Van den Berge, K., Bertagnolli, D., de Bézieux, H. R., Biancalani, T., Booeshaghi, A. S., Bravo, H. C., Casper, T., Colantuoni, C., Crabtree, J., Creasy, H., Crichton, K., Crow, M., Dee, N., Dougherty, E. L., Doyle, W. I., Dudoit, S., Fang, R., Felix, V., Fong, O., Giglio, M., Goldy, J., Hawrylycz, M., Herb, B. R., Hertzano, R., Hou, X., Hu, Q., Kancherla, J., Kroll, M., Lathia, K., Li, Y. E., Lucero, J. D., Luo, C., Mahurkar, A., McMillen, D., Nadaf, N. M., Nery, J. R., Nguyen, T. N., Niu, S.-Y., Ntranos, V., Orvis, J., Osteen, J. K., Pham, T., Pinto-Duarte, A., Poirion, O., Preissl, S., Purdom, E., Rimorin, C., Risso, D., Rivkin, A. C., Smith, K., Street, K., Sulc, J., Svensson, V., Tieu, M., Torkelson, A., Tung, H., Vaishnav, E. D., Vanderburg, C. R., van Velthoven, C., Wang, X., White, O. R., Huang, Z. J., Kharchenko, P. V., Pachter, L., Ngai, J., Regev, A., Tasic, B., Welch, J. D., Gillis, J., Macosko, E. Z., Ren, B., Ecker, J. R., Zeng, H. & Mukamel, E. A. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598,** 103–110 (2021).

178. Liu, H., Zeng, Q., Zhou, J., Bartlett, A., Wang, B.-A., Berube, P., Tian, W., Kenworthy, M., Altshul, J., Nery, J. R., Chen, H., Castanon, R. G., Zu, S., Li, Y. E., Lucero, J., Osteen, J. K., Pinto-Duarte, A., Lee, J., Rink, J., Cho, S., Emerson, N., Nunn, M., O'Connor, C., Wu, Z., Stoica, I., Yao, Z., Smith, K. A., Tasic, B., Luo, C., Dixon, J. R., Zeng, H., Ren, B., Behrens, M. M. & Ecker, J. R. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* **624,** 366–377 (2023).

179. Chien, J.-F., Liu, H., Wang, B.-A., Luo, C., Bartlett, A., Castanon, R., Johnson, N. D., Nery, J. R., Osteen, J., Li, J., Altshul, J., Kenworthy, M., Valadon, C., Liem, M., Claffey, N., O'Connor, C., Seeker, L. A., Ecker, J. R., Behrens, M. M. & Mukamel, E. A. Cell-type-specific effects of age and sex on human cortical neurons. *Neuron* (2024). doi:10.1016/j.neuron.2024.05.013

180. Bai, D., Zhang, X., Xiang, H., Guo, Z., Zhu, C. & Yi, C. Simultaneous single-cell analysis of 5mC and 5hmC with SIMPLE-seq. *Nat. Biotechnol.* (2024). doi:10.1038/s41587-024-02148-9

181. Cao, Y., Bai, Y., Yuan, T., Song, L., Fan, Y., Ren, L., Song, W., Peng, J., An, R., Gu, Q., Zheng, Y. & Xie, X. S. Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proc. Natl. Acad. Sci. U. S. A.* **120,** e2310367120 (2023).

182. Fabyanic, E. B., Hu, P., Qiu, Q., Berríos, K. N., Connolly, D. R., Wang, T., Flournoy, J., Zhou, Z., Kohli, R. M. & Wu, H. Joint single-cell profiling resolves 5mC and 5hmC and reveals their distinct gene regulatory effects. *Nat. Biotechnol.* (2023). doi:10.1038/s41587-023-01909-2

183. Viswanathan, R., Cheruba, E., Wong, P.-M., Yi, Y., Ngang, S., Chong, D. Q., Loh, Y.-H., Tan, I. B. & Cheow, L. F. DARESOME enables concurrent profiling of multiple DNA modifications with restriction enzymes in single cells and cell-free DNA. *Sci Adv* **9,** eadi0197 (2023).

184. Kozlenkov, A., Roussos, P., Timashpolsky, A., Barbu, M., Rudchenko, S., Bibikova, M., Klotzle, B., Byne, W., Lyddon, R., Di Narzo, A. F., Hurd, Y. L., Koonin, E. V. & Dracheva, S. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Res.* **42,** 109–127 (2014).

185. Kozlenkov, A., Wang, M., Roussos, P., Rudchenko, S., Barbu, M., Bibikova, M., Klotzle, B., Dwork, A. J., Zhang, B., Hurd, Y. L., Koonin, E. V., Wegner, M. & Dracheva, S. Substantial DNA methylation differences between two major neuronal subtypes in human brain. *Nucleic Acids Res.* **44,** 2593–2612 (2016).

186. Laguesse, S., Creppe, C., Nedialkova, D. D., Prévot, P.-P., Borgs, L., Huysseune, S., Franco, B., Duysens, G., Krusy, N., Lee, G., Thelen, N., Thiry, M., Close, P., Chariot, A., Malgrange, B., Leidel, S. A., Godin, J. D. & Nguyen, L. A Dynamic Unfolded Protein Response Contributes to the Control of Cortical Neurogenesis. *Dev. Cell* **35,** 553–567 (2015).

187. Shinmyo, Y., Asrafuzzaman Riyadh, M., Ahmed, G., Bin Naser, I., Hossain, M., Takebayashi, H., Kawasaki, H., Ohta, K. & Tanaka, H. Draxin from neocortical neurons controls the guidance of thalamocortical projections into the neocortex. *Nat. Commun.* **6,** 10232 (2015).

188. Islam, S. M., Shinmyo, Y., Okafuji, T., Su, Y., Naser, I. B., Ahmed, G., Zhang, S., Chen, S., Ohta, K., Kiyonari, H., Abe, T., Tanaka, S., Nishinakamura, R., Terashima, T., Kitamura, T. & Tanaka, H. Draxin, a repulsive guidance protein for spinal cord and forebrain commissures. *Science* **323,** 388–393 (2009).

189. Loerch, P. M., Lu, T., Dakin, K. A., Vann, J. M., Isaacs, A., Geula, C., Wang, J., Pan, Y., Gabuzda, D. H., Li, C., Prolla, T. A. & Yankner, B. A. Evolution of the aging brain transcriptome and synaptic regulation. *PLoS One* **3,** e3329 (2008).

190. Spiegel, I., Mardinly, A. R., Gabel, H. W., Bazinet, J. E., Couch, C. H., Tzeng, C. P., Harmin, D. A. & Greenberg, M. E. Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. *Cell* **157,** 1216–1229 (2014).

191. Zhou, Y., Gunput, R.-A. F., Adolfs, Y. & Pasterkamp, R. J. MICALs in control of the cytoskeleton, exocytosis, and cell death. *Cell. Mol. Life Sci.* **68,** 4033–4044 (2011).

192. Sekiguchi, M., Sobue, A., Kushima, I., Wang, C., Arioka, Y., Kato, H., Kodama, A., Kubo, H., Ito, N., Sawahata, M., Hada, K., Ikeda, R., Shinno, M., Mizukoshi, C., Tsujimura, K., Yoshimi, A., Ishizuka, K., Takasaki, Y., Kimura, H., Xing, J., Yu, Y., Yamamoto, M., Okada, T., Shishido, E., Inada, T., Nakatochi, M., Takano, T., Kuroda, K., Amano, M., Aleksic, B., Yamomoto, T., Sakuma, T., Aida, T., Tanaka, K., Hashimoto, R., Arai, M., Ikeda, M., Iwata, N., Shimamura, T., Nagai, T., Nabeshima, T., Kaibuchi, K., Yamada, K., Mori, D. & Ozaki, N. ARHGAP10, which encodes Rho GTPase-activating protein 10, is a novel gene for schizophrenia risk. *Transl. Psychiatry* **10,** 247 (2020).

193. Renick, S. E., Kleven, D. T., Chan, J., Stenius, K., Milner, T. A., Pickel, V. M. & Fremeau, R.

T., Jr. The mammalian brain high-affinity L-proline transporter is enriched preferentially in synaptic vesicles in a subpopulation of excitatory nerve terminals in rat forebrain. *J. Neurosci.* **19,** 21–33 (1999).

194. Sitbon, J., Nestvogel, D., Kappeler, C., Nicolas, A., Maciuba, S., Henrion, A., Troudet, R., Courtois, E., Grannec, G., Latapie, V., Barau, C., Le Corvoisier, P., Pietrancosta, N., Henry, C., Leboyer, M., Etain, B., Nosten-Bertrand, M., Martin, T. F. J., Rhee, J. & Jamain, S. CADPS functional mutations in patients with bipolar disorder increase the sensitivity to stress. *Mol. Psychiatry* **27,** 1145–1157 (2022).

195. Tang, S., Wang, T., Zhang, X., Guo, Y., Xu, P., Zeng, J., Luo, Z., Li, D., Zheng, Y., Luo, Y., Yu, C. & Xu, Z. Olfactomedin-3 Enhances Seizure Activity by Interacting With AMPA Receptors in Epilepsy Models. *Front Cell Dev Biol* **8,** 722 (2020).

196. Salinas, M., Duprat, F., Heurteaux, C., Hugnot, J. P. & Lazdunski, M. New modulatory alpha subunits for mammalian Shab K+ channels. *J. Biol. Chem.* **272,** 24371–24379 (1997).

197. An, W. F., Bowlby, M. R., Betty, M., Cao, J., Ling, H. P., Mendoza, G., Hinson, J. W., Mattsson, K. I., Strassle, B. W., Trimmer, J. S. & Rhodes, K. J. Modulation of A-type potassium channels by a family of calcium sensors. *Nature* **403,** 553–556 (2000).

198. Fu, W., Franchini, L. & Orlandi, C. Comprehensive Spatial Profile of the Orphan G Protein Coupled Receptor GPRC5B Expression in Mouse Brain. *Front. Neurosci.* **16,** 891544 (2022).

199. Reinhard, J. R., Kriz, A., Galic, M., Angliker, N., Rajalu, M., Vogt, K. E. & Ruegg, M. A. The calcium sensor Copine-6 regulates spine structural plasticity and learning and memory. *Nat. Commun.* **7,** 11613 (2016).

200. Li, Y., Zheng, H., Wang, Q., Zhou, C., Wei, L., Liu, X., Zhang, W., Zhang, Y., Du, Z., Wang, X. & Xie, W. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol.* **19,** 18 (2018).

201. Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol. Sex Differ.* **6,** 35 (2015).

202. Benjamin, K. J. M., Arora, R., Feltrin, A. S., Pertea, G., Giles, H. H., Stolz, J. M., D'Ignazio, L., Collado-Torres, L., Shin, J. H., Ulrich, W. S., Hyde, T. M., Kleinman, J. E., Weinberger, D. R., Paquola, A. C. M. & Erwin, J. A. Sex affects transcriptional associations with schizophrenia across the dorsolateral prefrontal cortex, hippocampus, and caudate nucleus. *Nat. Commun.* **15,** 3980 (2024).

203. Gegenhuber, B., Wu, M. V., Bronstein, R. & Tollkuhn, J. Gene regulation by gonadal hormone receptors underlies brain sex differences. *Nature* (2022). doi:10.1038/s41586-022-04686-1

204. Zuloaga, D. G., Zuloaga, K. L., Hinds, L. R., Carbone, D. L. & Handa, R. J. Estrogen receptor β expression in the mouse forebrain: age and sex differences. *J. Comp. Neurol.* **522,** 358–371 (2014).

205. Duncan, C. G., Grimm, S. A., Morgan, D. L., Bushel, P. R., Bennett, B. D., NISC Comparative Sequencing Program, Roberts, J. D., Tyson, F. L., Merrick, B. A. & Wade, P.

A. Dosage compensation and DNA methylation landscape of the X chromosome in mouse liver. *Sci. Rep.* **8,** 10138 (2018).

206. Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A. & Surani, M. A. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339,** 448–452 (2013).

207. Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., Gao, Q., Powell, B. E., Li, Z., Xu, M., Faull, K. F., Lyko, F. & Jaenisch, R. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24,** 310–323 (2013).

208. Hernandez Mora, J. R., Sanchez-Delgado, M., Petazzi, P., Moran, S., Esteller, M., Iglesias-Platas, I. & Monk, D. Profiling of oxBS-450K 5-hydroxymethylcytosine in human placenta and brain reveals enrichment at imprinted loci. *Epigenetics* **13,** 182–191 (2018).

209. Holm, T. M., Jackson-Grusby, L., Brambrink, T., Yamada, Y., Rideout, W. M., 3rd & Jaenisch, R. Global loss of imprinting leads to widespread tumorigenesis in adult mice. *Cancer Cell* **8,** 275–285 (2005).

210. Rainier, S., Johnson, L. A., Dobry, C. J., Ping, A. J., Grundy, P. E. & Feinberg, A. P. Relaxation of imprinted genes in human cancer. *Nature* **362,** 747–749 (1993).

211. Ogawa, O., Eccles, M. R., Szeto, J., McNoe, L. A., Yun, K., Maw, M. A., Smith, P. J. & Reeve, A. E. Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362,** 749–751 (1993).

212. Cevik, S. E., Skaar, D. A., Jima, D. D., Liu, A. J., Østbye, T., Whitson, H. E., Jirtle, R. L., Hoyo, C. & Planchart, A. DNA methylation of imprint control regions associated with Alzheimer's disease in non-Hispanic Blacks and non-Hispanic Whites. *Clin. Epigenetics* **16,** 58 (2024).

213. Mancino, S., Seneviratne, J., Mupo, A., Krueger, F., Oxley, D., Eckersley-Maslin, M. A. & da Rocha, S. T. Stability of Genomic Imprinting and X-Chromosome Inactivation in the Aging Brain. *bioRxiv* 2023.09.29.560184 (2023). doi:10.1101/2023.09.29.560184

214. Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D. & Horsthemke, B. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat. Genet.* **9,** 395–400 (1995).

215. Hadad, N., Masser, D. R., Blanco-Berdugo, L., Stanford, D. R. & Freeman, W. M. Early-life DNA methylation profiles are indicative of age-related transcriptome changes. *Epigenetics Chromatin* **12,** 58 (2019).

216. Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R. & Zoing, M. C. Amyotrophic lateral sclerosis. *The Lancet* **377,** 942–955 Preprint at https://doi.org/10.1016/s0140-6736(10)61156-7 (2011)

217. Al-Chalabi, A., van den Berg, L. H. & Veldink, J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat. Rev. Neurol.* **13,** 96–104 (2017).

218. Ryan, M., Heverin, M., McLaughlin, R. L. & Hardiman, O. Lifetime Risk and Heritability of

Amyotrophic Lateral Sclerosis. *JAMA Neurol.* **76,** 1367–1374 (2019).

219. Rademakers, R., Neumann, M. & Mackenzie, I. R. Advances in understanding the molecular basis of frontotemporal dementia. *Nat. Rev. Neurol.* **8,** 423–434 (2012).

220. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q. & Lee, V. M.-Y. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314,** 130–133 (2006).

221. Rohrer, J. D., Isaacs, A. M., Mizielinska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R. W., Hardy, J., Holton, J., Revesz, T., Rossor, M. N. & Warren, J. D. C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *Lancet Neurol.* **14,** 291–301 (2015).

222. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* **79,** 416–438 (2013).

223. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V.-M., Kaivorinne, A.-L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., ITALSGEN Consortium, Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J. & Traynor, B. J. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).

224. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y. R., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R. & Rademakers, R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).

225. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Front. Neurosci.* **14,** 42 (2020).

226. Smeyers, J., Banchi, E.-G. & Latouche, M. C9ORF72: What It Is, What It Does, and Why It Matters. *Front. Cell. Neurosci.* **15,** 661447 (2021).

227. Ash, P. E. A., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W.-L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., 3rd, Rademakers, R., Boylan, K. B., Dickson, D. W. & Petrucelli, L. Unconventional translation of C9ORF72 GGGGCC

expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77,** 639–646 (2013).

228. Mizielinska, S., Lashley, T., Norona, F. E., Clayton, E. L., Ridler, C. E., Fratta, P. & Isaacs, A. M. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol.* **126,** 845–857 (2013).

229. Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C. & Edbauer, D. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339,** 1335–1338 (2013).

230. Kim, G., Gautier, O., Tassoni-Tsuchida, E., Ma, X. R. & Gitler, A. D. ALS Genetics: Gains, Losses, and Implications for Future Therapies. *Neuron* **108,** 822–842 (2020).

231. Hodge, R. D., Miller, J. A., Novotny, M., Kalmbach, B. E., Ting, J. T., Bakken, T. E., Aevermann, B. D., Barkan, E. R., Berkowitz-Cerasano, M. L., Cobbs, C., Diez-Fuertes, F., Ding, S.-L., McCorrison, J., Schork, N. J., Shehata, S. I., Smith, K. A., Sunkin, S. M., Tran, D. N., Venepally, P., Yanny, A. M., Steemers, F. J., Phillips, J. W., Bernard, A., Koch, C., Lasken, R. S., Scheuermann, R. H. & Lein, E. S. Transcriptomic evidence that von Economo neurons are regionally specialized extratelencephalic-projecting excitatory neurons. *Nat. Commun.* **11,** 1172 (2020).

232. Conradi, S. & Ronnevi, L. O. Selective vulnerability of alpha motor neurons in ALS: relation to autoantibodies toward acetylcholinesterase (AChE) in ALS patients. *Brain Res. Bull.* **30,** 369–371 (1993).

233. Genç, B., Jara, J. H., Lagrimas, A. K. B., Pytel, P., Roos, R. P., Mesulam, M. M., Geula, C., Bigio, E. H. & Özdinler, P. H. Apical dendrite degeneration, a novel cellular pathology for Betz cells in ALS. *Sci. Rep.* **7,** 41765 (2017).

234. Brown, R. H. & Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *New England Journal of Medicine* **377,** 162–172 Preprint at https://doi.org/10.1056/nejmra1603471 (2017)

235. Chiò, A., Mazzini, L. & Mora, G. Disease-modifying therapies in amyotrophic lateral sclerosis. *Neuropharmacology* **167,** 107986 (2020).

236. Maekawa, S., Al-Sarraj, S., Kibble, M., Landau, S., Parnavelas, J., Cotter, D., Everall, I. & Leigh, P. N. Cortical selective vulnerability in motor neuron disease: a morphometric study. *Brain* **127,** 1237–1251 (2004).

237. Van Harten, A. C. M., Phatnani, H. & Przedborski, S. Non-cell-autonomous pathogenic mechanisms in amyotrophic lateral sclerosis. *Trends Neurosci.* **44,** 658–668 (2021).

238. Vahsen, B. F., Gray, E., Thompson, A. G., Ansorge, O., Anthony, D. C., Cowley, S. A., Talbot, K. & Turner, M. R. Non-neuronal cells in amyotrophic lateral sclerosis — from pathogenesis to biomarkers. *Nature Reviews Neurology* **17,** 333–348 Preprint at https://doi.org/10.1038/s41582-021-00487-8 (2021)

239. Rivara, C.-B., Sherwood, C. C., Bouras, C. & Hof, P. R. Stereologic characterization and spatial distribution patterns of Betz cells in the human primary motor cortex. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **270,** 137–151 (2003).

240. Hammer, R. P., Jr, Tomiyasu, U. & Scheibel, A. B. Degeneration of the human Betz cell due to amyotrophic lateral sclerosis. *Exp. Neurol.* **63,** 336–346 (1979).

241. Prudencio, M., Belzil, V. V., Batra, R., Ross, C. A., Gendron, T. F., Pregent, L. J., Murray, M. E., Overstreet, K. K., Piazza-Johnston, A. E., Desaro, P., Bieniek, K. F., DeTure, M., Lee, W. C., Biendarra, S. M., Davis, M. D., Baker, M. C., Perkerson, R. B., van Blitterswijk, M., Stetler, C. T., Rademakers, R., Link, C. D., Dickson, D. W., Boylan, K. B., Li, H. & Petrucelli, L. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat. Neurosci.* **18,** 1175–1182 (2015).

242. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S. & Gottardo, R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16,** 278 (2015).

243. Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., Engelborghs, S., Sieben, A., De Jonghe, P., Vandenberghe, R., Santens, P., De Bleecker, J., Maes, G., Bäumer, V., Dillen, L., Joris, G., Cuijt, I., Corsmit, E., Elinck, E., Van Dongen, J., Vermeulen, S., Van den Broeck, M., Vaerenberg, C., Mattheijssens, M., Peeters, K., Robberecht, W., Cras, P., Martin, J.-J., De Deyn, P. P., Cruts, M. & Van Broeckhoven, C. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol.* **11,** 54–65 (2012).

244. Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A. C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M. & Swarup, V. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53,** 1143–1155 (2021).

245. Khakh, B. S. & Sofroniew, M. V. Diversity of astrocyte functions and phenotypes in neural circuits. *Nat. Neurosci.* **18,** 942–952 (2015).

246. Sosunov, A. A., Wu, X., Tsankova, N. M., Guilfoyle, E., McKhann, G. M., 2nd & Goldman, J. E. Phenotypic heterogeneity and plasticity of isocortical and hippocampal astrocytes in the human brain. *J. Neurosci.* **34,** 2285–2298 (2014).

247. Gaur, N., Perner, C., Witte, O. W. & Grosskreutz, J. The Chitinases as Biomarkers for Amyotrophic Lateral Sclerosis: Signals From the CNS and Beyond. *Front. Neurol.* **11,** 377 (2020).

248. Vu, L., An, J., Kovalik, T., Gendron, T., Petrucelli, L. & Bowser, R. Cross-sectional and longitudinal measures of chitinase proteins in amyotrophic lateral sclerosis and expression of CHI3L1 in activated astrocytes. *J. Neurol. Neurosurg. Psychiatry* **91,** 350–358 (2020).

249. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41,** W77–83 (2013).

250. Schiweck, J., Eickholt, B. J. & Murk, K. Important Shapeshifter: Mechanisms Allowing Astrocytes to Respond to the Changing Nervous System During Development, Injury and Disease. *Front. Cell. Neurosci.* **12,** 261 (2018).

251. Phatnani, H. P., Guarnieri, P., Friedman, B. A., Carrasco, M. A., Muratet, M., O'Keeffe, S.,

Nwakeze, C., Pauli-Behn, F., Newberry, K. M., Meadows, S. K., Tapia, J. C., Myers, R. M. & Maniatis, T. Intricate interplay between astrocytes and motor neurons in ALS. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E756–65 (2013).

252. Endo, F., Komine, O., Fujimori-Tonou, N., Katsuno, M., Jin, S., Watanabe, S., Sobue, G., Dezawa, M., Wyss-Coray, T. & Yamanaka, K. Astrocyte-Derived TGF-β1 Accelerates Disease Progression in ALS Mice by Interfering with the Neuroprotective Functions of Microglia and T Cells. *Cell Reports* **11,** 592–604 Preprint at https://doi.org/10.1016/j.celrep.2015.03.053 (2015)

253. Chen, F., Lin, X., Xu, P., Zhang, Z., Chen, Y., Wang, C., Han, J., Zhao, B., Xiao, M. & Feng, X.-H. Nuclear Export of Smads by RanBP3L Regulates Bone Morphogenetic Protein Signaling and Mesenchymal Stem Cell Differentiation. *Mol. Cell. Biol.* **35,** 1700–1711 (2015).

254. Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., Petrucelli, L., Kim, H. J., Gao, F.-B. & Taylor, J. P. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525,** 129–133 (2015).

255. Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., 3rd, Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Van Den Bosch, L., Robberecht, W. & Gitler, A. D. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* **18,** 1226–1229 (2015).

256. Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., Gupta, S., Thomas, M. A., Hong, I., Chiu, S.-L., Huganir, R. L., Ostrow, L. W., Matunis, M. J., Wang, J., Sattler, R., Lloyd, T. E. & Rothstein, J. D. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525,** 56–61 (2015).

257. Coyne, A. N., Zaepfel, B. L., Hayes, L., Fitchman, B., Salzberg, Y., Luo, E.-C., Bowen, K., Trost, H., Aigner, S., Rigo, F., Yeo, G. W., Harel, A., Svendsen, C. N., Sareen, D. & Rothstein, J. D. G4C2 Repeat RNA Initiates a POM121-Mediated Reduction in Specific Nucleoporins in C9orf72 ALS/FTD. *Neuron* **107,** 1124–1140.e11 (2020).

258. Khakh, B. S. & McCarthy, K. D. Astrocyte Calcium Signaling: From Observations to Functions and the Challenges Therein. *Cold Spring Harbor Perspectives in Biology* **7,** a020404 Preprint at https://doi.org/10.1101/cshperspect.a020404 (2015)

259. Hipp, M. S., Kasturi, P. & Ulrich Hartl, F. The proteostasis network and its decline in ageing. *Nature Reviews Molecular Cell Biology* **20,** 421–435 Preprint at https://doi.org/10.1038/s41580-019-0101-y (2019)

260. Mogk, A., Bukau, B. & Kampinga, H. H. Cellular Handling of Protein Aggregates by Disaggregation Machines. *Molecular Cell* **69,** 214–226 Preprint at https://doi.org/10.1016/j.molcel.2018.01.004 (2018)

261. Dikic, I. Proteasomal and Autophagic Degradation Systems. *Annual Review of Biochemistry* **86,** 193–224 Preprint at https://doi.org/10.1146/annurev-biochem-061516-044908 (2017)

262. Bracher, A. & Verghese, J. The nucleotide exchange factors of Hsp70 molecular

chaperones. *Frontiers in Molecular Biosciences* **2,** Preprint at https://doi.org/10.3389/fmolb.2015.00010 (2015)

263. Dekker, S. L., Kampinga, H. H. & Bergink, S. DNAJs: more than substrate delivery to HSPA. *Frontiers in Molecular Biosciences* **2,** Preprint at https://doi.org/10.3389/fmolb.2015.00035 (2015)

264. Gamerdinger, M., Murat Kaya, A., Wolfrum, U., Clement, A. M. & Behl, C. BAG3 mediates chaperone-based aggresome-targeting and selective autophagy of misfolded proteins. *EMBO reports* **12,** 149–156 Preprint at https://doi.org/10.1038/embor.2010.203 (2011)

265. Rutledge, B. S., Choy, W.-Y. & Duennwald, M. L. Folding or holding?—Hsp70 and Hsp90 chaperoning of misfolded proteins in neurodegenerative disease. *Journal of Biological Chemistry* **298,** 101905 Preprint at https://doi.org/10.1016/j.jbc.2022.101905 (2022)

266. Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M. J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., Fiévet, N., Barberger-Gateau, P., Engelborghs, S., De Deyn, P., Mateo, I., Franck, A., Helisalmi, S., Porcellini, E., Hanon, O., European Alzheimer's Disease Initiative Investigators, de Pancorbo, M. M., Lendon, C., Dufouil, C., Jaillard, C., Leveillard, T., Alvarez, V., Bosco, P., Mancuso, M., Panza, F., Nacmias, B., Bossù, P., Piccardi, P., Annoni, G., Seripa, D., Galimberti, D., Hannequin, D., Licastro, F., Soininen, H., Ritchie, K., Blanché, H., Dartigues, J.-F., Tzourio, C., Gut, I., Van Broeckhoven, C., Alpérovitch, A., Lathrop, M. & Amouyel, P. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **41,** 1094–1099 (2009).

267. Gregory, J. M., Whiten, D. R., Brown, R. A., Barros, T. P., Kumita, J. R., Yerbury, J. J., Satapathy, S., McDade, K., Smith, C., Luheshi, L. M., Dobson, C. M. & Wilson, M. R. Clusterin protects neurons against intracellular proteotoxicity. *Acta Neuropathologica Communications* **5,** Preprint at https://doi.org/10.1186/s40478-017-0481-1 (2017)

268. Nguyen-Ba-Charvet, K. T. Multiple Roles for Slits in the Control of Cell Migration in the Rostral Migratory Stream. *Journal of Neuroscience* **24,** 1497–1506 Preprint at https://doi.org/10.1523/jneurosci.4729-03.2004 (2004)

269. Mariga, A., Mitre, M. & Chao, M. V. Consequences of brain-derived neurotrophic factor withdrawal in CNS neurons and implications in disease. *Neurobiology of Disease* **97,** 73–79 Preprint at https://doi.org/10.1016/j.nbd.2016.03.009 (2017)

270. Spilker, C. & Kreutz, M. R. RapGAPs in brain: multipurpose players in neuronal Rap signalling. *European Journal of Neuroscience* **32,** 1–9 Preprint at https://doi.org/10.1111/j.1460-9568.2010.07273.x (2010)

271. Limoni, G. Modelling and Refining Neuronal Circuits with Guidance Cues: Involvement of Semaphorins. *International Journal of Molecular Sciences* **22,** 6111 Preprint at https://doi.org/10.3390/ijms22116111 (2021)

272. Simonetti, M., Paldy, E., Njoo, C., Bali, K. K., Worzfeld, T., Pitzer, C., Kuner, T., Offermanns, S., Mauceri, D. & Kuner, R. The impact of Semaphorin 4C/Plexin-B2 signaling on fear memory via remodeling of neuronal and synaptic morphology. *Molecular Psychiatry* **26,** 1376–1398 Preprint at https://doi.org/10.1038/s41380-019-0491-4 (2021)

273. Zee, Y. Y. van der, van der Zee, Y. Y., Lardner, C. K., Parise, E. M., Mews, P., Ramakrishnan, A., Patel, V., Teague, C. D., Salery, M., Walker, D. M., Browne, C. J., Labonté, B., Parise, L. F., Kronman, H., Penã, C. J., Torres-Berrío, A., Duffy, J. E., de Nijs, L., Eijssen, L. M. T., Shen, L., Rutten, B., Issler, O. & Nestler, E. J. Sex-Specific Role for SLIT1 in Regulating Stress Susceptibility. *Biological Psychiatry* **91,** 81–91 Preprint at https://doi.org/10.1016/j.biopsych.2021.01.019 (2022)

274. Lowry, E. R., Kruyer, A., Norris, E. H., Cederroth, C. R. & Strickland, S. The GluK4 kainate receptor subunit regulates memory, mood, and excitotoxic neurodegeneration. *Neuroscience* **235,** 215–225 Preprint at https://doi.org/10.1016/j.neuroscience.2013.01.029 (2013)

275. Chestnut, B. A., Chang, Q., Price, A., Lesuisse, C., Wong, M. & Martin, L. J. Epigenetic Regulation of Motor Neuron Cell Death through DNA Methylation. *Journal of Neuroscience* **31,** 16619–16636 Preprint at https://doi.org/10.1523/jneurosci.1639-11.2011 (2011)

276. Štalekar, M., Yin, X., Rebolj, K., Darovic, S., Troakes, C., Mayr, M., Shaw, C. E. & Rogelj, B. Proteomic analyses reveal that loss of TDP-43 affects RNA processing and intracellular transport. *Neuroscience* **293,** 157–170 Preprint at https://doi.org/10.1016/j.neuroscience.2015.02.046 (2015)

277. Mihevc, S. P., Prpar Mihevc, S., Baralle, M., Buratti, E. & Rogelj, B. TDP-43 aggregation mirrors TDP-43 knockdown, affecting the expression levels of a common set of proteins. *Scientific Reports* **6,** Preprint at https://doi.org/10.1038/srep33996 (2016)

278. Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348,** 910–914 (2015).

279. Li, Y. E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O. B., Li, B., Chiou, J., Liu, H., Pinto-Duarte, A., Kubo, N., Yang, X., Fang, R., Wang, X., Han, J. Y., Lucero, J., Yan, Y., Miller, M., Kuan, S., Gorkin, D., Gaulton, K. J., Shen, Y., Nunn, M., Mukamel, E. A., Behrens, M. M., Ecker, J. R. & Ren, B. An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598,** 129–136 (2021).

280. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14,** 975–978 (2017).

281. Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. & Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–21936 (2010).

282. van Rheenen, W., van der Spek, R. A. A., Bakker, M. K., van Vugt, J. J. F. A., Hop, P. J., Zwamborn, R. A. J., de Klein, N., Westra, H.-J., Bakker, O. B., Deelen, P., Shireby, G., Hannon, E., Moisse, M., Baird, D., Restuadi, R., Dolzhenko, E., Dekker, A. M., Gawor, K., Westeneng, H.-J., Tazelaar, G. H. P., van Eijk, K. R., Kooyman, M., Byrne, R. P., Doherty, M., Heverin, M., Al Khleifat, A., Iacoangeli, A., Shatunov, A., Ticozzi, N., Cooper-Knock, J., Smith, B. N., Gromicho, M., Chandran, S., Pal, S., Morrison, K. E., Shaw, P. J., Hardy, J., Orrell, R. W., Sendtner, M., Meyer, T., Başak, N., van der Kooi, A. J., Ratti, A., Fogh, I., Gellera, C., Lauria, G., Corti, S., Cereda, C., Sproviero, D., D'Alfonso, S., Sorarù, G.,

Siciliano, G., Filosto, M., Padovani, A., Chiò, A., Calvo, A., Moglia, C., Brunetti, M., Canosa, A., Grassano, M., Beghi, E., Pupillo, E., Logroscino, G., Nefussy, B., Osmanovic, A., Nordin, A., Lerner, Y., Zabari, M., Gotkine, M., Baloh, R. H., Bell, S., Vourc'h, P., Corcia, P., Couratier, P., Millecamps, S., Meininger, V., Salachas, F., Mora Pardina, J. S., Assialioui, A., Rojas-García, R., Dion, P. A., Ross, J. P., Ludolph, A. C., Weishaupt, J. H., Brenner, D., Freischmidt, A., Bensimon, G., Brice, A., Durr, A., Payan, C. A. M., Saker-Delye, S., Wood, N. W., Topp, S., Rademakers, R., Tittmann, L., Lieb, W., Franke, A., Ripke, S., Braun, A., Kraft, J., Whiteman, D. C., Olsen, C. M., Uitterlinden, A. G., Hofman, A., Rietschel, M., Cichon, S., Nöthen, M. M., Amouyel, P., SLALOM Consortium, PARALS Consortium, SLAGEN Consortium, SLAP Consortium, Traynor, B. J., Singleton, A. B., Mitne Neto, M., Cauchi, R. J., Ophoff, R. A., Wiedau-Pazos, M., Lomen-Hoerth, C., van Deerlin, V. M., Grosskreutz, J., Roediger, A., Gaur, N., Jörk, A., Barthel, T., Theele, E., Ilse, B., Stubendorff, B., Witte, O. W., Steinbach, R., Hübner, C. A., Graff, C., Brylev, L., Fominykh, V., Demeshonok, V., Ataulina, A., Rogelj, B., Koritnik, B., Zidar, J., Ravnik-Glavač, M., Glavač, D., Stević, Z., Drory, V., Povedano, M., Blair, I. P., Kiernan, M. C., Benyamin, B., Henderson, R. D., Furlong, S., Mathers, S., McCombe, P. A., Needham, M., Ngo, S. T., Nicholson, G. A., Pamphlett, R., Rowe, D. B., Steyn, F. J., Williams, K. L., Mather, K. A., Sachdev, P. S., Henders, A. K., Wallace, L., de Carvalho, M., Pinto, S., Petri, S., Weber, M., Rouleau, G. A., Silani, V., Curtis, C. J., Breen, G., Glass, J. D., Brown, R. H., Jr, Landers, J. E., Shaw, C. E., Andersen, P. M., Groen, E. J. N., van Es, M. A., Pasterkamp, R. J., Fan, D., Garton, F. C., McRae, A. F., Davey Smith, G., Gaunt, T. R., Eberle, M. A., Mill, J., McLaughlin, R. L., Hardiman, O., Kenna, K. P., Wray, N. R., Tsai, E., Runz, H., Franke, L., Al-Chalabi, A., Van Damme, P., van den Berg, L. H. & Veldink, J. H. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* **53,** 1636–1648 (2021).

283. Fels, E., Muñiz-Castrillo, S., Vogrig, A., Joubert, B., Honnorat, J. & Pascual, O. Role of LGI1 protein in synaptic transmission: From physiology to pathology. *Neurobiol. Dis.* **160,** 105537 (2021).

284. Walz, W. Role of astrocytes in the clearance of excess extracellular potassium. *Neurochem. Int.* **36,** 291–300 (2000).

285. Pook, C., Ahrens, J. M. & Clagett-Dame, M. Expression pattern of Nav2 in the murine CNS with development. *Gene Expr. Patterns* **35,** 119099 (2020).

286. Macedo, A., Gómez, C., Rebelo, M. Â., Poza, J., Gomes, I., Martins, S., Maturana-Candelas, A., Pablo, V. G., Durães, L., Sousa, P., Figueruelo, M., Rodríguez, M., Pita, C., Arenas, M., Álvarez, L., Hornero, R., Lopes, A. M. & Pinto, N. Risk Variants in Three Alzheimer's Disease Genes Show Association with EEG Endophenotypes. *J. Alzheimers. Dis.* **80,** 209–223 (2021).

287. Kang, S. H., Li, Y., Fukaya, M., Lorenzini, I., Cleveland, D. W., Ostrow, L. W., Rothstein, J. D. & Bergles, D. E. Degeneration and impaired regeneration of gray matter oligodendrocytes in amyotrophic lateral sclerosis. *Nature Neuroscience* **16,** 571–579 Preprint at https://doi.org/10.1038/nn.3357 (2013)

288. Zhao, C., Devlin, A., Chouhan, A. K., Selvaraj, B. T., Stavrou, M., Burr, K., Brivio, V., He, X., Mehta, A. R., Story, D., Shaw, C. E., Dando, O., Hardingham, G. E., Miles, G. B. & Chandran, S. Mutant *C9orf72* human iPSC-derived astrocytes cause non-cell autonomous motor neuron pathophysiology. *Glia* **68,** 1046–1064 Preprint at https://doi.org/10.1002/glia.23761 (2020)

289. Alami, N. O., Schurr, C., Heuvel, F. O., Tang, L., Li, Q., Tasdogan, A., Kimbara, A., Nettekoven, M., Ottaviani, G., Raposo, C., Röver, S., Rogers-Evans, M., Rothenhäusler, B., Ullmer, C., Fingerle, J., Grether, U., Knuesel, I., Boeckers, T. M., Ludolph, A., Wirth, T., Roselli, F. & Baumann, B. NF-κB activation in astrocytes drives a stage-specific beneficial neuroimmunological response in ALS. *The EMBO Journal* **37,** Preprint at https://doi.org/10.15252/embj.201798697 (2018)

290. Lall, D., Lorenzini, I., Mota, T. A., Bell, S., Mahan, T. E., Ulrich, J. D., Davtyan, H., Rexach, J. E., Muhammad, A. K. M. G., Shelest, O., Landeros, J., Vazquez, M., Kim, J., Ghaffari, L., O'Rourke, J. G., Geschwind, D. H., Blurton-Jones, M., Holtzman, D. M., Sattler, R. & Baloh, R. H. C9orf72 deficiency promotes microglial-mediated synaptic loss in aging and amyloid accumulation. *Neuron* **109,** 2275–2291.e8 (2021).

291. Bakken, T. E., Jorstad, N. L., Hu, Q., Lake, B. B., Tian, W., Kalmbach, B. E., Crow, M., Hodge, R. D., Krienen, F. M., Sorensen, S. A., Eggermont, J., Yao, Z., Aevermann, B. D., Aldridge, A. I., Bartlett, A., Bertagnolli, D., Casper, T., Castanon, R. G., Crichton, K., Daigle, T. L., Dalley, R., Dee, N., Dembrow, N., Diep, D., Ding, S.-L., Dong, W., Fang, R., Fischer, S., Goldman, M., Goldy, J., Graybuck, L. T., Herb, B. R., Hou, X., Kancherla, J., Kroll, M., Lathia, K., van Lew, B., Li, Y. E., Liu, C. S., Liu, H., Lucero, J. D., Mahurkar, A., McMillen, D., Miller, J. A., Moussa, M., Nery, J. R., Nicovich, P. R., Niu, S.-Y., Orvis, J., Osteen, J. K., Owen, S., Palmer, C. R., Pham, T., Plongthongkum, N., Poirion, O., Reed, N. M., Rimorin, C., Rivkin, A., Romanow, W. J., Sedeño-Cortés, A. E., Siletti, K., Somasundaram, S., Sulc, J., Tieu, M., Torkelson, A., Tung, H., Wang, X., Xie, F., Yanny, A. M., Zhang, R., Ament, S. A., Behrens, M. M., Bravo, H. C., Chun, J., Dobin, A., Gillis, J., Hertzano, R., Hof, P. R., Höllt, T., Horwitz, G. D., Keene, C. D., Kharchenko, P. V., Ko, A. L., Lelieveldt, B. P., Luo, C., Mukamel, E. A., Pinto-Duarte, A., Preiss, S., Regev, A., Ren, B., Scheuermann, R. H., Smith, K., Spain, W. J., White, O. R., Koch, C., Hawrylycz, M., Tasic, B., Macosko, E. Z., McCarroll, S. A., Ting, J. T., Zeng, H., Zhang, K., Feng, G., Ecker, J. R., Linnarsson, S. & Lein, E. S. Author Correction: Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **604,** E8 (2022).

292. Nolan, M., Scott, C., Gamarallage, M. P., Lunn, D., Carpenter, K., McDonough, E., Meyer, D., Kaanumalle, S., Santamaria-Pang, A., Turner, M. R., Talbot, K. & Ansorge, O. Quantitative patterns of motor cortex proteinopathy across ALS genotypes. *Acta Neuropathol Commun* **8,** 98 (2020).

293. Fang, R., Xia, C., Close, J. L., Zhang, M., He, J., Huang, Z., Halpern, A. R., Long, B., Miller, J. A., Lein, E. S. & Zhuang, X. Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377,** 56–62 (2022).

294. Byrne, S., Elamin, M., Bede, P., Shatunov, A., Walsh, C., Corr, B., Heverin, M., Jordan, N., Kenna, K., Lynch, C., McLaughlin, R. L., Iyer, P. M., O'Brien, C., Phukan, J., Wynne, B., Bokde, A. L., Bradley, D. G., Pender, N., Al-Chalabi, A. & Hardiman, O. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. *Lancet Neurol.* **11,** 232–240 (2012).

295. Candelise, N., Salvatori, I., Scaricamazza, S., Nesci, V., Zenuni, H., Ferri, A. & Valle, C. Mechanistic Insights of Mitochondrial Dysfunction in Amyotrophic Lateral Sclerosis: An Update on a Lasting Relationship. *Metabolites* **12,** 233 Preprint at https://doi.org/10.3390/metabo12030233 (2022)

296. Petrozziello, T., Bordt, E. A., Mills, A. N., Kim, S. E., Sapp, E., Devlin, B. A., Obeng-Marnu,

A. A., Farhan, S. M. K., Amaral, A. C., Dujardin, S., Dooley, P. M., Henstridge, C., Oakley, D. H., Neueder, A., Hyman, B. T., Spires-Jones, T. L., Bilbo, S. D., Vakili, K., Cudkowicz, M. E., Berry, J. D., DiFiglia, M., Silva, M. C., Haggarty, S. J. & Sadri-Vakili, G. Targeting Tau Mitigates Mitochondrial Fragmentation and Oxidative Stress in Amyotrophic Lateral Sclerosis. *Mol. Neurobiol.* **59,** 683–702 (2022).

297. Smith, E. F., Shaw, P. J. & De Vos, K. J. The role of mitochondria in amyotrophic lateral sclerosis. *Neuroscience Letters* **710,** 132933 Preprint at https://doi.org/10.1016/j.neulet.2017.06.052 (2019)

298. Kepp, K. P. A quantitative model of human neurodegenerative diseases involving protein aggregation. *Neurobiol. Aging* **80,** 46–55 (2019).

299. Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Andrusivová, Ž., Saarenpää, S., Saiz-Castro, G., Cuevas, M., Watters, A., Lundeberg, J., Bonneau, R. & Phatnani, H. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* **364,** 89–93 (2019).

300. Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D. H. & Kriegstein, A. R. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364,** 685–689 (2019).

301. Jiang, Y., Matevossian, A., Huang, H.-S., Straubhaar, J. & Akbarian, S. Isolation of neuronal chromatin from brain tissue. *BMC Neurosci.* **9,** 42 (2008).

302. Ernst, A., Alkass, K., Bernard, S., Salehpour, M., Perl, S., Tisdale, J., Possnert, G., Druid, H. & Frisén, J. Neurogenesis in the striatum of the adult human brain. *Cell* **156,** 1072–1083 (2014).

303. Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keeffe, S., Phatnani, H. P., Guarnieri, P., Caneda, C., Ruderisch, N., Deng, S., Liddelow, S. A., Zhang, C., Daneman, R., Maniatis, T., Barres, B. A. & Wu, J. Q. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34,** 11929–11947 (2014).

304. Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M. M. & Lorincz, M. C. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6,** 6033 (2015).

305. Jara, J. H., Gautam, M., Kocak, N., Xie, E. F., Mao, Q., Bigio, E. H. & Özdinler, P. H. MCP1-CCR2 and neuroinflammation in the ALS motor cortex with TDP-43 pathology. *J. Neuroinflammation* **16,** 196 (2019).

306. Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J. & Bielas, J. H. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).

307. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep

generative model for unsupervised removal of background noise from scRNA-seq datasets. *bioRxiv* 791699 (2019). doi:10.1101/791699

308. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., Rogers, A. J., McElrath, J. M., Blish, C. A., Gottardo, R., Smibert, P. & Satija, R. Integrated analysis of multimodal single-cell data. *Cell* **184,** 3573–3587.e29 (2021).

309. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20,** 296 (2019).

310. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* (2019). doi:10.1038/s41592-019-0619-0

311. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018). at <http://arxiv.org/abs/1802.03426>

312. Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R. & Hemberg, M. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14,** 483–486 (2017).

313. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7,** (2018).

314. Miller, J. A., Gouwens, N. W., Tasic, B., Collman, F., van Velthoven, C. T., Bakken, T. E., Hawrylycz, M. J., Zeng, H., Lein, E. S. & Bernard, A. Common cell type nomenclature for the mammalian brain. *Elife* **9,** (2020).

315. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

316. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

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

318. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

319. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).

320. Satpathy, A. T., Granja, J. M., Yost, K. E., Qi, Y., Meschi, F., McDermott, G. P., Olsen, B. N., Mumbach, M. R., Pierce, S. E., Corces, M. R., Shah, P., Bell, J. C., Jhutty, D., Nemec, C. M., Wang, J., Wang, L., Yin, Y., Giresi, P. G., Chang, A. L. S., Zheng, G. X. Y., Greenleaf, W. J. & Chang, H. Y. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37,** 925–936 (2019).

321. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* **11,** 866 (2020).

322. Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y. & Greenleaf, W. J. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53,** 403–411 (2021).

323. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S. & Pe'er, D. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **0,** (2018).

324. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. & Liu, X. S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

325. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., Rynes, E., Reynolds, A., Nelson, J., Johnson, A., Frerker, M., Buckley, M., Kaul, R., Meuleman, W. & Stamatoyannopoulos, J. A. Global reference mapping of human transcription factor footprints. *Nature* **583,** 729–736 (2020).

326. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43,** W39–49 (2015).

327. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019).

328. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

329. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

330. Kozlenkov, A., Vermunt, M. W., Apontes, P., Li, J., Hao, K., Sherwood, C. C., Hof, P. R., Ely, J. J., Wegner, M., Mukamel, E. A., Creyghton, M. P., Koonin, E. V. & Dracheva, S. Evolution of regulatory signatures in primate cortical neurons at cell-type resolution. *Proc. Natl. Acad. Sci. U. S. A.* **117,** 28422–28432 (2020).

331. Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. & Manke, T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44,** W160–5 (2016).

332. Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C. & Carroll, J. S. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481,** 389–393 (2012).