GENETIC VARIANTS REGULATING MICRORNA EXPRESSION IN THE DEVELOPING HUMAN NEOCORTEX

Michael J. Lafferty

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Bioinformatics and Computational Biology in the School of Medicine.

Chapel Hill
2022

Approved by:

Jason L. Stein

Mohanish Deshmukh

Terry S. Furey

Yun Li

Michael I. Love

ABSTRACT

Michael J. Lafferty: Genetic Variants Regulating MicroRNA Expression in the Developing Human Neocortex
(Under the direction of Jason L. Stein)

Genome-wide association studies (GWAS) have revealed several genomic loci associated with risk for neuropsychiatric disorders and brain-related traits, but for many of these loci, the causal mechanisms are unknown. Using expression quantitative trait loci (eQTL) data, many of the risk loci can be linked to protein-coding genes or long non-coding RNAs. However, the mechanism underlying a significant proportion of loci remain unexplained. MicroRNAs (miRNAs) are poorly measured in standard eQTL studies yet have important influences on neurogenesis and have been found to be differentially expressed in brain tissue from patients with neuropsychiatric disorders. By linking miRNA-eQTLs to previously defined mRNA-eQTLs and genome-wide significant loci for brain traits, I aimed to reveal causal mechanisms by which common genetic variation influences gene regulation and inter-individual differences in brain-related traits. Additionally, I sought to find evidence for novel early brain-expressed miRNAs and define an expanded set of miRNAs associated with human neocortical neurogenesis and neuron maturity. Here, in 212 mid-gestation neocortical tissue samples, I profiled expression of 907 miRNAs, discovering 111 novel miRNAs and identified 85 local-miRNA-eQTLs. Colocalization of miRNA-eQTLs with GWAS summary statistics yielded one robust colocalization of miR-4707-3p expression with educational attainment and head size phenotypes, where the miRNA expression increasing allele was associated with decreased head size. After increasing expression of miR-4707-3p in primary human neural progenitor cultures and their differentiated progeny, I detected an increase in both proliferative and neurogenic gene markers by qPCR and ICC assays, implying an early cell-cycle exit and an increase in neurogenic divisions. Using 17 neocortical tissue samples micro-

dissected into progenitor rich germinal zone and neuron rich cortical plate, I identified 504 miRNAs associated with neurogenesis or neuron maturity. Finally, I was able to prioritize these neurogenesis associated miRNAs by looking at the differential expression of their predicted mRNA targets. In summary, this work shows how miRNA-eQTLs and neurogenesis associated miRNAs in developing neocortical tissue yielded insight into the causal mechanisms by which genetic variants influence brain traits.

ACKNOWLEDGEMENTS

I would first like to thank my advisor Dr. Jason L. Stein for encouraging me to become a more thoughtful and rigorous scientist. Thank you for allowing me to explore new ideas and techniques, and for giving me motivation and reinforcement when I needed it most. I'd also like to thank all the members of the Stein Lab at UNC for helpful conversations and needed distractions during this arduous journey: Oleh Krupa, Dan Liang, Rose Glass, Bran Le, Nil Aygün, Felix Kyere, Jordan Valone, Ian Curtin, Justin Wolter, Nana Matoba, Jessica McAfee, Alvaro Beltran, Carolyn McCormick, Niya Patel, and Angela Elwell.

Thank you to my committee members and the faculty members that contributed ideas and guidance throughout my training: Dr. Michael Love, Dr. Mohanish Deshmukh, Dr. Terry Furey, Dr. Yun Li, Dr. Scott Hammond, Dr. Kyle Roell, Dr. Luis de la Torre-Ubieta, and Dr. Karen Mohlke. Also, a heartfelt thank you to John Cornett without whom I would not have completed my degree. I would also like to thank Dr. Saskia Neher for allowing me to learn and grow as a young scientist in her lab. Her teachings and encouragement to explore continue to be the foundation upon which my scientific career is based.

Finally, thank you to my family for the love and support you've given me over the years. From my dad, John, who fostered my scientific curiosity. From my mom, Aline, who taught me to face my fears. And from my brother, Brian, who showed me perseverance is the key to success. And to my dog, Finley, who was by my side through the fun times and the stressful times making sure I was never lonely along the way.

CONTRIBUTIONS

In Chapter 2 and 3, donor tissue was obtained by the University of California Los Angeles (UCLA) Gene and Cell Therapy Core following IRB regulations under the direction of Daniel H. Geschwind. RNA-isolation, small-RNA sequencing, and total-RNA sequencing was conducted at UCLA in the lab of Daniel H. Geschwind. Luis de la Torre-Ubieta and Jason L. Stein advised in design and interpretation of bioinformatic analyses.

In Chapter 2, Genotyping of donor tissues was done at both UCLA in the lab of Daniel H. Geschwind and at the University of North Carolina at Chapel Hill in the lab of Jason L. Stein. Dan Liang was primarily responsible for the genotype data quality control. Dan Liang and Nil Aygün assisted in designing a robust mixed-model expression association analysis. Nil Aygün defined the splicing- and expression-QTLs for mRNA in fetal cortical tissue. Oleh Krupa, Justin Wolter, and Luis de la Torre-Ubieta advised in design and implementation of the miR-4707-3p over-expression experiments in primary human neural progenitor cells.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGO	Argonaute
CP	Cortical plate
DOX	Doxycycline
emiR	MicroRNA with a local eQTL
eQTL	Expression quantitative trait loci
FDR	False discover rate
GFP	Green fluorescent protein
GO	Gene ontology
GSA	Global cortical surface area
GW	Gestation week
GWAS	Genome-wide association study
GZ	Germinal zone
ICC	Immunocytochemistry
INT	Inverse normal transformation
LD	Linkage disequilibrium
lncRNA	Long non-coding RNA
miRNA	MicroRNA
mRNA	Messenger RNA
PC	Principal component
PCA	Principal component analysis
phNPC	Primary human neural progenitor cell
PSI	Percent spliced in
qPCR	Quantitative polymerase chain reaction
RISC	RNA-induced silencing complex

rRNA	Ribosomal RNA
sQTL	Splicing quantitative trait loci
TRE	Tetracycline response element
UTR	Untranslated region
VST	Variance stabilizing transformation

CHAPTER 1: INTRODUCTION

1.1 MicroRNAs Influence Cell Fate via mRNA Regulation

MicroRNAs (miRNAs) are small, 21-23 nucleotide, non-coding RNAs which regulate mRNA translation [1,2]. In the canonical miRNA biogenesis pathway, the primary miRNA transcript, pri-miR, originates from a genomic locus similar to protein-coding genes (Figure 1.1). The pri-miR is then processed within the nucleus by the proteins Drosha and DGCR8 into a stem-loop precursor called a miRNA hairpin. This hairpin sequence, consisting of a 5' sequence, 3' sequence, and "loop," is exported into the cytoplasm for further processing by the proteins Dicer and Argonaute (AGO). A mature miRNA, either the 5' or 3' strand (termed 5p and 3p), is loaded into AGO which gives the RNA-induced silencing complex (RISC) target specificity. Through homologous binding of the miRNA "seed" sequence (nucleotides 2-7 of the mature miRNA) with cytoplasmic mRNA, the RISC can repress translation and promote mRNA degradation. With canonical miRNA biogenesis and mRNA targeting, an increase in miRNA expression is often associated with a decrease in targeted mRNA expression [3,4].

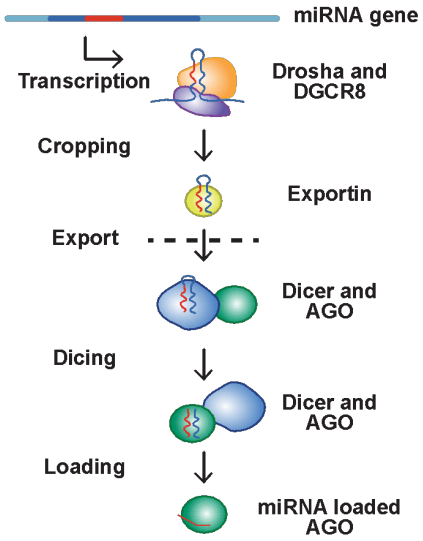


Figure 1.1. Canonical MicroRNA Biogenesis. Genomic loci containing a miRNA gene are transcribed into primary-miRNAs. Processing events in the nucleus result in export of a hairpin-miRNA. Mature miRNAs are further processed in the cytoplasm and loaded into AGO enabling the RISC target specificity.

Through post-transcriptional regulation of targeted mRNAs, miRNAs can have broad effects on gene regulation, cell fate decisions, and transcriptional programs [5–7]. Furthermore, given the crucial role miRNAs play on gene regulation, the miss-regulation of miRNAs is often seen with several disorders and diseases [8–12].

1.2 The Role of Neurogenesis during Human Neocortical Development

The human neocortex is the largest, and evolutionarily newest, portion of the human brain, and it is primarily responsible for our higher order cognitive functions [13]. From sensory perception, motor commands, spatial reasoning, and language our neocortex sets humans apart from all other mammalian species. Given the remarkable function of the neocortex, this portion of the brain is thought to play a major role in many behavioral traits and neuropsychiatric disorders [14,15]. Furthermore, previous studies have suggested that risk for neuropsychiatric disorder may begin during prenatal cerebral cortical

development [16–19]. Therefore, to understand the pathogenesis of these disorders, we must investigate neocortical development during mid gestation.

The primary cell type responsible for neocortical function is the neuron, and to understand the development of the neocortex, one must look at neurogenesis during fetal brain development [20,21]. Neurogenesis in the neocortex starts around the third gestation week with a pool of proliferating neuroepithelial cells that give rise to specialized neural progenitor cell types of the ventricular and subventricular zones (Figure 1.2) [22]. These progenitor cells, called apical and basal radial glial cells, divide symmetrically to expand the progenitor pool or asymmetrically to produce a newly born, immature neuron. Immature neurons then migrate along radial glial processes into the upper cortical laminae to form the cortical plate. By the 29th gestation week approximately 13 billion excitatory glutamatergic neurons are generated throughout the neocortex [23]. The delicate balance between symmetric progenitor proliferation and asymmetric neuronal differentiation is the basis of the radial unit hypothesis [24,25]. This hypothesis states that an increase in the progenitor pool would subsequently lead to a larger cortical surface area. While an increase in neurogenic differentiation could decrease the progenitor pool and lead to a decreased cortical surface area. Inter-individual differences in cortical surface area have been correlated to several brain traits and disorders making the study of neurogenesis and neocortical development crucial to understand the pathology of neuropsychiatric disorders and cognitive traits [26–28].

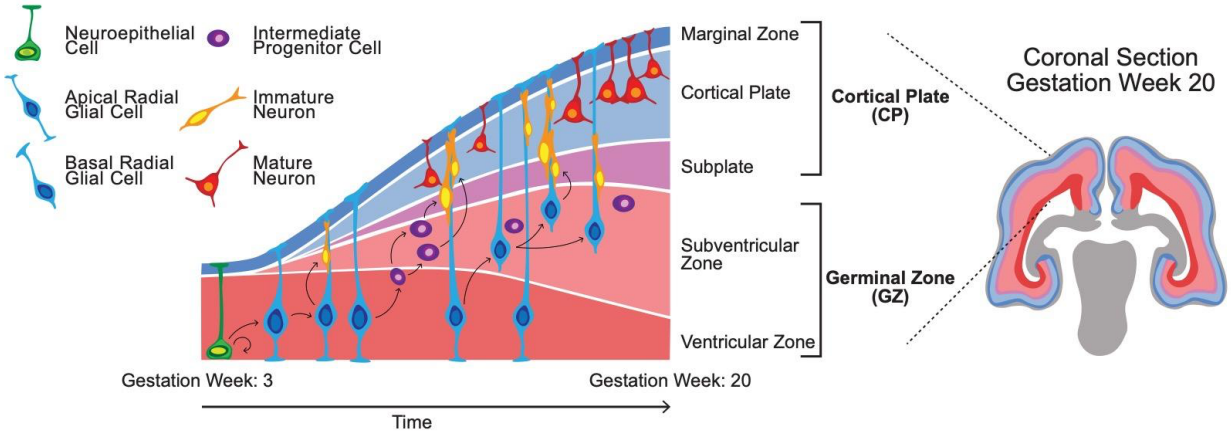


Figure 1.2. Neocortical Development. At early gestation weeks, neural progenitor cells expand and form the ventricular zone and subventricular zone laminae, collectively termed the germinal zone (GZ). Early born neurons migrate up radial glia processes into the subplate, cortical plate, and marginal zone, collectively termed the cortical plate (CP). Neurons remain in the cortical plate where they mature and eventually form the adult neocortex.

MicroRNAs have known roles in neurogenesis and neocortical development. Loss of functional Dicer in mouse radial glial cells during cortical neurogenesis shows an increase in the number of cortical neurons and structural changes in the resultant cortical plate [29,30]. MicroRNA-92b has been shown to be upregulated in cortical intermediate progenitor cells and necessary to maintain proliferative progenitors [31]. Increased miR-124 expression in neural progenitor cells leads to an increase in the number of differentiated neurons in cell-culture models [32]. Finally, miRNA expression and their targeted mRNAs undergo dynamic transitions which appear to control radial glia proliferation rates during human brain development [33]. Continued study of miRNA expression in human developing cortical tissue will yield further insight into the important role that miRNAs play during neurogenesis and neocortical development.

1.3 Effects of Genetic Variation on Brain Traits and Gene Expression

Despite high heritability and hundreds of genetic risk loci identified through genome-wide association studies (GWAS), relatively little is known about the molecular and cellular pathways in the brain which lead to behavioral traits and disorder phenotypes [34,35]. Linking genetic variation to causal

mechanisms is hindered by many factors, including high linkage disequilibrium at associated loci and the fact that many variants fall in non-coding regions of the genome [36]. Many of these variants are enriched within known regulatory regions, and therefore they likely function through gene expression regulation [27,37,38]. Thus, one way to overcome these problems is to find gene expression quantitative trait loci (eQTLs) that colocalize with GWAS variants thereby linking genomic variation to cellular and developmental pathways that may explain disorder phenotypes. Since eQTL variants are often dependent on cell type and developmental stages, we must characterize variants within tissues and timepoints relevant for disorder pathology [39,40].

Mapping eQTLs for protein-coding genes in human fetal cortical tissue during mid-gestation has revealed developmentally specific colocalizations with brain structure and cognitive traits [41–43]. However, there may be other mechanisms by which GWAS loci influence complex traits, and many loci remain unlinked to protein-coding genes [44–46]. MicroRNAs are poorly measured in standard eQTL studies because small RNAs are often lost during standard sequencing library preparation methods. However, using a specialized library preparation method to enrich for small RNA, miRNA expression can be studied in large sample sizes necessary for a miRNA-eQTL analysis. To date, relatively few miRNA-eQTL studies have been published [47–49]. The largest, in blood tissue, is not ideal to directly implicate brain-related phenotypes [50]. Given the important role miRNAs play in cell-fate decisions, there is a growing need to understand the genetic basis by which miRNAs are regulated during neurogenesis and cortical development [33,51].

CHAPTER 2: MICRORNA-EQTLS LINK MIR-4707-3P EXPRESSION TO BRAIN SIZE

2.1 Introduction

Genome-wide association studies (GWAS) have identified many genetic loci influencing human behavior, cognition, and brain structure [28,52–55]. Expression quantitative trait loci (eQTL) data is often used to link non-coding brain-trait associated loci with genes that putatively mediate their effects [41,43,56]. Brain eQTL studies are most often conducted in bulk adult post-mortem tissue and are focused on measuring mRNA expression levels from protein-coding genes. Though these methods have been successful in linking a subset of non-coding brain-trait associated loci to genes, there may be multiple mechanisms by which a single locus influences a complex trait and many loci are still unlinked to genes [42,44–46,57,58]. This suggests other types of RNAs, unmeasured in previous eQTL studies, may be mediating the genetic associations.

MicroRNAs (miRNAs) are poorly measured in standard eQTL studies because the library preparation methods effectively remove small RNAs. Library preparation methods have been developed that specifically measure small RNAs and allow measurement of miRNA expression in large sample sizes necessary for miRNA-eQTL studies. To date, relatively few miRNA-eQTL studies have been published, and those that have, are often underpowered or in tissues not directly implicated in brain-related phenotypes [47–50,59]. Given evidence that miRNAs are strongly involved in fate decisions during neurogenesis and brain development, there is an increasing need to understand the genetic basis by which miRNAs are regulated [30,33,51,60].

Previous studies have found enrichment of brain structure and cognition GWAS heritability within regulatory elements active during mid-fetal development [27,28,61]. Mapping mRNA-eQTLs in human mid-gestation cortical tissue or neural progenitor cells derived from that tissue has revealed novel

developmentally-specific colocalizations with brain structure and cognitive traits [42,43]. These findings are consistent with the radial unit hypothesis, which posits that increases in size of the neural progenitor pool, present only in mid-gestation, leads to increases in the size of the cortex [25,62]. The discovery of miRNA-eQTLs during prenatal cortical development may highlight additional molecular mechanisms by which non-coding loci influence brain-related traits.

In this study, we performed a local-miRNA-eQTL analysis in 212 mid-gestation human cortical tissue donors to discover the common genetic variation associated with expression of nearby miRNAs (Figure 2.1A). We identified 85 local-miRNA-eQTLs (variant - miRNA pairs) associated with expression of 70 miRNAs. These miRNAs were often found within host mRNAs (49 of 70 miRNAs), but the genetic signal associated with miRNA expression was seldom colocalized with a signal associated with mRNA expression (observed only in 3 of 49 loci). One robust colocalization was detected between a miRNA-eQTL for miR-4707-3p expression and GWAS signals for educational attainment and head size phenotypes. Experimental manipulation of miR-4707-3p expression within primary human neural progenitors during proliferation showed miR-4707-3p increased the number of proliferating cells. This example confirms the utility of miRNA-eQTLs in understanding how genetic variation may influence brain-related traits through regulation of miRNA expression.

2.2 Results

MicroRNA expression profiling

We profiled the expression of miRNAs across 223 fetal cortical tissue samples from donors between 14 and 21 gestation weeks using high-throughput small-RNA-sequencing. We used a specialized miRNA quantification algorithm, implemented in the miRge 2.0 package [63], to measure the expression of known miRNAs cataloged in miRBase release 22 (March 2018) [64]. Combined with total-RNA-sequencing using an rRNA depletion based library preparation in these same samples, collected in a previous study [42], we used rigorous quality control criteria (Methods) to eliminate 11 samples that were

possible swaps, mixtures, or expression outliers (Extended Data Figure 2.1A-C). Following batch effect correction for known technical confounding variables (Extended Data Fig 2.1A-B), a principal component analysis (PCA) on miRNA expression across the 212 remaining samples revealed the primary variation between samples was driven by gestation week at time of sample collection (Figure 2.1B). This finding is consistent with tissue composition differences between tissue from early gestation, composed of a greater proportion of neural progenitor cells, and late gestation, composed of a greater proportion of neurons [65,66]. This analysis shows that after controlling for known technical variation, miRNA expression captures expected biological variability across these samples.

We conducted a differential expression analysis to identify the miRNAs changing across gestation weeks (Figure 2.1C). We found 269 miRNAs positively correlated and 246 miRNAs negatively correlated with gestation week (false discovery rate (FDR) < 10%). Examples include miR-92b, which has known roles in maintaining stem cell proliferation and is higher expressed in neural progenitors relative to differentiated neurons [31,67]. By contrast, miR-124 is known to be higher expressed in post-mitotic neurons relative to neural progenitors and plays a role in promoting neuronal differentiation [32,68]. Consistent with these roles, we observed miR-92b was significantly upregulated in early gestation week samples, and miR-124 was significantly upregulated in late gestation week samples (Figure 2.1D and 2.1F). Furthermore, validated mRNA targets of miR-92b (SMAD7, TSC1, PER2, and CDKN1C) show differential expression between early and late gestation week samples consistent with targeting and downregulation of mRNA by a miRNA (Figure 2.1E) [69–74]. Validated mRNA targets of miR-124 (RHOA, PTBP1, ACTL6A, and SP1) also show consistent expression patterns in early and late gestation week samples (Figure 2.1G) [68,75–78].This differential expression analysis and the expression patterns of known miRNA-targeted mRNAs are expected given the cell-type compositions of cortical tissue during neuronal differentiation over developmental time.

In addition to quantifying the expression of known miRNAs in miRBase release 22, we quantified the expression of recently discovered miRNAs from studies by Friedländer et al [79] (72 miRNAs from the Friedländer dataset were detected in this study) and Nowakowski et al [33] (7 miRNAs

from the Nowakowski dataset were detected in this study). Finally, using two annotation packages, miRge 2.0 [63] and miRDeep2 [80], we discovered 111 putatively novel miRNAs that were not previously annotated in miRBase release 22, Friedländer et al, nor Nowakowski et al. Novel miRNAs discovered in this study showed sequencing read coverage plots consistent with known miRNAs and many were differentially expressed between early and late gestation week samples (Extended Data Figure 2.2). This represents a novel resource of miRNAs that may not have been previously detected due to unique expression of these miRNAs in developing brain tissue or lower read depth and sample size obtained in previous studies. Though these novel miRNAs have characteristic sequencing read patterns consistent with known miRNAs, they require validation to demonstrate they function as miRNAs [81].

Local-miRNA-eQTLs

Genotyping information from each of the 212 remaining donors revealed a sample of diverse ancestry (Extended Data Figure 2.1E). Following TOPMed mixed-ancestry imputation, 12.4 million genetic variants were combined with the expression of 907 known and novel miRNAs across 212 fetal cortical tissue samples to perform a local-miRNA-eQTL analysis (Figure 2.1A) [82,83]. To control for population stratification in our association testing, we used a mixed-effects linear model which included a kinship matrix as a random effect and 10 genotype principal components (PCs) as fixed effects [84,85]. We included 10 miRNA-expression PCs, technical variables such as sequencing-pool and RNA integrity number (RIN score), and the biological variables of sex and gestation week as additional fixed effect covariates in the model.

Following stringent local and global multiple testing correction (Methods), we identified 70 miRNAs with a local-eQTL, hereafter referred to as emiRs, using a hierarchical multiple comparisons threshold (FDR <5%, see Methods). Of these primary eQTLs, we identified an additional 14 loci with secondary eQTLs and one tertiary eQTL for a total of 85 conditionally-independent local-miRNA-eQTLs (Figure 2.2A). Of the 70 emiRs, one miRNA was cataloged in Friedländer et al, two in Nowakowski et al, and eight were novel miRNAs discovered within this fetal tissue dataset. To assess enrichments in

functionally annotated genomic regions, we also used a relaxed, global-only, multiple testing correction threshold (see Methods) which increased the number of local-miRNA-eQTLs to 200 across 153 emiRs (153, 30, 13, 3, and 1 eQTLs of degree primary, secondary, tertiary, quaternary, and quinary respectively). Discovery of these local-miRNA-eQTLs shows that genetic variation influences miRNA expression in the developing cortex, including the expression of previously unannotated miRNAs.

To characterize whether these miRNA-eQTLs are found in functionally annotated regions of the genome, we assessed whether eQTL signals were enriched in chromatin annotations that were from fetal tissue that were previously separated into male and female sexes [86]. We identified significant enrichments of miRNA-eQTLs within active transcription start sites (TssA) and chromatin associated with strong transcription (Tx), weak transcription (TxWk), enhancers (Enh), and ZNF genes and repeats (ZNF/Rpts). There was also a significant depletion of miRNA-eQTLs within quiescent chromatin (Quies) (Figure 2.2B). These enrichments were robust to either stringent or relaxed multiple testing correction methods used to declare significant eQTLs (Extended Data Figure 2.3). Enrichment of miRNA-eQTL signals within transcribed chromatin is expected given that most miRNAs are found within hosts or immediately adjacent to genes [87].

Colocalization of miRNA-eQTLs with mRNA-e/sQTLs

Since over 50% of miRNAs are found within host genes, we classified the miRNA-eQTLs based on whether the emiR is located within a host gene or intergenic [87]. Of the 70 emiRs at the stringent significance threshold, 49 are located within a host gene (100 of 153 emiRs are within hosts using the relaxed threshold). We found that miRNA-eQTLs are often close to their emiRs, and this trend is consistent whether or not the emiR is within a host gene (Figure 2.3A).

To further character these miRNA-eQTLs, we conducted a colocalization analysis to discover if the same genetic variants regulating miRNA expression also regulate mRNA expression (mRNA-eQTLs) and splicing (mRNA-sQTLs). mRNA-eQTLs and mRNA-sQTLs were discovered in an expanded set of fetal tissue samples largely overlapping with those samples used in our miRNA-eQTL analysis [42,43].

We found 17 and 12 colocalizations with eQTL and sQTLs, respectively. For emiRs within a host mRNA, we observed that miRNA expression was often positively correlated with mRNA expression (Figure 2.3B). Of these emiRs within hosts, we found 3 with mRNA-eQTL colocalizations and 4 with mRNA-sQTL colocalizations. Interestingly, expression of the few emiRs with a co-localized mRNA-eQTL were positively correlated with expression of their mRNA hosts, while expression of the few emiRs with a co-localized mRNA-sQTL were negatively correlated with their host mRNA expression.

This phenomenon is highlighted by a colocalization between a miRNA-eQTL of hsa-miR-1307-5p with an mRNA-sQTL for ATP5MK (Figure 2.3C). Hsa-mir-1307 sits within exon three of the 5' UTR of ATP5MK. In our dataset, we found evidence for five distinct intron excisions within the 5' UTR of ATP5MK (labeled SpliceA-SpliceE). Intron excision was quantified as percent spliced in (PSI) and normalized across all junctions in this cluster. Among these splice sites, we observed an association between genotypes (rs7911488 genotypes A and G) and splice site utilization at SpliceA and SpliceD. This same variant was associated with expression of miR-1307-5p (Figure 2.3D). The G allele was associated with an increased utilization of SpliceA and SpliceD, while these same samples showed a decreased expression of miR-1307-5p. These data are consistent with biogenesis of miR-1307-5p from an exon of its host gene, ATP5MK. Removal of exon three of ATP5MK results in more miR-1307-5p. In this case, common genetic variation influences on both splicing and miRNA expression led to understanding of this miRNA's biogenesis.

miRNA-eQTL tissue specificity

We next sought to quantify the degree to which our fetal cortical tissue miRNA-eQTLs are distinct from miRNA-eQTLs discovered in other tissues. We compared our brain miRNA-eQTLs (70 emiRs at the stringent testing correction threshold) to those of a large eQTL analysis in blood [50]. Of the 76 and 70 total emiRs in blood and brain tissue respectively, most are unique to a given tissue (65 and 54 are unique to blood and brain respectively; Figure 2.4A). There are only 11 miRNAs that are emiRs in both blood and brain tissue. Of these emiRs present in both blood and brain, only three of these eQTLs

have a colocalized genetic signal, implying the same causal variants affect miRNA expression in both tissues. The remaining eight emiRs found in both tissues do not share causal variants and therefore have tissue-specific genetic mechanisms regulating expression of these miRNAs. This shows that miRNA-eQTLs between blood and brain are highly tissue-specific, and the genetic variants regulating expression of shared emiRs can also be tissue-specific.

To further characterize the tissue-specificity of miRNA-eQTLs, we calculated the fraction of brain variant/miRNA pairs that are true associations within blood miRNA-eQTLs (π_1) [88]. Of the 76 blood emiRs, 52 were expressed at sufficient levels within our brain samples to test for genetic association. The fraction of blood associations also found in brain was 0.23 (+/- 0.2 95% conf. int.) (Figure 2.4B). By comparison, the fraction of mRNA-eQTLs from blood tissue that are true associations in brain was 0.51 (+/- 0.04 95% conf. int.) [89]. This provides further evidence that miRNA-eQTLs have stronger tissue specificity as compared to mRNA-eQTLs.

<u>miR-4707-3p is implicated in brain size and cognitive ability</u>

To determine if our miRNA-eQTLs may explain molecular mechanisms underlying disease risk imparted via GWAS loci associated with brain disorders or inter-individual differences in brain traits, we performed colocalization analyses between our 85 stringently defined local-miRNA-eQTLs and 21 GWAS summary statistics. We discovered one robust colocalization between an eQTL for miR-4707-3p and multiple brain phenotypes, including educational attainment, head size, and a subthreshold association for cortical surface area (Figure 2.5A). The eQTL for miR-4707-3p expression (rs4981455, alleles A/G) also co-localizes with an mRNA-eQTL for HAUS4 expression. Mir-4707 is located within the 5' UTR of the HAUS4 gene. Despite this, and in contrast to the above example for mir-1307, the allele associated with increased expression of miR-4707-3p is also associated with increased expression of its host gene, HAUS4 (Figure 2.5B and 2.5C). The index variant, rs4981455, is in high linkage disequilibrium (LD; $r^2 > 0.99$) with another variant (rs2273626, alleles C/A) which is within the "seed" sequence of miR-4707-3p (Extended Data Figure 2.4). The A allele at rs2273626, corresponding to index

variant allele G, would most likely change miR-4707-3p targeting. However, we did not detect any miR-4707-3p expression in samples with the G/G genotype at the index variant, therefore we did not study altered targeting of miR-4707-3p-G as it is not expressed (Figure 2.5B, Extended Data Figure 2.4B). This finding is unlikely to be caused by reference mapping bias, because the miRNA quantification algorithm we used, miRge 2.0, accounts for common genetic variants within mature miRNA sequences (Methods) [63]. We also performed an allele specific expression analysis for donors that were heterozygous at rs2273626 (Extended Data Figure 2.4C). rs2273636-A showed consistently lower expression, providing further support for the detected miRNA-eQTL ($p=3.63x10^{-14}$, using a paired, two-sided t-test). We detected only one read containing the A allele in three rs2273626 heterozygote donors. This indicates that miR-4707-3p is either expressed at levels too low for detection or not at all in chromosomes harboring the rs4981455 G allele.

In addition to the HAUS4 mRNA-eQTL colocalization, the miRNA-eQTL for miR-4707-3p expression is also co-localized with a locus associated with educational attainment (Figure 2.5A) [52]. In this case, the alleles associated with increased expression of miR-4707-3p are also associated with decreased educational attainment. We also highlight here associations to global cortical surface area (GSA) [28]. Although this locus does not have any genome-wide significant associations to GSA, a pattern of decreased p-values within the same LD block associated with miR-4707-3p expression implies that this locus may hold a significant association in future GSA GWAS with increased sample size. Variants associated with increased expression of miR-4707-3p are associated with decreased GSA. Supporting the association to cortical surface area, variants associated with increased miR-4707-3p expression are also co-localized with variants associated with decreased head size, a phenotype highly correlated with cortical surface area (data not shown from a publication in-preparation) [54]. This evidence suggests that the genetic risk for decreased head size and decreased cognitive abilities may be mediated through increased expression of miR-4707-3p in developing human cortex. Few publications imply a known function for miR-4707-3p, however, HAUS4 is known to play a role in mitotic spindle assembly during cell division and a potent regulator of proliferation [90–92]. Unifying these observations

lead us to a hypothesis, consistent with the radial unit hypothesis [62], whereby increased expression of miR-4707-3p may influence neural progenitor fate decisions during fetal cortical development ultimately leading to a decreased cortical surface area.

miR-4707-3p modulates proliferation in phNPCs

Given the genetic evidence implicating miR-4707-3p in cortical development and size, we next asked whether increased expression of miR-4707-3p in primary human neural progenitor cells (phNPCs), which model neurogenesis, influenced proliferation or cell fate decisions (Figure 2.6A) [93]. Using lentiviral transduction, we exogenously expressed mir-4707 in phNPCs derived from two genetically distinct donors. The cells were cultured in media with growth factors that retain the phNPCs in a proliferative state [61]. After confirming over-expression of miR-4707-3p (Figure 2.6B), we measured proliferation using an EdU pulse to label cells in S-phase of the cell cycle. At eight days post-transduction, we observed an increase in the number of EdU positive nuclei in samples which over-expressed miR-4707-3p, which indicates this miRNA causes an increased rate of proliferation (Figure 2.6D). Increases in proliferation could be due to either more neurogenic or more self-renewal fate decisions. Investigating further, we measured gene expression for a set of proliferation markers, progenitor markers, and neuronal markers in a time course experiment in phNPCs transduced with our expression construct (Figure 2.6E). At four, six, and eight days post transduction, we observed increased expression of the proliferation markers, KI67 and CCND1, which corroborate our findings using the EdU assay. We also observed an increase in the progenitor markers, PAX6 and SOX2, as well as the neuronal markers, DCX and TUJ1 (Beta-Tubulin III).

To measure differences in differentiation of phNPCs, we modified our experiments to quantify the number of progenitors and early born neurons after over-expression of miR-4707-3p. We used an *in vitro* differentiation model which removes the growth factors in the media that maintain phNPCs in a proliferative state [61,93]. We used stable lines of phNPCs which expressed the primary hairpin for mir-4707 under the control of a tetracycline response element (TRE; inducible with doxycycline [DOX];

Figure 2.7A). We assayed the resultant differentiated cells at one and two weeks post plating in differentiation media either continually overexpressing the mir-4707 (+DOX) or control (-DOX). RNA was extracted to confirm expression of miR-4707-3p, and immunocytochemistry (ICC) was used to test the fraction of cells labeled with the progenitor marker, GFAP (Figure 2.7B).

We detected an increase in miR-4707-3p expression at both one and two weeks post differentiation. With the inclusion of doxycycline, expression of miR-4707-3p was significantly greater in the D54 cell line, while expression was only moderately increased in the D88 cell line (Figure 2.7C). After ICC labeling, we detected a significant decrease in the total number of cells per well for both D54 and D88 lines with the inclusion of doxycycline at both the one and two week timepoints (Figure 2.7D). When accounting for the decrease in total number of cells, we see that the fraction of GFAP positive cells goes down with doxycycline (Figure 2.7E). This decrease in progenitor-labeled cells, indicates miR-4707-3p over-expression in phNPCs may lead to an increase in asymmetric divisions (neurogenic divisions) and reduced symmetric divisions resulting in depletion of progenitors and total number of cells. These observations are consistent with the radial-unit hypothesis and may explain how increased miR-4707-3p during neurogenesis leads to decreased GSA and head size (Figure 2.7F).

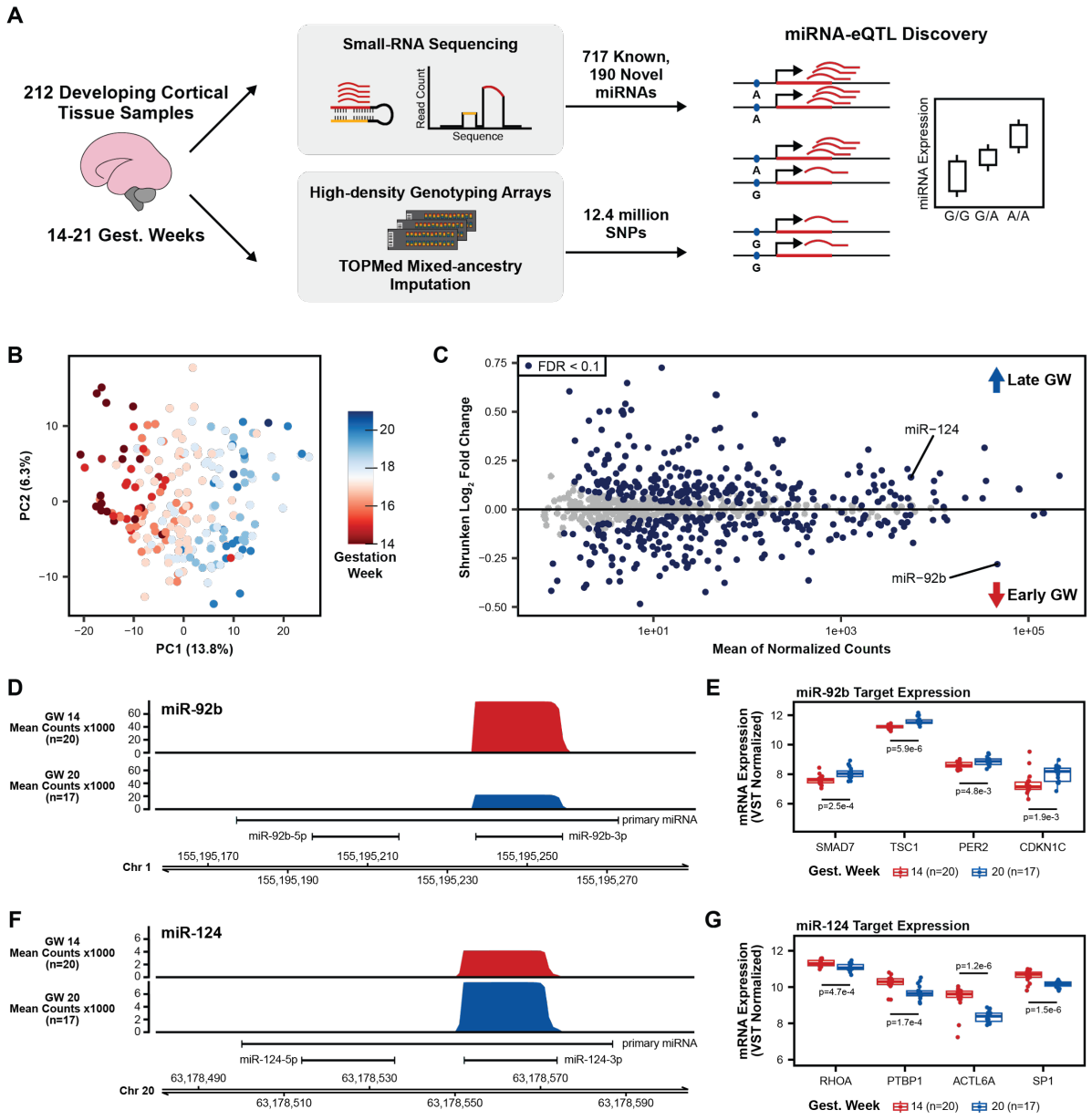


Figure 2.1. Study Design and miRNA Expression Analysis

A. Small-RNA sequencing was used on a total of 212 cortical tissue samples between 14 and 21 gestation weeks to look for novel miRNAs and quantify expression of known and novel miRNAs. Combined with imputed genotypes, a genetic association analysis was performed to discover local-miRNA-eQTLs.

B. Principal component analysis (PCA) on miRNA expression (miRBase release 22) after correcting for the known technical batch effects of sequencing pool and RNA purification method and removal of outlier samples (Extended Data Figure 2.1).

C. Differential miRNA expression analysis using gestation week as the comparison variable. MiR-124, with known roles in neurogenesis, is upregulated among late gestation week samples, miR-92b, with known roles in progenitor proliferation, is upregulated among early gestation week samples.

D. Coverage plot of mean small-RNA sequencing read counts at the miR-92b locus on chromosome 1 for a set of early gestation week samples (gest. week 14, n=20) and a set of late gestation week samples (gest. week 20, n=17).
E. Known miR-92b targets exhibit expression patterns consistent with down-regulation by a miRNA when separated between early and late gestation week samples. P-values when comparing mean expression using a two-sided t-test between early and late gestation week samples.
F. Coverage plot, as in D, but for the miR-124 locus on chromosome 20.
G. Known miR-124 targets expression patterns as in E.

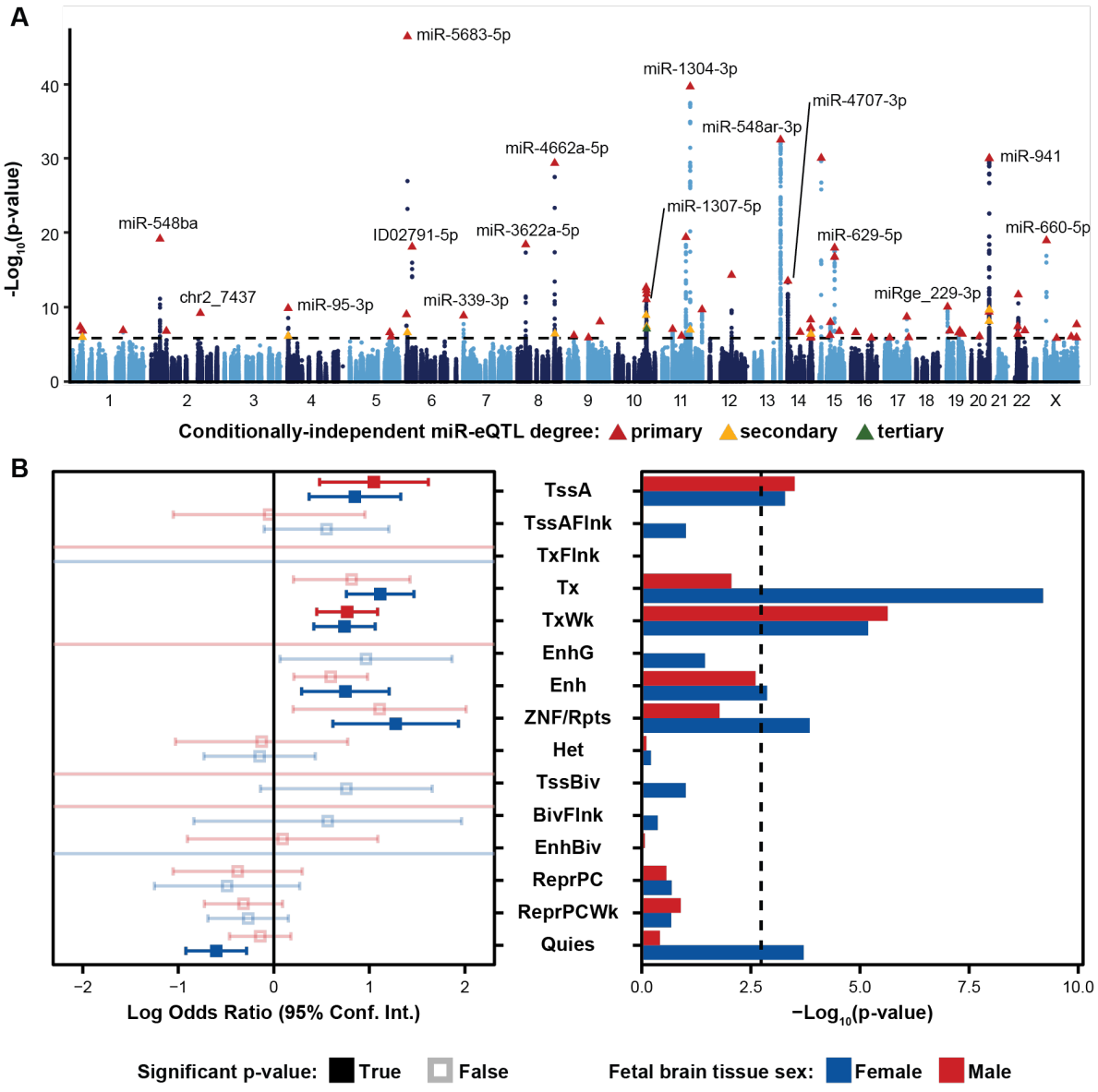


Figure 2.2. Local-miRNA-eQTLs and their Enrichment in Functional Chromatin

A. Manhattan plot showing -Log_{10}(p-values) for variants in a 2Mb window around each expressed miRNA. Index variants for 70 primary, 14 secondary, and 1 tertiary eQTL denoted by triangles and colored by degree by which the eQTL is conditionally-independent. Dotted line represents a stringent p-value significance threshold of 1.434×10^{-6} after local and global multiple testing correction at FDR < 5%.

B. Enrichment of local-miRNA-eQTLs within functionally annotated chromatin states. Local-miRNA-eQTLs passing the stringent p-value significance threshold were used for enrichment analysis. For other p-value significance thresholds, see Extended Data Figure 2.3. Chromatin states were predicted using ChromHMM for both male and female fetal brain tissue. Left, log odds ratio and 95% confidence interval for significant enrichments are in solid colors. Non-significant enrichments in faded color outlines. Right, -Log_{10}(p-values) for each enrichment test. Dotted line represents a multiple testing corrected p-value threshold of 1.83×10^{-3} (Bonferroni correction at $\alpha < 0.05$). miRNA-

eQTLs are significantly enriched within active transcription start sites (TssA) and chromatin associated with strong transcription (Tx), weak transcription (TxWk), enhancers (Enh), and ZNF genes and repeats (ZNF/Rpts). There is also a significant depletion of miRNA-eQTLs within quiescent chromatin (Quies). Other abbreviations: flanking active tss (TssAFlnk), transcription at gene 5' and 3' (TxFlnk), genic enhancers (EnhG), heterochromatin (Het), bivalent/poised TSS (TssBiv), flanking TssBiv (BivFlnk), bivalent enhancer (EnhBiv), repressed polycomb (ReprPC), and weak repressed polycomb (ReprPCWk).

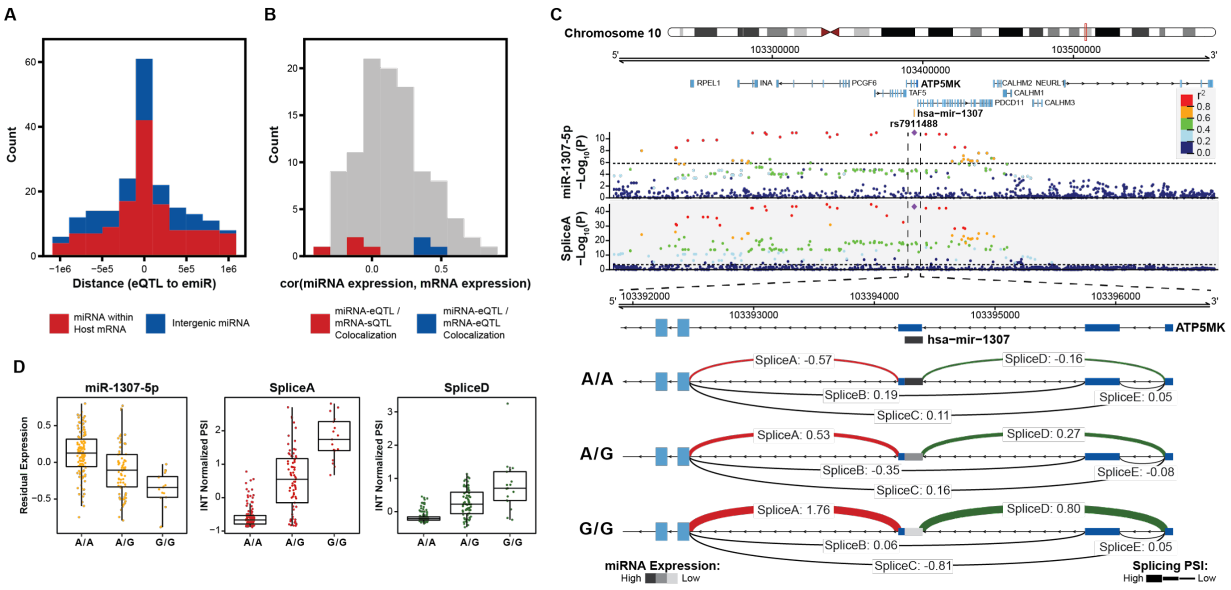


Figure 2.3. Colocalization of miRNA-eQTLs and mRNA-e/sQTLs

A. Histogram of distances from miRNA-eQTL to its significantly associated miRNA (eSNP to emiR distance in base pairs) for each of the 85 miRNA-eQTLs across 70 emiRs. Colored by whether the emiR resides within a host mRNA or intergenic.
B. For emiRs that resign within a host mRNA, histogram of expression correlation between miRNA and mRNA host. Colors represent correlations for miRNA-eQTLs that colocalize with mRNA-eQTLs or mRNA-sQTLs.
C. Example of a miRNA-eQTL, for miR-1307-5p expression, which colocalizes with a mRNA-sQTL for splice variants in the 5' UTR of ATP5MK. Top locus dot-plot indicates -Log_{10}(p-values) on variant associations to miR-1307-5p expression. Bottom locus dot-plot shows -Log_{10}(p-values) for SpliceA utilization measured as inverse normal transformed (INT) percent spliced in (PSI). Variants are colored by pairwise linkage disequilibrium (LD, r^2) to the index variant, rs7911488. Zoom-in shows the 5' UTR of ATP5MK and the INT normalized PSI for each of five splice sites, labeled A-E, separated by genotype at rs7911488. Lines representing splice junctions at SpliceA and SpliceD are colored and weighted for relative PSI at the indicated genotypes. The line representing mir-1307 is shaded based on relative miR-1307-5p expression at the indicated genotypes.
D. Residualized expression (after removal of known and unknown confounders) of miR-1307-5p at rs7911488 genotypes. INT normalized PSI values for SpliceA and SpliceD at rs7911488 genotypes.

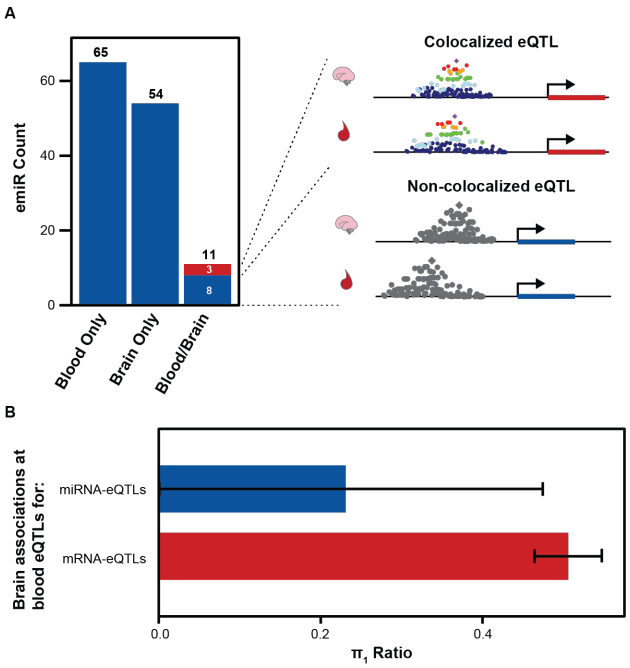


Figure 2.4. Comparison of Brain and Blood miRNA-eQTLs

A. MiRNAs with significant associations, emiRs, separated by unique to blood, unique to brain, or shared. Three of 11 shared emiRs have co-localized genetic signals in the blood and brain datasets. Cartoon locus dot-plot shows a representation of a colocalized genetic signal or a non-colocalized signal for a miRNA-eQTL that is present in both tissues.
B. The fraction of brain eQTL associations that are estimated true associations within the blood eQTL dataset (π_1) separated by miRNA-eQTLs and mRNA-eQTLs. Error bars represent 95% upper and lower confidence intervals after 100 bootstrap samplings.

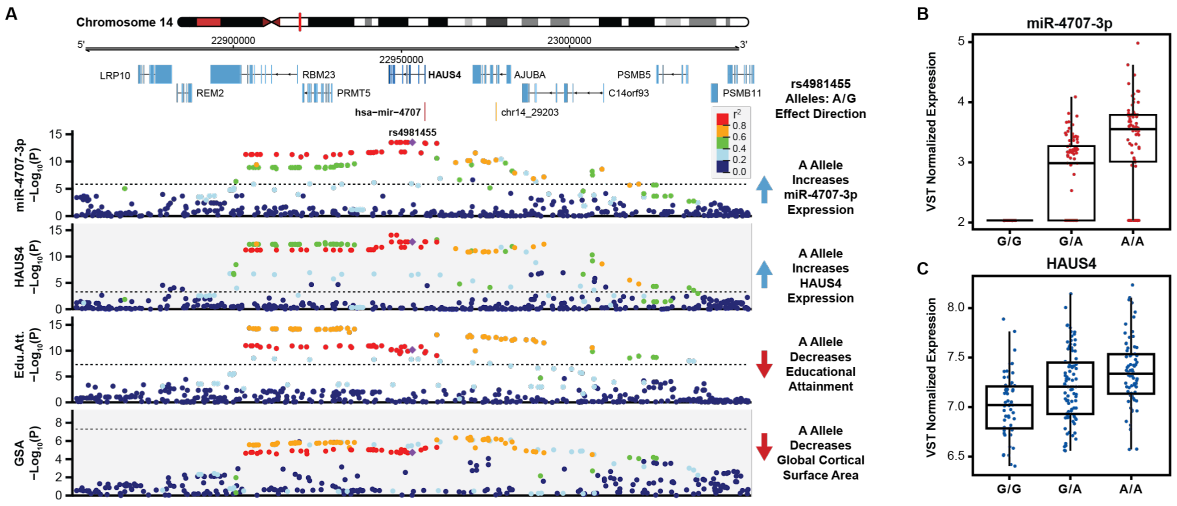


Figure 2.5. A miRNA-eQTL for miR-4707-3p Colocalizes with Multiple Brain Traits

A. A locus dot-plot surrounding hsa-mir-4707 on chromosome 14 showing -Log_{10}(p-value) for various associations. Top track shows associations for expression of miR-4707-3p with the index variant, rs4981455. The second track shows associations surrounding a significant mRNA-eQTL to HAUS4 expression. The third track shows associations at this locus for a GWAS to educational attainment. The fourth track shows variant associations to global cortical surface area (GSA). Variants are colored by pairwise linkage disequilibrium (LD, r^2) to rs4981455 in each of the four tracks. Dotted line in the first and second tracks represents a stringent p-value significance threshold of 1.43×10^{-6} and 8.17×10^{-4} respectively after local and global multiple testing correction at FDR < 5%. Dotted lines in bottom two tracks represents the global genome-wide significance threshold of 5×10^{-8}.

B. Boxplot showing variance stabilizing transformed (VST) expression of miR-4707-3p for samples with the indicated genotype at rs4981455. Samples with the G/G genotype show no expression of miR-4707-3p.

C. Boxplot showing VST expression of HAUS4, the host gene for mir-4707, at rs4981455.

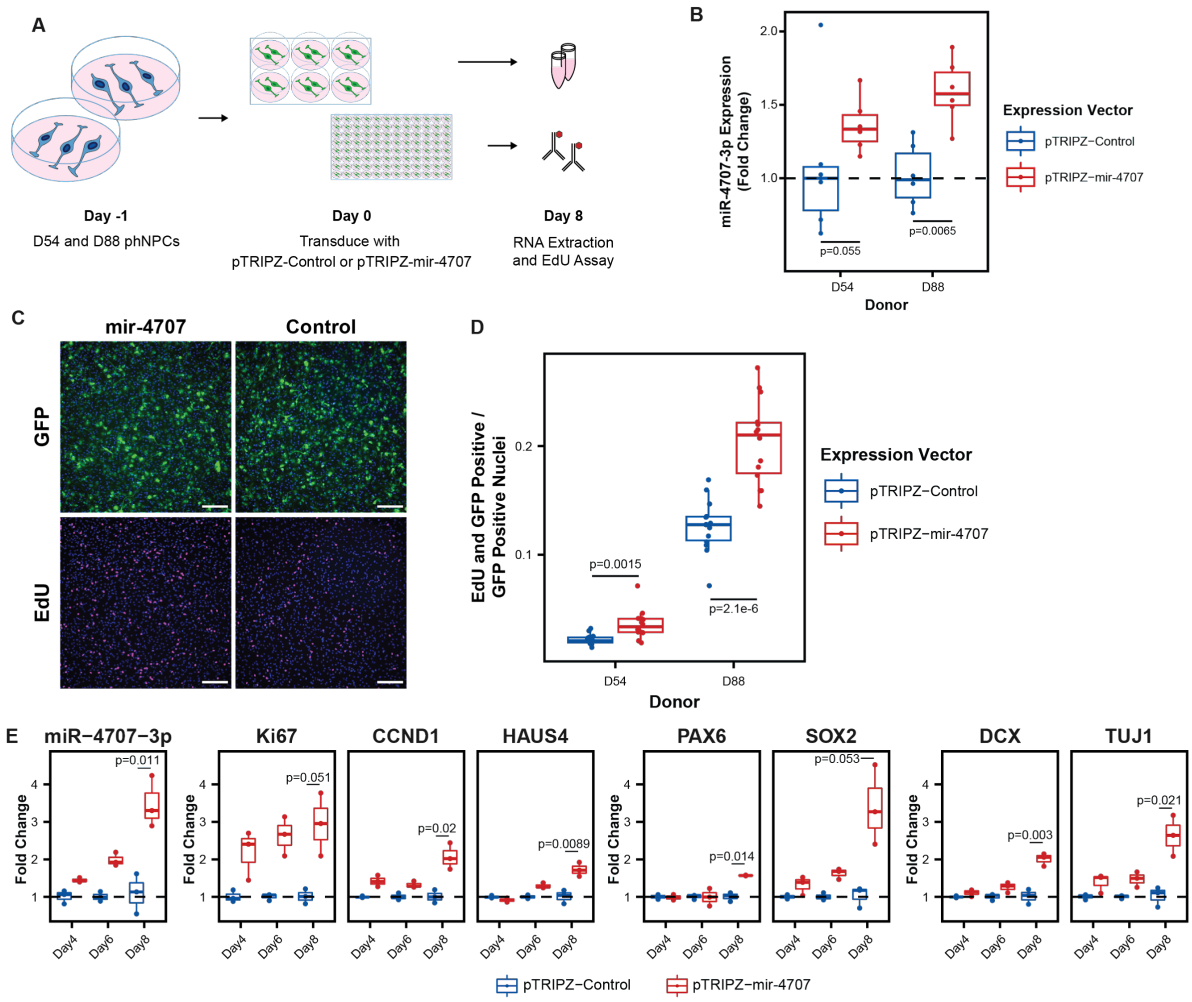


Figure 2.6. MiR-4707-3p Over-expression in Proliferating phNPCs

A. Experiment overview. Primary human neural progenitor cells (phNPCs) from two donor lines (D54 and D88) were maintained in proliferation media with growth factors that retain the phNPCs in a proliferative state and prevent differentiation into neurons. At day -1, cells were plated into 6-well and 96-well plates. The next day, day zero, cells are transduced with pTRIPZ-mir-4707-EGFP (pTRIPZ-4707) or a control pTRIPZ-EGFP (pTRIPZ-Control). Addition of doxycycline in the proliferation media turns on mir-4707 and EGFP expression. Cells were assayed at eight days post transduction.

B. At eight days post transduction, D54 and D88 transduced with pTRIPZ-4707 showed an increase of miR-4707-3p expression over cells transduced with control virus. P-values from a two-sided t-test on 6 samples per condition.

C. D54 GFP signal in both pTRIPZ-mir-4707 and pTRIPZ-Control at eight days post transduction. EdU pulse labeled nuclei also at eight days post transduction. Nuclei in all four images are also stained with DAPI (blue signal). Scale bars 200µm.

D. Proliferation assay using EdU pulse labeled cells. Both D54 and D88 cells transduced with pTRIPZ-4707 showed an increased number of nuclei stained with both EdU and GFP. Each well was normalized for the total number of GFP positive nuclei. P-values from a two-sided t-test on 14 technical replicates (wells) per condition.

E. Time course experiment using D88 cells transduced with pTRIPZ-4707 or control virus as in A, but assayed at four, six, and eight days post transduction. Gene expression measured with qPCR shows increased miR-4707-3p in cells transduced with pTRIPZ-4707 over pTRIPZ-Control. Increased expression is also seen in the proliferation markers (Ki67, CCND1, and HAUS4), the progenitor markers (PAX6 and SOX2), and the neuronal markers (DCX and TUJ1). P-values from a two-sided t-test on fold changes between samples transduced with pTRIPZ-4707 and pTRIPZ-Control at eight days post transduction for three samples in each condition.

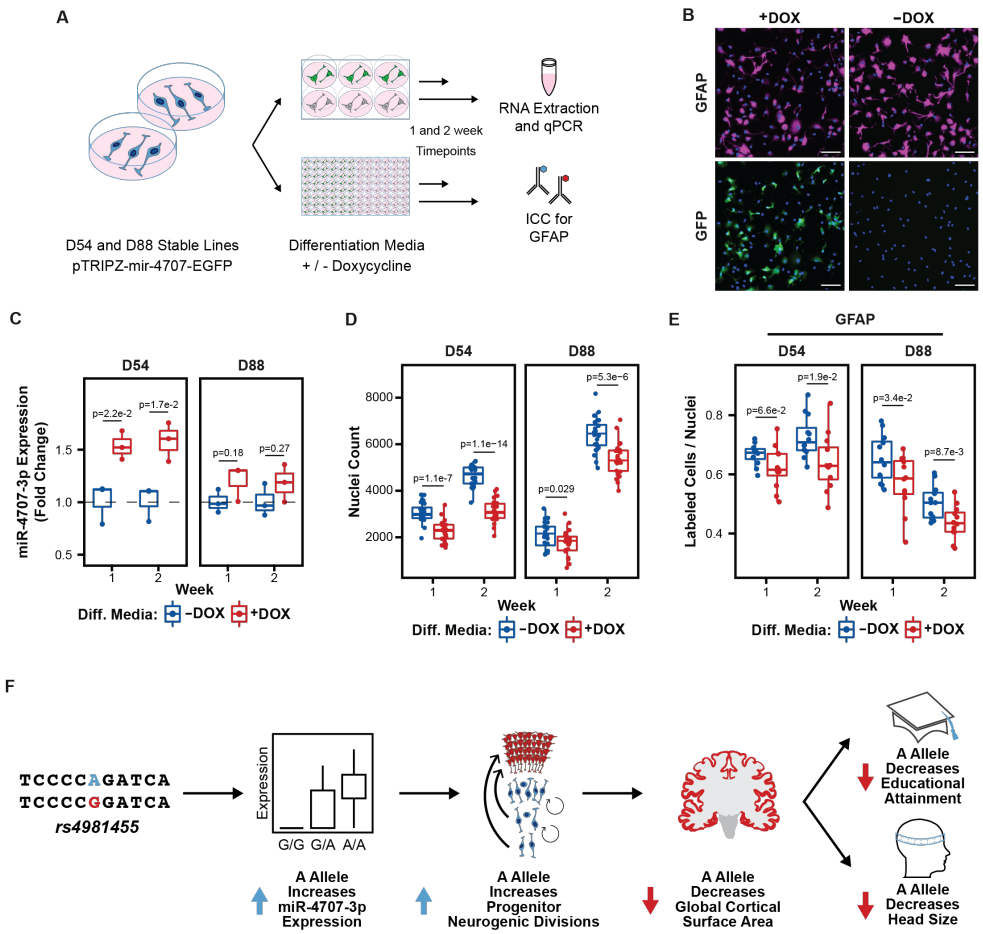


Figure 2.7. MiR-4707-3p Over-expression in Differentiated phNPCs

A. Experiment overview. Two genetically distinct donor cell lines (D54 and D88) of primary human neural progenitor cells (phNPCs) were transduced with pTRIPZ-mir-4707-EGFP and selected with puromycin to form stable lines in proliferation media. Stable lines were then plated in 6-well or 96-well plates with and without doxycycline to induce miR-4707-3p expression in differentiation media. Plates were assayed at 1 and 2 weeks post plating.

B. D54 cells at one week post plating in differentiation media with (+) or without (-) doxycycline (DOX). Cells were labeled with the progenitor marker, GFAP.

C. RNA for D54 and D88 was extracted at one and two weeks post differentiation and assayed for expression of miR-4707-3p using qPCR. Fold change in expression relative to mean expression without doxycycline (-DOX). P-values from a two-sided t-test on three technical replicates (wells) per condition.

D. D54 and D88 nuclei were labeled with DAPI and counted at one and two weeks post differentiation. P-values from a two-sided t-test for 48 technical replicates (wells) per condition.

E. D54 and D88 cells were labeled with GFAP and counted. Labeled cell counts were normalized to the number of nuclei per well. P-values from a two-sided t-test for 24 technical replicates (wells) per condition.

F. Hypothesis of how genetic variation at a locus significantly associated with multiple brain traits may be explained by a miRNA-eQTL. Genetic variation at rs4981455 influences expression of miR-4707-3p. Increased expression of miR-4707-3p leads to increased progenitor neurogenic divisions. Based

on the radial unit hypothesis, increased neurogenic divisions may lead to decreased global cortical surface area, which would in turn lead to decreased head size and decreased educational attainment.

2.3 Discussion

Using small-RNA-sequencing, we reveal robust miRNA expression across cortical tissue during mid-gestation, a stage and tissue which has not previously been captured in previous eQTL studies using standard RNA-sequencing techniques. In addition to the known roles of miR-92b and miR-124 on progenitor proliferation and neurogenesis, our differential expression analysis shows many other miRNAs likely play crucial roles in cortical development [31,68]. We were also able to find greater than 100 likely novel miRNAs and further evidence of recently discovered miRNAs within these tissue samples. These novel miRNAs have sequencing read coverage characteristic of known miRNAs, and they are differentially expressed across gestation weeks like miRNAs with known roles in neurogenesis. Investigating how these known and novel miRNAs function during neuronal differentiation may yield new gene regulatory mechanisms involved in human neurogenesis.

We also discovered 85 local-miRNA-eQTLs in a tissue-type and at a developmental time point with known influence on brain structure and cognitive traits. Despite many emiRs residing within host genes, miRNA-eQTLs seldom colocalize with mRNA-eQTLs for their host mRNAs. This implies a regulatory mechanism by which miRNA expression is largely independent of that which governs host mRNA expression, highlighting the unique information gained from miRNA-eQTLs that would otherwise be missed in standard mRNA-eQTL analyses. We found that the small subset of miRNA-eQTLs which colocalize with their host mRNA eQTLs have positively correlated expression, which would indicate a common genetic regulatory mechanism governing expression of both RNAs. While miRNA-eQTLs which colocalize with a host mRNA-sQTL have negatively correlated expression. The genetic regulatory mechanisms governing miRNA biogenesis and expression uncovered by our eQTL analysis provides mostly unique mechanisms when compared to current mRNA-eQTL datasets. These miRNA-eQTLs will be a continued resource in the pursuit of describing the genetic risk loci uncovered by current and future GWAS for brain traits and disorders.

In contrast to mRNA based eQTLs, miRNA-based eQTLs appear to map less frequently to known brain disease loci. The lack of colocalizations are highlighted when compared to mRNA-eQTLs in

the same tissue where 844 colocalizations with brain-trait and disorders were discovered across 18,667 mRNA-eQTLs despite a similar sample size (n=235 for mRNA-eQTLs vs n=212 for miRNA-eQTLs) [42,43]. This suggests that genetically regulated miRNAs either may not be a major contributor to neuropsychiatric disorders or current GWAS are underpowered to detect loci mediated through miRNA expression. We suspect that this may reflect the effect of purifying selection on miRNAs. MiRNAs are known to have broad downstream regulatory effects across hundreds or thousands of targeted mRNAs, and therefore the genetic mechanisms regulating miRNA expression may be more tightly regulated than for mRNA expression, as has previously been shown for transcription factors [94,95]. Rare variants, less subject to the influences of selective pressure, may be governing miRNA expression which this study did not have the power or methodology (i.e., whole genome or exome sequencing) to detect.

Nevertheless, we did find one colocalization between a miRNA-eQTL for miR-4707-3p expression and GWAS signals for head size phenotypes and educational attainment. This revealed a possible molecular mechanism by which expression differences in this miRNA may influence brain size and cognition. Experimental over-expression of miR-4707-3p in differentiating phNPCs showed a decrease in the number of proliferative cells resulting in a reduction in total number of cells. These findings are consistent with the radial-unit hypothesis which would explain a decreased cortical surface area by depletion of the neural progenitor pool of cells through increased neurogenesis [25,62].

An interesting feature of this genomic locus is the presence of both a miRNA-eQTL, for miR-4707-3p, and a mRNA-eQTL, for HAUS4. Although yet to be experimentally tested, the known effects on cell proliferation by the HAUS4 gene implies that increased expression of HAUS4 in neural progenitors would most likely lead to increased proliferation [91,92]. However, the effect of increased HAUS4 expression in newly born and maturing neurons is not known. It may be the case that the increased proliferative divisions driven by HAUS4 are totally counteracted by miR-4707-3p driven neurogenesis leading to decreased head size. Furthermore, we did not detect expression of miR-4707-3p in samples with the genotype G/G at the eQTL index variant rs4981455. This implies that the presence of the A allele turns on miR-4707-3p expression and individuals with the G/G genotype have no miR-4707-

3p expression in developing cortical tissue. This further highlights the utility of studying miRNA-eQTLs, as uncovering only the mRNA-eQTL at this locus would not reveal the full genetic mechanism leading to inter-individual differences in the head size and cognitive phenotypes.

Here we highlight one example of how miRNA expression leads to differences in brain size and cognition through altered neurogenesis during cortical development. We have yet to uncover the specific regulatory effects of miR-4707-3p (which genes are targeted or which pathways disrupted), but the effect on cellular behavior is clear. The lengthening of neurogenesis and associated expansion of the brain are hallmarks of the evolutionary differences between humans and other mammals [14,96–98]. Comparative genomic studies have revealed that human-specific gene regulatory differences in developing neocortex are associated with neurogenesis, brain complexity, and disease, and that primate-specific miRNAs have been shown to play a role in post transcriptionally regulating gene expression associated with these developmental processes [21,99–102]. Here, we show that miRNAs also play a role in differences in brain size between humans. Continued work on predictive miRNA targeting algorithms and on experimental methods to uncover miRNA regulatory networks will be crucial to further understanding the molecular pathways that lead to brain size or cognitive differences between individuals.

2.4 Methods

Tissue Procurement

Human prenatal cortical tissues samples were obtained from the UCLA Gene and Cell Therapy Core following IRB regulations for 223 genetically distinct donors (96 females, 127 males, 14-21 gestation weeks) following voluntary termination of pregnancy. Tissue samples were flash frozen after collection and stored at -80C. These tissue samples overlapped with those used in a previous mRNA-eQTL study. Cortical tissue samples from an additional 3 donors were micro-dissected into germinal zone and cortical plate sections as previously described [63] yielding 17 more tissue samples which were used for novel miRNA discovery but withheld from the miRNA-eQTL analysis.

Library Preparation and Sequencing

Total RNA was extracted using miRNeasy-mini (QIAGEN 217004) kits or was extracted using trizol with glycogen followed by column purification. Library preparation for small-RNA was conducted using TruSeq Small RNA Library Prep Kits (Illumina RS-200). RNA libraries were randomized into eight pools and run across eight lanes of an Illumina HiSeq2500 sequencer at 50 base-pair, single-end reads to a mean sequencing depth of 11.7 million reads per sample. mRNA library preparation and sequencing was previously described [42]. Briefly, library preparation for total RNA was conducted via TruSeq Stranded RNA Library Prep Kits (Illumina 20020597) with Ribozero Gold ribosomal RNA depletion. Libraries were sequenced with 50 base-pair, paired-end reads to a mean read depth of 60 million reads per sample.

MicroRNA Expression Analysis

Small RNA-sequencing FASTQ files were used as input to the miRge 2.0 [63] annotation workflow to quantify expression of known miRNAs from miRBase release 22 (March 2018) [64]. Briefly, sequencing reads were first quality controlled, adaptors removed, and collapsed into unique reads. The reads were then annotated against libraries of mature miRNA, hairpin miRNA, tRNA,

snoRNA, rRNA, and mRNA. The miRge 2.0 workflow protects against reference mapping bias by incorporating common genetic variants into the mature miRNA sequence library and allowing zero mismatched bases on the first pass of read annotation. A second pass of unannotated reads allows for mismatched bases to identify isomiRs. Unmapped reads were then used as input to a second annotation pipeline to quantify expression of recently discovered novel miRNAs from Friedländer et al [79] and Nowakowski et al [33]. Bowtie v1.2.2 [103] was used to map reads to the UCSC hg38 reference genome using the following options: -v 2 -5 1 -3 2 --norc -a --best -S --chunkmbs 512. Mapped reads were then counted using featureCounts [104] against a custom GTF file including the Friedländer and Nowakowski novel miRNAs using the following options: -s 0 -M -f -O.

To test for allele-specific expression of miR-4707-3p containing variants at rs2273626, we used Bowtie with a modified reference genome which only included miR-4707-3p mature sequence. Sequencing reads were allowed to map to either AGCCC**G**CCCCAGCCGAGGTTCT (reference allele C, complementary strand G) or AGCCC**T**CCCCAGCCGAGGTTCT (alternate allele A, complementary strand T) with no mismatches using the Bowtie options: -n 0 --norc --all -S. Mapped reads were then counted with featureCounts as above. Read counts, specific to each genotype, from samples heterozygous at rs2273626 were plotted in Extended Data Figure 2.4C.

We also defined an additional set of novel miRNAs discovered within our 240 sample dataset using miRge 2.0 and miRDeep2 [80] prediction pipelines. Putatively novel miRNAs predicted using miRge 2.0 (all predictions) and miRDeep2 (predictions with a score greater than zero) were removed if they overlapped with each other, known miRNAs from miRBase release 22, or recently discovered miRNAs from the Friedländer and Nowakowski datasets. Sequencing read coverage plots for each novel miRNA annotation were also created to visually inspect each annotation. Putatively novel miRNAs without a characteristic 5' and 3' mapping pattern (Extended Data Figure 2.2) were removed. Uniquely novel annotations, which passed visual inspection, were compiled and used to create a custom GTF file for use in the above annotation pipeline to quantify novel miRNA expression. Read counts of miRNAs

from miRBase release 22, Friedländer et al, Nowakowski et al, and putatively novel annotations from this study were combined into one count matrix for use in downstream analyses.

mRNA Expression Analysis

Total RNA-sequencing FASTQ files were first filtered and adapter trimmed using trim_galore and the following options: --length 20 --stringency 5. Filtered and trimmed reads were mapped to GRCh38 using STAR v2.5.4b [105]. Mapped reads were counted using featureCounts [104] against the Ensembl GRch38.p7 human gene annotations using the following options: -T 4 -p -t exon. Count data for each sample was combined into a count matrix for downstream analyses.

Genotyping

Genomic DNA was isolated using DNeasy Blood and Tissue Kit (QIAGEN 69504), and genotyping was performed on either HumanOmni2.5 or HumanOmni2.5Exome (Illumina) platforms in eight batches. SNP genotypes were exported and processed into PLINK format using PLINK v1.9 [106]. Quality control and pre-processing of genotypes was also done using PLINK v1.9 as previously described [61]. Briefly, SNPs were filtered based on Hardy-Weinberg equilibrium, minor allele frequency, individual missing genotype rate, and variant missing genotype rate (plink --hwe 1e-6 --maf 0.01 --mind 0.1 --geno 0.05) yielding a total of 1,760,704 genotyped SNPs.

Imputation

Sample genotypes were prepared for imputation using the McCarthy Group's HRC-1000G imputation preparation tool (https://www.well.ox.ac.uk/~wrayner/tools/): perl HRC-1000G-check-bim.pl -b AllSamplesQC.bim -f AllSamplesQC.frq -r 1000GP_Phase3_combined.legend -g -p ALL. This tool produces a script to separate genotype files by chromosome, filter variants to include only reference SNPs, and convert the Plink filesets into VCF files. Compressed VCF files were then uploaded to the

Michigan Imputation Server for use in the Minimac4 imputation pipeline [82,107]. TOPMed Freeze5 was used as the reference panel for imputation [83]. Imputed genotypes were filtered for an imputation quality score (R^2) greater than 0.3, Hardy-Weinberg equilibrium p-value greater than 1e-6, and a minor allele frequency greater than 0.01.

Sample Quality Control

Sample sex was called from the genotype data using PLINK v1.9 based on X chromosome heterozygosity. Sex was confirmed by checking expression of XIST within the mRNA-sequencing data. Of the 223 samples, two were declared female using the genotype data but male by XIST expression and were excluded from downstream analysis. We further sought to detect sample swaps or mixtures by evaluating the consistency of genotypes called via genotyping and those that can be detected by RNA-sequencing using VerifyBamID v1.1.3 [108]. We detected four samples that were mixtures or possible sample swaps with their assigned genotype data (FREEMIX > 0.04 or CHIPMIX > 0.04). Finally, five samples were identified as outliers via PCA analysis after accounting for known technical confounders (Extended Data Figure 2.1C). A total of 212 samples (93 females, 119 males, 14-21 gestation weeks) were used in the miRNA-eQTL analysis.

PCA and Differential miRNA Expression Analysis

Principal component analysis (PCA) was performed using the prcomp() function within the stats package of the R software language [109]. Only miRBase release 22 expressed miRNAs with at least 10 counts across at least 10 samples were included when doing PCA. MiRNA expression was first normalized using the variance-stabilizing transformation (VST) function within DESeq2 [110]. To correct for known batch effects (sequencing-pool and RNA-purification method; Extended Data Figure 2.1), we used the removeBatchEffect() function within the limma package on the VST transformed expression matrix [111]. While removing batch effects, we preserved the effect of gestation week using the design

option within removeBatchEffect(). After known batch effects were removed, PCA was repeated to confirm removal of technical variation across samples and to identify expression outliers (Extended Data Figure 2.1C).

Differential expression analysis was conducted on the 212 samples which survived quality control filtering using DESeq2 [110]. Expression of all known and novel miRNAs, which survived the above expression threshold, were used in the analysis. Gestation week was used as the treatment variable while controlling for technical confounding variables: sequencing-pool, RNA-purification method, RNA-integrity number (RIN), and RNA concentration after extraction. Differentially expressed miRNAs with a Benjamini-Hochberg adjusted p-value < 0.1 were deemed significant (false discovery rate (FDR) < 10%). For visualization of differentially expressed miRNAs (Figure 2.1C) $\log_2$(fold change) values were shrunk using the apeglm method [112].

local-miRNA-eQTL Mapping

We conducted local-eQTL mapping using tissue samples from 212 donors. A total of 866 miRNAs with an expression of at least 10 counts across at least 10 samples were included in this analysis. Since expressed miRNAs can originate from more than one genomic locus, associations were conducted at a total of 907 genomic loci. MiRNA counts were normalized using a variance-stabilizing transformation function within DESeq2 [110]. Normalized expression values were adjusted using a linear model accounting for population stratification as well as known and unknown confounders. Known confounders included: sequencing pool, rna purification method, rna integrity score (RIN), sex, and donor gestation week. Unobserved confounding variables on miRNA expression were controlled using the first 10 principal components from a PCA. Population stratification was controlled using the first 10 principal components from a genotype PCA using PLINK v1.9.

Association testing between variants and residual miRNA expression (after adjusting for confounders) was done using a linear mixed-effects model as implemented in the EMMAX package [84]. To further control for cryptic relatedness and population stratification, we included an identity-by-state

kinship matrix, also constructed using EMMAX (emmax-kin -v -h -s -d 10) by excluding variants located on the same chromosome as the variants tested in the association analysis (MLMe method [85]). In order to prevent a single outlier from driving the association results, variants were filtered before association testing to include only those which did not have exactly one homozygous minor sample and the number of heterozygous samples were greater than one. Variants within 1 Mb upstream of the mature miRNA start position or 1Mb downstream of the mature miRNA end position were tested for association. Using EMMAX, imputed variant dosages were used for association testing (emmax -v -d 10 -Z -t [variant_doages] -k [kinship_mat] -o [output_file] -p [expression_file]).

For multiple-testing adjustment we employed a two-stage analysis which accounts for linkage-disequilibrium between the variants tested (local adjustment) and the total number of miRNAs tested (global adjustment). First, p-values were adjusted locally using the eigenMT package [113]. Locally adjusted p-values were then further adjusted using the Benjamini-Hochberg multiple testing correction for a 5% false discovery rate (FDR < 0.05) across the 907 genomic loci tested. This yielded a nominal p-value of 1.434×10^{-6} as the stringent threshold for which variants were declared significantly associated with miRNA expression. We also declared a relaxed threshold using only global adjustment with a 5% FDR across the 6.3 million independent association tests which yielded a nominal p-value threshold of 2.804×10^{-5}.

To declare conditionally independent local-miRNA-eQTLs, primary eQTLs were first defined as the most significant variant/miRNA pair for each expressed miRNA with at least one variant below the given nominal p-value threshold. An emiR was defined as a miRNA that has at least one variant associated with it. For each emiR, variant association testing was repeated for variants within the original 2 Mb window with the genotypes of the primary eQTL added to the association equation. The most significant variants below the original nominal p-value threshold (1.434×10^{-6} for stringent or 2.804×10^{-5} for relaxed) within this secondary analysis (if any) were defined as secondary eQTLs. The process was repeated (with the inclusion of primary and secondary genotypes in the association equation to find

tertiary eQTLs, primary, secondary, and tertiary to find quaternary eQTLs, etc.) until no variants remained below the nominal p-value threshold.

Colocalization Analysis

Colocalization of local-miRNA-eQTLs with brain-relevant trait GWAS summary statistics (Supplementary Table 4), blood local-miRNA-eQTLs [50], and fetal brain local-mRNA-e/sQTLs from a largely overlapping sample [42] was done by first finding overlapping variants with LD $r^2 >= 0.8$ of each local-miRNA-eQTL index variant and the set of variants within LD $r^2 >= 0.8$ of each trait GWAS, disorder GWAS, eQTL, or sQTL index variant. LD for local-miRNA-eQTL, local-mRNA-eQTL, and local-mRNA-sQTL variants were calculated using the genotypes of the mixed-ancestry samples within each study. LD for GWAS and blood cis-miRNA-eQTL variants were calculated from 1000 Genomes phase 3 European genotypes [114]. After overlaps were detected, colocalization was confirmed by conditional analysis which incorporated the genotypes for a given overlapping index variant into the local-miRNA-eQTL association equation. A resultant increase in p-value beyond the stringent or relaxed p-value threshold confirmed a colocalization between the local-miRNA-eQTL index variant and the given trait GWAS, disorder GWAS, or QTL index variant.

eQTL Enrichment Analysis

Enrichment of local-miRNA-eQTLs within functionally annotated genomic regions was done using GARFIELD v2 [115] in order to control for the distance to transcription start sites, LD, minor allele frequency (MAF) of the tested variants, and the number of effective tests across multiple annotations. Functional annotations were derived from the Roadmap Epigenomics Project [86] for male and female fetal brain (E081 and E082), using the ChromHMM Core 15-state model [116]. MAF and LD for the variants were derived from the 212, mixed-ancestry samples in this study using PLINK v1.9. Minimum local-miRNA-eQTL p-values were used in cases where multiple association tests to different miRNAs

were performed at a given variant. Only p-values surviving the stringent significance threshold were used for Figure 2.2, while other thresholds, including the relaxed threshold, can be seen in Extended Data Figure 2.3.

Comparison to Blood miRNA-eQTLs

To assess the cell-type specificity of miRNA-eQTLs we calculated the π_1 statistic [88]. Blood miRNA-eQTLs were first defined as the emiR-variant pair with the lowest p-value for the 76 emiRs found in the blood miRNA-eQTL analysis [50]. Of the 76 emiRs, 52 of these miRNA were expressed in brain tissue (at least 10 counts in at least 10 samples). At the 52 blood miRNA-eQTLs, nominal p-values from brain miRNA-eQTL association analysis were used to compute the π_0 value using the qvalue() function in the qvalue package [117]. The π_1 statistic was then defined as 1 - π_0. To estimate the standard error, we did 100 bootstrap samplings and computed a 95% confidence interval for each π_1 statistic. An analogous calculation was done using mRNA-eQTLs from an overlapping cortical tissue dataset [42] and whole blood mRNA-eQTLs reported by GTEx [89].

Lentiviral Vector Cloning and Virus Production

To create the miR-4707 expression vector, the tetracycline-inducible, lentiviral expression vector pTRIPZ (ThermoFisher RHS4750) was modified by replacing the sequence for red fluorescent protein (RFP) with the sequence for enhanced green fluorescent protein (EGFP) between the AgeI and ClaI restriction enzyme cut sites. The stem-loop sequence for hsa-mir-4707 (miRbase release 22) was inserted into the multiple cloning site of pTRIPZ-EGFP using XhoI and EcoRI restriction enzymes.

pTRIPZ-mir-4707-EGFP lentivirus was produced in HEK293 cells. HEK293 cells were cultured to 90-95% confluency in DMEM (Life Technologies 11995081) supplemented with 10% FBS (Sigma-Aldrich F2442) and 1x Antibiotic-Antimycotic (Life Technologies 15240096) in 10cm tissue culture treated plates. Cells were triple transfected with 10 μg transfer plasmid, 7.5 μg PAX2 (Addgene plasmid

#12260), and 2.5 μg pMD2.G (Addgene plasmid #12259) using FUGENE HD (Promega E2311). After 24 hours, media was replaced with 12 mL of 1x proliferation base media without the growth factors EGF, FGF, LIF, and PDGF (see progenitor cell culture protocol below). At 24 hours post media change, culture supernatant was filtered through a 0.45 μm syringe filter, aliquoted, and stored at -80°C in single-use aliquots. Lentivirus was titered using the qPCR Lentivirus Titration Kit (Applied Biological Materials LV900).

Primary Human Neural Progenitor Cell Culture

Primary human Neural Progenitor Cells (phNPCs) derived from developing cortical tissue were cultured, as described previously [43,61,93]. Two donor phNPC lines were used: Donor 54 (D54, gestation week 15.5, male, genotype G/G at rs4981455) and Donor 88 (D88, gestation week 14, male, genotype A/A at rs4981455). PhNPCs were grown on tissue culture treated plates that were coated with Growth Factor Reduced Matrigel (Corning 354230) at 50 μg/mL in 1xPBS at 37°C for 1 hr. To maintain phNPCs in a proliferative state, they were cultured in 1x proliferation media consisting of Neurobasal A (Life Technologies 10888-022) with 100 μg/ml primocin (Invivogen, ant-pm-2), 10% BIT 9500 (Stemcell Technologies 09500), 1x glutamax (Life Technologies 35050061), 1 μg/ml heparin (Sigma-Aldrich, H3393-10KU), 20 ng/ml EGF (Life Technologies PHG0313), 20 ng/ml FGF (Life Technologies PHG0023), 2 ng/ml LIF (Life Technologies PHC9481), and 10 ng/ml PDGF (Life Technologies PHG1034). phNPCs lines were split once per week using 0.05% Trypsin-EDTA (Life Technologies 25300062) into 1x proliferation media. Every two or three days, half of the culture media was replaced with 2x proliferation media: Neurobasal A with 100 μg/ml primocin, 10% BIT 9500, 1x glutamax, 1 μg/ml heparin, 40 ng/ml EGF, 40 ng/ml FGF, 4 ng/ml LIF, and 20 ng/ml PDGF.

For phNPC proliferation experiments, cells were plated at $4x10^5$ and $2x10^4$ cells/well in Matrigel-coated 6-well (for RNA extractions: Corning 3516) and 96-well (for immunocytochemistry: Corning 3598) plates respectively with 1x proliferation media. At 24 hrs post plating, cells were transduced with pTRIPZ-mir-4707-EGFP or control (pTRIPZ-EGFP) at 20 IU/cell in 1x proliferation media with 1 μg/ml

doxycycline to express the miRNA (Sigma-Aldrich D9891). Media was changed at 24 hrs post transduction with 1x proliferation media with doxycycline, and plates were fed every two days by changing half of the culture media with 2x proliferation media with 2 μg/ml doxycycline. At 8 days post transduction, 6-well plates were used for RNA extraction, and 96-well plates were EdU labeled and fixed for immunofluorescence staining.

For differentiation experiments, phNPCs were first transduced with pTRIPZ-mir-4707-EGFP and a stable line was created to express mir-4707 under the control of doxycycline. phNPCs were plated in Matrigel-coated 10 cm dishes at $2.3x10^6$ cells/dish in 1x proliferation media. At 24 hrs post plating, cells were transduced with pTRIPZ-mir-4707-EGFP at 5 IU/cell. At 48 hrs post transduction, puromycin (FisherScientific A1113803) was added to the culture media at a final concentration of 400 ng/mL to select for cells expressing the viral construct. A control plate of phNPCs that were non-transduced were also fed with puromycin containing media to ensure non-transduced cells would not survive. After one week in puromycin containing media, all non-transduced cells had died. Stable lines were expanded in proliferation media (without puromycin) until ready to be plated for the differentiation experiments.

Stable lines of phNPCs containing pTRIPZ-mir-4707-EGFP were plated at $4.5x10^5$ and $1.5x10^4$ cells/well in Matrigel-coated 6-well and 96-well plates respectively in 1x differentiation media: Neurobasal A (Life Technologies 10888-022) with 100 μg/ml primocin (Invivogen ant-pm-2), 2% B27 (Life Technologies 17504-044), 1x glutamax (Life Technologies 35050061), 10 ng/ml NT-3 (Life Technologies PHC7036) and 10 ng/ml BDNF (Life Technologies PHC7074). Every two or three days, half of the culture media was replaced with 2x differentiation media: Neurobasal A with 100 μg/ml primocin, 2% B27, 1x glutamax, 20 ng/ml NT-3 and 20 ng/ml BDNF. To turn on expression of mir-4707, doxycycline was added at 1 μg/ml on the day of plating to 1x differentiation media. Doxycycline was maintained by addition of 2 μg/ml in 2x differentiation media. At one and two weeks post differentiation, 6-well plates were used for RNA extraction, and 96-well plates were fixed for immunofluorescence and imaging.

RNA Extraction and qPCR for miRNA and mRNA Expression

RNA was extracted from 6-well plates using miRNeasy Mini Kits (QIAGEN 217004) with the inclusion of an on-column DNase digestion (QIAGEN 79254). Following elution with RNase-Free water, RNA concentration and quality was assessed with a NanoDrop ND-1000 Spectrophotometer. For cDNA synthesis of mRNA, iScript cDNA Synthesis Kits (Bio-Rad 1708891) were used with an input of 200 ng of total RNA. For cDNA synthesis of miRNA, TaqMan Advanced miRNA cDNA Synthesis Kits (ThermoFisher Scientific A28007) were used with 10 ng of total RNA input. Delta cycle threshold values (ΔCt) were calculated using the housekeeping gene EIF4A2, and fold change was calculated using control samples.

To assay expression of mRNA, SsoAdvanced Universal SYBR Green Supermix (Bio-Rad 1725271) was used in 10 µL reactions on a 385-well plate. Reactions contained 2 µL of template cDNA (iScript reaction diluted 1:5 with water) and 500 nM of each forward and reverse primer (Supplementary Table 5). Primers were chosen from the PrimerBank database [118]. Reactions were placed in a QuantStudio 5 (Applied Biosystems) thermocycler and cycled for 40 cycles according to the SsoAdvanced protocol. To assay expression of miRNA, TaqMan Advanced miRNA Assays (ThermoFisher A25576) were used with probes against hsa-miR-361-5p as a housekeeping control and hsa-miR-4707-3p (Assay ID: 478056 and 479946). Reactions consisted of 2.5 µL of template cDNA (miRNA cDNA synthesis reaction diluted 1:10 with water), 0.5 µL of prob, and 5 µL of Taqman Fast Advanced Master Mix (ThermoFisher 4444557) in a 10 µL reaction. Reactions were placed in a QuantStudio 5 thermocycler and cycled for 45 cycles according to the TaqMan Advanced miRNA Assay protocol. Delta cycle threshold values were calculated using hsa-miR-361-5p expression as the housekeeping gene, and fold change was calculated using control samples with the above equations. Technical replicates were defined as independent wells on a 6-well plate. Technical replicates ranged from three to six; see figure legends for the number of technical replicates in each experiment. Significant differences were defined as a p-value < 0.05 from a two-sided t-test.

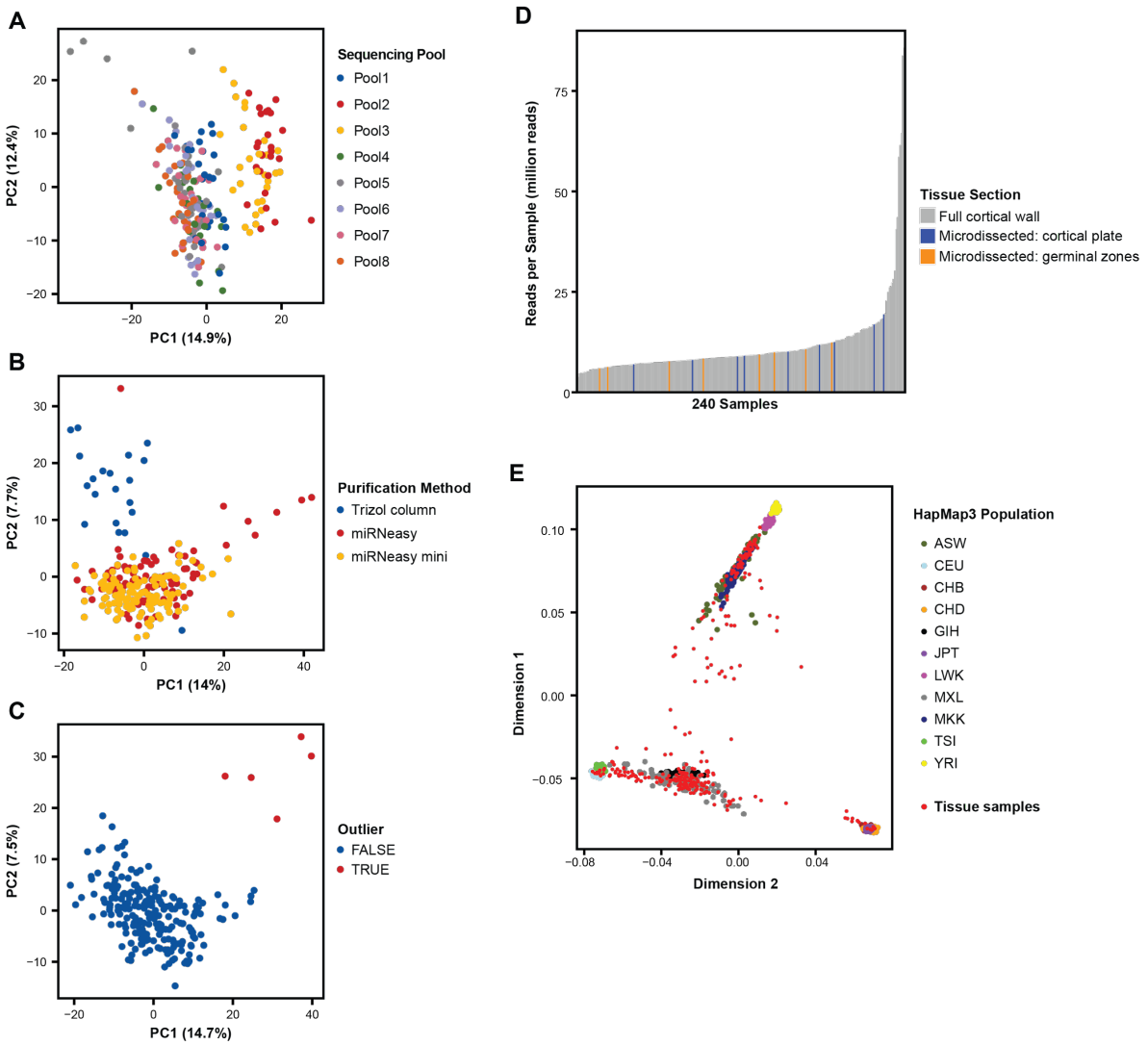
EdU Assay, Immunofluorescence Labeling and Imaging

To measure proliferating phNPCs we measured incorporation of 5-ethynyl-2'-deoxyuridine (EdU) into newly synthesized DNA using a Click-iT EdU Imaging Kit with Alexa Fluor 647 (Invitrogen C10340). At 8 days post transduction, phNPCs in 1x proliferation media in 96-well plates were labeled with the addition of 10 μM EdU at 37 °C for 2 hours. After the 2 hour incubation, cells were immediately fixed with 3.7% Paraformaldehyde (PFA) in PBS (FisherScientific 50-980-487) for 15 minutes. Fixed cells were then permeabilized with 0.5% Triton X-100 (Sigma-Aldrich T8787) for 20 minutes. Staining with Alexa Fluor 647 was conducted using the Click-iT EdU Imaging Kit protocol. phNPCs that were antibody labeled but not labeled with EdU (96-well differentiation experiments) were fixed and permeabilized with the above procedure.

Fixed and permeabilized phNPCs in 96-well plates (with or without EdU labeling) were blocked with 10% goat serum (MP Biomedicals 0219135680) in PBST: 0.02% Tween-20 (FisherScientific BP337500) in 1x PBS, for 1 hr at room temperature. Cells were labeled with primary antibody in 3% goat serum/PBST at 4 °C overnight using 1:500 chicken-anti-GFP (FisherScientific AB16901) and 1:1000 mouse-anti-Tuj1/TUBB3 (BioLegend 801202) to label neurons or 1:500 rabbit-anti-GFAP (ThermoFisher PA1-10019) to label progenitors. After overnight incubation, plates were washed three times for 5 minutes each with PBST. Secondary antibody labeling was done at 1:1000 dilution in 3% goat serum/PBST at room temperature for 1 hr using goat-anti-chicken-AF488 (ThermoFisher A11011) and goat-anti-mouse-AF647 (ThermoFisher A21236) or goat-anti-rabbit-AF647 (ThermoFisher A21245). After 1 hr incubation, a 1:1000 dilution of DAPI (ThermoFisher 62248) in PBST was added to the secondary antibody solution for 10 min at room temperature. Plates were then washed three times for 5 minutes each with PBST. Plates were stored at 4°C in 0.04% sodium azide (Sigma-Aldrich S2002) in PBS until imaging.

Plates were imaged using a Nikon Eclipse Ti2 microscope set up for high content image acquisition. Each well was imaged with 4 non-overlapping fields of view using 10x magnification and a 0.3 numerical aperture using 3 filter sets: DAPI (ex: 325-375, em: 435-485), GFP (ex: 450-490, em: 500-

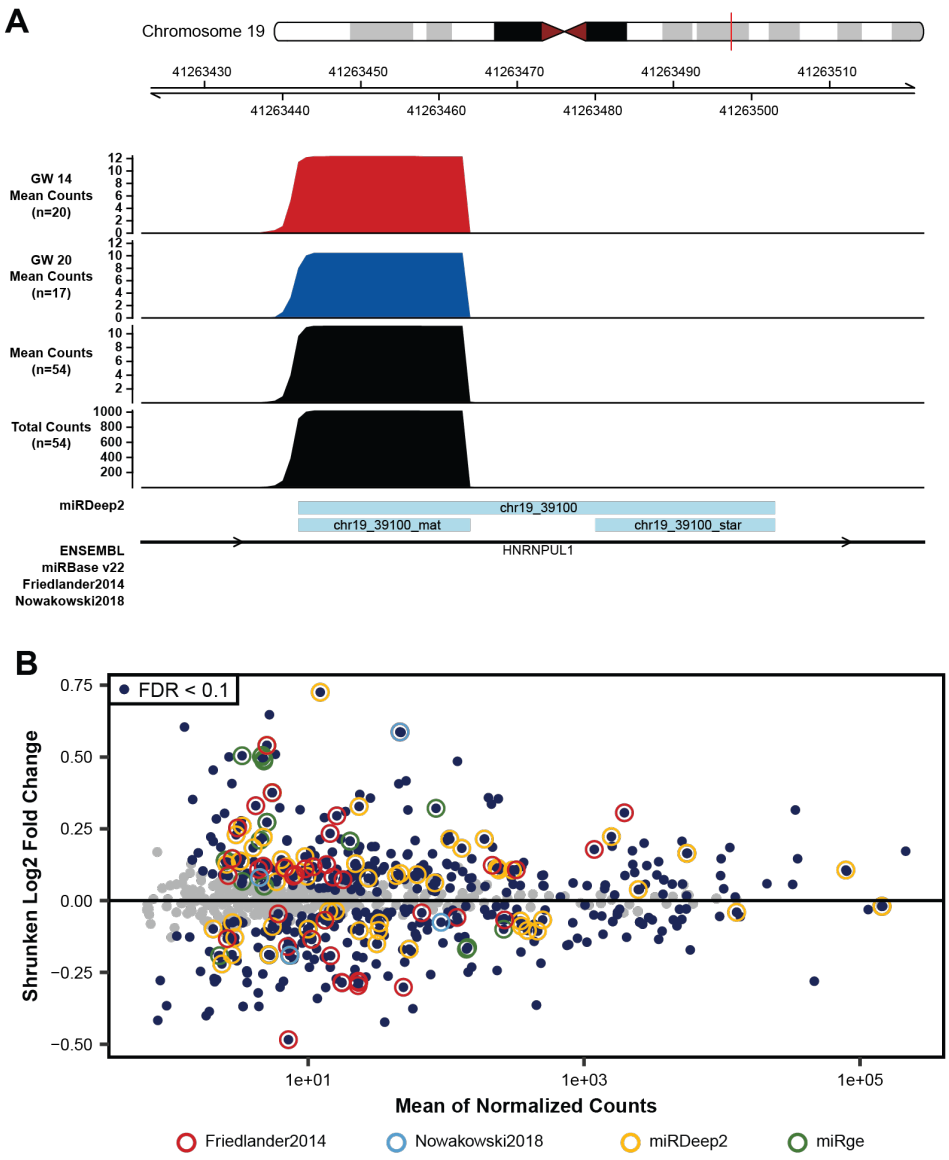
550), and AF647 (ex: 625-655, em: 665-715). Image sets were fed through an automated CellProfiler [119] pipeline to quantify the number of nuclei, GFP positive, and AF647 positive cells in each image. Technical replicates were defined as independent wells of a 96-well plate. Replicates ranged from 14 to 48 wells per condition; see figure legend for the number of technical replicates in each experiment. Significant differences were defined as a p-value < 0.05 from a two-sided t-test.



Extended Data Figure 2.1. Tissue Sample miRNA Expression and Sample Ancestry

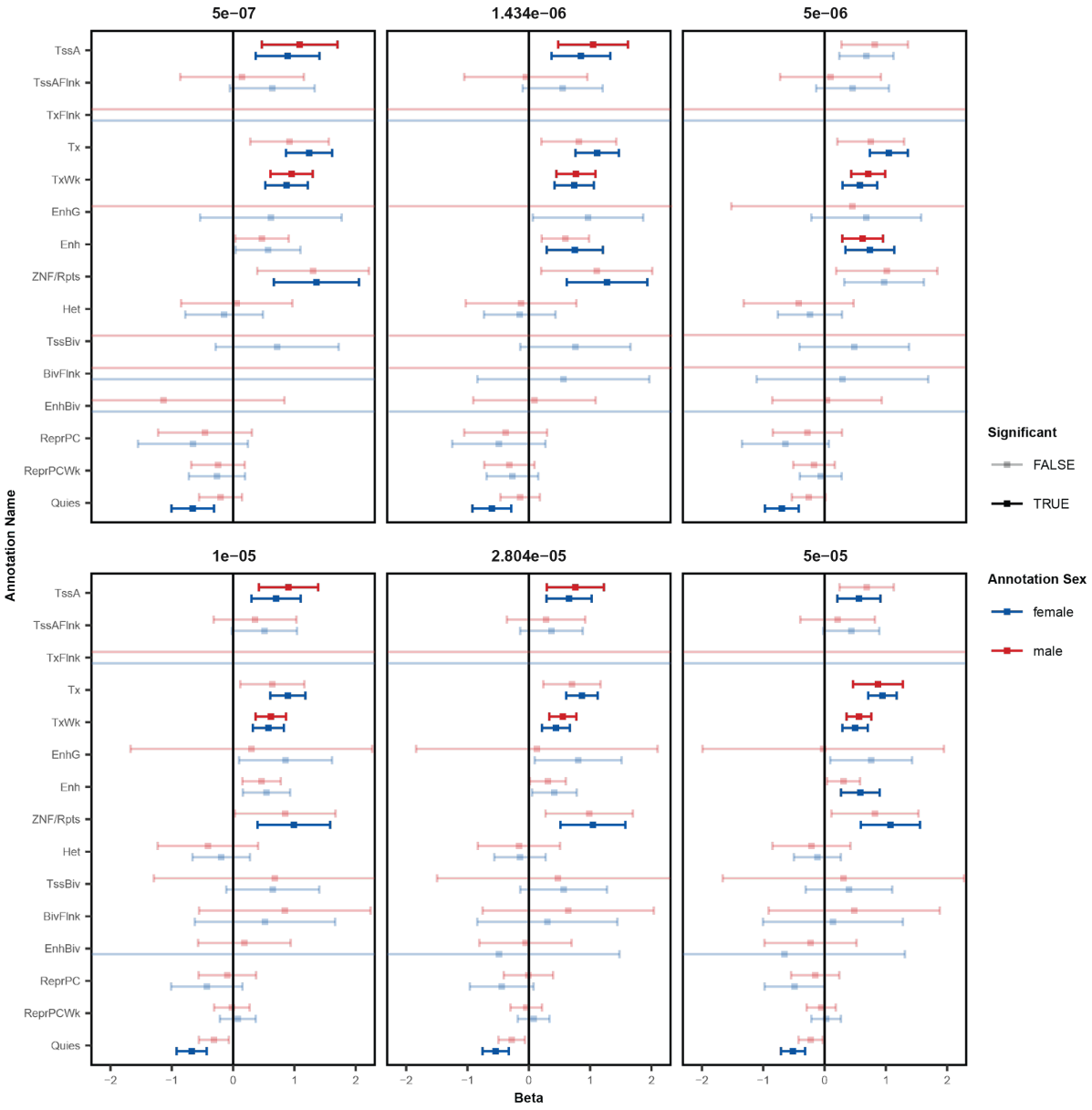
A. Principal component analysis (PCA) for uncorrected miRBase miRNA expression across the 217 quality controlled tissue samples shows sequencing-pool influences global miRNA expression and the primary driver of variation in PC1.
B. After correcting for sequencing-pool on miRNA expression using a linear model, PCA shows further influence by RNA purification method.
C. After correcting for sequencing-pool and purification method using a linear model, five samples remain as outliers in PCA and were removed from the study.
D. Small-RNA sequencing read depth across the 223 mid-gestation cortical wall samples used to discover novel miRNAs. Small-RNA sequencing data from 17 microdissected cortical plate and germinal zone samples were also included when detecting novel miRNAs.
E. Sample genotypes show diverse ancestry when overlaid on HapMap3 population samples. African ancestry in Southwest USA (ASW), Utah residents with Northern and Western European (CEU), Han Chinese in Beijing, China (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Gujarati

Indians in Houston, Texas (GIH), Japanese in Tokyo, Japan (JPT), Luhya in Webuye, Kenya (LWK), Mexican ancestry in Los Angeles, California (MXL), Maasai in Kinyawa, Kenya (MKK), Toscani in Italia (TSI), Yoruba in Ibadan, Nigeria (YRI)

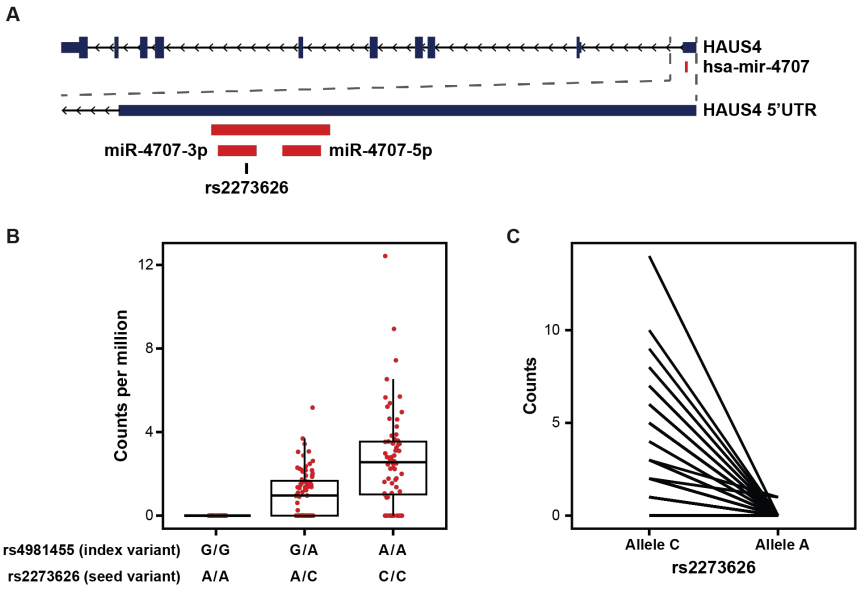


Extended Data Figure 2.2. Evidence for Novel miRNAs

A. Small-RNA sequencing read coverage for a putatively novel miRNA, detected using the miRDeep2 package, on chromosome 19 shows the characteristic read pile-up often seen from known miRNAs (Figure 2.1D and 2.1F).
B. Differential miRNA-expression analysis, as shown in Figure 2.1C, with miRNAs highlighted from Friedländer et al, Nowakowski et al, or novel miRNAs discovered in this study by either miRDeep2 or miRge2.0 packages.



Extended Data Figure 2.3. Enrichment of Local-miRNA-eQTLs within Functional Chromatin. Chromatin states as in Figure 2.2B at multiple miRNA-eQTL p-value significance thresholds. The stringent p-value significance threshold of 1.434×10^{-6} and the relaxed p-value threshold of 2.804×10^{-5} show similar enrichments. Abbreviations: active transcription start sites (TssA), chromatin associated with strong transcription (Tx), chromatin associated with weak transcription (TxWk), enhancers (Enh), ZNF genes and repeats (ZNF/Rpts), quiescent chromatin (Quies), flanking active tss (TssAFlnk), transcription at gene 5' and 3' (TxFlnk), genic enhancers (EnhG), heterochromatin (Het), bivalent/poised TSS (TssBiv), flanking TssBiv (BivFlnk), bivalent enhancer (EnhBiv), repressed polycomb (ReprPC), and weak repressed polycomb (ReprPCWk).



Extended Data Figure 2.4. MiR-4707-3p Alternate Allele Expression

A. The primary sequence for hsa-mir-4707 is located within the 5' UTR of the HAUS4 gene on chromosome 14. MiR-4707-3p contains a variant, rs2273626 (reference allele C, alternate allele A), within the seed sequence in high LD with the index variant, rs4981455 ($r^2 > 0.99$).
B. Sequencing read counts per million, as quantified by miRge 2.0, across the three genotypes for both the index variant and seed variant. All samples with index variant G/G showed zero reads.
C. Allele specific expression for heterozygous samples at the seed variant, rs2273626. A total of three reads across three independent donors were observed with the A allele. P-value=3.23×10^{-14} when using a paired, two-sided t-test.

CHAPTER 3: MICRORNAS ASSOCIATED WITH NEUROGENESIS IN THE DEVELOPING HUMAN NEOCORTEX

3.1 Introduction

The complex gene expression patterns that govern human neocortical development are being defined with increasing clarity [120–122]. However, the gene regulatory mechanisms that control these transcriptional programs are relatively understudied [123,124]. Progress has been made to define non-coding regions of the genome that may play a role in neocortical development by looking at open chromatin (through ATAC-seq) or 3D maps of chromatin contacts (via Hi-C) [27,125]. Yet, little work has been done in describing the post-transcriptional regulation of genes during neurogenesis and neocortical development. MicroRNAs (miRNAs) have broad effects on gene regulation, cell fate decisions, and transcriptional programs via post-transcriptional regulation [126,127]. Therefore, to fully understand the gene expression patterns that govern neocortical development, we must better understand the dynamics of miRNA expression during neurogenesis.

MiRNAs have been shown to play a role in progenitor proliferation, neuronal differentiation, and cortical wall development in mouse models [30,128]. However, relatively few studies have focused on the role of miRNAs during human neocortical development. In the most comprehensive study to date, miRNA expression in human fetal cortical tissue from nine samples revealed *in vivo* the miRNAs involved in human neurogenesis [33]. Here, we use small-RNA sequencing to quantify miRNA expression within developing human neocortex from 212 samples between 14 and 21 gestation weeks, allowing us to probe even further the extent to which miRNAs guide human neurogenesis. Furthermore, cortical wall tissue from three donors was divided into cortical plate (CP) and germinal zone (GZ) samples (see Figure 1.2). By comparing the differences between the progenitor rich GZ and neuronal rich CP tissues, we were able to resolve, with increased clarity, the expression patterns of miRNAs during

human *in vivo* neurogenesis. Combined with total-RNA sequencing, to capture mRNA and long non-coding RNA (lncRNA), we quantified the expression patterns of the mRNA targets of neurogenesis associated miRNAs. The analysis presented here represents the most thorough investigation to date of the miRNAs involved in human neocortical development.

The differences in mRNA expression between these GZ and CP samples have been previously shown to be primarily reflecting differences between neural progenitor cells of the germinal zones and newly born neurons in the cortical plate [27]. Therefore, we hypothesized that the miRNA expression differences between GZ and CP samples may also reflect human neurogenesis during neocortical development. It has also been seen that expression patterns within developing neocortical tissue are highly dynamic between early and late gestation weeks [42,93,121]. These expression differences may represent the complex transcriptional programs necessary to achieve the layer specific cell types of the fully developed neocortex [129]. We further hypothesized that miRNA expression differences of full cortical wall samples at early gestation weeks vs late gestation week samples may reflect these maturity differences between early born neurons and late born neurons. Here we quantify the miRNA expression differences associated with both neurogenesis and maturity and attempt to prioritize discovery of the miRNAs which may play a key role in regulating mRNA transcriptional programs necessary for human neocortical development.

3.2 Results

Gene Expression Profiling

Using small-RNA sequencing, we profiled miRNA expression in fetal cortical tissue from three donors, ranging from 17 to 19 gestation weeks. The cortical wall tissue was micro-dissected into two major anatomical divisions: the germinal zone (GZ), consisting of the ventricular zone, subventricular zones, and intermediate zone and the cortical plate (CP), consisting of subplate, cortical plate, and marginal zone (see Figure 1.2). Total-RNA sequencing was also used on these samples to profile

expression of mRNA and long non-coding RNA (lncRNA) within these same tissue samples. To determine if our miRNA and mRNA data reflected a biological difference between GZ and CP tissue, we performed a hierarchical clustering analysis (Figure 3.1A and 3.1B). Donor replicate tissue primarily clustered by donor but only after first clustering based on tissue type (GZ vs CP). This clustering pattern was seen for both miRNA and mRNA expression across these samples showing that expression differences between GZ and CP tissue will yield biologically significant information and not simply differences between donors.

In addition to the three donors which yielded micro-dissected tissue, we profiled miRNA and mRNA expression across 212 fetal cortical tissue samples from donors ranging from 14 to 21 gestation weeks. The miRNA expression across these samples was the basis for the miRNA-eQTL analysis in Chapter 2. Following correction for known batch effects, a principal component analysis (PCA) on both miRNA expression and mRNA expression revealed the primary variation between these samples was driven by donor gestation week (Figure 3.1C and 3.1D). Similar to the expression differences found between GZ and CP tissues, miRNA expression patterns across the various gestation week samples may yield biologically insightful information about neurogenesis and maturity associated miRNAs.

Differential Expression Analysis

To further probe the differences between GZ and CP tissue samples, we performed a differential expression analysis to identify the genes driving the biological differences between these samples (Figure 3.2A). We found 4,565 genes (mRNA and long non-coding RNA) with significantly higher expression in GZ relative to CP tissue. Conversely, 5,014 genes were significantly higher expressed in CP relative to GZ tissue. Genes known to be upregulated in neuronal progenitor cells, HES1, EOMES, and SOX2, are upregulated in GZ tissue. Genes known to be upregulated in neurons relative to progenitor cells, CALB1, HOMER1, STMN2, are upregulated in CP tissue. A previously published study, using these same tissue samples, revealed the gene expression differences between GZ and CP recapitulated the in vivo transition between ventricular zone and cortical plate tissue [27,121]. Furthermore, gene ontology analysis on genes

upregulated in GZ relative to CP reveal proliferative and mitotic cell cycle ontologies (Figure 3.2B). Analogously, genes upregulated in CP relative to GZ belong to ontologies associated with synaptic transmission and neuronal development. From this analysis, it is clear the difference in gene expression between GZ and CP tissue primarily reflects neurogenesis, or the cell type differences between neuronal progenitor cells and newly born neurons.

We next identified the miRNAs which may also be playing a role in neurogenesis. We found 134 and 85 miRNAs significantly upregulated in GZ and CP, respectively (Figure 3.2C). In accordance with the known role miR-92b plays in neural progenitor proliferation, miR-92b is upregulated in GZ tissue relative to CP tissue. Likewise, miR-124 is known to be upregulated in neurons relative to progenitors, and it is significantly upregulated in CP tissue relative to GZ tissue. Gene ontology analyses for groups of miRNAs to describe cellular processes have not yet been developed [130,131]. However, with the expression patterns of miR92b and miR-124, we can conclude differences in miRNA expression between GZ and CP tissue at least partly reflect changes associated with neurogenesis.

Using mRNA and miRNA expression differences across our 212 full cortical wall samples, we performed an analogous differential expression analysis to identify the genes differentially expressed in early gestation week samples as compared to late gestation week samples. The miRNA differential expression analysis is described in Chapter 2 and can be seen in Figure 2.1C. The differential expression analysis on mRNA and long non-coding RNA is described in the following section.

Neurogenesis and Maturity Associated miRNAs

Similar to the expression differences between GZ and CP tissue, early gestation week samples are associated with a greater proportion of progenitor cells while late gestation week samples are associated with a greater proportion of neurons [42]. However, it is possible our differential expression analysis also captures expression differences between early born neurons (and the younger progenitors that created them) and late born neurons (and the older progenitors that created them). These differences in expression would be associated with cell maturity and may support cell fate decisions associated with lower layer

(early born) and upper layer (late born) neurons [122,132]. To investigate whether the miRNA and mRNA expression patterns in our samples reflect these dynamic changes associated with neurogenesis and cellular maturity, we combined the GZ/CP differential expression analysis with the gestation week differential expression analysis.

First, we defined the set of genes that were significantly differentially expressed between GZ and CP tissue as "neurogenesis" associated. Genes that were significantly differentially expressed between early and late gestation week tissue, but were not previously defined as neurogenesis associated in the GZ/CP analysis, were defined as "maturation" associated. Plotting the fold change of both the GZ/CP analysis and the gestation week analysis revealed four sets of genes along two primary axes (Figure 3.3). Genes upregulated in the progenitor rich GZ tissue were labeled, "GZ>CP." Genes upregulated in the neuron rich CP tissue were labeled, "CP>GZ." Genes upregulated in early gestation week tissue, but not associated with differences in GZ and CP tissue, were labeled, "Early GW." Finally, genes upregulated in late gestation week tissue, but not associated with differences in GZ and CP tissue, were labeled, "Late GW." This categorization was repeated for miRNA differential expression and mRNA differential expression analyses. Within the mRNA data, markers of lower layer neurons (early born neurons), CTIP2 and FOXP2, are upregulated in early gestation week samples and fall in the Early GW group. Markers of upper layer neurons (late born neurons), CUX2 and BRN2, are upregulated in late gestation week samples and fall in the Late GW category. These examples in the mRNA differential expression categories show this analysis may reveal mRNA and miRNA not previously associated with neuronal maturity.

MicroRNA Target Enrichments

MiRNAs canonically regulate mRNA expression through targeted binding of the RISC complex guided by mRNA sequence homology to the miRNA "seed" sequence [3,4]. The binding of the RISC results in decreased mRNA expression of miRNA targets. We hypothesized the effect of an upregulated miRNA would be downregulation of mRNA targets and would be shown as upregulated in the opposite differential expression category. For example, miR-124 is categorized CP>GZ because it is upregulated in

CP tissue relative to GZ tissue. The targets of miR-124 would be downregulated in CP tissue, upregulated in GZ tissue, and categorized in our analysis as GZ>CP. miRNAs that play a significant role in either neurogenesis or maturity would have significant effects on expression across many targeted mRNAs. We sought to identify these miRNAs by looking for an enrichment of mRNA targets among the mRNAs that are differentially expressed.

To identify the mRNA targets of the differentially expressed miRNAs, we used miRNAtap to query five target prediction databases: DIANA, Miranda, PicTar, TargetScan, and miRDB [133–138]. mRNAs were considered targets of a given miRNA if at least two of the five prediction algorithms predicted a miRNA-mRNA interaction. For most miRNAs, hundreds to thousands of mRNAs are predicted to be targeted. For a miRNA categorized as GZ>CP, we would expect its targets to be enriched among the mRNAs categorized as CP>GZ. Furthermore, the mRNA targets of a miRNA categorized as GZ>CP would also be depleted among the mRNAs categorized as GZ>CP.

We performed this enrichment analysis across the four miRNA differential expression categories within the set of mRNAs in the corresponding neurogenesis or maturity differential expression categories (Figure 3.4). As expected, neurogenesis associated miRNAs upregulate in GZ (GZ>CP) have an enrichment of targets among mRNA upregulated in CP (CP>GZ) (Figure 3.4B). Also as expected, neurogenesis associated miRNAs upregulated in CP (CP>GZ) have an enrichment of targets among mRNA upregulated in GZ (GZ>CP) (Figure 3.4C). Not expected were the enrichments seen in Figure 3.4A and Figure 3.4D. We expected there to be a depletion of mRNA targets within the same differential expression category as any given miRNA.

Enrichment analyses within the maturity associated miRNA categories yielded few significant enrichments or depletions (Figure 3.4E-H). However, two maturity associated miRNAs upregulated in early gestation week tissue showed significant enrichment of targeted mRNAs within late gestation week tissue (Figure 3.4F). miR-106b-5p and miR-19a-3p have not previously been associated with neurogenesis or neuronal maturation, however their targeted mRNAs are enriched among mRNAs upregulated in late gestation week cortical tissue. Gene ontology analysis on the top 200 mRNA targets

upregulated in late gestation week tissue hint at nervous system development, transmembrane ion transport, and signal transduction which are necessary cellular processes associated with functional neurons in the developing cerebral cortex (Figure 3.5). This analysis shows how mRNA expression data, in conjunction with miRNA expression data, can prioritize a set of miRNAs involved in neurogenesis or neuronal maturity for follow up validation experiments.

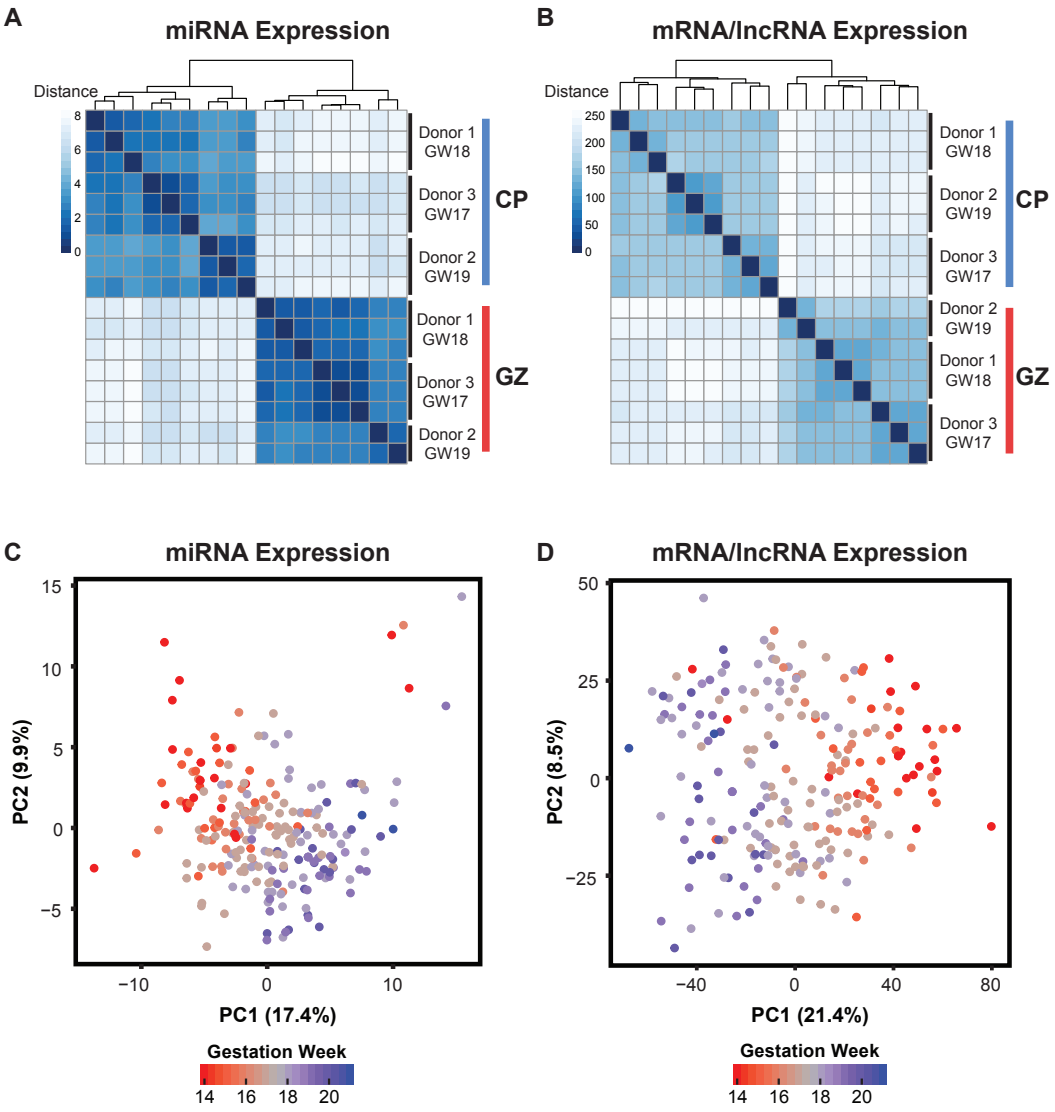


Figure 3.1. Expression Profiles for Developing Cortical Tissue Samples

A. Tissue samples from cortical plate (CP) and germinal zone (GZ) from 3 donors between gestation weeks (GW) 17-19 was used to profile miRNA expression. Hierarchical clustering based on Euclidean distance shows donor samples cluster together only after clustering based on tissue type (GZ or CP).
B. As in A, but using total-RNA sequencing to quantify mRNA and long non-coding RNA (lncRNA) expression.
C. Principal component analysis (PCA) on miRNA expression data across 212 tissue samples ranging from 14 to 21 gestation weeks. After correcting for known technical covariates, the first principal component (PC) is strongly associated with gestation week.
D. As in C, but using mRNA and long non-coding RNA expression.

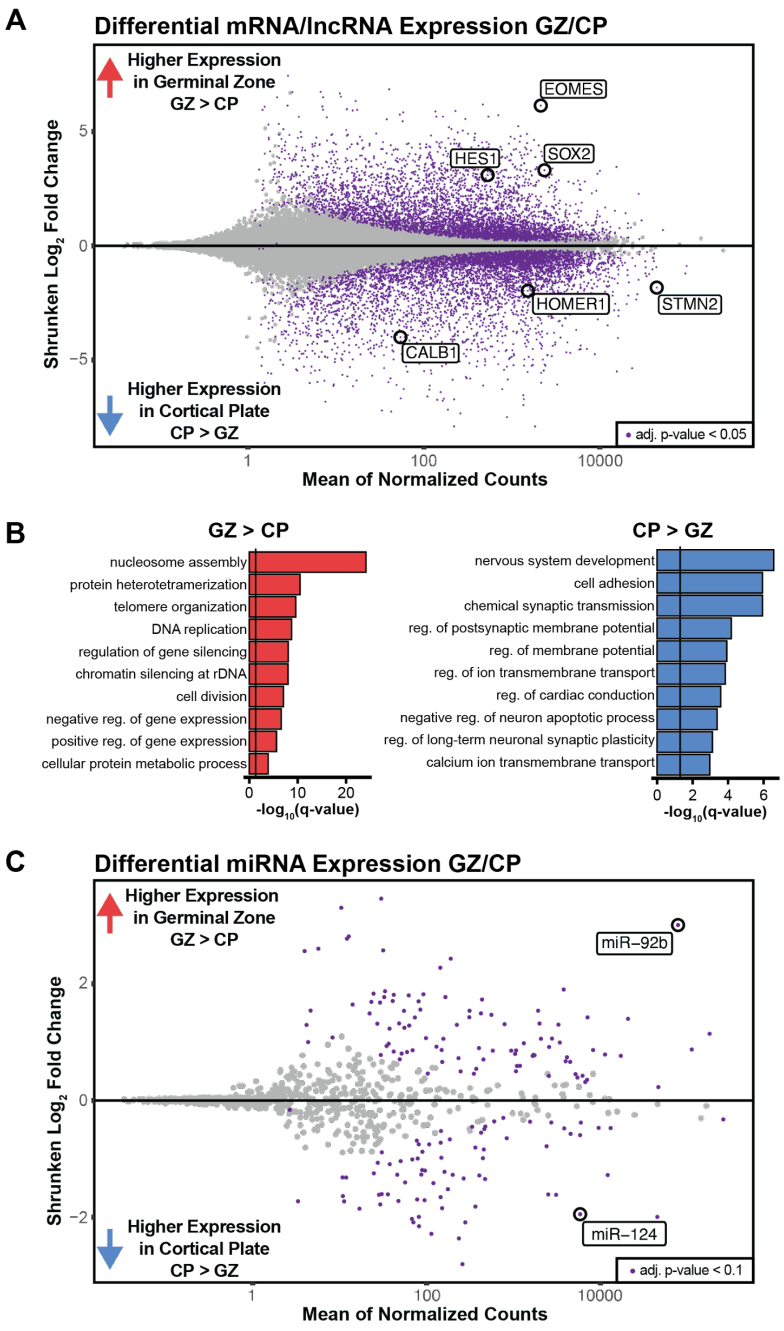


Figure 3.2. Differential Expression Analysis

A. Differential expression analysis between nine cortical plate (CP) samples and 8 germinal zone (GZ) samples for mRNA and long non-coding RNA. Genes significantly differentially expressed (FDR corrected p-value < 0.05) are shown in purple. Genes with known roles in nural progenitor proliferation, HES1, EOMES, and SOX2 are upregulated in GZ tissue. Genes with known roles in neurons, CALB1, HOMER1, and STMN2 are upregulated in CP tissue.
B. Gene ontology (GO) analysis using the top 200 upregulated genes in the GZ (GZ>CP) or the top 200 upregulated genes in the CP (CP>GZ). Showing the top 10 significantly enriched biological

process GO terms. Q-values are Bonferroni-corrected p-values, and the vertical line represents a corrected p-value of 0.05.

C. Differential expression analysis as in A but using miRNA expression. MiRNAs are declared significantly differentially expressed if their FDR corrected p-value is less than 0.1.

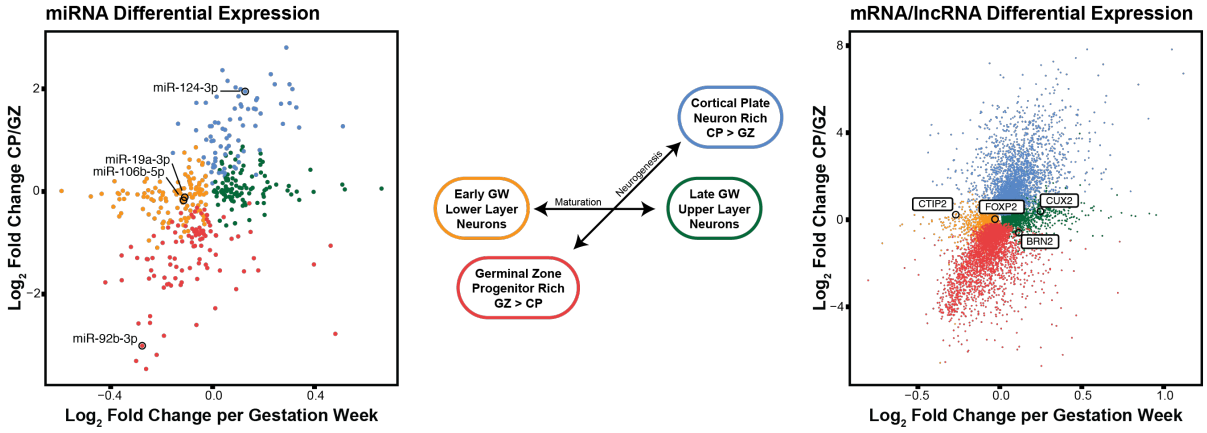


Figure 3.3. Neurogenesis and Maturity Associated Genes. Log2 fold-change from the differential expression analysis using CP and GZ tissue on the y-axis. Log2 fold-change from the differential expression analysis using gestation week as the comparison variable on the x-axis. Left graph shows miRNAs that were significantly differentially expressed in one or both analyses. Right graph shows mRNA and long non-coding RNA that were significantly differentially expressed in one or both analyses. Genes are colored based on neurogenesis categories "GZ>CP" or "CP>GZ" or by maturation categories "Early GW" or "Late GW."

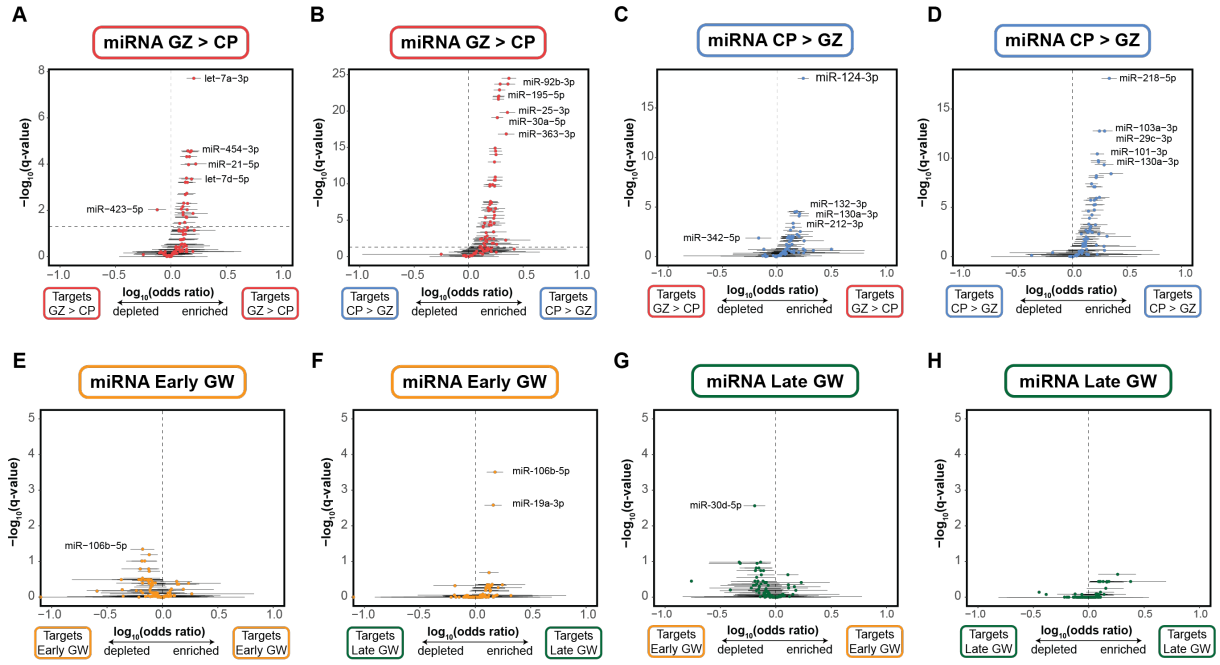


Figure 3.4. MiRNA Target Enrichment

A. The targets of miRNAs classified as GZ>CP were tested for enrichment among the mRNA/lncRNA upregulated in GZ tissue using as background the total set of expressed mRNA/lncRNA. Positive Log10(odds ratios) represent an enrichment while negative values represent depletion. Q-values are FDR corrected p-values and horizontal dotted line represents an FDR < 5%. Error bars are the 95% confidence interval from the Fisher's Exact enrichment test.
B. The targets of miRNAs classified as GZ>CP were tested for enrichment among the mRNA/lncRNA upregulated in CP tissue.
C. The targets of miRNAs classified as CP>GZ were tested for enrichment among the mRNA/lncRNA upregulated in GZ tissue.
D. The targets of miRNAs classified as CP>GZ were tested for enrichment among the mRNA/lncRNA upregulated in CP tissue.
E. The targets of miRNAs classified as Early GW were tested for enrichment among the mRNA/lncRNA upregulated in Early GW tissue.
F. The targets of miRNAs classified as Early GW were tested for enrichment among the mRNA/lncRNA upregulated in Late GW tissue.
G. The targets of miRNAs classified as Late GW were tested for enrichment among the mRNA/lncRNA upregulated in Early GW tissue.
H. The targets of miRNAs classified as Late GW were tested for enrichment among the mRNA/lncRNA upregulated in Late GW tissue.

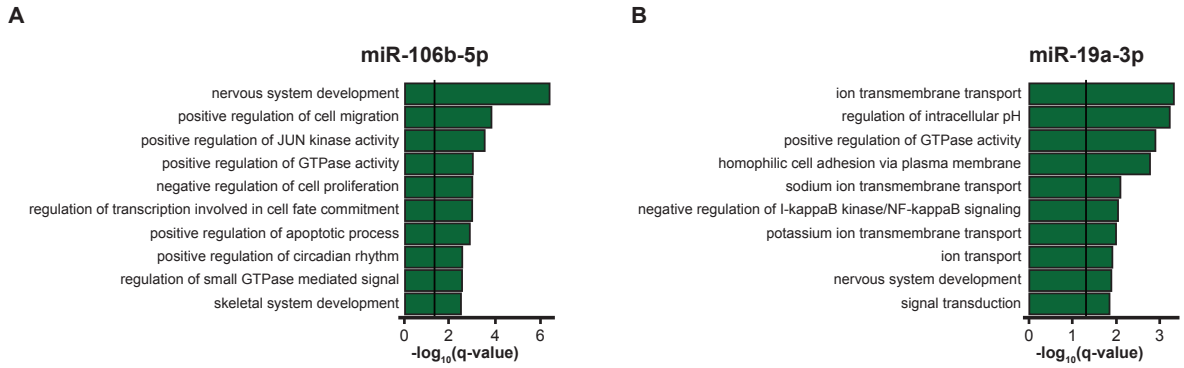


Figure 3.5. Gene Ontology Analysis on miRNA Targets

A. Gene ontology analysis against biological process GO terms for the top 200 genes upregulated in late gestation week tissue that are predicted targets of miR-106b-5p. The top 10 significantly enriched GO terms are shown. Q-values are Bonferroni-corrected p-values.
B. As in A, but for the predicted targets of miR-19a-3p.

3.3 Discussion

In this study, we utilized small-RNA sequencing to quantify miRNA expression in developing neocortical wall tissue samples between 14 and 21 gestation weeks; a tissue and timepoint associated with neural progenitor proliferation, differentiation, and neuronal migration [22]. By specifically profiling miRNA expression in micro-dissected germinal zone, rich in neural progenitors, and cortical plate, rich in neurons, we were able to define an expanded set of miRNAs associated with in vivo human neurogenesis than previously described [33,139]. Furthermore, investigating the miRNA differential expression patterns within full cortical wall tissues from early gestation week donors as compared to late gestation week donors, we were able to identify miRNAs that may play a role in neuronal maturation and neuron layer specificity.

The differential expression analysis between GZ and CP tissues and across early versus late gestation week tissues revealed miRNAs not previously associated with human neurogenesis. We also discovered putatively novel miRNAs differentially expressed between GZ and CP tissue and across gestation weeks. In Chapter 2, we reported on the novel miRNAs differentially expressed between early and late gestation week tissue. Here, we found 26 novel miRNAs that are differentially expressed between GZ and CP tissue. However, these novel miRNAs were not included in the results of this chapter because the miRNA target prediction databases we used did not allow for novel target prediction. In future analyses, we will seek to include these novel miRNAs by first computationally predicting their mRNA targets. The novel miRNAs discovered here may be specific to developing neocortical tissue and may prove significant to understanding neural progenitor proliferation and differentiation.

Despite nearly two decades of development on miRNA target prediction algorithms, putative mRNA target lists often contain many false positive or context-specific interactions [140,141]. To mitigate the false positives, we used five target prediction databases and declared mRNA targets when two or more algorithms predicted an interaction. Despite this strategy, the miRNAs in our analysis have hundreds to thousands of predicted mRNA targets. The extent to which false positive interactions are present in these predictions has not been determined and this may contribute to false positives or false

negatives in our enrichment analysis. We also attempted to use databases of known or experimentally validated miRNA-mRNA interactions in our enrichment analyses [142]. However, these databases are heavily populated with interactions of better studied miRNAs, and many of the neurogenesis associated miRNAs discovered here had few or no mRNA targets. Future analyses may improve on our approach by combining predictive tools with experimental interaction evidence before looking for enrichment of differentially expressed mRNA targets.

Identifying miRNAs associated with human neurogenesis will lead to a more complete understanding of the transcriptional programs which influence proliferative vs neurogenic fate decisions. Additionally, *in vitro* neuronal trans-differentiation methods are notoriously slow and require weeks to months to achieve neurons that transcriptionally resemble early born neurons [143,144]. The discovery of miRNAs that guide progenitor proliferation into later stage neurons or which quicken differentiation may lead to the development of higher fidelity *in vitro* differentiation methods. Since the exogenous manipulation of tens to hundreds of independent mRNAs within an in vitro system is prohibitive, modulating a small number of miRNAs may prove to be more effective in changing expression patterns across hundreds of neurogenesis associated genes. Here, we have prioritized neurogenesis and maturity associated miRNAs which have predicted mRNA targets that are also associated with neurogenesis and neuronal maturity. These prioritized miRNAs are ideal candidates for experimental manipulation within *in vitro* neural progenitor cell cultures to validate their effects on neurogenesis.

3.4 Methods

Tissue Procurement

The same tissue samples used in Chapter 2 were used in the Chapter 3 analyses. Human prenatal cortical tissues samples were obtained from the UCLA Gene and Cell Therapy Core following IRB regulations for 223 genetically distinct donors (96 females, 127 males, 14-21 gestation weeks) following voluntary termination of pregnancy. Cortical tissue samples from an additional 3 donors were micro-dissected into germinal zone and cortical plate sections as previously described yielding 17 more tissue samples. Tissue samples were flash frozen after collection and stored at -80C. These tissue samples overlapped with those used in a previous mRNA-eQTL study and gene regulatory studies [27,42,125].

Library Preparation and Sequencing

Total RNA was extracted using miRNeasy-mini (QIAGEN 217004) kits or was extracted using trizol with glycogen followed by column purification. Library preparation for small-RNA was conducted using TruSeq Small RNA Library Prep Kits (Illumina RS-200). RNA libraries were randomized into eight pools and run across eight lanes of an Illumina HiSeq2500 sequencer at 50 base-pair, single-end reads to a mean sequencing depth of 11.7 million reads per sample. mRNA library preparation and sequencing was previously described [42]. Briefly, library preparation for total RNA was conducted via TruSeq Stranded RNA Library Prep Kits (Illumina 20020597) with Ribozero Gold ribosomal RNA depletion. Libraries were sequenced with 50 base-pair, paired-end reads to a mean read depth of 60 million reads per sample.

MicroRNA Expression Analysis

Small RNA-sequencing FASTQ files were used as input to the miRge 2.0 [63] annotation workflow to quantify expression of known miRNAs from miRBase release 22 (March 2018) [64]. Briefly, sequencing reads were first quality controlled, adaptors removed, and collapsed into unique reads. The reads were then annotated against libraries of mature miRNA, hairpin miRNA, tRNA,

snoRNA, rRNA, and mRNA. The miRge 2.0 workflow protects against reference mapping bias by incorporating common genetic variants into the mature miRNA sequence library and allowing zero mismatched bases on the first pass of read annotation. A second pass of unannotated reads allows for mismatched bases to identify isomiRs. Unmapped reads were then used as input to a second annotation pipeline to quantify expression of recently discovered novel miRNAs from Friedländer et al [79] and Nowakowski et al [33]. Bowtie v1.2.2 [103] was used to map reads to the UCSC hg38 reference genome using the following options: -v 2 -5 1 -3 2 --norc -a --best -S --chunkmbs 512. Mapped reads were then counted using featureCounts [104] against a custom GTF file including the Friedländer and Nowakowski novel miRNAs using the following options: -s 0 -M -f -O.

We also defined an additional set of novel miRNAs discovered within our 240 sample dataset using miRge 2.0 and miRDeep2 [80] prediction pipelines (see Chapter 2 methods for details).

mRNA Expression Analysis

Total RNA-sequencing FASTQ files were first filtered and adapter trimmed using trim_galore and the following options: --length 20 --stringency 5. Filtered and trimmed reads were mapped to GRCh38 using STAR v2.5.4b [105]. Mapped reads were counted using featureCounts [105] against the Ensembl GRch38.p7 human gene annotations using the following options: -T 4 -p -t exon. Count data for each sample was combined into a count matrix for downstream analyses.

Sample Quality Control

See Chapter 2 for specifics on which samples were removed from the analysis. Of the 223 samples, two were declared female using the genotype data but male by XIST expression and were excluded from downstream analysis. We further sought to detect sample swaps or mixtures by evaluating the consistency of genotypes called via genotyping and those that can be detected by RNA-sequencing using VerifyBamID v1.1.3 [108]. We detected four samples that were mixtures or possible sample swaps

with their assigned genotype data (FREEMIX > 0.04 or CHIPMIX > 0.04). Finally, five samples were identified as outliers via PCA analysis after accounting for known technical confounders. A total of 212 cortical wall samples (93 females, 119 males, 14-21 gestation weeks) and 17 micro-dissected samples (1 male, 2 females, 17-19 gestation weeks) were used in the analysis.

PCA and Hierarchical Clustering

Principal component analysis (PCA) was performed using the prcomp() function within the stats package of the R software language [109]. Only miRBase release 22 expressed miRNAs with at least 10 counts across at least 10 samples were included when doing PCA. MiRNA expression was first normalized using the variance-stabilizing transformation (VST) function within DESeq2 [110]. To correct for known batch effects (sequencing-pool and RNA-purification method; see Chapter 2), we used the removeBatchEffect() function within the limma package on the VST transformed expression matrix [111]. While removing batch effects, we preserved the effect of gestation week using the design option within removeBatchEffect(). After known batch effects were removed, PCA was repeated to confirm removal of technical variation across samples and to identify expression outliers.

Differential Expression Analysis

Differential expression analysis for full cortical wall samples was conducted on the 212 samples which survived quality control filtering using DESeq2 [110]. Expression of all known and novel miRNAs, which survived the above expression threshold, were used in the analysis. Gestation week was used as the treatment variable while controlling for technical confounding variables: sequencing-pool, RNA-purification method, RNA-integrity number (RIN), and RNA concentration after extraction. Differentially expressed miRNAs with a Benjamini-Hochberg adjusted p-value < 0.1 were deemed significant (false discovery rate (FDR) < 10%). For visualization of differentially expressed miRNAs $\log_2$(fold change) values were shrunk using the apeglm method [112]. Differential expression analysis

using GZ and CP samples was controlled for sequencing-pool, RIN, RNA concentration, and donor while using the tissue section (GZ/CP) as the treatment variable.

miRNA Target Enrichment Analysis

We used miRNAtap to query five target prediction algorithms: DIANA, Miranda, PicTar, TargetScan, and miRDB[133–138]. mRNAs were considered targets of a given miRNA if two of the five prediction algorithms predicted a miRNA-mRNA interaction. Fisher's Exact used to test enrichment of miRNA target genes within the set of differentially expressed genes within the background set of all expressed genes. Positive $\log_{10}$(odds ratio) interpreted as enrichment, while negative $\log_{10}$(odds ratio) interpreted as depletion. Q-values calculated as FDR corrected p-values, and significant enrichment set at FDR < 5%. Error bars represent 95% confidence intervals.

CHAPTER 4: DISCUSSION

In my graduate work, I was able to profile the miRNA expression across hundreds of developing neocortical tissue samples and identify the genetic variants that are associated with miRNA expression. Additionally, I used two computational algorithms to identify putatively novel miRNAs that have not yet been catalogued in current miRNA databases. Furthermore, I compared the miRNA expression between germinal zone and cortical plate tissue to discover known and novel miRNAs involved in neurogenesis, and I compared the miRNA expression between early and late gestation week tissue to find miRNAs which may have roles in defining neuron layer specificity and maturity. This has important implications in furthering our understanding of how miRNAs guide and influence cell fate decisions, particularly the maintenance of neural progenitors and their differentiated progeny. Modern sequencing technology has made it possible to profile expression of small RNA without previous knowledge of which RNA species are present (as was needed for microarray assays for example). However, as we begin to answer questions about which miRNAs are expressed, which miRNAs have yet to be catalogued, and which genomic elements govern their expression, we uncover more questions about which transcriptional programs and molecular mechanisms are influenced by these miRNAs. The results presented here lead us to even more challenging questions and directions for future study.

4.1 Implications of Studying Human Genetics

The study of human genetics has a tenuous past that is rooted in the desire to "improve the inborn qualities of a race" as was described by Francis Galton in 1904 on the subject of eugenics [145]. It is tempting to think that our society no longer struggles with issues like eugenic sterilization laws or the forced sterilizations imparted by various eugenic boards in the second half of the 20th century [146,147].

However, if we are not careful, this new wave of genome-wide association studies and functional genomic annotation studies has the potential to start a new wave of eugenics under the guise of improving human health. Understanding the molecular mechanisms by which risk for heritable diseases is imparted through genetic variation has the potential to find treatments for common neuropsychiatric disorders that plague significant portions of our population [57,148]. But this information can also be easily used to look for "desirable" genetic markers that have small, but measurable improvements to human health. Should this information be accessible to parents of unborn babies? Should our society permit the design of babies with lower rates of mental health disorder? What about design of babies with greater intelligence?

GWAS for complex neuropsychiatric disorders and cognitive traits have unique challenges compared to studies looking for polygenic risk associated with adult onset phenotypes like cardiovascular disease or many types of cancer [149]. In diseases of the adult, knowing ones genetic risk can inform early or more frequent screening for disease [150]. Understanding the genetic mechanisms for polygenic risk can inform development of novel therapeutics with a deeper understanding of disease etiology [151]. Neuropsychiatric disorder GWAS may prove more challenging to implement treatments or screenings since these disorders often have origins during early fetal brain development [19,26,152–154]. The research presented here contributes in a small way to understanding how a genetic variant can have measured effects on brain size and educational attainment through changes in miRNA expression during development. As studies continue to find mechanisms by which genetic variants influence various cognitive traits, the temptation will increase to do something with that information.

The debate on where to draw the ethical line between reasonable and unreasonable interventions has just begun. From my point of view, I am hesitant to offer suggestions as to what we should permit as a society. However, as a student of genomics, I have experienced the difficulty in understanding these complicated issues. I realize that informing policy makers, healthcare providers, and parents how genetics and environmental factors both influence disease and disorder risk will be our greatest challenge. These important ethical questions will need to be faced as we struggle to understand the extent to which our

genetics predetermines our risk for disease or modulates our quality of life through cognitive traits, and I believe the first step in facing these issues is though thoughtful and creative education.

4.2 Limitations of miRNA-eQTLs

The advent of second-generation high throughput sequencing, it became possible to profile the expression of thousands of RNA species across many tissue types for relatively little cost. This has enabled the detection of thousands of eQTLs across an ever-growing number of tissues at various developmental timepoints [36,89]. However, RNA expression is only a proxy measurement of the translated proteins that actually modulate and influence cell fate decisions [155,156]. Furthermore, miRNA expression, as presented here, can modulate mRNA expression or protein translation across many expressed mRNAs [142,157]. The extent to which miRNA-eQTLs influence downstream target expression is difficult to measure with standard RNA-sequencing techniques.

MiRNA target prediction algorithms and databases of validated miRNA targets have multiple limitations [158,159]. Validated targets are biased to the miRNAs and mRNAs of particular interest to the researchers studying the interaction. Confounded with cell type specificity of mRNAs, target databases can be great positive evidence of miRNA-mRNA interaction but limited utility in determining which mRNAs are not targeted by a given miRNA [142,160]. Computational prediction algorithms for miRNA target identification have the promise to be unbiased. However, the six base-pair “seed” sequence of a mature miRNA giving the RISC complex mRNA target specificity, yields thousands of mRNA transcripts for which any given miRNA may bind. Algorithms vary in narrowing down these target lists by calculating RNA binding energies and predicting secondary structure in mRNA molecules [138,161]. The result is several prediction algorithms with widely different target prediction lists. These target prediction limitations make it difficult to determine if changes in mRNA expression, measured through RNA-sequencing, are direct results of miRNA expression and targeting.

4.3 Future Directions

The work presented here makes it clear that miRNAs are robustly expressed in developing cortical tissue during mid gestation. As shown in Chapters 2 and 3, global miRNA expression can segregate fetal cortical tissue samples by gestation week and cortical laminae of the developing neocortex. Global gene expression, mRNA and long non-coding RNA, also has the same ability to show biological differences between tissues samples at various gestation weeks and this variation is likely associated with neurogenesis [42,43]. The extent to which miRNA expression is also associated with neurogenesis is yet to be fully understood. A couple well studied miRNAs, miR-92b and miR-124, seem to show neurogenesis specific expression patterns, but the differential expression analysis shows many other miRNAs that may be influencing or guide neural progenitor proliferation and differentiation. This differential expression analysis serves as the basis for identifying known and novel miRNAs that have yet unknown roles in neurogenesis. The first step in validating these neurogenesis associated miRNAs would be repeating this differential expression analysis in an in vitro cell culture system with a homogenous population of neural progenitor cells (NPCs) and their differentiated progeny. Validation experiments could then be designed to see if modulation of neurogenesis associated miRNA expression influences the proliferation or differentiation of NPCs. The utility in discovering miRNAs which have specific roles during neurogenesis could be useful in designing better in vitro culture systems to model human neurogenesis.

Given the lack of validated miRNA targets, understanding the implications of the miRNA-eQTLs without disorder or trait colocalizations proves challenging. Do these miRNA-eQTLs have downstream effects on transcriptional programs important for neocortical development? Or are these miRNA-eQTLs simply noise across a genetic background with many common genetic variants? Cross-linked immunoprecipitation followed by next generation sequencing (CLIP-seq) techniques may be employed to better understand which mRNAs are targeted by miRNAs [162–164]. By immunoprecipitation of the AGO protein within the RISC silencing complex we could see which mRNAs are targets for degradation by miRNAs. Further advancements with CLASH-seq can identify specific miRNA-mRNA duplexes and

provide unparalleled evidence for targeting of a mRNA by a miRNA [165]. However, these assays have limitations. Optimization of the immunoprecipitation and RNA purification needs to be done in the desired cell type or tissue type and more starting material (a greater number of input cells) is needed than standard RNA purification methods. Furthermore, these assays are relatively new and untested, meaning there may be unforeseen bias in the types of miRNA-mRNA interactions that can be assayed. Despite these drawbacks, more evidence for miRNA-mRNA target interaction is needed to fully understand the effects of miRNA expression differences on cell fate decisions and transcriptional programs.

4.4 Summary

In this research, I was able to reveal robust miRNA expression of known and novel miRNAs across cortical tissue during mid-gestation, a stage and tissue which has not previously been captured in previous eQTL studies. Using this information, I was able to discover 85 miRNA-eQTLs in a tissue type and at a developmental timepoint with known influence on brain phenotypes and cognitive traits. Colocalization analysis yielded one robust colocalization between an eQTL for miR-4707-3p and variants with associations to multiple brain size phenotypes. Experimental over-expression of miR-4707-3p in primary human neural progenitor cells revealed a possible mechanism by which genetic variation associated with miRNA expression may modulate brain size and cognitive phenotypes during brain development. However, there are many questions that remain unanswered and significant challenges to understanding the effects of miRNA-eQTLs during neocortical development. I hope to take this valuable training forward into my career to continue to investigate the functional consequences of common genetic variation on human quantitative traits and diseases. As new approaches emerge to probe more specific cell-types, new developmental models, and more rare genetic variation, I seek to be on the forefront of functional genomics research.

REFERENCES

1. Hammond, S.M. (2015). An overview of microRNAs. Adv. Drug Deliv. Rev. *87*, 3–14.

2. O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of microrna biogenesis, mechanisms of actions, and circulation. Front Endocrinol (Lausanne) *9*, 402.

3. Neeb, Z.T., and Zahler, A.M. (2014). An expanding world of small RNAs. Dev. Cell *28*, 111–112.

4. Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet. *16*, 421–433.

5. Shenoy, A., and Blelloch, R.H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nat. Rev. Mol. Cell Biol. *15*, 565–576.

6. Ivey, K.N., and Srivastava, D. (2015). microRNAs as Developmental Regulators. Cold Spring Harb. Perspect. Biol. *7*, a008144.

7. Cursons, J., Pillman, K.A., Scheer, K.G., Gregory, P.A., Foroutan, M., Hediyeh-Zadeh, S., Toubia, J., Crampin, E.J., Goodall, G.J., Bracken, C.P., *et al.* (2018). Combinatorial Targeting by MicroRNAs Co-ordinates Post-transcriptional Control of EMT. Cell Syst. *7*, 77-91.e7.

8. Perkins, D.O., Jeffries, C.D., Jarskog, L.F., Thomson, J.M., Woods, K., Newman, M.A., Parker, J.S., Jin, J., and Hammond, S.M. (2007). microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. Genome Biol. *8*, R27.

9. Beveridge, N.J., Gardiner, E., Carroll, A.P., Tooney, P.A., and Cairns, M.J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. Mol. Psychiatry *15*, 1176–1189.

10. Saba, R., and Schratt, G.M. (2010). MicroRNAs in neuronal development, function and dysfunction. Brain Res. *1338*, 3–13.

11. Abu-Elneel, K., Liu, T., Gazzaniga, F.S., Nishimura, Y., Wall, D.P., Geschwind, D.H., Lao, K., and Kosik, K.S. (2008). Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics *9*, 153–161.

12. Fregeac, J., Colleaux, L., and Nguyen, L.S. (2016). The emerging roles of MicroRNAs in autism spectrum disorders. Neurosci. Biobehav. Rev. *71*, 729–738.

13. Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. Nat. Rev. Neurosci. *10*, 724–735.

14. Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18–36.

15. Rubenstein, J.L.R. (2011). Annual Research Review: Development of the cerebral cortex: implications for neurodevelopmental disorders. J. Child Psychol. Psychiatry *52*, 339–355.

16. Gilman, S.R., Chang, J., Xu, B., Bawa, T.S., Gogos, J.A., Karayiorgou, M., and Vitkup, D. (2012). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. Nat. Neurosci. *15*, 1723–1728.

17. Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, *et al.* (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. Cell *154*, 518–529.

18. Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008–1021.

19. Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., *et al.* (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

20. Noctor, S.C., Flint, A.C., Weissman, T.A., Dammerman, R.S., and Kriegstein, A.R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. Nature *409*, 714–720.

21. Namba, T., and Huttner, W.B. (2017). Neural progenitor cells and their role in the development and evolutionary expansion of the neocortex. Wiley Interdiscip. Rev. Dev. Biol. *6*.

22. Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The cellular and molecular landscapes of the developing human central nervous system. Neuron *89*, 248–268.

23. Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J. Comp. Neurol. *513*, 532–541.

24. Rakic, P. (2000). Radial unit hypothesis of neocortical expansion. Novartis Found. Symp. *228*, 30–42; discussion 42.

25. Geschwind, D.H., and Rakic, P. (2013). Cortical evolution: judge the brain by its cover. Neuron *80*, 633–647.

26. Hazlett, H.C., Gu, H., Munsell, B.C., Kim, S.H., Styner, M., Wolff, J.J., Elison, J.T., Swanson, M.R., Zhu, H., Botteron, K.N., *et al.* (2017). Early brain development in infants at high risk for autism spectrum disorder. Nature *542*, 348–351.

27. de la Torre-Ubieta, L., Stein, J.L., Won, H., Opland, C.K., Liang, D., Lu, D., and Geschwind, D.H. (2018). The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell *172*, 289-304.e18.

28. Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D.P., Lind, P.A., Pizzagalli, F., Ching, C.R.K., McMahon, M.A.B., *et al.* (2020). The genetic architecture of the human cerebral cortex. Science *367*.

29. Nowakowski, T.J., Mysiak, K.S., O'Leary, T., Fotaki, V., Pratt, T., and Price, D.J. (2013). Loss of functional Dicer in mouse radial glia cell-autonomously prolongs cortical neurogenesis. Dev. Biol. *382*, 530–537.

30. Nowakowski, T.J., Mysiak, K.S., Pratt, T., and Price, D.J. (2011). Functional dicer is necessary for appropriate specification of radial glia during early development of mouse telencephalon. PLoS ONE *6*, e23013.

31. Nowakowski, T.J., Fotaki, V., Pollock, A., Sun, T., Pratt, T., and Price, D.J. (2013). MicroRNA-92b regulates the development of intermediate cortical progenitors in embryonic mouse brain. Proc Natl Acad Sci USA *110*, 7056–7061.

32. Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol. Cell *27*, 435–448.

33. Nowakowski, T.J., Rani, N., Golkaram, M., Zhou, H.R., Alvarado, B., Huch, K., West, J.A., Leyrat, A., Pollen, A.A., Kriegstein, A.R., *et al.* (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. Nat. Neurosci. *21*, 1784–1792.

34. Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709.

35. Gratten, J., Wray, N.R., Keller, M.C., and Visscher, P.M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat. Neurosci. *17*, 782–790.

36. Nica, A.C., and Dermitzakis, E.T. (2013). Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120362.

37. Cockerill, P.N. (2011). Structure and function of active chromatin and DNase I hypersensitive sites. FEBS J. *278*, 2182–2210.

38. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., *et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

39. Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M., *et al.* (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. Science *325*, 1246–1250.

40. Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., *et al.* (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat. Genet. *50*, 956–967.

41. PsychENCODE Consortium, Akbarian, S., Liu, C., Knowles, J.A., Vaccarino, F.M., Farnham, P.J., Crawford, G.E., Jaffe, A.E., Pinto, D., Dracheva, S., *et al.* (2015). The PsychENCODE project. Nat. Neurosci. *18*, 1707–1712.

42. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell *179*, 750-771.e22.

43. Aygün, N., Elwell, A.L., Liang, D., Lafferty, M.J., Cheek, K.E., Courtney, K.P., Mory, J., Hadden-Ford, E., Krupa, O., de la Torre-Ubieta, L., *et al.* (2021). Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. Am. J. Hum. Genet. *108*, 1647–1668.

44. O'Brien, H.E., Hannon, E., Hill, M.J., Toste, C.C., Robertson, M.J., Morgan, J.E., McLaughlin, G., Lewis, C.M., Schalkwyk, L.C., Hall, L.S., *et al.* (2018). Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. Genome Biol. *19*, 194.

45. Uffelmann, E., and Posthuma, D. (2021). Emerging methods and resources for biological interrogation of neuropsychiatric polygenic signal. Biol. Psychiatry *89*, 41–53.

46. Umans, B.D., Battle, A., and Gilad, Y. (2021). Where Are the Disease-Associated eQTLs? Trends Genet. *37*, 109–124.

47. Borel, C., Deutsch, S., Letourneau, A., Migliavacca, E., Montgomery, S.B., Dimas, A.S., Vejnar, C.E., Attar, H., Gagnebin, M., Gehrig, C., *et al.* (2011). Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts. Genome Res. *21*, 68–73.

48. Civelek, M., Hagopian, R., Pan, C., Che, N., Yang, W., Kayne, P.S., Saleem, N.K., Cederberg, H., Kuusisto, J., Gargalovic, P.S., *et al.* (2013). Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. Hum. Mol. Genet. *22*, 3023–3037.

49. Gamazon, E.R., Ziliak, D., Im, H.K., LaCroix, B., Park, D.S., Cox, N.J., and Huang, R.S. (2012). Genetic architecture of microRNA expression: implications for the transcriptome and complex traits. Am. J. Hum. Genet. *90*, 1046–1063.

50. Huan, T., Rong, J., Liu, C., Zhang, X., Tanriverdi, K., Joehanes, R., Chen, B.H., Murabito, J.M., Yao, C., Courchesne, P., *et al.* (2015). Genome-wide identification of microRNA expression quantitative trait loci. Nat. Commun. *6*, 6601.

51. Volvert, M.L., Rogister, F., Moonen, G., Malgrange, B., and Nguyen, L. (2012). MicroRNAs tune cerebral cortical neurogenesis. Cell Death Differ. *19*, 1573–1581.

52. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., *et al.* (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. *50*, 1112–1121.

53. Davies, G., Lam, M., Harris, S.E., Trampush, J.W., Luciano, M., Hill, W.D., Hagenaars, S.P., Ritchie, S.J., Marioni, R.E., Fawns-Ritchie, C., *et al.* (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat. Commun. *9*, 2098.

54. Knol, M.J., Poot, R.A., Evans, T.E., Satizabal, C.L., Mishra, A., van der Auwera, S., Duperron, M.-G., Jian, X., Hostettler, I.C., van Dam-Nolen, D.H.K., *et al.* (2020). Genetic variants for head size share genes and pathways with cancer. BioRxiv.

55. Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J., *et al.* (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. *48*, 624–633.

56. Zeng, B., Bendl, J., Kosoy, R., Fullard, J.F., Hoffman, G.E., and Roussos, P. (2022). Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. Nat. Genet. *54*, 161–169.

57. Visscher, P.M., and Goddard, M.E. (2019). From R.A. fisher's 1918 paper to GWAS a century later. Genetics *211*, 1125–1130.

58. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. *51*, 1339–1348.

59. Li, J., Xue, Y., Amin, M.T., Yang, Y., Yang, J., Zhang, W., Yang, W., Niu, X., Zhang, H.-Y., and Gong, J. (2020). ncRNA-eQTL: a database to systematically evaluate the effects of SNPs on non-coding RNA expression across cancer types. Nucleic Acids Res. *48*, D956–D963.

60. Fineberg, S.K., Kosik, K.S., and Davidson, B.L. (2009). MicroRNAs potentiate neural development. Neuron *64*, 303–309.

61. Liang, D., Elwell, A.L., Aygün, N., Krupa, O., Wolter, J.M., Kyere, F.A., Lafferty, M.J., Cheek, K.E., Courtney, K.P., Yusupova, M., *et al.* (2021). Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. Nat. Neurosci. *24*, 941–953.

62. Rakic, P. (1988). Specification of cerebral cortical areas. Science *241*, 170–176.

63. Lu, Y., Baras, A.S., and Halushka, M.K. (2018). miRge 2.0 for comprehensive analysis of microRNA sequencing data. BMC Bioinformatics *19*, 275.

64. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162.

65. Fernández, V., Llinares-Benadero, C., and Borrell, V. (2016). Cerebral cortex expansion and folding: what have we learned? EMBO J. *35*, 1021–1044.

66. Nowakowski, T.J., Pollen, A.A., Sandoval-Espinosa, C., and Kriegstein, A.R. (2016). Transformation of the radial glia scaffold demarcates two stages of human cerebral cortex development. Neuron *91*, 1219–1227.

67. Roese-Koerner, B., Stappert, L., and Brüstle, O. (2017). Notch/Hes signaling and miR-9 engage in complex feedback interactions controlling neural progenitor cell proliferation and differentiation. Neurogenesis (Austin) *4*, e1313647.

68. Sun, Y., Luo, Z.-M., Guo, X.-M., Su, D.-F., and Liu, X. (2015). An updated role of microRNA-124 in central nervous system disorders: a review. Front. Cell. Neurosci. *9*, 193.

69. Sengupta, S., Nie, J., Wagner, R.J., Yang, C., Stewart, R., and Thomson, J.A. (2009). MicroRNA 92b controls the G1/S checkpoint gene p57 in human embryonic stem cells. Stem Cells *27*, 1524–1528.

70. Wu, Z.B., Cai, L., Lin, S.J., Lu, J.L., Yao, Y., and Zhou, L.F. (2013). The miR-92b functions as a potential oncogene by targeting on Smad3 in glioblastomas. Brain Res. *1529*, 16–25.

71. Long, M., Zhan, M., Xu, S., Yang, R., Chen, W., Zhang, S., Shi, Y., He, Q., Mohan, M., Liu, Q., *et al.* (2017). miR-92b-3p acts as a tumor suppressor by targeting Gabra3 in pancreatic cancer. Mol. Cancer *16*, 167.

72. Zhuang, L.K., Yang, Y.T., Ma, X., Han, B., Wang, Z.S., Zhao, Q.Y., Wu, L.Q., and Qu, Z.Q. (2016). MicroRNA-92b promotes hepatocellular carcinoma progression by targeting Smad7 and is mediated by long non-coding RNA XIST. Cell Death Dis. *7*, e2203.

73. Zhou, M.-H., Zhou, H.-W., Liu, M., and Sun, J.-Z. (2018). The role of miR-92b in cholangiocarcinoma patients. Int. J. Biol. Markers *33*, 293–300.

74. Lee, J., Heo, J., and Kang, H. (2019). miR-92b-3p-TSC1 axis is critical for mTOR signaling-mediated vascular smooth muscle cell proliferation induced by hypoxia. Cell Death Differ. *26*, 1782–1795.

75. Wang, Y., Wang, D., and Guo, D. (2016). MiR-124 Promote Neurogenic Transdifferentiation of Adipose Derived Mesenchymal Stromal Cells Partly through RhoA/ROCK1, but Not ROCK2 Signaling Pathway. PLoS ONE *11*, e0146646.

76. Xue, Q., Yu, C., Wang, Y., Liu, L., Zhang, K., Fang, C., Liu, F., Bian, G., Song, B., Yang, A., *et al.* (2016). miR-9 and miR-124 synergistically affect regulation of dendritic branching via the AKT/GSK3β pathway by targeting Rap2a. Sci. Rep. *6*, 26781.

77. Mondanizadeh, M., Arefian, E., Mosayebi, G., Saidijam, M., Khansarinejad, B., and Hashemi, S.M. (2015). MicroRNA-124 regulates neuronal differentiation of mesenchymal stem cells by targeting Sp1 mRNA. J. Cell. Biochem. *116*, 943–953.

78. Dash, S., Balasubramaniam, M., Martínez-Rivera, F.J., Godino, A., Peck, E.G., Patnaik, S., Suar, M., Calipari, E.S., Nestler, E.J., Villalta, F., *et al.* (2020). Cocaine-regulated microRNA miR-124 controls poly (ADP-ribose) polymerase-1 expression in neuronal cells. Sci. Rep. *10*, 11197.

79. Friedländer, M.R., Lizano, E., Houben, A.J.S., Bezdan, D., Báñez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B., González, J., Chen, K.C., *et al.* (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. Genome Biol. *15*, R57.

80. Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52.

81. Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat. Protoc. *9*, 263–293.

82. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., *et al.* (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.

83. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., *et al.* (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290–299.

84. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.

85. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. *46*, 100–106.

86. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

87. Liu, B., Shyr, Y., Cai, J., and Liu, Q. (2018). Interplay between miRNAs and host genes and their role in cancer. Brief. Funct. Genomics *18*, 255–266.

88. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci USA *100*, 9440–9445.

89. 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, *et al.* (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

90. Ghanbari, M., Iglesias, A.I., Springelkamp, H., van Duijn, C.M., Ikram, M.A., Dehghan, A., Erkeland, S.J., Klaver, C.C.W., Meester-Smoor, M.A., and International Glaucoma Genetics Consortium (IGGC) (2017). A Genome-Wide Scan for MicroRNA-Related Genetic Variants Associated With Primary Open-Angle Glaucoma. Invest. Ophthalmol. Vis. Sci. *58*, 5368–5377.

91. Lawo, S., Bashkurov, M., Mullin, M., Ferreria, M.G., Kittler, R., Habermann, B., Tagliaferro, A., Poser, I., Hutchins, J.R.A., Hegemann, B., *et al.* (2009). HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. Curr. Biol. *19*, 816–826.

92. Uehara, R., Nozawa, R., Tomioka, A., Petry, S., Vale, R.D., Obuse, C., and Goshima, G. (2009). The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells. Proc Natl Acad Sci USA *106*, 6998–7003.

93. Stein, J.L., de la Torre-Ubieta, L., Tian, Y., Parikshak, N.N., Hernández, I.A., Marchetto, M.C., Baker, D.K., Lu, D., Hinman, C.R., Lowe, J.K., *et al.* (2014). A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron *83*, 69–86.

94. Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*.

95. Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., *et al.* (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. *24*, 14–24.

96. Sousa, A.M.M., Meyer, K.A., Santpere, G., Gulden, F.O., and Sestan, N. (2017). Evolution of the human nervous system function, structure, and development. Cell *170*, 226–247.

97. Stepien, B.K., Naumann, R., Holtz, A., Helppi, J., Huttner, W.B., and Vaid, S. (2020). Lengthening Neurogenic Period during Neocortical Development Causes a Hallmark of Neocortex Expansion. Curr. Biol. *30*, 4227-4237.e5.

98. Otani, T., Marchetto, M.C., Gage, F.H., Simons, B.D., and Livesey, F.J. (2016). 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. Cell Stem Cell *18*, 467–480.

99. Florio, M., Borrell, V., and Huttner, W.B. (2017). Human-specific genomic signatures of neocortical expansion. Curr. Opin. Neurobiol. *42*, 33–44.

100. Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., Wimberger, P., Huttner, W.B., and Hiller, M. (2018). Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. eLife *7*.

101. Hu, H.Y., Guo, S., Xi, J., Yan, Z., Fu, N., Zhang, X., Menzel, C., Liang, H., Yang, H., Zhao, M., *et al.* (2011). MicroRNA expression and regulation in human, chimpanzee, and macaque brains. PLoS Genet. *7*, e1002327.

102. Barbash, S., Shifman, S., and Soreq, H. (2014). Global coevolution of human microRNAs and their target genes. Mol. Biol. Evol. *31*, 1237–1247.

103. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

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

105. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

106. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

107. Fuchsberger, C., Abecasis, G.R., and Hinds, D.A. (2015). minimac2: faster genotype imputation. Bioinformatics *31*, 782–784.

108. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. *91*, 839–848.

109. R Foundation for Statistical Computing, R.C.T. (2021). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).

110. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

111. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

112. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084–2092.

113. Davis, J.R., Fresard, L., Knowles, D.A., Pala, M., Bustamante, C.D., Battle, A., and Montgomery, S.B. (2016). An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. Am. J. Hum. Genet. *98*, 216–224.

114. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., *et al.* (2015). A global reference for human genetic variation. Nature *526*, 68–74.

115. Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganella, S., Min, J.L., Walter, K., Timpson, N.J., UK10K Consortium, Dunham, I., Birney, E., *et al.* (2019). GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat. Genet. *51*, 343–353.

116. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

117. Storey, J.D., Bass, A.J., Dabney, A., and Robinson, D. (2021). qvalue: Q-value estimation for false discovery rate control (R package version 2.26.0).

118. Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. Nucleic Acids Res. *40*, D1144-9.

119. McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., *et al.* (2018). CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol. *16*, e2005970.

120. 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., *et al.* (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489.

121. Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., *et al.* (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

122. Pollen, A.A., Nowakowski, T.J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C.R., Shuga, J., Liu, S.J., Oldham, M.C., Diaz, A., *et al.* (2015). Molecular identity of human outer radial glia during cortical development. Cell *163*, 55–67.

123. Nord, A.S., Pattabiraman, K., Visel, A., and Rubenstein, J.L.R. (2015). Genomic perspectives of transcriptional regulation in forebrain development. Neuron *85*, 27–47.

124. Visel, A., Taher, L., Girgis, H., May, D., Golonzhka, O., Hoch, R.V., McKinsey, G.L., Pattabiraman, K., Silberberg, S.N., Blow, M.J., *et al.* (2013). A high-resolution enhancer atlas of the developing telencephalon. Cell *152*, 895–908.

125. Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D., *et al.* (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538*, 523–527.

126. Fabian, M.R., Sundermeier, T.R., and Sonenberg, N. (2010). Understanding how miRNAs post-transcriptionally regulate gene expression. Prog. Mol. Subcell. Biol. *50*, 1–20.

127. Zhang, J.-G., Xu, C., Zhang, L., Zhu, W., Shen, H., and Deng, H.-W. (2019). Identify gene expression pattern change at transcriptional and post-transcriptional levels. Transcription *10*, 137–146.

128. Sun, E., and Shi, Y. (2015). MicroRNAs: Small molecules with big roles in neurodevelopment and diseases. Exp. Neurol. *268*, 46–53.

129. Mayer, S., Chen, J., Velmeshev, D., Mayer, A., Eze, U.C., Bhaduri, A., Cunha, C.E., Jung, D., Arjun, A., Li, E., *et al.* (2019). Multimodal Single-Cell Analysis Reveals Physiological Maturation in the Developing Human Neocortex. Neuron *102*, 143-158.e7.

130. Huntley, R.P., Sitnikov, D., Orlic-Milacic, M., Balakrishnan, R., D'Eustachio, P., Gillespie, M.E., Howe, D., Kalea, A.Z., Maegdefessel, L., Osumi-Sutherland, D., *et al.* (2016). Guidelines for the functional annotation of microRNAs using the Gene Ontology. RNA *22*, 667–676.

131. Huntley, R.P., Kramarz, B., Sawford, T., Umrao, Z., Kalea, A., Acquaah, V., Martin, M.J., Mayr, M., and Lovering, R.C. (2018). Expanding the horizons of microRNA bioinformatics. RNA *24*, 1005–1017.

132. Molnár, Z., Clowry, G.J., Šestan, N., Alzu'bi, A., Bakken, T., Hevner, R.F., Hüppi, P.S., Kostović, I., Rakic, P., Anton, E.S., *et al.* (2019). New insights into the development of the human cerebral cortex. J. Anat. *235*, 432–451.

133. Pajak, M, S.T. (2018). miRNAtap: microRNA Targets - Aggregated Predictions (R package version 1.16.0.).

134. Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., *et al.* (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res. *37*, W273-6.

135. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol. *5*, R1.

136. Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.-L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., Macmenamin, P., *et al.* (2006). A genome-wide map of conserved microRNA targets in C. elegans. Curr. Biol. *16*, 460–471.

137. Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

138. Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. *43*, D146-52.

139. Lang, M.-F., and Shi, Y. (2012). Dynamic Roles of microRNAs in Neurogenesis. Front. Neurosci. *6*, 71.

140. Quillet, A., Saad, C., Ferry, G., Anouar, Y., Vergne, N., Lecroq, T., and Dubessy, C. (2019). Improving Bioinformatics Prediction of microRNA Targets by Ranks Aggregation. Front. Genet. *10*, 1330.

141. Schäfer, M., and Ciaudo, C. (2020). Prediction of the miRNA interactome - Established methods and upcoming perspectives. Comput. Struct. Biotechnol. J. *18*, 548–557.

142. Hsu, S.-D., Lin, F.-M., Wu, W.-Y., Liang, C., Huang, W.-C., Chan, W.-L., Tsai, W.-T., Chen, G.-Z., Lee, C.-J., Chiu, C.-M., *et al.* (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. *39*, D163-9.

143. Studer, L., Vera, E., and Cornacchia, D. (2015). Programming and reprogramming cellular age in the era of induced pluripotency. Cell Stem Cell *16*, 591–600.

144. Qi, Y., Zhang, X.-J., Renier, N., Wu, Z., Atkin, T., Sun, Z., Ozair, M.Z., Tchieu, J., Zimmer, B., Fattahi, F., *et al.* (2017). Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. Nat. Biotechnol. *35*, 154–163.

145. Galton, F. (1904). Eugenics: Its Definition, Scope and Aims. Available at: https://galton.org/essays/1900-1911/galton-1904-am-journ-soc-eugenics-scope-aims.htm [Accessed March 14, 2022].

146. Stern, A.M. (2020). Forced sterilization policies in the US targeted minorities and those with disabilities – and lasted into the 21st century. Available at: https://ihpi.umich.edu/news/forced-sterilization-policies-us-targeted-minorities-and-those-disabilities-and-lasted-21st [Accessed March 14, 2022].

147. Comfort, N. (2012). The Eugenic Impulse. Available at: https://www.chronicle.com/article/the-eugenic-impulse/ [Accessed March 14, 2022].

148. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.

149. Lewis, C.M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. Genome Med. *12*, 44.

150. Lewis, A.C.F., and Green, R.C. (2021). Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues. Genome Med. *13*, 14.

151. Cano-Gamez, E., and Trynka, G. (2020). From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front. Genet. *11*, 424.

152. Devlin, B., and Scherer, S.W. (2012). Genetic architecture in autism spectrum disorder. Curr. Opin. Genet. Dev. *22*, 229–237.

153. Bassett, A.S., Chow, E.W., O'Neill, S., and Brzustowicz, L.M. (2001). Genetic insights into the neurodevelopmental hypothesis of schizophrenia. Schizophr. Bull. *27*, 417–430.

154. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

155. Koussounadis, A., Langdon, S.P., Um, I.H., Harrison, D.J., and Smith, V.A. (2015). Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Sci. Rep. *5*, 10775.

156. Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M., and Nilsson, P. (2009). Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics *10*, 365.

157. Ramachandran, S., Coffin, S.L., Tang, T.-Y., Jobaliya, C.D., Spengler, R.M., and Davidson, B.L. (2016). Cis-acting single nucleotide polymorphisms alter MicroRNA-mediated regulation of human brain-expressed transcripts. Hum. Mol. Genet. *25*, 4939–4950.

158. Peterson, S.M., Thompson, J.A., Ufkin, M.L., Sathyanarayana, P., Liaw, L., and Congdon, C.B. (2014). Common features of microRNA target prediction tools. Front. Genet. *5*, 23.

159. Riffo-Campos, Á.L., Riquelme, I., and Brebi-Mieville, P. (2016). Tools for Sequence-Based miRNA Target Prediction: What to Choose? Int. J. Mol. Sci. *17*.

160. Singh, N.K. (2017). miRNAs target databases: developmental methods and target identification techniques with functional annotations. Cell. Mol. Life Sci. *74*, 2239–2261.

161. Liu, W., and Wang, X. (2019). Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. Genome Biol. *20*, 18.

162. Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

163. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., *et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

164. Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274–287.

165. Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell *153*, 654–665.