Sites of active gene regulation in the developing human brain and their role in neuropsychiatric disorders

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

Manuela Rosa Kouakou

Acknowledgments

Acronyms

ADHD:	Attention deficit hyperactivity disorder
ASD:	Autism spectrum disorder
ATAC-Seq:	Assay of transposase-accessible chromatin with sequencing
ChIP-Seq:	Chromatin immunoprecipitation with sequencing
CNV:	Copy number variant
DMEM:	Dulbecco's modified Eagles's Medium
DMSO:	Dimethyl sulfoxide
DNA:	Deoxyribonucleic Acid
DHS:	DNase-hypersensitivity
DSM:	Diagnostic and Statistical Manual
ENCODE:	Encyclopaedia of DNA elements project
eQTL:	Expression quantitative trait loci
FACS:	Fluorescence-activated cell sorting
FANS	Fluorescence-activated nuclei sorting
FSC-A:	Forward scatter area
FSC-H:	Forward scatter height
GARFIELD:	GWAS analysis of regulatory or functional information enrichment with LD Correction
GCTA:	Genome-wide Complex Trait Analysis
GWAS:	Genome-wide association study
H3K4Me1	Monomethylation of histone 3 lysine 4
H3K4Me3	Trimethylation of histone 3 lysine 4
H3K27ac	Acetylation of histone 3 lysine 27
HDBR:	Human Developmental Biology Resource
ICD:	International Statistical Classification of Mental and Behavioural Disorders
IGV:	Integrative genomics viewer
Indels:	Insertions-deletion
iPSC:	Induced pluripotent stem cells
LD:	Linkage Disequilibrium
LDAK:	Linkage disequilibrium adjusted kinships
LDSC:	Linkage Disequilibrium score regression
MAF:	Minor allele frequency
NeuN+:	Neuron-enriched
NeuN-:	Neuron-depleted
PRS:	Polygenic risk score
PGS:	Polygenic score
PCR:	Polymerase chain reaction
REMC:	Roadmap Epigenomic Mapping Consortium
SCHEMA:	Schizophrenia Exome Meta-Analysis
SLDSC:	Stratified Linkage Disequilibrium score regression
sQTLs:	splice eQTLs
SNP:	Single nucleotide polymorphism
SNV:	Single nucleotide variant
TCF4:	Transcription factor 4

TF:	Transcription factor
TSS:	Transcription start sites
TWAS:	Transcriptome-wide association study
VCFS:	Velocardiofacial syndrome

Summary

Neuropsychiatric conditions such as schizophrenia, autism and attention deficit hyperactivity disorder (ADHD) are complex disorders with a hypothesized early neurodevelopmental component. Most common risk loci for these disorders are located in non-coding regions of the genome and are therefore likely to index functional variants that alter gene regulation rather than protein structure. It is possible to elucidate the biology underpinning these conditions by testing for enrichment of associated genetic variation within regulatory genomic regions operating in the developing human brain.

In this thesis, I have used ATAC-Seq to map open chromatin (an index of active regulatory genomic regions) in bulk tissue, neuron-enriched (NeuN+) and neuron-depleted (NeuN-) nuclei from the prenatal human frontal cortex, and tested enrichment of single nucleotide polymorphism (SNP) heritability for 5 neuropsychiatric disorders (autism spectrum disorder, ADHD, bipolar disorder, major depressive disorder and schizophrenia) within these regions.

I found significant enrichment of SNP heritability for schizophrenia within open chromatin regions mapped in bulk foetal frontal cortex, and for all 5 tested neuropsychiatric conditions when I restricted these sites to those overlapping histone modifications indicative of enhancers (H3K4me1) or promoters (H3K4me3) in foetal brain. SNP heritability for schizophrenia was significantly enriched in foetal NeuN+ nuclei overlapping H3K4me1 sites and for all 5 neuropsychiatric disorders in foetal NeuN- nuclei overlapping either H3K4me1 or H3K4me3 sites. I further used my mapped open chromatin regions to identify potentially functional SNPs at genome-wide significant risk loci for schizophrenia.

Table of Contents

Chapter 1: General introduction

Psychiatric disorders are highly debilitating conditions and a leading cause of disability worldwide. Despite recent global efforts to tackle this public health issue (Saxena, Funk, and Chisholm 2015), diagnostic tools and treatment options remain limited, mainly because their aetiology is unknown.

Generally, diagnoses are formulated in order to be able to suggest treatments and predict the course of the disorder. The vast majority of diagnostic systems label disease according to both clinical characteristics and aetiopathology. Given that the aetiology of psychiatric disorders remains obscure, the diagnostic system in psychiatry is based upon common clinical features, shared natural history and /or common treatment response. Individual conditions are considered as members of groups contained within a hierarchy; that is, a form of classification system. The two classification systems of psychiatric disorders mostly used for both clinical and research purposes are the Diagnostic and Statistical Manual (DSM) (American Psychiatric Association 2013) and the International Statistical Classification of Mental and Behavioural Disorders (ICD, World Health Organization, 2018), today at their fifth and eleventh editions respectively.

Indisputably, classification systems greatly facilitate clinical management. However, the reliability and validity of diagnoses purely based on symptoms, with no biomarkers, scans or other tests available, are rightly considered questionable by many (Clark, Watson, and Reynolds 1995). Moreover, in terms of therapeutic interventions, a substantial proportion of

patients show only partial or no response to first or second-line pharmacological treatments. Most drugs used today were discovered decades ago by serendipity and cause debilitating and possibly life-threatening side effects (Muench and Hamer 2010). Low compliance as well as treatment resistance are common. There is clearly a pressing need for new treatment strategies.

More effective diagnostic and treatment strategies for an illness can only be achieved when its underpinning molecular mechanisms are uncovered. Compared to other medical specialities, research in psychiatry has lagged behind for several reasons. Historically, the organic nature of conditions presenting with behavioural and psychological symptoms has often been questioned (Tyrer and Mackay 1986). Certainly, psychiatric disorders such as schizophrenia are not associated with overt brain pathology (Harrison 1999) as seen for neurodegenerative conditions. Additionally, investigation of pathologies that affect brain function is hindered by lack of tissue availability. However, recently, unprecedented progress in technology, as well as larger sample sizes, has allowed a deeper understanding of the aetiology of psychiatric disorders and a wide range of evidence indicates that there is a neural substrate for brain disorders that were once considered purely functional.

In the next section {1.1}, I will give a brief overview of the clinical characteristics, environmental risk factors and available treatment options of 5 major neuropsychiatric conditions: schizophrenia, attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), bipolar disorder and major depressive disorder. I will then move on to discuss advances in our understanding of their underpinning genetics.

1.1 Psychiatric disorders: clinical characteristic and epidemiology

1.1.1 Schizophrenia

Schizophrenia is one of the most devastating psychiatric illnesses, with an enormous personal and societal cost. It affects approximately 0.4% people worldwide (Bhugra 2005). The age of onset is typically late adolescence to early adulthood, with men manifesting the first symptoms earlier in life compared to women (Gogtay et al. 2011). The predominant clinical features of schizophrenia are two main symptom clusters, termed 'positive' and 'negative' symptoms, as well as cognitive and functional disabilities. Positive symptoms, so named for the presence of behaviours and sensations that should not be present, are also known as psychotic symptoms and consist of hallucinations, delusions and disorganised thinking. Negative symptoms are mainly deficit symptoms and include apathy, lack of drive and motivation, slowness, social withdrawal and the inability to feel and express emotions.

Medications for positive symptoms (antipsychotics) were initially discovered by serendipity in the mid-1950s. Since then, there has been little progress in treatment for negative and cognitive symptoms. Newer medications (formerly called atypical) can be useful for some treatment-resistant patients but are associated with potentially severe side effects. For example, neutropenia is observed in around 3% of people taking clozapine, requiring patients to undergo frequent blood tests, thus reducing compliance.

Various environmental factors such as urbanicity, migration, cannabis, childhood traumas, infectious agents, obstetrical complications and psychosocial factors have been associated with the risk of developing schizophrenia (Stilo and Murray 2010). Early hypotheses regarding

the pathology of schizophrenia emerged from studying the mechanisms of antipsychotic drugs. Since treatment of positive symptoms appears to be related to the antagonism of the dopamine D2-type receptor (Creese, Burt, and Snyder 1976), hallucinations and delusions have been postulated to be due to a hyperactive dopaminergic signal (the 'dopamine hypothesis' of schizophrenia). A neurodevelopmental hypothesis for the aetiopathogenesis of schizophrenia is now also widely accepted (see section 1.5.1 below).

1.1.2 Attention deficit hyperactivity disorder

Given its age of onset in childhood/adolescence, attention deficit hyperactivity disorder (ADHD) is considered a neurodevelopmental disorder by DSM-V. Its prevalence is quite variable among age groups and gender, with boys aged 6-11 years being the most affected overall (12.9% vs 5.6% of girls) (Danielson et al. 2018). The main clinical features are high levels of inattention, overactivity and impulsiveness, meaning a short attention span or lack of attention to details, an excess of movements in situations that require calm, and action without reflection. Given the continuous distribution of these clinical features in the population, strict diagnostic criteria were all the more needed for ADHD, in order to set a more tangible cut-off between pathology and what was considered 'just bad behaviour'. In this context, molecular biology research has been essential in providing evidence that a neuropathological background exists for a syndrome that was once labelled as purely psychosocial. Available pharmacological treatment for ADHD is aimed at reducing the most disruptive symptoms, aiding with concentration issues and impulsiveness. The most commonly used medication is methylphenidate, a stimulant that inhibits the reuptake of dopamine and norepinephrine, increasing the dopaminergic and noradrenergic activity in the

prefrontal cortex. Several environmental risk factors have been implicated in the aetiology of ADHD: prenatal substance exposures (tobacco, alcohol, drugs), heavy metal and chemical exposures, nutritional factors, low birth weight, as well as early traumatic events (Froehlich et al. 2011).

1.1.3 Autism spectrum disorder

Classified as a neurodevelopmental disability by the DSM-V, autism spectrum disorder (ASD) is characterised by abnormalities of social development and communication, and restriction of interests and behaviour. Individuals affected share some of these features; however, there is a wide variation in their expression and severity. For this reason, the disorder is considered a "spectrum condition", which affects people differently and to varying levels of impairment. For instance, abnormalities of communication may range from having difficulties in understanding nonverbal cues, such as facial expressions, hand gestures or vocal tones, to being unable to use and understand spoken language. Therapeutic interventions are mainly non-pharmacological and aimed at improving communication, speech and language skills. Although most children are not diagnosed before the age of 3, prodromal signs, such as difficulties with social communication, may manifest at an earlier stage. With a prevalence of approximately 1%, it is diagnosed approximately 4 times more frequently in males (Werling and Geschwind 2013). A wide range of environmental risk factors have been identified, including pre-, peri- and post-natal complications, exposure to toxins, migration and parental age among others (Kim et al. 2019).

1.1.4 Mood disorders: bipolar and major depressive disorder

Changes from baseline normal mood (that is, depression and / or elation) are the main feature of mood disorders. These changes differ from normal fluctuations in mood because they last longer and disrupt patients' social and occupational functioning.

The lifetime risk of bipolar disorder is in the range 0.6-1.4% (Merikangas et al. 2011), with a mean age of onset of about 20 years in community studies (Wittchen, Mhlig, and Pezawas 2003). Several environmental risk factors have been shown to increase the risk for bipolar disorder: prenatal viral infections, in particular seropositivity for *T. gondii* infection, prematurity (less than 32 weeks' gestation), childhood maltreatment as well as specific life events like early prenatal loss and childbirth and cannabis use (Bortolato et al. 2017).

BD is characterised by episodes of mania (i.e. elevated, expansive, or irritable mood often associated with sleep deprivation, self-important ideas, increased goal-directed activity and thought disorder). Manic episodes are clear periods of altered function that often require hospitalization and are associated with high rate of suicidal ideation and attempts. Manic episodes can alternate with episodes of major depression, in a cyclic manner. Pharmacological treatment includes mood stabilizers (lithium), antipsychotic and antidepressant medications. Although lithium is the most commonly used mood stabilizer in BD, its exact mechanism of action is unknown and due to its large number of side effects, its tolerability and compliance are low. The medications mostly prescribed today for depressive symptoms are selective serotonin reuptake inhibitors (SSRIs), such as fluoxetine, that increase availability of serotonin in the synaptic space.

Major depressive disorder, or unipolar depressive disorder, is characterised by low mood, lack of enjoyment, negative thinking and decreased energy. Depression may also present with psychotic features, typical of self-deprecatory or paranoid content. It is more common in females, with a point prevalence of around 12.9% (Lim et al. 2018). Several environmental risk factors have been posited to contribute to the development of depression, including but not limited to childhood maltreatment, childhood loss of a parent lack of adequate parental care , unemployment, lower educational attainment, lower social support, the absence of a partner, lower physical activity, and cannabis use (Köhler et al. 2018)

1.2 The genetics of psychiatric disorders

The genetic architecture of psychiatric disorders is complex and highly polygenic. Contrary to Mendelian disorders, where one genetic locus exerts a strong effect on risk, in complex disorders multiple genetic loci play a role, with different effect sizes. For this reason, first attempts to elucidate the genetics underlying psychiatric disorders through use of linkage studies and candidate gene association studies failed to reproducibly identify susceptibility variants. Large advances were only made in the early 2000s with the development of genotyping arrays, that allowed the genotyping of thousands of DNA variants in a fast and cost-effective manner. Use of genotyping arrays and next generation sequencing, together with very large sample sizes, are now uncovering the complex genetic architecture of psychiatric disorders, with both rare and common variants known to play a role (see section 1.2.1).

1.2.1 Genetic variation

DNA variation can take several forms. It includes changes at individual nucleotide bases (single nucleotide polymorphism [SNPs] or single nucleotide variants [SNVs]), small insertions or deletions of DNA sequence (indels), polymorphic repeat sequences (e.g., mononucleotide, dinucleotide and trinucleotide repeats) and large structural alterations such as copy number variants (CNVs), which are usually defined as insertions, deletions or duplications of 1 kilobase (kb) or larger.

1.2.2 Rare DNA variation and risk for neuropsychiatric disorders

Rare DNA variation is typically defined as variation that has a population frequency < 1%. Several recurrent CNVs with a frequency of <1% have been shown to increase the risk of developing schizophrenia (Marshall et al. 2017) and other neurodevelopmental disorders with large associated effect sizes (odds ratios of between 4 and 70). Involving large chromosomal regions, disease-associated CNVs often encompass a number of genes and are pleiotropic (i.e. associated with multiple phenotypes). A typical example is a deletion on chromosome 22q11.2 causing velocardiofacial (VCFS or Di George) syndrome, which is also associated with a much higher risk of developing schizophrenia (Murphy, Jones, and Owen 1999), ADHD (Schneider et al. 2014) and ASD (Clements et al. 2017). CNVs on chromosomes 1q21.1, 15q13.3 and at the *NRXN1* locus have also been shown to increase both ASD and schizophrenia risk (Mefford et al. 2008; Sharp et al. 2008; Kirov et al. 2008, 2009), suggesting shared pathways between some psychiatric disorders. An increased rate of rare CNVs has also been observed in ADHD (Williams et al., 2010). However, CNVs appear to be less important

in the aetiology of bipolar disorder (Charney et al. 2019) and major depressive disorder (Rucker et al. 2016).

While CNVs can be identified through standard SNP genotyping arrays, rare coding variation increasing risk for neuropsychiatric disorders has been explored using next generation sequencing of the exome in large numbers of samples. Rare loss-of-function (LoF) coding mutations have been shown to increase the risk for schizophrenia, autism and severe developmental disorders (Singh et al. 2016; Satterstrom et al. 2020). Recently, the Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium have identified 10 genes harbouring a genome-wide significant excess of rare coding variants in schizophrenia (Singh et al. 2020).

CNV analyses and exome sequencing studies have also highlighted an important role for *de novo* mutations (i.e. mutations not present in the germline) in the aetiology of developmental psychiatric disorders (Sanders et al. 2015; Fromer et al. 2014). Both *de novo* CNVs and SNVs have the potential to be more deleterious than inherited variants as they have undergone less selection pressure (Conrad et al. 2011).

1.2.3 Common DNA variation and risk for neuropsychiatric disorders

In recent years, large-scale genome-wide association studies (GWASs) have successfully identified hundreds of common single nucleotide polymorphisms (SNPs) that are associated with increased risk for individual neuropsychiatric disorders at high levels of statistical

confidence. In order to detect variants associated with a trait, allele frequencies greater than 1 million SNPs are typically compared between affected 'cases' and control populations. In order to control for the high number of tests performed in a GWAS analyses, a significance threshold of $P < 5 \times 10^{-8}$ (a Bonferroni correction, based on one million independent tests) is set for 'genome-wide significance'. Very large sample sizes are therefore required for a GWAS to detect SNPs of small effect size increasing risk for neuropsychiatric disorders at this stringent level of significance. For this purpose, international collaborations, such as the Psychiatric Genomics Consortium (PGC), were started in the joint effort to provide genotyping information for thousands of patients and controls. To date, a considerable number of genome-wide significant risk loci for neuropsychiatric disorders have been published: 145 for schizophrenia (Pardiñas et al. 2018), 44 for major depressive disorder (Wray et al. 2018), 30 for bipolar disorder (Stahl et al. 2019), 16 for ADHD (Demontis, Walters, Martin, Mattheisen, Als, Neale, et al. 2019) and 5 for ASD (Grove et al. 2019). In the section below, I will give a brief overview of ADHD, schizophrenia, bipolar, major depressive disorder and ASD, GWAS methods and findings.
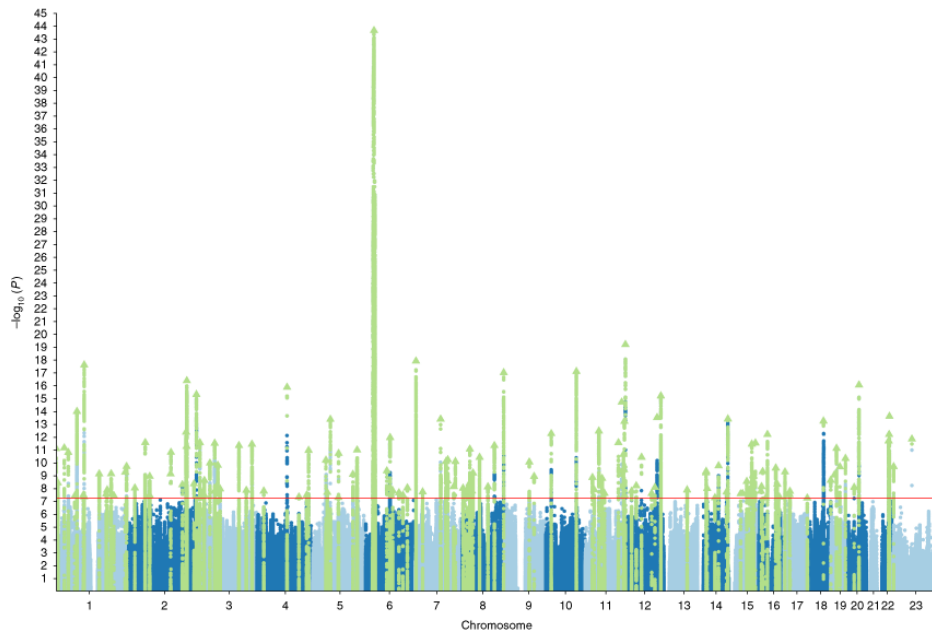
Figure 1.1. Taken from Pardinas et al, 2018. Manhattan plot of schizophrenia GWAS associations. The 145 genome-wide significant loci are highlighted in green. The red horizontal line indicates the genome-wide statistical significance threshold ($P = 5\times10^{-8}$)

ADHD GWAS (Demontis, Walters, Martin, Mattheisen, Als, Neale, et al. 2019)

Genotype array data for 20,183 individuals with ADHD and 35,191 controls were collected from 12 cohorts. These samples included a population-based cohort of 14,584 individuals with ADHD and 22,492 controls from Denmark and 11 European, North American and Chinese cohorts aggregated by the Psychiatric Genomics Consortium (PGC). Individuals with ADHD were diagnosed by psychiatrists at a psychiatric hospital according to ICD10 (F90.0). Non-genotyped markers were imputed using the 1000 Genomes Project Phase 3 reference panel. GWAS was conducted in each cohort using logistic regression with the imputed additive genotype dosages. The GWASs were then meta-analysed using an inverse-variance weighted fixed effects model. Association results were considered only for variants with an effective sample size >70% of the full meta-analysis, leaving 8,047,421 variants in the final meta-analysis. In total, 304 genetic variants in 12 loci surpassed the threshold for genome-wide

significance ($P < 5 \times 10^{-8}$). The authors used LD score regression and partitioned heritability to test for enrichment of the SNP heritability in cell-type-specific regulatory elements and observed a significant enrichment of the average per SNP heritability for variants located in central nervous system–specific regulatory elements. The authors estimated a liability-scale SNP heritability as $h^2_{SNP} = 0.216$ (SE = 0.014, $P = 8.18 \times 10^{-54}$), assuming a population prevalence of 5% for ADHD.

Schizophrenia GWAS (Pardiñas et al. 2018)

Following the seminal PGC schizophrenia GWAS, that initially identified 108 associated loci (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), Pardinas and colleagues (Pardiñas et al. 2018), reported a new genome-wide association study of schizophrenia that combined through meta-analysis the PGC datasets with an independent dataset. The authors collected blood samples from those with treatment-resistant schizophrenia in the UK through the mandatory clozapine blood-monitoring system for those taking clozapine, an antipsychotic licensed for treatment-resistant schizophrenia.

The final sample size was 11,260 cases and 24,542 controls (5,220 cases and 18,823 controls not in previous schizophrenia GWAS). Meta-analysis of the CLOZUK and the independent PGC datasets (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), excluding related and overlapping samples (total of 40,675 cases and 64,643 controls) identified 179 independent genome-wide significant SNPs ($P < 5 \times 10^{-8}$) mapping to 145 independent loci. The 145 associated loci included 93 of those that were genome-wide significant in the study of the PGC.

The polygenic variance on the liability scale[68] amounted to 5.7% for CLOZUK at the 0.05 *P*-value threshold.

The authors prioritised the SNPs that were more likely to be causal and found, among others, a number of SNPs mapping to regions identified as likely regulatory elements through chromosome conformation analysis performed in tissue from the developing brain using Hi-C physical interactions (Won et al. 2016).

Bipolar disorder GWAS (Stahl et al. 2019)

The authors performed a GWAS meta-analysis of 32 cohorts from 14 countries in Europe, North America and Australia, totalling 20,352 cases and 31,358 controls of European descent (effective sample size 46,582; Stahl et al. 2019). Variant dosages were imputed using the 1000 Genomes reference panel. The SNP heritability estimates for BD were 0.17–0.23 on the liability scale assuming population prevalence of 0.5–2%. Thirty autosomal loci achieved genome-wide significance ($P < 5 \times 10^{-8}$) in fixed-effect meta-analysis of the GWAS and follow-up samples. The authors tested for functional genomic enrichment using LD score regression and partitioned heritability (Finucane et al. 2015) and found enrichment in open chromatin annotations in the central nervous system.

Major depression GWAS (Wray et al. 2018)

The authors completed a GWA meta-analysis of seven MDD and major depression cohorts, 9.6 million imputed SNPs in 135,458 MDD and major depression cases and 344,901 controls. Cases included both directly evaluated subjects meeting standard criteria for major depressive disorder (MDD) and individuals identified using alternative methods (major depression).

The comparability of the seven cohorts was evaluated by estimating the common variant genetic correlations between them. The meta-analysis identified 44 independent loci that were statistically significant ($P<5\times10^{-8}$). Of these 44 loci, 30 are novel and 14 were significant in a prior study of MDD or depressive symptoms. Estimation of h2SNP was 8.7% (SE 0.004, liability scale, assuming lifetime risk 0.15), that this is about a quarter of h^2 estimated from twin or family studies (Sullivan, Neale, and Kendler 2000) . The authors used partitioned LD score regression (Finucane et al. 2015) to evaluate the enrichment of the major depression GWA findings in over 50 functional genomic annotations and found enrichment in regulatory activity, open chromatin in human brain and an epigenetic mark of active enhancers (H3K4me1).

ASD GWAS (Grove et al. 2019)

Imputation was performed using the 1000 Genomes Project phase 3 reference panel. Genotypes from 13,076 cases and 22,664 controls from the iPSYCH sample were included in the analysis. The primary analysis was a meta-analysis of the iPSYCH ASD results with five family-based trio samples of European ancestry from the Psychiatric Genomics Consortium (PGC; 5,305 cases and 5,305 pseudo controls). The main GWAS meta-analysis totalled 18,381 ASD cases and 27,969 controls, and applied an inverse variance-weighted fixed effects model. The SNP heritability (h2G) was estimated to be 0.118 (SE = 0.010), for a population prevalence of 0.012. The stratified LD score regression analysis (Finucane et al. 2015)identified significant enrichment of heritability in in brain and neuronal cell lines. The highest enrichment was observed in the developing brain, germinal matrix, cortex-derived neurospheres, and embryonic stem cell (ESC)-derived neurons, consistent with ASD as a neurodevelopmental disorder with largely prenatal origins.

All SNPs associated with neuropsychiatric disorders identified to date have low effect sizes (odds ratios typically < 1.1), consistent with individual common genetic variants conferring only small increases in susceptibility to these conditions. Indeed, it now appears that there exist thousands of common DNA variants conferring risk to neuropsychiatric disorders that are currently below the threshold of genome-wide significance. These can be captured through 'polygenic score' (PGS) methods. PRSs can be used to identify individuals at high risk of disease. This is achieved by combining hundreds of disease-associated variants carried by an individual into a single score that reflects their overall genetic risk (Chatterjee, Shi, and García-Closas 2016).

The integration of PGSs with epidemiological risk factors such as age, sex, smoking status, diet, or family history of disease could improve the stratification of individuals, potentially resulting in more effective clinical interventions (Torkamani, Wineinger, and Topol 2018). However, to date, the use of PGSs in the clinical setting remains limited, due to two main reasons. Firstly, prediction accuracy remains low. Secondly, PRSs are often based on European GWASs and their transferability between populations is low (A. R. Martin et al. 2019). However, polygenic risk scores represent the first quantitative biomarker of genetic liability and could be used to assess the validity of intermediate phenotypes (Riglin et al. 2017; Terwisscha van Scheltinga et al. 2013). Furthermore, PGSs today are a powerful tool to explore the genetic relationship between neuropsychiatric disorders. In 2009, the International Schizophrenia Consortium (I. S. Consortium 2009) showed a substantial polygenic component to risk of schizophrenia involving thousands of common alleles of very small effect (summed into a polygenic risk score) and showed that this component also contributes to risk of bipolar disorder, but not to multiple non-psychiatric diseases.

Substantial genetic overlap between various psychiatric disorders has been shown by several studies (B. Consortium et al. 2018; B. Bulik-Sullivan et al. 2015). Recently, the BrainStorm Consortium (B. Consortium et al. 2018) used genome-wide association data to quantify the degree of overlap for genetic risk factors of 25 common brain disorders. They found that common variant risk for psychiatric disorders was shown to correlate significantly, especially among attention deficit hyper-activity disorder, bipolar disorder, major depressive disorder, and schizophrenia.

The genetic correlation calculated using common SNPs (rg) is high between schizophrenia and bipolar disorder (0.68 ± 0.04 s.e.), moderate between schizophrenia and major depressive disorder (0.43 ± 0.06 s.e.), bipolar disorder and major depressive disorder (0.47 ± 0.06 s.e.), and ADHD and major depressive disorder (0.32 ± 0.07 s.e.), low between schizophrenia and ASD (0.16 ± 0.06 s.e.) (Lee et al. 2013).

1.2.4 Heritability

Elucidating the relative contribution of genetics versus environment for various diseases and disease-related complex traits can help better understand the causal mechanism of disease aetiology. A key quantity to evaluate the contribution of genetics versus environment is heritability, which measures the proportion of phenotypic variance explained by genetic factors. Two types of heritability can be estimated. The broad-sense heritability evaluates the proportion of phenotypic variance explained by all genetic factors, including additive effects, dominant effects, and epistasis effects. The narrow-sense heritability, on the other hand, evaluates the proportion of phenotypic variance explained by additive genetic effects.

The broad-sense heritability is estimated by the mean of twin studies. Twin studies compare the concordance of traits between monozygotic and dizygotic twins who are considered to share the same environment and known to share all and (on average) half of their genetic makeup, respectively. The difference in concordance is used to estimate the trait heritability. The broad-sense heritability of schizophrenia, ADHD, ASD and bipolar disorder is estimated at around 80%, while that of major depressive disorder has been estimated to be around 40% (Bray and O'Donovan 2018). It has to be noted that the twin study design makes several assumptions, most importantly that there is no greater environmental similarity of identical over non-identical twins.

Recently, large-scale GWASs, between apparently unrelated individuals, have identified many SNPs associated with various complex traits. However, the majority of identified SNPs only explain a small fraction of heritability for most traits, leading to a large fraction of unexplained heritability, commonly referred to "missing heritability" (Zuk et al. 2012). Many explanations have been proposed for missing heritability. For example, it is possible that rare variants with large effects contribute disproportionately to the phenotypic variance. In addition, twin studies may have overestimated heritability (H. Zhu and Zhou 2020). A prominent hypothesis suggests that current GWASs are underpowered and that many causal SNPs remain undetected below the stringent genome-wide significant threshold, which can vastly underestimate the phenotypic variance (Young 2019).

	Schizophrenia	Bipolar disorder	Major depressive disorder	ASD	ADHD
SNP heritability (s.e.)	0.23 (0.008)	0.25 (0.012)	0.21 (0.021)	0.17 (0.025)	0.28 (0.023)
Heritability estimated from twin studies	0.81	0.75	0.37	0.8	0.75

Table 1.1. Adapted from Lee et al. 2013). Compares heritability estimated from twin studies and SNP heritability for schizophrenia, bipolar disorder, major depressive disorder, autism and ADHD.

The SNP-heritability estimates for each disorder—schizophrenia, 0.23 (0.01 s.e.), bipolar disorder, 0.25 (0.01 s.e.), major depressive disorder, 0.21 (0.02), ASD, 0.17 (0.02 s.e.) and ADHD, 0.28 (0.02 s.e.)—are considerably less than the heritabilities estimated from twin studies (Table 1.1, (Lee et al. 2013). SNP-based values are a lower bound for narrow-sense heritability because they exclude contributions from some causal variants (mostly rare variants) not associated with common SNPs.

Partitioned heritability
A number of statistical methods have been developed to estimate the SNP heritability (the amount of phenotypic variance explained by SNPs) of a trait using either individual genotypes or summary statistics (Yang et al. 2010; B. K. Bulik-Sullivan et al. 2015) from GWAS.

Previous studies (Thurman et al. 2012; Trynka and Raychaudhuri 2013) have shown that heritability of complex disorders does not distribute uniformly across the whole genome; instead, different functional parts of the genome contribute disproportionally to the heritability. This gave rise to partitioning heritability approaches, such as stratified LD-score regression (SLDSC, Finucane et al. 2015), Genome-wide Complex Trait Analysis (GCTA, Yang et al. 2010) and Linkage disequilibrium adjusted kinships (LDAK, (Speed et al. 2012)) to estimate the proportion of phenotypic variance explained by all genome-wide SNPs for complex traits and test for a significant accumulation of trait heritability in different functional categories of the genome.

In the GCTA Model, all SNPs are expected to contribute equally heritability, therefore using the GCTA Model will result in biased estimates if causal variants are predominantly in regions of high or low linkage disequilibrium (LD). LD is defined as a non-random association of alleles (i.e. alternative genetic variants at the same genomic position) between different genomic positions (Slatkin, 2008). The importance of accounting for LD-dependent genetic architectures in analyses of heritability is now recognised (Gazal et al. 2017). Both LDAK and SLDSC models incorporate LD-dependent architectures, however the first LDAK Model did not take into account minor allele frequency (MAF). More recent version of the software do allow to account for MAF (Speed et al. 2017), however, the SLDSC is a more widely used and largely recommended method (Gazal et al. 2019).

SLDSC can be used to examine if some regions of the genome (e.g. regulatory regions operating in a particular tissue) are enriched for the SNP heritability of a trait, which can improve understanding of the genetic architecture and biology of complex disorders. Different genomic regions are defined as functional categories, including 24 non-specific annotations (coding regions, promoters, enhancers, introns, conserved elements and DHSs, among others) as well as histone modification profiles acquired from a variety of cell types, and these can provide a background control for assessment of user-defined functional annotations. The SNP heritability of variants in each category is defined as an enrichment score as the proportion of SNP heritability in a category divided by the proportion of SNPs in that category (Finucane et al. 2015).

One of the major advantages of LDSC, is that LDSC does not require any individual-level phenotype-genotype data. Instead, it uses only summary association statistics from the GWAS and an external LD reference panel to estimate heritability. However, an external LD

reference panel has to match to the population used in the GWAS, and a mismatch between LD estimates and the GWAS sample can bias LDSC estimates (Bulik-Sullivan et al., 2015). Summary-level data has three great advantages over the individual-level data. First, individual-level data is sensitive and privacy concerns often limit access to it, whereas from the summary-level data no single individual can be identified. Second, to increase power of the GWAS, many of the largest studies are conducted as a meta-analysis, where summary association statistics from multiple separate study cohorts are jointly analysed, and access to the original individual-level data is therefore usually not possible. Third, summary-level data is more compact and reduces the computational burden massively compared to individual-level data.

To partition heritability, SLDSC requires a GWAS summary statistics file and an annotation file that contains genomic coordinates for the functional partition. SLDSC exploits the predicted relationship that exists between the association statistics of a set of GWAS index SNPs and local linkage disequilibrium surrounding each index SNP for a polygenic trait, i.e. this relationship predicts that, on average, index SNPs with high LD r^2 scores are more likely to tag SNPs with higher association statistics than SNPs with low r^2 scores. As such, using a linear regression model, SLDSC generates an estimate of the proportion of SNP heritability that is captured by a given set of index SNPs, including any heritability explained by non-assayed SNPs within the haplotype block that each index SNP tags, as well as correcting for systematic biases.

1.3 Non-coding DNA and regulatory regions

It has to be noted that the majority of GWAS risk loci are located within non-coding regions and do not index variants in coding sequence which might impact on protein structure. For

example, in the seminal 2014 schizophrenia GWAS (S. W. G. of the P. G. Consortium 2014), of the 108 loci found associated with the disorder, only 10 potentially index coding SNPs changing amino acid sequence. It is therefore likely that these SNPs index genetic risk variants that affect gene regulation.

Non-coding DNA constitutes more than 98% of the human genome but is poorly characterised compared to protein coding sequence. The first major effort to map functional regions of the non-coding human genome was provided by the ENCODE (Encyclopedia of DNA elements) Consortium (Feingold et al. 2004), who used a variety of functional genomic technologies (e.g. RNA sequencing, chromatin immunoprecipitation followed by sequencing [ChIP-seq] and DNase-hypersensitivity I site [DHS] sequencing) to map functional regions of the genome in 147 human cell types. This was followed by the Roadmap Epigenomic Mapping Consortium (REMC) who released epigenomic data from 111 diverse human cell types and tissues in 2015 (Roadmap Epigenomics Consortium et al. 2015). Recently, the PsychENCODE Consortium have produced similar data from human neural cells and tissues (Wang et al. 2018). Sequences that are known to be important for gene regulation, and the technologies that have been used to detect them, are described in this section.

It is now known that a large proportion of non-coding DNA is represented by cis-regulatory elements (CRE), such as promoters, enhancers, insulators and silencers. These orchestrate the complex mechanisms of gene expression in a tissue- / cell- and time-specific manner. Transcription of a gene-coding DNA sequence into messenger RNA (mRNA) begins with the RNA Polymerase II (Pol II) binding to the promoter region, situated proximally to the gene they influence, at the transcription start site (TSS). All genes have at least one promoter

situated upstream of the 5' end of the gene; however, over 50% of genes have multiple, alternative promoters (Kimura et al. 2006; Cooper et al. 2006), which can produce different protein isoforms by producing different mRNA transcripts (Landry, Mager, and Wilhelm 2003; Xin, Hu, and Kong 2008).

The activity of Pol II is weak in the absence of enhancers. Located distally to the genes they regulate, enhancers increase the rate of expression of their target genes (Spitz and Furlong 2012). They interact with transcription start sites of promoters through chromatin looping, a biophysical phenomenon mediated by transcriptional co-activators and structural proteins, that allows sections of genomic sequence to form loops and bring distant regulatory elements into close spatial proximity with their target gene(s) (figure 1.2). Given these long-range interactions, it is difficult to accurately predict the genes that enhancers target, particularly as a single enhancer can modulate the activity of multiple genes. Furthermore, several enhancers can operate simultaneously on the same gene and clusters of enhancers have also been identified, namely *'super-enhancers'.*
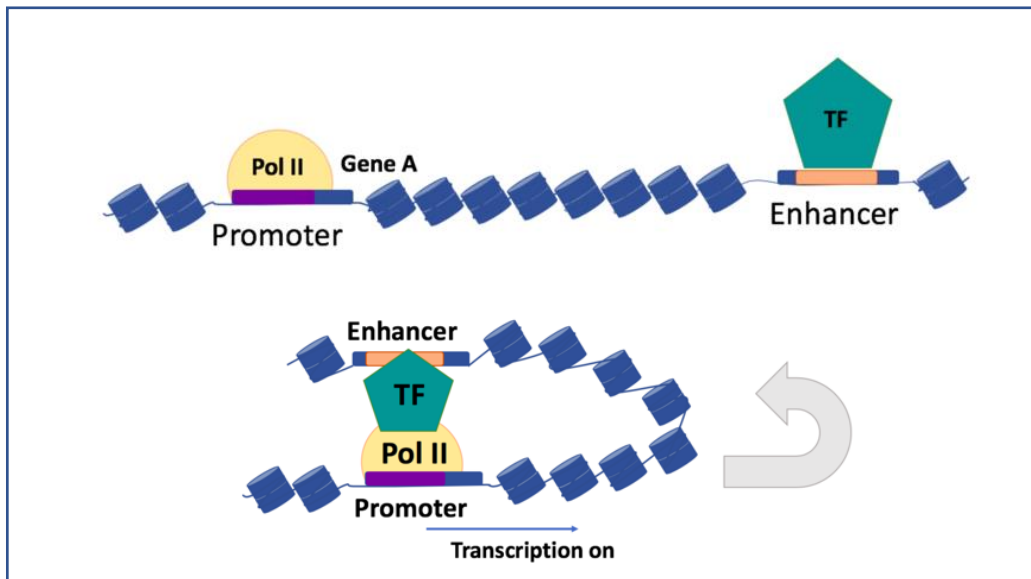
Figure 1.2. Diagrammatic representation of promoter – enhancer interaction by looping. The activity of Pol II is weak in the absence of enhancers. Located distally to the genes they regulate, enhancers increase the rate of expression of their target genes

CREs are defined by chromatin state, histone post-translational modifications and transcription factor binding, known to be associated with transcriptional regulatory activity.

When a CRE is inactive, the DNA containing it is wrapped around histone proteins into nucleosomes, forming heterochromatin (or so-called 'closed' chromatin). Some nuclear proteins (pioneer transcription factors, TFs) are able to access the DNA strand by opening up the chromatin, making the DNA accessible to other TFs. TFs can also recruit other regulatory proteins, including co-activators and co-repressors. Nucleosome-free or 'open' chromatin is therefore characteristic of an active regulatory region.

1.4 Features of regulatory regions

1.4.1 Open chromatin

Chromatin is a complex of DNA, RNA and protein that is located in the nucleus of all human cell types, and it is the main constituent of chromosomes. The core structural element of chromatin is the nucleosome, and each nucleosome consists of a 147 base pair section of DNA sequence that is coiled around 8 histone proteins. Nucleosomes are connected to one another via a short (~60bp) sections of DNA, called linker DNA, to form a repeating structure that resembles beads on a string. Each histone protein octamer is composed of two copies of the H2A, H2B, H3 and H4 histones, and these are bound on their external surface by a H1 linker histone (Luger et al. 1997; Robinson and Rhodes 2006). During cellular mitosis, chromatin tightly condenses such that it can be visualised microscopically. In order to regulate cellular gene expression, chromatin dynamically alters its physical state. Euchromatin, is relaxed, uncoiled, chromatin where histones are either well-spaced, or displaced entirely, exposing DNA to molecules (e.g. transcription factors; TFs) that regulate gene expression. By contrast, heterochromatin is tightly compacted, and histones are bunched together making DNA relatively inaccessible. Chromatin accessibility is a measure of the functional state of chromatin and is defined by the degree to which DNA is exposed. In any given cell type, accessible DNA constitutes ~2-3% of the entire genome (Thurman et al. 2012). In order for a gene regulatory event to take place chromatin must be uncoiled and in the euchromatic state to allow regulatory molecules to interact with DNA. When chromatin is heterochromatic most regulatory molecules cannot get access to DNA due to phenomena such as steric hindrance (Allis and Jenuwein 2016).

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a popular method used to measure genome-wide chromatin accessibility (J D Buenrostro et al. 2013). This method takes advantage of the fact that euchromatin is sensitive to enzymatic cleavage. In the ATAC-seq assay, nuclei are extracted from a cell of interest and incubated with a hyperactive Tn5 transpose enzyme, which selectively excises exposed DNA from all the open chromatin regions of the genome. Crucially, the enzyme has been modified to carry adapter sequences such that, whilst excising the DNA, adapters are simultaneously inserted at the ends of the DNA fragments. These adapters act as barcodes for the DNA fragments so they can be recognised after pooled sequencing. Due to the high efficiency of the Tn5 enzyme, ATAC-seq has largely superseded alternative methods for measuring open chromatin regions, such as DNAse-seq (Klemm, Shipony, and Greenleaf 2019), the micrococcal nuclease digestion with sequencing (MNase-Seq, (Schones et al. 2008) or the Formaldehyde-assisted identification of regulatory elements (FAIRE) followed by sequencing is also an open chromatin profiling method (Giresi et al. 2007), that uses crosslinking with formaldehyde of nucleosome-bound DNA. Once the DNA has been excised and barcoded, the DNA fragments are sequenced and the reads produced are aligned to a reference genome. The final read out of an ATAC-seq assay is a genome-wide set of 'peaks', representing sites of open chromatin that are potentially involved in the regulation of gene expression in the cell type of interest at the time the nuclei were extracted. Through deep sequencing of ATAC-Seq libraries, footprints left by nuclear proteins and transcription factors (TFs) bound to nucleosome-free promoters or enhancers can also be detected (Z. Li et al. 2019; Hesselberth et al. 2009; Karabacak Calviello et al. 2019). Machine learning methods can predict TF binding sites from a DNase-Seq or ATAC-Seq experiment based on the magnitude and shape of the TF footprint, that reflect the biophysics of the specific TF (Sherwood et al. 2014).

Open chromatin regions may be located either near the transcription start sites of genes or in non-coding regions located far from genes. This provides information on the genes that are likely to be active, or poised in preparation to be activated, within a cell type and the regulatory elements that have the potential to influence gene expression in these cells. It is also possible to identify specific DNA recognition sequences that gene regulatory proteins bind to within these loci. This provides a picture of the DNA-protein interactions that regulate gene expression in the cell of interest. Furthermore, when integrating chromatin accessibility data with other functional genomic data it is possible to make functional predictions of the pathologically relevant effects that genetic variation may have in particular tissues or cell types.
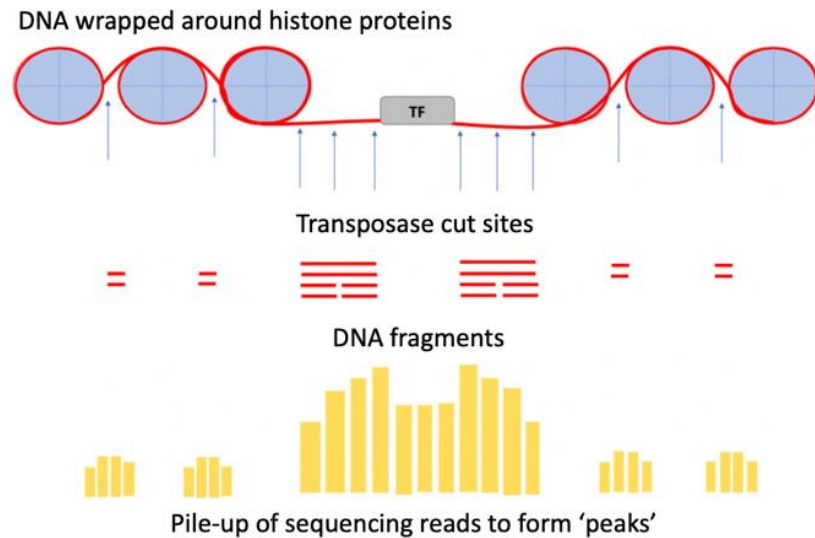
Figure 1.3. Principle of the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-Seq).
DNA at nucleosome-free, accessible chromatin regions, is cut by the hyperactive transposase Tn5 loaded with next generation sequencing adapters. Resulting tagged DNA fragments, sequenced and aligned to the human genome, pile up forming peaks at open chromatin sites. TF: transcription factor

1.4.2 Transcription factor binding

TF binding sites, 6-10bp long DNA sequences located within CREs, can be identified both computationally and experimentally. Sequence patterns preferentially bound by TFs are named TF motifs. They allow variation in nucleotides at some binding site positions but not others. Computational methods use enrichment of DNA sequences in putative promoters of co-expressed genes to identify motifs, as well as highly conserved regions between closely related species (Tompa et al. 2005). Whole genomes can then be scanned for TF motif matches.

High-throughput laboratory approaches make use of next generation sequencing technology to perform genome-scale analysis of epigenetic markers and chromatin states. In a chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiment, a TF is chemically crosslinked to its binding sites (Park 2009). Genomic DNA is then fragmented and precipitated with an antibody targeting the TF and TF binding sites fragments are then sequenced. Transcription factor sequence motifs are collected in databases such as TRANSFAC, JASPAR and UniPROBE.

1.4.3 Histone modification

Histone modifications are post-translational modifications that are typically covalently bound to the freely protruding N-terminus tail of histone proteins. They act dynamically by altering the structure and function of chromatin by direct interaction with DNA, or by recruitment of chromatin remodelling proteins, to influence gene expression. Histone modifications can be mapped to functionally annotate the genome and to predict the nature and activation status of the regulatory elements that they are proximal to (Kouzarides 2007; Tessarz and Kouzarides 2014).

ChIP-seq technology (figure 1.4) can be used to scan the genome for histone marks that are indicative of regulatory activity. Two histone modifications commonly used to annotate the human genome are methylation of histone 3 lysine 4 (H3K4Me1 / H3K4Me3) and acetylation of histone 3 lysine 27 (H3K27ac). Histone methylation is a reversible process that is facilitated by histone methyltransferases, which attach methyl groups to histones, and demethylases, which detach methyl groups. H3K4 trimethylation (H3K4me3) and H3K4 monomethylation

(H3K4me1), are associated with active transcription at promoters and enhancers respectively (Guenther et al. 2007; Bernstein et al. 2002; Heintzman et al. 2007). Repressed promoters and enhancers are both flanked by H3K27 trimethylation (H3K27me3), while H3K27 acetylation (H3K7ac) strongly correlates with active promoters and enhancers.

Computational and benchtop methods are also often combined to detect histone modifications. Machine learning has been used to integrate motif matching with binding sequences detected from experimental methods to predict novel enhancers (Gorkin et al. 2012).
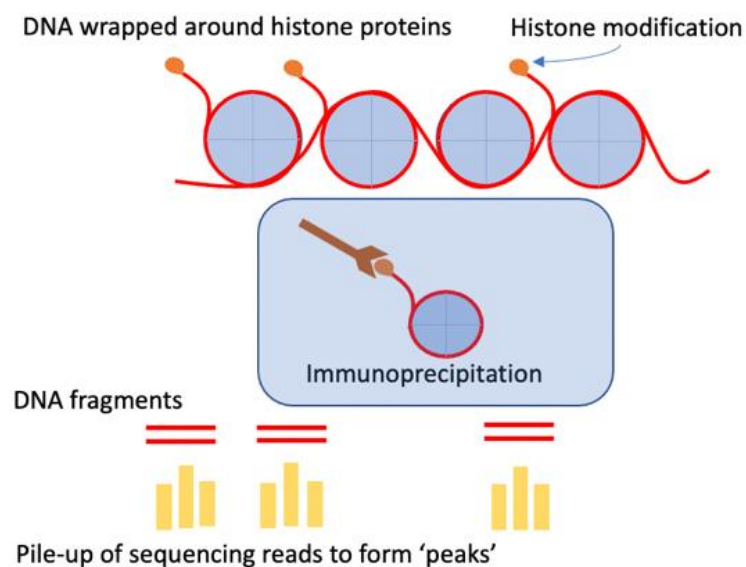
Figure 1.4. Diagrammatic representation of a chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiment. Histones are chemically crosslinked to DNA. DNA is then fragmented and precipitated with an antibody targeting the histone modification. Resulting DNA fragments, sequenced and aligned to the human genome, pile up forming 'peaks'.

1.4.4 DNA methylation

DNA methylation at promoter regions is associated with repressed gene expression (Suzuki and Bird 2008). DNA methylation is the most extensively studied epigenetic modification. It consists in the addition of a methyl or hydroxymethyl group to carbon five of the cytosine pyrimidine ring and is mainly found in the context of CG dinucleotides. It inhibits transcription by recruiting chromatin remodeler proteins, such as *MECP2*, that condense the chromatin into an inactive state. Although generally considered a repressive marker, evidence suggests that methylation within a gene body promotes gene expression (Ball et al. 2009). A variety of methods are available for DNA methylation analysis. The gold standard is considered to be bisulfite sequencing, a whole-genome single nucleotide resolution method (Kurdyukov and Bullock 2016). However due to its prohibitive costs, alternative methods measuring methylation levels at targeted loci such as BeadChip technologies are more widely used.

It has to be noted that no specific epigenetic marker or chromatin state has been demonstrated to consistently correlate with CRE activity (Klemm, Shipony, and Greenleaf 2019). Multi-level gene regulation is likely to be the norm and therefore multi-level approaches are needed to disentangle the regulatory landscape of different cells and tissues.

1.5 The importance of brain development

Given that regulatory mechanisms are known to exert gene expression changes in a tissue-type and time manner, investigation of the underlying molecular processes in disease-relevant tissue/cell types and time points is key to a better understanding of disease molecular pathology.

The main organ relevant to psychiatric disorders is assumed to be the brain. The human brain is the most complex known biological system and it undergoes enormous changes throughout life, with the largest changes taking place in early development (Dillman and Cookson 2014).

Around the third gestational week, the human brain starts to develop (Stiles and Jernigan 2010). Progenitor cells differentiate, migrate and mature into different cell types, organized into specialized circuits and layers. From early stages of development, neural morphology and function is dictated by extremely complex molecular processes. For these processes to take place, different proteins must be produced in different cells, at the right time, and for that to happen, gene expression must be meticulously orchestrated (Miller et al. 2014).

There is accumulating evidence to suggest that a proportion of both environmental and genetic risk factors for neuropsychiatric disorders operate during brain development, by disrupting the complex molecular processes underlying normal neural morphology and function (Brown 2012). Although only few psychiatric disorders are considered strictly developmental by the DSM-V (i.e., attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD)), with their age of onset in childhood or adolescence, emerging evidence indicates that brain insults occurring in early stages of life may predispose individuals to developing a larger range of mental health illnesses. The neurodevelopmental hypothesis of schizophrenia (Murray and Lewis 1987; Weinberger 1987) is widely accepted and evidence supporting a developmental component to bipolar disorder (BD) as well as major depressive disorder (MDD) is also starting to emerge (Hall et al. 2020; O'Brien et al. 2018; Kathuria et al. 2020; Wray et al. 2018).

1.5.1 The neurodevelopmental hypothesis of psychiatric disorders

The neurodevelopmental hypothesis of schizophrenia (Murray and Lewis 1987; Weinberger 1987) was proposed on the basis of a number of observations. These include: presence of structural brain abnormalities (such as enlarged ventricles) at the onset of illness; an absence of evidence for neurodegeneration in post-mortem studies; the frequent occurrence at a young age of premorbid deficits and early clinical abnormalities (such as neuromotor and speech delays, as well as morphological abnormalities) in children that would later develop psychotic symptoms; the higher incidence of maternal infections and obstetric complications in people who later develop schizophrenia; studies of primates that showed that neonatal lesions can have delayed effects on behaviour. More recently, the idea that adult-onset disorders could have their origins in early development has also been supported by reports of deficits in developmental connectivity in iPSC-derived neurons from people with schizophrenia compared with controls. For example, Brennand and colleagues reported an iPSC-based case/control study on familial schizophrenia which showed lower connectivity in schizophrenia iPSC-neurons, which also possessed fewer neurites and expressed lower levels of synaptic markers (Brennand et al. 2011).

According to this now widely accepted model, perturbation in early brain development leads to altered brain wiring that results in a vulnerable phenotype, sensitive to molecular changes associated with environmental insults as well as maturational changes, such as synaptic pruning and myelination within the frontal cortex, that occur in adolescence. Insults caused by both genetic and environmental risk factors, therefore, occur long before the first symptoms manifest as a result of decompensation in response to the increasing demand in

brain function. These brain perturbations that have a delayed impact on cortical function are thought to be less severe compared to the neural abnormalities responsible for early-onset neuropsychiatric disorders, like ASD (Owen and O'Donovan 2017).

As mentioned in section 1.5, neurodevelopmental disorders as grouped by DSM-5 (American Psychiatric Association 2013) only include autism, ADHD, learning, communication and motor disorders, and intellectual disability, and are defined as having an onset in the early developmental period, typically in early childhood. While categories can be useful for clinical decision making, psychiatric disorders appear to lie at the extreme of continuously distributed dimensions (J. Martin, Taylor, and Lichtenstein 2018).

There is in fact a need for clinicians and scientists to adopt a developmental perspective in clinical practice and research when dealing with all neuropsychiatric disorders, including adult-onset disorders like bipolar and major depression (Thapar and Riglin 2020). For example, the presence of associated structural abnormalities at birth or that become apparent during the early postnatal period. Genome-wide association studies (GWASs) have identified a number of genetic loci associated with BD (Stahl et al., 2019), including genes involved in neurodevelopment, such as NCAN, a proteoglycan involved in cell adhesion and neuronal migration, processes that are pivotal during neurodevelopment (O'Shea and McInnis 2016). Kathuria et al. (2020) used iPSCs from patients with BD to grow cerebral organoids ex vivo and investigate the disease biology of bipolar disorder. Analyses of transcriptomic data from the iPSC-derived cerebral organoids showed downregulation of genes involved in neurodevelopmental pathways. Furthermore, RNA-seq data comparing gene expression profiles in the organoids showed downregulation of pathways involved in neurodevelopment.

Evidence for a developmental component to depression has also started to emerge. Epidemiological studies have shown a number of prenatal perinatal, and postnatal characteristics associated with an increased risk of depression in offspring, including low birth weight, premature birth, small gestational age, maternal education, maternal smoking, paternal smoking, maternal stress, maternal anxiety, and prenatal depression (Su, D'Arcy, and Meng 2020). Furthermore, GWAS of depression implicated the *TCF4* (transcription factor 4) locus, also, strongly associated with schizophrenia, that is essential for normal brain development.

1.5.2 Gene expression in the foetal brain

In the last decade, several studies have assayed the transcriptome of different regions of the prenatal human brain across the lifespan. These studies show that the largest transcriptional changes occur early in foetal life.

In support of a prenatal component to schizophrenia, Gulsuner and colleagues (2013) found an enrichment of damaging *de novo* mutations in the disorder within neurogenesis-related gene co-expression networks (i.e. those involved in neuronal migration, synaptic transmission, signalling, transcriptional regulation, and transport) in the foetal brain (Gulsuner et al. 2013). Similarly, ASD risk genes mapped onto developmental co-expression networks show enrichment in modules implicating early transcriptional regulation and synaptic development (Parikshak et al. 2013).

Recently, lower sequencing costs and improved technologies have allowed production of single-cell resolution gene expression data. Single-cell RNA-seq from human prefrontal cortex tissue (gestational weeks 8 to 26) identified clusters of gene expression that correspond to six major cell-types (neuronal progenitor cells, excitatory neurons, interneurons, astrocytes, oligodendrocyte progenitor cells and microglia) and 35 subtypes within the foetal brain (Zhong et al. 2018). This and similar studies (Fan et al. 2020; Zeng et al. 2012; Darmanis et al. 2015) highlight the complexity of cell types in the prenatal brain and provide a valuable resource for exploring potential cell-specific effects of rare coding variation associated with neuropsychiatric disorders (Polioudakis et al. 2019).

1.5.3 Gene regulation in the foetal brain

Although gene expression changes in human foetal brain development have been extensively explored, the underlying regulatory genomic mechanisms were, at the start of this PhD, largely uncharacterised. The foetal brain is represented in the Roadmap Epigenomic Mapping Consortium (REMC) data but only for limited developmental time points and not in a cell-type specific manner (https://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/?display=100&search=fetal%20 brain&sort=acc). REMC data include bulk whole foetal brain open chromatin (DNase-seq), histone modifications, and methylation analysis. These foetal brain open chromatin maps (Roadmap Epigenomics Consortium et al. 2015) have been used to test for enrichment of disease-associated risk loci. For example, a significant two-fold enrichment of de novo mutations in foetal brain-active gene regulatory elements was found in developmental disorders (Short et al. 2018).

Maps of regulatory relationships in the developing human brain were produced by Won and colleagues (Won et al. 2016). Hi-C libraries were constructed from two zones of 17 and 18 post-conception week foetal brain tissue: the cortical plate and germinal zone, the former containing mainly mature neurons, the latter mostly formed of neuronal progenitors. This data was integrated with non-coding schizophrenia risk variants identified by GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014) identifying a number of potential schizophrenia risk genes. Identified genes were enriched for postsynaptic density, acetylcholine receptors, neuronal differentiation and chromatin remodelers, pathways consistent with brain development and function. Furthermore, de la Torre Ubieta and colleagues used ATAC-seq to map open chromatin in the same mid-gestational brain regions, the cortical plate and in the germinal zone, and found an enrichment of SNP heritability for schizophrenia, ADHD, depressive symptoms, neuroticism, educational attainment and intracranial volume in sites that were preferentially accessible in the germinal zone (de la Torre-Ubieta et al. 2018).

Recently, quantitative trait locus (QTL) studies have been successful in linking genotypic data with molecular markers, such as DNA methylation (mQTL), gene expression (eQTL) and alternative splicing (sQTL), in the foetal brain. Hannon and colleagues characterized DNA methylation quantitative trait loci (mQTLs) in human foetal brain samples from the first and second trimester of gestation, showing that variants associated with differential methylation are enriched for SNPs associated with schizophrenia at genome-wide significance (Hannon et al. 2016). O'Brien and colleagues (O'Brien et al. 2018) produced an eQTL dataset derived from the 120 second trimester human foetal brains and found that foetal eQTL are enriched among

risk variants for neuropsychiatric disorders including attention deficit hyperactivity disorder, schizophrenia, and bipolar disorder. In a similar, larger study, including 201 mid-gestational human brains, Walker and colleagues (Walker et al. 2019) identified eQTLs and splice eQTLs (sQTLs) specific to human prenatal brain development and showed that genetic risk for ADHD implicates prenatal eQTL and sQTL, while ASD risk showed a trending enrichment in sQTL.

1.6 Use of functional genomics to interpret genetic findings for neuropsychiatric disorders

It has to be noted that the most significant SNP at each GWAS risk locus (the 'index SNP') is not necessarily the causal variant. Due to linkage disequilibrium (LD), alleles at genetic variants located in close proximity to each other are often associated within a population. Further studies are therefore necessary to distinguish the functional SNP from correlated variants in high LD.

Establishing causality and elucidating the downstream molecular mechanisms of risk SNPs are the next challenges. Many GWAS causal variants have been proposed to influence disease risk by altering the function of cell type-specific regulatory elements, with ensuing changes in target gene expression. This hypothesis is supported by the overlap of risk SNPs with eQTLs (O'Brien et al. 2018). Moreover, the cell type in which the eQTL effect is observed often matches cell types already thought to be relevant to the disease in question (Lu et al. 2017).

In this context, variants may disrupt TF binding, promoter activity or enhancer function. Although there is no single straightforward strategy for functional studies, prioritization of candidate causal SNPs, using a combination of chromatin state, epigenomic marks and

computational methods, may represent a first step towards functional dissection of disease-associated loci. This multi-level approach has been shown to be successful at guiding downstream investigation (Huo et al. 2019).

1.7. Aims of this thesis

Given that several neuropsychiatric disorders are thought to have a neurodevelopmental origin (Weinberger 1987; Murray and Lewis 1987) and that gene regulation appears to play a key role in their aetiology, investigation of regulatory elements in the foetal brain is important to elucidate the biological basis of these conditions.

The aims of this thesis were:

1. To map open chromatin regions in the foetal frontal cortex and to test for enrichment of SNP heritability for neuropsychiatric disorders within these regions {Chapter 2}

2. To integrate open chromatin data with other epigenomic data from foetal brain to further explore enrichment of SNP heritability for neuropsychiatric disorders in active regulatory regions {Chapter 3}

3. To map open chromatin regions in neuronal and non-neuronal cell populations in the foetal frontal cortex and to test for enrichment of SNP heritability for neuropsychiatric disorders within these cell population-specific regions {Chapter 4}

4. To identify proxies for top schizophrenia GWAS SNPs within identified open chromatin regions of the foetal brain that may represent the underlying functional variants {Chapter 5}

Chapter 2. Chromatin accessibility in the human foetal brain

2.1 Introduction

The first objective of this chapter was to generate maps of 'open' or 'accessible' chromatin regions (OCRs) in the foetal frontal cortex using the assay for transposase accessible chromatin followed by sequencing (ATAC-Seq). This method provides an effective means of discovering gene regulatory elements through the identification of transposase-accessible DNA regions, visualised as open chromatin peaks (J D Buenrostro et al. 2013) (Figure 2.1).

Given that regulatory regions are believed to contain a large proportion of the common genetic component of complex traits, the second objective was to estimate the contribution of foetal brain gene regulatory elements to the heritability of psychiatric disorders.

I hypothesised that foetal brain open chromatin regions may be enriched in common genetic variants for a range of psychiatric disorders. In order to test this hypothesis, I used the stratified Linkage Disequilibrium Score Regression analysis for partitioning heritability (LDSC), (Finucane et al. 2015)), a widely used statistical method that exploits summary statistics from genome wide association studies (GWAS) to estimate the contribution of functional annotations to single nucleotide polymorphism heritability while accounting for linkage disequilibrium (LD).

The experimental procedure of ATAC is well-documented in the literature (J D Buenrostro et al. 2013). However, the efficiency of the transposase reaction may vary between primary tissues studied, due to factors including differences in membrane permeability and chromatin

state. To ensure consistent high-quality ATAC data, I therefore optimised the nuclei preparation and DNA library amplification protocol, as well as the post-sequencing bioinformatic pipeline.

Initially, snap frozen foetal brain tissue was used for this purpose. However, the resulting data showed a poor signal-to-noise ratio that compromised the peak calling process. These preliminary results suggested that the snap freezing method had caused damage to the chromatin structure. For this reason, fresh foetal brain tissue was acquired and dounce homogenised, and the derived cell suspensions were cryopreserved for subsequent ATAC-Seq investigation. The resulting data showed a much-improved signal-to-noise that allowed me to define high confidence open chromatin peaks.

In this chapter, I will first briefly discuss optimisation of the protocol. I will then describe the final experimental and bioinformatic methods I used to identify high confidence open chromatin sites in the human foetal frontal cortex, annotate these regions to the genome and test for their overlap with publicly available ATAC-Seq adult and foetal data. I will then illustrate and discuss results from the enrichment testing of single nucleotide polymorphism (SNP) heritability for 5 major neuropsychiatric disorders (ADHD, autism spectrum disorder [ASD], bipolar disorder, major depressive disorder and schizophrenia) within those regions.

2.2 Methods optimisation

2.2.1 ATAC-seq on snap frozen tissue

Two second trimester human foetal brain samples were available, one of 14 weeks, one of 17 weeks post conception. Samples were acquired from the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) (http://www.hdbr.org/). Tissue from 5 different brain regions (frontal cortex, striatum, thalamus, amygdala, hippocampus) was processed.

The ATAC library preparation was initially performed as follows. Between 20 and 150 mg of frozen brain tissue were homogenized in Phosphate Buffered Saline (PBS) using an ice-cold Dounce tissue homogenizer. After centrifugation, the cell pellet was re-suspended in PBS. Cells were then counted using an haemocytometer under an optical microscope and a total of 50,000 cells were collected and further centrifuged. The cell pellet was then re-suspended in cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL cat#CA-630) and the cell lysate was passed through a 21-gauge needle, before and after a 30 minute incubation on ice.

The preliminary results obtained from frozen brain tissue showed open chromatin peaks at the expected genomic regions (Figure 1). However, they showed a considerable background noise that compromised the peak calling process. To increase the signal-to-noise ratio, fresh foetal brain samples were obtained and used for the final analysis.
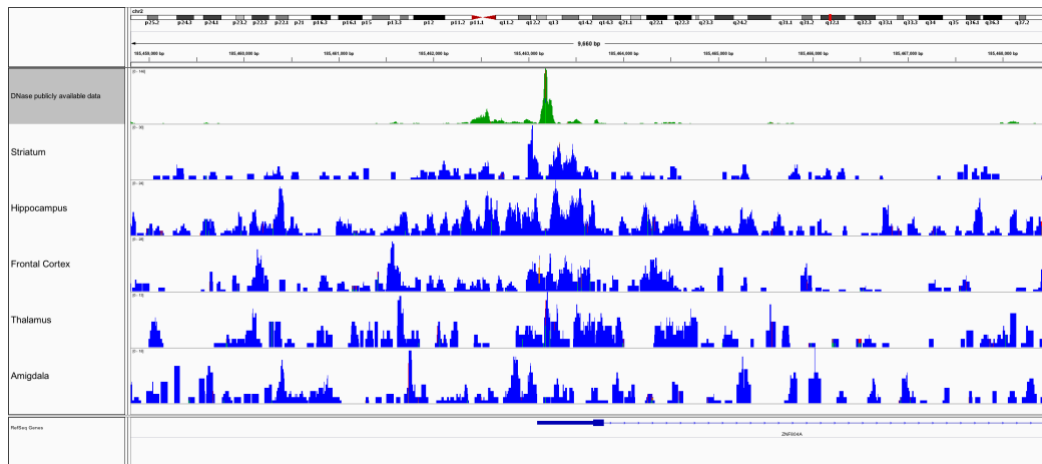
Figure 1. IGV image showing open chromatin peaks at the promoter region of ZNF804A. In green: DNase I Hypersensitivity publicly available (REMC) BAM alignments from 21 weeks foetal brain, brain region not specified. In blue: ATAC-seq BAM alignments (filtered for mitochondrial reads and PCR duplicates).

2.2.1 PCR optimisation

qPCR was performed to determine the number of thermal cycles needed for amplification of the DNA library. A total volume of 20 µl qPCR reaction was set-up (10 µl of 2X NEBNext Q5 Hot Start HiFi Master Mix, cat# M05435; 1 µl 10.000X SYBR green Nucleic acid gel, Invitrogen, cat#S7563; 0.65 µl of forward primer; 0.65 µl of reverse primer; 2.7 µl of MilliQ water; 5 µl of purified transposed DNA fragments). 40 thermal cycles were performed.

After amplification, the amplification plot was analysed on a linear graph. To calculate the number of additional cycles needed, the cycle number that corresponded to one third of the maximum fluorescent intensity was plotted, as described in figure 2. qPCR and analysis of the amplification plot showed that an ATAC reaction using 50,000 cells as starting material required a total of 15 PCR cycles.
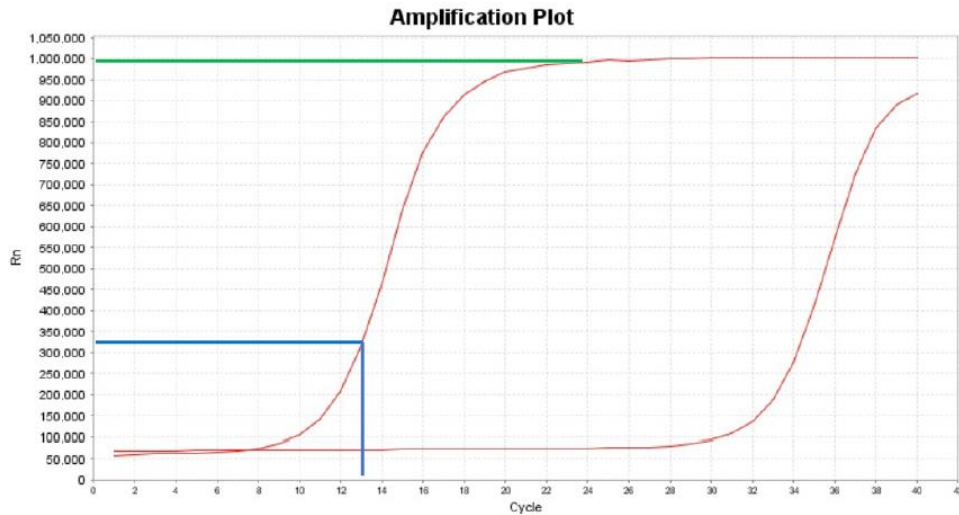
Figure 2.2. qPCR amplification plot. The green line indicates the maximum fluorescence intensity and the blue line plots the cycle number that corresponds to one third of the maximum fluorescent intensity.

Amplified library fragment sizes were analysed using Bioanalyzer and High Sensitivity DNA kit (cat# 5067-4626). To optimise sequencing on Illumina Hi-Seq 4000, that preferentially sequences amplicons <400 bp long, library fragments were firstly amplified by 8 cycles of PCR, then amplicons of 175 -250bp were size-selected using agarose gel electrophoresis on BluePippin 2% gel cassettes (Sage Science) and further amplified through 7 additional PCR cycles (see figure 3).
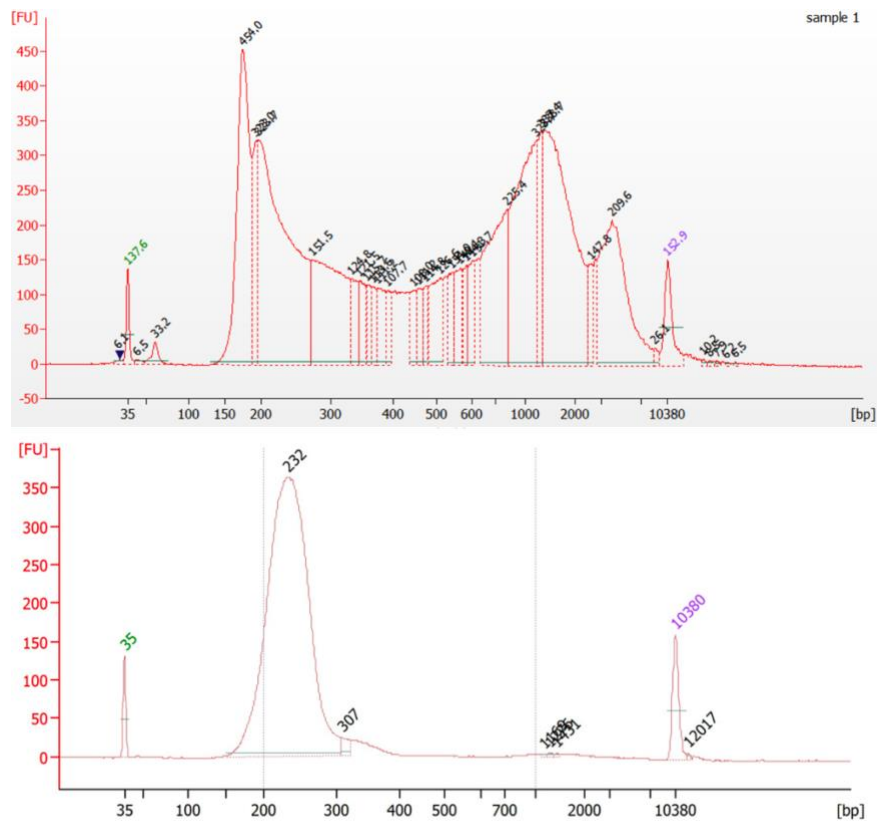
Figure 2.3. Upper image: Bioanalyzer trace of PCR-amplified ATAC-seq library DNA library. Lower image: Bioanalyzer trace of size-selected (175-250bp) PCR-amplified ATAC-seq library.

2.2.3 Peak sets optimisation

Table 1 shows the S-LDSR results for 3 different peak sets (peakset 1, peakset 2 and peakset 3) for comparison. Peakset 3, showing the highest proportion of SNPs and the highest heritability, is the one chosen for final analysis in this chapter.

"Peakset 1" ("merged" peak set) was derived by merging all technical replicates and biological replicates using Samtools merge and calling the peaks using MACS2 FDR<0.05. Artefact regions (defined as "Blacklist" regions by ENCODE; https://www.encodeproject.org/files/ENCFF001TDO/) with excessive unstructured

anomalous read mapping (e.g., regions of centromeres, telomeres and satellite repeats) were excluded. This resulted in 232,477 peaks.

To obtain "Peakset 2" ("consensus" peak set), first, the 2 technical replicate BAM files for each biological replicate were merged using Samtools to obtain 3 biological replicate BAM files. Peaks with FDR < 0.01 in each of the 3 biological replicates BAM files were then identified using MACS2 (Y. Zhang et al. 2008). Peaks observed in at least 2 of the 3 biological replicates were then identified using DiffBind v2.16.0 (Stark and Brown 2011) and finally "Blacklist" regions were removed. This produced 125,499 peaks.

"Peakset 3" ("downsampled" peak set) is the final peak set used for the final analysis and detailed results are described in section 2.4. To obtain Peakset 3, BAM reads from technical replicates from all 3 biological replicates were down-sampled to the lowest read count of any replicate (170 million) and merged using Samtools, resulting in a large, single down-sampled BAM file representing all samples. Peaks (FDR < 0.01) were identified in the down-sampled BAM file using MACS2 and those that intersected with the FDR < 0.01 peaks observed in at least 2 of the 3 biological replicates were taken as the high confidence OCRs. Blacklist regions were excludes. This produced 72,708 peaks.

	Prop_SNPs			Prop_h2			Enrichment			Enrichment p-value		
Trait	Peakset 1	Peakset 2	Peakset 3	Peakset 1	Peakset 2	Peakset 3	Peakset 1	Peakset 2	Peakset 3	Peakset 1	Peakset 2	Peakset 3
Schizophrenia	1.19E-02	1.00E-02	3.31E-02	9.43E-02	8.60E-02	1.99E-01	7.92E+00	8.25E+00	6.00E+00	4.83E-04	3.50E-04	3.30E-10
ADHD	1.19E-02	1.00E-02	3.31E-02	1.61E-01	1.29E-01	1.84E-01	1.35E+01	1.23E+01	5.56E+00	9.40E-03	2.20E-02	7.89E-03
Bipolar disorder	1.19E-02	1.00E-02	3.31E-02	5.45E-02	5.00E-02	1.53E-01	4.58E+00	4.90E+00	4.61E+00	2.40E-01	2.40E-01	5.16E-03
Autism spectrum disorder	1.19E-02	1.00E-02	3.31E-02	7.43E-02	2.70E-02	1.32E-01	6.24E+00	2.61E+00	3.99E+00	4.67E-01	7.60E-01	1.41E-01
Major depressive disorder	1.19E-02	1.00E-02	3.31E-02	6.44E-02	2.50E-02	1.47E-01	5.41E+00	2.35E+00	4.43E+00	2.84E-01	7.60E-01	8.40E-04
Blood triglyceride levels	1.19E-02	1.00E-02	3.31E-02	2.35E-02	4.80E-02	8.23E-02	1.98E+00	4.61E+00	2.48E+00	8.38E-01	4.80E-01	3.99E-01
Height	1.19E-02	1.00E-02	3.31E-02	1.55E-02	5.60E-03	4.76E-02	1.30E+00	5.40E-01	1.44E+00	8.55E-01	8.70E-01	6.64E-01

Table 2.1. S-LDSR results for 3 different peak sets (see text for how peak sets were obtained) Peakset 1: "merged", includes 232,477 peaks. Peakset 2: "consensus" , includes 125,499 peaks. Peakset 3: "downsampled", included 72,708 peaks.

2.3 Final method

2.3.1 Samples

Three fresh human fetal brain samples aged 16, 18 and 19 post-conception weeks were acquired from the MRC- Wellcome Trust Human Developmental Biology Resource (HDBR) (http://www.hdbr.org/) with ethical approval. Samples were obtained through elective terminations of pregnancy and were of normal karyotype (2 female and 1 male). The left frontal cortex from each foetus was dissected and dounce homogenized on ice to produce a single cell suspension. Prior to nuclei isolation, aliquots of 1 to 10 million cells were stored at -80°C in cryovials containing 1ml Hibernate-E media (ThermoFisher Scientific) supplemented with 6% dimethyl sulfoxide (DMSO), in Nalgene® Mr Frosty containers.

2.3.2. Nuclei isolation and Assay for Transposase-Accessible Chromatin

Two technical replicates were processed from each of the 3 bulk frontal cortex samples. Nuclei were isolated from cryopreserved cell suspensions by first centrifuging cells at 2500 RPM for 5 mins and resuspending them in ice-cold cell lysis buffer (sucrose 0.25M; KCl 25mM, MgCl2 5mM, Tris-Cl 10mM, dithiothereitol 1mM, 0.1% Triton) for 15 mins. Nuclei were then pelleted by centrifuging at 3200 RPM for 8 mins and resuspended in storage buffer (sucrose 0.25M, MgCl2 5mM, Tris-Cl 10mM, BSA 0.1%) by gentle pipetting for 10 secs. The nuclei suspension was then left on ice for 10 mins and passed through a 21G needle 10 times to thoroughly resuspend nuclei. Nuclei were then visualized under an optical microscope for quality control and counting. For each technical replicate, 50,000 nuclei were incubated in the transposition reaction mix (20µl nuclease-free water; 25µl 2X Tagment DNA Buffer,

Illumina Cat#FC-121-1030; 5µl Tn5 Transposase, Illumina Cat#FC-121-1030) at 37°C for 30 minutes, as described by Buenrostro et al (J D Buenrostro et al. 2013).

2.3.3 PCR amplification, size selection and sequencing of ATAC-Seq libraries

Transposase reactions were initially amplified by 8 cycles of PCR using primers described by Buenrostro et al (J D Buenrostro et al. 2013) and 2X NEBNext Q5 HotStart HiFi Master Mix (New England Biolabs, Cat# M0543). Amplicons of 175 -250bp were size-selected using agarose gel electrophoresis on BluePippin 2% gel cassettes (Sage Science) and further amplified through 7 additional PCR cycles. Fragment size analysis and quantification of ATAC-Seq libraries were performed using an Agilent Technologies 2100 Bioanalyzer and high sensitivity DNA kit and the Qubit dsDNA HS Assay Kit (ThermoFisher Cat#32854). ATAC-Seq libraries were sequenced on an Illumina Hi-Seq 4000 to a depth of at least 100 million paired-end 75bp reads per library.

		Temperature	Time
Holding stage		98°C	30 seconds
Cycling stage	Denaturation	98°C	10 seconds
	Annealing	65°C	30 seconds
	Elongation	72°C	30 seconds
		72°C	5 minutes

Table 2.2. PCR thermal cycles

2.3.4 ATAC-Seq data analysis

Sequencing quality was confirmed using FastQC (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc*). Reads were aligned to the GRCh37 (hg19) human genome reference sequence using Bowtie2 v2.2.9 (Langmead et al. 2009) following adapter trimming. This produced a SAM file for each replicate which was then converted into a coordinate-sorted BAM file of paired-end reads with Samtools v1.5 (Heng Li et al. 2009). Reads that mapped to more than one locus or were PCR duplicates were excluded. This yielded >90 million uniquely mapped, non-duplicated paired-end reads per technical replicate.

High-confidence OCRs were identified as described in section 2.2.3 for peakset 3, "downsampled". The bioinformatic pipeline is shown in figure 2.5.

Sample		Total_sequences	Avg_sequence_length	Percent_duplicates	Percent_mitochondrial_reads
Biol_rep_1	Tech_rep_1a	170878011	73.11	37.15	3.32
	Tech_rep_1b	286575310	71.22	23.61	3.12
Biol_rep_2	Tech_rep_2a	225897756	74.02	37.19	1.18
	Tech_rep_2b	230616836	73.51	41.26	1.23
Biol_rep_3	Tech_rep_3a	253595964	74.42	38.52	1.12
	Tech_rep_3b	239570038	73.99	33.20	1.34

Table 2.3. Biological and technical replicates metrics

File type	Software used
.FASTQ Sequencing reads	Bowtie2
.SAM alignment	Samtools
.BAM binary alignment	MACS2
.BED open chromatin peaks	
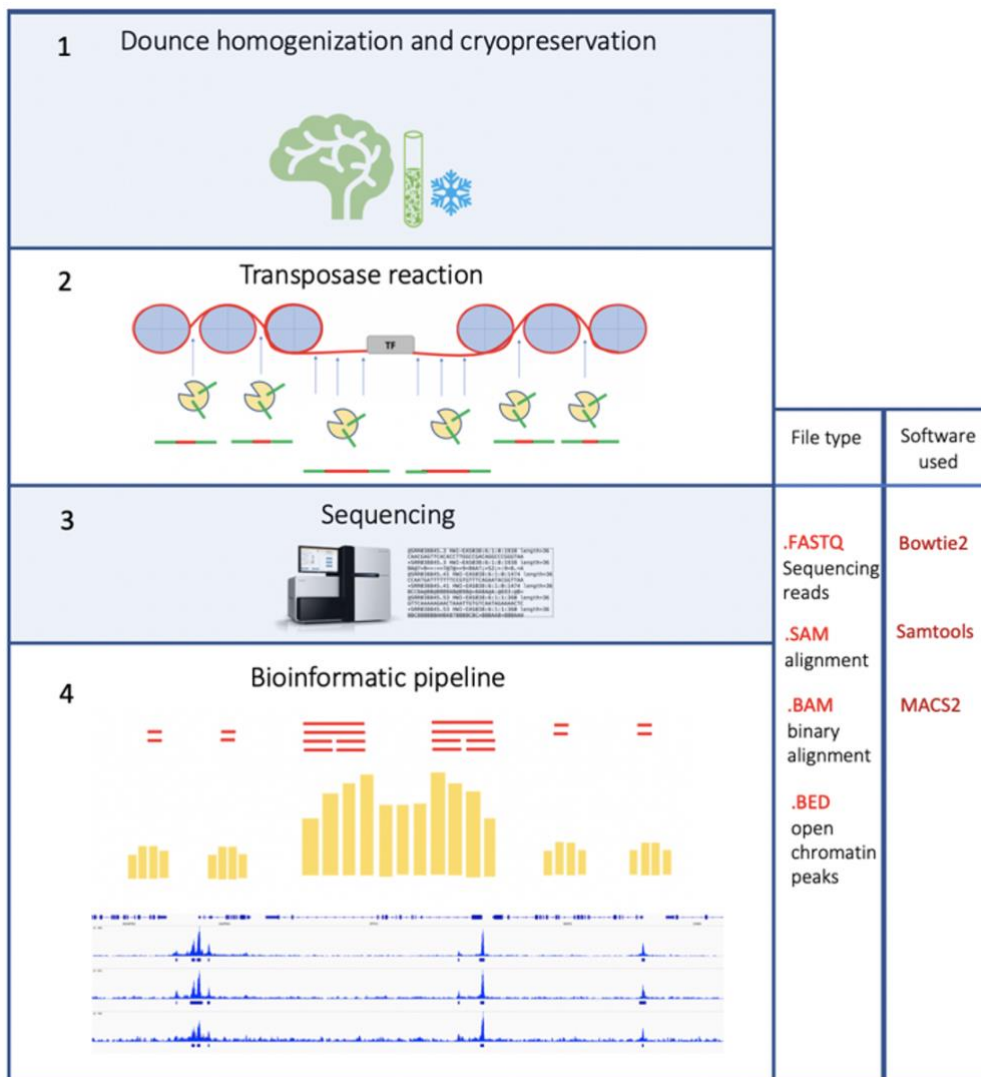
Figure 2.4. ATAC-Seq workflow. 1. Dounce homogenization and cryopreservation. Fresh foetal brain frontal cortex tissue is dounce homogenized and the resulting cell suspensions are cryopreserved. 2. Molecular mechanism of the ATAC reaction. The enzyme transposase simultaneously fragments and tags DNA in a single step. A PCR amplification then appends sequencing adapters and sample indexes to each DNA fragment 3. DNA library sequencing, adapter trimming and demultiplexing. The sequencer produces output data as FASTQ files. 4. Bioinformatic pipeline. Sequencing reads are transformed from FASTQ to BED file format using Bowtie2 (Langmead et al. 2009), Samtools (H Li et al. 2009) and MACS2 (Y. Zhang et al. 2008) software. The DNA fragments, sequenced and aligned to the human genome, pile up forming peaks at open chromatin sites.

```
##### Bowtie2_Alignment
bowtie2 --met-file .sam.metrics -x male.hg19.fa -1 fastq1 -2 fastq2 -S sam

#### Samtools_SamToBam
samtools view -b -S sam > bam

#### Samtools_MergeBamFiles
samtools merge merged.bam bam2 bam3

#### Samtools_Sort_MergedBamFiles
samtools sort merged.bam -o sorted.merged.bam

#### Samtools_Index_MergedBamFiles
samtools index sorted.merged.bam sorted.merged.bam.bai

#### Samtools_RemoveDuplicates
samtools rmdup sorted.merged.bam nodup.sorted.merged.bam

#### Samtools_Index_NoDup_MergedBamFiles
samtools index nodup.sorted.merged.bam nodup.sorted.merged.bam.bai

#### MACS2_PeakCalling
macs2 callpeak -t nodup.sorted.merged.bam -n NarrowPeaks --bdg --nomodel -g hs -f BAMPE --outdir ./ -q 0.01
```

Figure 2.5. ATAC-seq pipeline

2.3.5 Overlap between open chromatin regions identified in different tissues

ChIPpeakAnno (L. J. Zhu et al. 2010) was used to determine the numbers of bulk foetal frontal OCRs overlapping adult frontal cortex OCRs (Hoffman et al. 2019) and OCRs identified in the germinal zone / cortical plate of the foetal cortex (de la Torre-Ubieta et al. 2018). Consensus OCRs identified by ATAC-Seq in bulk adult frontal cortex (CMC_ATACSeq_consensusPeaks.bed) (Hoffman et al. 2019) were downloaded from https://www.synapse.org/#!Synapse:syn18134202 and converted to GRCh37 coordinates using https://genome.ucsc.edu/cgi-bin/hgLiftOver. OCRs identified by ATAC-Seq in the germinal zone / cortical plate of the foetal cortex (de la Torre-Ubieta et al. 2018) were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95023.

2.3.6 Bioinformatic annotation of open chromatin regions

Enrichment of transcription factor binding motifs within high confidence foetal frontal cortex OCRs was tested using HOMER (http://homer.ucsd.edu/homer/ngs/peakMotifs.html). Enrichment of biological process Gene Ontology (GO) annotations for genes with transcription start sites (TSS) closest to OCRs was tested using the comprehensive GOTERM_BP_FAT category in DAVID Bioinformatics Resources (Huang, Sherman, and Lempicki 2009). For the large number of OCRs uniquely observed in either foetal or adult (Hoffman et al. 2019) bulk frontal cortex, the GO analyses were restricted to genes with an OCR within -100bp and +50bp of their TSS (i.e. the core promoter). Enrichment of previously identified genome-wide significant (P < 5E-8) eQTL (O'Brien et al. 2018) operating in second trimester foetal brain within foetal frontal cortex OCRs was tested using GARFIELD (Iotchkova et al. 2019), a method that controls for minor allele frequency, linkage disequilibrium and local gene density.

2.3.7 Testing enrichment of SNP heritability for neuropsychiatric disorders in open chromatin regions

Stratified linkage disequilibrium score regression (SLDSC) is a statistical method designed to estimate the proportion of SNP heritability associated with a trait in the population. It also makes it possible to partition the genome by functional category and to test whether SNP heritability associated with a trait is attributable to certain functional categories more than others. SNP heritability in this context is defined as the proportion of genetic variance in a trait explained by all SNPs, as opposed to SNPs that reach the genome-wide threshold for significance. To partition heritability, SLDSR requires a GWAS summary statistics file and an

annotation file that contains genomic coordinates for the functional partition. SLDSR exploits the predicted relationship that exists between the association statistics of a set of GWAS index SNPs and local linkage disequilibrium surrounding each index SNP for a polygenic trait, i.e. this relationship predicts that, on average, index SNPs with high LD r^2 scores are more likely to tag SNPs with higher association statistics than SNPs with low r^2 scores. As such, using a linear regression model, SLDSC generates an estimate of the proportion of SNP heritability that is captured by a given set of index SNPs, including any heritability explained by non-assayed SNPs within the haplotype block that each index SNP tags, as well as correcting for systematic biases.

To calculate the GWAS SNP heritability associated within foetal brain open chromatin sites, an annotation file was generated by using bedtools to intersect the peak file with the SNPs contained in the 1000 genomes reference panel. This assigns a 1 or 0 to each SNP depending on whether or not it overlaps an open chromatin region. Next, using the lsdc.py script and plink files provided (code example is in the appendix), the latter containing genotype data from 329 European ancestry reference genomes, LD scores for all SNPs with a minor allele frequency >5% were taken by measuring the correlation between SNPs in 1cM windows in the plink files. However, only LD scores for SNPs that were located in both the GWAS summary stats file and the annotation file were retained for the next stage in the analysis. Next, again using the lsdc.py script, heritability was partitioned by functional category by regressing the χ^2 association statistics for each GWAS index SNP present in the annotation file against each SNP's local LD score. This produced a single regression co-efficient representing the per-SNP heritability for that functional category (note that this assumes uniform per-SNP heritability for all SNPs in the category). Finally, the total SNP

heritability for the category (h^2) was calculated by multiplying the regression coefficient by the total number of SNPs in the category and an enrichment score was produced. The significance of SNP heritability enrichment was tested by generating a further 200 SNP heritability regression coefficients from random, equally sized, blocks of SNPs. This provided a normal distribution of regression coefficients and made it possible to calculate a z-score for the SNP heritability coefficient of the functional annotation of interest relative to the normally distributed coefficients. This score includes a correction for the baseline model which includes 53 annotations representing intrinsic heritability that is attributable to common genomic signatures across cell types such as promoters proximal to housekeeping genes, shared chromatin features and evolutionarily conserved regions (Finucane et al. 2015). As these annotations are common between cell types, they would inflate the heritability score if not accounted for in the analysis. Z-scores were transformed and reported as p-values, with significance taken as $p < 0.05$.

SLDSC was used to test for enrichment of SNP heritability for neuropsychiatric disorders within foetal brain OCRs. Peaks were expanded by 500bp on each side prior to SLDSR, as recommended by Finucane et al (Finucane et al. 2015). Summary statistics from large-scale GWAS of 5 neuropsychiatric disorders (ADHD (Demontis, Walters, Martin, Mattheisen, Als, Team, et al. 2019), autism spectrum disorder (Grove et al. 2019), bipolar disorder (Stahl et al. 2019), major depressive disorder (Wray et al. 2018) and schizophrenia (Pardiñas et al. 2018) and two control traits with similar sample sizes (height (Lango Allen et al. 2010) and blood triglyceride levels (Teslovich et al. 2010b) were tested (Table 2.4). Fold-enrichment estimates are the proportion of SNP heritability for each trait explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation. Each annotation was

assessed against the baseline model provided by Finucane et al (Finucane et al. 2015) (consisting of 53 genomic annotations including coding regions, promoters, enhancers and conserved regions) and resulting Z-scores used to calculate two-tailed *P*-values. I more conservatively state the Z-score *P*-values (rather than raw fold-enrichment *P*-values) in the text, highlighting those that survive Bonferroni correction for the 7 tested traits (i.e., *P* < 0.0071); however, both z-score P values and raw enrichment *P*-values are also reported in table 2.5.

Trait		
Schizophrenia	**Source**	https://walters.psycm.cf.ac.uk
	File name	clozuk.sumstats
	Reference	PMID: 29483656
		Pardiñas, Holmans, Pocklington et al. (2018)
ADHD	**Source**	https://www.med.unc.edu/pgc/download-results/
	File name	adhd_jul2017
	Reference	PMID: 30478444
		Demontis, Walters, Martin et al. (2019)
Bipolar disorder	**Source**	https://www.med.unc.edu/pgc/download-results/
	File name	daner_PGC_BIP32b_mds7a_0416a
	Reference	PMID: 31043756
		Stahl, breen, Fostner et al (2019)
Autism spectrum disorder	**Source**	https://ipsych.dk/test/downloads/
	File name	iPSYCH-PGC_ASD_Nov2017-grove-ripke.sumstats
	Reference	PMID: 30804558
		Grove, Ripke, Als et al. (2019)
Major depressive disorder	**Source**	Restricted as includes 23andMe data. See https://www.med.unc.edu/pgc/download-results/
	File name	wray-ripke-mattheisen-2017-mdd-23me-summaryqc.tsv
	Reference	PMID: 29700475
		Wray, Ripke, Mattheinsen (2018)
Blood triglyceride levels	**Source**	https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
	File name	PASS_Triglycerides
	Reference	PMID: 20686565
		Teslovich, Musunuru, Smith et al. (2010)
Height	**Source**	https://data.broadinstitute.org/alkesgroup/sumstats_formatted/
	File name	PASS_Height1.sumstats
	Reference	PMID: 20881960
		Allen, Estrada, Lettre et al. (2010)

Table 2.4. Genome-wide association study (GWAS) summary statistics used in this study

2.4 Results

2.3.1 Open chromatin regions annotation

ATAC-Seq was used to map OCRs in the frontal cortex of 3 human foetuses aged 16, 18 and 19 post-conception weeks. An example of the ATAC-Seq data, aligned to the human genome, is shown in Figure 2.6. In total, 88,501 high confidence ATAC-Seq peaks were identified. The genomic distribution of peaks approximated that of OCRs identified in other tissues (Song et al. 2011), (Bryois et al. 2017), with 28% within 5kb of a TSS and 38% within gene bodies. Approximately 60% of the OCRs identified in foetal frontal cortex overlapped OCRs observed in adult frontal cortex (Hoffman et al. 2019) (Figure 2.7); genes with non-overlapping 'foetal-specific' OCRs at their TSS were most significantly enriched for the Gene Ontology term 'nervous system development' (1.7-fold enrichment, P = 7E-03), while the larger number of genes with 'adult-specific' OCRs at their TSS were most significantly enriched for the term 'cell surface receptor signalling pathway' (1.7-fold enrichment, P = 4.4E-12). There was also substantial overlap with OCRs previously observed in both the germinal zone and cortical plate of the human foetal brain (de la Torre-Ubieta et al. 2018) (Figure 2.8), although ~33% of my high-confidence OCRs were not observed in either of these regions. Bulk foetal frontal cortex OCRs were highly enriched for the CTCF binding motif (11.8-fold enrichment, P = 10E-5440), with notable enrichment of motifs for neural transcription factors such as Olig2 (2.2-fold enrichment, P = 10E-1067) and Oct6 (2.1-fold enrichment, P = 10E-449) also observed. Consistent with roles in gene regulation, foetal frontal cortex OCRs were enriched for expression quantitative trait loci (eQTLs) identified in the second trimester human foetal brain (2.1-fold enrichment, P = 3.47E-19).
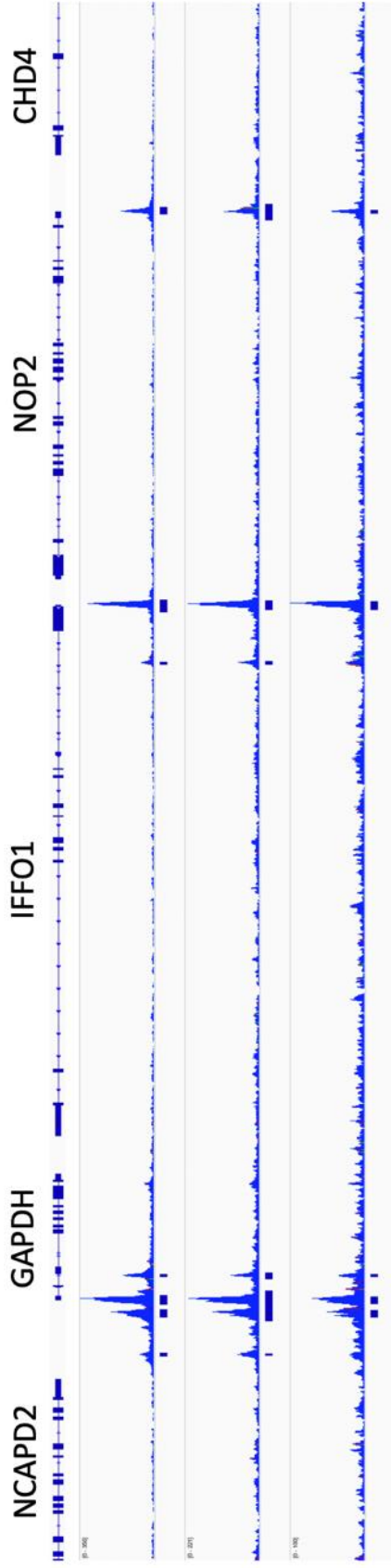
Figure 2.6. Example of open chromatin ATAC-Seq peaks from 3 second trimester foetal brain samples. BAM alignments (filtered for mitochondrial reads and PCR duplicates) were uploaded onto the Integrative Genome Browser
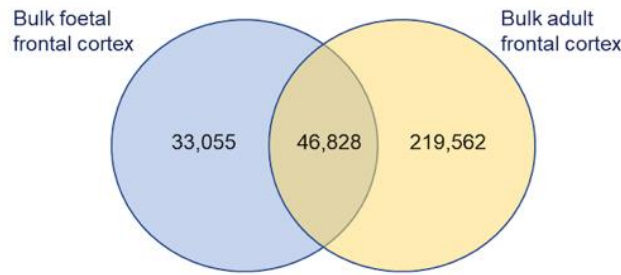
Figure 2.7. **Overlap between consensus open chromatin regions (OCRs) identified in bulk foetal frontal cortex and bulk adult frontal cortex** (Hoffman et al, 2019). Consensus OCRs identified by ATAC-Seq in bulk adult frontal cortex were downloaded from https://www.synapse.org/#!Synapse:syn18134202 and converted to Hg19 coordinates using https://genome.ucsc.edu/cgi-bin/hgLiftOver. Numbers of overlapping and unique peaks were determined using (L. J. Zhu et al. 2010). Note that this calculates the minimal number of intersected peaks, and therefore the summed number of OCRs will usually be smaller than the total number of OCRs in each set.
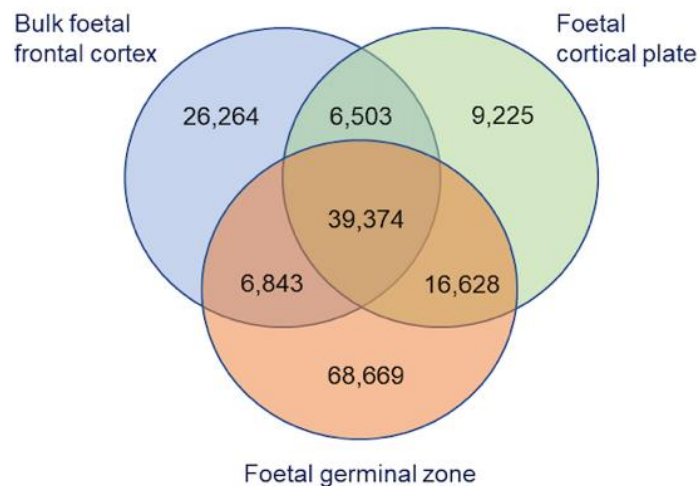
Figure 2.8. Overlap between consensus open chromatin regions (OCRs) identified in bulk fetal frontal cortex and those identified in cortical plate (CP) and germinal zone (GZ) of the fetal cortex in the study of de la Torre-Ubieta et al (2018). OCRs identified by ATAC-Seq in CP and GZ were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95023. Numbers of overlapping and unique peaks were determined using ChIPpeakAnno (Zhu et al,2010). Note that this calculates the minimal number of intersected peaks, and therefore the summed number of OCRs will usually be smaller than the total number of OCRs in each set.

2.4.2 Partitioned heritability analysis

Stratified linkage disequilibrium score regression (SLDSR) and partitioned heritability analyses (Finucane et al. 2015) were used to test for enrichment of single nucleotide polymorphism (SNP) heritability for 5 major neuropsychiatric disorders (ADHD, autism spectrum disorder [ASD], bipolar disorder, major depressive disorder and schizophrenia) within bulk foetal frontal cortex OCRs. A significant enrichment of SNP heritability for schizophrenia (6-fold enrichment, Z-score P = 3.5 X 10^{-8}) was observed within these regions. Enrichment of variants associated with the negative controls (triglyceride levels or height) (Table 2.5) was non-significant (Z-score P > 0.05). The level of enrichment for schizophrenia risk variation was similar to that reported in adult frontal cortex (Hoffman et al. 2019), with 3.3% of SNPs in foetal brain OCRs accounting for 19.9% of SNP heritability for the condition.

Peak set	Trait	SNPs			Enrichment			Coefficient	
Bulk ATAC peaks									
	Schizophrenia	Proportion of SNPs in annotation	3.314E-02		Enrichment	6.003E+00		Coefficient	2.800E-07
		Proportion SNP h2 explained	1.989E-01		Enrichment SE	7.744E-01		Coefficient SE	5.077E-08
					Enrichment P	3.299E-10		Coefficient Z-score	5.516E+00
								Z-score P (2-tailed)	3.479E-08
	ADHD	Proportion of SNPs in annotation	3.314E-02		Enrichment	5.562E+00		Coefficient	1.727E-07
		Proportion SNP h2 explained	1.843E-01		Enrichment SE	1.726E+00		Coefficient SE	6.607E-08
					Enrichment P	7.894E-03		Coefficient Z-score	2.613E+00
								Z-score P (2-tailed)	8.970E-03
	Bipolar disorder	Proportion of SNPs in annotation	3.314E-02		Enrichment	4.607E+00		Coefficient	1.142E-07
		Proportion SNP h2 explained	1.527E-01		Enrichment SE	1.298E+00		Coefficient SE	6.631E-08
					Enrichment P	5.158E-03		Coefficient Z-score	1.722E+00
								Z-score P (2-tailed)	8.506E-02
	Autism spectrum disorder	Proportion of SNPs in annotation	3.314E-02		Enrichment	3.989E+00		Coefficient	7.025E-08
		Proportion SNP h2 explained	1.322E-01		Enrichment SE	2.061E+00		Coefficient SE	6.489E-08
					Enrichment P	1.414E-01		Coefficient Z-score	1.083E+00
								Z-score P (2-tailed)	2.790E-01
	Major depressive disorder	Proportion of SNPs in annotation	3.314E-02		Enrichment	4.425E+00		Coefficient	1.974E-08
		Proportion SNP h2 explained	1.466E-01		Enrichment SE	1.007E+00		Coefficient SE	8.773E-09
					Enrichment P	8.405E-04		Coefficient Z-score	2.250E+00
								Z-score P (2-tailed)	2.445E-02
	Blood triglyceride levels	Proportion of SNPs in annotation	3.314E-02		Enrichment	2.484E+00		Coefficient	-1.742E-08
		Proportion SNP h2 explained	8.232E-02		Enrichment SE	1.795E+00		Coefficient SE	3.881E-08
					Enrichment P	3.990E-01		Coefficient Z-score	-4.488E-01
								Z-score P (2-tailed)	6.536E-01
	Height	Proportion of SNPs in annotation	3.314E-02		Enrichment	1.437E+00		Coefficient	-6.152E-08
		Proportion SNP h2 explained	4.761E-02		Enrichment SE	1.006E+00		Coefficient SE	3.827E-08
					Enrichment P	6.639E-01		Coefficient Z-score	-1.608E+00
								Z-score P (2-tailed)	1.079E-01

Table 2.5. Partitioned heritability results for open chromatin peaks in bulk tissue

2.5 Discussion

I applied ATAC-Seq to generate maps of regulatory genomic regions in second trimester foetal frontal cortex. Integration of these regulatory maps with summary statistics from large-scale GWAS showed enrichment of SNP heritability for schizophrenia.

The data presented in this chapter represents a substantial contribution to the growing online catalogue of tissue-specific functional annotations. A previous ATAC-Seq study by De la Torre-Ubieta and colleagues (de la Torre-Ubieta et al. 2018), mapped genomic regions of open chromatin in the germinal zone and cortical plate of the foetal cerebral cortex, reporting significant enrichment of genetic variation associated with ADHD, depressive symptoms, neuroticism and schizophrenia within sites defined as preferentially accessible in the germinal zone. The data presented in this chapter confirms enrichment of SNP heritability for schizophrenia within OCRs of the human frontal cortex during the second trimester of gestation.

Although open chromatin is necessary to promote gene expression, it is not in itself sufficient, requiring the binding of transcription factors and other regulatory molecules. In addition, the regulatory potential of these sites is modulated by histone modifications such as histone methylation and acetylation. In the following chapter {Chapter 3}, I integrate the high confidence OCRs I identified in the foetal frontal cortex with sites of histone modification previously identified in the human foetal brain (Consortium{ et al. 2015) in order to further assess the relevance of gene regulation in brain development to neuropsychiatric disorders

Chapter 3: Integration of foetal chromatin accessibility data with epigenomic data from foetal brain

3.1 Introduction

In the previous chapter, I generated open chromatin maps, indicative of active regulatory regions, in the second trimester foetal frontal cortex. Enrichment testing of SNP heritability within these regions suggested that they may play a role in the pathogenesis of schizophrenia. These chromatin accessible sites, however, located either near the transcription start site of genes or within intergenic or intronic genomic sites, were likely to encompass the whole range of regulatory elements (i.e. promoters, enhancers, silencers and insulators), that are known to play different roles in the transcriptional regulation of gene expression.

The first objective of chapter 3 was, therefore, to further define the foetal frontal cortex regulatory sites presented in chapter 2, by integrating my data with publicly available epigenomic annotations from the same tissue. For this purpose, I used foetal brain histone modification maps of H3K4me3 and H3K4me1, indicative of promoters and enhancers respectively, previously produced by the Roadmap Epigenomic Consortium (Roadmap Epigenomics Consortium et al. 2015).

As described in chapter 1, histone modifications have been widely used to functionally annotate the genome and to predict the nature and activation status of the regulatory elements that they are proximal to (Kouzarides 2007; Tessarz and Kouzarides 2014). The data produced with the method used to detect histone modifications (ChIP-Seq), is similar to the open chromatin data produced by the ATAC-Seq assay: sequencing reads aligned with the

reference genome pile up and can be visualised as peaks, that indicate the genomic loci associated with a particular histone modification. Integration of whole genome accessible chromatin and histone modification maps is therefore possible by detecting overlapping peaks between the two epigenomic annotations.

The second objective of this chapter was to explore the contribution of the identified foetal brain promoter and enhancer regions to the heritability of neuropsychiatric disorders. I hypothesised that restricting the enrichment analysis to active or poised promoters and enhancers would result in higher heritability enrichments. In order to test this hypothesis, I used the same method used in chapter 2, the Stratified Linkage Disequilibrium Score Regression analysis for partitioning heritability (LDSC) (Finucane et al. 2015), and summary statistics from the same 5 neuropsychiatric conditions: ADHD, autism spectrum disorder [ASD], bipolar disorder, major depressive disorder and schizophrenia.

In this chapter, I will describe the bioinformatic methods I used to integrate open chromatin foetal frontal cortex data with publicly available foetal brain histone modification data, indicative of active or poised promoters and enhancers (H3K4me3 and H3K4me1 respectively). I will then illustrate and discuss results from the enrichment testing of SNP heritability for the same neuropsychiatric disorders tested in chapter 1, within those promoter and enhancer regions.

3.2 Methods

3.2.1 Intersection with histone modification datasets from human foetal brain

Genomic regions marked by H3K4Me1 or H3K4Me3 in a human foetal brain sample of similar gestational age to those used in this study (sample E082; female, 17 post-conception weeks) were downloaded as BED files from the Roadmap epigenomics project (https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/) (Roadmap Epigenomics Consortium et al. 2015).

BEDTools (Quinlan and Hall 2010) v2.26.0 intersect (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html) was used to identify H3K4Me1 and H3K4Me3 sites overlapping open chromatin regions that I identified in nuclei from the foetal frontal cortex (Figure 3.1). Overlapping peaks were then merged using BEDTools merge (https://bedtools.readthedocs.io/en/latest/content/tools/merge.html), to combine overlapping or book-ended peaks into a single peak (Figure 3.2). Prior to being merged, the peak sets were sorted by chromosome and then by start position.

Bedtools intersect compares two or more BED files and identifies all the regions in the genome where the features in the two files overlap (that is, share at least one base pair in common).

Code examples provided in te appendix.
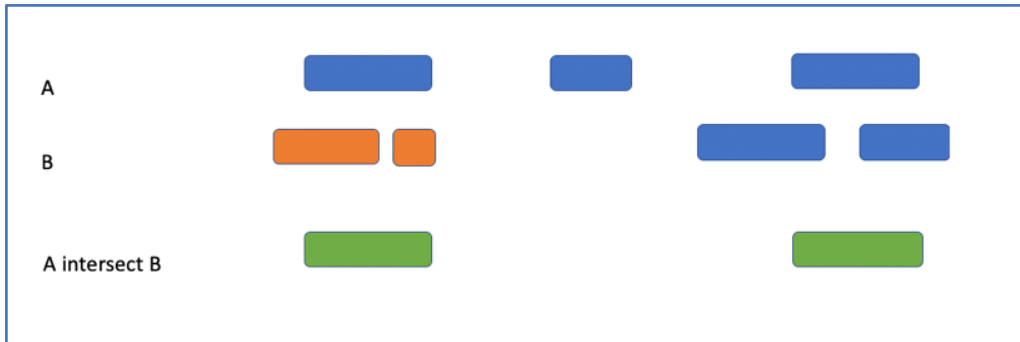
Figure 3.1. Adapted from
https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html. BEDTools Intersect
was used to identify overlapping peaks between foetal brain H3K4Me1 and H3K4Me3 sites
and open chromatin regions. The -wa option was used, to write the original entry in A for
each overlap.
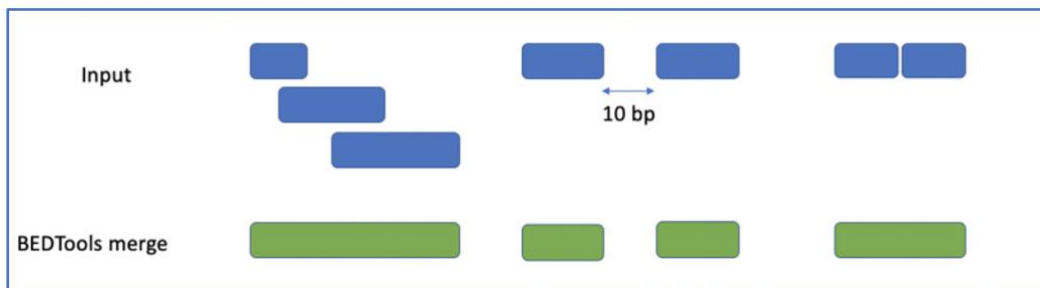
Figure 3.2. Adapted from
https://bedtools.readthedocs.io/en/latest/content/tools/merge.html. Bedtools merge was
used to combine overlapping and book-ended peaks into a single peak.

3.2.2 Testing enrichment of SNP heritability for neuropsychiatric disorders in open chromatin regions

I used stratified LD score regression (LDSC) (Finucane et al. 2015) to test for enrichment of

SNP heritability for neuropsychiatric disorders within H3K4Me1 and H3K4Me3 sites alone as

well as within open chromatin regions overlapping these sites. LDSC was performed using the

protocols described in section 2.2.7.

Peaks were expanded by 500bp on each side prior to SLDSR, as recommended by Finucane et al (Finucane et al. 2015) and BEDTools merge was used to combine overlapping and book-ended features (Figure 3.2).

Summary statistics from large-scale GWAS of 5 neuropsychiatric disorders (ADHD (Demontis, Walters, Martin, Mattheisen, Als, Team, et al. 2019), autism spectrum disorder (Grove et al. 2019), bipolar disorder (Stahl et al. 2019), major depressive disorder (Wray et al. 2018) and schizophrenia (Pardiñas et al. 2018)) and two control traits with similar sample sizes (height (Allen et al. 2010) and blood triglyceride levels (Teslovich et al. 2010a)) were tested (See chapter 1, table 1). Fold-enrichment estimates are the proportion of SNP heritability for each trait explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation. Each annotation was assessed against the baseline model provided by Finucane et al (Finucane et al. 2015) (consisting of 53 genomic annotations including coding regions, promoters, enhancers and conserved regions) and resulting Z-scores used to calculate two-tailed P-values. I more conservatively state the Z-score P-values (rather than raw fold-enrichment P-values) in the text and figures, highlighting those that survive Bonferroni correction for the 7 tested traits (i.e., $P < 0.0071$); however, both z-score P values and raw enrichment P-values are also reported in table 3.1.

3.3 Results

3.3.1 Intersection with histone modification datasets from human foetal brain

A total of 95,699 and 26,609 ChIP-Seq peaks markers of H3K4Me1 and H3K4Me3 respectively were downloaded from the Roadmap epigenomics project. The intersection and merge of

these peaks with 88,501 high confidence foetal frontal cortex ATAC-Seq peaks resulted in 36,850 and 22,833 genomic sites respectively.

3.3.2 Testing enrichment of SNP heritability for neuropsychiatric disorders in open chromatin regions overlapping H3K4Me1 and H3K4Me3 sites

I tested for enrichment of single nucleotide polymorphism (SNP) heritability for ADHD (Demontis, Walters, Martin, Mattheisen, Als, Team, et al. 2019), autism spectrum disorder [ASD], bipolar disorder (Stahl et al. 2019), major depressive disorder (Wray et al. 2018) and schizophrenia (Pardiñas et al. 2018) within H3K4Me1 and H3K4Me3 sites alone as well as within their overlap with foetal brain open chromatin regions using SLDSR (Finucane et al. 2015), controlling for general genomic annotations (e.g. coding regions, promoters, enhancers and conserved regions) included in the baseline model (Finucane et al. 2015) to obtain Z-score P-values.

Enrichment of SNP heritability for neuropsychiatric disorders within H3K4Me1 sites overlapping foetal brain open chromatin regions is shown in Figure 3.3. I observed strong enrichment of SNP heritability for all 5 tested neuropsychiatric disorders within H3K4Me1 sites overlapping foetal brain open chromatin regions (10.53-fold enrichment, Z-score P =1.2E-12 for schizophrenia, 11.7-fold enrichment , Z-score P = 1.20E-05 for ADHD, 10.51-fold enrichment, Z-score P = 1.2E-04 for bipolar disorder, 10.44-fold enrichment, Z-score P = 6.3E-03 for autism spectrum disorders, 8.7-fold enrichment, Z-score P = 1.1E-05 for major depressive disorder).

Enrichment of SNP heritability for neuropsychiatric disorders within H3K4Me3 sites overlapping foetal brain open chromatin regions is shown in Figure 3.4. SNP heritability for all 5 tested neuropsychiatric disorders was also significantly enriched within these sites, with estimates of enrichment for each condition even higher than for foetal frontal cortex open chromatin regions overlapping H3K4Me1 sites (14.08-fold enrichment, Z-score P = 10E-12 for schizophrenia, 15.65-fold enrichment, Z-score P = 2.1E-06 for ADHD, 14.71-fold enrichment, Z-score P = 7.4E-05 for bipolar disorder, 16.33-fold enrichment, Z-score P = 9E-05 for autism spectrum disorders, 9.35-fold enrichment, Z-score P = 6.3E-04 for major depressive disorder).

There was little evidence for an enrichment of variants associated with the control trait of triglyceride levels (4.29-fold enrichment, Z-score P = 8.7E-01 for H3K4me1 sites overlapping foetal brain open chromatin regions and 6.42-fold enrichment, Z-score P = 1 for H3K4me3 sites overlapping foetal brain open chromatin regions) when baseline annotations were considered.
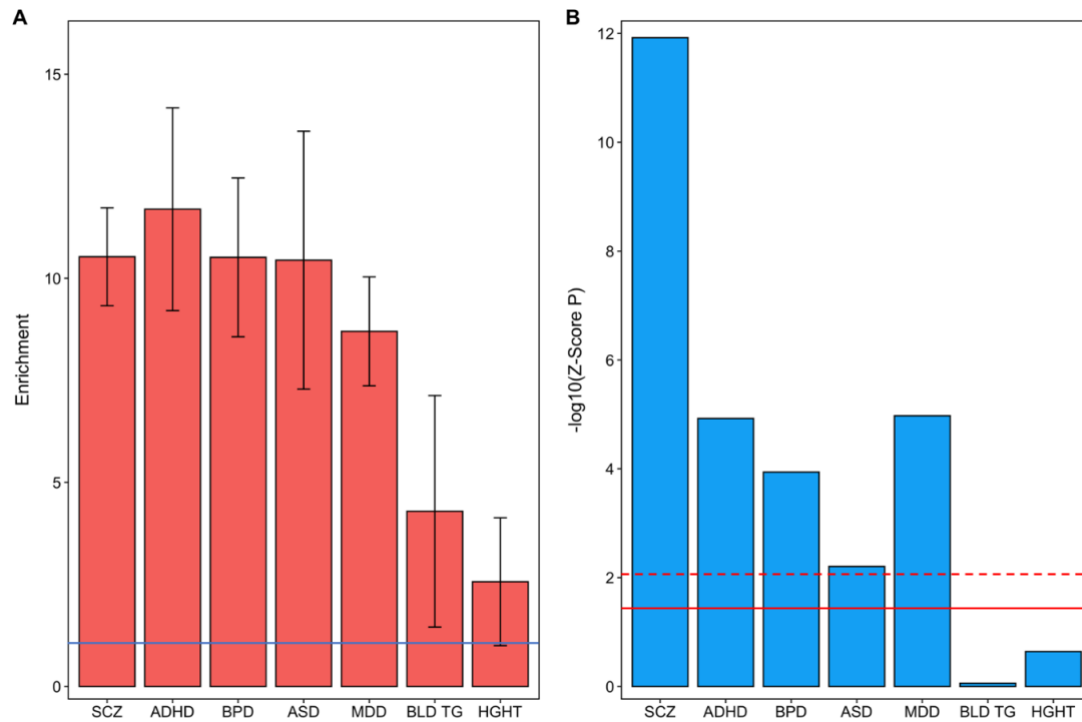
Figure 3.3. Taken from Kouakou et al., under review. Created by Darren Cameron. **Partitioned heritability for 5 neuropsychiatric disorders and 2 control traits within open chromatin regions identified in bulk foetal frontal cortex overlapping foetal brain H3K4Me1 sites.** a) Fold-enrichment estimates of SNP heritability for each trait (the proportion of SNP heritability explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation). Error bars represent standard error. The solid blue horizontal line indicates no enrichment. b) -Log$_{10}$ Z-score *P*-values for enrichment of SNP heritability, controlling for general genomic annotations included in the baseline model of Finucane and colleagues (Finucane et al. 2015). The solid red horizontal line indicates the Z-score *P*-value 0.05 threshold; the dashed red horizontal line indicates the threshold for Z-score *P*-values surviving Bonferroni correction for 7 tested traits. SCZ = schizophrenia; ADHD = attention deficit hyperactivity disorder; BPD = bipolar disorder; MDD = major depressive disorder; ASD = autism spectrum disorder; BLD TG = blood triglyceride levels; HGHT = height.

Figure 3.4. Taken from Kouakou et al., under review. Created by Darren Cameron. **Partitioned heritability for 5 neuropsychiatric disorders and 2 control traits within open chromatin regions identified in bulk foetal frontal cortex overlapping foetal brain H3K4Me3 sites.** a) Fold-enrichment estimates of SNP heritability for each trait (the proportion of SNP heritability explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation). Error bars represent standard error. The solid blue horizontal line indicates no enrichment. b) -Log$_{10}$ Z-score *P*-values for enrichment of SNP heritability, controlling for general genomic annotations included in the baseline model of Finucane and colleagues (Finucane et al. 2015). The solid red horizontal line indicates the Z-score *P*-value 0.05 threshold; the dashed red horizontal line indicates the threshold for Z-score *P*-values surviving Bonferroni correction for 7 tested traits. SCZ = schizophrenia; ADHD = attention deficit hyperactivity disorder; BPD = bipolar disorder; MDD = major depressive disorder; ASD = autism spectrum disorder; BLD TG = blood triglyceride levels; HGHT = height.

Figure 3.5. Taken from Kouakou et al., under review. **Enrichment of SNP heritability for 5 neuropsychiatric disorders within open chromatin regions identified in bulk foetal frontal cortex (ATAC), within previously identified sites of histone modification in the human foetal brain (H3K4me1 and H3K4me3)** (Roadmap Epigenomics Consortium et al. 2015) **and within bulk foetal frontal cortex open chromatin regions overlapping each histone modification (ATAC_ H3K4me1 and ATAC_ H3K4me3).** Fold-enrichment estimates of SNP heritability for each trait is the proportion of SNP heritability explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation. Error bars represent standard error. SCZ = schizophrenia; ADHD = attention deficit hyperactivity disorder; BPD = bipolar disorder; MDD = major depressive disorder; ASD = autism spectrum disorder.

Peak set	Trait	SNPs		Enrichment		Coefficient	
Bulk ATAC peaks intersecting foetal brain H3K4me1 sites							
	Schizophrenia	Proportion of SNPs in annotation	1.698E-02	Enrichment	1.053E+01	Coefficient	5.738E-07
		Proportion SNP h2 explained	1.788E-01	Enrichment SE	1.200E+00	Coefficient SE	8.185E-08
				Enrichment *P*	1.001E-13	Coefficient Z-score	7.011E+00
						Z-score *P* (2-tailed)	1.200E-12
	ADHD	Proportion of SNPs in annotation	1.698E-02	Enrichment	1.169E+01	Coefficient	4.208E-07
		Proportion SNP h2 explained	1.986E-01	Enrichment SE	2.484E+00	Coefficient SE	9.609E-08
				Enrichment *P*	1.804E-05	Coefficient Z-score	4.379E+00
						Z-score *P* (2-tailed)	1.190E-05
	Bipolar disorder	Proportion of SNPs in annotation	1.698E-02	Enrichment	1.051E+01	Coefficient	3.869E-07
		Proportion SNP h2 explained	1.786E-01	Enrichment SE	1.945E+00	Coefficient SE	1.003E-07
				Enrichment *P*	1.418E-06	Coefficient Z-score	3.856E+00
						Z-score *P* (2-tailed)	1.151E-04
	Autism spectrum disorder	Proportion of SNPs in annotation	1.698E-02	Enrichment	1.044E+01	Coefficient	2.669E-07
		Proportion SNP h2 explained	1.774E-01	Enrichment SE	3.160E+00	Coefficient SE	9.758E-08
				Enrichment *P*	2.623E-03	Coefficient Z-score	2.735E+00
						Z-score *P* (2-tailed)	6.240E-03
	Major depressive disorder	Proportion of SNPs in annotation	1.698E-02	Enrichment	8.700E+00	Coefficient	5.209E-08
		Proportion SNP h2 explained	1.478E-01	Enrichment SE	1.335E+00	Coefficient SE	1.182E-08
				Enrichment *P*	4.115E-08	Coefficient Z-score	4.405E+00
						Z-score *P* (2-tailed)	1.060E-05
	Blood triglyceride levels	Proportion of SNPs in annotation	1.698E-02	Enrichment	4.292E+00	Coefficient	-9.984E-09
		Proportion SNP h2 explained	7.289E-02	Enrichment SE	2.836E+00	Coefficient SE	6.184E-08
				Enrichment *P*	2.365E-01	Coefficient Z-score	-1.614E-01
						Z-score *P* (2-tailed)	8.717E-01
	Height	Proportion of SNPs in annotation	1.698E-02	Enrichment	2.567E+00	Coefficient	-6.975E-08
		Proportion SNP h2 explained	4.361E-02	Enrichment SE	1.566E+00	Coefficient SE	5.790E-08
				Enrichment *P*	3.171E-01	Coefficient Z-score	-1.205E+00
						Z-score *P* (2-tailed)	2.283E-01
Bulk ATAC peaks intersecting foetal brain H3K4me3 sites							
	Schizophrenia	Proportion of SNPs in annotation	1.112E-02	Enrichment	1.409E+01	Coefficient	8.910E-07
		Proportion SNP h2 explained	1.566E-01	Enrichment SE	1.633E+00	Coefficient SE	1.203E-07
				Enrichment *P*	7.893E-14	Coefficient Z-score	7.406E+00
						Z-score *P* (2-tailed)	1.000E-13
	ADHD	Proportion of SNPs in annotation	1.112E-02	Enrichment	1.564E+01	Coefficient	6.829E-07
		Proportion SNP h2 explained	1.739E-01	Enrichment SE	3.336E+00	Coefficient SE	1.440E-07
				Enrichment *P*	1.175E-05	Coefficient Z-score	4.741E+00
						Z-score *P* (2-tailed)	2.100E-06
	Bipolar disorder	Proportion of SNPs in annotation	1.112E-02	Enrichment	1.471E+01	Coefficient	6.130E-07
		Proportion SNP h2 explained	1.635E-01	Enrichment SE	2.709E+00	Coefficient SE	1.547E-07
				Enrichment *P*	9.284E-07	Coefficient Z-score	3.962E+00
						Z-score *P* (2-tailed)	0.0000742
	Autism spectrum disorder	Proportion of SNPs in annotation	1.112E-02	Enrichment	1.633E+01	Coefficient	5.198E-07
		Proportion SNP h2 explained	1.815E-01	Enrichment SE	3.923E+00	Coefficient SE	1.327E-07
				Enrichment *P*	1.491E-04	Coefficient Z-score	3.917E+00
						Z-score *P* (2-tailed)	8.970E-05
	Major depressive disorder	Proportion of SNPs in annotation	1.112E-02	Enrichment	9.346E+00	Coefficient	6.228E-08
		Proportion SNP h2 explained	1.039E-01	Enrichment SE	1.925E+00	Coefficient SE	1.821E-08
				Enrichment *P*	2.597E-05	Coefficient Z-score	3.421E+00
						Z-score *P* (2-tailed)	6.249E-04
	Blood triglyceride levels	Proportion of SNPs in annotation	1.112E-02	Enrichment	6.422E+00	Coefficient	1.713E-09
		Proportion SNP h2 explained	7.139E-02	Enrichment SE	4.141E+00	Coefficient SE	1.017E-07
				Enrichment *P*	1.885E-01	Coefficient Z-score	1.684E-02
						Z-score *P* (2-tailed)	9.866E-01
	Height	Proportion of SNPs in annotation	1.112E-02	Enrichment	5.921E+00	Coefficient	7.557E-09
		Proportion SNP h2 explained	6.582E-02	Enrichment SE	2.148E+00	Coefficient SE	8.415E-08
				Enrichment *P*	2.395E-02	Coefficient Z-score	8.980E-02
						Z-score *P* (2-tailed)	9.284E-01

Table 3.1: Partitioned heritability results for high confidence open chromatin regions in bulk foetal frontal cortex

I generated maps of foetal brain promoters and enhancers by integrating the second trimester regulatory regions presented in the previous chapter with foetal brain H3K4me3 and H3K4me1 sites (histone modifications indicative of poised or active enhancers and promoters, respectively), downloaded from the RoadMap Epigenomics Mapping Consortium (Consortium{ et al. 2015).

Although restricting my bulk ATAC-Seq peaks to those overlapping either foetal brain H3K4me1 or H3K4me3 sites (Roadmap Epigenomics Consortium et al. 2015) resulted in a fewer number of peaks (encompassing 1.1% and 1.7% of SNPs for ATAC-Seq regions overlapping H3K4me3 and H3K4me1 respectively), I observed strong enrichment of SNP heritability for all 5 tested neuropsychiatric disorders (but not for the two control traits), with Z-score *P*-values surviving Bonferroni correction in each case (Figure 3.4; Table 3.1). With the exception of enrichment of SNP heritability for major depressive disorder in H3K4Me3 sites, these enrichments were consistently greater than when based on foetal brain ATAC-Seq, H3K4Me1 or H3K4Me3 peaks alone (Figure 3.5), highlighting the value of additional epigenomic annotations to define regulatory regions of the genome. The amount of SNP heritability explained by these combined annotations (10-20%) was similar to that accounted for by foetal frontal cortex open chromatin regions alone, suggesting that the majority of relevant open chromatin regions are marked by these two histone modifications.

By integrating foetal brain open chromatin maps with other epigenomic data from the human foetal brain (Roadmap Epigenomics Consortium et al. 2015) and summary statistics from recent large-scale GWAS (Wray et al. 2018; Pardiñas et al. 2018; Grove et al. 2019; Stahl et al.

2019; Demontis, Walters, Martin, Mattheisen, Als, Team, et al. 2019), I provide evidence for an early neurodevelopmental component to a range of neuropsychiatric conditions.

Although bipolar disorder is not generally considered to be neurodevelopmental in origin, the present data, showing SNP heritability of bipolar disorder to be enriched within foetal brain OCRs at a similar level to that of schizophrenia and ADHD, are consistent with our previous finding of an enrichment of foetal brain eQTL within common genetic risk variants for the condition (O'Brien et al. 2018) and with recent evidence for altered expression of neurodevelopmental genes in iPSC-derived cerebral organoids generated from patients with bipolar disorder (Kathuria et al. 2020). Emerging evidence suggests that some of the common genetic influences on risk for major depressive disorder also operate *in utero* (Wray et al. 2018) *(*Hall et al. 2020), although we note that, in the present study, the enrichment of SNP heritability for the condition in foetal brain OCRs was consistently lower than that for bipolar disorder and schizophrenia. While rare genetic risk variants for ASD are known to disrupt genes functioning in the prenatal brain (Willsey et al. 2013; Parikshak et al. 2013; Satterstrom et al. 2020), only recently have ASD GWAS yielded sufficient signal for biological insights into the condition (Grove et al. 2019; Forsyth et al. 2020; Hall et al. 2020).

My findings are consistent with results from a recent study (Schork et al. 2019), that reported enrichment of SNP heritability for a broad neuropsychiatric phenotype encompassing ADHD, affective disorder, anorexia, ASD, bipolar disorder and schizophrenia within H3K4me1 and H3K4me3 sites identified by the RoadMap Epigenomics Mapping Consortium (Roadmap Epigenomics Consortium et al. 2015) in bulk human foetal brain tissue. However, it should be noted that the regulatory signal captured by the data presented here derives from a mixture

of different cell types, including progenitors, nascent neurons and non-neuronal cells. Investigation of cell type-specific OCRs is still needed to further disentangle the foetal brain regulatory signal relevant to neuropsychiatric disorders (see chapter 4).

Chapter 4: Chromatin accessibility in neuronal vs non-neuronal cell populations from foetal frontal cortex

4.1 Introduction

In the previous chapters, I provided evidence for an early neurodevelopmental component to a range of neuropsychiatric conditions, by integrating foetal brain open chromatin maps with other epigenomic data from the human foetal brain(Spiers et al. 2015) and summary statistics from recent large-scale GWAS (Grove et al. 2019; Demontis, Walters, Martin, Mattheisen, Als, Neale, et al. 2019) (Stahl et al. 2019) (Pardiñas et al. 2018) (Wray et al. 2018).

The data presented so far, however, were derived from homogenate brain tissue, containing a mixture of markedly different cell types, including progenitors, nascent neurons and non-neuronal cells. As gene regulatory processes such as chromatin accessibility, transcription factor expression and posttranslational modifications can be cell-type specific, the study of mixed cell populations can fail to detect the cell-type-specific regulatory signal that is essential for the maintenance of cell identity and function.

In the attempt to disentangle the human foetal brain regulatory signal, De la Torre-Ubieta and colleagues (de la Torre-Ubieta et al. 2018) produced ATAC-Seq data from two foetal brain areas, the germinal zone and the cortical plate, and performed enrichment analysis of SNP heritability for psychiatric disorders within those regions. Although the cortical areas analysed mostly contained neural progenitors and more mature neurons respectively, they nonetheless included mixed cell types. Therefore, the heritability enrichment analyses performed did not determine how specific cell types contributed to the risk burden.

Given that there is a current lack of functional genomic data derived from cells of the human foetal brain, the first objective of the work described in this chapter was to map cell-type specific gene regulatory sites in the developing human brain. To do this, I extracted neuronal and non-neuronal nuclei from the same cryopreserved foetal brain samples used in chapter 2, by means of fluorescence activated nuclei sorting (FANS) using the neuronal nuclei surface antigen NeuN. I then mapped regions of open chromatin in extracted nuclei using ATAC-Seq and integrated these data with the same histone modification data used in chapter 3, to identify promoters and enhancers.

The second objective of this chapter was to explore the contribution of neuronal and non-neuronal cell population to the heritability of psychiatric disorders. I hypothesised that both mature neurons as well as progenitor cells mediate some of the genetic risk for psychiatric disorders. In order to test this hypothesis, I used Stratified LD score regression (SLDSR) to test NeuN+ and NeuN- open chromatin regions for enrichment of risk SNPs from the same panel of brain disorder GWAS used in the previous chapters.

In the next section, I will illustrate the optimization carried out to ensure specificity of the NeuN antibody and then I will move on to describing the FANS experiments that allowed sorting of NeuN+ and NeuN- from the same bulk foetal brain samples used in chapter 2. I will then illustrate the bioinformatic methods used to identify high confidence open chromatin sites in neuronal and non-neuronal nuclei, annotate them to the genome and restrict those regions to those overlapping histone modifications indicative of promoters and enhancers. I will finally describe and discuss results from the SLDSR analysis within NeuN+ and NeuN- open chromatin regions.

4.2 Methods

4.2.1 FANS optimization

Fluorescence activated cell sorting (FACS) is a flow cytometry method that allows the separation of individual cell populations from a heterogenous mixture of cells based on cellular light scatter and fluorescence characteristics. During this process, mixed cells passing through the sorter are hydrodynamically focussed to shuttle past a laser beam in a steady stream, one cell at a time. As cells are moved through the sorter's interrogation point, light from the laser is transmitted onto each cell and the scattering patterns of the deflected photons are used to measure the size and structural complexity of the cells (Cossarizza et al. 2019). Cells can be further distinguished based on fluorescence characteristics. To facilitate this, antibodies that recognise a target feature on the outer membrane of a cell of interest are pre-labelled with fluorescent compounds called fluorophores, and then incubated with the mixed cell population. As the antibody should only bind to the cell of interest, when mixed cells pass the laser, electrons in the fluorophore are excited into higher orbitals by high energy photons in the laser light. Given that electrons in high orbitals are unstable, the electron quickly returns to its ground state and emits lower energy light, or fluoresces, as it does so. The light that the fluorophore emits can be used to distinguish fluorescently labelled cells from non-labelled cells.

For this project, an antibody targeting a nuclear antigen, rather than a cell surface antigen, was used. The anti-neuronal nuclei (NeuN) antibody had been previously used in the literature to isolate neuronal nuclei in the adult brain. However, it has never been used in the human foetal brain. The choice of an antibody targeting neuronal nuclei was based on the

hypothesis that neuronal cells play a pivotal role in the aetiology of several psychiatric disorders (Bryois et al. 2019).

Studies of the neural tissue proteome and immunocytochemical studies of nervous system organs have established that neurons contain a number of specific proteins, whose appearance in postmitotic cells is indicative of their neuronal differentiation. Some of these proteins are characteristic of only a number of specific neuronal types. Other specific proteins are present in the vast majority of neurons. One of them is the neuronal nuclear protein NeuN, which is often used as a marker of postmitotic neurons due to some of its properties (primarily nuclear localization) (Korzhevskii et al. 2009) . Monoclonal antibodies to the NeuN protein have been largely used in immunohistochemical studies of neuronal differentiation to assess the functional state of neurons. It is believed that NeuN emerges during early embryogenesis in postmitotic neuroblasts and remains in differentiating and terminally differentiated neurons throughout the whole subsequent ontogeny (Gusel'nikova and Korzhevskiy 2015).

In order to minimise unspecific staining, specificity of the anti-neuronal nuclei (NeuN) antibody was tested using cortical interneuron nuclei as a positive control and human white blood cells (WBCs) nuclei as a negative control. Mature (culture day 60) human pluripotent stem-cell-derived cortical interneurons from the H7 human cell line were provided by Prof Meng Li's lab. Interneurons were harvested using Accutase® enzyme, re-suspended in Dulbecco's modified Eagles's Medium (DMEM) and stored at -80 °C in a Mr. Frosty™ Freezing Container (ThermoFisher Scientific) with DMSO for cryopreservation. Human WBCs were isolated by incubating 10ml whole blood tissue mixed with 30ml cold lysis buffer (4g NH_4Cl, 0.5g KCl, 0.1ml 0.5M EDTA) for 15 minutes on ice. Following centrifugation, the supernatant

was removed, the cell pellet re-suspended in lysis buffer and further centrifuged. The resulting WBCs pellet was re-suspended in DMEM and cryopreserved using the same procedure as used for interneurons.

As illustrated in figure 4.1, at a final dilution of 1:2000, the Alexa Fluor 488 conjugated monoclonal antibody caused minimal non-specific staining of WBCs nuclei and substantial staining of neuronal nuclei.
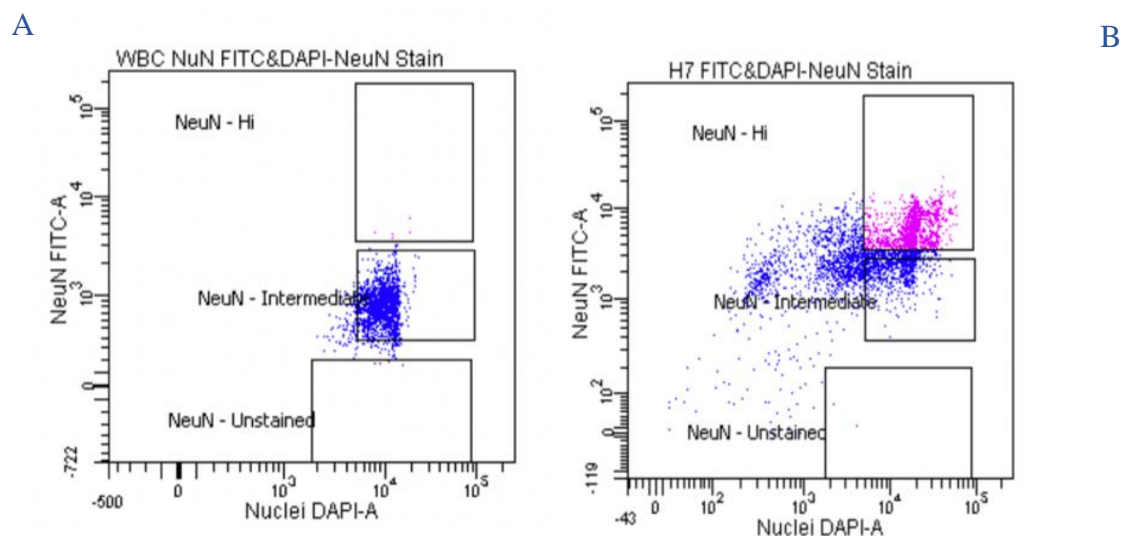
Figure 4.1. FANS optimization. A: NeuN-stained white blood nuclei. B: cortical interneurons. Y-axis is NeuN fluorescence. X-axis is DAPI fluorescence. Gates of NeuN fluorescence were divided in high, intermediate and unstained (low). WBC show intermediate fluorescence (due to non-specific staining), while the majority of cortical interneurons show high NeuN fluorescence.

4.2.2 FANS on foetal brain nuclei

The same 3 cryopreserved frontal cortex cell suspensions samples used in chapter 2 were used for FANS analysis and foetal brain nuclei were isolated as described in chapter 2. To block non-specific antibody binding, the resulting nuclei pellet was resuspended in buffer

containing sucrose 0.25 M, MgCl2 5mM, Tris-Cl 10mM and BSA 1%, and incubated for 30 minutes on ice. Nuclei were visualized under an optical microscope for quality control and counting. The samples were then centrifuged at 400g for 8 mins and the nuclei pellet re-suspended in FANS buffer (0.5% BSA in DPBS).

Before immunostaining, 100µl of the sample was transferred to a new tube and used as unstained control in the FANS analysis. The remaining sample was incubated with mouse anti-neuronal nuclei (NeuN) Alexa Fluor 488 conjugated monoclonal antibody (MerkMillipore Cat# MAB377X), at a 1:2000 dilution, and rotated at 4°C in the dark for 60 mins. After immunostaining, samples were washed 3 times in FANS buffer (400g for 5 min) to remove excess antibody. DAPI was then added to a final concentration of 1µg/ml and the nuclei suspension was filtered through a 35µm cell strainer to remove nuclei clumps and prevent clogging of the cytometer. DAPI positive neuronal (NeuN+) and non-neuronal (NeuN-) nuclei were sorted into tubes pre-coated with 5% BSA using a FACSAria flow cytometer (BD Biosciences) equipped with a 100µm nozzle. For both flow analysis and FANS, exclusion of debris using forward and side scatter pulse area parameters (FSC-A and SSC-A) were gated first, followed by exclusion of aggregates using pulse width (FSC-W and SSC-A) and DAPI-positive nuclei before gating populations based on NeuN fluorescence (Figure 4.2).
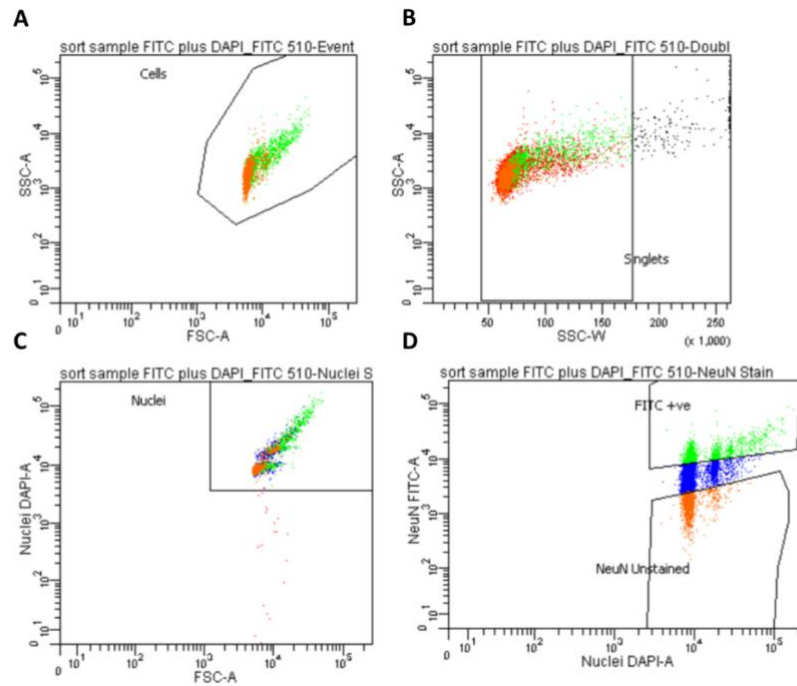
Figure 4.2. Isolation of NeuN+ and NeuN- nuclei using fluorescence-activated nuclei sorting (FANS). First, debris was excluded using forward and side scatter pulse area parameters (FSC-A and SSC-A) (A). This was followed by exclusion of nuclei aggregates using pulse width (FSC-W and SSC-A) (B) and isolation of DAPI-positive nuclei (C), before gating populations based on NeuN fluorescence (D).

4.2.3 Assay for Transposase-Accessible Chromatin and sequencing data analysis

Sorted NeuN+ and NeuN- nuclei were incubated in transposition reaction mix and PCR amplified as for bulk tissue (see chapter 2). Sufficient nuclei were recovered to perform two technical replicates for 2 of the 3 frontal cortex samples and 1 technical replicate for the other, resulting in 5 separate transposase reactions for NeuN+ nuclei (mean number of nuclei per reaction = 41,216) and 5 separate transposase reactions for NeuN- nuclei (mean number of nuclei per reaction = 44,066). Transposase reactions, PCR amplification, size selection and sequencing of ATAC-Seq libraries from neuronal and non-neuronal nuclei were performed exactly as described in chapter 2.

To identify high confidence NeuN+ and NeuN- open chromatin regions, I identified subsets of high confidence bulk open chromatin sites (produced in chapter 2) that could be conservatively attributed to NeuN+ and / or NeuN- fractions. In order to do this, I selected those open chromatin sites overlapping the (FDR < 0.01) peaks observed in at least 2 of the 3 biological replicates samples for each sorted nuclei population.

4.2.4 Bioinformatic annotation of open chromatin regions

Enrichment of biological process Gene Ontology (GO) annotations for genes with foetal frontal cortex OCRs within 30kb upstream and 100bp downstream of their TSS was tested using g:Profiler (Raudvere et al. 2019), correcting for multiple testing using the default g:SCS algorithm. g:Profiler is an online tool (https://biit.cs.ut.ee/gprofiler/) that performs functional enrichment analysis on gene lists, by mapping genes to known functional information datasets and detecting statistically significant enriched terms. g:Profiler allows for multiple testing correction, using different methods: the Bejamini Hochberg FDR, Bonferroni or g:SCS algorithm. g:SCS, which corresponds to an experiment-wide threshold of a=0.05, i.e. at least 95% of matches above threshold are statistically significant, was chosen because while Bonferroni correction and Benjamini-Hochberg FDR are designed for multiple tests that are independent of each other, a GO analysis consists of hierarchically related general and specific terms and genes associated to a given GO term are implicitly associated to all more general parents of this term.

4.2.5 Intersection with histone modification datasets from human foetal brain

Genomic regions marked by H3K4Me1 or H3K4Me3, downloaded from the Roadmap epigenomics project (https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/) (Roadmap Epigenomics Consortium et al. 2015) in chapter 3, were used for intersection with neuronal and non-neuronal foetal brain open chromatin regions. BEDTools (Quinlan and Hall 2010) v2.26.0 intersect (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html) 4.2.6 and BEDTools merge (https://bedtools.readthedocs.io/en/latest/content/tools/merge.html) were used exactly as described in chapter 3 to identify H3K4Me1 and H3K4Me3 sites overlapping open chromatin regions in neuronal and non-neuronal nuclei from the foetal frontal cortex.

4.2.6 Testing enrichment of SNP heritability for neuropsychiatric disorders in open chromatin regions

I used stratified LD score regression (SLDSR) (Finucane et al. 2015) to test for enrichment of SNP heritability for neuropsychiatric disorders (ADHD (Demontis, Walters, Martin, Mattheisen, Als, Team, et al. 2019), autism spectrum disorder (Grove et al. 2019), bipolar disorder (Stahl et al. 2019), major depressive disorder (Wray et al. 2018) and schizophrenia (Pardiñas et al. 2018)) and two control traits with similar sample sizes (height (Lango Allen et al. 2010) and blood triglyceride levels (Teslovich et al. 2010a)) in NeuN+ and NeuN- open chromatin sites and NeuN+ and NeuN- open chromatin sites overlapping H3K4Me1 and H3K4Me3 regions.

4.3 Results

In the adult brain, common genetic risk for schizophrenia has been reported to be primarily mediated through OCRs in NeuN+ (neuronal), rather than NeuN- (non-neuronal), nuclei (Fullard et al. 2018). To explore the cellular basis of genetic risk for neuropsychiatric disorders in the prenatal brain, I determined which of the previously identified foetal frontal cortex OCRs could be confidently attributed to NeuN+ and / or NeuN- fractions by fluorescence-activated sorting nuclei from the same 3 foetal frontal cortex samples. This identified 13,262 high confidence foetal frontal cortex OCRs that were also observed in sorted NeuN+ (neuron-enriched) nuclei and 20,782 such OCRs that were also observed in sorted NeuN- (neuron-depleted) nuclei.

4.3.1 Bioinformatic annotation of open chromatin regions

Genes with high confidence OCRs within 30kb upstream of their TSS that were detected in NeuN- but not NeuN+ nuclei were most significantly enriched for the GO term 'cellular metabolic process' (*P corrected* = 1.4E-25). In contrast, genes with high confidence OCRs within 30kb upstream of their TSS that were detected in NeuN+ but not NeuN- nuclei were most significantly enriched for the GO term 'anion transport' (*P corrected* = 9.0E-03), driven by neuronal markers such as *CACNA1A*, *RIMS1*, *SLC1A7*, *BDNF and GLS2*, consistent with successful enrichment of neuronal nuclei.

4.3.2 Intersection with histone modification datasets from human foetal brain

A total of 95,699 and 26,609 ChIP-Seq peaks markers of H3K4Me1 and H3K4Me3 respectively were downloaded from the Roadmap epigenomics project. The intersection and merge of these peaks with 13,262 high confidence foetal frontal cortex NeuN+ ATAC-Seq peaks resulted in 329 and 5,426 genomic sites respectively. The intersection and merge of H3K4Me1 and H3K4Me3 peaks with 20,782 high confidence foetal frontal cortex NeuN- ATAC-Seq peaks resulted in 2,531 and 14,167 genomic sites respectively.

4.3.3 Testing enrichment of SNP heritability for neuropsychiatric disorders in open chromatin regions

I tested for enrichment of single nucleotide polymorphism (SNP) heritability for 5 major neuropsychiatric disorders (ADHD, autism spectrum disorder [ASD], bipolar disorder, major depressive disorder and schizophrenia) within bulk foetal frontal cortex OCRs using Stratified Linkage Disequilibrium Score Regression (SLDSR), controlling for general genomic annotations (e.g., coding regions, promoters, enhancers and conserved regions) included in the baseline model (Finucane et al., 2015) to obtain Z-score P-values.

Although the relatively small number of foetal frontal cortex OCRs confidently attributed to NeuN+ and NeuN- nuclei (encompassing < 1% of SNPs) conferred limited statistical power to identify enrichment of SNP heritability by SLDSR, I nevertheless observed significant ($P < 0.05$, uncorrected) enrichment of SNP heritability for schizophrenia in NeuN+ OCRs overlapping H3K4me1 sites (Figure 4.3) and for all 5 tested neuropsychiatric disorders in NeuN- OCRs overlapping either H3K4me1 or H3K4me3 sites (Figure 4.4).

Estimated enrichments of SNP heritability were similar between bulk and NeuN- nuclei OCRs overlapping H3K4me1 sites for all neuropsychiatric disorders (approximately 10-fold). In contrast, ADHD showed a notable lack of enrichment within all foetal NeuN+ annotations (< 1-fold enrichment across all; Table 4.1), suggesting the lesser importance of nascent neurons within the prenatal frontal cortex in the genetic aetiology of this condition.
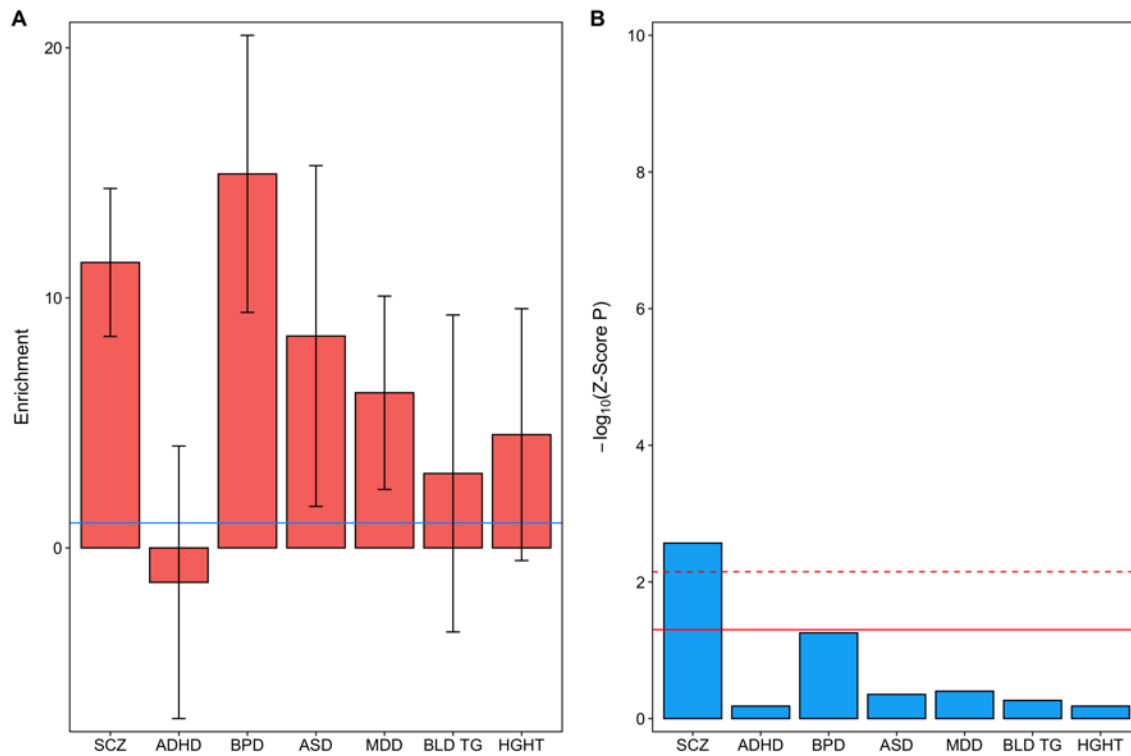
Figure 4.3. Taken from Kouakou et al., 2020, under review. Created by Darren Cameron. **Partitioned heritability for 5 neuropsychiatric disorders and 2 control traits within high confidence open chromatin regions observed in foetal brain NeuN+ nuclei.** a) Fold-enrichment estimates of SNP heritability for each trait (the proportion of SNP heritability explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation). Error bars represent standard error. The solid blue horizontal line indicates no enrichment. b) -Log$_{10}$ Z-score *P*-values for enrichment of SNP heritability, controlling for general genomic annotations included in the baseline model of Finucane and colleagues (Finucane et al., 2015). The solid red horizontal line indicates the Z-score *P*-value 0.05 threshold; the dashed red horizontal line indicates the threshold for Z-score *P*-values surviving Bonferroni correction for 7 tested traits. SCZ = schizophrenia; ADHD = attention deficit hyperactivity disorder; BPD = bipolar disorder; MDD = major depressive disorder; ASD = autism spectrum disorder; BLD TG = blood triglyceride levels; HGHT = height.

Figure 4.4. Taken from Kouakou et al., 2020, under review. Created by Darren Cameron. **Partitioned heritability for 5 neuropsychiatric disorders and 2 control traits within high confidence open chromatin regions observed in foetal brain NeuN- nuclei.** a) Fold-enrichment estimates of SNP heritability for each trait (the proportion of SNP heritability explained by SNPs within the annotation divided by the proportion of genome-wide SNPs within the annotation). Error bars represent standard error. The solid blue horizontal line indicates no enrichment. b) -Log$_{10}$ Z-score *P*-values for enrichment of SNP heritability, controlling for general genomic annotations included in the baseline model of Finucane and colleagues (Finucane et al., 2015). The solid red horizontal line indicates the Z-score *P*-value 0.05 threshold; the dashed red horizontal line indicates the threshold for Z-score *P*-values surviving Bonferroni correction for 7 tested traits. SCZ = schizophrenia; ADHD = attention deficit hyperactivity disorder; BPD = bipolar disorder; MDD = major depressive disorder; ASD = autism spectrum disorder; BLD TG = blood triglyceride levels; HGHT = height.

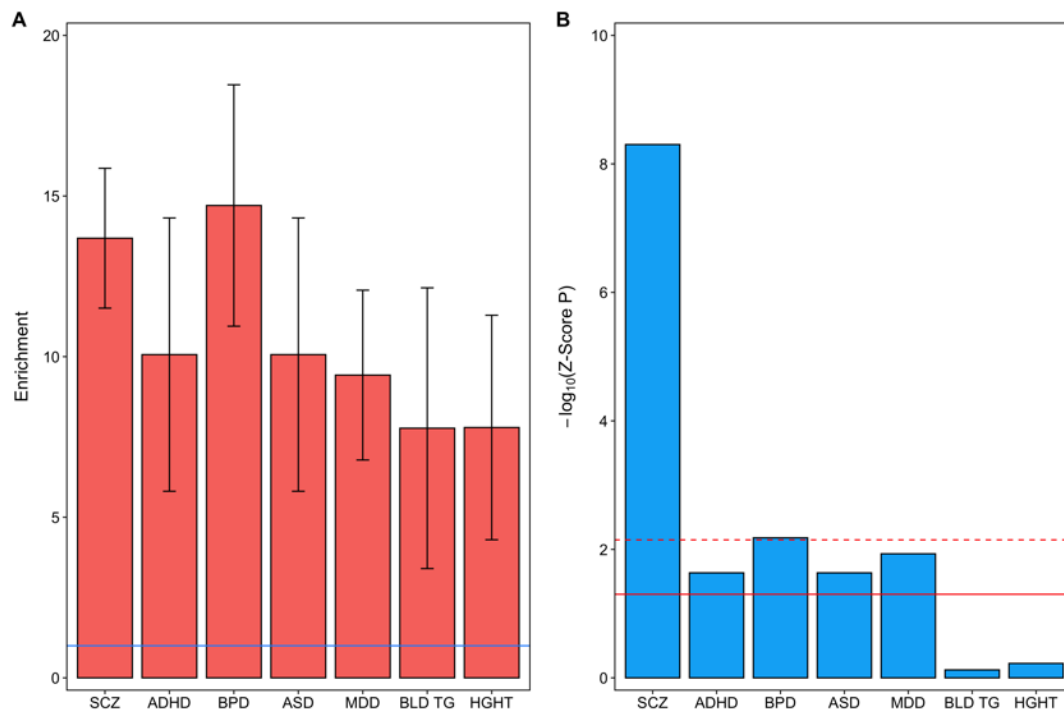| Peak set	Trait	SNPs		Enrichment		Coefficient	
NeuN+ ATAC peaks							
	Schizophrenia	Proportion of SNPs in annotation	6.024E-03	Enrichment	4.577E+00	Coefficient	1.609E-07
		Proportion SNP h2 explained	2.757E-02	Enrichment SE	1.979E+00	Coefficient SE	1.273E-07
				Enrichment P	7.283E-02	Coefficient Z-score	1.264E+00
						Z-score P (2-tailed)	2.063E-01
	ADHD	Proportion of SNPs in annotation	6.024E-03	Enrichment	6.757E-01	Coefficient	-2.888E-09
		Proportion SNP h2 explained	4.070E-03	Enrichment SE	3.598E+00	Coefficient SE	1.35401593717e
				Enrichment P	9.282E-01	Coefficient Z-score	-2.133E-02
						Z-score P (2-tailed)	9.830E-01
	Bipolar disorder	Proportion of SNPs in annotation	6.024E-03	Enrichment	8.632E+00	Coefficient	2.707E-07
		Proportion SNP h2 explained	5.200E-02	Enrichment SE	3.532E+00	Coefficient SE	1.676E-07
				Enrichment P	2.937E-02	Coefficient Z-score	1.615E+00
						Z-score P (2-tailed)	1.062E-01
	Autism spectrum disorder	Proportion of SNPs in annotation	6.024E-03	Enrichment	2.212E+00	Coefficient	2.774E-09
		Proportion SNP h2 explained	1.332E-02	Enrichment SE	4.558E+00	Coefficient SE	1.414E-07
				Enrichment P	7.889E-01	Coefficient Z-score	1.962E-02
						Z-score P (2-tailed)	9.843E-01
	Major depressive disorder	Proportion of SNPs in annotation	6.024E-03	Enrichment	2.066E+00	Coefficient	1.606E-09
		Proportion SNP h2 explained	1.244E-02	Enrichment SE	2.421E+00	Coefficient SE	1.953E-08
				Enrichment P	6.600E-01	Coefficient Z-score	8.219E-02
						Z-score P (2-tailed)	9.345E-01
	Blood triglyceride levels	Proportion of SNPs in annotation		Enrichment	1.698E+00	Coefficient	-5.716E-08
		Proportion SNP h2 explained	1.023E-02	Enrichment SE	4.236E+00	Coefficient SE	8.838E-08
				Enrichment P	8.679E-01	Coefficient Z-score	-6.468E-01
						Z-score P (2-tailed)	5.178E-01
	Height	Proportion of SNPs in annotation	6.024E-03	Enrichment	1.656E+00	Coefficient	-9.906E-08
		Proportion SNP h2 explained	9.973E-03	Enrichment SE	2.823E+00	Coefficient SE	9.758E-08
				Enrichment P	8.158E-01	Coefficient Z-score	-1.015E+00
						Z-score P (2-tailed)	3.100E-01
NeuN+ ATAC peaks intersecting foetal brain H3K4me1 sites							
	Schizophrenia	Proportion of SNPs in annotation	3.169E-03	Enrichment	1.141E+01	Coefficient	5.780E-07
		Proportion SNP h2 explained	3.617E-02	Enrichment SE	2.959E+00	Coefficient SE	1.924E-07
				Enrichment P	5.268E-04	Coefficient Z-score	3.003E+00
						Z-score P (2-tailed)	2.671E-03
	ADHD	Proportion of SNPs in annotation	3.169E-03	Enrichment	-1.377E+00	Coefficient	-9.450E-08
		Proportion SNP h2 explained	-4.364E-03	Enrichment SE	5.450E+00	Coefficient SE	2.122E-07
				Enrichment P	6.641E-01	Coefficient Z-score	-4.454E-01
						Z-score P (2-tailed)	6.561E-01
	Bipolar disorder	Proportion of SNPs in annotation	3.169E-03	Enrichment	1.496E+01	Coefficient	5.119E-07
		Proportion SNP h2 explained	4.740E-02	Enrichment SE	5.541E+00	Coefficient SE	2.674E-07
				Enrichment P	1.094E-02	Coefficient Z-score	1.914E+00
						Z-score P (2-tailed)	5.557E-02
	Autism spectrum disorder	Proportion of SNPs in annotation	3.169E-03	Enrichment	8.473E+00	Coefficient	1.628E-07
		Proportion SNP h2 explained	2.685E-02	Enrichment SE	6.814E+00	Coefficient SE	2.120E-07
				Enrichment P	2.574E-01	Coefficient Z-score	7.677E-01
						Z-score P (2-tailed)	4.427E-01
	Major depressive disorder	Proportion of SNPs in annotation	3.169E-03	Enrichment	6.204E+00	Coefficient	2.765E-08
		Proportion SNP h2 explained	1.966E-02	Enrichment SE	3.867E+00	Coefficient SE	3.266E-08
				Enrichment P	1.777E-01	Coefficient Z-score	8.467E-01
						Z-score P (2-tailed)	3.971E-01
	Blood triglyceride levels	Proportion of SNPs in annotation	3.169E-03	Enrichment	2.976E+00	Coefficient	-8.477E-08
		Proportion SNP h2 explained	9.430E-03	Enrichment SE	6.339E+00	Coefficient SE	1.389E-07
				Enrichment P	7.503E-01	Coefficient Z-score	-6.103E-01
						Z-score P (2-tailed)	5.417E-01
	Height	Proportion of SNPs in annotation	3.169E-03	Enrichment	4.530E+00	Coefficient	-8.117E-08
		Proportion SNP h2 explained	1.435E-02	Enrichment SE	5.037E+00	Coefficient SE	1.824E-07
				Enrichment P	4.825E-01	Coefficient Z-score	-4.449E-01
						Z-score P (2-tailed)	6.564E-01
NeuN+ ATAC peaks intersecting foetal brain H3K4me3 sites							
	Schizophrenia	Proportion of SNPs in annotation	3.462E-03	Enrichment	8.194E+00	Coefficient	3.482E-07
		Proportion SNP h2 explained	2.836E-02	Enrichment SE	2.849E+00	Coefficient SE	1.839E-07
				Enrichment P	1.206E-02	Coefficient Z-score	1.894E+00
						Z-score P (2-tailed)	5.821E-02
	ADHD	Proportion of SNPs in annotation	3.462E-03	Enrichment	-3.103E-01	Coefficient	-5.250E-08
		Proportion SNP h2 explained	-1.074E-03	Enrichment SE	5.271E+00	Coefficient SE	2.068E-07
				Enrichment P	8.041E-01	Coefficient Z-score	-2.539E-01
						Z-score P (2-tailed)	7.996E-01
	Bipolar disorder	Proportion of SNPs in annotation	3.462E-03	Enrichment	1.333E+01	Coefficient	4.200E-07
		Proportion SNP h2 explained	4.614E-02	Enrichment SE	5.322E+00	Coefficient SE	2.591E-07
				Enrichment P	1.902E-02	Coefficient Z-score	1.621E+00
						Z-score P (2-tailed)	1.049E-01
	Autism spectrum disorder	Proportion of SNPs in annotation	3.462E-03	Enrichment	7.326E+00	Coefficient	1.262E-07
		Proportion SNP h2 explained	2.536E-02	Enrichment SE	6.181E+00	Coefficient SE	1.993E-07
				Enrichment P	2.964E-01	Coefficient Z-score	6.335E-01
						Z-score P (2-tailed)	5.264E-01
	Major depressive disorder	Proportion of SNPs in annotation	3.462E-03	Enrichment	5.066E+00	Coefficient	1.772E-08
		Proportion SNP h2 explained	1.754E-02	Enrichment SE	3.787E+00	Coefficient SE	3.226E-08
				Enrichment P	2.821E-01	Coefficient Z-score	5.492E-01
						Z-score P (2-tailed)	5.828E-01
	Blood triglyceride levels	Proportion of SNPs in annotation	3.462E-03	Enrichment	5.789E+00	Coefficient	-2.842E-08
		Proportion SNP h2 explained	2.004E-02	Enrichment SE	6.377E+00	Coefficient SE	1.391E-07
				Enrichment P	4.385E-01	Coefficient Z-score	-2.043E-01
						Z-score P (2-tailed)	8.381E-01
	Height	Proportion of SNPs in annotation	3.462E-03	Enrichment	5.394E+00	Coefficient	-5.553E-08
		Proportion SNP h2 explained	1.867E-02	Enrichment SE	4.770E+00	Coefficient SE	1.760E-07
				Enrichment P	3.564E-01	Coefficient Z-score	-3.154E-01
						Z-score P (2-tailed)	7.524E-01

Table 4.1. Partitioned heritability results for high confidence open chromatin regions in NeuN+ cells from foetal frontal cortex

Peak set	Trait	SNPs		Enrichment		Coefficient	
NeuN- ATAC peaks							
	Schizophrenia	Proportion of SNPs in annotation	9.211E-03	Enrichment	8.089E+00	Coefficient	4.092E-07
		Proportion SNP h2 explained	7.451E-02	Enrichment SE	1.601E+00	Coefficient SE	1.068E-07
				Enrichment P	1.892E-05	Coefficient Z-score	3.832E+00
						Z-score P (2-tailed)	1.271E-04
	ADHD	Proportion of SNPs in annotation	9.211E-03	Enrichment	7.377E+00	Coefficient	2.747E-07
		Proportion SNP h2 explained	6.795E-02	Enrichment SE	3.141E+00	Coefficient SE	1.231E-07
				Enrichment P	4.059E-02	Coefficient Z-score	2.231E+00
						Z-score P (2-tailed)	2.568E-02
	Bipolar disorder	Proportion of SNPs in annotation	9.211E-03	Enrichment	1.011E+01	Coefficient	3.464E-07
		Proportion SNP h2 explained	9.311E-02	Enrichment SE	2.739E+00	Coefficient SE	1.441E-07
				Enrichment P	9.709E-04	Coefficient Z-score	2.404E+00
						Z-score P (2-tailed)	1.620E-02
	Autism spectrum disorder	Proportion of SNPs in annotation	9.211E-03	Enrichment	1.072E+01	Coefficient	2.908E-07
		Proportion SNP h2 explained	9.873E-02	Enrichment SE	4.193E+00	Coefficient SE	1.309E-07
				Enrichment P	1.816E-02	Coefficient Z-score	2.221E+00
						Z-score P (2-tailed)	2.633E-02
	Major depressive disorder	Proportion of SNPs in annotation	9.211E-03	Enrichment	5.644E+00	Coefficient	3.053E-08
		Proportion SNP h2 explained	5.199E-02	Enrichment SE	2.122E+00	Coefficient SE	1.864E-08
				Enrichment P	2.975E-02	Coefficient Z-score	1.638E+00
						Z-score P (2-tailed)	1.015E-01
	Blood triglyceride levels	Proportion of SNPs in annotation	9.211E-03	Enrichment	4.940E+00	Coefficient	-1.862E-09
		Proportion SNP h2 explained	4.550E-02	Enrichment SE	3.328E+00	Coefficient SE	6.930E-08
				Enrichment P	2.364E-01	Coefficient Z-score	-2.687E-02
						Z-score P (2-tailed)	9.786E-01
	Height	Proportion of SNPs in annotation	9.211E-03	Enrichment	5.546E+00	Coefficient	3.554E-08
		Proportion SNP h2 explained	5.108E-02	Enrichment SE	2.409E+00	Coefficient SE	8.984E-08
				Enrichment P	6.051E-02	Coefficient Z-score	3.956E-01
						Z-score P (2-tailed)	6.924E-01
NeuN- ATAC peaks intersecting foetal brain H3K4me1 sites							
	Schizophrenia	Proportion of SNPs in annotation	5.905E-03	Enrichment	1.369E+01	Coefficient	7.978E-07
		Proportion SNP h2 explained	8.081E-02	Enrichment SE	2.177E+00	Coefficient SE	1.363E-07
				Enrichment P	1.842E-08	Coefficient Z-score	5.855E+00
						Z-score P (2-tailed)	4.765E-09
	ADHD	Proportion of SNPs in annotation	5.905E-03	Enrichment	1.006E+01	Coefficient	3.946E-07
		Proportion SNP h2 explained	5.942E-02	Enrichment SE	4.254E+00	Coefficient SE	1.738E-07
				Enrichment P	3.242E-02	Coefficient Z-score	2.270E+00
						Z-score P (2-tailed)	2.319E-02
	Bipolar disorder	Proportion of SNPs in annotation	5.905E-03	Enrichment	1.470E+01	Coefficient	5.496E-07
		Proportion SNP h2 explained	8.683E-02	Enrichment SE	3.758E+00	Coefficient SE	2.024E-07
				Enrichment P	2.793E-04	Coefficient Z-score	2.716E+00
						Z-score P (2-tailed)	6.603E-03
	Autism spectrum disorder	Proportion of SNPs in annotation	5.905E-03	Enrichment	1.006E+01	Coefficient	3.946E-07
		Proportion SNP h2 explained	5.942E-02	Enrichment SE	4.254E+00	Coefficient SE	1.738E-07
				Enrichment P	3.242E-02	Coefficient Z-score	2.270E+00
						Z-score P (2-tailed)	2.319E-02
	Major depressive disorder	Proportion of SNPs in annotation	5.905E-03	Enrichment	9.425E+00	Coefficient	5.948E-08
		Proportion SNP h2 explained	5.565E-02	Enrichment SE	2.643E+00	Coefficient SE	2.360E-08
				Enrichment P	1.538E-03	Coefficient Z-score	2.520E+00
						Z-score P (2-tailed)	1.172E-02
	Blood triglyceride levels	Proportion of SNPs in annotation	5.905E-03	Enrichment	7.772E+00	Coefficient	2.851E-08
		Proportion SNP h2 explained	4.589E-02	Enrichment SE	4.370E+00	Coefficient SE	9.021E-08
				Enrichment P	1.170E-01	Coefficient Z-score	3.160E-01
						Z-score P (2-tailed)	7.520E-01
	Height	Proportion of SNPs in annotation	5.905E-03	Enrichment	7.794E+00	Coefficient	7.039E-08
		Proportion SNP h2 explained	4.602E-02	Enrichment SE	3.494E+00	Coefficient SE	1.328E-07
				Enrichment P	5.444E-02	Coefficient Z-score	5.301E-01
						Z-score P (2-tailed)	5.960E-01
NeuN- ATAC peaks intersecting foetal brain H3K4me3 sites							
	Schizophrenia	Proportion of SNPs in annotation	6.149E-03	Enrichment	1.159E+01	Coefficient	6.557E-07
		Proportion SNP h2 explained	7.125E-02	Enrichment SE	2.211E+00	Coefficient SE	1.442E-07
				Enrichment P	2.681E-06	Coefficient Z-score	4.548E+00
						Z-score P (2-tailed)	5.400E-06
	ADHD	Proportion of SNPs in annotation	6.149E-03	Enrichment	1.073E+01	Coefficient	4.486E-07
		Proportion SNP h2 explained	6.601E-02	Enrichment SE	4.200E+00	Coefficient SE	1.806E-07
				Enrichment P	1.967E-02	Coefficient Z-score	2.484E+00
						Z-score P (2-tailed)	1.300E-02
	Bipolar disorder	Proportion of SNPs in annotation	6.149E-03	Enrichment	1.549E+01	Coefficient	6.037E-07
		Proportion SNP h2 explained	9.523E-02	Enrichment SE	3.833E+00	Coefficient SE	2.175E-07
				Enrichment P	1.911E-04	Coefficient Z-score	2.775E+00
						Z-score P (2-tailed)	5.517E-03
	Autism spectrum disorder	Proportion of SNPs in annotation	6.149E-03	Enrichment	1.740E+01	Coefficient	5.277E-07
		Proportion SNP h2 explained	1.070E-01	Enrichment SE	5.234E+00	Coefficient SE	1.637E-07
				Enrichment P	1.099E-03	Coefficient Z-score	3.224E+00
						Z-score P (2-tailed)	1.265E-03
	Major depressive disorder	Proportion of SNPs in annotation	6.149E-03	Enrichment	8.615E+00	Coefficient	5.385E-08
		Proportion SNP h2 explained	5.297E-02	Enrichment SE	2.723E+00	Coefficient SE	2.578E-08
				Enrichment P	5.587E-03	Coefficient Z-score	2.089E+00
						Z-score P (2-tailed)	3.672E-02
	Blood triglyceride levels	Proportion of SNPs in annotation	6.149E-03	Enrichment	9.205E+00	Coefficient	5.664E-08
		Proportion SNP h2 explained	5.660E-02	Enrichment SE	4.619E+00	Coefficient SE	1.021E-07
				Enrichment P	7.690E-02	Coefficient Z-score	5.546E-01
						Z-score P (2-tailed)	5.791E-01
	Height	Proportion of SNPs in annotation	6.149E-03	Enrichment	9.042E+00	Coefficient	1.156E-07
		Proportion SNP h2 explained	5.560E-02	Enrichment SE	3.340E+00	Coefficient SE	1.294E-07
				Enrichment P	1.728E-02	Coefficient Z-score	8.932E-01
						Z-score P (2-tailed)	3.718E-01

Table 4.2. Partitioned heritability results for high confidence open chromatin regions in NeuN- cells from foetal frontal cortex

4. 4 Discussion

In this chapter, I provide the first maps of open chromatin in NeuN+ and NeuN- cell populations of the human foetal brain. Neuronal-enriched nuclei were separated from homogenate brain tissue by FANS, targeting the neuronal nuclear antigen NeuN, encoded by the RBFOX3 gene. Although the anti-NeuN antibody had been previously used for FANS in adult brain tissue (Fullard et al., 2017), its specificity was tested here to ensure minimal non-specific staining and maximum fluorescence in a negative and a positive control respectively.

In addition, bioinformatic annotation of open chromatin peaks confirmed successful enrichment of neuronal nuclei. GO analysis of genes with high confidence OCRs within 30kb upstream of their TSS that were detected in NeuN+ but not NeuN- nuceli showed enrichment of the GO term 'anion transport' (*P corrected* = 9E-03), driven by neuronal markers (*CACNA1A*, *RIMS1*, *SLC1A7*, *BDNF and GLS2*).By integrating NeuN+ and NeuN- open chromatin maps with other epigenomic data from the human foetal brain (Roadmap Epigenomics Consortium et al. 2015) and summary statistics from recent large-scale GWAS (Pardiñas et al. 2018) (Wray et al. 2018) (Demontis, Walters, Martin, Mattheisen, Als, Neale, et al. 2019) (Grove et al. 2019) (Stahl et al. 2019),

I highlight an important role for regulatory regions active within non-neuronal cells of the prenatal brain in susceptibility to neuropsychiatric disorders. In contrast to the adult brain, where genetic risk for schizophrenia appears to be largely mediated by neuronal (NeuN+) cells (Fullard et al. 2018), I find that, within the foetal brain, genetic risk for this and other

neuropsychiatric conditions is at least as strongly enriched within high confidence OCRs observed in NeuN- nuclei. However, whereas NeuN- cells in the adult brain largely consist of mature glia (astrocytes and oligodendrocytes), those of the second trimester foetal brain encompass a variety of developing cells, including cycling neural progenitors, radial glia and oligodendrocyte precursors. My data are therefore consistent with the findings of de la Torre-Ubieta and colleagues (de la Torre-Ubieta et al. 2018), who report enrichment of SNP heritability for neuropsychiatric disorders in OCRs of the neural progenitor cell-containing germinal zone, and of Schork and colleagues (Schork et al. 2019), who report enriched expression of fine-mapped candidates genes for a broad neuropsychiatric phenotype in foetal radial glia.

Although I was able perform ATAC-Seq in NeuN+ and NeuN- nuclei, these still encompass diverse cell populations (Fan et al., 2018). It is now possible to combine ATAC-Seq with single cell profiling (Cusanovich et al Cell. 2018), which, in future, should provide greater cellular resolution of regulatory genomic sites in the foetal brain with which to test for enrichment of genetic signal for neuropsychiatric and other brain disorders.

Genome-wide mapping of regulatory elements in large ensembles of cells have uncovered substantial variation in chromatin structure across cell types, particularly at distal regulatory regions (Feingold et al. 2004). Single cell-based chromatin accessibility studies can be used for identification of subpopulations in a heterogeneous biological sample, and for identification of the regulatory elements active in each subpopulation.

Several single cell sequencing techniques have been developed to profile the chromatin accessibility in single cells. For instance, one approach relies on isolation of cell using microfluidic devices (Fluidigm, C1) (Jason D Buenrostro et al. 2015). Another type of

approach involves combinatorial indexing to simultaneously analyse tens of thousands of

cells (Cusanovich et al. 2015).

Chapter 5: Prioritization of putative schizophrenia SNPs for further functional genomic analyses

5.1 Introduction

Given that the majority of common genetic loci for complex disorders are located within non-coding DNA, in the previous chapters, I investigated non-coding genomic regions in a tissue and time point relevant to several brain conditions. I mapped human foetal frontal cortex regulatory genomic regions and showed that they are enriched in GWAS genetic risk variants for a number of psychiatric disorders. It is hypothesised that these risk variants perturb the expression of genes in the developing brain, by disrupting active promoters and enhancers.

An important limitation of GWAS is that, due to linkage disequilibrium, several SNPs at a locus typically display a similar level of association and therefore identifying the causal variant is a major challenge. Since performing functional studies on all the genome-wide significant variants would be extremely costly and time consuming, prioritizing potentially causal SNPs is essential.

The main objective of this chapter was therefore to prioritize potentially functional risk SNPs for schizophrenia. To do this, I identified SNPs within foetal brain open chromatin regions that were in high linkage disequilibrium ($r2 > 0.8$) with the most significant (index) SNP at each genome-wide significant schizophrenia risk locus. In order to further characterise these SNPs, I integrated my foetal brain open chromatin data with foetal brain eQTL data from a previous study (O'Brien et al., 2018) and identified those open chromatin SNPs that are significant eQTLs for transcripts in the foetal brain and the degree of linkage disequilibrium between them and the most significant eQTLs for those transcripts.

In the next section, I will describe the method used to detect SNPs located within open chromatin regions and identify those in high likage disequilibrium with both a schizophrenia index SNP and a top eQTL for a foetal brain transcript. I will finally provide the list of potentially functional SNPs.

5.2 Methods

Identification of SNPs within open chromatin regions in strong linkage disequilibrium with schizophrenia index SNPs

GARFIELD (GWAS analysis of regulatory or functional information enrichment with linkage disequilibrium correction) (Iotchkova et al. 2019) was used to identify SNPs associated with schizophrenia at genome-wide significance (Pardinas et al., 2018) that overlapped bulk foetal frontal cortex open chromatin regions. SNP information was provided with the GARFIELD software and was derived from the UK10K sequence data (Walter et al. 2015).

GARFIELD tests for enrichment of SNPs associated with the trait at specific p-value thresholds, by comparing these to an independent set of SNPs matched for minor allele frequency, distance to the nearest transcription start site, number of linkage disequilibrium proxies and GC content.

Bedtools (Quinlan et al., 2010) was used to intersect SNP base positions with open chromatin regions derived from the ATAC-Seq experiments, and each SNP was annotated with a 1 or 0 to indicate whether or not it fell within an open chromatin region. GWAS summary statistics were prepared using the garfield-create-input- gwas.sh (code example provided in the appendix) script provided to retain only the base position and trait-associated p- value information for each SNP. GARFIELD required four additional files, all provided with the

software, for the analyses. These included two LD tag files, derived from the UK10K study, that contain LD information for each SNP at two different thresholds ($r^2 >= 0.1$ and $r^2 >= 0.8$).

To identify SNPs that overlapped regulatory regions for the GARFIELD analyses, custom annotation files were created. SNP information was provided with the GARFIELD software and was derived from the UK10K sequence data (Walter et al. 2015). Bedtools (Quinlan et al., 2010) was used to intersect SNP base positions with open chromatin regions derived from the ATAC-seq experiments, and each SNP was annotated with a 1 or 0 to indicate whether or not it fell within an open chromatin region. GWAS summary statistics were prepared using the garfield-create-input- gwas.sh script provided to retain only the base position and trait-associated p- value information for each SNP. GARFIELD required four additional files, all provided with the software, for the analyses. These included two LD tag files, derived from the UK10K study, that contain LD information for each SNP at two different thresholds ($r^2 >= 0.1$ and $r^2 >= 0.8$). To reduce the set of GWAS variants included in the analysis to a more computationally tractable independent set of SNPs, the lower LD r^2 threshold was used for a process called greedy pruning. This involved partitioning the genome into 1Mb windows, and for each window, sequentially removing SNPs with r^2 value > 0.1 relative to the most significant trait- associated variant and retaining the next most significant independent SNP (with $r^2 < 0.1$ being taken as an approximate measure of SNP independence) until a pruned set of independent SNPs was obtained. The higher threshold was used as the inclusion criteria for SNPs overlapping a functional region, i.e. a SNP was considered overlapping if it, or one of its LD proxies within 500kb ($r^2 >= 0.8$), was located within the open chromatin region. Finally,

files to match variants by minor allele frequency and distance to the nearest transcription start site were included, the latter to correct for bias driven by local gene density.

The online tool LDpair was used within the NIH national Cancer Institute LDlink (https://ldlink.nci.nih.gov/) to identify those SNPs in OCRs that were in strong linkage disequilibrium (r2 > 0.8) with the most significant (index) SNP at each genome-wide significant schizophrenia risk locus in the CEU population.

Publicly available reference haplotypes from Phase 3 of the 1000 Genomes Project are used by LDlink to calculate population-specific measures of linkage disequilibrium. Haplotypes are available for continental populations (ex: European, African, and Admixed American). The CEU population Northwestern Europeans was used given that all data used in this study is mainly derived from European populations.

To further prioritise potentially functional OCR SNPs, I identified those for which there was evidence that they act as eQTL in the human foetal brain ($P < 5E-05$) (O'Brien et al. 2018), and used LDpair to determine the r2 between them and the most significant eQTL for the implicated transcript in the CEU population.

5.3 Results

A limitation of GWAS approaches to complex disorders is that identification of the functional genetic variants underlying associations is often complicated by linkage disequilibrium (resulting in multiple variants at a locus displaying similar levels of association) and incomplete functional characterisation of non-coding regions of the genome.

To refine potentially functional risk SNPs for schizophrenia, I identified those located within my high confidence foetal frontal cortex open chromatin regions that are in strong linkage disequilibrium with the most significant (index) schizophrenia-associated SNP at the 145 genome-wide significant loci reported by Pardiñas and colleagues (Pardiñas et al. 2018).

In keeping with my SLDSR results, genome-wide significant (P < 5E-08) SNPs for schizophrenia were enriched within foetal frontal cortex OCRs (fold enrichment = 2.8; P = 2.35E-08). I identified SNPs in 41 bulk foetal frontal cortex OCRs that are in strong linkage disequilibrium (r2 > 0.8) with the index SNPs at 26 schizophrenia-associated loci (Table 1), further characterising these in terms of whether they are also in foetal brain H3K4me1 and / or H3K4me3 sites, can be confidently attributed to NeuN+ and / or NeuN- nuclei and if they have been found to be a high confidence eQTL for a transcript in foetal brain (O'Brien et al. 2018) (Table 5.1).

Bulk OCR Hg19 co-ordinates	H3K4me1	H3K4me3	NeuN+	NeuN-	SNP in OCR	Schizophrenia index SNP (Pardiñas et al, 2018)	r2 between OCR SNP and SZ index SNP (CEU)	Transcripts regulated by OCR SNP at P <5E-05 in foetal brain	Associated gene symbol	P for OCR SNP as transcript eQTL	Most significant eQTL for transcript	r2 between OCR SNP and top eQTL for transcript (CEU)
chr2:145141366-145141557					rs12991836	rs12991836	1					
chr2:198364425-198365226		Y	Y	Y	rs1116734	rs6434928	1					
chr2:200820058-200820737		Y	Y	Y	rs281766	rs2949006	0.97					
chr2:200838548-200838698					rs11895190	rs2949006	0.97	ENST00000493181	TYW5	6.13E-08	rs281767	0.4671
chr2:225449920-225450695		Y	Y	Y	rs3768886	rs11685299	0.9776					
chr2:233561708-233562451		Y	Y	Y	rs4144797	rs4144797	1					
					rs1104764	rs4144797	0.9381					
chr3:180630079-180630562		Y	Y	Y	rs34455584	rs34796896	0.9689					
chr6:26189217-26189541		Y		Y	rs13204572	rs3130820	0.9421					
chr6:26328298-26328421		Y			rs34107459	rs3130820	0.9421					
					rs13204572	rs3130820	0.9421					
					rs34107459	rs3130820	0.9421					
chr6:26568945-26569150		Y	Y	Y	rs36162392	rs3130820	0.8373					
chr6:27446514-27446777		Y			rs7749305	rs3130820	0.887					
chr6:27759104-27759280		Y			rs35037868	rs3130820	0.9421					
chr6:27805119-27805358		Y		Y	rs34706883	rs3130820	0.9421					
chr6:27759104-27759280		Y			rs35037868	rs3130820	0.9421	ENST00000316606	ZSCAN26	1.22E-07	rs1778484	0.2882
								ENST00000529104	ZSCAN12P1	1.64E-08	rs13217620	0.9392
								ENST00000618305	H4C13	9.46E-06	rs200483	0.9421
								ENST00000316606	ZSCAN26	1.22E-07	rs1778484	0.2882
								ENST00000529104	ZSCAN12P1	1.64E-08	rs13217620	0.9392
								ENST00000618305	H4C13	9.46E-06	rs200483	0.9421
chr6:27805119-27805358		Y		Y	rs34706883	rs3130820	0.9421					
chr6:28058841-28058968	Y	Y			rs35902873	rs3130820	0.9421	ENST00000316606	ZSCAN26	1.22E-07	rs1778484	0.2882
								ENST00000529104	ZSCAN12P1	1.64E-08	rs13217620	0.9392
								ENST00000618305	H4C13	9.46E-06	rs200483	0.9421
chr6:28192980-28193268		Y	Y		rs34243448	rs3130820	0.9421	ENST00000316606	ZSCAN26	8.73E-08	rs1778484	0.2882
								ENST00000529104	ZSCAN12P1	1.56E-07	rs13217620	0.9392
								ENST00000618305	H4C13	3.65E-05	rs200483	0.9421
chr6:28833714-28834785		Y		Y	rs3118358	rs3130820	1					
chr6:28855698-28855896		Y			rs3135309	rs3130820	1	ENST00000375688	VWA7	3.06E-05	rs2736426	0.1018
								ENST00000424176	BAG6	4.18E-05	rs3130490	0.8216
								ENST00000428956	C4A	2.23E-09	rs431204	rs431204 not in 1000G reference panel
								ENST00000464392	HLA-DMA	5.42E-06	rs73409691	0.3083
								ENST00000475627	HLA-DMA	4.64E-05	rs73396800	0.3083
								ENST00000477541	HLA-DMA	1.78E-06	rs59576053	0.3083
								ENST00000498020	HLA-DMB	5.18E-06	rs73396800	0.3083
chr6:32577666-32577857					rs1339227	rs1339227	1					

Table 5.1. Taken from Kouakou et al., 2020, under review. SNPs in high confidence bulk foetal frontal cortex open chromatin regions that are in strong linkage disequilibrium (r2 > 0.8) with schizophrenia GWAS index SNPs (Pardiñas et al, 2018) and their associated effects as eQTL in foetal brain (O'Brien et al, 2018)

Table 5.1 (continued)

Region			SNP1	SNP2	r²	ENST	Gene	p-value	eQTL SNP	value
chr7:105283888-105284171		Y	rs6952386	rs7789569	0.8053					
chr7:86758416-86758600			rs73203000	rs147922658	1					
chr7:86781363-86782353	Y		rs113458968	rs147922658	1					
chr8:10030709-10031396		Y	rs11787370	rs11993663	0.9763					
chr8:38033922-38034608	Y		rs2270376	rs10156310	0.852	ENST00000531344	DDHD2	3.25E-05	rs10958683	0.852
			rs55962869	rs10156310	0.852					
chr10:104613893-104614454	Y	Y	rs3824754	rs12416331	0.9181					
chr11:24408226-24408357			rs72875878	rs1899543	1					
chr11:57434766-57435779	Y	Y	rs11602109	rs7129727	1					
chr12:123717362-123718041	Y	Y	rs368847118	rs2851447	0.9503					
chr12:123738533-123738980	Y	Y	rs1568427	rs2851447	0.9018	ENST00000315580	ARL6IP4	2.21E-05	rs6195196	0.6577
						ENST00000425637	C12orf65	3.68E-06	rs4275659	0.6577
						ENST00000542678	ABCB9	1.25E-13	rs1790094	0.6577
						ENST00000543935	ABCB9	5.78E-07	rs1727295	0.6577
						ENST00000622723	ABCB9	1.07E-07	rs1790094	0.6577
chr12:92256610-92256842	Y	Y	rs10777340	rs4240748	1					
			rs10777341	rs4240748	1					
chr13:79973811-79974270			rs9545101	rs9545047	1	ENST00000450187	LINC01068	8.25E-13	rs2760102	0.8829
						ENST00000457858	LINC01038	3.86E-08	rs9545047	1
						ENST00000620874	LINC01068	2.72E-13	rs1212082	0.8829
chr14:104313796-104314383	Y	Y	rs10083370	rs10083370	1	ENST00000553286	KLC1	1.83E-06	rs1799796	0.4669
			rs10083367	rs10083370	1					
chr15:85201384-85201509	Y	Y	rs12904605	rs12908161	0.9753	ENST00000459938	RN7SL417P	1.04E-09	rs2019611	0.6413
						ENST00000512109	GOLGA6L4	7.96E-06	rs2019611	0.6413
						ENST00000558195	UBE2QL	9.73E-07	rs62028142	0.6277
						ENST00000558801	DNM1P51	2.01E-07	rs55639064	0.086
						ENST00000560634	CSPG4P11	3.56E-06	rs11633897	rs11633897 not in 1000G reference panel
						ENST00000561247	AC243562.2	1.07E-19	rs4338765	0.6496
chr15:85201522-85201703	Y	Y	rs1596061174	rs12908161	1					
chr16:29984768-29985320	Y	Y	rs4318227	rs11646127	1					
chr16:71323066-71323543	Y	Y	rs3096379	rs2161711	0.8287					
chr17:2206641-2207409	Y	Y	rs408067	rs7216638	0.8167					
chr19:11849671-11850047	Y	Y	rs72986630	rs72986630	1					
chr19:19516258-19516665	Y	Y	rs2916068	rs2905432	0.9177					
chr19:50183923-50180576	Y	Y	rs975484	rs7508148	0.9405	ENST00000594587	CPT1C	4.02E-05	rs892149	0.1564

I highlight the example of SNP rs9545101, located in a foetal frontal cortex OCR within the first intron of the *RBM26* gene, which is in strong linkage disequilibrium with both the schizophrenia index SNP at this locus (rs9545047; r2 = 1) and the most significant eQTLs for the long noncoding RNAs *LINC01068* (r2 = 0.88; *P* eQTL of OCR SNP = 2.7E-13) and *LINC01038* (r2 = 1; *P* eQTL of OCR SNP= 3.9E-08) in the CEU population (Figure 5.1).
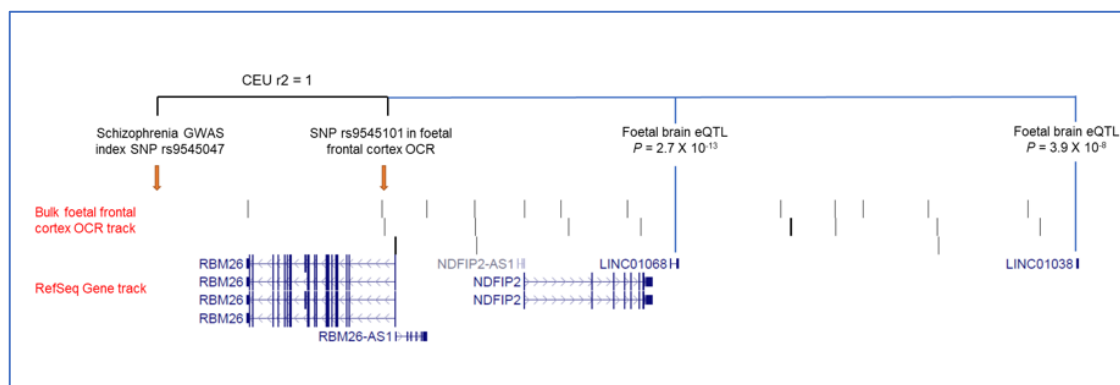

Figure 5.1. (Taken from Kouakou et al. 2020 – under review) Created by Nicholas J. Bray. **A single nucleotide polymorphism (SNP) within a foetal frontal cortex open chromatin region that is in perfect linkage disequilibrium (r2 = 1) with the schizophrenia GWAS index SNP at the locus**(Pardiñas et al. 2018) **and an eQTL for *LINC01068* and *LINC01038* in the human foetal brain** (O'Brien et al. 2018)**.** Image generated using the UCSC Genome Browser (https://genome.ucsc.edu/index.html) and an uploaded BED file for identified bulk foetal frontal cortex open chromatin as a custom track.

5.4. Discussion

Although GWASs have successfully identified a large number of genetic loci that impact risk for complex disorders, these successes are tempered by two major limitations that make GWAS data interpretation difficult. Firstly, functional characterisation of non-coding regions of the genome in different cell types and developmental stages is incomplete. Secondly, identification of the functional genetic variants underlying associations is often complicated by linkage disequilibrium, a phenomenon by which SNPs in close spatial proximity are inherited non-randomly, resulting in multiple variants at a locus displaying similar levels of association.

Therefore, as most risk loci reside in poorly characterised non-coding regions, it is a challenge for researchers to ascertain exactly how risk is conferred by these loci. Furthermore, with the spread, and volume, of associated loci across the genome, the cost of functionally examining every genetic variant individually would be substantial. For these reasons, prioritisation of the genetic risk loci that are most likely to confer risk is essential.

In this chapter, I used the foetal brain open chromatin regions to identify potentially functional SNPs at genome-wide significant risk loci for schizophrenia. I identified a total of 41 SNPs within open chromatin regions and provide additional information that will be useful to guide further functional studies.

An open chromatin SNP that is in strong linkage disequilibrium with both the schizophrenia index SNP and the most significant eQTLs for a transcript is more likely to be the causal SNP (SNP rs9545101, for instance). In contrast, when the OCR SNP is a significant eQTL for a transcript but is not in strong LD with the top eQTL for that transcript (for instance rs11895190

and TYW5), even if the open chromatin SNP is the genuine schizophrenia risk SNP, it probably doesn't operate through expression of that gene in foetal brain. I highlighted the example of SNP rs9545101 on chromosome 13, which is in high LD with both the schizophrenia index SNP at the locus and the top eQTL for the long intergenic non-coding RNAs *LINC01068 / LINC01038.* Although these two genes have yet to be functionally characterised, long non-coding RNAs are known to participate in various neurodevelopmental processes (L. Li et al. 2019). I also note foetal OCR SNPs at 2 other loci that are in strong LD with both the schizophrenia index SNP and top eQTL for at least one transcript. Several OCR SNPs were in strong LD with the index SNP rs3130820 at the major histocompatibility complex (MHC) locus on chromosome 6, and these were in high LD with the top eQTL SNPs for multiple transcripts (*ZSCAN12P1*, *H4C13* and *BAG6*), consistent with extensive LD in this region. In addition, SNP rs2270376 on chromosome 8 is in strong LD with the schizophrenia index SNP rs10156310 and also with the most significant eQTL for a transcript of *DDHD2*. *DDHD2 Domain Containing 2* encodes a phospholipase enzyme associated with hereditary spastic paraplegia and intellectual disability (Inloes et al. 2014). Interestingly, reduced expression of DDHD2 in the foetal brain was found to be associated with ADHD, autism spectrum disorder, bipolar disorder, major depressive disorder, and schizophrenia in a recent transcriptome-wide association study (TWAS) (Hall et al. 2020).

Chapter 6. General Discussion

The purpose of annotating the genome for integration with GWAS data is to identify and prioritise the tissue, cell-types, regulatory processes and risk variants that drive the pathophysiology of complex disorders. In the work described in this thesis, I provide annotations of the genome relevant to a key tissue and time point implicated in several psychiatric disorders.

The general aim of this thesis was to investigate whether gene regulatory processes in the human prenatal brain contribute to genetic risk for psychiatric disorders. In order to achieve this aim, firstly, I have mapped open chromatin regions in the second trimester bulk foetal frontal cortex nuclei. Then, using stratified linkage disequilibrium regression analysis, I tested whether these open chromatin sites were enriched for SNP heritability associated with 5 psychiatric disorders. Results showed enrichment of SNP heritability for schizophrenia.

In order to test whether the enrichment signal was driven by a specific type of regulatory element, I restricted my bulk ATAC-Seq peaks to those overlapping either foetal brain H3K4me1 or H3K4me3 sites, indicative of active or poised promoters and enhancers respectively. This showed significant enrichment of SNP heritability for all 5 tested neuropsychiatric disorders in both histone modification annotations. The heritability enrichments were stronger than when open chromatin sites were considered alone.

Furthermore, given that the regulatory processes are known to operate in a cell-type specific manner, I explored the regulatory landscape of neuronal and non-neuronal nuclei populations, by extracting NeuN+ and NeuN- nuclei from the same prenatal human frontal

cortex samples used for the bulk analyses, using FANS. Integrating open chromatin data from FANS extracted nuclei with foetal brain H3K4me1 or H3K4me3 sites and GWAS summary statistics showed that genetic risk for the tested neuropsychiatric conditions is at least as strongly enriched within open chromatin regions observed in NeuN- nuclei as in NeuN+ nuclei.

Recently, GWASes have grown significantly both in sample size and in the number of investigated traits. Despite the success of GWASes in identifying common genetic variants associated with neuropsychiatric disorders, the biological insights derived from these results have been limited. This is due to the difficulty of interpreting GWAS associations. Neighbouring genetic variants are often correlated with one another within a population, a phenomenon known as linkage disequilibrium (LD). LD results in multiple variants at a locus being present in the same individuals, and therefore traits are often associated with multiple variants at a locus at similar levels of statistical significance. In addition, most trait-associated SNP GWAS are in non-coding regions of the genome (F. Zhang and Lupski 2015), for which functionally annotation is limited. These two features make it difficult to distinguish the causal (functional) variants underpinning genetic associations at GWAS risk loci.

Use of functional annotations of the non-coding genome, such as open chromatin regions, can help prioritise potential causal variants at GWAS risk loci, while use of eQTL maps can further indicate the functionally altered (e.g., mis-regulated) gene(s) (Broekema, Bakker, and Jonkers 2020). In chapter 5, I produced a list of potentially functional non-coding SNPs tagged by genome-wide significant risk SNPs for schizophrenia, which could guide future functional genomics studies (e.g., CRISPR-Cas9 genome editing). I identified SNPs within foetal brain open chromatin regions that were in high linkage disequilibrium ($r2 > 0.8$) with the most significant (index) SNP at each genome-wide significant schizophrenia risk locus and

integrated these data with foetal brain eQTL data from a previous study (O'Brien et al, 2018). In this way, I identified open chromatin SNPs that are in strong LD with both the index SNP for schizophrenia and the most significant eQTLs for transcripts in the foetal brain.

Producing novel functional genomic data from primary human brain tissue, and integrating it with robust genetic associations, is essential to enhance the understanding of how non-coding genetic variation contributes to the onset and progression of psychiatric disorders. Of particular importance is determining the cellular contexts in which risk variants are active. I have generated the first open chromatin maps of distinct populations of cells in human foetal brain. This has allowed me to identify specific patterns of gene regulation in neuronal and non-neuronal cells. However, my analysis was restricted to the study of two broad populations of cells – neurons and non-neurons. Both populations are themselves very heterogeneous, consisting of a variety of neural and non-neuronal sub-types. Furthermore, because the isolation of non-neuronal nuclei relies on negative selection, this population is more heterogeneous and includes several diverse cell types such as cycling neural progenitors, radial glia and oligodendrocyte precursors. The difference in NeuN- population between foetal and adult brain, which largely consist of mature glia (astrocytes and oligodendrocytes), may explain the contrasting results of Fullard and colleagues (Fullard et al. 2018) who found that genetic risk for schizophrenia was mainly mediated by neuronal (NeuN+) cells.

In this thesis, I have performed a comprehensive assessment of chromatin accessibility during human cortical neurogenesis. The analysis also has important implications for

understanding the pathophysiology of neuropsychiatric disorder, guiding future genetic discovery and disease modelling. However, while epigenomic methods such as ATAC-Seq can be exploited to better understand the cellular basis of complex traits, it does not in itself indicate which genes are functionally impacted by risk variation and how they are altered.

Enhancers are often located hundreds of kilobases away from their target gene, so physical interaction with the promoter sequence occurs through looping of the DNA strand. Attributing enhancers to their target gene(s) is therefore a challenging task. Chromosome conformation capture methods, such as 3C, 4C and Hi-C (Lieberman-Aiden et al. 2009) are used for this purpose. They capture intra- and inter-chromosomal physical contacts, by fixing looped DNA with formaldehyde and then amplifying and sequencing the interacting DNA fragments. Using these methods, genome-wide interaction maps can be constructed. The original plan for this thesis was to perform a 3C experiment on one of the foetal brain samples to visualise the interaction between a foetal brain active enhancer and its target gene. Unfortunately, due to the current COVID-19 pandemic, my work in the lab couldn't be carried out and I opted for a computational chapter using existing foetal brain eQTL data to indicate functionally impacted genes instead (chapter 5). Future work should consider producing Hi-C maps from the same foetal brain samples and neuronal and non-neuronal nuclei in particular. Using next generation sequencing technology, the Hi-C technique, an extension of 3C, enables the identification of the chromosome conformation at a genome wide scale (Schmitt, Hu, and Ren 2016; Ay and Noble 2015). Compared to other variant of the 3C technique, the Hi-C technique is the first method to capture chromosome conformation on a "all versus all" basis —that is, it can profile interactions for all read pairs in an entire genome.

Hi-C maps of the foetal brain do exist (Won et al. 2016), but not at cell-type resolution. This would allow to map the putative functional variants to their target genes in different cell population.

It has to be noted that the results from heritability enrichment analyses performed in this study are influenced by statistical power: both GWAS sample sizes and number of SNPs laying within the open chromatin annotation have an effect on the partitioned heritability analyses. In fact, while the level of SNP heritability enrichment for all 5 tested psychiatric disorders is similar (ranging between 10.4 and 19.9 in unsorted nuclei), the significance level of the enrichment of larger GWASs (such as schizophrenia) consistently showed much lower p-values. It is therefore likely that using more powerful GWASs would allow me to more accurately gauge the enrichment for these disorders. Moreover, the proportion of accessible chromatin in sorted nuclei was considerably lower compared to the open chromatin found in homogenate tissue. Consequently, the low number of genome-wide significant SNPs lying within cell-type-specific open chromatin regions (between 0.3% and 0.6% for NeuN+ and 0.6% and 0.9% for NeuN-) provided less statistical power for the enrichment analysis using LD score regression.

My data provide evidence for an early neurodevelopmental component to a range of neuropsychiatric conditions and highlight an important role for regulatory genomic regions active within non-neuronal cells of the prenatal brain. Although I have focused on neuropsychiatric phenotypes, the maps of open chromatin in the foetal frontal cortex produced by this project are likely to be useful in exploring early neurodevelopmental antecedents to a variety of brain-related conditions. As we move into the era of whole

genome sequencing, maps of functional elements within the non-coding genome will continue to be important.

Epigenomic characterisation of the human brain is particularly challenging due to its complex histological structure and diversity of cell types. Future studies should take advantage of advances in single cell sequencing, which now make it possible to discriminate between individual cell types of the brain, including heterogeneous cellular subtypes that may have distinct transcriptomic and epigenomic profiles in discrete brain regions. Single-cell genomic assays allow profiling gene expression, chromatin accessibility, and TF occupancy with single-cell resolution (Ma et al. 2020). These assays can resolve the cellular composition of complex organs and tissues and are used to assemble cells into reference tissue atlases. The high resolution of single-cell genomic maps makes them a promising resource for SNP heritability enrichment analysis. The increasing availability of functional data for more cell types and states is expected to improve the accuracy of these enrichment signals. This will allow us to more confidently nominate the specific cell types and developmental states causally involved in disease.

References

Allen, Hana Lango, Karol Estrada, Guillaume Lettre, Sonja I. Berndt, Michael N. Weedon, Fernando Rivadeneira, Cristen J. Willer, et al. 2010. "Hundreds of Variants Clustered in Genomic Loci and Biological Pathways Affect Human Height." *Nature* 467 (7317). https://doi.org/10.1038/nature09410.

Allis, C David, and Thomas Jenuwein. 2016. "The Molecular Hallmarks of Epigenetic Control." *Nature Reviews Genetics* 17 (8): 487–500. https://doi.org/10.1038/nrg.2016.59.

American Psychiatric Association. 2013. *American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders Fifth Edition*. Arlington.

Ay, Ferhat, and William S Noble. 2015. "Analysis Methods for Studying the 3D Architecture of the Genome." *Genome Biology* 16 (1): 183. https://doi.org/10.1186/s13059-015-0745-7.

Ball, Madeleine P, Jin Billy Li, Yuan Gao, Je-Hyuk Lee, Emily M LeProust, In-Hyun Park, Bin Xie, George Q Daley, and George M Church. 2009. "Targeted and Genome-Scale Strategies Reveal Gene-Body Methylation Signatures in Human Cells." *Nature Biotechnology* 27 (4): 361–68. https://doi.org/10.1038/nbt.1533.

Bernstein, Bradley E, Emily L Humphrey, Rachel L Erlich, Robert Schneider, Peter Bouman, Jun S Liu, Tony Kouzarides, and Stuart L Schreiber. 2002. "Methylation of Histone H3 Lys 4 in Coding Regions of Active Genes." *Proceedings of the National Academy of Sciences* 99 (13): 8695 LP – 8700. https://doi.org/10.1073/pnas.082249499.

Bhugra, Dinesh. 2005. "The Global Prevalence of Schizophrenia." *PLoS Medicine* 2 (5): e151–75. https://doi.org/10.1371/journal.pmed.0020151.

Bortolato, Beatrice, Cristiano A Köhler, Evangelos Evangelou, Jordi León-Caballero, Marco Solmi, Brendon Stubbs, Lazaros Belbasis, et al. 2017. "Systematic Assessment of Environmental Risk Factors for Bipolar Disorder: An Umbrella Review of Systematic Reviews and Meta-Analyses." *Bipolar Disorders* 19 (2): 84–96. https://doi.org/https://doi.org/10.1111/bdi.12490.

Bray, Nicholas J, and Michael C O'Donovan. 2018. "The Genetics of Neuropsychiatric Disorders." *Brain and Neuroscience Advances* 2 (January): 2398212818799271. https://doi.org/10.1177/2398212818799271.

Brennand, Kristen J, Anthony Simone, Jessica Jou, Chelsea Gelboin-Burkhart, Ngoc Tran, Sarah Sangar, Yan Li, et al. 2011. "Modelling Schizophrenia Using Human Induced Pluripotent Stem Cells." *Nature* 473 (7346): 221–25. https://doi.org/10.1038/nature09915.

Broekema, R V, O B Bakker, and I H Jonkers. 2020. "A Practical View of Fine-Mapping and Gene Prioritization in the Post-Genome-Wide Association Era." *Open Biology* 10 (1): 190221. https://doi.org/10.1098/rsob.190221.

Brown, Alan S. 2012. "Epidemiologic Studies of Exposure to Prenatal Infection and Risk of Schizophrenia and Autism." *Developmental Neurobiology* 72 (10): 1272–76. https://doi.org/10.1002/dneu.22024.

Bryois, Julien, Melanie E Garrett, Lingyun Song, Alexias Safi, Paola Giusti-Rodriguez, Graham D Johnson, Alfonso Buil Demur, et al. 2017. "Evaluation Of Chromatin Accessibility In Prefrontal Cortex Of Schizophrenia Cases And Controls." *BioRxiv*, January. http://biorxiv.org/content/early/2017/05/25/141986.abstract.

Bryois, Julien, Nathan G Skene, Thomas Folkmann Hansen, Lisette Kogelman, Hunna J Watson, Leo Brueggeman, Gerome Breen, et al. 2019. "Genetic Identification of Cell

Types Underlying Brain Complex Traits Yields Novel Insights Into the Etiology of Parkinson's Disease." *BioRxiv*, January, 528463. https://doi.org/10.1101/528463.

Buenrostro, J D, P G Giresi, L C Zaba, H Y Chang, and W J Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nat Methods* 10. https://doi.org/10.1038/nmeth.2688.

Buenrostro, Jason D, Beijing Wu, Ulrike M Litzenburger, Dave Ruff, Michael L Gonzales, Michael P Snyder, Howard Y Chang, and William J Greenleaf. 2015. "Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation." *Nature* 523 (7561): 486–90. https://doi.org/10.1038/nature14590.

Bulik-Sullivan, Brendan, Hilary K Finucane, Verneri Anttila, Alexander Gusev, Felix R Day, Po-Ru Loh, Laramie Duncan, et al. 2015. "An Atlas of Genetic Correlations across Human Diseases and Traits." *Nature Genetics* 47 (11): 1236–41. https://doi.org/10.1038/ng.3406.

Bulik-Sullivan, Brendan K, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, Benjamin M Neale, and Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47 (3): 291–95. https://doi.org/10.1038/ng.3211.

Charney, Alexander W, Eli A Stahl, Elaine K Green, Chia-Yen Chen, Jennifer L Moran, Kimberly Chambert, Richard A Belliveau, et al. 2019. "Contribution of Rare Copy Number Variants to Bipolar Disorder Risk Is Limited to Schizoaffective Cases." *Biological Psychiatry* 86 (2): 110–19. https://doi.org/https://doi.org/10.1016/j.biopsych.2018.12.009.

Chatterjee, Nilanjan, Jianxin Shi, and Montserrat García-Closas. 2016. "Developing and Evaluating Polygenic Risk Prediction Models for Stratified Disease Prevention." *Nature Reviews Genetics* 17 (7): 392–406. https://doi.org/10.1038/nrg.2016.27.

Clark, L A, D Watson, and S Reynolds. 1995. "Diagnosis and Classification of Psychopathology: Challenges to the Current System and Future Directions." *Annual Review of Psychology* 46 (1): 121–53. https://doi.org/10.1146/annurev.ps.46.020195.001005.

Clements, Caitlin C, Tara L Wenger, Alisa R Zoltowski, Jennifer R Bertollo, Judith S Miller, Ashley B de Marchena, Lauren M Mitteer, et al. 2017. "Critical Region within 22q11.2 Linked to Higher Rate of Autism Spectrum Disorder." *Molecular Autism* 8 (1): 58. https://doi.org/10.1186/s13229-017-0171-7.

Conrad, Donald F, Jonathan E M Keebler, Mark A DePristo, Sarah J Lindsay, Yujun Zhang, Ferran Casals, Youssef Idaghdour, et al. 2011. "Variation in Genome-Wide Mutation Rates within and between Human Families." *Nature Genetics* 43 (7): 712–14. https://doi.org/10.1038/ng.862.

Consortium{, Roadmap Epigenomics, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "ARTICLE Integrative Analysis of 111 Reference Human Epigenomes." https://doi.org/10.1038/nature14248.

Consortium, Brainstorm, Verneri Anttila, Brendan Bulik-Sullivan, Hilary K Finucane, Raymond K Walters, Jose Bras, Laramie Duncan, et al. 2018. "Analysis of Shared Heritability in Common Disorders of the Brain." *Science (New York, N.Y.)* 360 (6395): eaap8757. https://doi.org/10.1126/science.aap8757.

Consortium, International Schizophrenia. 2009. "Common Polygenic Variation Contributes

to Risk of Schizophrenia That Overlaps with Bipolar Disorder." *Nature* 460 (7256): 748–52. https://doi.org/10.1038/nature08185.

Consortium, Schizophrenia Working Group of the Psychiatric Genomics. 2014. "Biological Insights from 108 Schizophrenia-Associated Genetic Loci." *Nature* 511 (7510): 421–27. http://dx.doi.org/10.1038/nature13595.

Cooper, Sara J, Nathan D Trinklein, Elizabeth D Anton, Loan Nguyen, and Richard M Myers. 2006. "Comprehensive Analysis of Transcriptional Promoter Structure and Function in 1% of the Human Genome." *Genome Research* 16 (1): 1–10. https://doi.org/10.1101/gr.4222606.

Cossarizza, Andrea, Hyun-Dong Chang, Andreas Radbruch, Andreas Acs, Dieter Adam, Sabine Adam-Klages, William W Agace, et al. 2019. "Guidelines for the Use of Flow Cytometry and Cell Sorting in Immunological Studies (Second Edition)." *European Journal of Immunology* 49 (10): 1457–1973. https://doi.org/10.1002/eji.201970107.

Creese, I, D R Burt, and S H Snyder. 1976. "Dopamine Receptor Binding Predicts Clinical and Pharmacological Potencies of Antischizophrenic Drugs." *Science* 192 (4238): 481 LP – 483. https://doi.org/10.1126/science.3854.

Cusanovich, Darren A, Riza Daza, Andrew Adey, Hannah A Pliner, Lena Christiansen, Kevin L Gunderson, Frank J Steemers, Cole Trapnell, and Jay Shendure. 2015. "Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing." *Science (New York, N.Y.)* 348 (6237): 910–14. https://doi.org/10.1126/science.aab1601.

Danielson, Melissa L, Rebecca H Bitsko, Reem M Ghandour, Joseph R Holbrook, Michael D Kogan, and Stephen J Blumberg. 2018. "Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016." *Journal of Clinical Child and Adolescent Psychology : The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53* 47 (2): 199–212. https://doi.org/10.1080/15374416.2017.1417860.

Darmanis, Spyros, Steven A Sloan, Ye Zhang, Martin Enge, Christine Caneda, Lawrence M Shuer, Melanie G Hayden Gephart, Ben A Barres, and Stephen R Quake. 2015. "A Survey of Human Brain Transcriptome Diversity at the Single Cell Level." *Proceedings of the National Academy of Sciences* 112 (23): 7285 LP – 7290. https://doi.org/10.1073/pnas.1507125112.

Demontis, Ditte, Raymond K. Walters, Joanna Martin, Manuel Mattheisen, Thomas D. Als, Esben Agerbo, Gísli Baldursson, et al. 2019. "Discovery of the First Genome-Wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder." *Nature Genetics* 51 (1). https://doi.org/10.1038/s41588-018-0269-7.

Demontis, Ditte, Raymond K Walters, Joanna Martin, Manuel Mattheisen, Thomas D Als, Esben Agerbo, Gísli Baldursson, et al. 2019. "Discovery of the First Genome-Wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder." *Nature Genetics* 51 (1): 63–75. https://doi.org/10.1038/s41588-018-0269-7.

Dillman, Allissa A, and Mark R Cookson. 2014. "Transcriptomic Changes in Brain Development." *International Review of Neurobiology* 116: 233–50. https://doi.org/10.1016/B978-0-12-801105-8.00009-6.

Fan, Xiaoying, Yuanyuan Fu, Xin Zhou, Le Sun, Ming Yang, Mengdi Wang, Ruiguo Chen, et al. 2020. "Single-Cell Transcriptome Analysis Reveals Cell Lineage Specification in Temporal-Spatial Patterns in Human Cortical Development." *Science Advances* 6 (34): eaaz2978. https://doi.org/10.1126/sciadv.aaz2978.

Feingold, E. A., P. J. Good, M. S. Guyer, S. Kamholz, L. Liefer, K. Wetterstrand, F. S. Collins, et

al. 2004. "The ENCODE (ENCyclopedia of DNA Elements) Project." *Science*. https://doi.org/10.1126/science.1105136.

Finucane, Hilary K, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, et al. 2015. "Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics." *Nat Genet* 47 (11): 1228–35. http://dx.doi.org/10.1038/ng.3404.

Forsyth, Jennifer K., Daniel Nachun, Michael J. Gandal, Daniel H. Geschwind, Ariana E. Anderson, Giovanni Coppola, and Carrie E. Bearden. 2020. "Synaptic and Gene Regulatory Mechanisms in Schizophrenia, Autism, and 22q11.2 Copy Number Variant–Mediated Risk for Neuropsychiatric Disorders." *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2019.06.029.

Froehlich, Tanya E, Julia S Anixt, Irene M Loe, Vilawan Chirdkiatgumchai, Lisa Kuan, and Richard C Gilman. 2011. "Update on Environmental Risk Factors for Attention-Deficit/Hyperactivity Disorder." *Current Psychiatry Reports* 13 (5): 333–44. https://doi.org/10.1007/s11920-011-0221-3.

Fromer, Menachem, Andrew J Pocklington, David H Kavanagh, Hywel J Williams, Sarah Dwyer, Padhraig Gormley, Lyudmila Georgieva, et al. 2014. "De Novo Mutations in Schizophrenia Implicate Synaptic Networks." *Nature* 506 (7487): 179–84. https://doi.org/10.1038/nature12929.

Fullard, John F, Mads E Hauberg, Jaroslav Bendl, Gabor Egervari, MARIA DANIELA CIRNARU, Sarah M Reach, Jan Motl, Michelle E Ehrlich, Yasmin L Hurd, and Panos Roussos. 2018. "An Atlas of Chromatin Accessibility in the Adult Human Brain." *Genome Research*, June, gr.232488.117. https://doi.org/10.1101/gr.232488.117.

Gazal, Steven, Hilary K Finucane, Nicholas A Furlotte, Po-Ru Loh, Pier Francesco Palamara, Xuanyao Liu, Armin Schoech, et al. 2017. "Linkage Disequilibrium-Dependent Architecture of Human Complex Traits Shows Action of Negative Selection." *Nature Genetics* 49 (10): 1421–27. https://doi.org/10.1038/ng.3954.

Gazal, Steven, Carla Marquez-Luna, Hilary K Finucane, and Alkes L Price. 2019. "Reconciling S-LDSC and LDAK Functional Enrichment Estimates." *Nature Genetics* 51 (8): 1202–4. https://doi.org/10.1038/s41588-019-0464-1.

Giresi, P G, J Kim, R M McDaniell, V R Iyer, and J D Lieb. 2007. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin." *Genome Res* 17. https://doi.org/10.1101/gr.5533506.

Gogtay, Nitin, Nora S Vyas, Renee Testa, Stephen J Wood, and Christos Pantelis. 2011. "Age of Onset of Schizophrenia: Perspectives from Structural Neuroimaging Studies." *Schizophrenia Bulletin* 37 (3): 504–13. https://doi.org/10.1093/schbul/sbr030.

Gorkin, David U, Dongwon Lee, Xylena Reed, Christopher Fletez-Brant, Seneca L Bessling, Stacie K Loftus, Michael A Beer, William J Pavan, and Andrew S McCallion. 2012. "Integration of ChIP-Seq and Machine Learning Reveals Enhancers and a Predictive Regulatory Sequence Vocabulary in Melanocytes." *Genome Research* 22 (11): 2290–2301. https://doi.org/10.1101/gr.139360.112.

Grove, Jakob, Stephan Ripke, Thomas D Als, Manuel Mattheisen, Raymond K Walters, Hyejung Won, Jonatan Pallesen, et al. 2019. "Identification of Common Genetic Risk Variants for Autism Spectrum Disorder." *Nature Genetics* 51 (3): 431–44. https://doi.org/10.1038/s41588-019-0344-8.

Guenther, Matthew G, Stuart S Levine, Laurie A Boyer, Rudolf Jaenisch, and Richard A Young. 2007. "A Chromatin Landmark and Transcription Initiation at Most Promoters in

Human Cells." *Cell* 130 (1): 77–88. https://doi.org/10.1016/j.cell.2007.05.042.

Gulsuner, Suleyman, Tom Walsh, Amanda C. Watts, Ming K. Lee, Anne M. Thornton, Silvia Casadei, Caitlin Rippey, et al. 2013. "XSpatial and Temporal Mapping of de Novo Mutations in Schizophrenia to a Fetal Prefrontal Cortical Network." *Cell*. https://doi.org/10.1016/j.cell.2013.06.049.

Gusel'nikova, V V, and D E Korzhevskiy. 2015. "NeuN As a Neuronal Nuclear Antigen and Neuron Differentiation Marker." *Acta Naturae* 7 (2): 42–47. https://pubmed.ncbi.nlm.nih.gov/26085943.

Hall, Lynsey S, Oliver Pain, Heath E O'Brien, Richard Anney, James T R Walters, Michael J Owen, Michael C O'Donovan, and Nicholas J Bray. 2020. "Cis-Effects on Gene Expression in the Human Prenatal Brain Associated with Genetic Risk for Neuropsychiatric Disorders." *Molecular Psychiatry*. https://doi.org/10.1038/s41380-020-0743-3.

Hannon, Eilis, Helen Spiers, Joana Viana, Ruth Pidsley, Joe Burrage, Therese M Murphy, Claire Troakes, et al. 2016. "Methylation QTLs in the Developing Brain and Their Enrichment in Schizophrenia Risk Loci." *Nature Neuroscience* 19 (1): 48–54. https://doi.org/10.1038/nn.4182.

Harrison, Paul J. 1999. "The Neuropathology of Schizophrenia: A Critical Review of the Data and Their Interpretation." *Brain* 122 (4): 593–624. https://doi.org/10.1093/brain/122.4.593.

Heintzman, Nathaniel D, Rhona K Stuart, Gary Hon, Yutao Fu, Christina W Ching, R David Hawkins, Leah O Barrera, et al. 2007. "Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome." *Nature Genetics* 39 (3): 311–18. https://doi.org/10.1038/ng1966.

Hesselberth, J R, X Chen, Z Zhang, P J Sabo, R Sandstrom, A P Reynolds, R E Thurman, et al. 2009. "Global Mapping of Protein-DNA Interactions in Vivo by Digital Genomic Footprinting." *Nat Methods* 6. https://doi.org/10.1038/nmeth.1313.

Hoffman, Gabriel E., Jaroslav Bendl, Georgios Voloudakis, Kelsey S. Montgomery, Laura Sloofman, Ying Chih Wang, Hardik R. Shah, et al. 2019. "CommonMind Consortium Provides Transcriptomic and Epigenomic Data for Schizophrenia and Bipolar Disorder." *Scientific Data* 6 (1). https://doi.org/10.1038/s41597-019-0183-6.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2009. "Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources." *Nature Protocols* 4 (1). https://doi.org/10.1038/nprot.2008.211.

Huo, Yongxia, Shiwu Li, Jiewei Liu, Xiaoyan Li, and Xiong-Jian Luo. 2019. "Functional Genomics Reveal Gene Regulatory Mechanisms Underlying Schizophrenia Risk." *Nature Communications* 10 (1): 670. https://doi.org/10.1038/s41467-019-08666-4.

Inloes, Jordon M, Ku-Lung Hsu, Melissa M Dix, Andreu Viader, Kim Masuda, Thais Takei, Malcolm R Wood, and Benjamin F Cravatt. 2014. "The Hereditary Spastic Paraplegia-Related Enzyme DDHD2 Is a Principal Brain Triglyceride Lipase." *Proceedings of the National Academy of Sciences* 111 (41): 14924 LP – 14929. https://doi.org/10.1073/pnas.1413706111.

Iotchkova, Valentina, Graham R S Ritchie, Matthias Geihs, Sandro Morganella, Josine L Min, Klaudia Walter, Nicholas John Timpson, et al. 2019. "GARFIELD Classifies Disease-Relevant Genomic Features through Integration of Functional Annotations with Association Signals." *Nature Genetics* 51 (2): 343–53. https://doi.org/10.1038/s41588-018-0322-6.

Karabacak Calviello, Aslıhan, Antje Hirsekorn, Ricardo Wurmus, Dilmurat Yusuf, and Uwe Ohler. 2019. "Reproducible Inference of Transcription Factor Footprints in ATAC-Seq and DNase-Seq Datasets Using Protocol-Specific Bias Modeling." *Genome Biology* 20 (1): 42. https://doi.org/10.1186/s13059-019-1654-y.

Kathuria, Annie, Kara Lopez-Lengowski, Magdalena Vater, Donna McPhie, Bruce M Cohen, and Rakesh Karmacharya. 2020. "Transcriptome Analysis and Functional Characterization of Cerebral Organoids in Bipolar Disorder." *Genome Medicine* 12 (1): 34. https://doi.org/10.1186/s13073-020-00733-6.

Kim, Jong Yeob, Min Ji Son, Chei Yun Son, Joaquim Radua, Michael Eisenhut, Florence Gressier, Ai Koyanagi, et al. 2019. "Environmental Risk Factors and Biomarkers for Autism Spectrum Disorder: An Umbrella Review of the Evidence." *The Lancet. Psychiatry* 6 (7): 590–600. https://doi.org/10.1016/S2215-0366(19)30181-6.

Kimura, Kouichi, Ai Wakamatsu, Yutaka Suzuki, Toshio Ota, Tetsuo Nishikawa, Riu Yamashita, Jun-ichi Yamamoto, et al. 2006. "Diversification of Transcriptional Modulation: Large-Scale Identification and Characterization of Putative Alternative Promoters of Human Genes." *Genome Research* 16 (1): 55–65. https://doi.org/10.1101/gr.4039406.

Kirov, George, Dilihan Gumus, Wei Chen, Nadine Norton, Lyudmila Georgieva, Murat Sari, Michael C O'Donovan, et al. 2008. "Comparative Genome Hybridization Suggests a Role for NRXN1 and APBA2 in Schizophrenia." *Human Molecular Genetics* 17 (3): 458–65. https://doi.org/10.1093/hmg/ddm323.

Kirov, George, Dan Rujescu, Andres Ingason, David A Collier, Michael C O'Donovan, and Michael J Owen. 2009. "Neurexin 1 (NRXN1) Deletions in Schizophrenia." *Schizophrenia Bulletin* 35 (5): 851–54. https://doi.org/10.1093/schbul/sbp079.

Klemm, Sandy L, Zohar Shipony, and William J Greenleaf. 2019. "Chromatin Accessibility and the Regulatory Epigenome." *Nature Reviews Genetics* 20 (4): 207–20. https://doi.org/10.1038/s41576-018-0089-8.

Köhler, Cristiano A, Evangelos Evangelou, Brendon Stubbs, Marco Solmi, Nicola Veronese, Lazaros Belbasis, Beatrice Bortolato, et al. 2018. "Mapping Risk Factors for Depression across the Lifespan: An Umbrella Review of Evidence from Meta-Analyses and Mendelian Randomization Studies." *Journal of Psychiatric Research* 103: 189–207. https://doi.org/https://doi.org/10.1016/j.jpsychires.2018.05.020.

Korzhevskii, D É, E S Petrova, O V Kirik, and V A Otellin. 2009. "Assessment of Neuron Differentiation during Embryogenesis in Rats Using Immunocytochemical Detection of Doublecortin." *Neuroscience and Behavioral Physiology* 39 (6): 513–16. https://doi.org/10.1007/s11055-009-9164-0.

Kouzarides, Tony. 2007. "Chromatin Modifications and Their Function." *Cell*. https://doi.org/10.1016/j.cell.2007.02.005.

Kurdyukov, Sergey, and Martyn Bullock. 2016. "DNA Methylation Analysis: Choosing the Right Method." *Biology* 5 (1): 3. https://doi.org/10.3390/biology5010003.

la Torre-Ubieta, Luis de, Jason L Stein, Hyejung Won, Carli K Opland, Dan Liang, Daning Lu, and Daniel H Geschwind. 2018. "The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis." *Cell* 172 (1): 289-304.e18. https://doi.org/10.1016/j.cell.2017.12.014.

Landry, Josette-Renée, Dixie L Mager, and Brian T Wilhelm. 2003. "Complex Controls: The Role of Alternative Promoters in Mammalian Genomes." *Trends in Genetics* 19 (11): 640–48. https://doi.org/https://doi.org/10.1016/j.tig.2003.09.014.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Lango Allen, Hana, Karol Estrada, Guillaume Lettre, Sonja I Berndt, Michael N Weedon, Fernando Rivadeneira, Cristen J Willer, et al. 2010. "Hundreds of Variants Clustered in Genomic Loci and Biological Pathways Affect Human Height." *Nature* 467 (September): 832. http://dx.doi.org/10.1038/nature09410.

Lee, S Hong, Stephan Ripke, Benjamin M Neale, Stephen V Faraone, Shaun M Purcell, Roy H Perlis, Bryan J Mowry, et al. 2013. "Genetic Relationship between Five Psychiatric Disorders Estimated from Genome-Wide SNPs." *Nature Genetics* 45 (9): 984–94. https://doi.org/10.1038/ng.2711.

Li, H, B Handsaker, A Wysoker, T Fennell, J Ruan, N Homer, G Marth, G Abecasis, and R Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25. https://doi.org/10.1093/bioinformatics/btp352.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics (Oxford, England)* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Li, Ling, Yingliang Zhuang, Xingsen Zhao, and Xuekun Li. 2019. "Long Non-Coding RNA in Neuronal Development and Neurological Disorders   ." *Frontiers in Genetics* . https://www.frontiersin.org/article/10.3389/fgene.2018.00744.

Li, Zhijian, Marcel H Schulz, Thomas Look, Matthias Begemann, Martin Zenke, and Ivan G Costa. 2019. "Identification of Transcription Factor Binding Sites Using ATAC-Seq." *Genome Biology* 20 (1): 45. https://doi.org/10.1186/s13059-019-1642-2.

Lieberman-Aiden, Erez, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science* 326 (5950): 289 LP – 293. http://science.sciencemag.org/content/326/5950/289.abstract.

Lim, Grace Y, Wilson W Tam, Yanxia Lu, Cyrus S Ho, Melvyn W Zhang, and Roger C Ho. 2018. "Prevalence of Depression in the Community from 30 Countries between 1994 and 2014." *Scientific Reports* 8 (1): 2861. https://doi.org/10.1038/s41598-018-21243-x.

Lu, Qiongshi, Ryan L Powles, Sarah Abdallah, Derek Ou, Qian Wang, Yiming Hu, Yisi Lu, et al. 2017. "Systematic Tissue-Specific Functional Annotation of the Human Genome Highlights Immune-Related DNA Elements for Late-Onset Alzheimer's Disease." *PLOS Genetics* 13 (7): e1006933. https://doi.org/10.1371/journal.pgen.1006933.

Luger, K, A W Mader, R K Richmond, D F Sargent, and T J Richmond. 1997. "Crystal Structure of the Nucleosome Core Particle at 2.8 A Resolution." *Nature* 389. https://doi.org/10.1038/38444.

Ma, Sai, Bing Zhang, Lindsay M LaFave, Andrew S Earl, Zachary Chiang, Yan Hu, Jiarui Ding, et al. 2020. "Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin." *Cell* 183 (4): 1103-1116.e20. https://doi.org/https://doi.org/10.1016/j.cell.2020.09.056.

Marshall, Christian R, Daniel P Howrigan, Daniele Merico, Bhooma Thiruvahindrapuram, Wenting Wu, Douglas S Greer, Danny Antaki, et al. 2017. "Contribution of Copy Number Variants to Schizophrenia from a Genome-Wide Study of 41,321 Subjects." *Nature Genetics* 49 (1): 27–35. https://doi.org/10.1038/ng.3725.

Martin, Alicia R, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M Neale, and Mark J Daly. 2019. "Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities." *Nature Genetics* 51 (4): 584–91. https://doi.org/10.1038/s41588-019-0379-x.

Martin, Joanna, Mark J Taylor, and Paul Lichtenstein. 2018. "Assessing the Evidence for Shared Genetic Risks across Psychiatric Disorders and Traits." *Psychological Medicine* 48 (11): 1759–74. https://doi.org/10.1017/S0033291717003440.

Mefford, Heather C, Andrew J Sharp, Carl Baker, Andy Itsara, Zhaoshi Jiang, Karen Buysse, Shuwen Huang, et al. 2008. "Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes." *New England Journal of Medicine* 359 (16): 1685–99. https://doi.org/10.1056/NEJMoa0805384.

Merikangas, Kathleen R, Robert Jin, Jian-Ping He, Ronald C Kessler, Sing Lee, Nancy A Sampson, Maria Carmen Viana, et al. 2011. "Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative." *Archives of General Psychiatry* 68 (3): 241–51. https://doi.org/10.1001/archgenpsychiatry.2011.12.

Miller, Jeremy A, Song-Lin Ding, Susan M Sunkin, Kimberly A Smith, Lydia Ng, Aaron Szafer, Amanda Ebbert, et al. 2014. "Transcriptional Landscape of the Prenatal Human Brain." *Nature* 508 (7495): 199–206. https://doi.org/10.1038/nature13185.

Muench, John, and Ann M Hamer. 2010. "Adverse Effects of Antipsychotic Medications." *American Family Physician* 81 (5): 617—622. http://europepmc.org/abstract/MED/20187598.

Murphy, Kieran C, Lisa A Jones, and Michael J Owen. 1999. "High Rates of Schizophrenia in Adults With Velo-Cardio-Facial Syndrome." *Archives of General Psychiatry* 56 (10): 940–45. https://doi.org/10.1001/archpsyc.56.10.940.

Murray, R M, and S W Lewis. 1987. "Is Schizophrenia a Neurodevelopmental Disorder?" *British Medical Journal (Clinical Research Ed.)* 295 (6600): 681–82. https://doi.org/10.1136/bmj.295.6600.681.

O'Brien, Heath E., Eilis Hannon, Matthew J. Hill, Carolina C. Toste, Matthew J. Robertson, Joanne E. Morgan, Gemma McLaughlin, et al. 2018. "Expression Quantitative Trait Loci in the Developing Human Brain and Their Enrichment in Neuropsychiatric Disorders." *Genome Biology* 19 (1). https://doi.org/10.1186/s13059-018-1567-1.

O'Shea, K Sue, and Melvin G McInnis. 2016. "Neurodevelopmental Origins of Bipolar Disorder: IPSC Models." *Molecular and Cellular Neuroscience* 73: 63–83. https://doi.org/https://doi.org/10.1016/j.mcn.2015.11.006.

Owen, Michael J, and Michael C O'Donovan. 2017. "Schizophrenia and the Neurodevelopmental Continuum:Evidence from Genomics." *World Psychiatry* 16 (3): 227–35. https://doi.org/https://doi.org/10.1002/wps.20440.

Pardiñas, Antonio F., Peter Holmans, Andrew J. Pocklington, Valentina Escott-Price, Stephan Ripke, Noa Carrera, Sophie E. Legge, et al. 2018. "Common Schizophrenia Alleles Are Enriched in Mutation-Intolerant Genes and in Regions under Strong Background Selection." *Nature Genetics* 50 (3). https://doi.org/10.1038/s41588-018-0059-2.

Parikshak, Neelroop N., Rui Luo, Alice Zhang, Hyejung Won, Jennifer K. Lowe, Vijayendran Chandran, Steve Horvath, and Daniel H. Geschwind. 2013. "Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism." *Cell* 155 (5): 1008–21. https://doi.org/10.1016/j.cell.2013.10.031.

Park, P J. 2009. "ChIP-Seq: Advantages and Challenges of a Maturing Technology." *Nat Rev Genet* 10. https://doi.org/10.1038/nrg2641.

Polioudakis, Damon, Luis de la Torre-Ubieta, Justin Langerman, Andrew G Elkins, Xu Shi, Jason L Stein, Celine K Vuong, et al. 2019. "A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-Gestation." *Neuron* 103 (5): 785-801.e8. https://doi.org/10.1016/j.neuron.2019.06.011.

Quinlan, A R, and I M Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26. https://doi.org/10.1093/bioinformatics/btq033.

"Rare Chromosomal Deletions and Duplications in Attention-Deficit Hyperactivity Disorder: A Genome-Wide Analysis." 2010. *British Dental Journal* 209 (11): 567. https://doi.org/10.1038/sj.bdj.2010.1106.

Raudvere, Uku, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, and Jaak Vilo. 2019. "G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update)." *Nucleic Acids Research* 47 (W1): W191–98. https://doi.org/10.1093/nar/gkz369.

Riglin, Lucy, Stephan Collishaw, Alexander Richards, Ajay K Thapar, Barbara Maughan, Michael C O'Donovan, and Anita Thapar. 2017. "Schizophrenia Risk Alleles and Neurodevelopmental Outcomes in Childhood: A Population-Based Cohort Study." *The Lancet Psychiatry* 4 (1): 57–62. https://doi.org/10.1016/S2215-0366(16)30406-0.

Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, et al. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature*. https://doi.org/10.1038/nature14248.

Robinson, Philip J J, and Daniela Rhodes. 2006. "Structure of the '30nm' Chromatin Fibre: A Key Role for the Linker Histone." *Current Opinion in Structural Biology* 16 (3): 336–43. https://doi.org/https://doi.org/10.1016/j.sbi.2006.05.007.

Rucker, James J H, Katherine E Tansey, Margarita Rivera, Dalila Pinto, Sarah Cohen-Woods, Rudolf Uher, Katherine J Aitchison, et al. 2016. "Phenotypic Association Analyses With Copy Number Variation in Recurrent Depressive Disorder." *Biological Psychiatry* 79 (4): 329–36. https://doi.org/10.1016/j.biopsych.2015.02.025.

Sanders, Stephan J, Xin He, A Jeremy Willsey, A Gulhan Ercan-Sencicek, Kaitlin E Samocha, A Ercument Cicek, Michael T Murtha, et al. 2015. "Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci." *Neuron* 87 (6): 1215–33. https://doi.org/10.1016/j.neuron.2015.09.016.

Satterstrom, F. Kyle, Jack A. Kosmicki, Jiebiao Wang, Michael S. Breen, Silvia De Rubeis, Joon Yong An, Minshi Peng, et al. 2020. "Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism." *Cell*. https://doi.org/10.1016/j.cell.2019.12.036.

Saxena, S, M K Funk, and D Chisholm. 2015. "Comprehensive Mental Health Action Plan 2013-2020." *Eastern Mediterranean Health Journal = La Revue de Santé de La Méditerranée Orientale = Al-Majallah Al-Ṣiḥḥīyah Li-Sharq Al-Mutawassiṭ* 21 (July): 461–63. https://doi.org/10.26719/2015.21.7.461.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. 2014. "Biological Insights from 108 Schizophrenia-Associated Genetic Loci." *Nature* 511 (7510): 421–27. https://doi.org/10.1038/nature13595.

Schmitt, Anthony D, Ming Hu, and Bing Ren. 2016. "Genome-Wide Mapping and Analysis of Chromosome Architecture." *Nature Reviews Molecular Cell Biology* 17 (12): 743–55. https://doi.org/10.1038/nrm.2016.104.

Schneider, Maude, Martin Debbané, Anne S Bassett, Eva W C Chow, Wai Lun Alan Fung, Marianne van den Bree, Michael Owen, et al. 2014. "Psychiatric Disorders from

Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome." *The American Journal of Psychiatry* 171 (6): 627–39. https://doi.org/10.1176/appi.ajp.2013.13070864.

Schork, Andrew J, Hyejung Won, Vivek Appadurai, Ron Nudel, Mike Gandal, Olivier Delaneau, Malene Revsbech Christiansen, et al. 2019. "A Genome-Wide Association Study of Shared Risk across Psychiatric Disorders Implicates Gene Regulation during Fetal Neurodevelopment." *Nature Neuroscience* 22 (3): 353–61. https://doi.org/10.1038/s41593-018-0320-0.

Sharp, Andrew J, Heather C Mefford, Kelly Li, Carl Baker, Cindy Skinner, Roger E Stevenson, Richard J Schroer, et al. 2008. "A Recurrent 15q13.3 Microdeletion Syndrome Associated with Mental Retardation and Seizures." *Nature Genetics* 40 (3): 322–28. https://doi.org/10.1038/ng.93.

Sherwood, Richard I, Tatsunori Hashimoto, Charles W O'Donnell, Sophia Lewis, Amira A Barkal, John Peter van Hoff, Vivek Karun, Tommi Jaakkola, and David K Gifford. 2014. "Discovery of Directional and Nondirectional Pioneer Transcription Factors by Modeling DNase Profile Magnitude and Shape." *Nature Biotechnology* 32 (2): 171–78. https://doi.org/10.1038/nbt.2798.

Short, Patrick J., Jeremy F. McRae, Giuseppe Gallone, Alejandro Sifrim, Hyejung Won, Daniel H. Geschwind, Caroline F. Wright, et al. 2018. "De Novo Mutations in Regulatory Elements in Neurodevelopmental Disorders." *Nature*. https://doi.org/10.1038/nature25983.

Singh, Tarjinder, Mitja I Kurki, David Curtis, Shaun M Purcell, Lucy Crooks, Jeremy McRae, Jaana Suvisaari, et al. 2016. "Rare Loss-of-Function Variants in SETD1A Are Associated with Schizophrenia and Developmental Disorders." *Nature Neuroscience* 19 (4): 571–77. https://doi.org/10.1038/nn.4267.

Singh, Tarjinder, Timothy Poterba, David Curtis, Huda Akil, Mariam Al Eissa, Jack Barchas, Nicholas Bass, et al. 2020. "Exome Sequencing Identifies Rare Coding Variants in 10 Genes Which Confer Substantial Risk for Schizophrenia." medRxiv. https://doi.org/10.1101/2020.09.18.20192815.

Song, L, Z Zhang, L L Grasfeder, A P Boyle, P G Giresi, B K Lee, N C Sheffield, et al. 2011. "Open Chromatin Defined by DNaseI and FAIRE Identifies Regulatory Elements That Shape Cell-Type Identity." *Genome Res* 21. https://doi.org/10.1101/gr.121541.111.

Speed, Doug, Na Cai, Michael R Johnson, Sergey Nejentsev, David J Balding, and the UCLEB Consortium. 2017. "Reevaluation of SNP Heritability in Complex Human Traits." *Nature Genetics* 49 (7): 986–92. https://doi.org/10.1038/ng.3865.

Speed, Doug, Gibran Hemani, Michael R Johnson, and David J Balding. 2012. "Improved Heritability Estimation from Genome-Wide SNPs." *American Journal of Human Genetics* 91 (6): 1011–21. https://doi.org/10.1016/j.ajhg.2012.10.010.

Spiers, Helen, Eilis Hannon, Leonard C Schalkwyk, Rebecca Smith, Chloe C Y Wong, Michael C O'Donovan, Nicholas J Bray, and Jonathan Mill. 2015. "Methylomic Trajectories across Human Fetal Brain Development." *Genome Research* 25 (3): 338–52. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=25650246.

Spitz, François, and Eileen E M Furlong. 2012. "Transcription Factors: From Enhancer Binding to Developmental Control." *Nature Reviews Genetics* 13 (9): 613–26. https://doi.org/10.1038/nrg3207.

Stahl, Eli A, Gerome Breen, Andreas J Forstner, Andrew McQuillin, Stephan Ripke, Vassily Trubetskoy, Manuel Mattheisen, et al. 2019. "Genome-Wide Association Study Identifies 30 Loci Associated with Bipolar Disorder." *Nature Genetics* 51 (5): 793–803. https://doi.org/10.1038/s41588-019-0397-8.

Stark, Rory, and Gordon Brown. 2011. "DiffBind : Differential Binding Analysis of ChIP-Seq Peak Data." *Bioconductor*.

Stiles, Joan, and Terry L Jernigan. 2010. "The Basics of Brain Development." *Neuropsychology Review* 20 (4): 327–48. https://doi.org/10.1007/s11065-010-9148-4.

Stilo, Simona A, and Robin M Murray. 2010. "The Epidemiology of Schizophrenia: Replacing Dogma with Knowledge." *Dialogues in Clinical Neuroscience* 12 (3): 305–15. https://doi.org/10.31887/DCNS.2010.12.3/sstilo.

Su, Yingying, Carl D'Arcy, and Xiangfei Meng. 2020. "Research Review: Developmental Origins of Depression – a Systematic Review and Meta-Analysis." *Journal of Child Psychology and Psychiatry* n/a (n/a). https://doi.org/https://doi.org/10.1111/jcpp.13358.

Sullivan, Patrick F, Michael C Neale, and Kenneth S Kendler. 2000. "Genetic Epidemiology of Major Depression: Review and Meta-Analysis." *American Journal of Psychiatry* 157 (10): 1552–62. https://doi.org/10.1176/appi.ajp.157.10.1552.

Suzuki, Miho M, and Adrian Bird. 2008. "DNA Methylation Landscapes: Provocative Insights from Epigenomics." *Nature Reviews Genetics* 9 (6): 465–76. https://doi.org/10.1038/nrg2341.

Terwisscha van Scheltinga, Afke F, Steven C Bakker, Neeltje E M van Haren, Eske M Derks, Jacobine E Buizer-Voskamp, Heleen B M Boos, Wiepke Cahn, et al. 2013. "Genetic Schizophrenia Risk Variants Jointly Modulate Total Brain and White Matter Volume." *Biological Psychiatry* 73 (6): 525–31. https://doi.org/10.1016/j.biopsych.2012.08.017.

Teslovich, Tanya M., Kiran Musunuru, Albert V. Smith, Andrew C. Edmondson, Ioannis M. Stylianou, Masahiro Koseki, James P. Pirruccello, et al. 2010a. "Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids." *Nature* 466 (7307). https://doi.org/10.1038/nature09270.

Teslovich, Tanya M, Kiran Musunuru, Albert V Smith, Andrew C Edmondson, Ioannis M Stylianou, Masahiro Koseki, James P Pirruccello, et al. 2010b. "Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids." *Nature* 466 (August): 707. http://dx.doi.org/10.1038/nature09270.

Tessarz, Peter, and Tony Kouzarides. 2014. "Histone Core Modifications Regulating Nucleosome Structure and Dynamics." *Nature Reviews Molecular Cell Biology* 15 (11): 703–8. https://doi.org/10.1038/nrm3890.

Thapar, Anita, and Lucy Riglin. 2020. "The Importance of a Developmental Perspective in Psychiatry: What Do Recent Genetic-Epidemiological Findings Show?" *Molecular Psychiatry* 25 (8): 1631–39. https://doi.org/10.1038/s41380-020-0648-1.

Thurman, R E, E Rynes, R Humbert, J Vierstra, M T Maurano, E Haugen, N C Sheffield, et al. 2012. "The Accessible Chromatin Landscape of the Human Genome." *Nature* 489. https://doi.org/10.1038/nature11232.

Tompa, Martin, Nan Li, Timothy L Bailey, George M Church, Bart De Moor, Eleazar Eskin, Alexander V Favorov, et al. 2005. "Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites." *Nature Biotechnology* 23 (1): 137–44. https://doi.org/10.1038/nbt1053.

Torkamani, Ali, Nathan E Wineinger, and Eric J Topol. 2018. "The Personal and Clinical Utility

of Polygenic Risk Scores." *Nature Reviews Genetics* 19 (9): 581–90. https://doi.org/10.1038/s41576-018-0018-x.

Trynka, Gosia, and Soumya Raychaudhuri. 2013. "Using Chromatin Marks to Interpret and Localize Genetic Associations to Complex Human Traits and Diseases." *Current Opinion in Genetics & Development* 23 (6): 635–41. https://doi.org/10.1016/j.gde.2013.10.009.

Tyrer, Peter, and Angus Mackay. 1986. "Schizophrenia: No Longer a Functional Psychosis." *Trends in Neurosciences* 9: 537–38. https://doi.org/https://doi.org/10.1016/0166-2236(86)90169-4.

Walker, Rebecca L, Gokul Ramaswami, Christopher Hartl, Nicholas Mancuso, Michael J Gandal, Luis de la Torre-Ubieta, Bogdan Pasaniuc, Jason L Stein, and Daniel H Geschwind. 2019. "Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms." *Cell* 179 (3): 750-771.e22. https://doi.org/10.1016/j.cell.2019.09.021.

Walter, Klaudia, Josine L Min, Jie Huang, Lucy Crooks, Yasin Memari, Shane McCarthy, John R B Perry, et al. 2015. "The UK10K Project Identifies Rare Variants in Health and Disease." *Nature* 526 (7571): 82–90. https://doi.org/10.1038/nature14962.

Wang, Daifeng, Shuang Liu, Jonathan Warrell, Hyejung Won, Xu Shi, Fabio C P Navarro, Declan Clarke, et al. 2018. "Comprehensive Functional Genomic Resource and Integrative Model for the Human Brain." *Science* 362 (6420): eaat8464. https://doi.org/10.1126/science.aat8464.

Weinberger, Daniel R. 1987. "Implications of Normal Brain Development for the Pathogenesis of Schizophrenia." *Archives of General Psychiatry* 44 (7): 660–69. https://doi.org/10.1001/archpsyc.1987.01800190080012.

Werling, Donna M, and Daniel H Geschwind. 2013. "Sex Differences in Autism Spectrum Disorders." *Current Opinion in Neurology* 26 (2). https://journals.lww.com/co-neurology/Fulltext/2013/04000/Sex_differences_in_autism_spectrum_disorders.6.aspx.

Willsey, A. Jeremy, Stephan J. Sanders, Mingfeng Li, Shan Dong, Andrew T. Tebbenkamp, Rebecca A. Muhle, Steven K. Reilly, et al. 2013. "XCoexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism." *Cell*. https://doi.org/10.1016/j.cell.2013.10.020.

Wittchen, Hans-Ulrich, Stephan Mhlig, and Lukas Pezawas. 2003. "Natural Course and Burden of Bipolar Disorders." *The International Journal of Neuropsychopharmacology* 6 (2): 145—154. https://doi.org/10.1017/s146114570300333x.

Won, Hyejung, Luis de la Torre-Ubieta, Jason L Stein, Neelroop N Parikshak, Jerry Huang, Carli K Opland, Michael J Gandal, et al. 2016. "Chromosome Conformation Elucidates Regulatory Relationships in Developing Human Brain." *Nature* 538 (7626): 523–27. https://doi.org/10.1038/nature19847.

World Health Organization. 2018. *International Classification of Sideases for Mortality and Morbidity Statistics (11th Revision)*. https://icd.who.int/browse11/l-m/en.

Wray, Naomi R, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski, Enda M Byrne, Abdel Abdellaoui, Mark J Adams, et al. 2018. "Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression." *Nature Genetics* 50 (5): 668–81. https://doi.org/10.1038/s41588-018-0090-3.

Xin, Dedong, Landian Hu, and Xiangyin Kong. 2008. "Alternative Promoters Influence Alternative Splicing at the Genomic Level." *PloS One* 3 (6): e2377–e2377. https://doi.org/10.1371/journal.pone.0002377.

Yang, Jian, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, et al. 2010. "Common SNPs Explain a Large Proportion of the Heritability for Human Height." *Nature Genetics* 42 (7): 565–69. https://doi.org/10.1038/ng.608.

Young, Alexander I. 2019. "Solving the Missing Heritability Problem." *PLoS Genetics* 15 (6): e1008222–e1008222. https://doi.org/10.1371/journal.pgen.1008222.

Zeng, Hongkui, Elaine H. Shen, John G. Hohmann, Seung Wook Oh, Amy Bernard, Joshua J. Royall, Katie J. Glattfelder, et al. 2012. "Large-Scale Cellular-Resolution Gene Profiling in Human Neocortex Reveals Species-Specific Molecular Signatures." *Cell* 149 (2): 483–96. https://doi.org/https://doi.org/10.1016/j.cell.2012.02.052.

Zhang, Feng, and James R Lupski. 2015. "Non-Coding Genetic Variants in Human Disease." *Human Molecular Genetics* 24 (R1): R102–10. https://doi.org/10.1093/hmg/ddv259.

Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology* 9 (9): R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhong, Suijuan, Shu Zhang, Xiaoying Fan, Qian Wu, Liying Yan, Ji Dong, Haofeng Zhang, et al. 2018. "A Single-Cell RNA-Seq Survey of the Developmental Landscape of the Human Prefrontal Cortex." *Nature* 555 (March): 524. http://dx.doi.org/10.1038/nature25980.

Zhu, Huanhuan, and Xiang Zhou. 2020. "Statistical Methods for SNP Heritability Estimation and Partition: A Review." *Computational and Structural Biotechnology Journal* 18: 1557–68. https://doi.org/https://doi.org/10.1016/j.csbj.2020.06.011.

Zhu, Lihua J., Claude Gazin, Nathan D. Lawson, Hervé Pagès, Simon M. Lin, David S. Lapointe, and Michael R. Green. 2010. "ChIPpeakAnno: A Bioconductor Package to Annotate ChIP-Seq and ChIP-Chip Data." *BMC Bioinformatics* 11. https://doi.org/10.1186/1471-2105-11-237.

Zuk, Or, Eliana Hechter, Shamil R Sunyaev, and Eric S Lander. 2012. "The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability." *Proceedings of the National Academy of Sciences* 109 (4): 1193 LP – 1198. https://doi.org/10.1073/pnas.1119675109.

Appendix

LDSC code example

```
. /ldsc/ldsc.py \
--l2 \
--bfile ./ldsc/1000G_plinkfilesPhase1/1000G.mac5eur.$i \
--ld-wind-cm 1 \
--annot /home/c1654239/ldsc/my.annot.phase1/my.annot.$i.annot \
--out /home/c1654239/ldsc/my.annot.phase1.$i \
--print-snps /ldsc/hapmap3_snps/hm.$i.snp
```

SLDSC code example

```
./ldsc/ldsc.py \
--h2 ./ldsc/SCZ.sumstats \
--ref-ld-chr

/ldsc/baselineLD_v1.1/baselineLD.,./ldsc/my.annotation/my.annot. \
--w-ld-chr ./ldsc/1000G_Phase3_weights_hm3_no_MHC/weights.hm3_noMHC. \
--overlap-annot \
--frqfile-chr /ldsc/1000G_Phase3_frq/1000G.EUR.QC. \
--out SCZ_my.annot. \
--print-coefficients
```

Bedtools code example

```
bedtools intersect -wa -a ATAC_peaks -b H3K4Me1_peaks  > ATAC_H3K4Me1_intersection_peaks

bedtools merge -i ATAC_H3K4Me1_intersection_peaks > Merged_ ATAC_H3K4Me1_intersection_peaks
```

GARFIELD code example

```
INPUTNAME=ADHD
DATADIR=./garfield/garfield-data

PRUNETAGSDIR=$DATADIR/tags/r01
CLUMPTAGSDIR=$DATADIR/tags/r08
MAFTSSDDIR=$DATADIR/maftssd
PVALDIR=$DATADIR/pval/$INPUTNAME
ANNOTDIR=$DATADIR/myannot
OUTDIR=$DATADIR/output/$INPUTNAME-myannot
mkdir -p $OUTDIR

ANNOTLINKFILE=$ANNOTDIR/link_file.txt
PTHRESH=1e-5,5e-8
```

```
BINNING=m5,n5,t5
CONDITION=0
SUBSET="1-1005"

F1=$OUTDIR/garfield.prep.$INPUTNAME.out
F0=$OUTDIR/garfield.Meff.$INPUTNAME.out

echo 'Prune and Clump'
echo -n > $F1
for CHR in `seq 1 22` #X
do
        echo 'CHR'$CHR
        ./garfield/garfield-v2/garfield-prep-chr -ptags $PRUNETAGSDIR/chr$CHR -ctags
$CLUMPTAGSDIR/chr$CHR -maftss $MAFTSSDDIR/chr$CHR -pval $PVALDIR/chr$CHR -ann
$ANNOTDIR/chr$CHR -excl 895,975,976,977,978,979,980 -chr $CHR -o $F1 || { echo 'Failure!'; }
done

echo 'Calculate effective number of annotations'
Rscript ./garfield/garfield-v2/garfield-Meff-Padj.R -i $F1 -o $F0
NEA=$(head -1 $F0 |awk '{print $2}')
Padj=$(tail -1 $F0 |awk '{print $2}')

echo 'Calculate Enrichment and Significance'
F2=$OUTDIR/garfield.test.$INPUTNAME.out
Rscript ./garfield/garfield-v2/garfield-test.R -i $F1 -o $F2 -l $ANNOTLINKFILE -pt $PTHRESH -b $BINNING -s
$SUBSET -c $CONDITION
echo 'GARFIELD single annotation analysis complete'

echo 'Create Plots'
Rscript ./garfield/garfield-v2/garfield-plot.R -i $F2 -o $F2 -l $ANNOTLINKFILE -t " " -f 10 -padj $Padj

echo 'Prioritize relevant annotations by conditional analysis'
# Additional prioritization of annotations
CONDITION=1
CONDITIONTHRESH=0.05
Rscript ./garfield/garfield-v2/garfield-test.R -i $F1 -o $F2 -l $ANNOTLINKFILE -pt $PTHRESH -b $BINNING -s
$SUBSET -c $CONDITION -ct $CONDITIONTHRESH -padj $Padj
echo 'GARFIELD model selection complete'

# Step to extract variants driving the enrichment signals
echo 'Extracting variants driving enrichment analysis signals'
PTHRESH=5e-8, 1e-5, ## GWAS threshold
PENRICH=0.05 ## Enrichment significance threshold
GARFIELD_significant_annotations=$F2.significant.annotations.$PTHRESH.$PENRICH
GARFIELD_VARS=${GARFIELD_significant_annotations}.variants
```