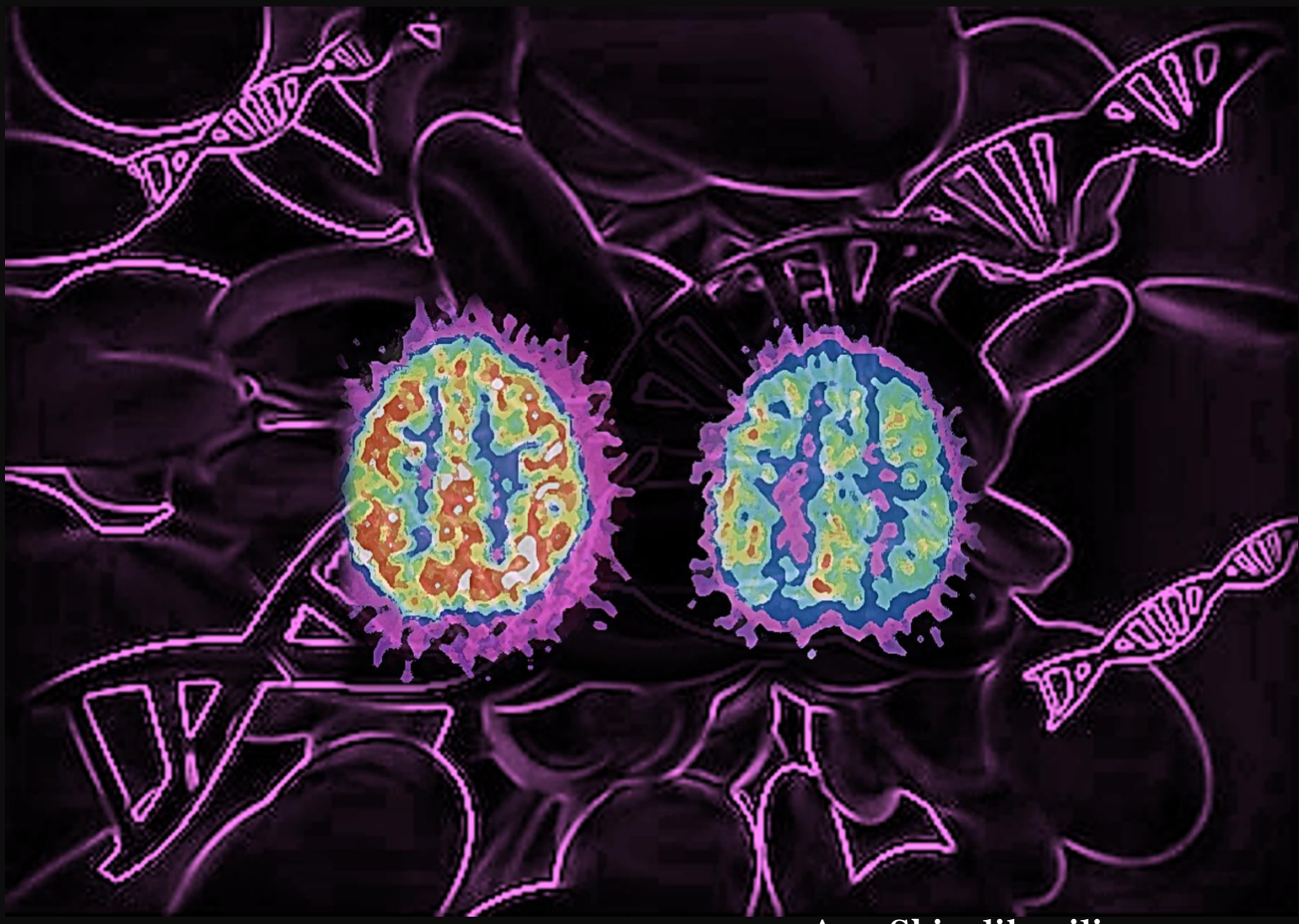
UNIVERSITAT DE BARCELONA

Genetics and epigenetics of attention-deficit/hyperactivity disorder and comorbid conditions

Anu Shivalikanjli

Genetics and epigenetics of attention-deficit/hyperactivity disorder and comorbid conditions



Anu Shivalikanjli

2020

Genetics and epigenetics of attention-deficit/hyperactivity disorder and comorbid conditions

Doctoral Thesis presented by

Anu Shivalikanjli

To apply for the title of

Doctor of the University of Barcelona

Programme of Genetics

Department of Genetics, Microbiology and Statistics

Neurogenetics Group

Dr. Bru Cormand Rifà

Supervisor/Tutor

Dr. Stephen V. Faraone

Supervisor

Anu Shivalikanjli

Barcelona, December 2020

UNIVERSITAT DE BARCELONA

This Thesis is

dedicated to my Father,

the strongest man in my life

Who taught me to read,

and whose Dream

this Doctorate

has been all

along

*

"If I could go back in time,
I would still choose to do a PhD"

...said no one ever

PREFACE

The following funding sources have sponsored this doctoral thesis:

- The European Union H2020 Program [H2020/2014-2020] under grant agreement n° 643051. Project: MiND (Mastering skills in the training Network for attention deficit hyperactivity and autism spectrum Disorders)

- The European Union H2020 Program [H2020/2014-2020] under grant agreement n° 667302. Project: CoCA (Comorbid Conditions of ADHD)

- The European Union H2020 Program [H2020/2014-2020] under grant agreement n° 728018. Project: Eat2beNICE

- The Spanish 'Ministerio de Economía y Competitividad' (SAF2015-68341-R, RTI2018-100968-B-I00)

- AGAUR, 'Generalitat de Catalunya' (2017-SGR-738).

THESIS SUMMARY

The broad objectives of this work are the identification of genes that contribute to the susceptibility to attention-deficit/hyperactivity disorder (ADHD) and cocaine dependence, two disorders that co-occur in patients. In this Thesis, we describe (i) the contribution to ADHD of allele-specific methylation (ASM), an epigenetic mechanism that involves single single-nucleotide polymorphisms (SNPs) correlating with differential levels of DNA methylation at CpG sites, (ii) the role of microRNA (miRNA) genes in ADHD, and (iii) a genome-wide association meta-analysis of cocaine dependence. We also explore the common genetic basis that explains the comorbidity between these disorders.

The main results from the three studies include:

(i) Common genetic risk variants for ADHD identified in a previous genome-wide association study (GWAS) that included 20,000 cases and 35,000 controls are enriched in SNPs that correlate with levels of DNA methylation. Eight ASM SNPs were found significantly associated with ADHD and correlated with differential methylation at six CpG sites *in cis* in different brain areas. These six CpG sites are located at possible promoter regions of six genes expressed in brain: *ARTN*, *C2orf82*, *NEUROD6*, *PIDD1*, *RPLP2* and *GAL*. Based on the bioinformatic functional analyses of these genes, our study highlights the candidacy of *ARTN*, *C2orf82* and *PIDD1* genes as potential contributors to ADHD susceptibility.

(ii) We conducted a case-control association study to investigate the contribution to ADHD of common genetic variation in 1,761 autosomal miRNAs using pre-existing GWAS data from 20,000 cases and 35,000 controls. We identified significant associations of SNPs with ADHD that highlight 12 miRNA genes, all located within protein-coding genes. The associated variants are located in the putative regulatory regions of the miRNA genes or in the promoter region of the host protein-coding gene. We inspected the target genes, brain expression, homologs for the miRNAs and we propose miR-7-1 and miR-3666 as promising candidates since both are brain expressed, have validated brain-expressed targets, and homologs in model species. Pathway analysis of ADHD-associated miRNAs revealed miRNA-mediated regulation of serotonin receptor genes, well-known contributors to neurological functions and diseases.

(iii) We performed the largest cocaine dependence GWAS meta-analysis in individuals of European ancestry, including 2,100 cases and 4,300 controls. Although SNP-based analysis revealed no genome-wide significant associations with cocaine dependence, probably due to limited sample size, gene-based analysis identified the *HIST1H2BD* gene, previously associated with schizophrenia. The estimated SNP-based heritability of cocaine dependence was estimated as 30%. A significant genetic correlation was found between cocaine dependence and ADHD, schizophrenia, major depressive disorder and risk-taking behaviour, suggesting a shared genetic basis across pathologies and traits. Polygenic risk score (PRS) analysis shows that all the comorbid features analysed (ADHD, schizophrenia, major depressive disorder, aggressiveness, antisocial personality or risk-taking behaviour) can predict cocaine dependence.

Overall, we identified common genetic and epigenetic risk factors that underlie the susceptibility to ADHD and to cocaine dependence. The results reinforce the idea that epigenetic mechanisms dictate the differential expression of genes that may be causal to ADHD. Cocaine dependence, which has been widely believed to occur under environmental and epigenetic influences, is also in part genetically determined. Finally, ADHD and cocaine dependence are comorbid disorders, and the observed genetic correlation between these conditions can reflect biological pleiotropy.

RESUM DE LA TESI

Aquest treball té com a objectiu principal la identificació de gens que contribueixen a la susceptibilitat al trastorn per dèficit d'atenció amb hiperactivitat (TDAH) i a la dependència de cocaïna, dos trastorns que es presenten amb freqüència conjuntament en pacients. En aquesta Tesi es descriu (i) la contribució al TDAH de la metilació específica de l'al·lel (ASM), un mecanisme epigenètic pel qual variants polimòrfiques presenten correlació amb nivells diferencials de metilació de l'ADN en llocs CpG, (ii) el paper dels gens de microRNAs (miRNAs) en el TDAH, i (iii) una meta-anàlisi d'estudis d'associació a escala genòmica de la dependència de cocaïna. També explorem la base genètica comuna que explica la comorbiditat entre aquests dos trastorns.

Els principals resultats dels tres estudis són:

(i) Les variants genètiques comunes de risc al TDAH identificades en un estudi previ d'associació a escala genòmica (GWAS) amb 20.000 casos i 35.000 controls estan enriquides en variants de canvi d'un sol nucleòtid (SNPs) que tenen influència sobre la metilació de l'ADN. Vuit SNPs de tipus ASM estan associats significativament amb el TDAH i presenten correlació amb la metilació diferencial de sis dinucleòtids CpG en *cis* en diferents àrees cerebrals. Aquests sis llocs CpG estan en possibles regions promotores de sis gens que s'expressen al cervell: *ARTN, C2orf82, NEUROD6, PIDD1, RPLP2* i *GAL*. En base a anàlisis bioinformàtiques d'aquests gens a nivell funcional, el nostre estudi prioritza els gens *ARTN, C2orf82* i *PIDD1* com a possibles contribuents a la susceptibilitat al TDAH.

(ii) Hem dut a terme un estudi d'associació cas-control per investigar la contribució al TDAH de la variació genètica comuna en 1.761 miRNA autosòmics utilitzant dades GWAS preexistents de 20.000 casos i 35.000 controls. Hem identificat associacions significatives de SNPs amb el TDAH que assenyalen 12 gens de miRNAs, tots situats dins de gens que codifiquen proteïnes. Les variants associades estan situades en suposades regions reguladores dels gens de miRNA o a la regió promotora del gen hoste. Hem inspeccionat els gens diana dels miRNAs, la seva expressió en cervell i els gens homòlegs en altres espècies, i proposem els gens miR-7-1 i miR-3666 com a candidats prometedors, ja que tots dos són s'expressen al sistema nerviós central, tenen dianes validades que s'expressen també en cervell i tenen homòlegs en espècies model. L'anàlisi de vies a partir dels miRNAs associats al TDAH ha

assenyalat gens de receptors de serotonina regulats pels nostres miRNAs, la relació dels quals amb funcions i malalties neurològiques és ben coneguda.

(iii) Hem realitzat la metaanàlisi més gran fins ara de dades GWAS de dependència de cocaïna en individus d'ascendència europea, amb 2.100 casos i 4.300 controls. Tot i que l'anàlisi basada en SNPs no ha revelat cap associació significativa amb la dependència de cocaïna, probablement a causa de la mida mostral limitada, l'anàlisi basada en gens ha permès identificar el gen *HIST1H2BD*, anteriorment associat a l'esquizofrènia. Hem calculat també l'heretabilitat basada en SNPs de la dependència de cocaïna, que seria d'un 30%. Hem detectat una correlació genètica significativa entre la dependència de cocaïna i el TDAH, l'esquizofrènia, el trastorn depressiu major i els comportaments de risc, tot suggerint que hi ha una base genètica compartida entre patologies i trets. L'anàlisi de la puntuació de risc poligènic (PRS) mostra que totes les característiques comòrbides analitzades (TDAH, esquizofrènia, trastorn depressiu major, agressivitat, personalitat antisocial o comportaments de risc) prediuen la dependència de la cocaïna.

En resum, hem identificat factors de risc genètics i epigenètics freqüents a la població que contribueixen a la susceptibilitat al TDAH i a la dependència de cocaïna. Els resultats reforcen la idea que els mecanismes epigenètics estan relacionats amb l'expressió diferencial de gens que poden contribuir al TDAH. La dependència de cocaïna, que fins ara s'havia relacionat amb factors de risc ambientals i epigenètics, també estaria determinada, en part, per factors genètics. Finalment, el TDAH i la dependència de la cocaïna són trastorns comòrbids i la correlació genètica observada entre aquestes afeccions pot reflectir pleiotropia biològica.

Contents

ABBREVIATIONS

3′UTR: Three prime untranslated region

5-HT: 5-hydroxytryptamine (serotonin)

AA: African-American

AAED1: AhpC/TSA antioxidant enzyme domain containing 1

ADE: Adverse drug events

ADGRL3: Adhesion G protein-coupled receptor L3

ADHD: Attention-deficit/hyperactivity disorder

AGO1: Argonaute RISC component 1

AMPAR: α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor

ARTN: Artemin

ASD: Autism spectrum disorder

ASE: Allele-specific expression

ASM: Allele-specific methylation

BDNF: Brain derived neurotrophic factor

BeCOME: Biological classification of mental disorders

Bhlhb2: Basic helix-loop-helix family member e40

BPD: Borderline personality disorder

BTN3A2: Butyrophilin subfamily 3 member A2

C2orf82: Secondary ossification center associated regulator of chondrocyte maturation

CADD: Combined annotation dependent depletion

CBT: Cognitive-behavioral therapy

CD: Conduct disorder

CDCV: Common disease/common variant

CDRV: Common disease/rare variant

CHRNA5: Cholinergic receptor nicotinic alpha5 subunit

CNR1: Cannabinoid receptor 1

CNS: Central nervous system

CNV: Copy number variant

COL10A1: Collagen type X alpha 1 chain

COMT: Catechol-O-methyl transferase

CRH: Corticotropin releasing hormone

CRISPR: Clustered regularly interspaced short palindromic repeats

CSF: Cerebrospinal fluid

CTD: Chronic tic disorder

CYP2D6: Cytochrome P450 family 2 subfamily D member 6

CYP2C19: Cytochrome P450 family 2 subfamily C member 19

DA: Dopamine

DANN: Deleterious annotation of genetic variants using neural networks

DAT: Dopamine transporter

dbGAP: Database of genotypes and phenotypes

dbSNP: Single nucleotide polymorphism database

DCC: Deleted in colorectal cancer netrin 1 receptor

DIRAS: Di-ras

DISC1: Disrupted-in-schizophrenia 1 scaffold protein

DNA: Deoxyribonucleic acid

DNMT: DNA methyltransferase

DRD: Dopamine receptor

DSM: Diagnostic and statistical manual of mental disorders

DUSP6: Dual specificity phosphatase 6

EA: European-American

ED: Emotion dysregulation

EE: Enriched environment

EGFR: Epidermal growth factor receptor

EHR: Electronic health records

EIF4E: Eukaryotic translation initiation factor 4E

eMERGE: Electronic medical records and genomics

ENCODE: Encyclopedia of DNA elements

ENIGMA: Enhancing neuro imaging genetics through meta-analysis

eQTL: Expression quantitative trait loci

EUR: European

FAM53B: Family with sequence similarity 53 member B

FDR: False discovery rate

FGFR: Fibroblast growth factor receptor

FOXP2: Forkhead box P2

FUMA GWAS: Functional mapping and annotation of genome-wide association studies

FUSION: Functional summary-based imputation

GABA: Gamma-aminobutyric acid

GABRB1: Gamma-aminobutyric acid type A receptor subunit beta1

GCTA: Genome-wide complex trait analysis

GECCO: Geoscience and health cohort consortium

GGHDC: Global geo health data center

GO: Gene ontology

GREML: Genome-based restricted maximum likelihood

GRM4: Glutamate metabotropic receptor 4

GRS: Genetic risk scores

GSMA: Genome scan meta-analysis

GTEx: Genotype-tissue expression

GTP: Guanosine triphosphate

GWAS: Genome-wide association study

GWPA: Genome-wide pathway analysis

GWS: Genome-wide significant

GxE: Gene-environment interaction

GxG: Gene-gene interaction

h2: Heritability

H3K27: Histone H3 on lysine 27

H3K27ac: Histone H3 lysine 27 acetylation

H3K4: Histone H3 on lysine 4

H3K4me1: Histone H3 lysine 4 methylation

H3K4me3: Histone H3 lysine 4 trimethylation

H3K9: Histone H3 lysine 9

H3K9ac: Histone H3 lysine 9 acetylation

HDAC: Histone deacetylase inhibitors

HI: Hyperactive/impulsive

HIST1H2AK: H2A clustered histone 15

HIST1H2BD: H2B clustered histone 5

HIVEP2: Human immunodeficiency virus type I enhancer binding protein 2

HMGA2: High mobility group AT-hook 2

HMT: Histone demethylase inhibitors

hsa-miR: Human microRNA

HT1B: 5-hydroxytryptamine receptor 1B

HTR1D: 5-hydroxytryptamine receptor 1D

HTT: Solute carrier family 6 member 4

HWE: Hardy-Weinberg equilibrium

ICD: International classification of diseases

ID: Intellectual disability

iN: Induced neuron

iPSC: Induced pluripotent stem cell

IQ: Intelligence quotient

iMEGES: Integrated mental-disorder genome Score

INFERNO: Inferring the molecular mechanisms of noncoding genetic variants

IPA: Ingenuity pathway analysis

LD: Linkage disequilibrium

LDSC: Linkage disequilibrium score

LgD: Learning disabilities

lncRNA: Long non-coding RNA

LPHN3: Latrophilin3

MAF: Minor allele frequency

MAOA: Monoamine oxidase

MAP1A: Microtubule associated protein 1A

MDD: Major depressive disorder

MDS: Multidimensional scaling

MEF2C: Myocyte enhancer factor 2C

MEOX2: Mesenchyme homeobox 2

MGH: Massachusetts General Hospital

MHC: Major histocompatibility complex

miR: MicroRNA

miRNA: MicroRNA

mQTL: Methylation quantitative trait loci

MR: Mendelian randomization

mRNA: Messenger RNA

MRI: Magnetic resonance imaging

MTAG: Multi-trait analysis of GWAS

ncRNAs: Non-coding RNAs

NCS-R: National comorbidity survey replication

NGS: Next-generation sequencing

NHGRI-EBI: National human genome research institute-European bioinformatics institute

NIH: National institutes of health

NLGN3: Neuroligin 3

NOTCH: Notch receptor

NOTCH2: Notch receptor 2

Nr3c1: Nuclear receptor subfamily 3 group C member 1

NT5DC1: 5'-nucleotidase domain containing 1

OCD: Obsessive-compulsive disorder

ODD: Oppositional defiant disorder

OMIM: Online Mendelian inheritance in man

PC: Principal component

PCA: Principal component analysis

PD: Personality disorder

PFC: Prefrontal cortex

PGC: Psychiatric genomics consortium

PheWAS: Phenome-wide association studies

PICK1: Protein interacting with protein kinase C alpha type (PRKCA) 1

PIDD1: p53-induced death domain protein 1

PLCB1: Phospholipase C beta 1

POLR3C: RNA polymerase III subunit C

pri-miR: Primary microRNA

PRSS16: Serine protease 16

PSD: Pleckstrin and Sec7 domain containing

PTSD: Post-traumatic stress disorder

QC: Quality control

QTL: Quantitative trait loci

RARG: Retinoic acid receptor gamma

RBFOX1: RNA binding fox-1 homolog 1

RORa: RAR related orphan receptor A

RNA: Ribonucleic acid

SAGE: Study of addiction: genetics and environment

SCID: Structured clinical interview for DSM disorders

SEC23IP: SEC23 interacting protein

SEMA6D: Semaphorin 6D

SERT: Serotonin receptor

SHR: Spontaneously hypertensive rat

SLC17A7: Solute carrier family 17 member 7

SLC6A3: Solute carrier family 6 member 3

SLC6A4: Solute carrier family 6 member 4

SNAP: Synaptosome associated protein

SNCA: Synuclein alpha

SNP: Single nucleotide polymorphism

SNP-h2: Single nucleotide polymorphism-based heritability

SNV: Single nucleotide variant

SORCS3: Sortilin related VPS10 domain containing receptor 3

ss-eQTL: Sex-specific expression quantitative trait loci

SUD: Substance use disorder

SVM: Support vector machine

TAC1: Tachykinin precursor 1

TC: Tic disorder

TS: Tourette syndrome

TSS: Transcription start site

TWAS: Transcriptome-wide association studies

VAMP2: Vesicle associated membrane protein 2

VEP: Variant effect predictor

VNTR: Variable number of tandem repeats

WES: Whole exome sequencing

WGS: Whole genome sequencing

YWHAG: Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein gamma

ZCCHC4: Zinc finger CCHC-type containing 4

ZNF184: Zinc finger protein 184

ZSCAN31: Zinc finger and SCAN domain containing 31

ZCCHC4: Zinc finger CCHC-type containing 4

ZNF184: Zinc finger protein 184

ZSCAN31: Zinc finger and SCAN domain containing

LIST OF FIGURES

INTRODUCTION

CHAPTER 1. OVERVIEW

Attention-deficit/hyperactivity disorder (ADHD) is a common childhood-onset neurodevelopmental psychiatric condition. It is characterized by impairing symptoms of age-inappropriate inattention, impulsivity and hyperactivity. The disorder affects around 5-7% of children and adolescents worldwide[1,2]. According to longitudinal data, ADHD symptoms persist into adulthood in nearly 65% of the affected individuals[3], making it a lifelong state. The prevalence of adult ADHD is 2.5-5%[1,3,4]. The questions on causation, risk, mediating factors and lifespan trajectory of this disorder remain still poorly understood.

1.1 Clinical symptomatology

In line with the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V)[5], ADHD symptoms must be discernible before the age of 12 years with significant debilitation of child's integration in more than one environment, e.g., school or work, or family and peers. The previous criterion, according to DSM-IV, for the age of onset of ADHD symptoms was 7 years[6]. Based on the associated behavioral symptoms, there are three subtypes of ADHD (DSM-IV): predominantly hyperactive/impulsive (HI), predominantly inattentive, or combined. However, DSM-V puts less emphasis on these distinctions[7]. It is relatively easier to notice the contrasting subtypes of inattention and HI. However, the presence of the combined form of ADHD severely disturbs the functioning of an individual[8]. A greater percentage of males fall within the HI spectrum, while females tend to present more inattentive symptoms and often go undiagnosed[8,9]. This may explain, in part, the disparity in the reported incidence of ADHD between both genders.

1.2 Comorbidities

It is well known that ADHD is highly comorbid with other psychiatric and non-psychiatric conditions, throughout the life span[4]. Major comorbidities observed during several stages of a subject with ADHD are learning disorders, tics, autism spectrum disorders (ASD), conduct disorder (CD), oppositional defiant disorder (ODD), mood and anxiety disorders, antisocial behaviors, sleep disorders, major depressive disorder (MDD), substance use disorders (SUDs) and obsessive-compulsive disorder (OCD)[4,9,10].

The comorbidities and their impact often change during the lifetime of patients[10]. For example, children with ADHD might show more CD, ODD and anxiety; while adolescents with ADHD

are likely to exhibit OCD and SUD[4,11]. These co-occurring behaviors can be direct outcomes of the ADHD and increase the severity of the disorder. The relatively poor performance in education or in work spaces may trigger anxiety and depression in ADHD individuals[11]. Social exclusion may lead to increased loneliness or to antisocial behavior at later stages[12].

1.3 Social problems

The disorder poses impediments that begin in childhood and continue throughout the lifespan generating a personal burden that can be highly impairing. HI children are more likely to present aggressiveness or behaviors that are inadequate in sociocultural contexts. This increases their odds of facing social rejection from their non-ADHD counterparts[13]. Adolescents with ADHD have troubles building peer relationships because of the reported low esteem, which also exposes them to a greater risk of substance abuse[14].

While the medication benefits the condition, it is certainly not a substitute for interpersonal and social skills development. There are concealed struggles by ADHD-afflicted in achieving high working memory, decision making, time management, impulse control, endurance in personal or professional tasks and follow-through on ideas. In most instances, not all these parameters are satisfactorily met.

In addition, the stigma attached to ADHD as with many other psychiatric conditions, challenges the integration of the affected into the society, which is presently still not prepared to accept the neurodiverse community. Despite the global validity of ADHD as a psychiatric condition, it is still not prioritized in some societies, which hinders the conditioning and progress of the suffering individuals.

1.4 Misdiagnosis of ADHD

Multiple issues can complicate and interfere with a clear and uniform diagnosis of ADHD. First, the clinical presentation of ADHD may vary according to gender, age and stage of development[8,15,16]. Second, the uniformity of ADHD diagnosis is hindered by the presence of cultural differences in the expected activity and inattention levels[8,15,16]. Third, the presence of comorbid conditions may mask the ADHD symptoms and lead to misdiagnosis[16]. In the presence of multiple manifestations, detailed screenings are needed to figure out the causal condition and the effects[9]. Fourth, the inattentive form of ADHD gets more easily unnoticed, especially in the absence of comorbidities.

1.5 Therapeutic approaches

ADHD treatment involves a combination of pharmacological and non-pharmacological interventions in children, adolescents and adults[4]. Pharmacotherapy relies mainly on the use of stimulant medications, e.g. methylphenidate, pemoline and dextroamphetamine[17,18]. Mostly considered safe in structured doses, these molecules may improve the core symptoms of inattention, hyperactivity and impulsivity[19]. The affected children and adolescents benefit from an increased academic performance and social functioning at school and at home, and adults tend to cope better with occupational and social dysfunctions while under stimulant medication[20–23]. Nearly all the treatments show the same efficacy, with around 70% of ADHD patients responding to the available treatments that target mainly the dopaminergic and noradrenergic systems[17,19]. However, there is a need for research into the causal biological pathways in ADHD to address new targets and for a better outreach. In children, non-pharmacological interventions involving cognitive-behavioral therapy (CBT) are perceived to be less efficient than stimulant medication, if used alone[24]. However, CBT can assist in managing comorbid behavior disruptions like CD[25]. CBT has been shown to be more effective in adolescents with ADHD than in children[26,27].

CHAPTER 2. ETIOLOGY OF ADHD

ADHD etiology is believed to be multifactorial, with genetics contributing significantly to the cause of the disorder, together with environmental risk factors. ADHD is one of the psychiatric disorders with the strongest genetic basis according to familial, twin, and SNP-based statistics[3,4].

2.1 Heritability estimates in ADHD

Heritability is a measure to understand how much of the variation in a given trait can be attributed to genetic factors. Two main approaches are implemented to quantify heritability. The classic method employs twin studies, where the concordance (or correlation) of a phenotype is assessed in monozygotic twins (sharing 100% of their genomes) versus dizygotic twins (sharing 50% of their genomes). Another method identifies the impact of the environment in addition to the genetic contribution in twins. The environmental risks may not be shared among siblings. This estimates the genetic (A), shared environment (C) and unique environment (E) in the twin studies and is therefore termed as 'ACE' model[3]. Twenty twin studies in ADHD estimated the heritability of this psychiatric disorder as 76%[28] and the latest estimated mean heritability from across 37 twin studies is 74%[29], retaining consistency over the decade (Figure 1). This indicates that three quarters of the phenotype variation in ADHD is due to genetic variation, and therefore ADHD is among the most heritable of psychiatric disorders.

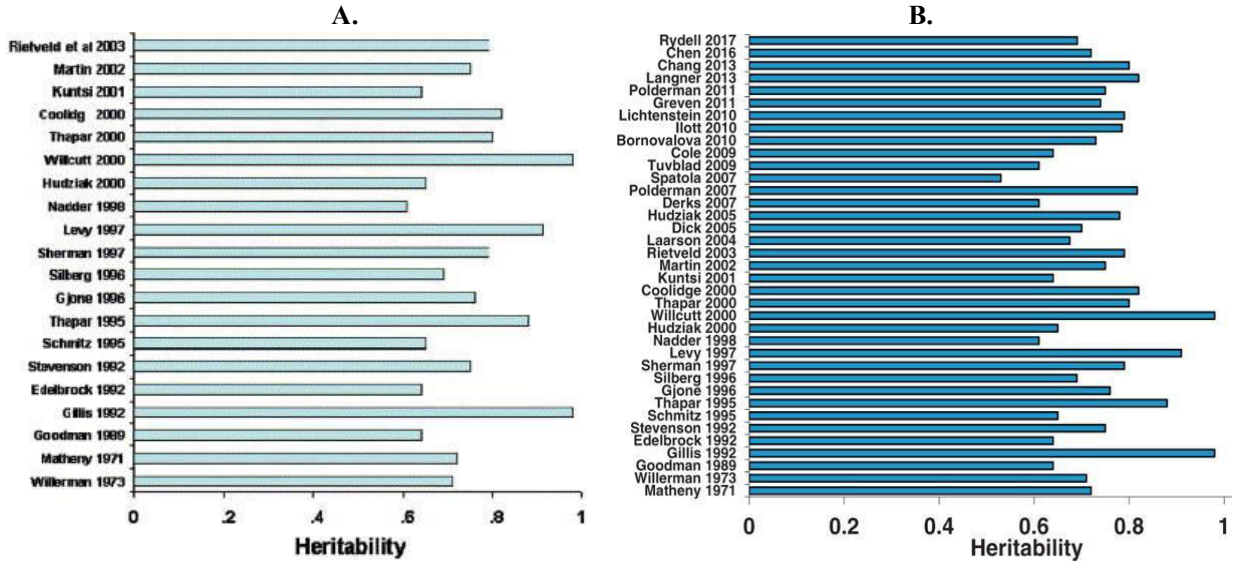


Figure 1. A: Estimated heritability of ADHD, based on pooled results from 20 twin studies. B: Heritability of ADHD from twin studies of ADHD diagnoses or symptom counts.

Adapted from Faraone *et al.*, 2005 and Faraone *et al.*, 2018.

Although several countries maintain twin registries, especially in Scandinavia and in the USA, finding an adequate number of twins can be challenging since human twin occurrences are not very common. Thus, an alternative approach estimates single nucleotide polymorphism (SNP)-based heritability (SNP-h2) through genome-wide complex trait analysis (GCTA). The starting point is a genome-wide association study (GWAS), where SNPs covering the whole genome at regular intervals are genotyped -typically- in a group of patients and a group of controls and pairwise genetic and phenotypic correlations are computed by employing linkage disequilibrium (LD) score regression models. The additive effect of common SNPs to the phenotype is a predictor of SNP-h2. In ADHD, the most recent estimation for SNP-h2 is 0.22 (standard error = 0.01)[1]. It is important to note that SNP-based heritability accounts for the genetic contribution to the trait of a certain type of variation (SNPs) with a certain frequency (greater than 1% or than 5%, depending on the study), neglecting the genetic effects of other variants (e.g. copy number variants -CNVs- or rare variants).

2.2 Is there an environmental basis for ADHD?

Pre-, peri- and postnatal risk factors may aggravate the development of ADHD symptoms: 1) Maternal characteristics during pregnancy are counted as prenatal influences, for instance-presence of illnesses, stressors, exposure to chemicals or substance use. 2) Perinatal factors involve birth-related indicators like type of delivery, hypoxia, child's birth weight, infections in child following birth, among others. 3) Exposures and experiences during early childhood build the postnatal environment.

Maternal exposure to cigarettes and alcohol prenatally may induce alterations in the cerebellum[30,31], a brain region monitoring some of the cognitive functions that are impaired in ADHD. It may also increase hyperactivity, impulsivity and disruptive behaviours in children, potentially exposing them to developing psychiatric conditions. Maternal smoking escalates the ADHD risk in children by twice[32] and a positive relation is reported between smoking doses and hyperactivity[33]. Other prenatal factors can lead to very low birth weight in children, which is associated with a two-fold increase in ADHD[34,35].

Perinatal events like complications during delivery are associated with increased incidence of ADHD symptoms in children[36]. Hypoxic occurrences adversely affect developing brains[37]. Postnatal risks are diverse but mostly reflected by the social conditioning of children and their dietary imbalances. For instance, a role for iron-deficient diets and unbalanced consumption of essential fatty acids (omega-3 and omega-6) have been suggested impact ADHD

development[38–40]. However, further evidences are needed to validate these speculations. Exposure to environmental toxins and social adversities like stressful social life at early stages may also interfere with the neurobehavioral progression and trigger ADHD symptomatology[41].

2.3 Quantifying gene-environment interactions

As several risk factors for ADHD have been identified in addition to the genetic ones, the models to study the etiology of ADHD have accordingly evolved and are complex. These models intend to quantify the interactions between genes and environmental exposures. For instance – It is known that an altered activity of nicotinic receptors disrupts dopaminergic function[42], a functional pathway known to be relevant to ADHD. So, it is enticing to speculate on the impact of prenatal smoking to ADHD risk on those subjects bearing genetic risk factors for ADHD in dopaminergic genes. An example model to estimate the ADHD outcome considers ADHD risk variants in *DRD4* -encoding a dopaminergic receptor- and *DAT1* -the dopamine transporter- together with the intensity of prenatal smoking[43]. Such models may also help to predict the severity of ADHD symptoms and also its clinical subtypes. Similarly, males homozygous for the 10-repeat allele of the 40-bp variable number of tandem repeats (VNTR) polymorphism in *DAT1* who grow up in a context of psychosocial adversity exhibit higher hyperactivity-impulsivity than non-homozygous *DAT1* males or those who grow up in less adverse conditions. Similarly, significant interactions of the 30-bp VNTR and *DAT1* haplotype with psychosocial adversity on ADHD symptoms have also surfaced[44].

CHAPTER 3. MOLECULAR GENETIC STUDIES

3.1 Genome-wide linkage scans

Seven independent genome-wide linkage scans have identified susceptibility loci for ADHD on chromosomes 5p13, 14q12 and 17p11[4]. However, the linkage signals on 5p and 17p could not be replicated in a later high-density SNP linkage scan and suggestive linkage signals were reported on chromosomes 16q23 and 9q22 in this same study[45]. Other linkage spots appeared on 4q13.2, 5q33.3, 8q11.23, 11q22, and 17p11[46] and 16p13[47]. A genome scan meta-analysis (GSMA) was conducted on all seven independent genome-wide linkage scans in ADHD that identified a genome-wide significant (GWS) linkage on chromosome 16 (16q22–16q24), and nine genomic regions showing nominal linkage[48]. Follow-up studies on these linkage signals are limited and only a few genes behave been highlighted. For instance - The 9q22 locus pointed at genetic variation in the promoter region of the brain-expressed *DIRAS2* gene, encoding a GTPase of the Ras family, as a risk factor for ADHD and impulsive disorders[49]. In addition, fine mapping of the genomic region on chromosome 4 found previously linked to ADHD narrowed down a segment encompassing exons 4 to 19 of the *LPHN3* gene, containing several functional domains and variants with a potential impact on splice isoform variability[50]. *LPHN3* encodes a G-protein coupled receptor involved in cell-to-cell adhesion.

3.2 Common genetic variation in ADHD

3.2.1 Candidate gene association studies/Hypothesis-driven studies

Research through this approach has focused on neurobiological pathways that are suspected to be involved in ADHD. Genes that belong to these pathways are proposed to be candidates for the disorder and subsequently examined in both cases and controls for differences in the frequency of genetic variants.

Most classic candidate gene studies in ADHD have tagged neurotransmitter systems, mainly dopamine and serotonin. The fact that the main pharmacological treatments for the disorder target dopaminergic receptors and the dopamine transporter, has motivated an extensive scrutiny of this system, with several associations with childhood or adult ADHD found at genetic loci containing *DRD4*, *DRD5*, *DAT1 5HTT, HTR1B,* and *SNAP25*[51].

However, an important limiting factor hampers the use of the candidate gene approach: the knowledge we have on the disease mechanisms is still scarce and fragmented. Also, the

reliability of this approach is limited, as the chances of obtaining false positive results are high, making it difficult to replicate the findings[3]. It is worth to note that none of the hypothesis-driven findings have emerged as Bonferroni significant ($P < 5 \times 10^{-8}$) in GWASs so far, with the only exception of *FOXP2*, a language-related gene that encodes a transcription factor and found associated with adult ADHD in a study published in 2012[52].

3.2.2 Genome-wide association studies / Hypothesis-free studies

Initial attempts on identifying genome-wide associations of SNPs with ADHD through GWAS have been performed on nine independent datasets[4]. Three of these GWASs were specific to the persistent form[53–55], and another one focused on conduct disorder with ADHD[56], and studied families. Several of these datasets were part of the first two meta-analyses in ADHD[57,58]. However, neither the individual GWASs nor the meta-analyzed associations resulted in genome-wide significant loci. Key functions of the top genetic loci from these ADHD GWASs are central nervous system development, neuronal progression involving differentiation and activity, neurite outgrowth, synaptic transmission, axon guidance, ligand binding e.g. to FGFR, and also the activation of ion channels[4].

The very first GWS hits in ADHD have been unraveled in a meta-analysis published in 2019 that comprises 20,183 ADHD cases and 35,191 controls from 12 datasets including both children and adults[1]. This study highlights 12 independent GWS loci containing 304 SNP variants. The well annotated genes include *FOXP2* (mentioned in the previous section, a transcription factor involved in learning disabilities and language), *SORCS3* (a brain-expressed receptor with a role in brain plasticity and neuronal development), *DUSP6* (with a role in dopamine homeostasis), *SEMA6D* (possibly involved in the inhibition of axon growth, associated with educational attainment) and *MEF2C* (a transcription factor associated with other brain conditions). However, there are additional genes (e.g. miR-3666) and intergenic variants located in these GWS loci[1] that are not well annotated, but they might be potentially relevant in understanding ADHD.

The latest findings from the GWAS and, especially, those from the previous linkage studies or the CGAS continue to present inconsistencies. Presumably, with studies that utilize an even greater sample size, we expect to achieve replications and more confidence in the findings. Indeed, a preliminary GWAS that is an extension of the one published in 2019[1] and includes more than 100,000 ADHD cases and over 120,000 controls was presented at the World

Congress of Psychiatric Genetics 2019 and has not been published yet. This study released around 100 independent GWS hits.

3.3 Rare genetic variation and structural variants in ADHD

The contribution of rare genetic variants (minor allele frequency (MAF) <1%) and of structural variants, also known as copy-number variants or CNVs (involving DNA segments >1 Kb) to ADHD heritability can be weighted through whole-genome, whole-exome or targeted sequencing, but also by using genotyping arrays and other methodologies. Rare mutations (either-single nucleotide variants -SNVs- or CNVs) and high-frequency CNVs are suspected to explain a considerable fraction of the so-called missing heritability of ADHD. Those genetic effects would add to those explained by common single-nucleotide variation (SNPs), estimated to be around 20% for this disorder.

Large rare CNVs (> 100 kb or > 500 kb) are at a greater burden in both child and adult ADHD patients[59–61]. Many of the risk CNVs are found in genomic regions related to neurodevelopmental processes and are shared with other brain disorders, like intellectual disability (ID), autism and schizophrenia. Individuals carrying rare risk CNVs may require a lesser load of common risk variants to develop ADHD[59].

Rare SNVs in psychiatric disorders have been studied mainly in autism[62] and schizophrenia[63] through whole-exome (WES) or whole-genome sequencing (WGS). A recent exome analysis on a large number of ADHD and ASD cases has ascertained a significantly greater burden of rare protein-truncating variants, and associated *MAP1A*, with both disorders[64]. A WES study explored a prioritized set of 52 candidate risk genes in ADHD and found that rare missense and disruptive variants in these genes were more than twice as prevalent in patients with persistent ADHD compared to controls[65]. A combined linkage analysis and WES approach identified 38 rare variants within 25 genes where these genes altogether were significantly associated with persistent ADHD. The *AAED1* gene that can possibly regulate DAT trafficking through PICK1 binding emerged as gene-wide significant, and a rare variant in *AAED1* (rs151326868) segregated with ADHD[66]. Also, an exome-wide scan of rare coding variants for adult ADHD revealed four significant candidate loci at 6q22.1, where *NT5DC1* and *COL10A1* reside, along with the *SEC23IP*, *PSD* and *ZCCHC4* loci[67]. Putative functional rare SNVs associated with hyperactivity and inattention have been detected in the context of one of the major common variants of *DRD4*: a repeat of a 48-bp unit in exon 3 of the gene (*DRD4*-7R). It is worth speculating whether these rare variants in *DRD4* can be a stronger

contributor to ADHD symptoms than the classically investigated *DRD4* common variation[68]. The brain-derived neurotropic factor gene *BDNF,* previously related to impulsive symptoms, is enriched in putatively functional, rare SNVs[69]. An investigation on the etiology of sporadic ADHD highlighted six brain-expressed genes as candidates for the disorder, based on the load of *de novo* missense SNVs[70].

3.4 Epigenetics and ADHD

Epigenetic modulators dynamically regulate the expression of many genes, including those that control the neural cell phenotype and brain function. Thus, modifications of DNA methylation, histone code, nucleosome positioning, and noncoding RNA-based mechanisms are recognized epigenetic regulators, all of them relevant to the identification of disease mechanisms. Each epigenetic modification can contribute multiplicatively to the disease risk[71]. Given the phenomenon of transgenerational epigenetic inheritance[71], a better scrutiny of the epigenetic landscape may help to elucidate a proportion of the "missing heritability" that accompanies many common and complex human diseases[72]. The first epigenetic studies in ADHD have already established links between the epigenetic modification of genes and ADHD etiology.

3.4.1 MicroRNAs (miRNAs, miRs) and ADHD

3.4.1.1 MicroRNAs

MiRNAs are 19-21bp long non-coding RNAs (ncRNAs) and have been extensively studied. They bind to target mRNAs which might lead to either complete degradation of the mRNA or to an alteration in its translation to protein. A single miRNA can target one or more mRNAs, and many miRNAs can simultaneously target a single mRNA molecule[73]. This therefore constitutes a critical form of posttranscriptional regulation of gene expression that involves around 60% of all the protein-coding genes in the human genome[74].

The precise identification of miRNAs and their probable targets is not an easy task. The database miRBase v.21 (http://www.miRBase.org), released in 2014, reports the latest figures on identified human precursor and mature miRNAs in humans, mice, and other species[75]. In humans, 1,881 precursor miRNAs have been identified, resulting in the generation of 2,588 mature miRNAs.

Investigating the mechanisms of action of miRNAs is of upmost importance to our understanding of the regulation of the cellular function. The human brain contains 70% of all known miRNAs, making miRNAs a crucial component in neuropsychiatry[76,77]. MiRNAs

define the overall multifold molecular, structural and functional development of the nervous system[73]. Therefore, disentangling the functionality of miRNAs in brain may specifically trace the underlying neuropsychiatric pathways (Figure 2).

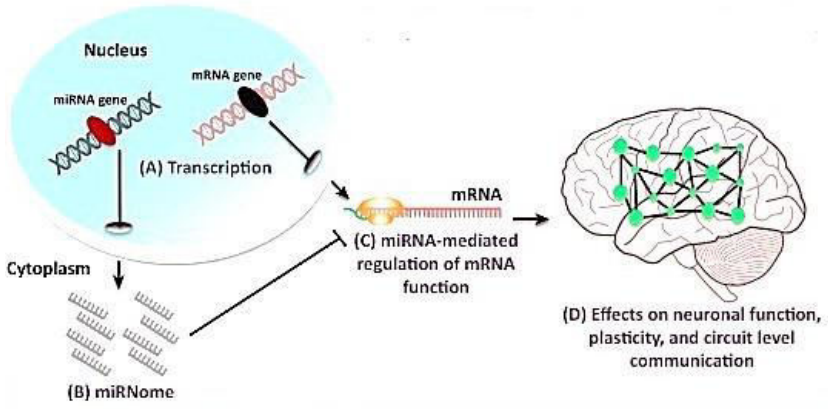


Figure 2. The effect of the miRNome on brain functions.

Adapted from O'Connor et al., 2016.

One of the approaches used involves comparing the mRNA transcriptome and the miRNA levels from brain biopsies of patients against those of controls. However, because of the invasive nature of brain biopsies, this technique has ethical issues and necropsies from post-mortem brains are used instead. An alternate minimally invasive approach might be to study the levels of cell-free circulating miRNAs in body fluids like saliva, serum, urine, and cerebrospinal fluid (CSF). MiRNA profiles are highly stable in nearly all fluids, as these molecules are resistant to RNAse degradation in bloodstream[78,79].

Human plasma carries nearly 10% of all known human miRNAs, including mirtrons, unconventional precursors to miRNA biogenesis pathways[79]. However, how these circulating miRNAs contribute to the normal and/or altered physiology, remains unclear. Expression profiling of miRNAs can be valuable in estimating the risk or progression of diseases.

3.4.1.2 MiRNAs and ADHD

In order to evaluate circulating miRNAs as biomarkers in neurodevelopmental conditions, studies have been attempted in individuals with ADHD, ASD, schizophrenia, anxiety disorder, bipolar disorder and Tourette syndrome[80,81]. Regarding ADHD, we know that several miRNAs modulate the expression of genes that have been linked to the disorder, for instance *BDNF*, which controls neuronal activity, and *DAT1*, *HTR2C*, *HTR1B* and *SNAP-25*, involved in neurotransmitter mediation[82]. Also, the levels of several miRNAs have been found altered in peripheral tissues of ADHD patients and in animal models[82].

Deregulated miRNAs in ADHD

Five dysregulated miRNAs have been reported in serum samples with significant predictive values to discriminate between the ADHD and control groups (upregulated: hsa-miR-101-3p, hsa-miR-130a-3p, hsa-miR-138-5p and hsa-miR-195-5p and downregulated hsa-miR-106b-5p)[83]. miR let-7 is most extensively addressed for its role in neuronal adaptations and neurodegeneration[84]. Reports of miR let-7 in the brains of an ADHD rat model (SHR, Spontaneously Hypertensive Rat) have confirmed its elevated levels and its role as a downstream regulator of tyrosine hydroxylase, a critical molecule to dopamine metabolism[85]. The glucocorticoid receptor Nr3c1 is known to undergo a complex miRNA-Bhlhb2 regulation. In the prefrontal cortex (PFC) of this ADHD animal model, Nr3c1 inhibits the expression of miR-296, 34c, 138 and 138, which in turn target *Bhlhb2*, encoding a transcription factor, that becomes overexpressed. Targeted silencing of *Bhlhb2* significantly improves the hyperactivity behavior in the SHRs[86]. In the most recent attempts to establish circulating blood miRNAs as ADHD biomarkers, 13 miRNAs have been reported using next-generation sequencing (NGS)[87], and also miR-26b-5p, miR-185-5p, and miR-191-5p in a genome-wide miRNA expression analysis[88].

Genetic variation in the miRNA machinery in ADHD

Presence of polymorphisms in miRNA sequences or their target sites may disrupt the binding strength of miRNAs-mRNAs[89]. Moreover, variation in regions involved in the regulation of miRNA expression may also be functionally relevant. These variants may be located in *cis* or in *trans* with respect to the miRNA gene which expression is affected. Two ADHD-associated SNPs are located in the 3′UTR of the *SNAP-25* gene, which is also a predicted binding region for miR-510 and miR-641[90,91]. A SNP in the pri-miR-34b/c locus has been associated with ADHD, and their mature forms miR-34b and miR-34c are differentially expressed in the blood of ADHD subjects. This pri-miR loci targets ADHD-associated genes, either validated (*NOTCH2, HMGA2)* or just predicted targets (*HTR2C* and *VAMP2)*[92]. The 3'UTR of *DAT1,* encoding the dopamine transporter, has a 40bp-VNTR that contains binding sites for four miRNAs: mir-1972, miR-30b-5p, miR-1301 and miR-6070[93]. Interestingly, two SNPs located downstream from miR-96 in the miR-183–96–182 cluster have been associated with ADHD without substance use disoders[94]. The members of this cluster may target the serotonin receptor gene *HTR1B* and also *RARG*, implicated in the control of the dopamine signaling pathway[95]. Finally, a study explored the contribution of genetic variants in the miRNA biogenesis machinery, i.e. in genes and found a suggestive association between a SNP in the *AGO1* gene

and ADHD[96]. The authors claim for innovation in the diagnostic strategies in ADHD, in the sense that there is a need to interrogate also the 'regulators of the regulators' (Figure 3).

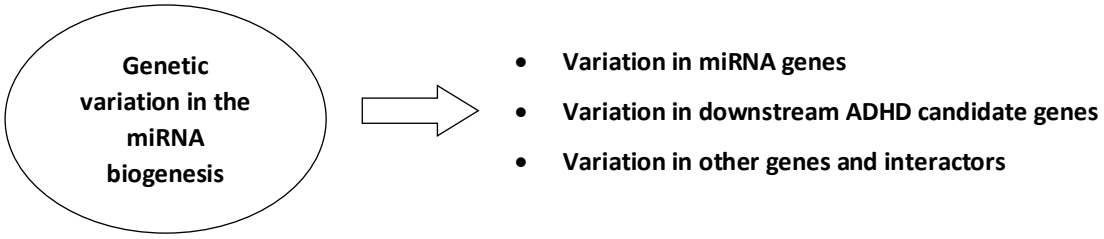


Figure 3. Novel candidates for exploration of genetic risk factors in ADHD: miRNA biogenesis. Adapted from Karakas *et al.*, 2017.

3.4.2 DNA methylation and Allele-Specific methylation (ASM)

The methylation of cytosine molecules in CpG dinucleotides through a covalent modification is known as DNA methylation. It is a stable epigenetic modification with a direct role in defining chromatin-histone patterns, and in the processes of X-chromosome inactivation, genomic imprinting, gametogenesis, embryogenesis and silencing of repetitive DNA elements.

Research on DNA methylation patterns in candidate genes for distinct disorders have allowed identification of aberrant patterns of methylation in several genomic regions and tissues. Often, these patterns are tissue- or even cell line-specific. In addition, allele-specific DNA methylation can lead to allele-specific gene expression (ASE)[97].

Multiple studies have tested whether the patterns of DNA methylation of certain candidate genes for a disorder can be synchronous with the symptoms of the disorder. For instance, a study conducted on newborns has reported negative associations between childhood ADHD symptomology and the methylation of genomic regions encompassing *DRD4* and *5-HTT*[98]. In some instances, the co-occurring conditions may weaken or enhance the associations between the DNA methylation levels and ADHD scores[98].

ASM is a prevalent epigenetic mechanism across the genome where different alleles at a polymorphic site can skew the patterns of DNA methylation. The most widespread is the *cis*-effect ASM, where a local genotype is concomitant with allelic DNA methylation on the same DNA molecule. On the contrary, a minority of ASM events are non-*cis*, resulting from trans-acting elements. Even though ASM is cell-type specific, it occurs on 23%~37% heterozygous SNPs in any given cell line[97]. The heterozygous SNPs in the regions of CpG dinucleotides can

account for up to 80% of ASM regions[97]. In consequence, the methylation potential of these CpG dinucleotides is perturbed, which alters the expression of the regulated genes. Deciphering the ASM variants and integrating this information with GWAS data is a valid approach to mine the functional connections from variants to phenotype[99].

3.4.3 Challenges in epigenetic research

Epigenetics research is a promising way to digest the chunks of available information and to infer the most out of GWAS datasets[100]. Efforts in establishing abnormal epigenetic marks in the genome of ADHD patients are subject to some practical limitations. To begin with, a major drawback is the difficult access to human brain tissues[101]. Second, the epigenetic patterning is tissue or even cell-specific, which may limit reproducibility. Third, the human epigenome still lacks the rich annotation as of the human genome, for instance, (i) an accurate estimation of human miRNA genes, (ii) validated targets of known miRNAs or (iii) miRNA genes still poorly explored in terms of their potential as quantitative trait loci (QTL). Nonetheless, owing to the causal relevance of epigenetic research, alternate methods like the use of proxy tissues and newer molecular techniques like miRNA-Seq continue to ensure its feasibility[101].

3.5 Neuroimaging profiles and genetics in ADHD

ADHD has been postulated as a disorder of impaired neurocognition. The identification of neuroanatomical changes in ADHD is thus fundamental to ADHD research. Neuroimaging scans recognize abnormalities in brain anatomy and function in individuals with ADHD. For instance, smaller volumes and abnormal surface morphologies in basal ganglia have been associated with ADHD[102,103]. In addition to these static changes, brain structures may undergo delayed maturation development, and these developmental patterns correlate with the severity and subtypes of ADHD trajectories[104]. The enrichment of neurodevelopmental genes found in the association studies[29] suggest that these genes may underlie the impaired brain structures and function revealed by neuroimaging studies. Verification of this hypothesis may concatenate the functionality of the highlighted genetic loci and the brain phenotypes.

CHAPTER 4. COMORBID PROFILES IN ADHD

More than 10 psychiatric and behavioral phenotypes are known to be comorbid with ADHD in both children and adults. The various presentations of ADHD, hyperactive-impulsive, inattentive and combined, differ in the comorbidity profile [10,105]. Of these three presentations, the combined subtype presents a greater risk of developing other psychiatric conditions[105]. Thus, psychiatric comorbidity can be a marker of more severe ADHD as reflected by the combined subtype symptoms[105]. Several of the comorbid disorders appear to be more strongly associated with ADHD in females[106].

4.1 Autism Spectrum Disorders (ASD)

ASD can co-occur with ADHD in up to 50% of the children with ADHD and these disorders share impairments in developmental and cognitive domains. For instance, patterns of social and empathy deficits that directly affect the ability to build peer relationships are found in individuals having ADHD and ASD[107]. More is known about the ADHD-ASD comorbidity in children than in adults[4]. Children with ADHD and ASD present augmented ADHD symptoms along with worsened executive functions (e.g. verbal working memory) when compared to those with ASD alone[4,107,108]. It has also been reported that individuals with the 22q11 deletion syndrome present with combined ADHD and also ASD, and they may suffer behavioral and/or learning problems[109].

4.2 Intellectual disability (ID) and learning disabilities (LgD)

ID is also 5-10 times more prevalent in children with ADHD than in those without ADHD[4]. Subjects with both ADHD and ID exhibit increased diagnoses of ODD and CD[110]. Learning disorders are characterized by difficulties in reading, writing and arithmetical skills, which are experienced by 25–40% of the individuals with ADHD[4]. Such difficulties are presumably driven more by inattentive symptoms in ADHD than by hyperactivity/impulsivity. The prevalence estimates of LgD with ADHD and vice versa vary considerably given the heterogeneous LgD identification criteria[111]. ADHD+LgD females are also at greater risk to experience cognitive depression[112].

4.3 Tic disorders (TDs)

TDs are neurodevelopmental disorders characterized by an onset of multiple episodes of motor and/or vocal tics. TDs include Tourette's disorder, chronic tic disorder (CTD) and transient tic disorder[113]. About 20% of children with ADHD also suffer from TD[114] and the likeliness of developing CTD peaks between the ages of 7 to 10 years[115]. Compared to non-ADHD children, the incidence of CTD in ADHD kids is nearly three to four times[115]. A Tourette syndrome (TS)-based study on adults and children could also diagnose ADHD in 60% of the TS patients, making ADHD the most common comorbidity in TS, together with learning disabilities, social skill deficits and anger control issues. ADHD+TD may aggregate in families as ADHD and TD do separately.

4.4 Aggression and rule-breaking behaviors

Aggression has been repeatedly reported to co-occur with ADHD[116], and this correlation becomes stronger at older ages. Aggressive behavior results in rule-breaking at later stages[117]. Direct associations have also surfaced between ADHD and rule-breaking behaviors. In legal systems, offenders with ADHD tend to re-offend twice than the non-ADHD offenders, leading to increased chances of legal confinement[118]. Externalizing behaviors of aggression and rule breaking, to some extent, are manifestations of emotion dysregulation (ED) typical of ADHD[119].

4.5 Mood Disorders

Having ADHD may elevate the risk of developing bipolar disorder by 10% in children and adolescents and other mood disorders by up to 40%[120,121]. The occurrence of bipolar disorder in ADHD can range between 5 and 47%[122]. Lifetime ADHD is more comorbid with bipolar disorder type I than with type II[120,123]. Restlessness, talkativeness, distractibility and fidgeting are the usual symptoms in cases with both manic bipolar disorder and ADHD[9]. The prevalence rates of comorbid depression in ADHD individuals range between 18 to 53%, and nearly 14% of the children with ADHD may exhibit clinical depression[9,124]. ADHD is a common comorbidity in individuals with bipolar disorder or MDD, with a higher comorbidity rate in the former group[123]. Individuals with both ADHD and bipolar disorder have an earlier age of onset of the mood disorder of around 5 years[120].

4.6 Substance use disorders (SUDs)

SUD is one of the most common comorbid conditions of ADHD, particularly alcohol and/or nicotine, cannabis, cocaine use. Substance abuse and dependence are nearly twice more prevalent in individuals with ADHD[9]. Subjects with persistent ADHD are 4.6 times more likely to develop SUD, when compared to controls in a major longitudinal study[125]. More severe physical dependency for nicotine has been established in ADHD[9]. While various neurobiological or psychosocial factors trigger the increased incidence of SUDs in individuals with ADHD, novelty-seeking behavior and higher impulsivity in this cohort remain major drivers towards the initial contact with the drug as of self-medication[9].

The use of stimulant medication like methylphenidate in children can postpone the onset of substance abuse (cigarette smoking, drug and alcohol misuse) at later ages[9,125,126]. However, prescribing stimulant medications to high-risk ADHD-SUD groups can be controversial because of the increased risk of substance misuse and diversion among this group[126,127]. The presence of ADHD in individuals with SUD may lower the onset age of substance abuse and rates of treatment obedience while increasing the odds of developing self-harms and multiple substance abuse.

4.7 Sleep disorders

Sleep disorders in ADHD children are characterized by significant and consistent disturbances in the patterns of sleep initiation and sleep duration. Irregular circadian rhythms and melatonin production are believed to underlie to the occurrence of sleep disorders. Nearly 25-50% of ADHD patients experience sleep troubles, and worsened ADHD symptoms are reported in individuals with sleep deficiencies. The latter group is reported to show increased daytime and cognitive impairments. Circadian rhythm sleep disorder, insomnia, narcolepsy, restless leg syndrome, sleep-disordered breathing are primary forms of comorbid sleep disorders[128].

4.8 Obesity and food addiction

Obesity or overweight is a comorbid condition that presents in both child and adult ADHD, more often reported in males[10]. The percentage of reported obesity in individuals with ADHD can be as high as 50% and it has significantly augmented since the first recorded ADHD-obesity relationship[129,130]. Impulsivity and inattention lead to abnormal eating patterns which results in weight gain[131]. Whether the inattentive or the HI ADHD group show a higher

prevalence of obesity is still unclear[10,132]. ADHD is associated with binge eating and food addiction, and this is more evident in adults than in children[133]. The association between obesity and ADHD is stated to be bidirectional and attributed to shared underlying neurobiological mechanisms[131]. However, the role of 'reward-deficiency syndrome' due to altered dopaminergic receptors has been documented in addictions and in ADHD[132,134].

4.9 Conduct Disorder (CD) and Personality Disorder (PD)

Behavioral disruptions if not resolved can perpetuate into personality disorders. 37% of adults with ADHD have one PD, and 27% have two or more PDs, specifically in cluster C and B PD[135,136]. Increased levels of ED and symptoms of ODD are also present in subjects with PD+ADHD[136]. Higher number of PD symptoms are perceived in adolescents with ADHD[137], more frequently in girls than in boys[137]. The HI ADHD group shows a higher incidence of CD in childhood and antisocial personality traits in late adolescence[10]. The prevalence of borderline personality disorder (BPD) in ADHD individuals can range between 10-45% and lifetime comorbidity of BPD with ADHD is around 33%[138,139]. BPD reflects patterns of unstable identity and interpersonal relationships, pronounced impulsivity and ED[140]. Conducts typically associated with CD and PD like kleptomania or substance abuse can have criminal, judicial and financial repercussions. Therefore, these comorbidities are detrimental to the social environment in addition to the life of sufferers[10].

OBJECTIVES

The broad objectives of this work are the identification of genes that contribute to the susceptibility to attention-deficit/hyperactivity disorder (ADHD) and cocaine dependence, two disorders that co-occur in patients. We propose to focus on epigenetic risk factors (allele-specific methylation and miRNAs) in ADHD, and to scrutinize the genetic basis of cocaine dependence. We also aim at exploring the common genetic basis that explains the comorbidity between these disorders.

The specific aims are outlined below:

Chapter 1. Exploring genetic variation that influences brain methylation in ADHD

1.1 Assessment of the contribution to ADHD of allele-specific methylation (ASM), an epigenetic mechanism that involves SNPs correlating with differential levels of DNA methylation at CpG sites.

1.2 Assessment of the possible effects of identified ASM variants on gene expression and on brain volumes to identify new genes contributing to ADHD.

Chapter 2. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions

2.1. Investigation of the variants that underlie cocaine-dependence by meta-analyzing available genome-wide association study (GWAS) datasets.

2.2. Investigation of shared genetic risk factors between cocaine dependence and its comorbid conditions, including ADHD.

Chapter 3. Exploring the impact of common variation in micro-RNA genes in attention-deficit/hyperactivity disorder

3.1. Assessment of the contribution to ADHD of variation in miRNA genes and their putative regulatory elements through case-control association studies.

3.2. Deciphering of the miRNA-mediated pathways that regulate the expression of genes potentially causal in ADHD.

RESULTS

REPORT FROM THE THESIS SUPERVISORS ON THE CONTRIBUTION OF THE PhD STUDENT TO THE PUBLICATIONS OF THIS DOCTORAL THESIS

Thesis title: "Genetics and epigenetics of attention-deficit/hyperactivity disorder and comorbid conditions"

Author: Anu Shivalikanjli

Supervisors: Dr. Bru Cormand i Rifà, Dr. Stephen V. Faraone

CHAPTER 1: Exploring genetic variation that influences brain methylation in attention-deficit/hyperactivity disorder

Article 1

Pineda-Cirera L*, Shivalikanjli A*, Cabana-Domínguez J, Demontis D, Rajagopal VM, Børglum A, Faraone SV, Cormand B, Fernàndez-Castillo N. Exploring Genetic Variation that Influences Brain Methylation in Attention-Deficit/Hyperactivity Disorder. *Transl Psychiatry* **9,** 242 (2019). https://doi.org/10.1038/s41398-019-0574-7

* Equally contributed

IF_{2018}: 5.182, Q1 (18/146, PSYCHIATRY)

Contribution of the autor to the article: Accessing the ADHD GWAS and brain volumes summary statistics, inspection of the variants in the summary data. Carrying out the statistical analyses. Participation in the functional annotations for the identified associations using bioinformatic methods. Preparing the first draft of the manuscript, and tables, and participation in the final edits.

CHAPTER 2: Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions

Article 2

Cabana-Domínguez J, Shivalikanjli A, Fernàndez-Castillo N, Cormand B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Prog Neuropsychopharmacol Biol Psychiatry*. 2019 Aug 30;94:109667. doi: 10.1016/j.pnpbp.2019.109667.

IF_{2018}: 4.315, Q1 (38/199 CLINICAL NEUROLOGY; 41/267 PHARMACOLOGY & PHARMACY)

Contribution of the author to the article: Accessing and cleaning of datasets. High-computing data and pipeline management. Participated in the quality control, association analysis, generation of plots. Participated in the final edits of paper.

CHAPTER 3: Exploring the impact of common variation in micro-RNA genes in attention-deficit/hyperactivity disorder

Article 3

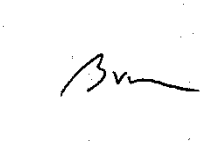
Shivalikanjli A*, Reinbold C*, Cabana-Domínguez J, Stangier S, Demontis D, Borglum A, Faraone SV, Nöthen MM, Mattheisen M, Fernàndez-Castillo N, Forstner A, Cormand B. Exploring the impact of common variation in microRNA genes in attention-deficit/hyperactivity disorder. *Manuscript in preparation*.

*Equally contributed

Contribution of the author to the article: Selection of genome-wide miRNA genes. Retrieval of genomic coordinates for the miRNAs and their regulatory regions. Liftover of the genomic coordinates and resolving the liftover inconsistencies through manual inspection. Performing the tagSNP selection. Mining the variants in the GWAS summary statistics for SNP-based analysis. Bioinformatic follow-up of highlighted associations. Preparing the first draft of the manuscript, figures and tables, and participating in the final edits.

Barcelona, December 10, 2020

Signature of supervisors



Dr. Bru Cormand Rifà

Dr. Stephen V. Faraone

SUMMARY ARTICLE 1

"Explorant la variació genètica amb efectes sobre la metilació en cervell en el dèficit d'atenció amb hiperactivitat

El trastorn per dèficit d'atenció i hiperactivitat (TDAH) és un trastorn del neurodesenvolupament causat per la interacció entre factors genètics i ambientals. L'epigenètica és crucial en la generació de canvis persistents que afecten l'expressió gènica al cervell. Estudis recents suggereixen que la metilació de l'ADN tindria un paper clau en el TDAH. Hem explorat la contribució al TDAH de la metilació específica d'al·lel (ASM), un mecanisme epigenètic pel qual variacions de tipus SNP presenten correlació amb nivells diferencials de metilació de l'ADN en llocs CpG. Hem seleccionat 3.896 tagSNPs que sabem que tenen una influencia sobre la metilació al cervell i hem realitzat un estudi d'associació de tipus cas-control utilitzant els resultats de la meta-anàlisi GWAS més gran realitzada en TDAH, que comprèn 20.183 casos i 35.191 controls. Hem observat un enriquiment de variants genètiques que confereixen risc al TDAH en el conjunt de SNPs de tipus ASM, i hem identificat associacions significatives amb vuit tagSNPs (FDR=5%). Aquests SNPs presenten correlació amb la metilació de llocs CpG situats a les regions promotores de sis gens. Atès que la metilació pot afectar l'expressió gènica, hem investigat si aquests SNPs, juntament amb 52 SNPs en alt desequilibri de lligament, són eQTLs en teixits cerebrals i hem observat que tenen un impacte sobre l'expressió de tres d'aquests gens. Els al·lels de risc al TDAH presenten correlació amb una major expressió (i disminució de la metilació) d'*ARTN* i *PIDD1* i amb una disminució de l'expressió (i augment de la metilació) de *C2orf82*. A més, hem predit que aquests tres gens tindrien una expressió alterada en pacients amb TDAH, i variants genètiques a *C2orf82* presenten correlació amb el volum de determinades regions del cervell. En resum, hem seguit una estratègia sistemàtica per identificar variants de risc al TDAH que correlacionen amb la *cis*-metilació diferencial, tot identificant tres nous gens que contribueixen al trastorn.

Reference

Pineda-Cirera L*, Shivalikanjli A*, Cabana-Domínguez J, Demontis D, Rajagopal VM, Børglum A, Faraone SV, Cormand B, Fernàndez-Castillo N. Exploring Genetic Variation that Influences Brain Methylation in Attention-Deficit/Hyperactivity Disorder. *Transl Psychiatry*. 2019 Oct 3;9(1):242. doi: 10.1038/s41398-019-0574-7.

* Equally contributed

Translational Psychiatry

ARTICLE Open Access

Exploring genetic variation that influences brain methylation in attention-deficit/hyperactivity disorder

Laura Pineda-Cirera[1,2,3], Anu Shivalikanjli[1,3,4], Judit Cabana-Domínguez[1,2,3,4], Ditte Demontis[5,6,7], Veera M. Rajagopal[5,6,7], Anders D. Børglum[5,6,7], Stephen V. Faraone[8], Bru Cormand[1,2,3,4] and Noèlia Fernàndez-Castillo[1,2,3,4]

Abstract

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder caused by an interplay of genetic and environmental factors. Epigenetics is crucial to lasting changes in gene expression in the brain. Recent studies suggest a role for DNA methylation in ADHD. We explored the contribution to ADHD of allele-specific methylation (ASM), an epigenetic mechanism that involves SNPs correlating with differential levels of DNA methylation at CpG sites. We selected 3896 tagSNPs reported to influence methylation in human brain regions and performed a case-control association study using the summary statistics from the largest GWAS meta-analysis of ADHD, comprising 20,183 cases and 35,191 controls. We observed that genetic risk variants for ADHD are enriched in ASM SNPs and identified associations with eight tagSNPs that were significant at a 5% false discovery rate (FDR). These SNPs correlated with methylation of CpG sites lying in the promoter regions of six genes. Since methylation may affect gene expression, we inspected these ASM SNPs together with 52 ASM SNPs in high LD with them for eQTLs in brain tissues and observed that the expression of three of those genes was affected by them. ADHD risk alleles correlated with increased expression (and decreased methylation) of *ARTN* and *PIDD1* and with a decreased expression (and increased methylation) of *C2orf82*. Furthermore, these three genes were predicted to have altered expression in ADHD, and genetic variants in *C2orf82* correlated with brain volumes. In summary, we followed a systematic approach to identify risk variants for ADHD that correlated with differential *cis*-methylation, identifying three novel genes contributing to the disorder.

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder with a worldwide prevalence of around 5%[1]. Its main symptoms include inattention and/or hyperactivity-impulsivity (DSM-V)[2]. ADHD is among the most heritable psychiatric disorders, with about 76% of its etiology accounted by genetic risk factors[3] and with single-nucleotide polymorphisms (SNPs) explaining around 22% of the phenotypic variance[4]. Furthermore, there is molecular evidence of shared genetic risk factors across many psychiatric disorders[5]. In ADHD, a recent genome-wide association study (GWAS) meta-analysis of 12 sample groups unraveled some of the specific genetic underpinnings of this polygenic disorder for the first time[4]. One of the challenges of GWAS is to establish the causal relationship between the associated genetic variants, especially those

Correspondence: Bru Cormand (bcormand@ub.edu) or Noèlia Fernàndez-Castillo (noefernandez@ub.edu)
[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain
[2]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain
Full list of author information is available at the end of the article
These authors contributed equally: Laura Pineda-Cirera, Anu Shivalikanjli
These authors jointly supervised this work: Bru Cormand, Noèlia Fernàndez-Castillo

located outside genes, and the disorder. In this regard, the use of epigenetic information can improve the interpretation of functionality of non-coding genetic variation[6]. In addition, some studies have hypothesized the importance of sub-threshold variants derived from GWAS[7,8], particularly those located in enhancer regions, with a potential impact on gene regulation[9,10].

DNA methylation is one of the most stable epigenetic mechanisms, involving mainly cytosines of CpG dinucleotides. This mechanism plays an important role in the regulation of neurogenesis, differentiation, and brain development[11]. Furthermore, epigenetic alterations have been hypothesized to contribute to neurodevelopmental disorders[12], including ADHD[13], autism spectrum disorders (ASD)[14,15], or borderline personality disorder[16].

DNA methylation can be complementary if it involves both alleles, or non-complementary when it affects only one allele, as in chromosome X inactivation in females or allele-specific methylation (ASM)[6]. ASM is a common mechanism by which single nucleotide variants determine differential methylation levels of CpG sites. ASM can alter promoter activity, leading to allele-specific expression[17] in combination with other still quite unknown factors, such as environmental effects[6]. It is quantitative and heterogeneous across tissues and individuals[6]. The environment affects DNA methylation leading to changes in gene regulation, although the underlying mechanism is still not well understood[18]. It has been suggested that, during embryonic development, ASM regions could be especially sensitive to environmental effects[6]. Investigating SNPs that display ASM could help to identify risk variants for common diseases, including neuropsychiatric disorders[19], as shown by recent studies of bipolar disorder (BD) and schizophrenia[10,20].

The present study investigated the possible contribution of ASM to ADHD using data from the largest GWAS meta-analysis performed to date in ADHD[4]. We also assessed its possible effect on gene expression and on brain volumes to identify new genes contributing to the disorder.

Materials and methods

Selection of ASM SNPs

SNP selection was made based on the results of two previous studies[21,22], which identified ASM variants in multiple brain regions of post-mortem human samples. Gibbs et al.[21], considered four brain regions (cerebellum, frontal cortex, caudal pons, and temporal cortex) of 150 subjects and Zhang et al.[22], used only the cerebellum of 153 subjects. Gibbs et al.[21], unlike Zhang et al.[22], excluded those sequences of probes with significant correlation with methylation that contained polymorphisms. To discard possible artifacts in our results, we checked and confirmed that none of the probes used to detect the six highlighted CpG sites target genomic regions with SNP variants. The genotyping platforms used in the two studies were different (Gibbs et al.[21] used Infinium HumanHap550 Beadchips and Zhang et al.[22] used Affymetrix GeneChip Mapping 5.0K Array). Both studies evaluated DNA methylation using the HumanMethylation27 Beadchips, and performed linear regression analyses by PLINK[23] to determine the correlation between each SNP and methylation of any CpG site[21,22]. Zhang et al.[22], unlike Gibbs et al.[21] applied quantile normalization to the residuals prior to the linear regression analyses.

In the study by Zhang et al.[22], a total of 12,117 SNP–CpG pairs associations were reported in cerebellum, and Gibbs et al.[21] listed a total of 12,135 SNP–CpG pairs in frontal cortex, 11,374 in caudal pons, 16,734 in temporal cortex, and 12,102 in cerebellum (Fig. 1). We combined the information from both studies and obtained a total of 43,132 SNP–CpG pairs involving 33,944 different SNPs and 5306 CpG sites (Fig. 1). We considered all the ASM SNPs from all the tissues in the two studies, as there are multiple SNP–CpG pairs in common between them (Fig. S1).

We subsequently applied different filters to generate a sub-list of 3896 SNPs (Figs. 1 and S2) out of these 33,944 variants to minimize redundancy: associations in *cis* between the SNP and the CpG site, correlation of the SNP with methylation levels of the CpG (R^2) ≥ 0.2, as performed in Gibbs et al. (2010)[21]. We considered only autosomal SNPs and selected tagSNPs for each CpG site ($r^2 \geq 0.85$), by assessing linkage disequilibrium (LD) with Haploview software[24] using the Central European (CEU) reference panel from 1000 Genomes Project Phase 3[25].

Case-control GWAS datasets

We explored the selected ASM SNPs in the summary statistics from a meta-analysis of 11 independent GWAS of ADHD conducted by the Psychiatric Genomics Consortium (PGC) and iPSYCH. This case-control study investigated 8,047,420 markers in 20,183 cases and 35,191 controls from Europe, USA, Canada, and China, with patients diagnosed according to the criteria detailed in Demontis et al. (2019)[4].

Statistical analysis

To test whether risk variants for ADHD are enriched in ASM SNPs, we carried out an enrichment analysis using the Fisher's exact test in R[26] at *p*-value thresholds ranging from 5E−02 to 5E−08 considering the total number of ASM SNPs available from the ADHD GWAS meta-analysis[4] (32,884 out of 33,944 SNPs).

From our selection of 3896 ASM tagSNPs, we could retrieve information on the association with ADHD of 3771 SNPs (96.8%) that were present in the summary

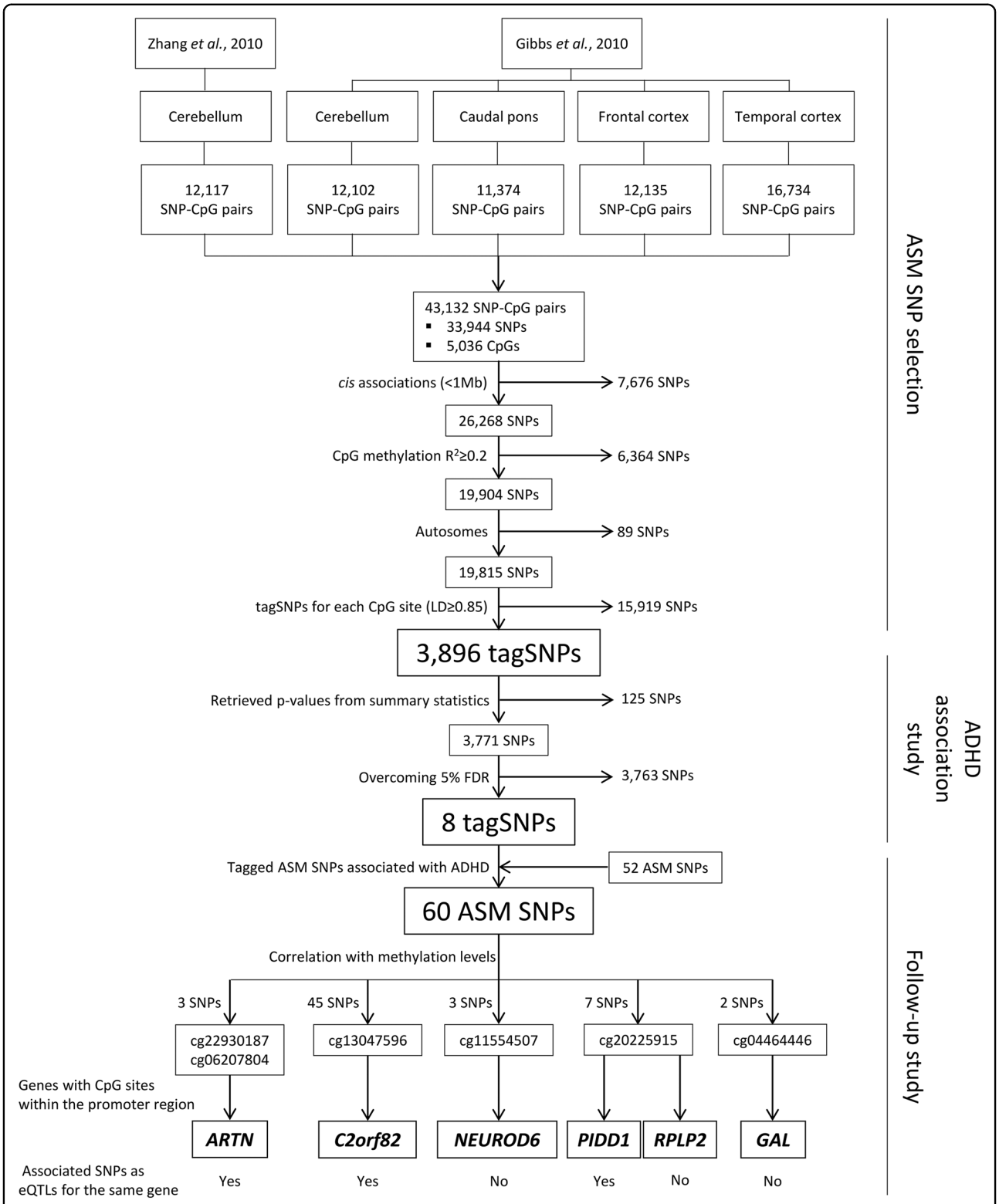


Fig. 1 Selection of allele-specific methylation (ASM) SNPs and association results obtained for ASM variants in ADHD. SNPs tested in the ADHD GWAS meta-analysis and multiple testing correction. SNPs correlating with differential methylation of CpG sites and eQTLs in brain regions (only for genes in which the CpG site lies <5 kb from the transcription start site) are depicted

statistics of the ADHD GWAS meta-analysis (Fig. 1)[4]. False discovery rate (FDR) was applied to correct for multiple testing. We used the *q*-value package for R[27] and obtained a threshold *p*-value of 6.78E−05 corresponding to a 5% FDR. CpG sites highlighted by SNPs that were significant at this FDR threshold were followed-up in further analyses (Fig. 1). Additionally, we performed corrections for multiple testing, using Bonferroni and Genetic type 1 Error Calculator (GEC) methods (http://grass.cgs.hku.hk/gec/)[28]. The Bonferroni-corrected threshold was set at $p \leq 1.32E{-}05$, which considered all the SNPs and tests to be independent (0.05/3771 SNPs). The GEC established the significance threshold at 1.98E−05, which addressed multiple testing for the set of 3771 dependent SNPs by estimating the independent number of tests. The LD between SNPs was calculated according to the 1000 Genomes EUR reference data[25].

Finally, we considered and retrieved *p*-values of those tagged ASM SNPs in high LD ($r^2 \geq 0.85$) with the previous ones that also correlated in *cis* with the methylation levels of the same CpG sites ($R^2 \geq 0.2$) (Fig. 1).

Functional annotation of associated ASM SNPs

We applied four methods to obtain information about the possible functional impact of the ASM SNPs that were associated with ADHD. First, we evaluated the presence of possible enhancer or promoter regions using the Haploreg v4.1 tool[29]. To do this, we considered histone modifications related to enhancer regions (H3K4me1 and H3K27ac) and promoters (H3K4me3 and H3K9ac) of 10 different brain regions (hippocampus middle, substantia nigra, anterior caudate, cingulate gyrus, inferior temporal lobe, angular gyrus, dorsolateral prefrontal cortex, germinal matrix, and male and female fetal brain). Second, we evaluated the effect on gene expression through an eQTL analysis using GTEx data (Release V7)[30]. We considered eQTL information for all available brain tissues: amygdala, anterior cingulate cortex (BA24), caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia, spinal cord cervical c-1, and substantia nigra. Third, we considered all the SNPs, not only ASMs, located within ±1 Mb from the transcription start site (TSS) of each gene to infer if the genetically determined expressions of genes of interest correlated with ADHD. This analysis was carried out using MetaXcan[31], the input being the summary statistics of the ADHD GWAS meta-analysis[4] and prediction models trained with RNA-Seq data of 10 GTEx[30] brain tissues and CommonMind[32] dorsolateral prefrontal cortex. The SNP covariance matrices were generated using the 1000 Genomes Project Phase 3[25] EUR genotypes of the prediction model SNPs. Bonferroni correction for multiple testing was considered ($p \leq 2.27E{-}03$; 0.05/22 tests). Finally, we examined the possible influence of the identified variants on subcortical brain structures. We obtained the summary statistics of a GWAS meta-analysis of eight MRI volumetric measures (nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen, and thalamus) produced by the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium[33]. This ENIGMA2 discovery sample included 13,171 subjects of European ancestry and contained association results between seven million markers and variance in the volumes of the mentioned structures[33]; we applied the Bonferroni correction ($p \leq 1E{-}03$; 0.05/50 SNPs).

Results

We investigated the possible association with ADHD of SNPs that show ASM in brain regions. Starting from two previous studies[21,22] that describe ASM in brain tissues we obtained 43,132 SNP–CpG pairs involving 33,944 SNPs and 5306 CpG sites (Figs. 1 and S1). Genetic risk variants for ADHD are enriched in those ASM SNPs, as observed through enrichment analysis at different association *p*-value thresholds (Table S1).

We detected some overlaps and redundancies between studies and tissues (Fig. S1), so we performed a selection process ending up with a list of 3896 ASM tagSNPs (Figs. 1 and S2). Eight ASM tagSNPs were significantly associated with ADHD after correcting for multiple comparisons (5% FDR, $p \leq 6.78E{-}05$) (Fig. 1 and Table S2). These eight SNPs correlated with differential methylation at six CpG sites in *cis* (three for cg20225915, two for both cg22930187 and cg06207804, and one for each of cg13047596, cg11554507, and cg04464446) in different brain areas (Figs. 2–4 and regional associational plots Figs. S3–S10, Table S2). Three of the eight ASM tagSNPs remained associated with ADHD after applying the Bonferroni and GEC corrections, all correlating with differential methylation at the cg20225915 site (Table S2).

As considering only tagSNPs may overlook true causal variants, we retrieved association results from all the 52 ASM SNPs tagged by the previous ones (LD; $r^2 \geq 0.85$), ending up with 60 variants in eight LD blocks that show association with ADHD and correlate with methylation levels at six CpG sites (Figs. 2–4 and S11–S15 and Table S3). We also selected, for each LD block, the SNP showing the highest number of functional annotations (Table 1), as a putative causal SNP.

Consistently, the direction of the effect of the risk alleles on methylation levels is the same for all the SNPs correlating with the same CpG site. Thus, the risk alleles correlate with decreased methylation of cg22930187, cg06207804, cg11554507 and cg20225915 and with increased methylation of cg13047596 and cg04464446[21,22] (Figs. 2–4 and Tables 1, S2, and S3).

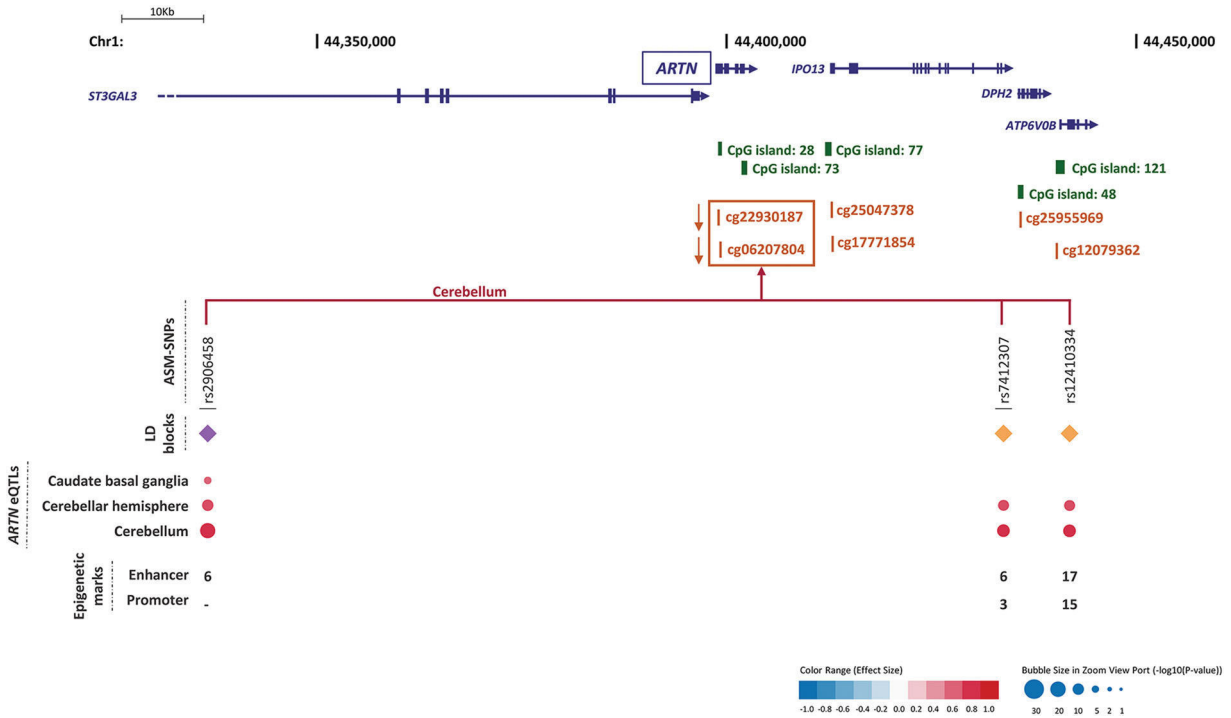


Fig. 2 Genomic context of ASM variants, and methylation and eQTL information for cg22930187 and cg06207804. Genes are depicted in dark blue, showing the direction of transcription with an arrow; CpG sites inspected in the reference studies appear in brown; framed CpG sites indicate those sites showing differential levels of methylation for the associated ASM SNPs, and brown arrows indicate the effect on methylation of the ADHD risk variants, with indication of the brain regions where the ASMs were described. The tagSNPs are underscored. The colored rhombuses show the LD blocks present in each region. The colored dots for eQTLs indicate the effect on gene expression of the ADHD risk allele, according to the legend (red: over-expression, blue: under-expression). The number of enhancer (H3K4me1 and H3K27ac) and promoter (H3K4me3 and H3K9ac) histone marks found in the different brain areas are displayed for each SNP. '-' indicates no known enhancer or promoter histone marks

All six CpG sites are located in possible promoter regions (<5000 bp upstream from a TSS) of six genes (Table 1), all of them expressed in brain: *ARTN* (cg22930187 and cg06207804), *C2orf82* (cg13047596), *NEUROD6* (cg11554507), *PIDD1* (cg20225915), *RPLP2* (cg20225915), and *GAL* (cg04464446) (Figs. 2–4). Furthermore, 85% of the 60 ASM SNPs are located within a region with enhancer or promoter histone marks in at least one brain area (Figs. 2–4 and Tables S4–S8). All putative causal SNPs selected from each LD block lie within a region with histone marks, ranging from 3 to 17 in enhancer regions and from 4 to 16 in promoter regions (Table 1).

We subsequently assessed the possible effect of those 60 SNPs on gene expression and observed that 57 of them are eQTLs for different genes in brain regions (Table S3). Seven out of the eight putative causal SNPs are eQTLs in brain for at least one gene (Table 1). We focused on methylation in promoter regions, which is well established to inversely correlate with gene expression. The eQTLs for *ARTN*, *C2orf82*, and *PIDD1* correlated with methylation of CpG sites lying on their possible promoter regions, showing opposite directions for methylation and gene expression levels (Figs. 2–4 and Tables 1 and S3). The ADHD risk alleles are associated with increased expression of *ARTN* (in cerebellum and a subcortical region) and *PIDD1* (in cerebellum and cortex) and with decreased expression of *C2orf82* (in cortical, subcortical, and cerebellar regions) (Figs. 2–4 and Tables 1 and S3).

Consistently, the predicted direction of the effect on gene expression for these three genes is the same when we consider all variants within ±1 MB from the TSS (and not only the ASM SNPs). We found significant associations of gene expression with ADHD for the same three genes in multiple brain tissues using MetaXcan: *ARTN*, *PIDD1* showed increased expression ($3.57 < Z\text{-score} < 4.19$ and $3.57 < Z\text{-score} < 5.37$, respectively) and *C2orf82* with a decreased expression ($-3.64 < Z\text{-score} < -3.07$) (Table S7), all of them surviving the Bonferroni correction.

We also evaluated the correlation of the 60 ADHD-associated SNPs with subcortical brain volume changes in ENIGMA2 data. SNPs correlating with methylation at cg13047596 and at cg04464446 correlate with nucleus accumbens and/or caudate nucleus volumes, while the only SNP correlating with cg11554507, which is present in ENIGMA2, correlates with thalamus volume (Table S10).

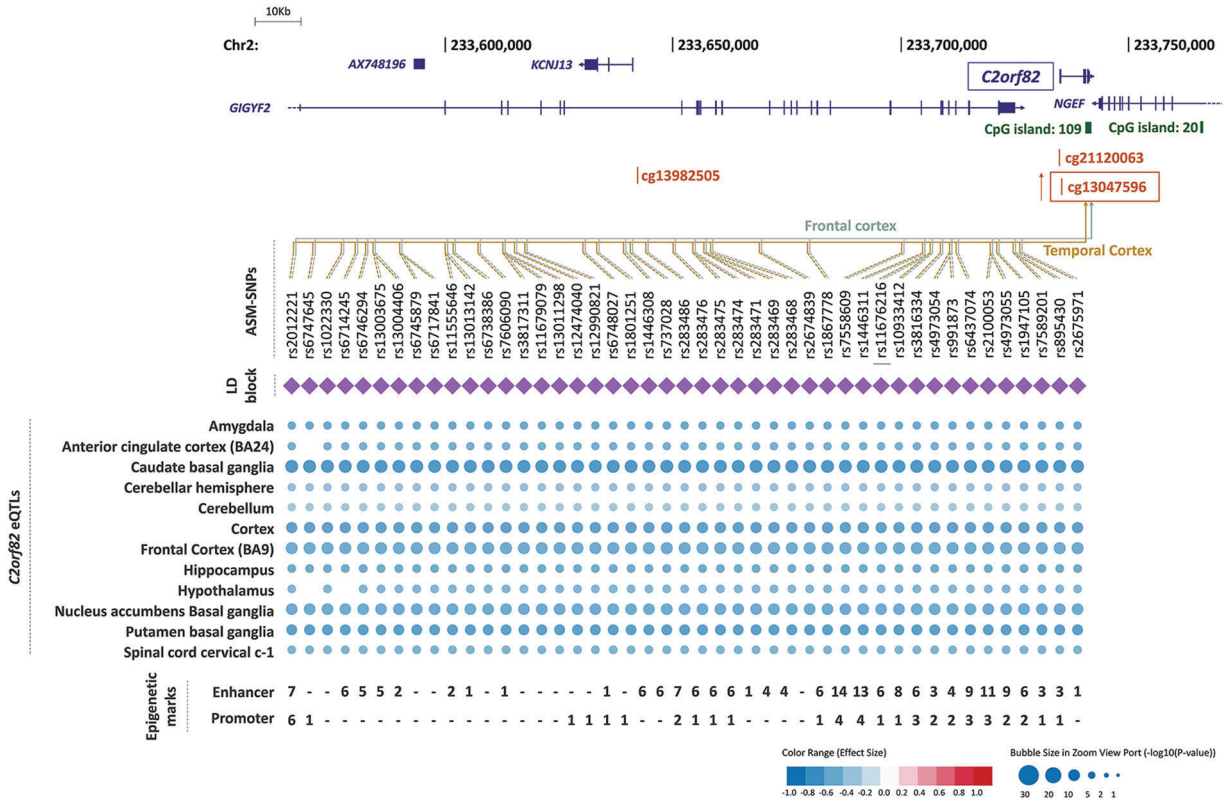


Fig. 3 Genomic context of ASM variants, and methylation and eQTL information for cg13047596. Genes are depicted in dark blue, showing the direction of transcription with an arrow; CpG sites inspected in the reference studies appear in brown; framed CpG sites indicate those sites showing differential levels of methylation for the associated ASM SNPs, and brown arrows indicate the effect on methylation of the ADHD risk variants, with indication of the brain regions where the ASMs were described. The tagSNPs are underscored. The colored rhombuses show the LD blocks present in each region. The colored dots for eQTLs indicate the effect on gene expression of the ADHD risk allele, according to the legend (red: over-expression, blue: under-expression). The number of enhancer (H3K4me1 and H3K27ac) and promoter (H3K4me3 and H3K9ac) histone marks found in the different brain areas are displayed for each SNP. '-' indicates no known enhancer or promoter histone marks

Three of the putative causal SNPs showed correlation with brain volumes (Table 1).

Interestingly, the majority of ASM SNPs that correlate with methylation levels of cg13047596, located in the promoter region of *C2orf82*, are eQTLs in brain for this gene, lie in a region with histone marks and correlate with volume changes of nucleus accumbens and caudate nucleus (Figs. 2–4, Tables 1 and S3–S8 and S10). All this functional evidence highlights the *C2orf82* gene as a good candidate for contributing to ADHD.

Discussion

This study is the first comprehensive assessment of the contribution to ADHD of genetic variants altering methylation in the brain. We identified a total of 60 variants from eight LD blocks associated with ADHD that correlate with differential levels of methylation at six different CpG sites[21,22] (Tables 1 and S3). All the variants from six out of the eight LD blocks alter the methylation of CpG sites lying at potential promoter regions and are also eQTLs for one of the following three genes in multiple brain regions: *ARTN*, *C2orf82*, and *PIDD1* (Figs. 2–4 and Tables 1 and S3). It is well known that DNA methylation in promoter regions inversely correlates with levels of gene expression[18], and all these ASM variants associated with ADHD in our study are concordant with this statement.

The *ARTN* gene, highlighted by two tagSNPs, encodes Artemin, a ligand of the *GDNF* family (glial cell line-derived neurotrophic factor). Artemin supports the survival of sensory and sympathetic peripheral neurons in culture by interacting with GFRα3-RET and possibly also of dopaminergic neurons of the ventral mid-brain through activation of GFRα1-RET complex[34]. Gene Ontology (GO) pathways link it to key neurodevelopmental functions: axon guidance (GO:0007411), neuroblast proliferation (GO:0007405), and peripheral nervous system development (GO:0007422). Risk alleles for ADHD lead to an overexpression of *ARTN*. Previously, overexpression of *ARTN* has been studied in transgenic mice and been linked to an increase of neuron excitability that leads to hypersensitivity[35,36]. Another study in *ARTN*

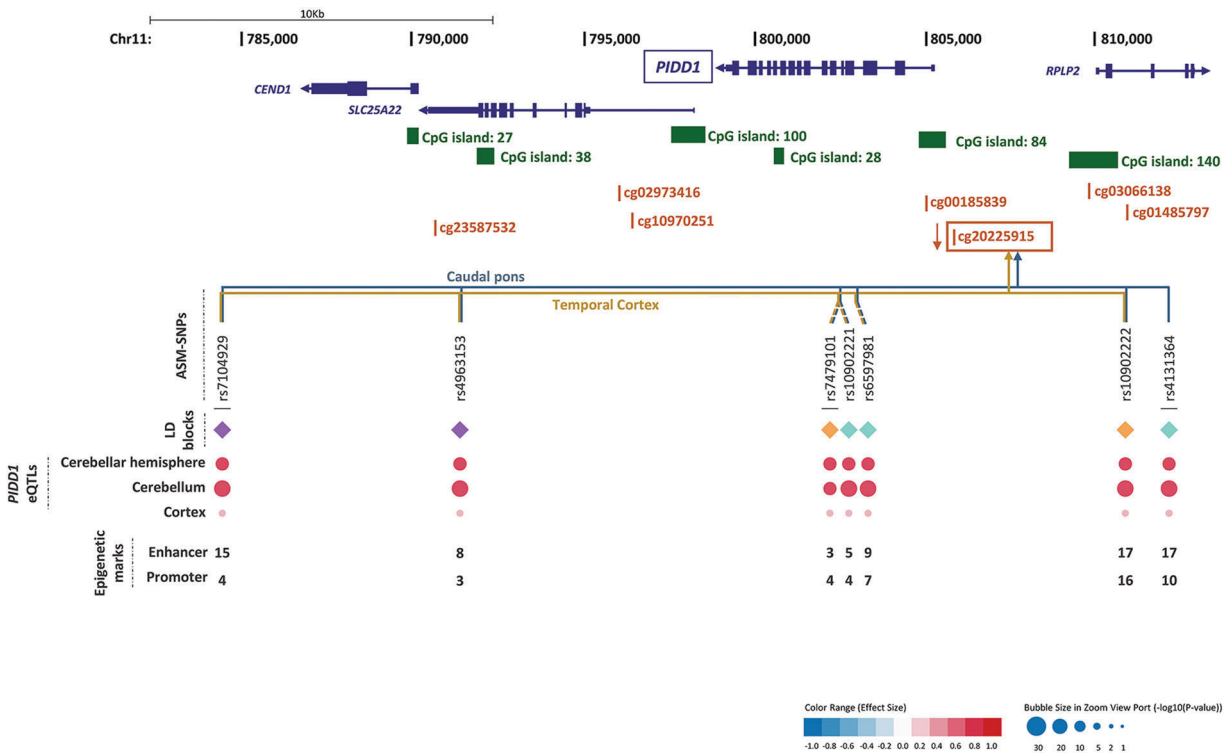


Fig. 4 Genomic context of ASM variants, and methylation and eQTL information for cg20225915. Genes are depicted in dark blue, showing the direction of transcription with an arrow; CpG sites inspected in the reference studies appear in brown; framed CpG sites indicate those sites showing differential levels of methylation for the associated ASM SNPs, and brown arrows indicate the effect on methylation of the ADHD risk variants, with indication of the brain regions where the ASMs were described. The tagSNPs are underscored. The colored rhombuses show the LD blocks present in each region. The colored dots for eQTLs indicate the effect on gene expression of the ADHD risk allele, according to the legend (red: over-expression, blue: under-expression). The number of enhancer (H3K4me1 and H3K27ac) and promoter (H3K4me3 and H3K9ac) histone marks found in the different brain areas are displayed for each SNP. '-' indicates no known enhancer or promoter histone marks

knockout mice reported aberrations in the sympathetic nervous system related to migration and axonal projection[37]. The *C2orf82* gene (also known as *SNORC*) was highlighted by one tagSNP and it encodes a proteoglycan transmembrane protein that is expressed in brain more than in other tissues[30]. Little is known about its function. Finally, *PIDD1* was highlighted by three tagSNPs. It is a cell life regulator gene and it has been linked to apoptotic and anti-apoptotic pathways. The PIDD protein initiates apoptosis as a component of the PIDDosome together with RAIDD (RIP-associated ICH-1/ECD3-homologous protein with a death domain) and procaspase-2[38] and it also activates an anti-apoptotic pathway involving the transcription factor NF-κB in response to genotoxic stress[39].

Alterations in the expression of these three genes (upregulation of *ARTN* and *PIDD1* and downregulation of *C2orf82*) in different brain regions seem to be related to ADHD. Interestingly, most of these regions are relevant for this disorder. Neuroimaging studies have implicated the cerebellum, subcortical and prefrontal regions in ADHD, suggesting a link to problems in the processing of temporal information[40]. Structural anomalies in the cerebellum have been reported in ADHD individuals through neuroimaging studies[41–43]. Cerebellar developmental trajectories and hippocampal volumes are linked to the severity of ADHD symptoms[44–46]. Structural and functional abnormalities in cerebellum and basal ganglia have been associated with motor impairments[47], which are frequent in nearly half of ADHD cases[48]. Subcortical regions identified through our expression analyses have also been related to ADHD, for instance: (i) remarkably different shapes of caudate-putamen basal ganglia and smaller volumes have been reported in ADHD boys[49–52]; (ii) in adult males with ADHD, right caudate volume correlates with poor accuracy on sensory selection tasks[53] and also with hyperactivity/impulsivity[54,55]; (iii) nucleus accumbens, caudate nucleus, putamen, amygdala, and hippocampus are structurally altered in the brains of ADHD patients[56]. Remarkably, all the ASM SNPs in the LD block for *C2orf82* with available information nominally correlate with increased volumes of nucleus accumbens and caudate nucleus subcortical regions. Also, the eQTL

Table 1 Selection of putative causal ASM SNPs associated with ADHD according to functional annotations

SNP	Association with ADHD[a]		Effect on methylation[b]	Epigenetic marks[c]		Effect on expression (GTEx data)[d]	Effect on brain volumes[e]
	Risk allele	*p*-value		Enhancer	Promoter		
rs2906458	G	3.01E−05	↓ cg22930187, ↓ cg06207804	6	0	↑ ***ARTN***	–
rs12410334	A	2.87E−05		17	15		–
rs7558609	A	7.06E−05	↑ cg13047596	14	4	↓ ***C2orf82***	↑ NAc ↑ CN
rs4140961	A	6.05E−05	↓ cg11554507	3	0	–	↑ T
rs7104929	G	7.89E−06	↓ cg20225915	15	4	↑ ***PIDD1*** ↓ *PNPLA2*	?
rs10902222	T	2.03E−06		17	16		–
rs4131364	A	1.60E−06		17	10		–
rs1054252	G	3.86E−05	↑ cg04464446	4	0	↑ *MRPL21*, ↑ *MRGPRD* ↓ *IGHMBP2*	↓ NAc ↓ CN

ASM: Allele-specific methylation, *SNP*: single nucleotide polymorphism, *NAc*: nucleus accumbens, *CN*: caudate nucleus, *T*: thalamus. Risk allele: all alleles are reported in the forward strand; Underlined: significant associations between ASM tagSNPs and ADHD overcoming Bonferroni correction for multiple testing and *p*-value threshold determined using independent number of tests (GEC); ↑: Hypermethylation/overexpression/increased brain volume; ↓: Hypomethylation/underexpression/decreased brain volume; "–": No significant data for the SNP; "?": No information available for the SNP; Enhancer: Number of H3K4me1 and H3K27ac marks; Promoter: Number of H3K4me3 and H3K9ac marks; In bold: genes with the reported CpG sites lying in their possible promoter region

[a]Data obtained from the PGC+iPSYCH ADHD GWAS meta-analysis[4]

[b]Described in Zhang et al. [22] and Gibbs et al. [21]

[c]Histone marks found in brain areas

[d]eQTL information for brain tissues

[e]Data from the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium[33]

effect sizes of these SNPs are the largest for caudate basal ganglia, which volume correlates with the SNP genotype variation. There is evidence about the role in ADHD of cortical thickness, cortical volume and functional connectivity in the anterior cingulate cortex, a region involved in cognitive control, attention, affect and drive[57–63]. Furthermore, delayed cortical development, e.g. in prefrontal regions has been reported in ADHD patients[64,65] and this appears to be stronger in ADHD children with below median intelligence quotient[66]. All the above mentioned fronto-subcortical structures and pathways are rich in catecholamines, the molecular targets in pharmacological treatments for ADHD[48,52,64,67].

Interestingly, the methylation of cg20225915 has also been associated with *PIDD1* expression in peripheral blood[68], turning it into a good candidate as a biomarker. The expression of *ARTN* was found to be altered in blood of major depressive disorder (MDD) patients[69] and the *C2orf82* gene has been associated to schizophrenia[70,71]. Furthermore, *C2orf82* was highlighted in a cross-disorder GWAS of eight psychiatric conditions, including ADHD and schizophrenia[72], with the rs778353 lead SNP, located 47 kb downstream from the gene, showing a genome-wide significant association with the phenotype. All three genes overlap with several CNVs that contribute to autism, intellectual disability or aggressive behavior, conditions often comorbid with ADHD (Table S11). It is noteworthy that some of the CNVs reported in *ARTN*, *C2orf82*, and *PIDD1* are related to brain-specific and overall developmental delay at both fetal and postnatal stages. Thus, it is reasonable to assume that altered expression of these genes might affect brain volumes and cognition. Overall, the fact that these genes have previously been related to neuropsychiatric disorders that are often comorbid with ADHD[73] make them appealing candidates to be pursued.

ARTN is the only gene highlighted in our study that is present in one of the top regions reported in the ADHD GWAS meta-analysis[4], although it did not contain SNPs surviving genome-wide significance. The GWAS findings in the region could be accounted for by one of several genes: *ST3GAL3*, *PTPRF*, *KDM4A*, *RP11-184I16.4*, *XR_246316.1*, *KDM4A-AS1*, and *SLC6A9*. *ST3GAL3* had the most signals. Although two of the reported ASM variants associated with ADHD are intronic to *ST3GAL3*, this gene was not highlighted in our study as none of the associated variants correlated with differential methylation of CpG sites near the *ST3GAL3* TSS (distance from the nearest CpG site: 197 kb) or were eQTLs for the gene in brain tissues. Instead, these SNPs correlated with a nearby gene, *ARTN*, both in terms of methylation and gene expression. This suggests the importance of finding functional connections between disease-associated SNPs and genes, besides considering the genes in the physical vicinity of variants. Furthermore, another of the highlighted genes, *PIDD1*, although not being among the top

findings in the ADHD GWAS meta-analysis[4], it is pointed out by the gene-based association analysis performed in the same study.

Genetic variants surpassing genome-wide significance in GWAS explain only a small part of the SNP-based heritability and associations not reaching the significance threshold also contribute to disease susceptibility[4,9]. An omnigenic model has been recently proposed suggesting that the sub-threshold variants could point at regulatory elements of core genes[7,8]. Indeed, a previous study on a cardiovascular cardiac phenotype reported that nominally significant associations are enriched in enhancer regions[9], consistent with our findings. Therefore, although none of the variants that we identified in our study display genome-wide significant association with ADHD, they may contribute to the susceptibility to ADHD, as they do have a functional impact (methylation, expression, and in some cases brain structure) via genes that are expressed in brain.

Brain-specific ASM information has also been utilized to detect key genes and pathways in BD[20]. Also, a higher enrichment of brain ASM was observed in a schizophrenia GWAS in comparison to non-psychiatric GWAS[10]. This, together with the enrichment of ASM in ADHD-associated variants found in the present study, reinforces the rationale of utilizing ASM SNPs to highlight genes that are relevant to psychiatric disorders from GWAS data.

There are some strengths and limitations in our study that should be discussed. Strengths: (i) We used the largest GWAS meta-analysis of ADHD performed so far, including around 20,000 cases and 35,000 controls. (ii) The genetic variants identified as associated with ADHD have a functional impact on epigenetic regulation, expression or brain volumes. (iii) Two of the highlighted genes in this study, *ARTN* and *C2orf82*, have previously been associated with other psychiatric disorders. (iv) For two of the genes there is more than one LD block showing the same effect on CpG site methylation. (v) Our results are concordant with eQTL information that had been assessed in an independent sample, with all the SNPs showing the opposite effect on methylation of the promoter region and on the expression of a given gene in brain (more promoter methylation and less gene expression or vice versa), even for the different LD blocks from each region. Limitations: (i) We did not perform a follow-up study to replicate the association findings in an independent sample. (ii) The previous studies that we used for the selection of ASM SNPs were performed on different genotyping platforms that do not include all the existing SNPs in the genome, and therefore we could not test all possible ASMs. (iii) We only considered *cis*-associated ASM variants, which are the vast majority, although non-*cis* ASM also occurs. (iv) There is an overrepresentation of ASM SNPs from cerebellum compared to the other studied tissues.

To conclude, the present study points to the *ARTN*, *C2orf82*, and *PIDD1* genes as potential contributors to ADHD susceptibility. The identified risk variants have an impact on the methylation levels of different CpG sites located in promoter regions and they inversely correlate with expression of the corresponding genes in brain. This finding is supported by a prediction of increased expression of *ARTN* and *PIDD1*, and a decreased expression of *C2orf82* in ADHD. Moreover, variants correlating with methylation at cg13047596 (near *C2orf82*) influence the volumes of nucleus accumbens and/or caudate nucleus. Further studies are required to elucidate the mechanisms by which these genes contribute to ADHD.

Acknowledgements

Major financial support for this research was received by B.C. from the Spanish 'Ministerio de Economía y Competitividad' (SAF2015-68341-R, RTI2018-100968-B-I00) and AGAUR, 'Generalitat de Catalunya' (2017-SGR-738). The research leading to these results has also received funding from the European Union Seventh Framework Program [FP7/2007–2013] under grant agreement no. 602805 and from the European Union H2020 Program [H2020/2014-2020] under grant agreements nos. 667302 and 643051, the latter supporting the contract of A.S. L.P.-C. and J.C.-D. were supported by 'Generalitat de Catalunya' (2016 FI_B 00728 and 2015 FI_B 00448, respectively). L.P.-C. was also supported by 'Ministerio de Educación, Cultura y Deporte' (FPU15/03867). N.F.-C. was supported by contracts of the 'Centro de Investigación Biomédica en Red de Enfermedades Raras' (CIBERER). V.R. was supported by the Graduate School of Health from the University of Aarhus. The iPSYCH team acknowledges support from the Lundbeck Foundation. Finally, S.F. was supported by the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement no. 602805, the European Union's Horizon 2020 research and innovation program under grant agreements nos. 667302 and 728018 and NIMH grants 5R01MH101519 and U01 MH109536-01. We are thankful to Roser Corominas (Universitat de Barcelona) for helpful advice. We are also grateful to the ADHD Working Group of the Psychiatric Genomics Consortium (PGC) and the iPSYCH team for distributing the summary statistics of the ADHD GWAS meta-analysis. This study is part of the International Multicentre persistent ADHD Collaboration (IMpACT); www.impactadhdgenomics.com). IMpACT unites major research centres working on the genetics of ADHD persistence across the lifespan and has participants in The Netherlands, Germany, Spain, Norway, the United Kingdom, the United States, Brazil and Sweden. Principal investigators of IMpACT are: Barbara Franke (chair), Andreas Reif (co-chair), Stephen V. Faraone, Jan Haavik, Bru Cormand, J. Antoni Ramos-Quiroga, Marta Ribasés, Philip Asherson, Klaus-Peter Lesch, Jonna Kuntsi, Claiton H.D. Bau, Jan Buitelaar, Alejandro Arias-Vásquez, Tetyana Zayats, Henrik Larsson, Alysa Doyle, and Eugenio H. Grevet. Access to the PGC ADHD data was obtained through dbGaP project number 10608 that includes the following datasets: phs000016.v2.p2, phs000407.v1.p1, phs000358.v1.p1, and phs000490.v1.p1. We thank the ENIGMA consortium for sharing the summary statistics of genome-wide association meta-analyses of MRI phenotypes.

Author details

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain. [2]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Madrid, Spain. [3]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain. [4]Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Barcelona, Catalonia, Spain. [5]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [6]Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. [7]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [8]Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA

Conflict of interest
The authors declare that they have no conflict of interest.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-019-0574-7).

Received: 15 August 2018 Revised: 3 July 2019 Accepted: 30 July 2019
Published online: 03 October 2019

References

1. Polanczyk, G., De Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* **164**, 942–948 (2007).
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Arlington, VA, US: American Psychiatric Publishing, Inc. (2013).
3. Faraone, S. V. et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **57**, 1313–1323 (2005).
4. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
5. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
6. Meaburn, E. L., Schalkwyk, L. C. & Mill, J. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics* **5**, 578–582 (2010).
7. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
8. Liu, X., Li, Y. I. & Pritchard, J. K. Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034 (2019).
9. Wang, X. et al. Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* **5**, e10557 (2016).
10. Gagliano, S. A. et al. Allele-skewed DNA modification in the brain: relevance to a Schizophrenia GWAS. *Am. J. Hum. Genet.* **98**, 956–962 (2016).
11. Ladd-Acosta, C. et al. DNA methylation signatures within the human brain. *Am. J. Hum. Genet.* **81**, 1304–1315 (2007).
12. Jang, H. S., Shin, W. J., Lee, J. E. & Do, J. T. CpG and non-CpG methylation in epigenetic gene regulation and brain function. *Genes* **8**, E148 (2017).
13. Xu, Y. et al. Multiple epigenetic factors predict the attention deficit/hyperactivity disorder among the Chinese Han children. *J. Psychiatr. Res.* **64**, 40–50 (2015).
14. Schanen, N. C. Epigenetics of autism spectrum disorders. *Hum. Mol. Genet.* **15**, R138–R150 (2006).
15. Grayson, D. R. & Guidotti, A. Merging data from genetic and epigenetic approaches to better understand autistic spectrum disorder. *Epigenomics* **8**, 85–104 (2016).
16. Dammann, G. et al. Increased DNA methylation of neuropsychiatric genes occurs in borderline personality disorder. *Epigenetics* **6**, 1454–1462 (2011).
17. Gaur, U., Li, K., Mei, S. & Liu, G. Research progress in allele-specific expression and its regulatory mechanisms. *J. Appl. Genet.* **54**, 271–283 (2013).
18. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a019133 (2014).
19. Do, C. et al. Genetic-epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol.* **18**, 120 (2017).
20. Chuang, L.-C., Kao, C.-F., Shih, W.-L. & Kuo, P.-H. Pathway analysis using information from allele-specific gene methylation in genome-wide association studies for bipolar disorder. *PLoS ONE* **8**, e53092 (2013).
21. Gibbs, J. R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.* **6**, e1000952 (2010).
22. Zhang, D. et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am. J. Hum. Genet.* **86**, 411–419 (2010).
23. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
24. Barrett, J. C. Haploview: visualization and analysis of snp genotype. *Data. Cold Spring Harb. Protoc.* **4**, 1–5 (2009).
25. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
26. R Core Team. *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/ (2014).
27. Storey, John D. with contributions from Bass, Andrew J. AD and, Robinson D. qvalue: *Q*-value estimation for false discovery rate control. R Package Version 2.6.0. http://github.com/jdstorey/qvalue (2015).
28. Li, M.-X., Yeung, J. M. Y., Cherny, S. S. & Sham, P. C. Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).
29. Ward, L. D. & Kellis, M. HaploRegv4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
30. GTEx Consortium et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Publ. Gr.* **45**, 580–585 (2013).
31. Barbeira, A. N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1825 (2018).
32. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
33. Hibar, D. P. et al. Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
34. Baloh, R. H. et al. Artemin, a novel member of the GDNF ligand family, supports peripheral and central neurons and signals through the GFRα3-RET receptor complex. *Neuron* **21**, 1291–1302 (1998).
35. Elitt, C. M. et al. Artemin overexpression in skin enhances expression of TRPV1 and TRPA1 in cutaneous sensory neurons and leads to behavioral sensitivity to heat and cold. *J. Neurosci.* **26**, 8578–8587 (2006).
36. Elitt, C. M., Malin, S. A., Koerber, H. R., Davis, B. M. & Albers, K. M. Overexpression of artemin in the tongue increases expression of TRPV1 and TRPA1 in trigeminal afferents and causes oral sensitivity to capsaicin and mustard oil. *Brain Res.* **1230**, 80–90 (2008).
37. Honma, Y. et al. Artemin is a vascular-derived neurotropic factor for developing sympathetic neurons. *Neuron* **35**, 267–282 (2002).
38. Tinel, A. & Tschopp, J. The PIDDosome, a protein complex implicated in activation of caspase-2 in response to genotoxic stress. *Science* **304**, 843–846 (2004).
39. Janssens, S., Tinel, A., Lippens, S. & Tschopp, J. PIDD mediates NF-kappaB activation in response to DNA damage. *Cell* **123**, 1079–1092 (2005).
40. Toplak, M. E., Dockstader, C. & Tannock, R. Temporal information processing in ADHD: findings to date and new methods. *J. Neurosci. Methods* **151**, 15–29 (2006).
41. Berquin, P. C. et al. Cerebellum in attention-deficit hyperactivity disorder: a morphometric MRI study. *Neurology* **50**, 1087–1093 (1998).
42. Castellanos, F. X. et al. Quantitative brain magnetic resonance imaging in attention-deficit hyperactivity disorder. *Arch. Gen. Psychiatry* **53**, 607–616 (1996).
43. Castellanos, F. X. et al. Quantitative brain magnetic resonance imaging in girls with attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiatry* **58**, 289–295 (2001).
44. Mackie, S. et al. Cerebellar development and clinical outcome in attention deficit hyperactivity disorder. *Am. J. Psychiatry* **164**, 647–655 (2007).
45. Plessen, K. J. et al. Hippocampus and amygdala morphology in attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiatry* **63**, 795–807 (2006).
46. Friedman, L. A. & Rapoport, J. L. Brain development in ADHD. *Curr. Opin. Neurobiol.* **30**, 106–111 (2015).
47. Pasini, A. & D'agati, E. Pathophysiology of NSS in ADHD. *World J. Biol. Psychiatry* **10**, 495–502 (2009).
48. Curatolo, P., D'Agati, E. & Moavero, R. The neurobiological basis of ADHD. *Ital. J. Pediatr.* **36**, 79 (2010).
49. Qiu, A. et al. Basal ganglia volume and shape in children with attention deficit hyperactivity disorder. *Am. J. Psychiatry* **166**, 74–82 (2009).
50. Greven, C. U. et al. Developmentally stable whole-brain volume reductions and developmentally sensitive caudate and putamen volume alterations in those with attention-deficit/hyperactivity disorder and their unaffected siblings. *JAMA Psychiatry* **72**, 490–499 (2015).
51. Sobel, L. J. et al. Basal ganglia surface morphology and the effects of stimulant medications in youth with attention deficit hyperactivity disorder. *Am. J. Psychiatry* **167**, 977–986 (2010).

52. Rubia, K., Alegría, A. A. & Brinson, H. Brain abnormalities in attention-deficit hyperactivity disorder: a review. *Rev. Neurol.* **58**(Suppl. 1), S3–S16 (2014).
53. Casey, B. J. et al. Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **36**, 374–383 (1997).
54. Singh, A., Yeh, C. J., Verma, N. & Das, A. K. Overview of attention deficit hyperactivity disorder in young children. *Heal Psychol. Res.* **3**, 23–35 (2015).
55. Onnink, A. M. H. et al. Brain alterations in adult ADHD: Effects of gender, treatment and comorbid depression. *Eur. Neuropsychopharmacol.* **24**, 397–409 (2014).
56. Hoogman, M. et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry* **4**, 310–319 (2017).
57. Shaw, P. et al. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiatry* **63**, 540–549 (2006).
58. Ambrosino, S., de Zeeuw, P., Wierenga, L. M., van Dijk, S. & Durston, S. What can cortical development in attention-deficit/hyperactivity disorder teach us about the early developmental mechanisms involved? *Cereb. Cortex* **27**, 4624–4634 (2017).
59. Castellanos, F. X. & Aoki, Y. Intrinsic functional connectivity in attention-deficit/hyperactivity disorder: a science in development. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 253–261 (2016).
60. Makris, N. et al. Cortical thinning of the attention and executive function networks in adults with attention-deficit/hyperactivity disorder. *Cereb. Cortex* **17**, 1364–1375 (2007).
61. Makris, N. et al. Anterior cingulate volumetric alterations in treatment-naïve adults with ADHD: a pilot study. *J. Atten. Disord.* **13**, 407–413 (2010).
62. Yang, X.-R., Carrey, N., Bernier, D. & MacMaster, F. P. Cortical thickness in young treatment-naive children with ADHD. *J. Atten. Disord.* **19**, 925–930 (2015).
63. Qiu, M. et al. Changes of brain structure and function in ADHD children. *Brain Topogr.* **24**, 243–252 (2011).
64. Cortese, S. The neurobiology and genetics of Attention-Deficit/Hyperactivity Disorder (ADHD): what every clinician should know. *Eur. J. Paediatr. Neurol.* **16**, 422–433 (2012).
65. Shaw, P. et al. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proc. Natl Acad. Sci. USA* **104**, 19649–19654 (2007).
66. de Zeeuw, P. et al. Differential brain development with low and high IQ in attention-deficit/hyperactivity disorder. *PLoS ONE* **7**, e35770 (2012).
67. Pliszka, S. R., Lancaster, J., Liotti, M. & Semrud-Clikeman, M. Volumetric MRI differences in treatment-naive vs chronically treated children with ADHD. *Neurology* **67**, 1023–1027 (2006).
68. Wu, Y. et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat. Commun.* **9**, 918 (2018).
69. Otsuki, K. et al. Altered expression of neurotrophic factors in patients with major depression. *J. Psychiatr. Res.* **42**, 1145–1153 (2008).
70. Schizophrenia Working Group of the Psychiatric Genomics. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
71. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
72. Lee, P. H. et al. Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *bioRxiv*. 528117 (2019).
73. Brown T. E. *ADHD Comorbidities: Handbook for ADHD Complications in Children and Adults*. Washington, DC, US: American Psychiatric Press (2009).

Supplementary Material

EXPLORING GENETIC VARIATION THAT INFLUENCES BRAIN METHYLATION IN ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

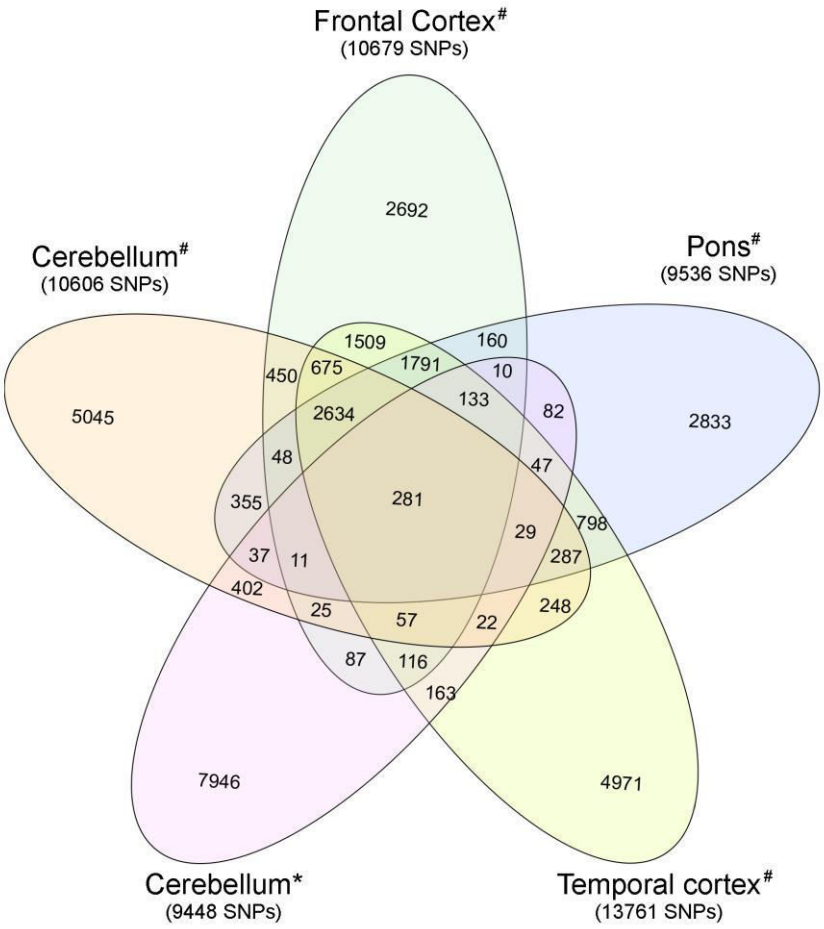


Figure S1. Venn diagram showing the overlaps among the initial 33,944 ASM SNPs selected in the different brain areas. *Cerebellum from the study by Zhang *et al.* 2010. #Brain areas from the study by Gibbs *et al.* 2010. The overlap accounts for 31% of the ASM SNPs; 12.5% ASM SNPs are shared between two tissues, 9% are shared between three tissues, 8.5% are shared between four tissues and only 1% are shared between all the tissues. Sixteen percent of the 9,448 ASM SNPs identified by Zhang et al., 2010 overlap with the ASM SNPs identified in the Gibbs et al., 2010 study.

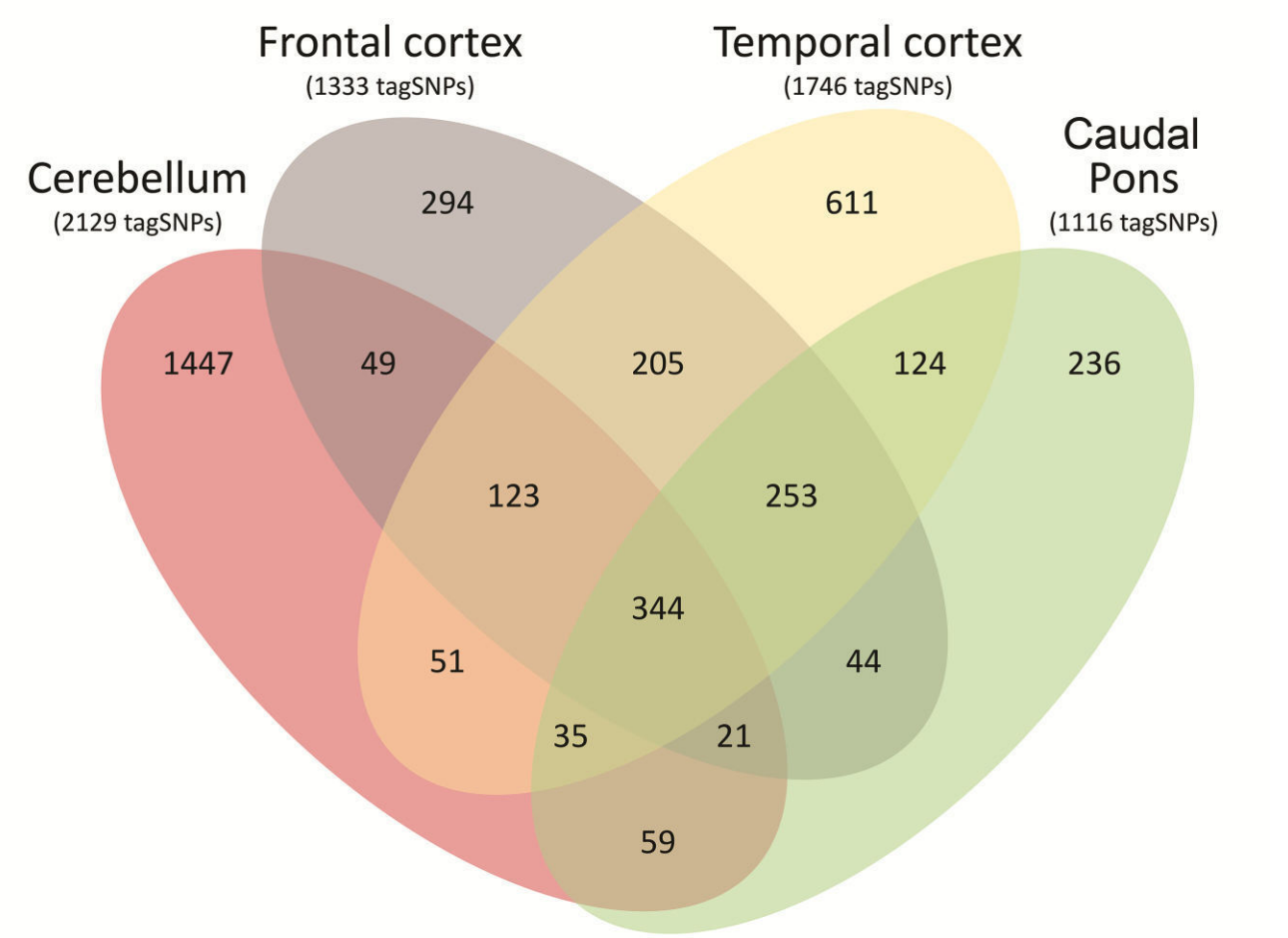


Figure S2. Venn diagram showing the overlaps among the 3,896 ASM tagSNPs selected in the different brain areas. The overlap accounts for 33.5% of tagSNPs; 13.5% tagSNPs are shared between two tissues, 11% are shared between three tissues and 9% are shared between all the tissues.

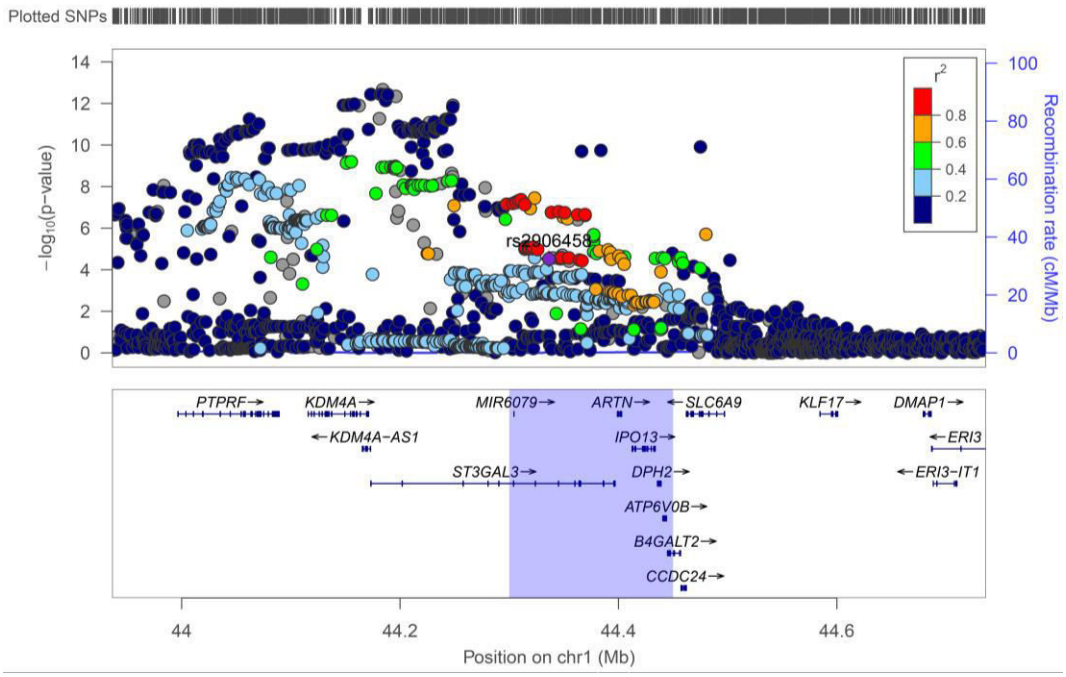


Figure S3. Regional association plot for rs2906458. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 2.

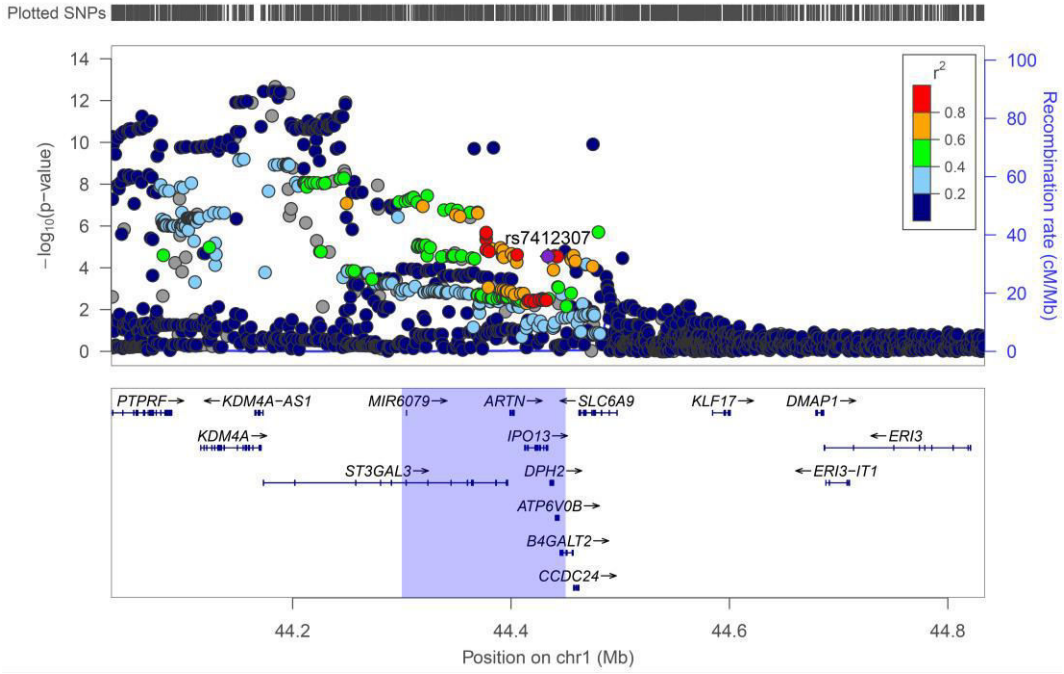


Figure S4. Regional association plot for rs7412307. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 2.

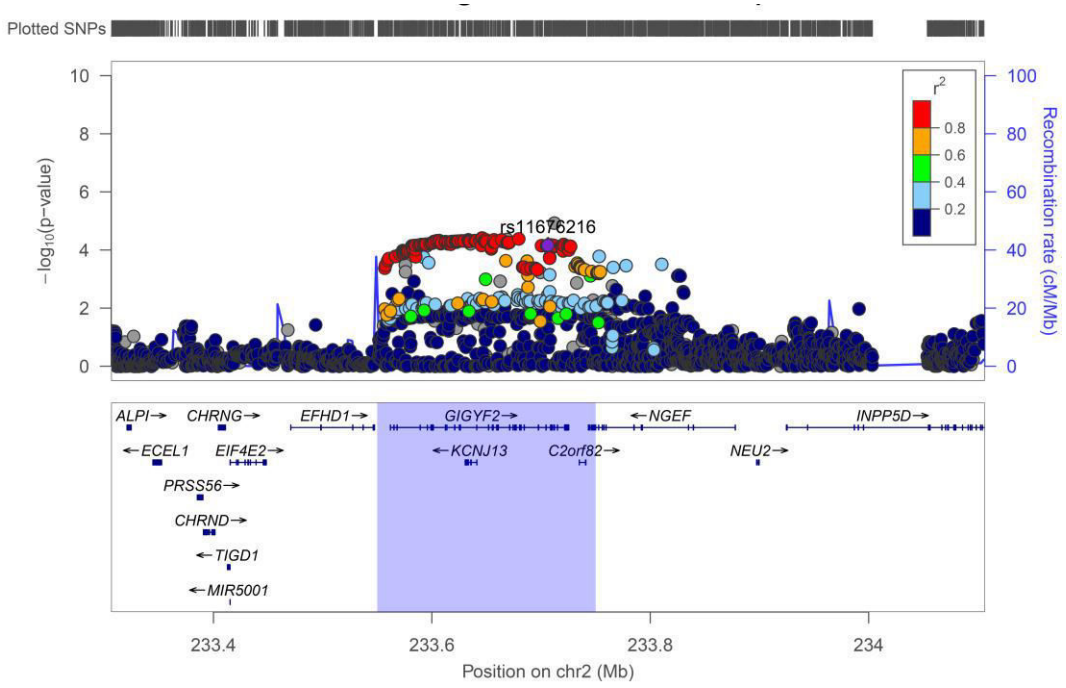


Figure S5. Regional association plot for rs11676216. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 3.

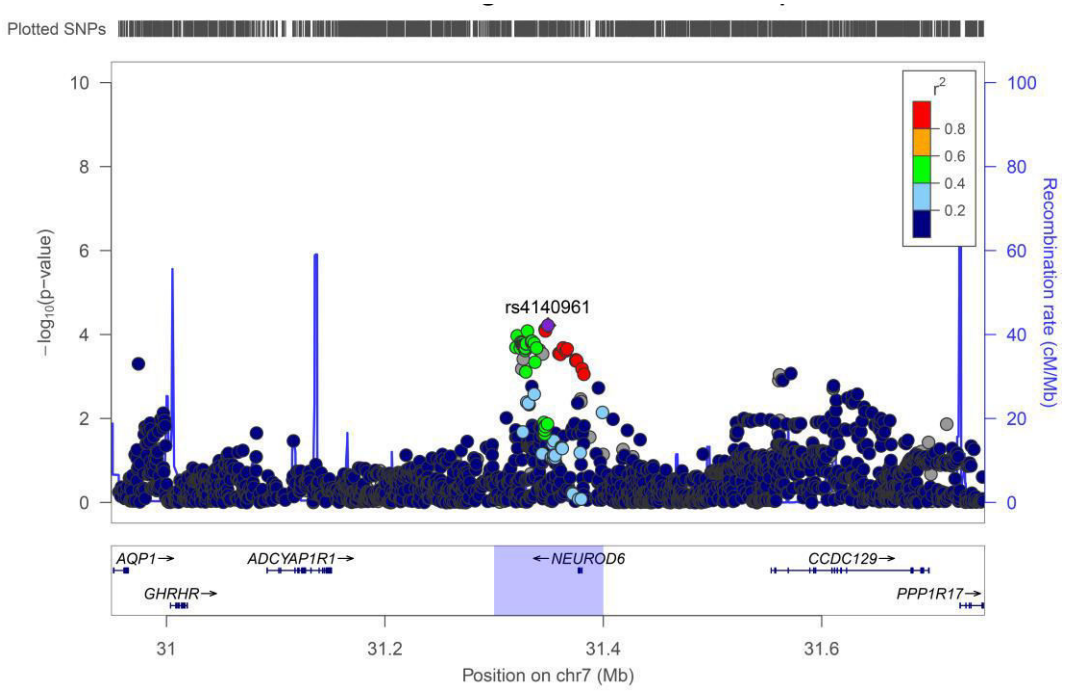


Figure S6. Regional association plot for rs4140961. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: region containing the CpG site and the ASM-SNPs of interest.

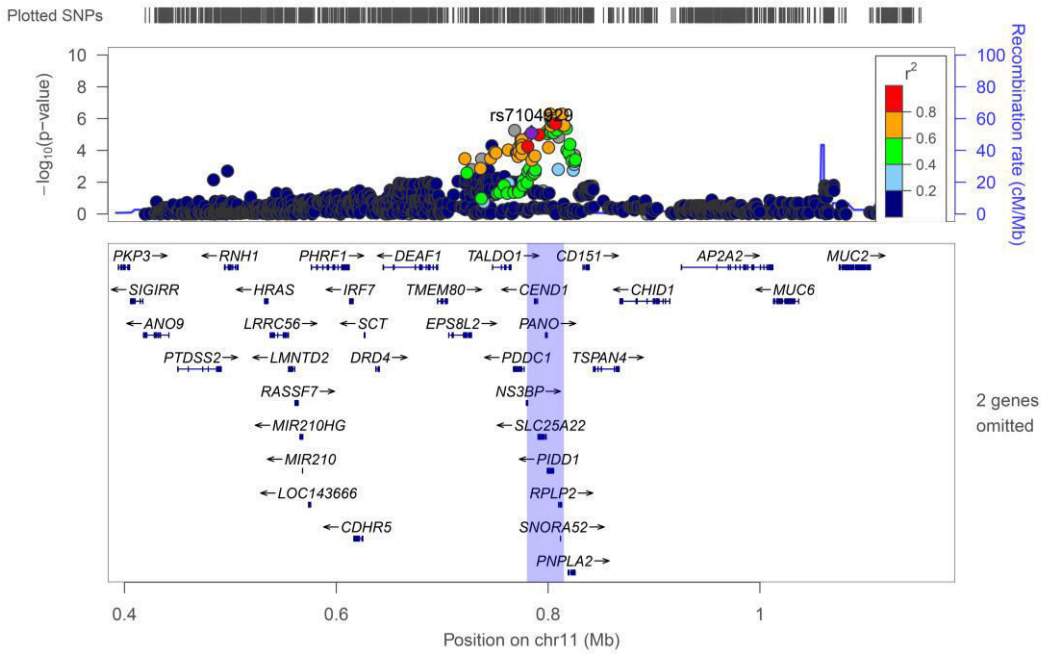


Figure S7. Regional association plot for rs7104929. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 4.

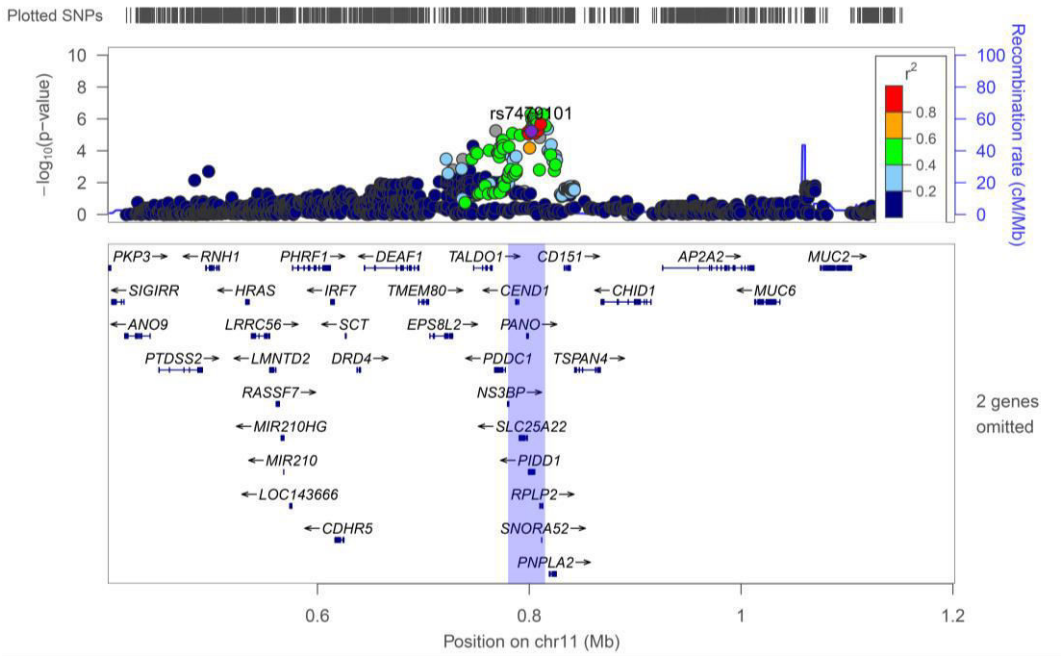


Figure S8. Regional association plot for rs7479101. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 4.

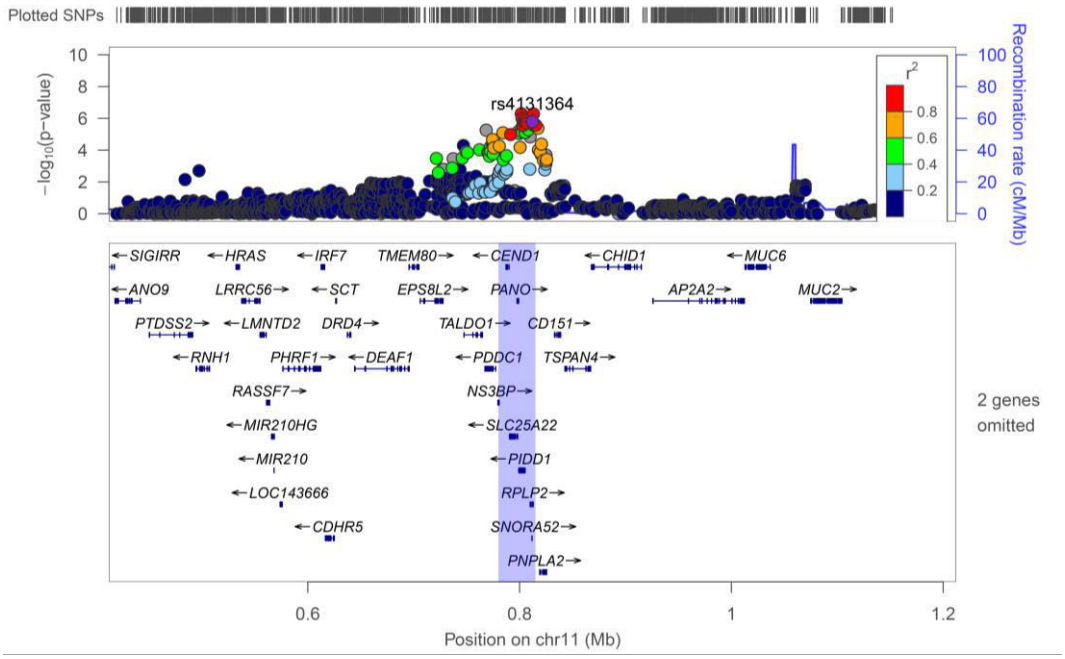


Figure S9. Regional association plot for rs4131364. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: Region represented in Figure 4.

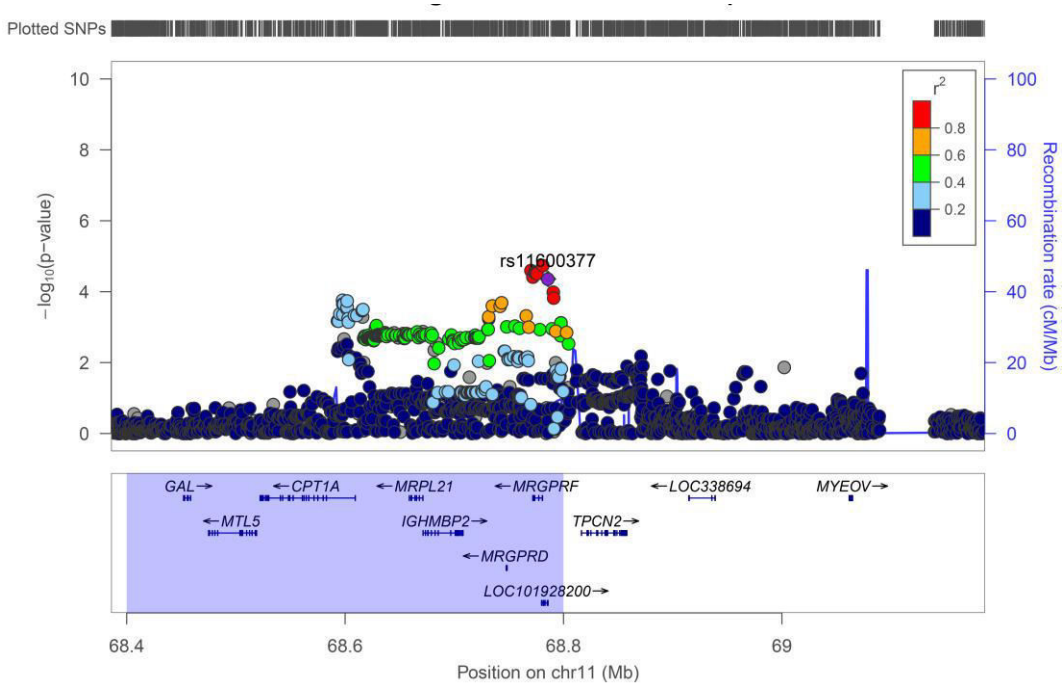


Figure S10. Regional association plot for rs11600377. The SNP represented in the regional plot is depicted in purple. Highlighted in blue: region containing the CpG site and the ASM-SNPs of interest.

Figure S11. Linkage disequilibrium (LD, r^2) patterns between the three SNPs associated with ADHD that also correlate with differential methylation at two CpG sites, cg22930187 and cg06207804, located in the possible promoter region of *ARTN*.

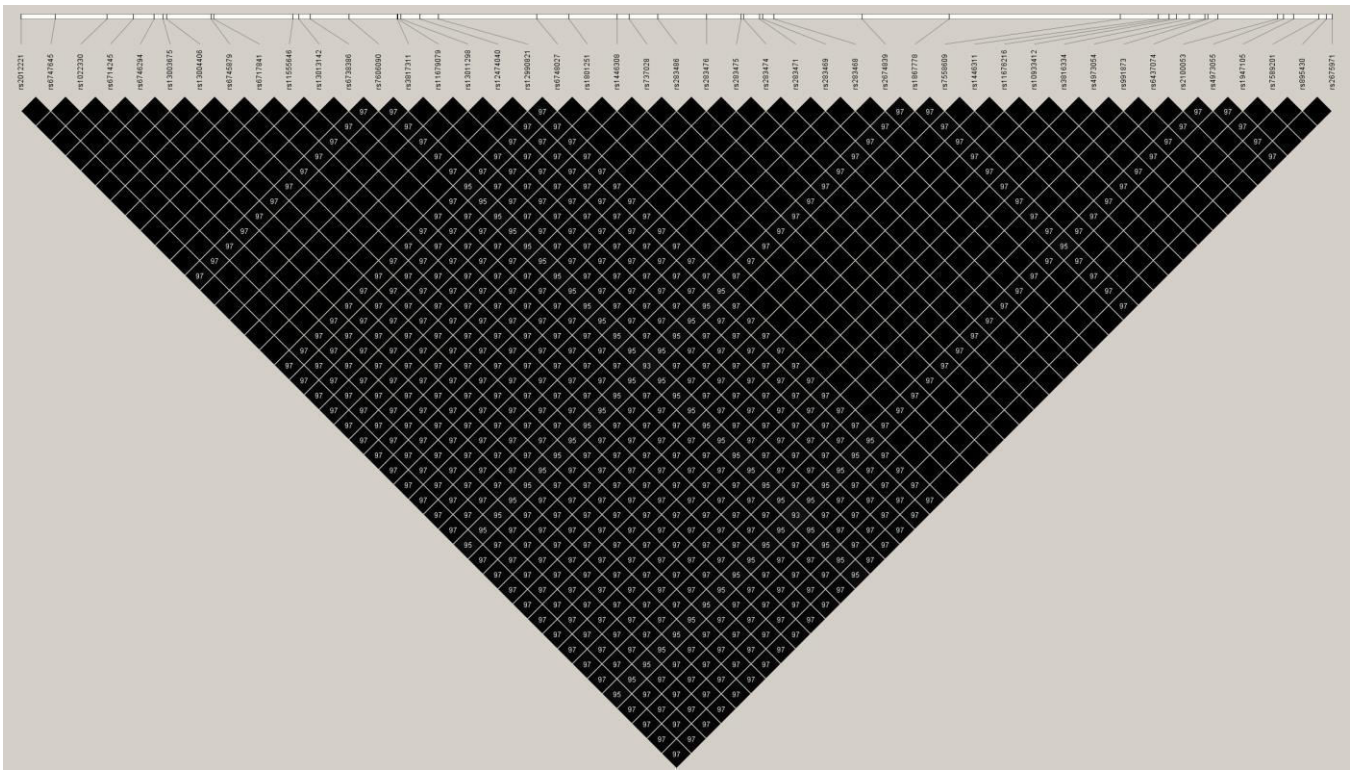


Figure S12. Linkage disequilibrium (LD, r^2) patterns between the 45 SNPs associated with ADHD that also correlate with differential methylation at the CpG site cg13047596, located in the possible promoter region of *C2orf82*.

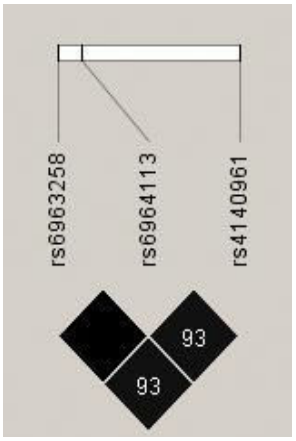


Figure S13. Linkage disequilibrium (LD, r^2) patterns between the three SNPs associated with ADHD that also correlate with differential methylation at the CpG site cg11554507, located in the possible promoter region of *NEUROD6*.

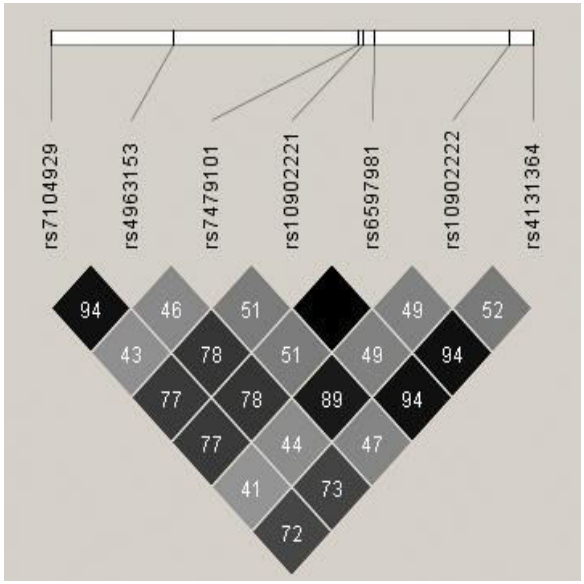


Figure S14. Linkage disequilibrium (LD, r^2) patterns between the 7 SNPs associated with ADHD that also correlate with differential methylation at the CpG site cg20225915, located in the possible promoter region of *PIDD1*.

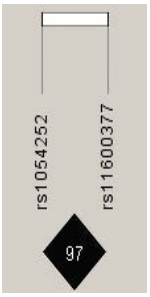


Figure S15. Linkage disequilibrium (LD, r^2) patterns between the two SNPs associated with ADHD that also correlate with differential methylation at the CpG site cg04464446, located in the possible promoter region of *GAL*.

Table S1. Enrichment analysis of ASM SNPs at different significance thresholds in the ADHD GWAS meta-analysis by Demontis et al. (2019).

Significance Threshold	N SNPs	N ASM SNPs	p-value	OR
5.00E-08	303	6	1.70E-03	4.92
5.00E-07	945	8	4.30E-02	2.08
5.00E-06	2,122	15	3.15E-02	1.74
5.00E-05	6,970	35	1.31E-01	1.23
5.00E-04	25,288	139	4.58E-04	1.35
5.00E-03	115,681	527	6.94E-03	1.12
5.00E-02	651,772	2790	5.54E-03	1.05

ASM: Allele-specific methylation; N SNPs: Significant SNPs in the ADHD GWAS meta-analysis for the corresponding significance threshold; N ASM SNPs: Significant ASM SNPs in the GWAS meta-analysis; Underlined: Significant enrichment of ASM SNPs in the list of ADHD-associated SNPs; OR: Odds ratio.

Table S2. ASM tagSNPs associated with ADHD.

	SNP information				[1]Association with ADHD				[2]Correlation with methylation			
			Alleles		Freq A1							
SNP	Chr	Pos	A1	A2	Cases	Controls	OR	p-value	CpG site	Tissue	Effect	p-value
rs2906458	1	44336389	A	G	0.74	0.756	0.94	3.01E-05	cg22930187	Crbl	↓	5.55E-10
									cg06207804	Crbl	↓	1.10E-12
rs7412307	1	44433864	C	G	0.185	0.172	1.07	2.82E-05	cg22930187	Crbl	↓	3.59E-11
									cg06207804	Crbl	↓	1.76E-15
rs11676216	2	233706368	T	C	0.646	0.654	0.95	6.78E-05	cg13047596	Tctx	↑	1.09E-10
										Fctx	↑	2.35E-10
rs4140961	7	31349352	A	G	0.597	0.592	1.06	6.05E-05	cg11554507	Pons	↓	8.45E-25
										Tctx	↓	4.72E-23
rs7104929	11	784340	C	G	0.512	0.526	0.94	**7.89E-06**	cg20225915	Pons	↓	8.17E-10
										Tctx	↓	1.51E-08
rs7479101	11	802115	A	G	0.317	0.33	0.93	**5.90E-06**	cg20225915	Pons	↓	2.15E-14
										Tctx	↓	3.37E-14
rs4131364	11	812188	A	G	0.517	0.502	1.07	**1.60E-06**	cg20225915	Pons	↓	2.49E-10
rs11600377	11	68785803	A	G	0.731	0.72	1.06	4.38E-05	cg04464446	Crbl	↑	3.14E-08

ASM: Allele-specific methylation; [1]Data obtained from the PGC+iPSYCH ADHD GWAS meta-analysis (Demontis et al., 2019); [2]Described in Zhang et al., 2010 and Gibbs et al., 2010; SNP: Single Nucleotide Polymorphism; Chr: Chromosome; Pos: Position (build hg19); A1: Allele 1; A2: Allele 2; All alleles are reported in the forward strand; Freq A1: Frequency of allele 1; OR: Odds Ratio (calculated on A1); Effect: Direction of the risk allele effect on DNA methylation levels; Underlined allele: Risk allele for ADHD; In bold: Significant associations for the association between ASM tagSNPs and ADHD p-values overcoming Bonferroni correction for multiple testing and p-value threshold determined using independent number of tests (GEC); Crbl: Cerebellum; Tctx: Temporal cortex; Fctx: Frontal cortex.

Table S3. ASM SNPs associated with ADHD.

					Alleles		[1] Association with ADHD					
SNP	**TagSNP**	**LD (R^2)**	**Chr**	**Pos**	**A1**	**A2**	**Freq A1 cases**	**Freq A1 controls**	**OR**	**p-value**	**[2] ADHD risk allele effect on CpG methylation (ASM studies)**	**[3] ADHD risk allele effect on gene expression (GTEx data)**
rs2906458*	rs2906458	1	1	44336389	A	G	0.74	0.756	0.94	3.01E-05		
rs7412307*	rs7412307	1	1	44433864	C	G	0.185	0.172	1.07	2.82E-05	↓ cg22930187, cg06207804	↑ *ARTN*
rs12410334		0.96	1	44442521	A	C	0.184	0.171	1.07	2.87E-05		
rs2012221		0.977	2	233566848	T	C	0.642	0.651	0.95	1.67E-04		
rs6747645		0.977	2	233571033	A	G	0.637	0.647	0.95	1.39E-04		
rs1022330		0.977	2	233577330	A	G	0.357	0.348	1.05	1.15E-04		
rs6714245		0.977	2	233580505	T	C	0.643	0.652	0.95	1.12E-04		
rs6746294		0.977	2	233583050	C	G	0.645	0.653	0.95	9.06E-05		
rs13003675		0.977	2	233584109	T	C	0.355	0.346	1.06	6.88E-05		
rs13004406		0.977	2	233584557	A	T	0.642	0.651	0.95	1.49E-04		
rs6745879		0.977	2	233590007	A	G	0.645	0.654	0.95	6.66E-05		
rs6717841		0.977	2	233590255	T	C	0.355	0.346	1.06	6.42E-05		
rs11555646		0.977	2	233599904	A	C	0.645	0.654	0.95	5.80E-05		
rs13013142		0.977	2	233600606	A	G	0.645	0.654	0.95	5.73E-05		
rs6738386		0.977	2	233602028	T	C	0.645	0.654	0.95	5.06E-05		
rs7606090		0.954	2	233606740	T	C	0.645	0.654	0.95	4.90E-05		
rs3817311		0.977	2	233612557	T	C	0.355	0.346	1.06	5.13E-05		
rs11679079		0.977	2	233612656	T	C	0.355	0.346	1.06	5.29E-05		
rs13011298		0.977	2	233612996	A	G	0.645	0.654	0.95	5.30E-05		
rs12474040		0.977	2	233615345	T	G	0.645	0.654	0.95	5.28E-05		
rs12990821		0.977	2	233617585	C	G	0.645	0.654	0.95	5.09E-05		
rs6748027		1.0	2	233629552	T	C	0.645	0.654	0.95	4.86E-05		
rs1801251		1.0	2	233633460	A	G	0.355	0.346	1.06	4.42E-05		
rs1446308		1.0	2	233639309	T	C	0.645	0.654	0.95	4.67E-05		
rs737028		1.0	2	233640750	T	C	0.644	0.653	0.95	4.74E-05		
rs283486	rs11676216	1.0	2	233644223	A	G	0.645	0.653	0.95	4.73E-05	↑ cg13047596	↓ *C2orf82*
rs283476		1.0	2	233650168	A	G	0.657	0.666	0.95	5.09E-05		
rs283475		1.0	2	233654381	T	G	0.342	0.334	1.06	9.06E-05		
rs283474		1.0	2	233654627	A	G	0.656	0.665	0.95	7.96E-05		
rs283471		1.0	2	233656627	A	G	0.342	0.333	1.06	5.30E-05		
rs283469		1.0	2	233656997	T	G	0.353	0.345	1.06	4.45E-05		
rs283468		1.0	2	233658309	T	C	0.342	0.333	1.06	5.26E-05		
rs2674839		1.0	2	233669040	C	G	0.354	0.346	1.06	5.56E-05		
rs1867778		0.977	2	233679644	T	C	0.646	0.655	0.94	4.11E-05		
rs7558609		1.0	2	233700379	A	G	0.354	0.346	1.06	7.06E-05		
rs1446311		1.0	2	233705071	A	G	0.646	0.654	0.95	6.97E-05		
rs11676216*		1.0	2	233706368	T	C	0.646	0.654	0.95	6.78E-05		
rs10933412		1.0	2	233707226	C	G	0.646	0.654	0.95	6.76E-05		
rs3816334		1.0	2	233708806	A	G	0.354	0.346	1.06	6.66E-05		
rs4973054		1.0	2	233710713	C	G	0.646	0.654	0.95	6.82E-05		
rs991873		1.0	2	233711046	A	G	0.354	0.346	1.06	6.89E-05		
rs6437074		1.0	2	233712296	A	G	0.646	0.654	0.95	7.05E-05		
rs2100053		1.0	2	233719516	T	G	0.645	0.653	0.95	8.18E-05		
rs4973055		0.977	2	233720283	T	G	0.639	0.647	0.95	1.07E-04		
rs1947105		1.0	2	233721455	A	G	0.645	0.654	0.95	7.31E-05		
rs7589201		1.0	2	233724536	A	G	0.355	0.346	1.06	7.49E-05		
rs895430		1.0	2	233725483	A	C	0.646	0.655	0.95	9.84E-05		
rs2675971		1.0	2	233726154	A	G	0.355	0.346	1.06	7.31E-05		
rs6963258		0.937	7	31346832	A	T	0.401	0.405	0.95	7.43E-05		
rs6964113	rs4140961	0.937	7	31347163	C	G	0.598	0.593	1.05	8.16E-05	↓ cg11554507	-
rs4140961*		1	7	31349352	A	G	0.597	0.592	1.06	6.05E-05		
rs7104929*	rs7104929	1.0	11	784340	C	G	0.512	0.526	0.94	**7.89E-06**		
rs4963153		0.941	11	791462	A	G	0.492	0.478	1.06	**1.04E-05**		
rs7479101*	rs7479101	1.0	11	802115	A	G	0.317	0.33	0.93	**5.90E-06**		
rs10902222		0.896	11	810882	T	G	0.697	0.683	1.07	**2.03E-06**	↓cg20225915	↑ *PIDD1* ↓ *PNPLA2*
rs10902221		0.941	11	802379	T	C	0.478	0.492	0.93	**9.70E-07**		
rs6597981	rs4131364	0.941	11	803017	A	G	0.471	0.485	0.94	**2.77E-06**		
rs4131364*		1.0	11	812188	A	G	0.517	0.502	1.07	**1.60E-06**		
rs1054252	rs11600377	0.973	11	68772072	A	G	0.27	0.281	0.94	3.86E-05	↑ cg04464446	↑ *MRPL21, MRGPRD* ↓ *IGHMBP2*
rs11600377*		1.0	11	68785803	A	G	0.731	0.72	1.06	4.38E-05		

ASM: Allele-specific methylation; [1] Data obtained from the PGC+iPSYCH ADHD GWAS meta-analysis (Demontis et al., 2019); [2]Described in Zhang et al., 2010 and Gibbs et al., 2010; [3]eQTL information for brain tissues; SNP: Single Nucleotide Polymorphism; LD: linkage disequilibrium; Chr: Chromosome; Pos: Position (build hg19); A1: Allele 1; A2: Allele 2; All alleles are reported in the forward strand; Freq A1: Frequency of allele 1; OR: Odds Ratio (calculated on A1); Underlined allele: Risk allele for ADHD; ↑: Hypermethylation/Overexpression; ↓: Hypomethylation/Downexpression; "-": No significant data for the SNP; *: Significant tagSNPs overcoming 5% FDR, the other SNPs are ASM SNPs in LD with these significant tagSNPs; In bold: Significant associations for the association between ASM tagSNPs and ADHD p-values overcoming Bonferroni correction for multiple testing and p-value threshold determined using independent number of tests (GEC).

Table S4. Histone marks for the ASM SNPs correlating with differential methylation of cg22930187 and cg06207804.

SNP			Hippocampus Middle	Substantia Nigra	Anterior Caudate	Cingulate Gyrus	Inferior Temporal Lobe	Angular Gyrus	Dorsolateral Prefrontal Cortex	Germinal Matrix	Fetal Brain Female	Fetal Brain Male
rs2906458	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1			H3K4me1		H3K4me1	H3K4me1		H3K4me1	
		H3K27ac							H3K27ac			
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs7412307	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1										
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3	H3K4me3					H3K4me3				
		H3K9ac			H3K9ac							
rs12410334	Chromatin state	15-state model	EnhG	TxFlnk	EnhG	TxFlnk	EnhG	EnhG	TxFlnk	EnhG	TxFlnk	EnhG
		25-state model	TxReg	TxReg	TxReg	TxReg	TxReg	TxReg	TxReg	TxEnh5	TxEnh5	TxReg
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	
		H3K9ac		H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac			

ASM: Allele-specific methylation; H3K4me1 and H3K27ac: Histone marks related to enhancer regions; H3K4me3 and H3K9ac: Histone marks related to promoter regions; EnhG: Genic enhancer; TxFlnk: Transcription at gene 5' and 3'; TxReg: Transcribed and regulatory (Promoter/Enhancer); TxEnh5: Transcribed 5' preferential and enhancers; Gray cells: Experiments from Roadmap are not available.

Table S5. Histone marks for the ASM SNPs correlating with differential methylation of cg13047596.

SNP			Hippocampus Middle	Substantia Nigra	Anterior Caudate	Cingulate Gyrus	Inferior Temporal Lobe	Angular Gyrus	Dorsolateral Prefrontal Cortex	Germinal Matrix	Fetal Brain Female	Fetal Brain Male
rs2012221	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1			
		H3K27ac					H3K27ac					
	Promoter histone marks	H3K4me3						H3K4me3	H3K4me3			
		H3K9ac		H3K9ac	H3K9ac		H3K9ac	H3K9ac				
rs6747645	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1										
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac						H3K9ac				
rs1022330	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1										
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs6714245	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1	H3K4me1			
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs6746294	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1			H3K4me1	H3K4me1			
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs13003675	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1			H3K4me1	H3K4me1			
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs13004406	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1			H3K4me1			H3K4me1				
		H3K27ac										

	Promoter histone marks	H3K4me3									
		H3K9ac									
rs6745879	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1									
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs6717841	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1									
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs11555646	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1	H3K4me1						H3K4me1		
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs13013142	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1							H3K4me1		
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs6738386	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1									
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs7606090	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1		H3K4me1							
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs3817311	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1									
		H3K27ac									
	Promoter histone	H3K4me3									

	marks	H3K9ac											
rs11679079	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac											
rs13011298	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac											
rs12474040	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs12990821	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs6748027	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac						H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac									
rs1801251	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs1446308	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1					
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac											

rs737028	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1	H3K4me1		
		H3K27ac									
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs283486	Chromatin state	15-state model		Enh				Enh			
		25-state model		EnhW2	EnhW2	EnhW2	EnhW2	EnhAF	EnhW2		
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1		
		H3K27ac		H3K27ac							
	Promoter histone marks	H3K4me3									
		H3K9ac		H3K9ac	H3K9ac						
rs283476	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1		H3K4me1		H3K4me1					
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac					
	Promoter histone marks	H3K4me3									
		H3K9ac			H3K9ac						
rs283475	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1	H3K4me1								
		H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3									
		H3K9ac			H3K9ac						
rs283474	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1	H3K4me1								
		H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3									
		H3K9ac			H3K9ac						
rs283471	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1									
		H3K27ac					H3K27ac				
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs283469	Chromatin state	15-state model									
		25-state model									
	Enhancer histone marks	H3K4me1			H3K4me1						
		H3K27ac	H3K27ac				H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3									
		H3K9ac									
rs283468	Chromatin state	15-state model									

	Chromatin state	25-state model											
	Enhancer histone marks	H3K4me1			H3K4me1								
		H3K27ac	H3K27ac	H3K27ac				H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac											
rs2674839	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac											
	Promoter histone marks	H3K4me3											
		H3K9ac											
rs1867778	Chromatin state	15-state model				EnhG							
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1		H3K4me1				
		H3K27ac						H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac					H3K9ac						
rs7558609	Chromatin state	15-state model	EnhG	Enh				EnhG					
		25-state model	TxEnh3	TxEnh3	TxEnh3	TxEnh3	TxEnh3	TxEnh3	TxEnh3				
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1				
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac	H3K9ac	H3K9ac		H3K9ac					
rs1446311	Chromatin state	15-state model		EnhG									
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1	H3K4me1				
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac	H3K9ac		H3K9ac	H3K9ac					
rs11676216	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1										
		H3K27ac		H3K27ac	H3K27ac		H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs10933412	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1			H3K4me1								
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs3816334	Chromatin state	15-state model					EnhG						
		25-state model			TxEnh3		TxEnh3	TxEnh3	TxEnh3				

	Enhancer histone marks	H3K4me1											
		H3K27ac	H3K27ac		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac		H3K9ac	H3K9ac					
rs4973054	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac			H3K27ac	H3K27ac		H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac			H3K9ac					
rs991873	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1			H3K4me1								
		H3K27ac			H3K27ac	H3K27ac		H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac			H3K9ac					
rs6437074	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1		H3K4me1	H3K4me1	H3K4me1			H3K4me1				
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac					
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac	H3K9ac			H3K9ac					
rs2100053	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1			H3K4me1					
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac	H3K9ac			H3K9ac					
rs4973055	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1	H3K4me1		H3K4me1								
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac				H3K9ac					
rs1947105	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac	H3K27ac				
	Promoter histone marks	H3K4me3											
		H3K9ac		H3K9ac	H3K9ac								
rs7589201	Chromatin state	15-state model											
		25-state model											
	Enhancer histone	H3K4me1											

	marks	H3K27ac	H3K27ac	H3K27ac		H3K27ac							
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs895430	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac	H3K27ac	H3K27ac		H3K27ac							
	Promoter histone marks	H3K4me3											
		H3K9ac			H3K9ac								
rs2675971	Chromatin state	15-state model											
		25-state model											
	Enhancer histone marks	H3K4me1											
		H3K27ac		H3K27ac									
	Promoter histone marks	H3K4me3											
		H3K9ac											

ASM: Allele-specific methylation; H3K4me1 and H3K27ac: Histone marks related to enhancer regions; H3K4me3 and H3K9ac: Histone marks related to promoter regions; Enh: Enhancers; EnhW2: Weak Enhancer 2; EnhG: Genic enhancer; EnhAF: Active enhancer Flank; TxEnh3: Transcribed 3' preferential and enhancers; Gray cells: Experiments from Roadmap are not available.

Table S6. Histone marks for the ASM SNPs correlating with differential methylation of cg11554507.

SNP			Hippocampus Middle	Substantia Nigra	Anterior Caudate	Cingulate Gyrus	Inferior Temporal Lobe	Angular Gyrus	Dorsolateral Prefrontal Cortex	Germinal Matrix	Fetal Brain Female	Fetal Brain Male
rs6963258	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1									H3K4me1	H3K4me1
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs6964113	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1									H3K4me1	H3K4me1
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs4140961	Chromatin state	15-state model								Enh	Enh	Enh
		25-state model										EnhAc
	Enhancer histone marks	H3K4me1								H3K4me1	H3K4me1	H3K4me1
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										

ASM: Allele-specific methylation; H3K4me1 and H3K27ac: Histone marks related to enhancer regions; H3K4me3 and H3K9ac: Histone marks related to promoter regions; Enh: Enhancers; EnhAc: Primary H3K27ac possible enhancer; Gray cells: Experiments from Roadmap are not available.

Table S7. Histone marks for the ASM SNPs correlating with differential methylation of cg20225915.

SNP			Hippocampus Middle	Substantia Nigra	Anterior Caudate	Cingulate Gyrus	Inferior Temporal Lobe	Angular Gyrus	Dorsolateral Prefrontal Cortex	Germinal Matrix	Fetal Brain Female	Fetal Brain Male
rs7104929	Chromatin state	15-state model									Enh	Enh
		25-state model									EnhW2	EnhAc
	Enhancer histone marks	H3K4me1	H3K4me1		H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1		H3K4me1	H3K4me1
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3							H3K4me3			
		H3K9ac			H3K9ac		H3K9ac		H3K9ac			
rs4963153	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1							H3K4me1			H3K4me1
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac		H3K27ac			
	Promoter histone marks	H3K4me3										
		H3K9ac			H3K9ac			H3K9ac	H3K9ac			
rs7479101	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1							H3K4me1			
		H3K27ac						H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3										
		H3K9ac			H3K9ac	H3K9ac	H3K9ac	H3K9ac				
rs10902221	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1							H3K4me1			H3K4me1
		H3K27ac	H3K27ac					H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3										
		H3K9ac			H3K9ac	H3K9ac	H3K9ac	H3K9ac				
rs6597981	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1						H3K4me1	H3K4me1	H3K4me1		H3K4me1
		H3K27ac	H3K27ac			H3K27ac	H3K27ac	H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3					H3K4me3					
		H3K9ac		H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac			
rs10902222	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac			
	Promoter histone marks	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3	H3K4me3
		H3K9ac		H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac	H3K9ac			
rs4131364	Chromatin state	15-state model	EnhG			EnhG			EnhG	EnhG		
		25-state model	TxEnh5		TxEnh5	TxEnh5	TxEnh5	TxEnh5	TxEnh5	TxEnh5	TxEnh5	
	Enhancer histone marks	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1	H3K4me1
		H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac	H3K27ac			
	Promoter histone	H3K4me3	H3K4me3				H3K4me3		H3K4me3	H3K4me3	H3K4me3	

ASM: Allele-specific methylation; H3K4me1 and H3K27ac: Histone marks related to enhancer regions; H3K4me3 and H3K9ac: Histone marks related to promoter regions; Enh: Enhancers; EnhW2: Weak Enhancer 2; EnhAc: Primary H3K27ac possible enhancer; EnhG: Genic enhancer; TxEnh5: Transcribed 5' preferential and enhancers; Gray cells: Experiments from Roadmap are not available.

Table S8. Histone marks for the ASM SNPs correlating with differential methylation of cg04464446.

SNP			Hippocampus Middle	Substantia Nigra	Anterior Caudate	Cingulate Gyrus	Inferior Temporal Lobe	Angular Gyrus	Dorsolateral Prefrontal Cortex	Germinal Matrix	Fetal Brain Female	Fetal Brain Male
rs1054252	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1							H3K4me1			
		H3K27ac	H3K27ac			H3K27ac			H3K27ac			
	Promoter histone marks	H3K4me3										
		H3K9ac										
rs11600377	Chromatin state	15-state model										
		25-state model										
	Enhancer histone marks	H3K4me1										
		H3K27ac										
	Promoter histone marks	H3K4me3										
		H3K9ac										

ASM:Allele-specific methylation; H3K4me1 and H3K27ac: Histone marks related to enhancer regions; H3K4me3 and H3K9ac: Histone marks related to promoter regions; Gray cells: Experiments from Roadmap are not available.

Table S9. MetaXcan prediction of gene expression effects on ADHD for multiple brain tissues.

Gene	Brain tissue*	Z-score	p-value	N SNPs in model	N SNPs used	N ASM SNPs	Predicted R^2
ARTN	Caudate basal ganglia	1.54	1.20E-01	19	16	0	0.10
	Cerebellar hemisphere	4.19	**2.50E-05**	15	15	4	0.37
	Cerebellum	3.57	**3.50E-04**	31	31	5	0.36
	Cortex	3.94	**8.10E-05**	5	4	2	0.14
	Frontal cortex	1.42	1.50E-01	29	24	2	0.09
	Hippocampus	1.55	1.10E-01	140	129	2	0.16
C2orf82	Dorsolateral prefrontal cortex	-3.50	**4.50E-04**	58	41	9	0.43
	Amygdala	-3.07	**2.00E-03**	92	89	42	0.40
	Anterior cingulate cortex	-3.55	**3.00E-04**	11	11	7	0.19
	Caudate basal ganglia	-3.54	**3.00E-04**	11	11	8	0.44
	Cerebellar hemisphere	-3.64	**2.00E-04**	37	35	6	0.19
	Cerebellum	-3.50	**4.00E-04**	30	30	7	0.28
	Cortex	-3.46	**5.00E-04**	59	54	7	0.27
	Frontal cortex	-3.59	**3.00E-04**	29	28	6	0.45
	Hippocampus	-3.40	**6.00E-04**	25	23	8	0.24
	Hypothalamus	-3.37	**7.00E-04**	30	28	13	0.17
	Nucleus accumbens basal ganglia	-3.59	**3.00E-04**	29	29	11	0.33
	Putamen basal ganglia	-3.20	**1.00E-03**	52	43	9	0.52
PIDD1	Dorsolateral prefrontal cortex	4.71	**2.41E-06**	4	2	1	0.06
	Cerebellar hemisphere	5.00	**4.20E-07**	32	27	4	0.53
	Cerebellum	5.37	**7.60E-08**	36	27	5	0.49
	Cortex	3.57	**3.40E-04**	64	47	2	0.03

ASM: Allele-specific methylation; *ADHD prediction models were only available for some tissues and genes; Z-score: Number of standard deviations change in gene expression in ADHD; p-value: Significance of the association between predicted expression levels and ADHD; N SNPs in model: Number of SNPs used in the training of prediction models for each gene; N SNPs used: Number of SNPs used from the ADHD GWAS meta-analysis summary statistics; N ASM SNPs: Number of ASM SNPs included in the model; Predicted R^2: Correlation between the predicted and observed gene expression during prediction model training; In bold: Significant p-values overcoming Bonferroni correction for multiple testing.

Table S10. Correlations with sub-cortical brain volumes of the ASM SNPs associated with ADHD.

				Nucleus Accumbens		Amygdala		Caudate nucleus		Hippocampus		Pallidum		Putamen		Thalamus	
SNP	**ASM for CpG site**	**A1**	**A2**	**Effect**	**p-value**	**Effect**	**p-value**	**Effect**	**p-value**	**Effect**	**p-value**	**Effect**	**p-value**	**Effect**	**p-value**	**Effect**	**p-value**
rs2906458*	cg22930187 cg06207804	A	G	0.2	0.885	0.7	0.784	0.8	0.881	-4.1	0.442	3.9	0.073	8.9	0.191	3.5	0.628
rs7412307*		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs12410334		A	C	-0.6	0.651	2.9	0.341	-1.6	0.795	-3.0	0.624	3.9	0.119	3.9	0.619	-4.0	0.623
rs2012221		T	C	2.5	**0.041**	1.7	0.494	10.8	**0.040**	-2.5	0.618	0.2	0.923	8.2	0.202	-4.9	0.471
rs6747645		A	G	2.5	**0.030**	1.8	0.473	11.1	**0.035**	-2.7	0.589	0.5	0.812	8.5	0.187	-4.6	0.496
rs1022330		A	G	2.5	**0.029**	1.8	0.479	10.8	**0.038**	-2.6	0.609	0.2	0.908	8.2	0.200	-4.8	0.471
rs6714245		T	C	2.5	**0.038**	1.8	0.470	10.7	**0.040**	-2.8	0.577	0.3	0.880	8.1	0.202	-4.9	0.467
rs6746294		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs13003675		T	C	2.5	**0.036**	1.8	0.465	10.5	**0.044**	-2.2	0.663	0.3	0.897	7.9	0.220	-4.9	0.467
rs13004406		A	T	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs6745879		A	G	2.5	**0.031**	1.8	0.468	10.6	**0.040**	-2.7	0.585	0.3	0.881	8.1	0.201	-4.9	0.456
rs6717841		T	C	2.4	**0.044**	2.2	0.377	10.1	0.052	-2.4	0.628	0.2	0.941	7.1	0.268	-4.8	0.471
rs11555646		A	C	2.5	**0.028**	1.7	0.478	10.6	**0.039**	-2.7	0.586	0.2	0.923	7.9	0.207	-5.1	0.435
rs13013142		A	G	2.5	**0.028**	1.7	0.478	10.6	**0.039**	-2.7	0.587	0.2	0.922	7.9	0.207	-5.1	0.435
rs6738386		T	C	2.5	**0.028**	1.8	0.476	10.6	**0.039**	-2.7	0.587	0.2	0.920	7.9	0.208	-5.1	0.435
rs7606090		T	C	2.5	**0.027**	1.8	0.470	10.7	**0.037**	-2.7	0.580	0.2	0.912	8.0	0.202	-5.2	0.428
rs3817311		T	C	2.5	**0.029**	1.8	0.472	10.6	**0.038**	-2.7	0.577	0.2	0.915	7.8	0.212	-5.5	0.405
rs11679079		T	C	2.5	**0.029**	1.8	0.472	10.6	**0.038**	-2.7	0.576	0.2	0.916	7.8	0.213	-5.5	0.405
rs13011298		A	G	2.5	**0.029**	1.8	0.472	10.6	**0.037**	-2.7	0.576	0.2	0.916	7.8	0.213	-5.5	0.404
rs12474040		T	G	2.5	**0.029**	1.8	0.473	10.6	**0.038**	-2.8	0.567	0.2	0.918	7.7	0.217	-5.6	0.395
rs12990821		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs6748027		T	C	2.5	**0.025**	1.6	0.512	10.5	**0.040**	-3.0	0.536	0.2	0.938	7.6	0.224	-5.5	0.396
rs1801251		A	G	2.5	**0.025**	1.6	0.516	10.5	**0.040**	-3.1	0.531	0.2	0.940	7.6	0.224	-5.6	0.394
rs1446308		T	C	2.5	**0.025**	1.6	0.518	10.5	**0.040**	-3.1	0.532	0.2	0.937	7.6	0.222	-5.5	0.398
rs737028		T	C	2.5	**0.024**	1.6	0.519	10.5	**0.040**	-3.1	0.530	0.2	0.933	7.7	0.218	-5.4	0.404
rs283486	cg13047596	A	G	2.6	**0.023**	1.6	0.517	10.5	**0.039**	-3.1	0.525	0.1	0.944	8.0	0.201	-5.2	0.427
rs283476		A	G	2.1	0.068	2.6	0.301	10.5	**0.041**	-2.1	0.671	-0.7	0.725	6.5	0.307	-3.2	0.634
rs283475		T	G	2.2	0.065	2.6	0.289	10.4	**0.044**	-2.1	0.665	-0.7	0.716	6.4	0.311	-3.2	0.626
rs283474		A	G	2.2	0.065	2.6	0.301	10.5	**0.042**	-2.1	0.664	-0.7	0.714	6.5	0.304	-3.2	0.633
rs283471		A	G	2.2	0.062	2.9	0.252	10.7	**0.039**	-2.1	0.676	-0.7	0.736	6.3	0.318	-2.9	0.667
rs283469		T	G	2.6	**0.025**	1.6	0.506	10.3	**0.044**	-3.0	0.541	0.1	0.974	8.4	0.179	-5.0	0.449
rs283468		T	C	2.2	0.064	2.6	0.299	10.5	**0.042**	-2.2	0.659	-0.8	0.707	6.5	0.308	-3.2	0.630
rs2674839		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs1867778		T	C	2.5	**0.030**	1.6	0.516	10.3	**0.044**	-3.2	0.508	0.1	0.973	7.9	0.203	-5.8	0.373
rs7558609		A	G	2.6	**0.019**	1.5	0.537	10.6	**0.037**	-3.4	0.484	0.1	0.974	8.0	0.202	-6.4	0.326
rs1446311		A	G	2.6	**0.020**	1.5	0.526	10.3	**0.043**	-3.1	0.522	0.1	0.967	7.9	0.202	-6.1	0.348
rs11676216*		T	C	2.6	**0.020**	1.5	0.527	10.5	**0.039**	-2.9	0.555	0.0	0.992	8.1	0.194	-6.1	0.348
rs10933412		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs3816334		A	G	2.6	**0.019**	1.6	0.508	10.4	**0.042**	-3.0	0.536	0.1	0.959	8.1	0.194	-6.2	0.339
rs4973054		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs991873		A	G	2.6	**0.019**	1.6	0.507	10.4	**0.042**	-3.0	0.533	0.1	0.960	8.1	0.194	-6.3	0.337
rs6437074		A	G	2.6	**0.020**	1.7	0.483	9.9	0.052	-3.0	0.536	0.1	0.948	7.8	0.210	-6.6	0.316
rs2100053		T	G	2.6	**0.020**	1.6	0.524	10.1	**0.047**	-3.2	0.514	0.1	0.952	7.9	0.204	-6.0	0.358
rs4973055		T	G	2.6	**0.024**	2.1	0.398	9.6	0.063	-3.1	0.526	0.3	0.881	7.8	0.214	-5.0	0.444
rs1947105		A	G	2.6	**0.021**	1.6	0.499	10.1	**0.048**	-3.2	0.508	0.1	0.959	7.7	0.213	-6.0	0.361
rs7589201		A	G	2.6	**0.020**	1.6	0.519	10.2	**0.046**	-3.2	0.507	0.1	0.956	7.8	0.209	-5.9	0.363
rs895430		A	C	2.6	**0.020**	1.6	0.517	10.2	**0.045**	-3.2	0.510	0.1	0.951	7.8	0.208	-5.9	0.367
rs2675971		A	G	2.6	**0.019**	1.6	0.516	10.2	**0.045**	-3.1	0.522	0.2	0.937	7.9	0.206	-5.8	0.375
rs6963258		A	T	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs6964113	cg11554507	C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs4140961*		A	G	1.6	0.147	1.7	0.467	1.9	0.700	7.0	0.147	2.2	0.273	5.8	0.341	13.5	**0.035**
rs7104929*		C	G	?	?	?	?	?	?	?	?	?	?	?	?	?	?
rs4963153		A	G	0.1	0.931	1.7	0.514	-2.4	0.643	-0.6	0.899	0.5	0.795	-3.5	0.588	8.0	0.232
rs7479101*		A	G	0.4	0.766	2.5	0.365	-2.0	0.733	-1.4	0.808	0.6	0.798	-1.7	0.814	-2.1	0.777
rs10902222	cg20225915	T	G	0.5	0.715	1.6	0.557	-2.0	0.727	-2.4	0.668	0.5	0.840	-2.1	0.761	-2.6	0.726
rs10902221		T	C	0.1	0.940	1.4	0.574	-1.2	0.825	-2.1	0.681	0.6	0.773	2.8	0.663	2.9	0.662
rs6597981		A	G	0.2	0.881	1.6	0.530	-0.7	0.890	-1.9	0.714	0.7	0.734	3.4	0.596	3.6	0.593
rs4131364*		A	G	0.1	0.941	1.2	0.622	-0.9	0.862	-3.3	0.511	0.6	0.749	0.0	0.998	0.7	0.915
rs1054252	cg04464446	A	G	-2.5	**0.037**	-1.2	0.651	-11.2	**0.038**	-2.8	0.587	-1.5	0.477	-3.6	0.583	-1.7	0.804
rs11600377*		A	G	-2.5	**0.038**	-0.9	0.741	-11.7	**0.030**	-2.9	0.579	-1.9	0.375	-3.7	0.573	-2.2	0.751

ASM: Allele-specific methylation; A1: Allele 1; A2: Allele 2; All alleles are reported in the forward strand; Effect: Effect sizes are given in units of mm3 per risk allele; *: Significant tagSNPs overcoming 5% FDR, the other SNPs are ASM SNPs in LD with these significant tagSNPs; Underlined allele: Risk allele for ADHD; In bold: Nominally significant p-values; "?": No values retrieved for these SNPs.

Table S11. Copy number variants overlapping with the highlighted genes for ADHD.

Gene	Variant	Type	Sex	Inheritance	Phenotype(s)
ARTN	1:36831723-61417768	Deletion	46XX	Unknown	Hypoplasia of the corpus callosum, Intellectual disability
	1:41126365-45133712	Deletion	46XY	De novo constitutive	Congenital microcephaly
	1:43827703-45107379	Deletion	46XX	De novo constitutive	Autism, Intellectual disability
	1:43870211-45107379	Deletion	46XX	De novo constitutive	Autism, Delayed speech and language development, Intellectual disability
C2orf82	2:216527137-242783384	Duplication	46XX	De novo constitutive	Intellectual disability
	2:217860965-242783384	Deletion	46XX	Unknown	Hypodysplasia of the corpus callosum
	2:228207994-236247812	Deletion	46XY	Unknown	Intellectual disability
	2:231683468-235003677	Duplication	unknown	De novo constitutive	Aggressive behavior, ADHD, Dysgraphia, Self-injurious behavior
	2:231869040-243014630	Duplication	other	De novo constitutive	Aggressive behavior, Dysarthria, Intellectual disability, Microcephaly, Small for gestational age
	2:232756751-238245191	Duplication	46XY	De novo constitutive	Intellectual disability
	2:232763439-242717043	Duplication	46XY	De novo constitutive	Intellectual disability
	2:232831650-239972530	Deletion	46XY	De novo constitutive	Macrocephaly, Mild global developmental delay
	2:232981279-243007359	Deletion	46XX	De novo constitutive	Aggressive behavior, Behavioral abnormality, Failure to thrive, Global developmental delay, Growth delay, Intrauterine growth
	2:233508667-241070353	Duplication	46XX	De novo constitutive	Intellectual disability, Seizures.
	2:233527432-242852625	Duplication	46XY	De novo mosaic	Delayed speech and language development, Global developmental delay, Motor delay
	2:233613670-233742556	Duplication	46XY	Unknown	Cognitive impairment
PIDD1	11:1-18716078	Duplication	46XX	De novo constitutive	Intellectual disability
	11:1-4061418	Duplication	46XX	Maternally inherited, constitutive in mother	Macrocephaly, Postnatal growth retardation
	11:192372-3653923	Triplication	46XX	De novo constitutive	Intellectual disability
	11:196966-3254236	Duplication	46XY	De novo constitutive	Intellectual disability, Small for gestational age
	11:198510-2261562	Duplication	46XX	De novo constitutive	Global developmental delay
	11:203788-3131418	Duplication	46XX	Unknown	Delayed speech and language development, Global developmental delay
	11:210100-2425271	Duplication	46XY	Imbalance arising from a balanced parental rearrangement	Delayed speech and language development, Muscular hypotonia
	11:210300-3363804	Duplication	46XX	De novo constitutive	Autism
	11:353347-872723	Duplication	46XY	Maternally inherited, constitutive in mother	Delayed speech and language development
	11:662317-1157708	Duplication	46XX	Maternally inherited, constitutive in mother	Global developmental delay

Variant: CNV position (build hg19); Source: DECIPHER (v9.27) database; The reported phenotypes have been restricted to those comorbid to ADHD or neurodevelopmental processes.

SUMMARY ARTICLE 2

"Meta-anàlisi de GWAS de dependència de cocaïna: base genètica compartida amb trastorns comòrbids"

La dependència de cocaïna és un trastorn neuropsiquiàtric complex que presenta un elevat grau de comorbiditat amb altres trets psiquiàtrics. Els resultats obtinguts en estudis d'associació suggereixen que les variants genètiques comunes podrien tenir un paper rellevant en la susceptibilitat a la dependència de cocaïna. D'altra banda, cada vegada hi ha més evidències de l'existència de variants genètiques de risc compartides entre trastorns psiquiàtrics. En aquest estudi hem realitzat una meta-anàlisi de dependència de cocaïna emprant dades GWAS de quatre estudis disponibles a la base de dades dbGaP (2.085 casos i 4.293 controls, tots ells d'ascendència europea). Tot i que no s'ha identificat cap variant que superi el llindar de significació GWAS, en l'anàlisi basada en gens (*gene-based analysis*) s'ha identificat associació significativa entre el gen *HISTH2BD* i la dependència de cocaïna, que supera correccions per *False Discovery Rate* del 10%. Aquest gen està al cromosoma 6 en una regió enriquida en gens d'histones, la qual ha estat prèviament associada a l'esquizofrènia. Els SNPs amb menor valor p de la regió, el rs806973 i rs56401801 (P = 3,14 x 10^{-6} i 3,44 x 10^{-6}, respectivament), són eQTLs (*expression Quantitative Trait Loci*) per diferents gens de la regió en múltiples àrees cerebrals. D'altra banda, s'ha identificat correlació genètica entre la dependència de cocaïna i el trastorn per dèficit d'atenció amb hiperactivitat (TDAH), esquizofrènia, trastorn depressiu major (TDM) i personalitat amb tendència a assolir riscos, emprant l'aproximació LDSC (*Linkage Disequilibrium Score Regression*). A més, s'ha vist que tots els fenotips testats permeten predir si un individu és cas o control per la dependència de cocaïna mitjançant una anàlisi PRS (*Polygenic Risk Score*): esquizofrènia (R^2 = 2,28%; P = 1,21 x 10^{-26}), TDAH (R^2 = 1,39%; P = 4,5 x 10^{-17}), personalitat amb tendència a assumir riscos (R^2 = 0,60%; P = 2,7 x 10^{-08}), TDM (R^2 = 1,21%; P = 4,35 x 10^{-15}), comportament agressiu en nens (R^2 = 0,3%; P = 8,8 x 10^{-05}) i personalitat antisocial (R^2 = 1,33%; P = 2,2 x 10^{-16}). Aquesta és la meta-anàlisi de GWAS de dependència de cocaïna més gran publicada fins ara. Tot i les limitacions de l'estudi a causa de la mida mostral limitada, s'han identificat regions potencialment implicades en la dependència de cocaïna i mostrem evidències que hi ha factors genètics de risc comuns entre aquesta patologia i les condicions comòrbides testades.

Reference

Cabana-Domínguez J, Shivalikanjli A, Fernàndez-Castillo N, Cormand B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. Prog Neuropsychopharmacol Biol Psychiatry. 2019 Aug 30;94:109667. doi: 10.1016/j.pnpbp.2019.109667.

Progress in Neuropsychopharmacology & Biological Psychiatry 94 (2019) 109667

Contents lists available at ScienceDirect

Progress in Neuropsychopharmacology & Biological Psychiatry

journal homepage: www.elsevier.com/locate/pnp

Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions

Judit Cabana-Domínguez[a,b,c,d], Anu Shivalikanjli[a,b,c,d], Noèlia Fernàndez-Castillo[a,b,c,d,*,1], Bru Cormand[a,b,c,d,*,1]

[a] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain
[b] Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Spain
[c] Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain
[d] Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Catalonia, Spain

ARTICLE INFO

Keywords:
Cocaine addiction
Cocaine dependence
Genetic correlation
GWAS
Poligenic risk score
Meta-analysis

ABSTRACT

Cocaine dependence is a complex psychiatric disorder that is highly comorbid with other psychiatric traits. Twin and adoption studies suggest that genetic variants contribute substantially to cocaine dependence susceptibility, which has an estimated heritability of 65–79%. Here we performed a meta-analysis of genome-wide association studies of cocaine dependence using four datasets from the dbGaP repository (2085 cases and 4293 controls, all of them selected by their European ancestry). Although no genome-wide significant hits were found in the SNP-based analysis, the gene-based analysis identified *HIST1H2BD* as associated with cocaine-dependence (10% FDR). This gene is located in a region on chromosome 6 enriched in histone-related genes, previously associated with schizophrenia (SCZ). Furthermore, we performed LD Score regression analysis with comorbid conditions and found significant genetic correlations between cocaine dependence and SCZ, ADHD, major depressive disorder (MDD) and risk taking. We also found, through polygenic risk score analysis, that all tested phenotypes are significantly associated with cocaine dependence status: SCZ (R^2 = 2.28%; P = 1.21e-26), ADHD (R^2 = 1.39%; P = 4.5e-17), risk taking (R^2 = 0.60%; P = 2.7e-08), MDD (R^2 = 1.21%; P = 4.35e-15), children's aggressive behavior (R^2 = 0.3%; P = 8.8e-05) and antisocial behavior (R^2 = 1.33%; P = 2.2e-16). To our knowledge, this is the largest reported cocaine dependence GWAS meta-analysis in European-ancestry individuals. We identified suggestive associations in regions that may be related to cocaine dependence and found evidence for shared genetic risk factors between cocaine dependence and several comorbid psychiatric traits. However, the sample size is limited and further studies are needed to confirm these results.

1. Introduction

Cocaine is one of the most used illicit drugs worldwide and its abuse produces serious health problems. In Europe, around 5.2% of adults (from 15 to 64 years old) have tried cocaine (EMCDDA, 2017), but at most 20% will develop addiction (Wagner and Anthony, 2002). This information allows us to estimate the prevalence of cocaine dependence in the European population around 1.1%, similar to the prevalence observed in American populations (Compton et al., 2007).

Cocaine dependence is a complex psychiatric disorder that results from the interaction of environmental and genetic risk factors. It is one of the most heritable psychiatric conditions, with an estimated heritability of 65–79% (Ducci and Goldman, 2012). Although many case-control association studies in candidate genes have been performed, only a few risk variants for cocaine dependence have been identified and replicated so far, such as rs16969968 in the *CHRNA5* gene, encoding the cholinergic receptor nicotinic alpha 5 subunit, and rs806368 in *CNR1*, coding for the cannabinoid receptor 1 (Bühler et al., 2015). To date, only one genome-wide association study (GWAS) on cocaine dependence has been performed in European- and African-American individuals (Gelernter et al., 2014). When combining the two populations, one genome-wide finding was identified in the *FAM53B* gene, using a symptom count approach, but this hit could not be replicated in a subsequent study (Pineda-Cirera et al., 2018).

Several studies have shown that substance use disorders (SUD), and especially cocaine dependence, is highly comorbid with other

* Corresponding authors at: Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 643, edifici Prevosti, 3ª planta, 08028 Barcelona, Catalonia, Spain.
E-mail addresses: noefernandez@ub.edu (N. Fernàndez-Castillo), bcormand@ub.edu (B. Cormand).
[1] Equally contributed.

https://doi.org/10.1016/j.pnpbp.2019.109667
Received 8 February 2019; Received in revised form 7 May 2019
Available online 15 June 2019

psychiatric disorders and related phenotypes like aggressive, antisocial or risk-taking behaviors (Bezinović and Malatestinić, 2009; Hasin and Kilcoyne, 2012). For example, the occurrence of lifetime SUD in patients with schizophrenia (SCZ) is 70–80%, in attention deficit/hyperactivity disorder (ADHD) it is 39.2% and in major depressive disorder (MDD) it is 16.1% (Currie et al., 2005; Piñeiro-Dieguez et al., 2016; Westermeyer, 2006). Conversely, about 81% of substance abuse/dependence patients have at least one comorbid mental disorder: 33% MDD, 11% SCZ and 9% personality disorders (Shantna et al., 2009). Such comorbidity is associated with an increase of severity for all disorders, although it is unclear whether this relationship is causal or the result of shared genetic and/or environmental risk factors. Some studies have started to inspect these relationships using both genetic correlation and polygenic risk score approaches, supporting the hypothesized role of shared genetic risk factors in the lifetime co-occurrence of several psychiatric disorders and SUD (Carey et al., 2016; Du Rietz et al., 2017; Hartz et al., 2017; Reginsson et al., 2018).

Here we performed a GWAS meta-analysis of cocaine dependence in samples with European ancestry using datasets from the dbGaP repository. Then we investigated the shared genetics between cocaine dependence and other psychiatric traits.

2. Experimental procedures

Detailed description of the materials and methods used, as well as supplementary figures, can be found in the Supplementary Information.

2.1. Subjects

The case-control GWAS meta-analysis was performed using four datasets from the dbGaP repository (https://www.ncbi.nlm.nih.gov/gap) under the project 17,170 (Table 1). All cases used met DSM-IV criteria for cocaine dependence, although most of them are also dependent to other drugs of abuse. Diagnoses for schizophrenia, bipolar affective disorder or other major psychotic illnesses or gross cognitive impairment were exclusion criteria for all samples except for SAGE. Drug abuse or dependence were discarded only in controls of the SAGE sample, whereas in the other studies general population individuals were used as controls (see Supplementary Information for detailed description of subjects used in this study).

Since samples 2–4 did not have enough controls to perform the association studies, we used controls from other datasets, strategy previously followed by others (Johnson et al., 2016). In all cases, patients and controls were from the same geographic area, they were genotyped with the same array and using the same genome assembly. In order to reduce bias, we merged controls with cases prior to quality control (QC) and imputation (Mitchell et al., 2014; Uh et al., 2012). The analyses performed to control for population stratification is summarized in Fig. S1.

The study was approved by the ethics committee of our institution, in accordance with the Helsinki Declaration and with the dbGaP protocols.

2.2. Quality control and association analyses

Prior to analysis, stringent QC was performed on the genotyped markers and individuals in each sample using the standardized pipeline "Ricopili" (Ripke, 2014). Very briefly, subjects and SNPs were selected according to "Ricopili" default parameters: SNP and subject call rate > 0.98, SNP Hardy-Weinberg equilibrium ($P > 1e-06$ in controls or $P > 1e-10$ in cases), autosomal heterozygosity deviation ($|F_{het}| < 0.2$) and sex check based on X chromosome heterozygosity. European-ancestry individuals were selected based on principal component analysis (PCA): PC1 and PC2 were used to define a genetic homogenous population, excluding individuals with PC values greater than three standard deviations from the reference population (European individuals from 1000 Genomes Project Phase 3 (1KGP3)). Related individuals and genetic outliers were excluded. A permutation test for between-group IBS differences with fixed 10,000 permutations was performed to discard population stratification between cases and controls (T1 p-value < 0.05).

After QC, non-genotyped markers were imputed using the European individuals from the 1KGP3 reference panel in MINIMAC3 (https://genome.sph.umich.edu/wiki/Minimac3).

In order to identify overlapping or related individuals across all datasets, we ran the "Ricopili" PCA module considering the four samples together, and one individual of each pair was excluded.

For each sample, case-control GWAS was conducted using logistic regression under the additive model in PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/). The 10 firsts PCs were included as covariates to correct for population stratification, and only variants with imputation INFO score > 0.8 and minor allele frequency (MAF) > 0.01 were considered. In all samples the genomic inflation factor (λ) was lower than 1.05.

2.3. GWAS meta-analysis

In total, 2085 cases and 4293 controls were meta-analyzed using an inverse-variance weighted fixed effects model in METAL (http://csg.sph.umich.edu//abecasis/Metal/)(Willer et al., 2010). Association results were considered only for variants with an effective sample size [$N = 2/((1/Ncases) + (1/Ncontrols))$] > 70% of the full meta-analysis. Heterogeneity across studies was tested with the Cochran's Q test and quantified with the I^2 heterogeneity index, implemented in METAL.

Manhattan plot and Q-Q plot from each sample and the meta-analysis results were generated with the library "qqman" implemented in R.

2.4. LD score intercept evaluation

LD score (LDSC) regression analysis was used to calculate LDSC intercept by regressing the chi-square statistics from GWAS against LD scores for each SNP (downloaded from GitHub website, https://github.com/bulik/ldsc) (Bulik-Sullivan et al., 2015b).

2.5. SNP heritability

The proportion of phenotypic variance explained by common SNPs

Table 1
Description of dbGaP samples used in the cocaine dependence GWAS meta-analysis.

Sample	N cases	%F cases	N controls	%F controls	Genotyping chip	dbGaP code
Sample 1 (SAGE)	468	39.3%	1284	69.9%	Illumina ILMN_Human-1	phs000092.v1.p1 (cases/controls)
Sample 2	609	45.9%	410	50.2%	Illumina HumanOmni1-Quad_v1-0_B	phs000952.v1.p1 (cases/controls) + phs000179.v5.p2 (controls)
Sample 3	504	44.3%	1190	62%	Illumina Human660W-Quad_v1_A	phs000277.v1.p1 (cases/controls) + phs000170.v2.p1 (controls)
Sample 4	504	36.7%	1409	41.9%	Illumina HumanOmni1-Quad_v1-0_B	phs000425.v1.p1 (cases/controls) + phs000524.v1.p1 (controls)
Total	2085		4293			

%F = percentage of females

in the liability scale was assessed using two different methodologies: LDSC 1.0.0 (https://github.com/bulik/ldsc/) and the Genome-based REstricted Maximum Likelihood analysis implemented in the tool Genome-wide Complex Trait Analysis (GCTA-GREML; https://cnsgenomics.com/software/gcta/#Overview) (Lee et al., 2011). In both analyses, a population prevalence for cocaine dependence of 1.1% was considered (Compton et al., 2007). The GCTA-GREML analysis was adjusted for the 10 first PCs and a dummy variable indicating genotyping-study.

Partitioned heritability analysis was performed using LDSC 1.0.0 based on 24 functional overlapping annotations described previously (Finucane et al., 2015). Enrichment in the heritability of a functional category was defined as the proportion of SNP heritability explained divided by the proportion of SNPs. The threshold for significance was calculated using the Bonferroni correction for multiple testing ($P < 2e\text{-}03$).

2.6. Functional annotation of risk SNPs

SNPs were functionally annotated using FUMA (http://fuma.ctglab.nl/)(Watanabe et al., 2017). FUMA define lead SNPs as signals that are significantly associated with the phenotype of interest (we considered suggestive associations ($P < 1e\text{-}05$)) and independent to each other ($r^2 < 0.1$). For each lead SNP, a "Genomic risk locus" is defined, including all independent signals that are physically close or overlapping in a single locus. The variants located in a "Genomic risk locus" were explored considering the following functional annotations: eQTL (from GTEx v6/v7 and BRAINEAC), CADD_v1.3, ANNOVAR, RegulomeDB_v1.1, 15-core chromatin state and GWAS-catalog e91.

2.7. Gene-based and gene-set association analyses

Gene-based and gene-set association analyses were performed with MAGMA 1.05b (Willer et al., 2010) using the summary statistics from the cocaine dependence GWAS meta-analysis. For gene-based analysis, the SNP-wise mean model was used as the statistic test, considering *p*-values for SNPs located within the transcribed region. For multiple testing corrections, 10% False Discovery Rate (FDR) was considered.

Gene-set analysis was used to test for enrichment in association signals in genes belonging to specific biological pathways or processes. We performed a competitive test using: "All Canonical Pathways" (1329 gene sets), "GO" (4436 gene sets) and "BioCarta" (217 gene sets) provided by MsigDB 5.1 (https://software.broadinstitute.org/gsea/msigdb/)(Subramanian et al., 2005). Multiple testing corrections were applied to each gene set separately. When gene sets are strongly overlapping, the Bonferroni correction can be quite conservative, and for that reason, we used an empirical multiple testing correction implemented in MAGMA, based on a permutation procedure.

2.8. Shared genetic factors between cocaine dependence and comorbid conditions

2.8.1. Subjects

We studied shared genetic factors between cocaine dependence and six previously described comorbid conditions using publicly available GWAS summary statistics of SCZ, ADHD, MDD, children's aggressive behavior, antisocial behavior and risk-taking behavior (Table 2). Summary statistics from the vitamin D levels GWAS of the UK Biobank was used as a negative control.

2.8.2. LDSC genetic correlation

Genetic correlation between cocaine dependence and the six selected comorbid disorders/traits was calculated using LDSC 1.0.0 (Bulik-Sullivan et al., 2015a). Summary statistics from all samples and pre-computed LD scores from HapMap3_SNPs calculated on 378 phased European-ancestry individuals from 1KGP3 were used (available at

Table 2
Description of samples used to inspect shared genetic factors between cocaine dependence and comorbid conditions.

GWAS	Abbreviation	N individuals	Reference	Website
Schizophrenia European meta-analysis	SCZ	32,405 cases and 42,221 controls 1235 parent affected-offspring trios	Ripke et al., 2014	Psychiatric Genomics Consortium (PGC): https://www.med.unc.edu/pgc/results-and-downloads
Attention-deficit/hyperactivity disorder European meta-analysis	ADHD	19,099 cases and 34,194 controls	Demontis et al., 2019	
Major depressive disorder meta-analysis	MDD	59,851 cases and 113,154 controls	Wray et al., 2018	
Children's aggressive behavior GWAS from EAGLE	Child-Aggre	18,988 individuals	Pappa et al., 2016	
Antisocial behavior meta-analysis	ASB	16,400 individuals	Tielbeek et al., 2017	BroadABC: http://broadabc.ctglab.nl/summary_statistics
Risk-taking behavior dataset	RT	325,821 individuals		UK Biobank: https://sites.google.com/broadinstitute.org/ukbbgwasresults/home?authuser=0

https://github.com/buligk/ldsc). After Bonferroni correction, the significance threshold was adjusted to $P < 7.1e-03$.

Furthermore, the genetic correlation of cocaine dependence with other traits available at LD Hub (http://ldsc.broadinstitute.org/ldhub/) (Zheng et al., 2017) was evaluated (Bonferroni correction threshold $P < 7.2e-05$).

2.8.3. Polygenic risk scores for cocaine dependence

Poligenic risk score (PRS) analysis was performed using the PRSice 2.1.0 software (https://github.com/choishingwan/PRSice) (Euesden et al., 2015).

The four cocaine dependence datasets were merged using PLINK v1.9 and used as a target sample. After merging, strict QC was applied resulting in 5,957,307 SNPs in 2083 cases and 4287 controls. To assess population stratification PCA was performed, and the 10 first PCs and a dummy variable indicating genotyping-study was included in the PRS analysis as a covariate.

Summary statistics of the seven phenotypes described above were used as discovery samples, which were clumped ($r^2 < 0.1$ in a 250-kb window) to remove SNPs in linkage disequilibrium (LD). Then, PRSs were estimated for each discovery sample using a wild range of meta-analysis *p*-value thresholds (P_T) between $P_T = 1e-04$ and $P_T = 1$ at increments of 5e-05. For each P_T, the proportion of variance explained (R^2) by each discovery sample was computed by comparing the full model (PRS + covariates) score to a reduced model (covariates only). The reported R^2 value is the difference between R^2 from the two models. For quantitative traits we performed linear regression analysis, and for qualitative traits we used logistic regression and *Nagelkerke's pseudo-R^2* values are shown.

As recommended, we used the significance threshold of $P = 0.004$ (Euesden et al., 2015). Bonferroni correction was applied considering the seven tested phenotypes ($P < 5.7e-04$).

3. Results

3.1. GWAS results

We performed a GWAS meta-analysis of cocaine dependence using four datasets from the dbGaP repository. In total, we meta-analized 9,290,362 common genetic variants in 2085 cases and 4293 controls of European ancestry. No marker demonstrated significant heterogeneity between studies (Fig. S2 and Table S1). The Q-Q plot (Fig. 1A) displayed a λ of 1.06, comparable to other GWAS. The LDSC analyses estimated an intercept of 1.01 (SE = 0.006; $P = 0.1$), not significantly > 1, discarding residual population stratification or cryptic relatedness (Bulik-Sullivan et al., 2015b).

None of the analyzed variants reached the threshold for genome-wide significance ($P < 5e-08$) in the SNP-based analysis, although we identified several suggestive associations ($P < 1e-05$) (Table S1; Fig. 1B).

3.2. Polygenic architecture of cocaine dependence

We applied two approaches to assess the proportion of phenotypic variance explained by common SNPs. For LDSC, the estimated SNP heritability in liability scale was $h^2_{snp} = 0.30$ (SE = 0.06; $P = 2.4e-07$), and for GCTA-GREML $h^2_{snp} = 0.27$ (SE = 0.03, $P < 0.01$). Then we performed partitioned heritability analysis on LDSC based on functional genomic categories and found significant enrichment in the heritability by SNPs located in intronic regions (enrichment = 2.17; SE = 0.45; $P = 1.2e-03$), and a nominal result for conserved genomic regions (enrichment = 23.63; SE = 8.57; $P = 4e-03$) (Fig. S3).

3.3. Functional annotation of risk SNPs

To identify potentially interesting regions with FUMA we considered the SNPs showing suggestive associations ($P < 1e-05$), as the SNP-based analysis did not reveal genome-wide significant hits ($P < 5e-08$). We identified 23 lead SNPs which correspond to 22 genomic risk loci including 112 genes (Table 3). Interestingly, the risk locus located on chromosome 6 overlaps with a genomic region previously associated with schizophrenia. This region is defined by two lead SNPs (rs806973 and rs56401801, GWAS $P = 3.1e-06$ and 3.4e-06, respectively) and it includes 77 genes and 458 nominally associated SNPs. Moreover, most of the SNPs in this region (447) are brain eQTLs for at least one member of a small group of 12 genes, including *BTN3A2*, *HIST1H2AK*, *ZSCAN31*, *PRSS16* and *ZNF184* (Figs. 2 and S4).

3.4. Gene-based and gene-set analyses

The gene-based analysis mapped 3,418,270 SNPs from the GWAS meta-analysis to 18,069 protein-coding genes (Fig. S5 and Table S2). The *HIST1H2BD* gene, located in a genomic region on chromosome 6 that showed a suggestive association in the SNP-based analysis, displayed a significant gene-wise association with cocaine dependence (10% FDR), although it did not overcome the Bonferroni correction for multiple testing. Then we performed competitive gene-set tests for all BioCarta, GO and Canonical Pathways. No gene sets attained a significant association with cocaine dependence after correction for multiple testing (Tables S3–5), although the BioCarta immunity pathway "BIOCARTA TNFR2 PATHWAY" showed a trend (uncorrected $P = 5.38e-04$, corrected $P = 0.09$), being also the most significantly associated canonical pathway gene set. Furthermore, from the 10 GO gene sets with lower *p*-values, seven were related to synapse organization, glutamatergic neurotransmission and brain functions.

3.5. Cocaine dependence and shared genetic factors with comorbid conditions

Cocaine dependence is highly comorbid with other psychiatric disorders like SCZ, ADHD and MDD, and also with other phenotypes like aggressive behavior in children, antisocial behavior or risk taking. In order to investigate whether these phenotypic correlations are genetically mirrored, we performed a genetic correlation analysis using LDSC analysis and found significant genetic correlations ($P < 7.1e-03$) between cocaine dependence and SCZ (rg = 0.2; SE = 0.05; $P = 1e-04$), ADHD (rg = 0.5; SE = 0.08; $P = 1.6e-09$), MDD (rg = 0.4; SE = 0.08; $P = 6.6e-07$) and risk taking (rg = 0.35; SE = 0.06; $P = 9.1e-08$) (Fig. 3A). No significant correlations were found with children's aggression or antisocial behavior, nor with a negative control (Vitamin D levels).

On the other hand, we tested genetic correlations between cocaine dependence and all the GWAS summary statistics publicly available in the LD Hub. From the 690 tested traits, 109 demonstrated significant correlations after applying the Bonferroni correction for multiple testing, including negative correlations with educational achievements (e.g. college completion) or with reproductive traits (e.g. age at first child) and positive correlations with familiar situation (e.g. tobacco smoke exposure at home) or with several psychological or psychiatric traits like neuroticism, depressive symptoms or loneliness (Fig. S6 and Table S6). The high number of significant associations may be explained by the high redundancy of the phenotypes of the LD Hub.

We also investigated the shared genetic etiology between cocaine dependence and comorbid phenotypes through PRS analysis, and tested whether these phenotypes are associated with cocaine dependence status. For all the discovery samples tested, PRS was significantly associated with cocaine dependence: SCZ (pseudo-$R^2 = 2.28\%$, $P_T = 0.4911$, $P = 1.21e-26$), ADHD (pseudo-$R^2 = 1.39\%$, $P_T = 0.04275$, $P = 4.5e-17$), antisocial behavior (pseudo-$R^2 = 1.33\%$, $P_T = 0.4055$, $P = 2.2e-16$), MDD (pseudo-$R^2 = 1.21\%$, $P_T = 0.0129$, $P = 4.35e-15$), risk taking ($R^2 = 0.60\%$, $P_T = 0.00135$, $P = 2.7e-08$) and children's aggressive behavior ($R^2 = 0.30\%$, $P_T = 0.3552$,

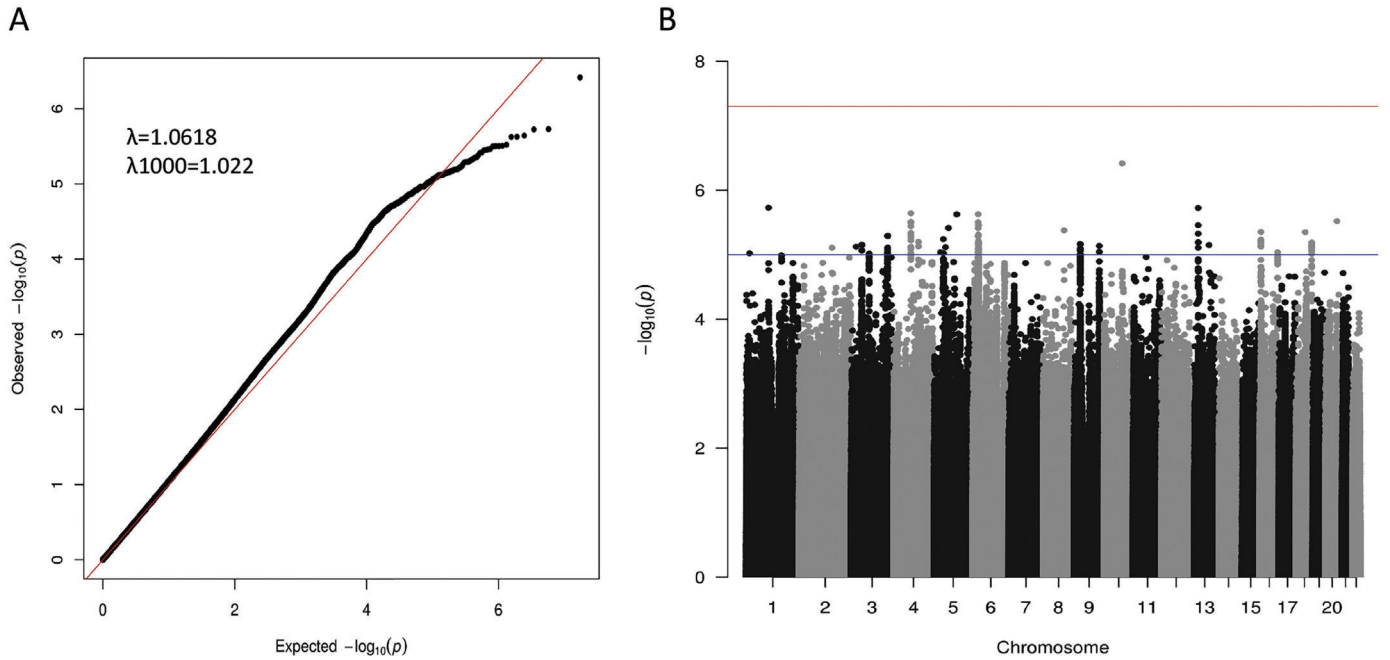


Fig. 1. Results from the GWAS meta-analysis on cocaine dependence. A) Q-Q plot and B) Manhattan plot. Red line: threshold for genome-wide significance ($P < 5e-08$). Blue line: threshold for suggestive associations ($P < 1e-05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
FUMA analysis of genetic risk loci for cocaine dependence as identified from the GWAS meta-analysis.

Genomic locus	Position[a]	Lead SNP	SNP position	GWAS P-value	nSNPs	nGenes	Genes
1	chr1:15,916,714–16,049,893	rs148069235	15,950,064	9.52e-06	236	6	*DNAJC16, AGMAT, DDI2, RSC1A1, PLEKHM2, SLC25A34*
2	chr1:104,966,872–104,966,872	rs72685414	104,966,872	1.87e-06	1	0	
3	chr3:50,615,472–51,356,217	rs148179194	50,798,652	7.03e-06	152	3	*C3orf18, HEMK1, DOCK3*
4	chr3:84,862,871–84,961,810	rs6767407	84,955,841	9.61e-06	27	0	
5	chr3:172,170,104–172,214,001	rs57361543	172,212,144	5.09e-06	28	2	*GHSR, TNFSF10*
6	chr4:82,943,149–83,005,137	rs7675557	82,970,816	2.28e-06	16	1	*RASGEF1B*
7	chr4:117,560,979–117,607,439	rs67769911	117,607,070	6.33e-06	18	1	*MIR1973*
8	chr5:30,884,915–31,001,774	rs62357000	30,884,915	9.20e-06	4	1	*CDH6*
9	chr5:44,051,305–44,130,776	rs4410642	44,129,423	9.42e-06	61	0	
10	chr5:54,436,897–54,754,893	rs334878	54,519,878	7.67e-06	90	9	*ESM1, CDC20B, GPX8, MCIDAS, CCNO, DHX29, SKIV2L2, PLPP1, DDX4*
11	chr5:106,788,817–106,796,026	rs71575441	106,788,817	2.37e-06	2	1	*EFNA5*
12	chr6:26,148,326–28,301,195	rs806973; rs56401801	26,148,326; 27,301,762	3.14e-06; 3.44e-06	458	77	Locus too broad
13	chr8:99,193,765–99,226,821	rs4734387	99,193,765	4.20e-06	35	1	*NIPAL2*
14	chr9:28,890,331–28,993,271	rs35735220	28,963,289	6.86e-06	113	0	
15	chr9:118,176,789–118,273,407	rs10121366	118,244,022	7.33e-06	54	0	
16	chr13:36,947,826–37,019,186	rs79309473	36,972,202	1.89e-06	26	5	*SOHLH2, SPG20, SPG20-AS1, CCNA1, SERTM1*
17	chr13:88,150,884–88,150,884	rs7332726	88,150,884	7.10e-06	1	0	
18	chr16:6,654,017–6,695,032	rs112252907	6,675,141	4.43e-06	42	1	*RBFOX1*
19	chr16:84,581,684–84,590,225	rs247831	84,581,893	9.17e-06	8	2	*TLDC1, COTL1*
20	chr18:43,206,985–43,231,622	rs1484873	43,206,985	4.45e-06	14	1	*SLC14A2*
21	chr18:73,743,937–73,775,398	rs73973283	73,757,906	6.46e-06	88	0	
22	chr20:54,516,338–54,516,338	rs11086525	54,516,338	3.01e-06	1	0	

[a] Gene coordinates based on GRCh37/hg19. NSNPs: Number of nominally associated SNPs per genomic locus.

$P = 8.8e-05$) (Figs. 3B and S7). In all cases, the quantile plot shows the positive nature of this relationship as cocaine dependence increases with greater polygenic load at each discovery dataset (Fig. S8). As expected, the negative control based on vitamin D levels was not associated with cocaine dependence ($R^2 = 0.07\%$, $P_T = 0.03075$, $P = 0.06$).

4. Discussion

To our knowledge, this is the largest GWAS meta-analysis on cocaine dependence performed so far in individuals with European ancestry, although the sample size is still limited. No genome-wide significant (GWS) signals emerged from the SNP-based analysis, but the gene-based study identified *HIST1H2BD* as significantly associated with cocaine dependence. This gene is located in a region of chromosome 6 enriched in immune system and histone-related genes. These pathways

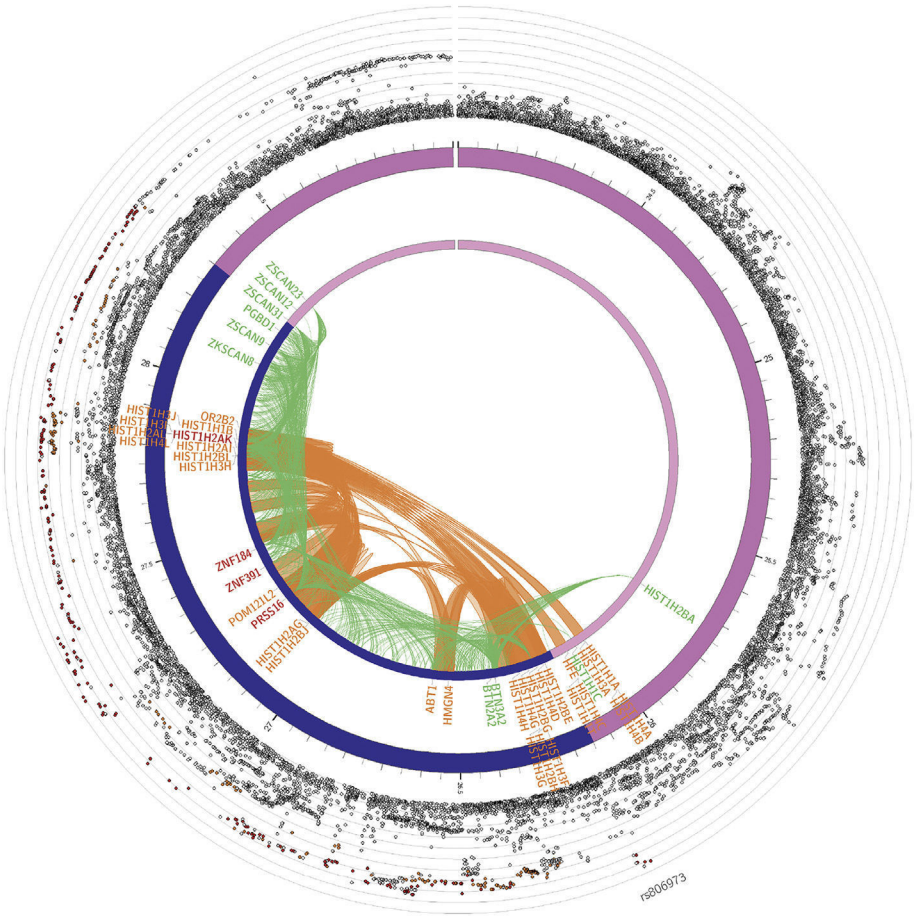


Fig. 2. Circo-plot from genomic risk loci on chromosome 6 by FUMA. The most outer layer is the Manhattan plot (only SNPs with $P < 0.05$ are displayed). SNPs in genomic risk loci are colour-coded as a function of their maximum r^2 to the lead SNPs in the locus, as follows: red ($r^2 > 0.8$), orange ($r^2 > 0.6$), green ($r^2 > 0.4$), blue ($r^2 > 0.2$) and grey ($r^2 \leq 0.2$). The rs ID of the top SNPs in the risk locus is displayed in the most outer layer. Y-axis is ranged between 0 to the maximum -$\log_{10}$(p-value) of the SNPs. The second layer is the chromosome ring, with the genomic risk locus highlighted in blue. Here genes are mapped by chromatin interactions (orange) or eQTLs (green). When the gene is mapped by both, it is colored in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

have previously been associated with other psychiatric disorders like schizophrenia, major depressive disorder and bipolar disorder (O'Dushlaine, 2015). Despite the lack of GWS findings in this region, we identified many subthreshold variants. Based on the omnigenic model that has recently been proposed for complex disorders, these variants could point at regulatory elements of core genes (Boyle et al., 2017; Wray et al., 2018) and, therefore, contribute to the susceptibility to cocaine dependence, as most of them are brain eQTLs. Interestingly, this region overlaps with the genomic region that has been most consistently associated with schizophrenia. Indeed, it contains five SNPs (rs16897515, rs17693963, rs34706883, rs41266839 and rs55834529) nominally associated with cocaine dependence ($P < 1e-04$) that had previously been associated with schizophrenia and with bipolar disorder, being the risk allele the same in all studies. The genetic variant most consistently associated with schizophrenia is rs17693963, reported in five different studies (Bergen et al., 2012; Ripke et al., 2011; 2014; Ruderfer et al., 2014; Sleiman et al., 2013), which is a brain eQTL for *PRSS16*, *ZSCAN9, ZNF184* and *ZSCAN31*. Furthermore, a transcriptomic study performed in lymphoblastoid cell lines of 413 individuals with schizophrenia and 446 controls found that the top differentially expressed genes are located in this region (e.g. *HIST1H2BD*, *HIST1H2BC*, *HIST1H2BH*, *HIST1H2BG* and *HIST1H4K*) (Sanders et al., 2013).

Cocaine dependence is a highly heritable disorder (around 65–79% (Ducci and Goldman, 2012)). Our analyses estimate that SNPs significantly capture more than one third of cocaine dependence heritability, as estimated using two different methods (LDSC $h^2_{SNP} = 0.30$; GCTA-GREML $h^2_{SNP} = 0.27$). Interestingly, studies with comparable sample sizes obtained similar results for cocaine dependence ($h^2_{TOTAL} = 0.65–0.79$; $h^2_{SNP} = 0.28$) (Huggett and Stallings, 2019), alcohol dependence ($h^2_{TOTAL} = 0.55–0.69$; $h^2_{SNP} = 0.33$) (Mbarek et al., 2015) and for other psychiatric disorders like ADHD ($h^2_{TOTAL} = 0.77–0.88$; $h^2_{SNP} = 0.28$) and schizophrenia ($h^2_{TOTAL} = 0.7–0.8$; $h^2_{SNP} = 0.28$) (Cross-Disorder Group of the PGC et al., 2013). Increasing sample size has resulted in lower SNP-heritability estimates in some studies (e.g. alcohol dependence, $h^2_{SNP} = 0.09$ (Walters et al., 2018) and MDD, $h^2_{SNP} = 0.12$ (Wray et al., 2018)), but in others they remained similar (e.g. schizophrenia, $h^2_{SNP} = 0.26$ (Ripke et al., 2014) and ADHD, $h^2_{SNP} = 0.22$ (Demontis et al., 2019)). For this reason, larger samples are needed to confirm our results.

It is well known that most psychiatric disorders are highly comorbid. About 73.4% of cocaine abuse/dependence patients have comorbid mental disorders: 49.7% have personality disorders (e.g. 5.3% schizoid and 17% antisocial personality) and 61.5% other mental disorders (e.g. 23.4% MDD and 20.5% anxiety) (Arias et al., 2013). However, the reasons for these covariations remain largely unknown. We investigated whether the phenotypic correlations between cocaine dependence and six comorbid psychiatric traits are genetically mirrored by performing genetic correlation analyses using LDSC. For the first time we found significant genetic correlation with ADHD, SCZ, MDD and risk-taking behavior, although these results should be taken with caution and need to be followed up in a larger sample of cocaine-dependent individuals. Furthermore, we used the PRS method that, in contrast to LDSC, uses individual-level SNP data, resulting in higher statistical power and allowing for direct testing of interaction effects. According to our results, all the tested comorbid conditions are associated with cocaine dependence status, suggesting that cocaine dependence is more likely in individuals with many risk alleles for the tested conditions than in those with fewer risk alleles. To our knowledge, this is the first report of a shared genetic etiology between cocaine dependence and ADHD, antisocial behavior, risk-taking behavior and children's aggressive behavior based on genome-wide data. Previous

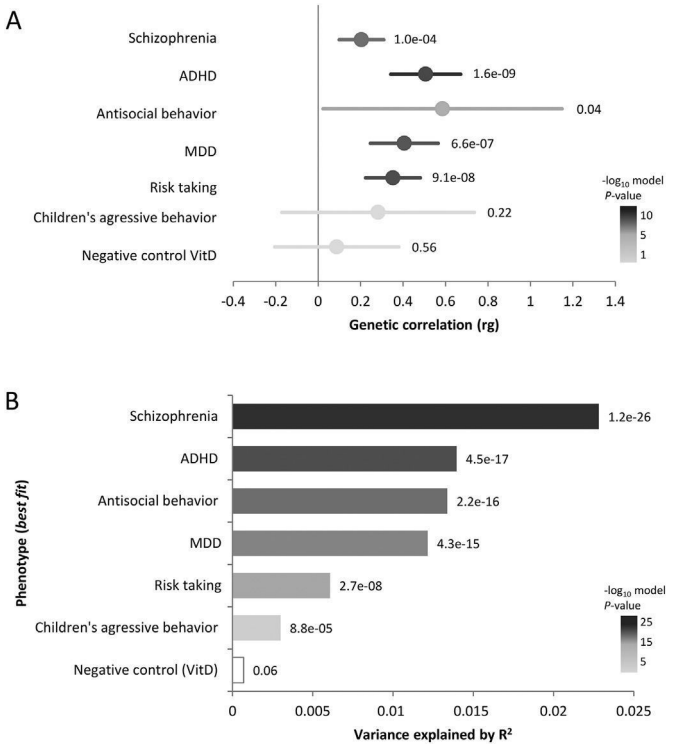


Fig. 3. Shared genetic factors between cocaine dependence and comorbid conditions. A) Results from LD Score (LDSC) regression analysis showing genetic correlation (rg) between cocaine dependence and several traits. Error bars indicate 95% confidence limits. The significance threshold was set to $P < 7.1e\text{-}03$. B) Best fit results from Polygenic Risk Score (PRS) analysis for each tested phenotype. Values displayed next to each bar represent the p-value for significance for the most predictive models. The significance threshold was set at $P < 5.7e\text{-}04$.

studies have reported significant PRS associations between cocaine dependence and SCZ or MDD (Carey et al., 2016; Hartz et al., 2017; Reginsson et al., 2018), and also between SUD and other psychiatric disorders (Du Rietz et al., 2017; Gurriarán et al., 2018), although our study used the largest sample of cocaine dependence for this type of analysis so far. This correlation can reflect biological pleiotropy, where similar genetic mechanisms influence more than one trait, or mediated pleiotropy, where one phenotype is causally related to a second phenotype, so that the variants associated with this phenotype are indirectly associated with the second one. A recent study performed a GWAS meta-analysis of eight psychiatric disorders, and found that 75% of the LD-independent associated regions (109 out of 146) were associated with more than one disorder (Cross-Disorder Group of the PGC et al., 2019). Supporting the idea that the high comorbidity of psychiatric disorders can be explained, in part, because some of the genetic risk factors are shared among them (Brainstorm Consortium et al., 2018; Martin et al., 2018).

An important and controversial consideration in association studies for substance dependence is the selection of the control sample. Some studies use control individuals that have been exposed to the drug at least once (Agrawal et al., 2018; Gelernter et al., 2014). In this case, the association study would capture the predisposing genetic component involved in the transition from use to addiction, but not the ones related to the initiation in the consumption (e.g. impulsivity or risk-taking behavior). Our study shows evidence of shared genetic risk factors between cocaine dependence and risk-taking behavior, estimating a high genetic correlation (rg = 0.35) and identifying a significant association also in the PRS analysis. For the above reasons, we have used unscreened controls from the general population in our study (except for the SAGE controls, where dependence was discarded). It is also true that this approach could eventually dilute positive findings due to the presence of some cases in the control sample. However, based on the prevalence of cocaine dependence in the general population (about 1.1%), the probability of false negative results due to this effect is low. Similar controls were used in other GWAS of drug addiction (Ikeda et al., 2013; Johnson et al., 2016).

This study has strengths and limitations that need some discussion. We performed a GWAS meta-analysis using all the cocaine dependence datasets available at the dbGaP repository, but we could not find any GWS association at the SNP-based level, as expected given the limited sample size, with a total of around 6000 subjects, one third of them cases. However, these data allowed us to detect genetic correlations between cocaine dependence and several co-occurring conditions. Also, we calculated polygenic risk scores that explain a small fraction of the variance in the target phenotype, with figures that are similar to those obtained for other pairs of psychiatric conditions. To obtain a more comprehensive picture of the etiological overlap between cocaine dependence and comorbid conditions, larger studies will be needed, and other genetic factors should be included in the analyses (e.g. CNVs and rare variants). It is important to note that the high comorbidity across the tested traits could influence our results on genetic correlation and PRS. However, several studies have shown that this high co-occurrence is due, at least in part, to shared genetic risk factors (Brainstorm Consortium et al., 2018; Martin et al., 2018). On the other hand, some of the dbGaP datasets used included only cases but not control individuals. For this reason, we used controls from other datasets that can introduce potential biases into the experimental design. Nevertheless, we performed very strict quality controls to avoid population stratification: the paired case and control samples were genotyped with the same platform and are from the same geographical area, the merging of the different datasets was performed prior to quality control and imputation, and after that a permutation test was performed to discard population stratification (Mitchell et al., 2014). Population admixture is one of the main sources of false positive findings in association studies. For this reason, we performed ancestry selection using genetic data, which allowed us to discard a relatively high number of individuals with non-European ancestry (ranging from 8 to 30% depending on the dataset). This highlights the importance of using genetic data rather than self-reported ancestry, as previously discussed by others (Shraga et al., 2017). Moreover, the LDSC analyses confirmed that most of the observed inflation ($\lambda = 1.06$) can be attributed to polygenicity rather than to residual population stratification or cryptic relatedness (Bulik-Sullivan et al., 2015b). Finally, the disease phenotype has not been excluded in most of the control samples, which may potentially dilute positive findings in the association study (but not lead to false positive results).

5. Conclusion

In conclusion, we reported the largest cocaine dependence GWAS meta-analysis on individuals of European ancestry, even though no GWS hits could be identified. Enlarging the sample size of this study would increase the chances to detect significant associations. However, the fact that our analyses highlighted a region on chromosome 6 that also pops-up in several schizophrenia GWAS supports the idea of shared genetic risk factors in these two comorbid disorders. This is in line with the significant results derived from the genetic correlation and PRS analyses in our study and in others. Finally, it would also be interesting to investigate the genetic pathways and neurobiological mechanisms that underlie the genetic overlap between cocaine dependence and comorbid traits.

Role of funding source

Major financial support for this research was received by BC from the Spanish 'Ministerio de Ciencia, Innovación y Universidades'

(RTI2018-100968-B-100), 'Ministerio de Economía y Competitividad' (SAF2015-68341-R) and AGAUR, 'Generalitat de Catalunya' (2017-SGR-738). The research leading to these results has also received funding from the European Union Seventh Framework Program [FP7/2007-2013] under grant agreement n° 602805 (Aggressotype) and from the European Union H2020 Program [H2020/2014-2020] under grant agreements n° 667302 (CoCA), Eat2beNICE (728018) and 643051 (MiND). JC-D was supported by 'Generalitat de Catalunya'(2015 FI_B 00448) and by the H2020 CoCA project, AS by the H2020 MiND project and NF-C by 'Centro de Investigación Biomédica en Red de Enfermedades Raras' (CIBERER) and by an EMBO short-term fellowship (ASTF 573-2016).

Ethical statement

The study was approved by the ethics committee of our institution, in accordance with the Helsinki Declaration and with the dbGaP protocols.

Contributors

JC-D and AS contributed to the acquisition of genotype data; JC-D performed the GWAS meta-analysis and secondary analyses with the assistance of AS; JC-D prepared the first draft of the manuscript and all figures and tables; BC and NF-C coordinated the study and supervised the manuscript preparation. All authors contributed to the design of the study and approved the final manuscript.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgments

We are really thankful to Paula Rovira, Marta Ribasés, Nina Roth Mota, Concepció Arenas and Oscar Lao for their support with the GWAS analysis. We are also very grateful to the developers of the software tools used, especially Stephan Ripke and Raymond Walters.

The authors acknowledge the contribution of data from dbGAP accessed through project number 17170: the Study of Addiction: Genetics and Environment (SAGE), one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI (supported by U01 HG004422, U01 HG004446, U01HG004438, HHSN268200782096C; and support for collection of datasets and samples was provided by the COGA; U10 AA008401, COGEND; P01 CA089392 and FSCD; R01 DA013423); GWAS of Heroin Dependence (supported by R01DA17305); GWAS of Cocaine Dependence in Two Populations (supported by RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535); CRIC Study (supported by U01DK060990, U01DK060984, U01DK061022, U01DK061021, U01DK061028, U01DK60980, U01DK060963 and U01DK060902); CIDR-Gelernter Study, a genome-wide association study funded as part of the Genetics of Alcohol Dependence in American Populations (supported by U01HG004438 and HHSN268200782096C); the COPDGene study (supported by U01HL089856 and U01HL089897, and the COPD Foundation through contributions made by an Industry Advisory Board comprised of Pfizer, AstraZeneca, Boehringer Ingelheim, Novartis, and Sunovion); and Personalized Medicine Research Project (PMRP) (supported by U01HG004608, U01HG004438 and U01HG004603).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pnpbp.2019.109667.

References

Agrawal, A., Chou, Y.-L., Carey, C.E., Baranger, D.A.A., Zhang, B., Sherva, R., Wetherill, L., Kapoor, M., Wang, J.-C., Bertelsen, S., Anokhin, A.P., Hesselbrock, V., Kramer, J., Lynskey, M.T., Meyers, J.L., Nurnberger, J.I., Rice, J.P., Tischfield, J., Bierut, L.J., Degenhardt, L., Farrer, L.A., Gelernter, J., Hariri, A.R., Heath, A.C., Kranzler, H.R., Madden, P.A.F., Martin, N.G., Montgomery, G.W., Porjesz, B., Wang, T., Whitfield, J.B., Edenberg, H.J., Foroud, T., Goate, A.M., Bogdan, R., Nelson, E.C., 2018. Genome-wide association study identifies a novel locus for cannabis dependence. Mol. Psychiatry 23, 1293–1302. https://doi.org/10.1038/mp.2017.200.

Arias, F., Szerman, N., Vega, P., Mesias, B., Basurte, I., Morant, C., Ochoa, E., Poyo, F., Babin, F., 2013. Abuso o dependencia a la cocaína y otros trastornos psiquiátricos. Estudio Madrid sobre la prevalencia de la patología dual. Rev. Psiquiatr. Salud Ment. 6, 121–128. https://doi.org/10.1016/j.rpsm.2012.09.002.

Bergen, S.E., O'Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., Moran, J.L., Chambert, K.D., Handsaker, R.E., Backlund, L., Ösby, U., McCarroll, S., Landen, M., Scolnick, E.M., Magnusson, P.K.E., Lichtenstein, P., Hultman, C.M., Purcell, S.M., Sklar, P., Sullivan, P.F., 2012. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol. Psychiatry 17, 880–886. https://doi.org/10.1038/mp.2012.73.

Bezinović, P., Malatestinić, D., 2009. Perceived exposure to substance use and risk-taking behavior in early adolescence: cross-sectional study. Croat. Med. J. 50, 157–164. https://doi.org/10.3325/CMJ.2009.50.157.

Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.

Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., Malik, R., Patsopoulos, N.A., Ripke, S., Wei, Z., Yu, D., Lee, P.H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., Berr, C., Letenneur, L., Hannequin, D., Amouyel, P., Boland, A., Deleuze, J.-F., Duron, E., Vardarajan, B.N., Reitz, C., Goate, A.M., Huentelman, M.J., Kamboh, M.I., Larson, E.B., Rogaeva, E., St George-Hyslop, P., Hakonarson, H., Kukull, W.A., Farrer, L.A., Barnes, L.L., Beach, T.G., Demirci, F.Y., Head, E., Hulette, C.M., Jicha, G.A., Kauwe, J.S.K., Kaye, J.A., Leverenz, J.B., Levey, A.I., Lieberman, A.P., Pankratz, V.S., Poon, W.W., Quinn, J.F., Saykin, A.J., Schneider, L.S., Smith, A.G., Sonnen, J.A., Stern, R.A., Van Deerlin, V.M., Van Eldik, L.J., Harold, D., Russo, G., Rubinsztein, D.C., Bayer, A., Tsolaki, M., Proitsi, P., Fox, N.C., Hampel, H., Owen, M.J., Mead, S., Passmore, P., Morgan, K., Nöthen, M.M., Rossor, M., Lupton, M.K., Hoffmann, P., Kornhuber, J., Lawlor, B., McQuillin, A., Al-Chalabi, A., Bis, J.C., Ruiz, A., Boada, M., Seshadri, S., Beiser, A., Rice, K., van der Lee, S.J., De Jager, P.L., Geschwind, D.H., Riemenschneider, M., Riedel-Heller, S., Rotter, J.I., Ransmayr, G., Hyman, B.T., Cruchaga, C., Alegret, M., Winsvold, B., Palta, P., Farh, K.-H., Cuenca-Leon, E., Furlotte, N., Kurth, T., Ligthart, L., Terwindt, G.M., Freilinger, T., Ran, C., Gordon, S.D., Borck, G., Adams, H.H.H., Lehtimäki, T., Wedenoja, J., Buring, J.E., Schürks, M., Hrafnsdottir, M., Hottenga, J.-J., Penninx, B., Artto, V., Kaunisto, M., Vepsäläinen, S., Martin, N.G., Montgomery, G.W., Kurki, M.I., Hämäläinen, E., Huang, H., Huang, J., Sandor, C., Webber, C., Muller-Myhsok, B., Schreiber, S., Salomaa, V., Loehrer, E., Göbel, H., Macaya, A., Pozo-Rosich, P., Hansen, T., Werge, T., Kaprio, J., Metspalu, A., Kubisch, C., Ferrari, M.D., Belin, A.C., van den Maagdenberg, A.M.J.M., Zwart, J.-A., Boomsma, D., Eriksson, N., Olesen, J., Chasman, D.I., Nyholt, D.R., Avbersek, A., Baum, L., Berkovic, S., Bradfield, J., Buono, R., Catarino, C.B., Cossette, P., De Jonghe, P., Depondt, C., Dlugos, D., Ferraro, T.N., French, J., Hjalgrim, H., Jamnadas-Khoda, J., Kälviäinen, R., Kunz, W.S., Lerche, H., Leu, C., Lindhout, D., Lo, W., Lowenstein, D., McCormack, M., Møller, R.S., Molloy, A., Ng, P.-W., Oliver, K., Privitera, M., Radtke, R., Ruppert, A.-K., Sander, T., Schachter, S., Schankin, C., Scheffer, I., Schoch, S., Sisodiya, S.M., Smith, P., Sperling, M., Striano, P., Surges, R., Thomas, G.N., Visscher, F., Whelan, C.D., Zara, F., Heinzen, E.L., Marson, A., Becker, F., Stroink, H., Zimprich, F., Gasser, T., Gibbs, R., Heutink, P., Martinez, M., Morris, H.R., Sharma, M., Ryten, M., Mok, K.Y., Pulit, S., Bevan, S., Holliday, E., Attia, J., Battey, T., Boncoraglio, G., Thijs, V., Chen, W.-M., Mitchell, B., Rothwell, P., Sharma, P., Sudlow, C., Vicente, A., Markus, H., Kourkoulis, C., Pera, J., Raffeld, M., Silliman, S., Boraska Perica, V., Thornton, L.M., Huckins, L.M., William Rayner, N., Lewis, C.M., Gratacos, M., Rybakowski, F., Keski-Rahkonen, A., Raevuori, A., Hudson, J.I., Reichborn-Kjennerud, T., Monteleone, P., Karwautz, A., Mannik, K., Baker, J.H., O'Toole, J.K., Trace, S.E., Davis, O.S.P., Helder, S.G., Ehrlich, S., Herpertz-Dahlmann, B., Danner, U.N., van Elburg, A.A., Clementi, M., Forzan, M., Docampo, E., Lissowska, J., Hauser, J., Tortorella, A., Maj, M., Gonidakis, F., Tziouvas, K., Papezova, H., Yilmaz, Z., Wagner, G., Cohen-Woods, S., Herms, S., Julià, A., Rabionet, R., Dick, D.M., Ripatti, S., Andreassen, O.A., Espeseth, T., Lundervold, A.J., Steen, V.M., Pinto, D., Scherer, S.W., Aschauer, H., Schosser, A., Alfredsson, L., Padyukov, L., Halmi, K.A., Mitchell, J., Strober, M., Bergen, A.W., Kaye, W., Szatkiewicz, J.P., Cormand, B., Ramos-Quiroga, J.A., Sánchez-Mora, C., Ribasés, M., Casas, M., Hervas, A., Arranz, M.J., Haavik, J., Zayats, T., Johansson, S., Williams, N., Dempfle, A., Rothenberger, A., Kuntsi, J., Oades, R.D., Banaschewski, T., Franke, B., Buitelaar, J.K., Arias Vasquez, A., Doyle, A.E., Reif, A., Lesch, K.-P., Freitag, C., Rivero, O., Palmason, H., Romanos, M., Langley, K., Rietschel, M., Witt, S.H., Dalsgaard, S., Børglum, A.D., Waldman, I., Wilmot, B., Molly, N., Bau, C.H.D., Crosbie, J., Schachar, R., Loo, S.K., McGough, J.J., Grevet, E.H., Medland, S.E., Robinson, E., Weiss, L.A., Bacchelli, E., Bailey, A., Bal, V., Battaglia, A., Betancur, C., Bolton, P., Cantor, R., Celestino-Soper, P., Dawson, G., De Rubeis, S., Duque, F., Green, A., Klauck, S.M., Leboyer, M., Levitt, P., Maestrini, E., Mane, S., De-Luca, D.M., Parr, J., Regan, R., Reichenberg, A., Sandin, S., Vorstman, J., Wassink, T., Wijsman, E., Cook, E., Santangelo, S., Delorme, R., Rogé, B., Magalhaes, T., Arking, D., Schulze, T.G., Thompson, R.C., Strohmaier, J., Matthews,

K., Melle, I., Morris, D., Blackwood, D., McIntosh, A., Bergen, S.E., Schalling, M., Jamain, S., Maaser, A., Fischer, S.B., Reinbold, C.S., Fullerton, J.M., Guzman-Parra, J., Mayoral, F., Schofield, P.R., Cichon, S., Mühleisen, T.W., Degenhardt, F., Schumacher, J., Bauer, M., Mitchell, P.B., Gershon, E.S., Rice, J., Potash, J.B., Zandi, P.P., Craddock, N., Ferrier, I.N., Alda, M., Rouleau, G.A., Turecki, G., Ophoff, R., Pato, C., Anjorin, A., Stahl, E., Leber, M., Czerski, P.M., Cruceanu, C., Jones, I.R., Posthuma, D., Andlauer, T.F.M., Forstner, A.J., Streit, F., Baune, B.T., Air, T., Sinnamon, G., Wray, N.R., MacIntyre, D.J., Porteous, D., Homuth, G., Rivera, M., Grove, J., Middeldorp, C.M., Hickie, I., Pergadia, M., Mehta, D., Smit, J.H., Jansen, R., de Geus, E., Dunn, E., Li, Q.S., Nauck, M., Schoevers, R.A., Beekman, A.T., Knowles, J.A., Viktorin, A., Arnold, P., Barr, C.L., Bedoya-Berrio, G., Bienvenu, O.J., Brentani, H., Burton, C., Camarena, B., Cappi, C., Cath, D., Cavallini, M., Cusi, D., Darrow, S., Denys, D., Derks, E.M., Dietrich, A., Fernandez, T., Figee, M., Freimer, N., Gerber, G., Grados, M., Greenberg, E., Hanna, G.L., Hartmann, A., Hirschtritt, M.E., Hoekstra, P.J., Huang, A., Huyser, C., Illmann, C., Jenike, M., Kuperman, S., Leventhal, B., Lochner, C., Lyon, G.J., Macciardi, F., Madruga-Garrido, M., Malaty, I.A., Maras, A., McGrath, L., Miguel, E.C., Mir, P., Nestadt, G., Nicolini, H., Okun, M.S., Pakstis, A., Paschou, P., Piacentini, J., Pittenger, C., Plessen, K., Ramensky, V., Ramos, E.M., Reus, V., Richter, M.A., Riddle, M.A., Robertson, M.M., Roessner, V., Rosário, M., Samuels, J.F., Sandor, P., Stein, D.J., Tsetsos, F., Van Nieuwerburgh, F., Weatherall, S., Wendland, J.R., Wolanczyk, T., Worbe, Y., Zai, G., Goes, F.S., McLaughlin, N., Nestadt, P.S., Grabe, H.-J., Depienne, C., Konkashbaev, A., Lanzagorta, N., Valencia-Duarte, A., Bramon, E., Buccola, N., Cahn, W., Cairns, M., Chong, S.A., Cohen, D., Crespo-Facorro, B., Crowley, J., Davidson, M., DeLisi, L., Dinan, T., Donohoe, G., Drapeau, E., Duan, J., Haan, L., Hougaard, D., Karachanak-Yankova, S., Khrunin, A., Klovins, J., Kučinskas, V., Lee Chee Keong, J., Limborska, S., Loughland, C., Lönnqvist, J., Maher, B., Mattheisen, M., McDonald, C., Murphy, K.C., Nenadic, I., van Os, J., Pantelis, C., Pato, M., Petryshen, T., Quested, D., Roussos, P., Sanders, A.R., Schall, U., Schwab, S.G., Sim, K., So, H.-C., Stögmann, E., Subramaniam, M., Toncheva, D., Waddington, J., Walters, J., Weiser, M., Cheng, W., Cloninger, R., Curtis, D., Gejman, P.V., Henskens, F., Mattingsdal, M., Oh, S.-Y., Scott, R., Webb, B., Breen, G., Churchhouse, C., Bulik, C.M., Daly, M., Dichgans, M., Faraone, S.V., Guerreiro, R., Holmans, P., Kendler, K.S., Koeleman, B., Mathews, C.A., Price, A., Scharf, J., Sklar, P., Williams, J., Wood, N.W., Cotsapas, C., Palotie, A., Smoller, J.W., Sullivan, P., Rosand, J., Corvin, A., Neale, B.M., Schott, J.M., Anney, R., Elia, J., Grigoroiu-Serbanescu, M., Edenberg, H.J., Murray, R., 2018. Analysis of shared heritability in common disorders of the brain. Science 360. https://doi.org/10.1126/science.aap8757.

Bühler, K.M., Giné, E., Echeverry-Alzate, V., Calleja-Conde, J., De Fonseca, F.R., López-Moreno, J.A., 2015. Common single nucleotide variants underlying drug addiction: more than a decade of research. Addict. Biol. 20, 845–871. https://doi.org/10.1111/adb.12204.

Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., Daly, M.J., Price, A.L., Neale, B.M., 2015a. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241. https://doi.org/10.1038/ng.3406.

Bulik-Sullivan, B., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., Neale, B.M., 2015b. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295. https://doi.org/10.1038/ng.3211.

Carey, C.E., Agrawal, A., Bucholz, K.K., Hartz, S.M., Lynskey, M.T., Nelson, E.C., Bierut, L.J., Bogdan, R., 2016. Associations between polygenic risk for psychiatric disorders and substance involvement. Front. Genet. 7, 149. https://doi.org/10.3389/fgene.2016.00149.

Compton, W.M., Thomas, Y.F., Stinson, F.S., Grant, B.F., 2007. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: results from the national epidemiologic survey on alcohol and related conditions. Arch. Gen. Psychiatry 64, 566–576. https://doi.org/10.1001/archpsyc.64.5.566.

Cross-Disorder Group of the PGC, Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O.A., Anjorin, A., Anney, R., Anttila, V., Arking, D.E., Asherson, P., Azevedo, M.H., Backlund, L., Badner, J.A., Bailey, A.J., Banaschewski, T., Barchas, J.D., Barnes, M.R., Barrett, T.B., Bass, N., Battaglia, A., Bauer, M., Bayés, M., Bellivier, F., Bergen, S.E., Berrettini, W., Betancur, C., Bettecken, T., Biederman, J., Binder, E.B., Black, D.W., Blackwood, D.H.R., Bloss, C.S., Boehnke, M., Boomsma, D.I., Breen, G., Breuer, R., Bruggeman, R., Cormican, P., Buccola, N.G., Buitelaar, J.K., Bunney, W.E., Buxbaum, J.D., Byerley, W.F., Byrne, E.M., Caesar, S., Cahn, W., Cantor, R.M., Casas, M., Chakravarti, A., Chambert, K., Choudhury, K., Cichon, S., Cloninger, C.R., Collier, D.A., Cook, E.H., Coon, H., Cormand, B., Corvin, A., Coryell, W.H., Craig, D.W., Craig, I.W., Crosbie, J., Cuccaro, M.L., Curtis, D., Czamara, D., Datta, S., Dawson, G., Day, R., De Geus, E.J., Degenhardt, F., Djurovic, S., Donohoe, G.J., Doyle, A.E., Duan, J., Dudbridge, F., Duketis, E., Ebstein, R.P., Edenberg, H.J., Elia, J., Ennis, S., Etain, B., Fanous, A., Farmer, A.E., Ferrier, I.N., Flickinger, M., Fombonne, E., Foroud, T., Frank, J., Franke, B., Fraser, C., Freedman, R., Freimer, N.B., Freitag, C.M., Friedl, M., Frisén, L., Gallagher, L., Gejman, P.V., Georgieva, L., Gershon, E.S., Geschwind, D.H., Giegling, I., Gill, M., Gordon, S.D., Gordon-Smith, K., Green, E.K., Greenwood, T.A., Grice, D.E., Gross, M., Grozeva, D., Guan, W., Gurling, H., De Haan, L., Haines, J.L., Hakonarson, H., Hallmayer, J., Hamilton, S.P., Hamshere, M.L., Hansen, T.F., Hartmann, A.M., Hautzinger, M., Heath, A.C., Henders, A.K., Herms, S., Hickie, I.B., Hipolito, M., Hoefels, S., Holmans, P.A., Holsboer, F., Hoogendijk, W.J., Hottenga, J.J., Hultman, C.M., Hus, V., Ingason, A., Ising, M., Jamain, S., Jones, E.G., Jones, I., Jones, L., Tzeng, J.Y., Kähler, A.K., Kahn, R.S., Kandaswamy, R., Keller, M.C., Kennedy, J.L., Kenny, E., Kent, L., Kim, Y., Kirov, G.K., Klauck, S.M., Klei, L., Knowles, J.A., Kohli, M.A., Koller, D.L., Konte, B., Korszun, A., Krabbendam, L., Krasucki, R., Kuntsi, J., Kwan, P., Landén, M., Långström, N., Lathrop, M., Lawrence, J., Lawson, W.B., Leboyer, M., Ledbetter, D.H., Lee, P.H., Lencz, T., Lesch, K.P., Levinson, D.F., Lewis, C.M., Li, J., Lichtenstein, P., Lieberman, J.A., Lin, D.Y., Linszen, D.H., Liu, C., Lohoff, F.W., Loo, S.K., Lord, C., Lowe, J.K., Lucae, S., Macintyre, D.J., Madden, P.A.F., Maestrini, E., Magnusson, P.K.E., Mahon, P.B., Maier, W., Malhotra, A.K., Mane, S.M., Martin, C.L., Martin, N.G., Mattheisen, M., Matthews, K., Mattingsdal, M., Mccarroll, S.A., Mcghee, K.A., Mcgough, J.J., Mcgrath, P.J., Mcguffin, P., Mcinnis, M.G., Mcintosh, A., Mckinney, R., Mclean, A.W., Mcmahon, F.J., Mcmahon, W.M., Mcquillin, A., Medeiros, H., Medland, S.E., Meier, S., Melle, I., Meng, F., Meyer, J., Middeldorp, C.M., Middleton, L., Milanova, V., Miranda, A., Monaco, A.P., Montgomery, G.W., Moran, J.L., Moreno-De-Luca, D., Morken, G., Morris, D.W., Morrow, E.M., Moskvina, V., Muglia, P., Mühleisen, T.W., Muir, W.J., Müller-Myhsok, B., Murtha, M., Myers, R.M., Myin-Germeys, I., Neale, M.C., Nelson, S.F., Nievergelt, C.M., Nikolov, I., Nimgaonkar, V., Nolen, W.A., Nöthen, M.M., Nurnberger, J.I., Nwulia, E.A., Nyholt, D.R., O'dushlaine, C., Oades, R.D., Olincy, A., Oliveira, G., Olsen, L., Ophoff, R.A., Osby, U., Owen, M.J., Palotie, A., Parr, J.R., Paterson, A.D., Pato, C.N., Pato, M.T., Penninx, B.W., Pergadia, M.L., Pericak-Vance, M.A., Pickard, B.S., Pimm, J., Piven, J., Posthuma, D., Potash, J.B., Poustka, F., Propping, P., Puri, V., Quested, D.J., Quinn, E.M., Ramos-Quiroga, J.A., Rasmussen, H.B., Raychaudhuri, S., Rehnström, K., Reif, A., Ribasés, M., Rice, J.P., Rietschel, M., Roeder, K., Roeyers, H., Rossin, L., Rothenberger, A., Rouleau, G., Ruderfer, D., Rujescu, D., Sanders, A.R., Sanders, S.J., Santangelo, S.L., Sergeant, J.A., Schachar, R., Schalling, M., Schatzberg, A.F., Scheftner, W.A., Schellenberg, G.D., Scherer, S.W., Schork, N.J., Schulze, T.G., Schumacher, J., Schwarz, M., Scolnick, E., Scott, L.J., Shi, J., Shilling, P.D., Shyn, S.I., Silverman, J.M., Slager, S.L., Smalley, S.L., Smit, J.H., Smith, E.N., Sonuga-Barke, E.J.S., St. Clair, D., State, M., Steffens, M., Steinhausen, H.C., Strauss, J.S., Strohmaier, J., Stroup, T.S., Sutcliffe, J.S., Szatmari, P., Szelinger, S., Thirumalai, S., Thompson, R.C., Todorov, A.A., Tozzi, F., Treutlein, J., Uhr, M., Van Den Oord, E.J.C.G., Van Grootheest, G., Van Os, J., Vicente, A.M., Vieland, V.J., Vincent, J.B., Visscher, P.M., Walsh, C.A., Wassink, T.H., Watson, S.J., Weissman, M.M., Werge, T., Wienker, T.F., Wijsman, E.M., Willemsen, G., Williams, N., Willsey, A.J., Witt, S.H., Xu, W., Young, A.H., Yu, T.W., Zammit, S., Zandi, P.P., Zhang, P., Zitman, F.G., Zöllner, S., Devlin, B., Kelsoe, J.R., Sklar, P., Daly, M.J., O'donovan, M.C., Craddock, N., Sullivan, P.F., Smoller, J.W., Kendler, K.S., Wray, N.R., 2013. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. 45, 984–994. https://doi.org/10.1038/ng.2711.

Cross-Disorder Group of the PGC, Lee, P.H., Anttila, V., Won, H., Feng, Y.-C.A., Rosenthal, J., Zhu, Z., Tucker-Drob, E.M., Nivard, M.G., Grotzinger, A.D., Posthuma, D., Wang, M.M.-J., Yu, D., Stahl, E., Walters, R.K., Anney, R.J.L., Dun, L.E., Smoller, J.W., 2019. Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. bioRxiv. https://doi.org/10.1101/528117.

Currie, S.R., Patten, S.B., Williams, J.V.A., Wang, J., Beck, C.A., El-Guebaly, N., Maxwell, C., 2005. Comorbidity of major depression with substance use disorders. Can. J. Psychiatr. 50, 660–666. https://doi.org/10.1177/070674370505001013.

Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J.I., Grasby, K.L., Grove, J., Gudmundsson, O.O., Hansen, C.S., Hauberg, M.E., Hollegaard, M.V., Howrigan, D.P., Huang, H., Maller, J.B., Martin, A.R., Martin, N.G., Moran, J., Pallesen, J., Palmer, D.S., Pedersen, C.B., Pedersen, M.G., Poterba, T., Poulsen, J.B., Ripke, S., Robinson, E.B., Satterstrom, F.K., Stefansson, H., Stevens, C., Turley, P., Walters, G.B., Won, H., Wright, M.J., Andreassen, O.A., Asherson, P., Burton, C.L., Boomsma, D.I., Cormand, B., Dalsgaard, S., Franke, B., Gelernter, J., Geschwind, D., Hakonarson, H., Haavik, J., Kranzler, H.R., Kuntsi, J., Langley, K., Lesch, K.-P., Middeldorp, C., Reif, A., Rohde, L.A., Roussos, P., Schachar, R., Sklar, P., Sonuga-Barke, E.J.S., Sullivan, P.F., Thapar, A., Tung, J.Y., Waldman, I.D., Medland, S.E., Stefansson, K., Nordentoft, M., Hougaard, D.M., Werge, T., Mors, O., Mortensen, P.B., Daly, M.J., Faraone, S.V., Børglum, A.D., Neale, B.M., 2019. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. 51, 63–75. https://doi.org/10.1038/s41588-018-0269-7.

Du Rietz, E., Coleman, J., Glanville, K., Choi, S.W., O'Reilly, P.F., Kuntsi, J., 2017. Association of polygenic risk for attention-deficit/hyperactivity disorder with co-occurring traits and disorders. Biol. Psychiatry Cogn. Neurosci. Neuroimaging. https://doi.org/10.1016/J.BPSC.2017.11.013.

Ducci, F., Goldman, D., 2012. The genetic basis of addictive disorders. Psychiatr. Clin. North Am. https://doi.org/10.1016/j.psc.2012.03.010.

Euesden, J., Lewis, C.M., O'Reilly, P.F., 2015. PRSice: polygenic risk score software. Bioinformatics 31, 1466–1468. https://doi.org/10.1093/bioinformatics/btu848.

European Monitoring Centre for Drugs and Drug Addiction, 2017. European Drug Report 2017: Trends and Developments. Publications Office of the European Union, Luxembourg. https://doi.org/10.2810/610791.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R.B., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M., Price, A.L., Patterson, N., Neale, B.M., Price, A.L., 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235. https://doi.org/10.1038/ng.3404.

Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H.R., Farrer, L., 2014. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. Mol. Psychiatry 19, 717–723. https://doi.org/10.1038/mp.2013.99.

Gurriarán, X., Rodríguez-López, J., Flórez, G., Pereiro, C., Fernández, J.M., Fariñas, E., Estévez, V., Arrojo, M., Costas, J., GenPol Study Group, 2018. Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores.

Genes Brain Behav., e12504. https://doi.org/10.1111/gbb.12504.
Hartz, S.M., Horton, A.C., Oehlert, M., Carey, C.E., Agrawal, A., Bogdan, R., Chen, L.-S., Hancock, D.B., Johnson, E.O., Pato, C.N., Pato, M.T., Rice, J.P., Bierut, L.J., 2017. Association between substance use disorder and polygenic liability to schizophrenia. Biol. Psychiatry 82, 709–715. https://doi.org/10.1016/j.biopsych.2017.04.020.
Hasin, D., Kilcoyne, B., 2012. Comorbidity of psychiatric and substance use disorders in the United States: current issues and findings from the NESARC. Curr. Opin. Psychiatry 25, 165–171. https://doi.org/10.1097/YCO.0b013e3283523dcc.
Huggett, S.B., Stallings, M.C., 2019. Cocaine'omics: genome-wide and transcriptome-wide analyses provide biological insight into cocaine use and dependence. Addict. Biol. https://doi.org/10.1111/adb.12719.
Ikeda, M., Okahisa, Y., Aleksic, B., Won, M., Kondo, N., Naruse, N., Aoyama-Uehara, K., Sora, I., Iyo, M., Hashimoto, R., Kawamura, Y., Nishida, N., Miyagawa, T., Takeda, M., Sasaki, T., Tokunaga, K., Ozaki, N., Ujike, H., Iwata, N., 2013. Evidence for shared genetic risk between methamphetamine-induced psychosis and schizophrenia. Neuropsychopharmacology 38, 1864–1870. https://doi.org/10.1038/npp.2013.94.
Johnson, E.O., Hancock, D.B., Levy, J.L., Gaddis, N.C., Page, G.P., Glasheen, C., Saccone, N.L., Bierut, L.J., Kral, A.H., 2016. KAT2B polymorphism identified for drug abuse in African Americans with regulatory links to drug abuse pathways in human prefrontal cortex. Addict. Biol. 21, 1217–1232. https://doi.org/10.1111/adb.12286.
Lee, S.H., Wray, N.R., Goddard, M.E., Visscher, P.M., 2011. Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88, 294–305. https://doi.org/10.1016/j.ajhg.2011.02.002.
Martin, J., Taylor, M.J., Lichtenstein, P., 2018. Assessing the evidence for shared genetic risks across psychiatric disorders and traits. Psychol. Med. 48, 1759–1774. https://doi.org/10.1017/S0033291717003440.
Mbarek, H., Milaneschi, Y., Fedko, I.O., Hottenga, J.-J., de Moor, M.H.M., Jansen, R., Gelernter, J., Sherva, R., Willemsen, G., Boomsma, D.I., Penninx, B.W., Vink, J.M., 2015. The genetics of alcohol dependence: twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 168, 739–748. https://doi.org/10.1002/ajmg.b.32379.
Mitchell, B.D., Fornage, M., McArdle, P.F., Cheng, Y.C., Pulit, S.L., Wong, Q., Dave, T., Williams, S.R., Corriveau, R., Gwinn, K., Doheny, K., Laurie, C.C., Rich, S.S., de Bakker, P.I.W., 2014. Using previously genotyped controls in genome-wide association studies (GWAS): application to the stroke genetics network (SiGN). Front. Genet. 5, 95. https://doi.org/10.3389/fgene.2014.00095.
O'Dushlaine, C., The Network and Pathway Analysis Subgroup of the Psychiatric GenomicsConsortium, 2015. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat. Neurosci. 18, 199–209. https://doi.org/10.1038/nn.3922.
Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M.G., Nolte, I.M., Tiesler, C.M.T., Bakermans-Kranenburg, M.J., Davies, G.E., Evans, D.M., Geoffroy, M.C., Grallert, H., Groen-Blokhuis, M.M., Hudziak, J.J., Kemp, J.P., Keltikangas-Järvinen, L., McMahon, G., Mileva-Seitz, V.R., Motazedi, E., Power, C., Raitakari, O.T., Ring, S.M., Rivadeneira, F., Rodriguez, A., Scheet, P.A., Seppälä, I., Snieder, H., Standl, M., Thiering, E., Timpson, N.J., Veenstra, R., Velders, F.P., Whitehouse, A.J.O., Smith, G.D., Heinrich, J., Hypponen, E., Lehtimäki, T., Middeldorp, C.M., Oldehinkel, A.J., Pennell, C.E., Boomsma, D.I., Tiemeier, H., 2016. A genome-wide approach to children's aggressive behavior: the EAGLE consortium. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. 171, 562–572. https://doi.org/10.1002/ajmg.b.32333.
Pineda-Cirera, L., Cabana-Domínguez, J., Roncero, C., Cozar, M., Grau-López, L., Abad, A.C., Martínez-Luna, N., Robles-Martínez, M., Sánchez-Mora, C., Ramos-Quiroga, J.A., Casas, M., Ribasés, M., Fernàndez-Castillo, N., Cormand, B., 2018. Evaluation of previous substance dependence genome-wide significant findings in a Spanish sample. Drug Alcohol Depend. 187, 358–362. https://doi.org/10.1016/j.drugalcdep.2018.03.013.
Piñeiro-Dieguez, B., Balanzá-Martínez, V., García-García, P., Soler-López, B., Domingo, M.A., Labarra, J.D.A., Lobato, P.A., Salamanca, A.A., Bes, J.A., Fernández, F.J.B., Moraga, R.B., Blanco, J.B., Perona, A.B., Boniatti, T.C., Gras, J.M.C., Martínez, J.M.C., Machado, I.C., Pena, M.C.C., Flores, C.C., Dono, C.C., Taracena, M.T.D.L., García, B.D.P., Martín, P.F., Del Valle, E.F.G., Fontova, E., Eroles, F.F., Chapa, R.F., Tejedor, I.G., Palomares, A.G., Campayo, J.G., Moreno, J.M.G., Pérez, X.G., Escrig, M.D.C.G., Vives, S.G., Queralt, E.G., Gonochategui, A.G., Escajedo, A.G., Labrador, R.G., Mendoza, A.H., Lorenzo, F.G.I., Iribarren, M.M., Yepez, G.J.D.F., Vazquez, M.A.L., Borrajo, J.M.L., Arribas, S.L., García, J.M.M., Flores, A.M., Fernández-Mayoralas, D.M., Navarro, N.M., Recuero, L.M., Gras, I.M., Raga, J.M., López, P.A.M., Horche, J.L.M., Flores, P.L.M., Diago, E.O., Panicali, F., Pastor, F.P., Alfaro, G.P., Rosati, S.P., Sanz, M.P., Presa, A.M.I.Q., Alandes, C.R., Solano, J.J.R., Buguña, J.R.S., Izquierdo, G.S.N., García, M.S., De La Garza, C.L.S., Granado, O.S., Murla, F.S., Ceballos, P.S., Cabra, O.S., Rodríguez, P.A.S., Sender, V.T., Moreira, R.V., Abarquero, M.A.V., Peri, R.V., Mateo, L.V., Zambrano, D., Ramos, J.M.Z., 2016. Psychiatric comorbidity at the time of diagnosis in adults with ADHD: the CAT study. J. Atten. Disord. 20, 1066–1075. https://doi.org/10.1177/1087054713518240.
Reginsson, G.W., Ingason, A., Euesden, J., Bjornsdottir, G., Olafsson, S., Sigurdsson, E., Oskarsson, H., Tyrfingsson, T., Runarsdottir, V., Hansdottir, I., Steinberg, S., Stefansson, H., Gudbjartsson, D.F., Thorgeirsson, T.E., Stefansson, K., 2018. Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. Addict. Biol. 23, 485. https://doi.org/10.1111/ADB.12496.
Ripke, S., Consortium, S.P.G, 2011. Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. 43, 969–976. https://doi.org/10.1038/ng.940.
Ripke, S., Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427. https://doi.org/10.1038/nature13595.
Ruderfer, D.M., Fanous, A.H., Ripke, S., McQuillin, A., Amdur, R.L., Gejman, P.V., O'Donovan, M.C., Andreassen, O.A., Djurovic, S., Hultman, C.M., Kelsoe, J.R., Jamain, S., Landén, M., Leboyer, M., Nimgaonkar, V., Nurnberger, J., Smoller, J.W., Craddock, N., Corvin, A., Sullivan, P.F., Holmans, P., Sklar, P., Kendler, K.S., 2014. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol. Psychiatry 19, 1017–1024. https://doi.org/10.1038/mp.2013.138.
Sanders, A.R., Göring, H.H.H., Duan, J., Drigalenko, E.I., Moy, W., Freda, J., He, D., Shi, J., MGS, Gejman, P.V., 2013. Transcriptome study of differential expression in schizophrenia. Hum. Mol. Genet. 22, 5001–5014. https://doi.org/10.1093/hmg/ddt350.
Shantna, K., Chaudhury, S., Verma, A., Singh, A., 2009. Comorbid psychiatric disorders in substance dependence patients: a control study. Ind. Psychiatry J. 18, 84–87. https://doi.org/10.4103/0972-6748.62265.
Shraga, R., Yarnall, S., Elango, S., Manoharan, A., Rodriguez, S.A., Bristow, S.L., Kumar, N., Niknazar, M., Hoffman, D., Ghadir, S., Vassena, R., Chen, S.H., Hershlag, A., Grifo, J., Puig, O., 2017. Evaluating genetic ancestry and self-reported ethnicity in the context of carrier screening. BMC Genet. 18, 99. https://doi.org/10.1186/s12863-017-0570-y.
Sleiman, P., Wang, D., Glessner, J., Hadley, D., Gur, R.E., Cohen, N., Li, Q., Hakonarson, H., Sleiman, P., Glessner, J., Hadley, D., Kao, C., Wu, Z., Kim, C., Thomas, K., Hakonarson, H., Wang, D., Favis, R., Fu, D.-J., Chung, H., Savitz, A., Gopal, S., Cohen, N., Li, Q., 2013. GWAS meta analysis identifies TSNARE1 as a novel schizophrenia/bipolar susceptibility locus. Sci. Rep. 3, 3075. https://doi.org/10.1038/srep03075.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
Tielbeek, J.J., Johansson, A., Polderman, T.J.C., Rautiainen, M.-R., Jansen, P., Taylor, M., Tong, X., Lu, Q., Burt, A.S., Tiemeier, H., Viding, E., Plomin, R., Martin, N.G., Heath, A.C., Madden, P.A.F., Montgomery, G., Beaver, K.M., Waldman, I., Gelernter, J., Kranzler, H.R., Farrer, L.A., Perry, J.R.B., Munafò, M., LoParo, D., Paunio, T., Tiihonen, J., Mous, S.E., Pappa, I., de Leeuw, C., Watanabe, K., Hammerschlag, A.R., Salvatore, J.E., Aliev, F., Bigdeli, T.B., Dick, D., Faraone, S.V., Popma, A., Medland, S.E., Posthuma, D., 2017. Genome-wide association studies of a broad spectrum of antisocial behavior. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2017.3069.
Uh, H.W., Deelen, J., Beekman, M., Helmer, Q., Rivadeneira, F., Hottenga, J.J., Boomsma, D.I., Hofman, A., Uitterlinden, A.G., Slagboom, P.E., Böhringer, S., Houwing-Duistermaat, J.J., 2012. How to deal with the early GWAS data when imputing and combining different arrays is necessary. Eur. J. Hum. Genet. 20, 572–576. https://doi.org/10.1038/ejhg.2011.231.
Wagner, F.A., Anthony, J.C., 2002. From first drug use to drug dependence: developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. Neuropsychopharmacology 26, 479–488. https://doi.org/10.1016/S0893-133X(01)00367-0.
Walters, R.K., Polimanti, R., Johnson, E.C., McClintick, J.N., Adams, M.J., Adkins, A.E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J.M., Bigdeli, T.B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A.R., Edwards, A.C., Fontanillas, P., Foo, J.C., Fox, L., Frank, J., Giegling, I., Gordon, S., Hack, L.M., Hartmann, A.M., Hartz, S.M., Heilmann-Heimbach, S., Herms, S., Hodgkinson, C., Hoffmann, P., Jan Hottenga, J., Kennedy, M.A., Alanne-Kinnunen, M., Konte, B., Lahti, J., Lahti-Pulkkinen, M., Lai, D., Ligthart, L., Loukola, A., Maher, B.S., Mbarek, H., McIntosh, A.M., McQueen, M.B., Meyers, J.L., Milaneschi, Y., Palviainen, T., Pearson, J.F., Peterson, R.E., Ripatti, S., Ryu, E., Saccone, N.L., Salvatore, J.E., Sanchez-Roige, S., Schwandt, M., Sherva, R., Streit, F., Strohmaier, J., Thomas, N., Wang, J.-C., Webb, B.T., Wedow, R., Wetherill, L., Wills, A.G., Boardman, J.D., Chen, D., Choi, D.-S., Copeland, W.E., Culverhouse, R.C., Dahmen, N., Degenhardt, L., Domingue, B.W., Elson, S.L., Frye, M.A., Gäbel, W., Hayward, C., Ising, M., Keyes, M., Kiefer, F., Kramer, J., Kuperman, S., Lucae, S., Lynskey, M.T., Maier, W., Mann, K., Männistö, S., Müller-Myhsok, B., Murray, A.D., Nurnberger, J.I., Palotie, A., Preuss, U., Räikkönen, K., Reynolds, M.D., Ridinger, M., Scherbaum, N., Schuckit, M.A., Soyka, M., Treutlein, J., Witt, S., Wodarz, N., Zill, P., Adkins, D.E., Boden, J.M., Boomsma, D.I., Bierut, L.J., Brown, S.A., Bucholz, K.K., Cichon, S., Costello, E.J., de Wit, H., Diazgranados, N., Dick, D.M., Eriksson, J.G., Farrer, L.A., Foroud, T.M., Gillespie, N.A., Goate, A.M., Goldman, D., Grucza, R.A., Hancock, D.B., Harris, K.M., Heath, A.C., Hesselbrock, V., Hewitt, J.K., Hopfer, C.J., Horwood, J., Iacono, W., Johnson, E.O., Kaprio, J.A., Karpyak, V.M., Kendler, K.S., Kranzler, H.R., Krauter, K., Lichtenstein, P., Lind, P.A., McGue, M., MacKillop, J., Madden, P.A.F., Maes, H.H., Magnusson, P., Martin, N.G., Medland, S.E., Montgomery, G.W., Nelson, E.C., Nöthen, M.M., Palmer, A.A., Pedersen, N.L., Penninx, B.W.J.H., Porjesz, B., Rice, J.P., Rietschel, M., Riley, B.P., Rose, R., Rujescu, D., Shen, P.-H., Silberg, J., Stallings, M.C., Tarter, R.E., Vanyukov, M.M., Vrieze, S., Wall, T.L., Whitfield, J.B., Zhao, H., Neale, B.M., Gelernter, J., Edenberg, H.J., Agrawal, A., Agrawal, A., 2018. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat. Neurosci. 21, 1656–1669. https://doi.org/10.1038/s41593-018-0275-1.
Watanabe, K., Taskesen, E., Van Bochoven, A., Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826. https://doi.org/10.1038/s41467-017-01261-5.
Westermeyer, J., 2006. Comorbid schizophrenia and substance abuse: a review of epidemiology and course. Am. J. Addict. https://doi.org/10.1080/10550490600860114.
Willer, C.J., Li, Y., Abecasis, G.R., 2010. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.
Wray, N.R., The Major Depressive Disorder Working Group of the Psychiatric Genomics

Consortium, 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681. https://doi.org/10.1038/s41588-018-0090-3.

Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J., Visscher, P.M., 2018. Common disease is more complex than implied by the core gene omnigenic model. Cell 173, 1573–1580. https://doi.org/10.1016/j.cell.2018.05.051.

Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.St., Warrington, N.M., Finucane, H.K., Price, A.L., Bulik-Sullivan, B.K., Anttila, V., Paternoster, L., Gaunt, T.R., Evans, D.M., Neale, B.M., 2017. LD hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272–279. https://doi.org/10.1093/bioinformatics/btw613.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY MATERIALS AND METHODS

1. Detailed description of samples

A brief overview of the samples included can be found in Table 1.

All datasets used in this work were obtained from dbGaP (database of Genotypes and Phenotypes / National Center for Biotechnology Information, National Library of Medicine (NCBI/NLM), https://www.ncbi.nlm.nih.gov/gap) under the project "17170: Meta-analysis of cocaine dependence GWAS".

All affected subjects included in this study meet DSM-IV criteria for cocaine dependence, although most of them are also dependent to other drugs of abuse. Diagnoses for schizophrenia, bipolar affective disorder or other major psychotic illnesses or gross cognitive impairment were exclusion criteria. Only for the SAGE controls, drug abuse or dependence could be discarded; for the other studies general population individuals were used as controls.

Sample 1 (SAGE)

The data presented in the current publication was downloaded from the dbGaP website, under dbGaP accession phs000092.v1.p1. SAGE (Study of Addiction: Genetics and Environment) and it is part of the Gene Environment Association Studies initiative (GENEVA, www.genevastudy.org) funded by the National Human Genome Research Institute. Cases and controls included in this study belong to three large, complementary datasets: the Collaborative Study on the Genetics of Alcoholism (COGA), the Family Study of Cocaine Dependence (FSCD), and the Collaborative Genetic Study of Nicotine Dependence (COGEND). All three studies include measures of basic socio-demographic variables (e.g: age, sex,

race/ethnicity, family income…) and other important covariates and/or potential moderators of genetic effects (e.g: comorbid addictions and age at initiation of use for cigarettes, alcohol and drugs). Coding for both individual variables and indices has been standardized across studies and all subjects were assessed in person by trained research assistants. In total this dataset includes 1,897 European-American and African-American alcohol dependent subjects and 1,932 unrelated controls [1]. All cases met criteria for DSM-IV alcohol dependence, but some of them also met the criteria for other drug dependencies such as cocaine, tobacco or cannabis.

Genomic DNA was extracted from whole blood samples and genotyping was performed on an Illumina ILMN_Human-1 chip (Illumina, Inc., San Diego, CA, USA). This dataset includes 1,130 cocaine-dependent individuals and 1,967 controls. After quality control (QC) and ancestry selection, 468 cases and 1,284 controls were selected.

The following datasets did not have controls that meet the inclusion criteria for the association study, so general population individuals with European ancestry genotyped with the same chip were obtained for each dataset of cases:

Sample 2

Cases: The data for the present analysis was downloaded from the dbGaP website, under dbGaP accession phs000952.v1.p1 (Substance Dependence GWAS in European- and African-Americans). Samples were collected in the course of three projects: studies focused on alcohol dependence genetics, on cocaine and opioid dependence genetics. This dataset includes 1,531 self-reported African-American subjects and 1,339 self-reported European-American subjects that meet DSM-IV criteria for opioid, cocaine or alcohol dependence.

Genotyping was performed on an Illumina HumanOmni1-Quad_V1-0_B chip. This dataset includes 2,433 cocaine-dependent individuals and 61 controls.

Controls: The data presented in the current publication is based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000179.v5.p2 (Genetic Epidemiology of COPD (COPDGene) funded by the National Heart, Lung, and Blood Institute). Only control individuals genotyped with the Illumina HumanOmni1-Quad_V1-0_B chip from this dataset were used for our association study (493 controls).

Case and control samples were merged, after QC and ancestry selection, 609 cases and 410 controls were selected.

Sample 3

Cases: The data presented in the current publication is based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000277.v1.p1 (GWAS of Heroin Dependence). This dataset is a collaboration of Australian and American researchers. Cases and controls were obtained from several large investigations including: the Comorbidity and Trauma Study, Heroin Dependence in Western Australia, the OZ-ALC Study, a Twin Study of Mole Development in Adolescence, and ongoing genetic studies of substance dependence conducted by researchers at Yale and collaborating institutions. Cases met lifetime DSM-IV criteria for heroin dependence, but also for cocaine, alcohol or cannabis. Controls included screened individuals who did not meet DSM-IV heroin dependence criteria and unscreened general population controls.

Genomic DNA was extracted from whole blood samples. Individuals used in this study were genotyped on an Illumina Human660W-Quad_v1_A chip (933 cocaine-dependent individuals and 349 controls). Although this dataset includes controls, we added more individuals to increase the statistical power.

Controls: The data presented in the current publication is based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000170.v2.p1. (eMERGE

Genome-Wide Association Study on Cataract and Low HDL cholesterol). Only control individuals from this dataset were used for our association study, all of them genotyped on an Illumina Human660W-Quad_v1_A chip (1,370 controls).

Case and control samples were merged, and after QC and ancestry assessment, 504 cases and 1,190 controls were selected.

Sample 4

Cases: The data presented in the current publication are based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs000425.v1.p1 (Alcohol Dependence GWAS in European- and African Americans). The sample includes 1,889 African-American subjects (1,397 meet DSM-IV criteria for alcohol dependence and 491 are controls) and 1,020 European-American (1,010 meet the criteria for alcohol dependence and 9 are controls). All cases met criteria for DSM-IV alcohol dependence, but some of them also met criteria for cocaine, nicotine and opioid dependence.

Genotyping was performed with the Illumina HumanOmni1-Quad_V1-0_B chip. This dataset includes 1,920 cocaine-dependent individuals and 500 controls.

Controls: The data presented in the current publication are based on the use of study data downloaded from the dbGaP web site, under dbGaP accession phs000524.v1.p1 (Chronic Renal Insufficiency Cohort Study (CRIC) GWAS). As this is a cohort study, all individuals were included in our analysis (3,541 controls), all of them genotyped with the Illumina HumanOmni1-Quad_V1-0_B chip.

Case and control samples were merged, and after QC and ancestry selection, 504 cases and 1,409 controls were selected.

Addictive Diseases

The data presented in the current publication is based on the use of study data downloaded from the dbGaP website, under dbGaP accession phs001109.v1.p1 (Addictions: Genotypes, Polymorphisms, and Function/Human Genetic Correlates of Addictive Diseases). The sample includes 864 cases and 797 controls. All cases met the DSM-IV for at least one of the following drug of abuse: opioid, cocaine, nicotine, cannabis, stimulant, sedative and other drugs. Only a few individuals of European ancestry meet criteria for cocaine dependence, and this caused problems in the analysis of the data; consequently, this dataset was not included in the meta-analysis.

2. Bioinformatics pipeline for quality control and association analyses

Pre-imputation quality control and imputation were performed using the bioinformatics pipeline "Ricopili", developed by the Psychiatric Genomics Consortium (PGC) Statistical Analysis Group [2]. This pipeline generates high quality imputed data to perform association analyses and meta-analyses. To avoid potential study effects, all samples were processed separately.

2.1 Pre-imputation quality control

To reduce batch effect bias, cases and controls from samples 2-4 were merged prior to quality control and imputation using PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) [3]. In the three samples, cases and controls were genotyped with the same genotyping array and were built on the same genomic assembly [4–7]. Flowchart in figure S1 summarizes all steps to control population stratification.

Subjects and SNPs were included in the analyses based on the following default quality control parameters: SNP call rate > 0.95 (before sample removal), subject call rate > 0.98, autosomal heterozygosity deviation ($|F_{het}| < 0.2$), SNP call rate > 0.98 (after sample removal), difference

in SNP missingness between cases and controls < 0.02, and SNP Hardy-Weinberg equilibrium (HWE) (*P* > 1e−06 in controls or *P* > 1e−10 in cases). Furthermore, chromosome X data was used to determine sex based on heterozygosity rates, and individuals were removed from the analyses if gender in phenotype and genotype data did not match.

2.2 European subjects selection

In order to select European subjects from our samples, a principal component analysis (PCA) was performed using smartPCA incorporated in the software Eigensoft [8,9] using the reference panel data from 1,092 individuals from the 1000 Genomes Project, phase 3 (1KGP3, high quality 3,382,774 variants). Only a set of high quality markers were used: autosomal SNPs found in all datasets, minor allele frequency (MAF) > 0.05, HWE *P* > 1e-03, SNP call rate > 0.98, which were pruned for linkage disequilibrium (LD) (r^2 < 0.2 in a 200Kb SNPs window). For this analysis AT/GC SNPs (strand ambiguous SNPs), the MHC region (6:25-35Mb) and Chr8 inversion (8:7-13Mb) were excluded. The subsample of European individuals from 1KGP3 (combined CEU, FIN, GBR, IBS, TSI) was used to define the center of an ellipsoid based on the mean values of principal component (PC) 1 and 2. Then, PC1 and PC2 for all individuals in our sample were used to define a genetically homogeneous population, excluding individuals with PC values greater than three standard deviations from the reference population.

Relatedness was tested with the same set of markers using identity-by-descendent (IBD) analysis in PLINK v1.9, and one individual was excluded in pairs of subjects with pi_hat > 0.2 (cases preferred over controls).

After relatedness and population stratification analysis, another quality control step was performed, using the same parameters described above.

2.3 Genotype imputation

Imputation was performed to obtain information about non-genotyped markers. We used the pre-phasing software SHAPEIT [10] to estimate haplotypes and subsequently MINIMAC3 (https://genome.sph.umich.edu/wiki/Minimac3) [11] for imputing the genotypes. Imputation was done in chunks of 3 Mb using default parameters. The imputation reference data consisted of 2,504 phased haplotypes from the European individuals of the 1000 Genomes Project, phase 3 (1KGP3; October 2014, 81,706,022 variants, release 20130502) [12,13].

2.4 Genetic outliers and population stratification

After imputation, population stratification was reevaluated using a set of high quality markers (autosomal markers with MAF > 0.05, HWE *P* > 1e-04 and SNP call rate > 0.98), which were pruned for LD (r^2 < 0.075) resulting in a set of ~30,000 pruned markers. Based on genome-wide identity-by-state (IBS) information, PLINK generates metrics to detect the outliers. For a given individual, PLINK ranks all other individuals on the basis of how similar (in IBS terms) they are to this particular proband individual. Then it checks if the proband's closest neighbour is significantly more distant to the proband than all other individuals' nearest neighbour is to them. In other words, from the distribution of 'nearest neighbour' scores, one for each individual, it can calculate a sample mean and variance and transform this measure into a Z-score. If an individual has an extremely low Z-score (less than 4 standard deviation units), this would indicate that this individual is an outlier with respect to the rest of the sample and would be removed.

Second, a permutation test for between group IBS differences was performed with fixed 10,000 permutations. To test stratification effects between cases and controls, we reported the p-value of testing whether, on average, an individual was less similar to another phenotypically-discordant individual than would be expected by chance (denoted as T1 in

PLINK). In samples 2 and 3, population stratification was detected (T1 p-value < 0.05). Then, for each sample, PCA was repeated following the steps described previously but, in this case, individuals with PC1 and PC2 values greater than two standards deviations from the mean obtained for European individuals were excluded. After this correction, population stratification analysis was repeated and the T1 p-value was > 0.05 in both cases.

Then, we ran again the "Ricopili" PCA module considering the four samples together to test overlap or relatedness across all datasets. This module uses IBD analysis in PLINK v1.9 to identify pairs of subjects with pi_hat > 0.2, and one individual for each pair was excluded (cases preferred over controls).

Finally, for each sample PCA was redone after exclusion of genetic outliers. The first 20 PCs were tested for association with the phenotype using logistic regression and their impact on the genome-wide test statistics was evaluated using λ. The first 10 PCs (PC1 - PC10) were included as covariates for all samples.

3. GWAS and meta-analysis

Case-control association analyses using the imputed markers (INFO > 0.8 and MAF > 0.01) were performed for each sample by logistic regression under the additive model, with the derived 10 first PCs as covariates using PLINK v1.9. The summary statistics obtained from the different GWASs (in total 2,085 cases and 4,293 controls) were meta-analysed using and inverse-variance weighted fixed effects model implemented in METAL software (http://csg.sph.umich.edu//abecasis/Metal/) [14]. Finally, the GWAS meta-analysis results were filtered by N effective (N_{eff}), so only the markers which were supported by an effective sample size greater than 70% (n ≥ 1,964.76) were included (a total of 9,290,362 markers).

$$Neff = \frac{2}{\frac{1}{Ncases} + \frac{1}{Ncontrols}}$$

Manhattan plot and quantile-quantile plot (Q-Q plot) from each sample and the meta-analysis results were performed using the library qqman implemented in R [15] (Figure 1 and Table S1).

Heterogeneity across studies was tested with the Cochran's Q test and quantified with the I^2 heterogeneity index, in METAL (Figure S2).

3.1 LD Score intercept evaluation

A light deviation from null was observed in the distribution of the test statistics in the Q-Q plot of the results from the GWAS meta-analysis (Figure 1). This deviation can appear because of polygenicity but also due to other confounding factors such as cryptic relatedness, population stratification or model misspecification. LD score regression (LDSC) analysis allows to differentiate between these two components [16]. Under this model when regressing the chi-square statistics from GWAS against LD scores for each SNP (pre-computed LD-scores downloaded from the GitHub website available on https://github.com/bulik/ldsc), the intercept minus one is an estimator for the mean contribution of confounding bias to the inflation in the statistic tests. Based on regression of the test statistics from GWAS meta-analysis, we estimated an intercept close to one (intercept = 1.01 (SE = 0.0068; *P* = 0.1)). Additionally, the ratio (ratio = (intercept-1)/(mean(chi^2)-1)) measures the proportion of the inflation in the mean chi-square that the LDSC intercept ascribes to causes other than polygenic heritability. Results estimated a ratio = 0.24 (SE = 0.09), that indicates that most of the inflation in the distribution of the test statistics is caused by polygenicity, but other confounding factors are also present.

3.2 SNP heritability

Two approaches were used to estimate SNP heritability (h^2) in order to evaluate how much of the variation in the phenotypic trait could be ascribed to common additive genetic variation.

On the one hand, we used LDSC 1.0.0 (https://github.com/bulik/ldsc/). Only HapMap-3 SNPs from summary statistics of the GWAS meta-analyses and pre-computed LD scores (available on https://github.com/bulik/ldsc) were used in the analyses.

On the other hand, we used the genome-based restricted maximum likelihood analysis implemented in the software tool genome-wide complex traits analysis (GCTA-GREML; https://cnsgenomics.com/software/gcta/#Overview) (Lee et al., 2011). The samples (1-4) used for the cocaine dependence GWAS meta-analysis were merged using PLINK v1.9 and filtered using strict quality controls, keeping only SNPs with: MAF > 0.01, SNP call rate > 0.98 and individual call rate > 0.98 (resulting in 5,957,307 SNPs in 2,083 cases and 4,287 controls). To account for population stratification we performed PCA using smartPCA included in Eigensoft (previously described), and 10 first PCs and a dummy variable indicating genotyping-study were included in the analysis as covariates.

The SNP heritability was calculated on the liability scale ($h^2_{liability}$) using a prevalence of cocaine dependence of 1.1% in the population [17]. For LDSC it was estimated to be $h^2_{liability}$= 0.30 (SE = 0.06, *P* = 2.4e-07) and for GCTA-GREML $h^2_{liability}$= 0.26 (SE = 0.03, *P* < 0.01).

3.3 Partitioning heritability by functional annotation

Partitioning of the heritability by functional categories was done based on 53 functional overlapping annotations described in Finucane *et al.* (2015), but only 24 annotations were considered (annotations of 500bp-windows around each functional category were not considered). The pre-computed LDSC, the baseline model LD scores, regression weights and allele frequencies (based on the 1KGP3 European ancestry samples) were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/. Enrichment in the heritability of a functional category was defined as the proportion of SNP heritability explained divided by the

proportion of SNPs [18]. Significance threshold was calculated using the Bonferroni correction to control for multiple testing (P = 0.05/24 = 2e-03). The analysis revealed significant enrichment in the heritability by SNPs located in intronic regions (enrichment = 2.17; SE = 0.45; P = 1.2e-03), and nominal association for conserved regions (enrichment = 23.63; SE = 8.57; P = 4e-03) (Figure S3).

4. Functional annotation of SNPs

4.1 FUMA

Functional annotation of SNPs was performed using the FUMA web application (http://fuma.ctglab.nl/) [19]. This tool can be used to annotate, prioritize, visualize and interpret GWAS results. FUMA defines lead SNPs as signals that are significantly associated with the disorder (we used P < 1e-05) and independent to each other at $r^2 < 0.1$. For each lead SNP, FUMA defines a "Genomic risk locus", including all independent signals that are physically close or overlapping in a single locus. To evaluate the potential impact of the variants in the "Genomic risk locus", we considered annotations of functional consequences for those variants based on external reference data. In particular, we explored:

- eQTL: evaluation of expression quantitative trait loci using gene expression data from GTEx v6/v7 (https://www.gtexportal.org/home/) [20,21] and BRAINEAC (http://www.braineac.org/) [22].
- CADD v1.3: A deleterious score of variants computed by integrating 63 functional annotations. The higher the score, the more deleterious the variant (12.37 is the suggested threshold to be deleterious) (http://cadd.gs.washington.edu/) [23] .
- ANNOVAR: A variant annotation tool used to obtain functional consequences of SNPs on gene function (http://annovar.openbioinformatics.org/en/latest/) [24].

- RegulomeDB v1.1: A categorical score (from 1 to 7) representing regulatory functionality of SNPs based on eQTLs and chromatin marks. Score 1a means that those SNPs are most likely affecting regulatory elements and 7 means that those SNPs do not have any annotations (http://regulomedb.org/index) [25].
- 15-core chromatin state: The chromatin state represents accessibility of genomic regions (every 200bp) with 15 categorical states predicted by ChromHMM based on 5 chromatin marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) for 127 epigenomes. In this study we only used data available for the 13 brain tissues (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html) [26,27].

Finally we explored the GWAS-catalog e91 2018-02-06 (https://www.ebi.ac.uk/gwas/) [28], a database of reported SNP-trait associations, to see if the identified SNPs were previously associated to other traits (Table 2; Figure 2 and S4).

4.2 Gene-based association analysis

Gene-based association with cocaine dependence was estimated by MAGMA 1.05b [14] using the summary statistics from the GWAS meta-analysis. The SNP-wise mean model was used, in which the test statistic used was the sum of -log(SNP p-value) for SNPs located within the transcribed region (defined using NCBI 37.3 gene definitions). The gene p-value was calculated using a known approximation of the sampling distribution [29]. MAGMA accounts for gene size, number of SNPs in a gene and LD between markers. When using summary statistics in estimating gene-based p-values, MAGMA corrects for LD based on estimates from reference data with similar ancestry; for this we used the 1KGP3, European ancestry samples, as the reference [13]. We applied no padding around genes.

A total of 18,069 genes were analysed, and *HIST1H2BD* gene demonstrated significant gene-wise association with cocaine dependence (surpassing 10% FDR) (Figure S5 and Table S2).

4.3 Gene-set association analysis: Canonical pathways, BioCarta and GO gene-sets

Gene-set analyses were performed using MAGMA 1.05b. Based on the gene-based p-values generated as described in the previous section, we analysed sets of genes in order to test for enrichment in association signals in genes belonging to specific biological pathways or processes. MAGMA applies a competitive test to analyse whether the genes of a gene set are more strongly associated with the trait than other genes, while correcting for a series of confounding effects such as gene length and size of the gene set. In our analysis only genes on autosomes were included. For gene sets we used "All Canonical Pathways" (1,329 gene sets), "Gene Ontology" (4,436 gene sets) and "BioCarta" (217 gene sets) provided from MsigDB 5.1 (https://software.broadinstitute.org/gsea/msigdb/) [30]. Multiple testing corrections were performed for each gene set separately. When gene sets strongly overlap, the Bonferroni correction can be quite conservative, and for this reason we used an empirical multiple testing correction implemented in MAGMA, based on a permutation procedure.

None of the gene sets remained significantly associated with the disorder after correction for multiple testing (Table S3-5). Interestingly, from the first 10 Gene Ontology gene sets, 7 related to synapse organization, glutamatergic neurotransmission or brain functions.

5. Genetic correlation of cocaine dependence with comorbid conditions

5.1 Description of the summary statistics from comorbid conditions

We performed genetic correlation studies between cocaine dependence and previously described comorbid disorders or associated phenotypes using publicly available summary statistics (Table 2):

- Schizophrenia (SCZ) European meta-analysis: 34,241 cases and 45,604 controls, and 1,235 parent-affected offspring trios. In total: 15,358,498 SNPs [2].
- Attention deficit/hyperactivity disorder (ADHD) European meta-analysis: 19,099 cases and 34,194 controls. In total: 8,094,095 SNPs [31].
- Major depressive disorder (MDD): 59,851 cases and 113,154 controls. In total: 13,554,490 SNPs [32].
- Children's aggressive behavior (Child-Aggre) GWAS from EAGLE (Early Genetics and Lifecourse Epidemology Consortium): 18,988 individuals. Prior to analysis, data from this dataset was converted from hg18 to hg19 using the liftOver tool (http://genome.ucsc.edu/goldenPath/help/hg18ToHg19LiftOver.html). In total: 2,200,951 SNPs [33].

All of them are available on the PGC website, https://www.med.unc.edu/pgc/results-and-downloads.

- Antisocial behavior (ASB) meta-analysis: 16,400 individuals. In total: 7,795,277 SNPs [34]. Available on BroadABC website (http://broadabc.ctglab.nl/summary_statistics)
- Risk taking (RT) behavior from the UK Biobank: 325,821 individuals. In total: 10,894,597 SNPs. Available on https://sites.google.com/broadinstitute.org/ukbbgwasresults/home?authuser=0.

As a negative control we used summary statistics of vitamin D levels from the UK Biobank: 335,591 individuals. In total: 10,894,597 SNPs.

5.2 LDSC Genetic correlation

Genetic correlations (r_g) between cocaine dependence and six comorbid disorders/phenotypes (ADHD, SCZ, MDD, Child-Aggre, ASB, RT) were calculated using LDSC 1.0.0 [35]. In these analyses we used summary statistics from all samples and pre-computed LD scores from

HapMap3 SNPs, calculated on 378 phased European-ancestry individuals from the 1000Genomes Project (available on https://github.com/bulik/ldsc). Only results for markers present in the HapMap3 SNPs list with an imputation INFO score > 0.90 (this filter was applied only in datasets where this information was available) were included in the analysis. We selected a conservative significance threshold to control for multiple testing by applying the Bonferroni correction. As we tested genetic correlation between cocaine dependence and 7 phenotypes, significance threshold was set at $P < 7.1e-03$ ($P < 0.05/7$).

We found significant genetic correlations between cocaine dependence and SCZ ($r_g = 0.2$; SE = 0.05; $P = 1e-04$), ADHD ($r_g = 0.5$; SE = 0.08; $P = 1.6e-09$), MDD ($r_g = 0.4$; SE = 0.08; P = 6.6e-07) and RT ($r_g = 0.35$; SE = 0.06; $P = 9.1e-08$) but not with Child-Aggre ($r_g = 0.28$; SE = 0.23; $P = 0.22$) or ASB ($r_g = 0.58$; SE = 0.28; $P = 0.04$) (Figure 3A). No significant results were found for the negative control ($r_g = 0.08$; SE = 0.15; $P = 0.55$).

Furthermore, the genetic correlation of cocaine dependence with other traits available at LD Hub (http://ldsc.broadinstitute.org/ldhub/) [36] were evaluated. In total, 832 phenotypes were tested for genetic overlap with cocaine dependence, but we obtained valid results only for 690 and 109 demonstrated significant correlation after Bonferroni correction ($P < 7.24e-05$). Detailed information about significant genetic correlations can be found in Table S6 and Figure S6.

5.3 Polygenic risk scores for cocaine dependence

Polygenic Risk Scores (PRS) can be used to investigate the shared genetic etiology between cocaine dependence and comorbid phenotypes, and to test how these phenotypes can predict cocaine dependence. Using GWAS summary statistics results, the PRS on the discovery phenotype are calculated, and these are used as predictors of a target phenotype in a regression analysis. Using PRSice 2.1.0 software (https://github.com/choishingwan/PRSice)

[37] we analysed the proportion of genetic aetiology shared between cocaine dependence and comorbid psychiatric disorders (ADHD, SCZ and MDD) or associated phenotypes (RT, Child-Aggre and ASB).

In order to perform PRS analysis, the samples (1-4) used for the cocaine dependence GWAS meta-analysis were merged using PLINK v1.9 and were used as a target sample. After merging, quality control was performed and only SNPs with MAF > 0.01, SNP call rate > 0.98 and individual call rate > 0.98 were used (resulting in 5,957,307 SNPs in 2,083 cases and 4,287 controls). To assess population stratification we performed PCA using smartPCA included in Eigensoft (previously described), and 10 first PCs and a dummy variable indicating genotyping-study were included in the PRS analysis as covariates.

We used the summary statistics of the comorbid conditions (described above) as independent discovery samples. The discovery samples were clumped (r^2 < 0.1 in a 250-kb window) to remove SNPs in LD. Both variants with an imputation INFO score < 0.9 and ambiguous strand variants were removed from the analysis. Then, PRSs were estimated for each discovery sample using a wild range of meta-analysis p-value thresholds (P_T) between P_T = 1e-04 and P_T = 1 at increments of 5e-05. Summing over the markers abiding by the p-value threshold in the discovery set and weighting by the additive scale effect measure of the marker (log(OR) or β). For each P_T, the proportion of variance explained (R^2) by each discovery sample was computed by comparing the full model (PRS + covariates (10 PCs and study)) score to a reduced model (covariates only). The reported R^2 value is the difference between R^2 from the two models. For quantitative traits we performed linear regression analysis, and for qualitative traits we performed traits logistic regression and *Nagelkerke's pseudo-*R^2 values are shown.

We selected a conservative significance threshold to control for multiple testing by applying a Bonferroni correction. Euesden and colleagues recommend using a significance threshold of at least P = 0.004 in order to control for the high-resolution scoring approach of selecting the

most predictive PRS [37]. As we tested the most predictive PRS across each of the 7 discovery phenotypes, we divided the p-value by the number of tests performed (P = 0.004/7), which resulted in a significance threshold of P < 5.7e-04.

For all discovery samples, PRS significantly predict cocaine dependence: SCZ (pseudo-R^2 = 2.28%, P_T = 0.4911, P = 1.21e-26), ADHD (pseudo-R^2 = 1.39%, P_T = 0.04275, P = 4.5e-17), ASB (R^2 = 1.33%, P_T = 0.4055, P = 2.2e-16), MDD (pseudo-R^2 = 1.21%, P_T = 0.0129, P = 4.35e-15), RT (R^2 = 0.60%, P_T = 0.00135, P = 2.7e-08) and Child-Aggre (R^2 = 0.3%, P_T = 0.3552, P = 8.8e-05). No significant results were found for the negative control (R^2 = 0.07%, P_T = 0.03075, P = 0.06) (Figure 3B, S7 and S8).

References

1. Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, et al. A genome-wide association study of alcohol dependence. Proc Natl Acad Sci U S A. 2010;107: 5082–7.
2. Ripke S, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511: 421–427.
3. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81: 559–75.
4. Van Iperen EPA, Hovingh GK, Asselbergs FW, Zwinderman AH. Extending the use of GWAS data by combining data from different genetic platforms. PLoS One. 2017;12: e0172082.
5. Zuvich RL, Armstrong LL, Bielinski SJ, Bradford Y, Carlson CS, Crawford DC, et al. Pitfalls of merging GWAS data: Lessons learned in the eMERGE network and quality control procedures to maintain high data quality. Genet Epidemiol. 2011;35: 887–898.
6. Uh HW, Deelen J, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, et al. How to deal with the early GWAS data when imputing and combining different arrays is necessary. Eur J Hum Genet. 2012;20: 572–576.
7. Mitchell BD, Fornage M, McArdle PF, Cheng YC, Pulit SL, Wong Q, et al. Using previously genotyped controls in genome-wide association studies (GWAS): Application to the Stroke Genetics Network (SiGN). Front Genet. 2014;5: 95.
8. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38: 904–909.
9. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006;2: 2074–2093.
10. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012;9: 179–181.
11. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48: 1284–1287.
12. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An

integrated map of structural variation in 2,504 human genomes. Nature. 2015;526: 75–81.
13. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74.
14. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26: 2190–2191.
15. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots [Internet]. bioRxiv. 2014 May.
16. Bulik-Sullivan B, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47: 291–295.
17. Compton WM, Thomas YF, Stinson FS, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. Arch Gen Psychiatry. 2007;64: 566–576.
18. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47: 1228–1235.
19. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8: 1826.
20. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (80-). 2015;348: 648–660.
21. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550: 204–213.
22. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci. 2014;17: 1418–1428.
23. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46: 310–315.
24. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38: e164–e164.
25. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22: 1790–1797.
26. Ernst J, Kellis M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat Biotechnol. 2015;33: 364–376.
27. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518: 317–330.
28. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45: D896–D901.
29. Hou CD. A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. Stat Probab Lett. 2005;73: 179–187.
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci. 2005;102: 15545–15550.
31. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. Nat. Genet. 51, 63–75
32. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic

architecture of major depression. Nat Genet. 2018; 167577.
33. Pappa I, St Pourcain B, Benke K, Cavadino A, Hakulinen C, Nivard MG, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. Am J Med Genet Part B Neuropsychiatr Genet. 2016;171: 562–572.
34. Tielbeek JJ, Johansson A, Polderman TJC, Rautiainen M-R, Jansen P, Taylor M, et al. Genome-Wide Association Studies of a Broad Spectrum of Antisocial Behavior. JAMA Psychiatry. 2017;
35. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47: 1236–1241.
36. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33: 272–279.
37. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31: 1466–1468.

SUPPLEMENTARY FIGURES

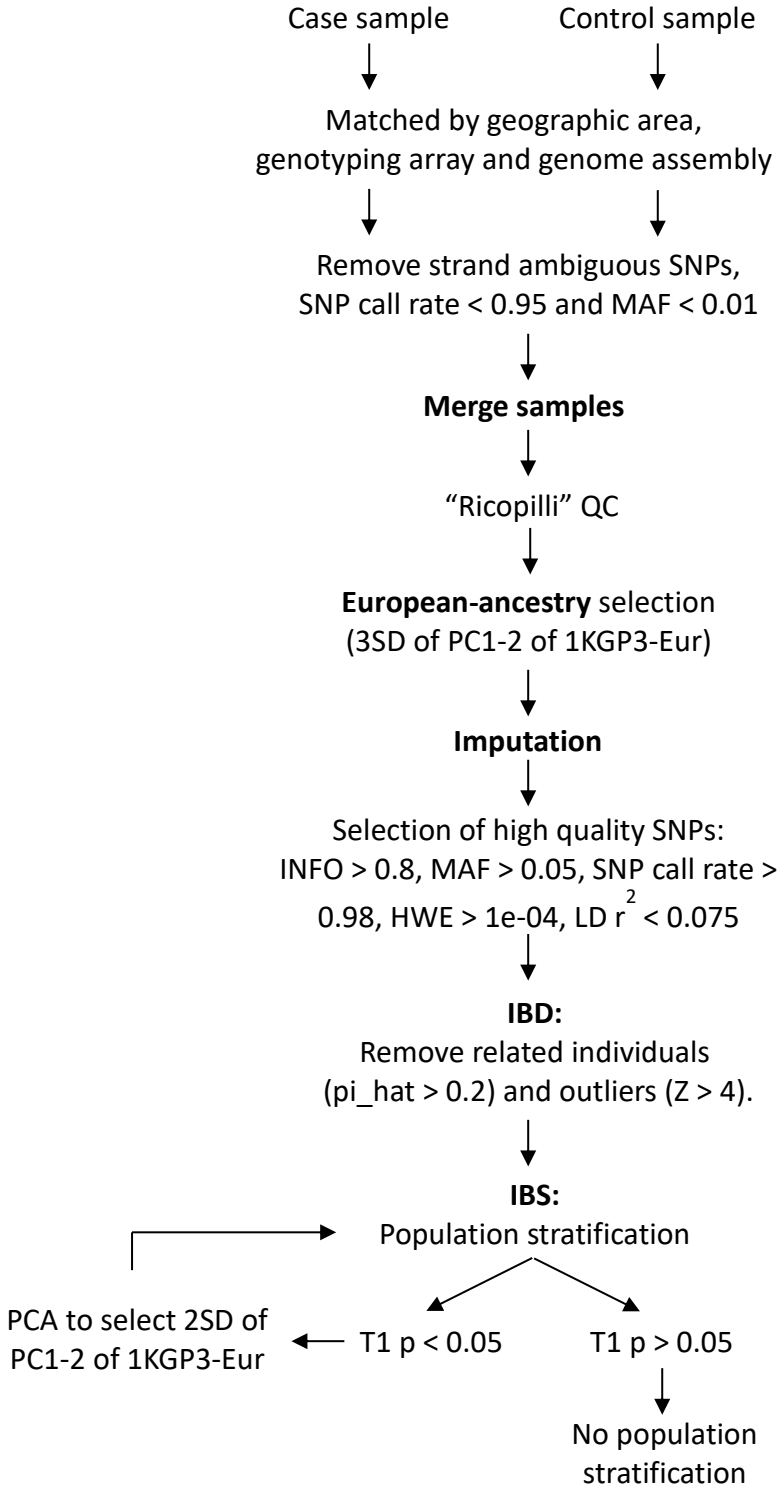


Figure S1. Flowchart illustrating steps to control for population stratification. MAF: minor allele frequency; 1KGP3-Eur: European individuals from 1000 Genomes Project Phase 3 (combined CEU, FIN, GBR, IBS, TSI); INFO: imputation info score; HWE: Hardy Weinberg equilibrium; IBD: identical by descent; IBS: identical by state

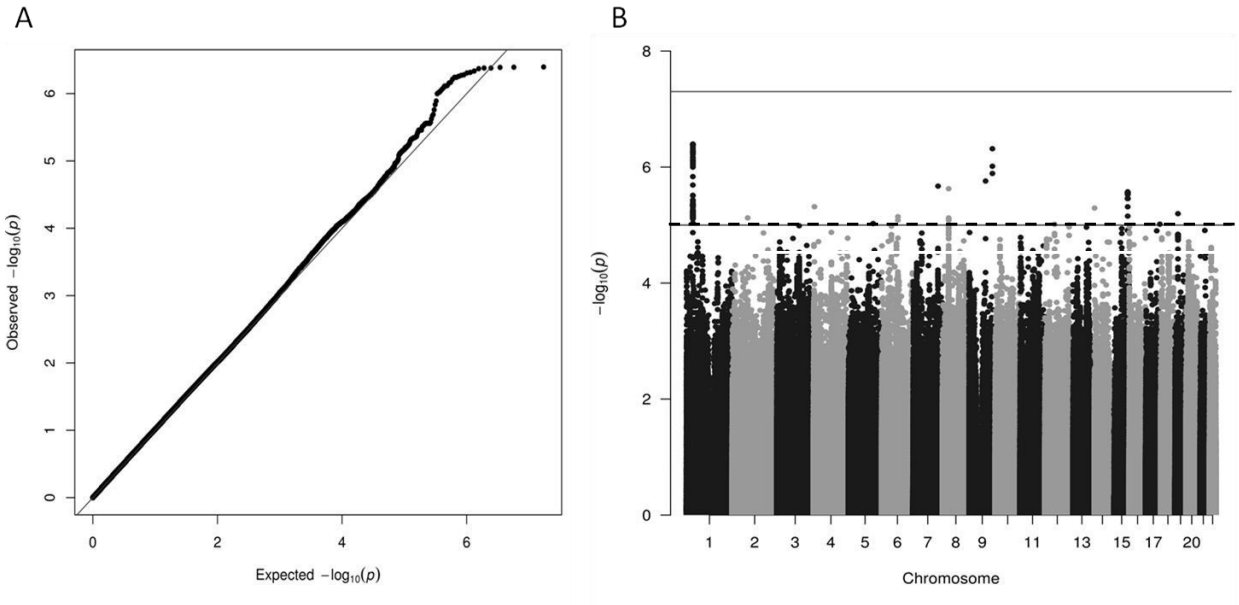


Figure S2. Plots from test for heterogeneity between samples in the GWAS meta-analysis. A) Q-Q plot and B) SNP-based Manhattan plot obtained for the heterogeneity test between samples in the cocaine dependence GWAS meta-analysis. Continuous line: threshold for genome-wide significance (*P* < 5e-08). Discontinuous line: threshold for suggestive associations (*P* < 1e-05).

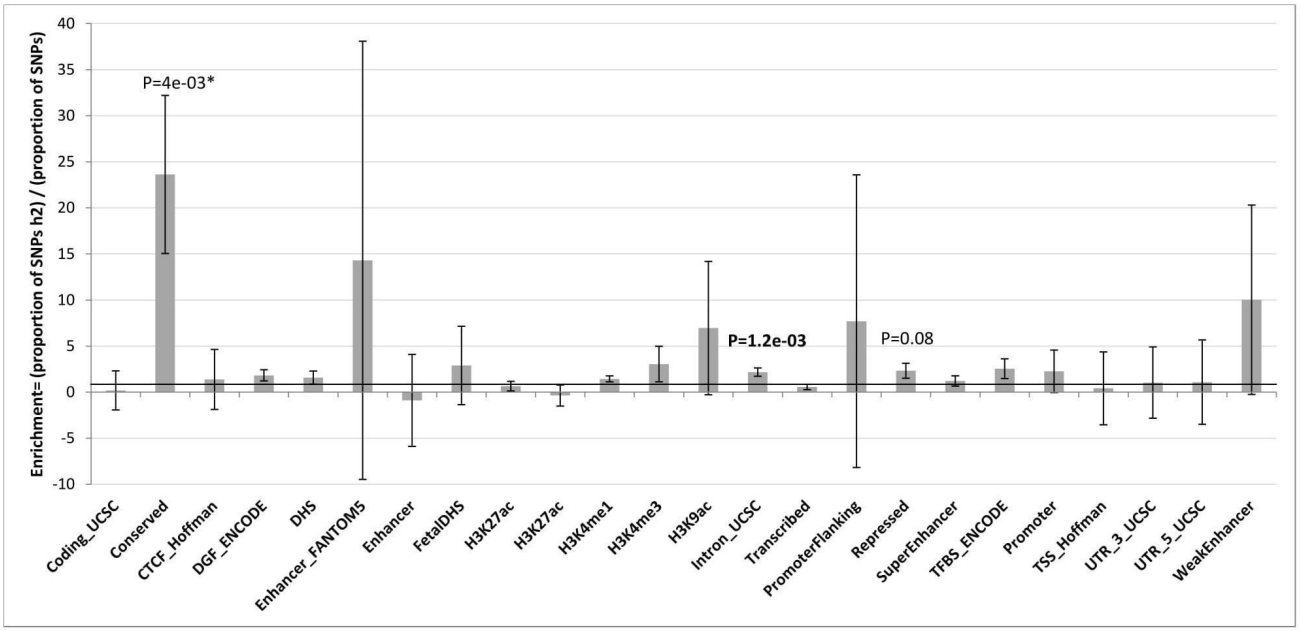


Figure S3. Partitioning of heritability (h2) by functional annotations. Enrichment by 24 functional annotations defined by Finucane et al. (2015). Error bars represent 95% confidence intervals. P-values for annotation categories with nominally significant enrichment are shown and * indicates significance after Bonferroni correction (*P* < 2e-03). The horizontal black line indicates no enrichment.

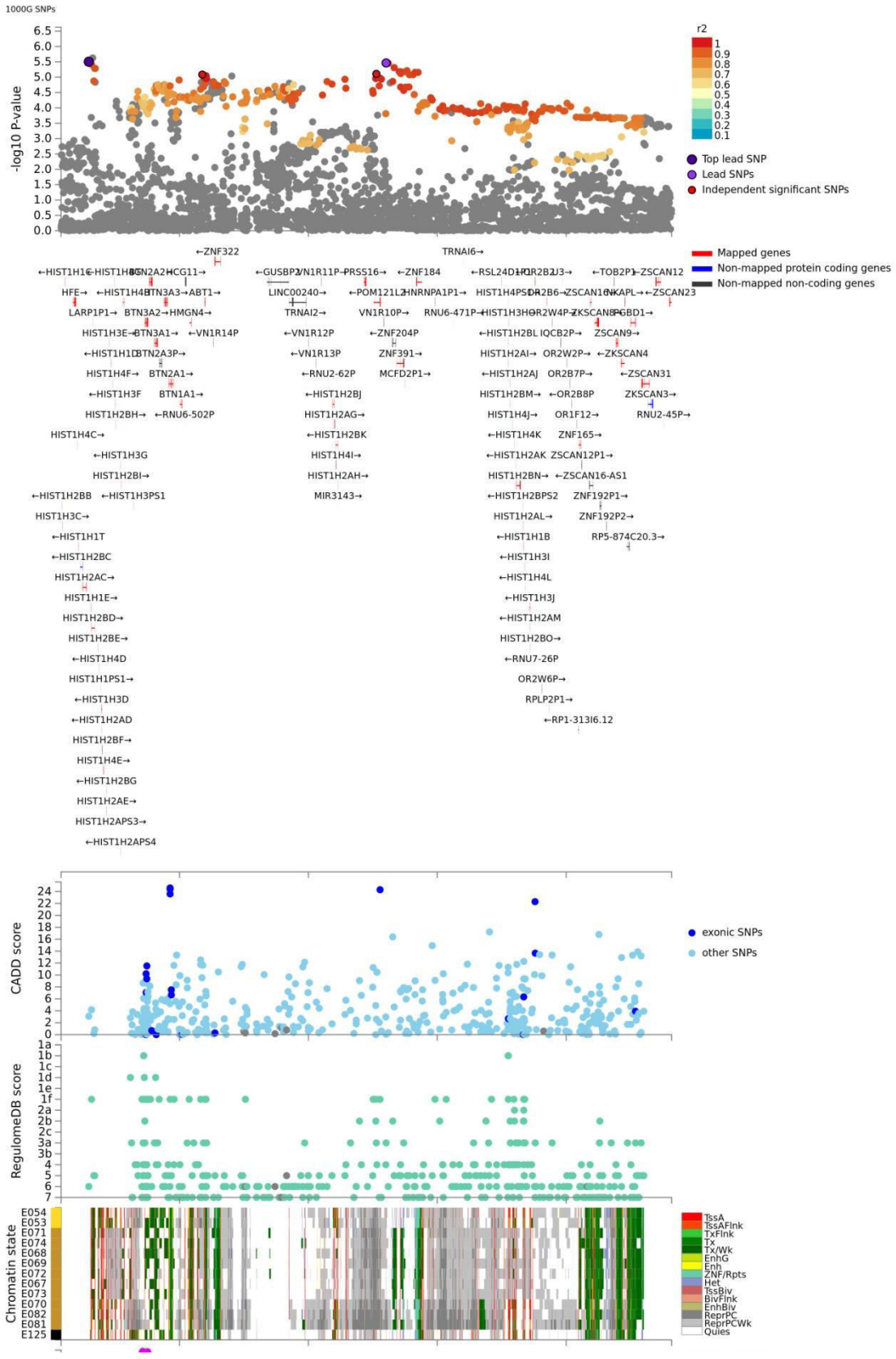

1000G SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
CADD score
exonic SNPs
other SNPs
RegulomeDB score
Chromatin state
TssA
TssAFlnk
TxFlnk
Tx
Tx/Wk
EnhG
Enh
ZNF/Rpts
Het
TssBiv
BivFlnk
EnhBiv
ReprPC
ReprPCWk
Quies

Figure S4. Regional plot with genes and functional annotations of genomic risk loci on Chr6 by FUMA. Genes prioritized by FUMA are highlighted in red. From the top, GWAS p-value (SNPs are colored based on r^2), CADD score (coding SNPs and other SNPs are colored blue and light blue, respectively), RegulomeDB score, 15-core chromatin state in the brain, eQTL p-value and chromatin interactions. eQTLs are plotted per gene and colored based on tissue type. Tissue/cell 19 types of epigenome ID are the following; E054: Ganglion eminence derived

primary cultured neurospheres, E053: Cortex derived primary cultured neurospheres, E071: Hippocampus middle, E074: Substantia nigra, E068: Anterior caudate, E069: Cingulate gyrus, E072: Inferior temporal lobe, E067: Angular gyrus, E073: Dorsolateral prefrontal cortex, E070: Germinal matrix, E082: Fetal brain female, E081: Fetal brain male and E125: NH-A Astrocytes primary cells. eQTLs are plotted per gene and colored based on tissue type.

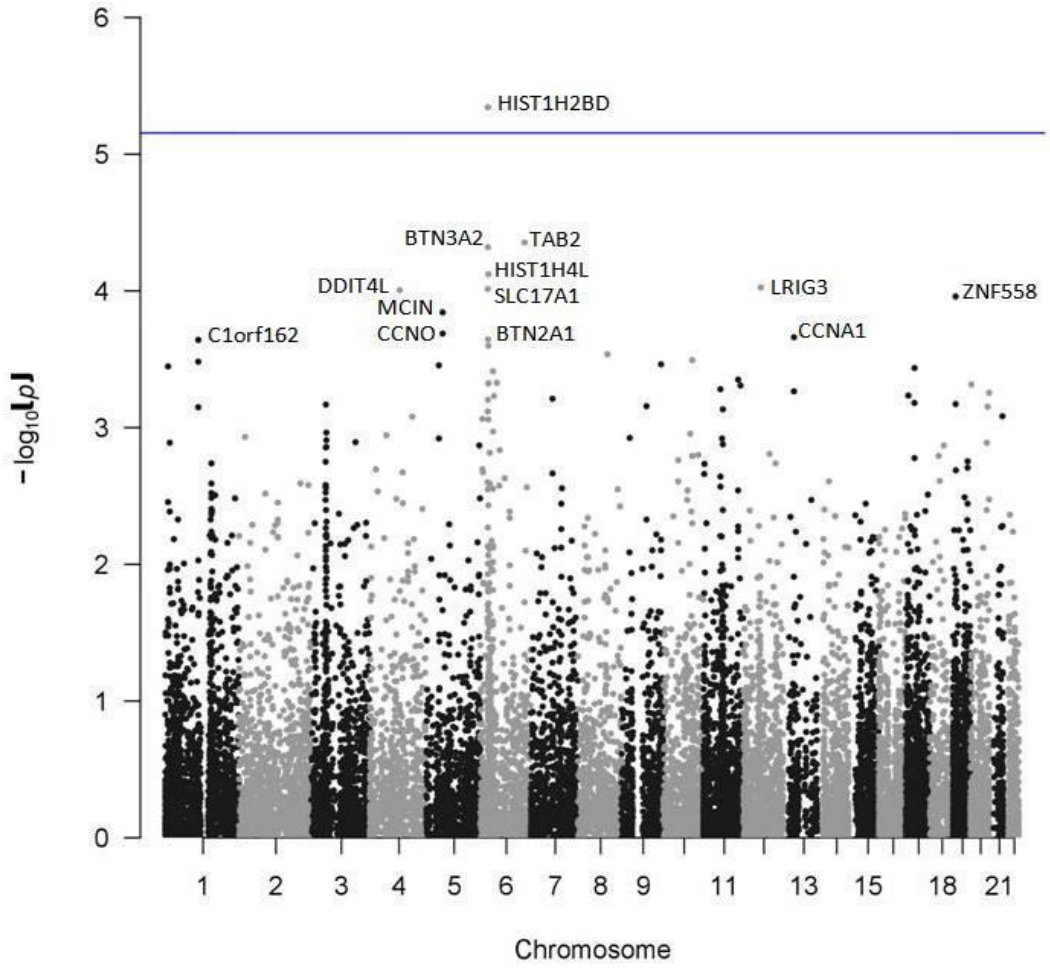


Figure S5. Manhattan plot of gene-based p-values in the cocaine dependence meta-analysis. Blue line: threshold for 10% FDR significance.

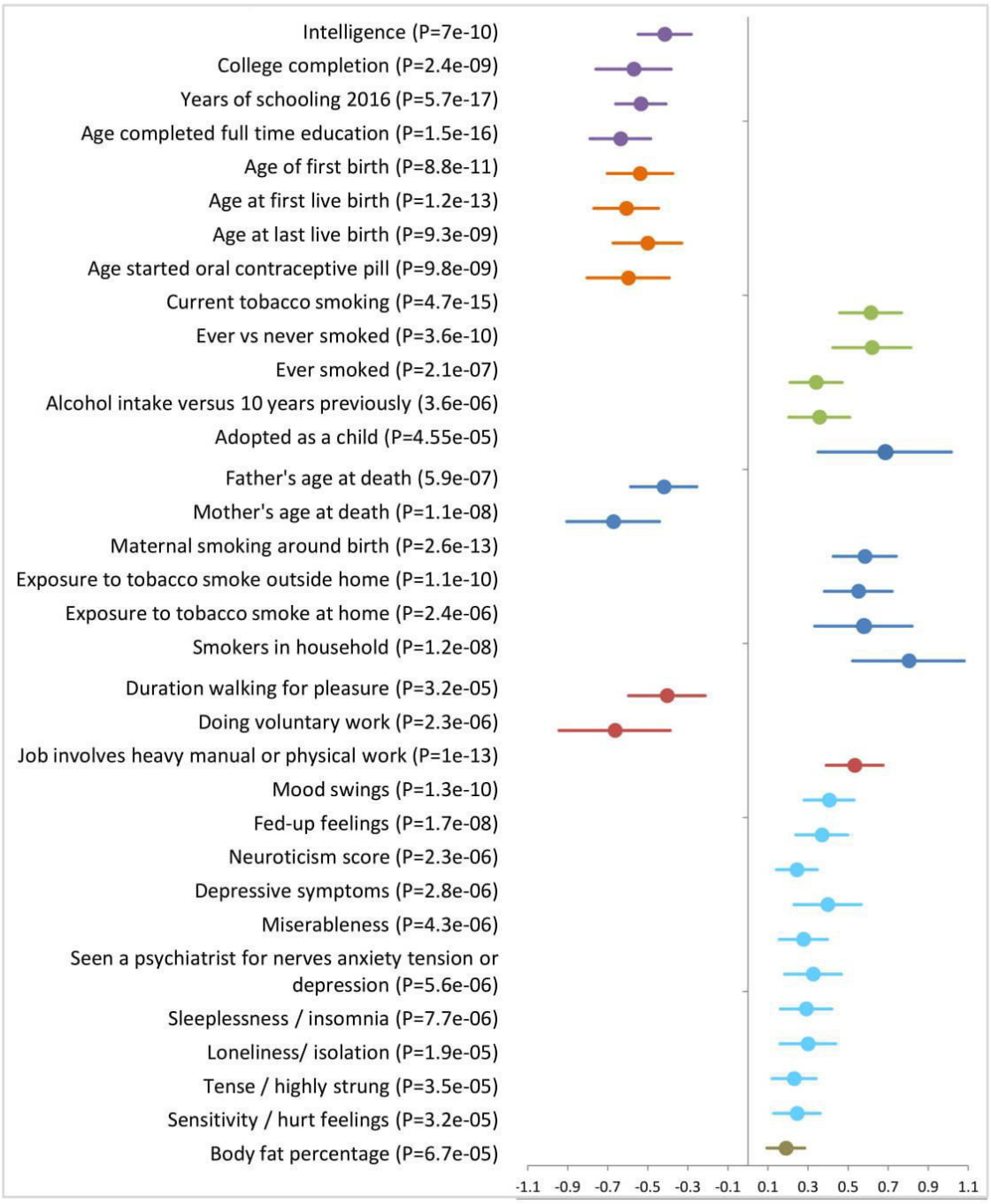


Figure S6. Genetic correlation of cocaine dependence with other traits from LDhub. Significant genetic correlations between cocaine dependence and other traits after Bonferroni correction for testing a total of 690 traits available at LDhub. The most interesting results are shown here; see table S6 for the full output of this analysis. Groups defined by colours, from the top: In purple, educational achievements; in orange, reproductive traits; in green, alcohol and tobacco exposure; in dark blue, familiar situation; in red, exercise; in light blue, psychological and psychiatric traits; in brown, physical condition. Error bars indicate 95% confidence limits. The significance threshold was set at $P < 7.24e\text{-}05$.

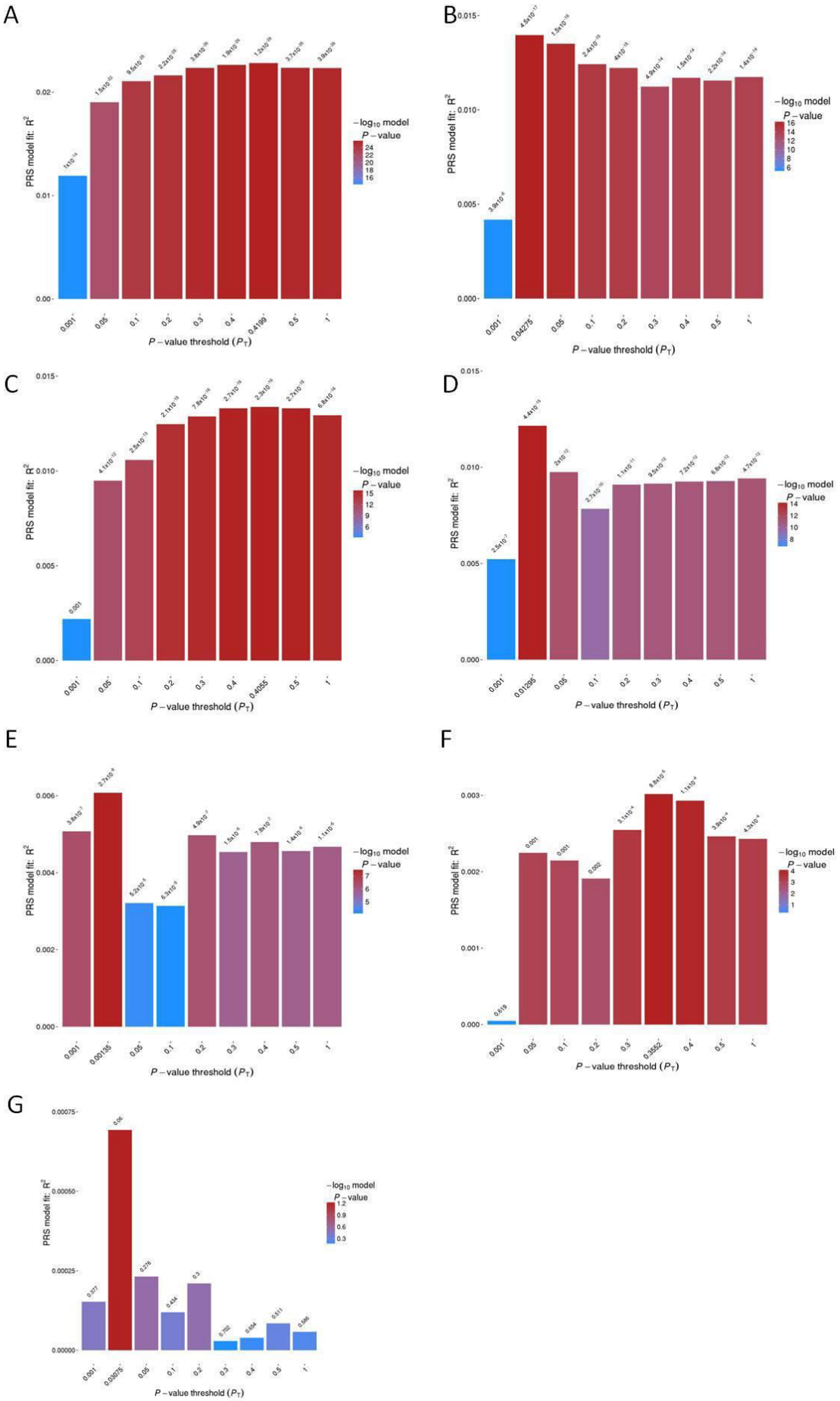

A
B
C
D
E
F
G
PRS model fit: R^2
P – value threshold (P_T)
$-\log_{10}$ model P – value

Figure S7. Polygenic risk score results from the seven tested phenotypes. A) Schizophrenia, B) Attention deficit/hyperactivity disorder, C) Antisocial behavior, D) Major depressive disorder, E) Risk-taking behavior, F) Children's aggressive behavior and G) Negative control (Vitamin D levels). P-value threshold (P_T) represents the p-value at the cut-off for inclusion of SNPs in the polygenic risk score. Values on top of the bars represent p-values for the regression models. The significance threshold was set at $P < 5.7e-04$.

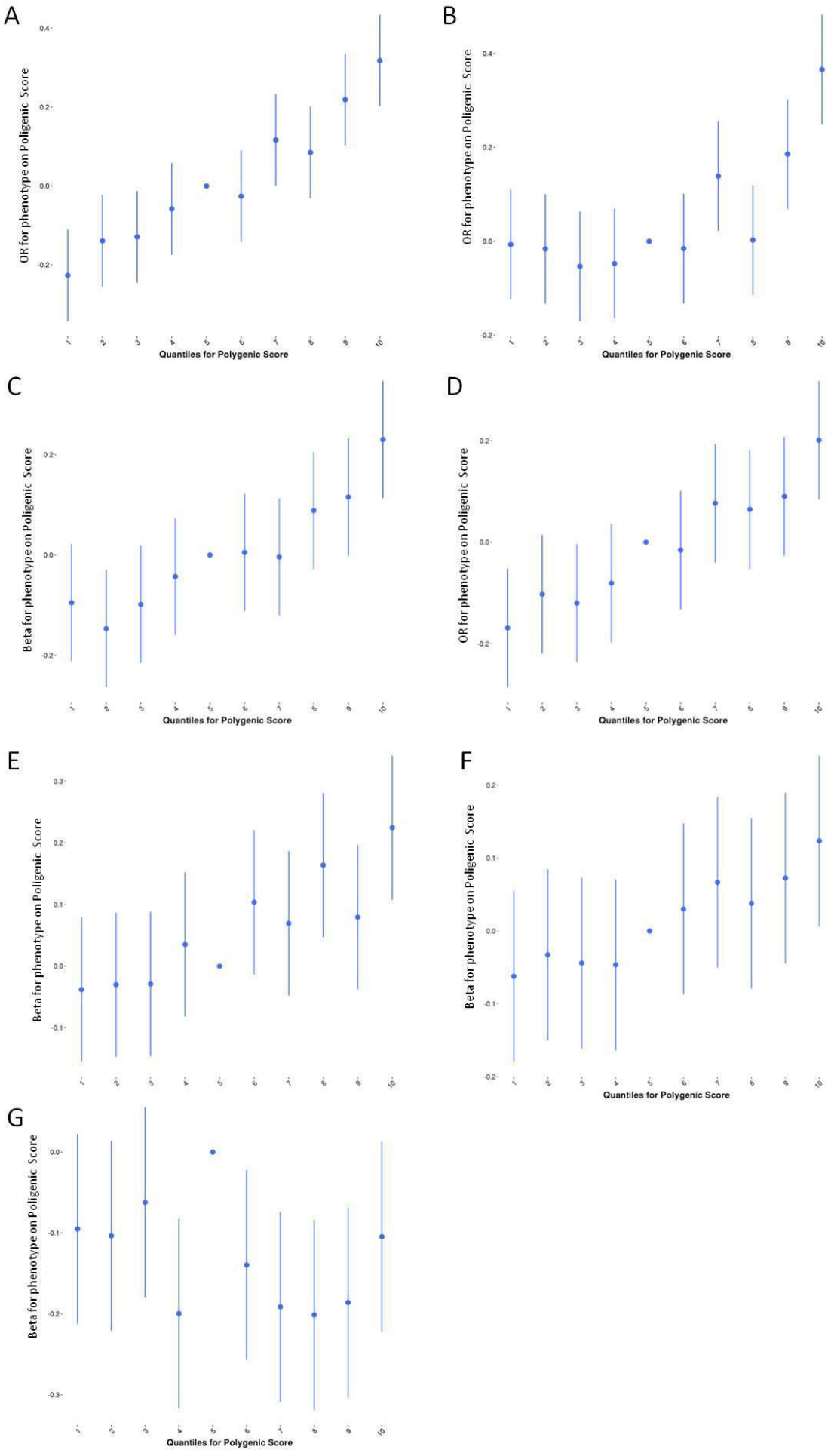

A
OR for phenotype on Poligenic Score
Quantiles for Polygenic Score
B
OR for phenotype on Poligenic Score
Quantiles for Polygenic Score
C
Beta for phenotype on Poligenic Score
Quantiles for Polygenic Score
D
OR for phenotype on Poligenic Score
Quantiles for Polygenic Score
E
Beta for phenotype on Poligenic Score
F
Beta for phenotype on Poligenic Score
Quantiles for Polygenic Score
G
Beta for phenotype on Poligenic Score
Quantiles for Polygenic Score

Figure S8. Deciles of polygenic risk scores plotted against effects on phenotypes A) Schizophrenia, B) Attention deficit/hyperactivity disorder, C) Antisocial behavior, D) Major depressive disorder, E) Risk-taking behavior, F) Children's aggressive behavior and G) Negative control (Vitamin D levels). A regression is performed with phenotype as outcome and each 10% quantile separately, whereby the effect size of each quantile is compared to the central quantile as a reference, such that each polygenic score in the quantile in question is coded 1 and each polygenic score in the reference quantile is coded 0. In each regression, the covariates used in the main analyses are included. OR, odds ratio. Error bars indicate 95% confidence limits.

SUMMARY ARTICLE 3

"Explorant l'impacte de la variació comuna en gens de miRNAs en el trastorn per dèficit d'atenció amb hiperactivitat"

El trastorn per dèficit d'atenció amb hiperactivitat (TDAH) és un trastorn neuropsiquiàtric multifactorial prevalent l'etiologia del qual és encara força desconeguda. Contribueixen a la patologia tant factors genètics com ambientals, amb diversos mecanismes epigenètics també implicats. La modulació epigenètica per microRNAs (miRNAs), una classe d'ARN no codificant, s'ha revelat com un procés clau en l'origen i desenvolupament dels trastorns neuropsiquiàtrics. Els miRNAs actuen com a reguladors de l'expressió gènica en el sistema nerviós, tot modulant el desenvolupament del cervell i la neuroplasticitat. En aquest estudi hem dut a terme estudis d'associació basats en SNPs i en gens que interroguen tots els miRNAs autosòmics coneguts (~ 1700) utilitzant les dades procedents de l'estudi GWAS més gran realitzat fins ara en TDAH (20.000 casos i 35.000 controls). Ens hem centrat en variants comunes situades en gens que codifiquen miRNAs i en els seus possibles elements reguladors. Hem identificat associacions significatives entre el TDAH i 19 SNPs de 12 miRNAs intragènics (situats dins de gens que codifiquen proteïnes). Quatre dels miRNA assenyalats (miR-6079, miR-6734, miR-6735, miR-3666) també mostren una associació significativa amb el fenotip en l'anàlisi basada en gens. Les variants associades estan situades en potencials regions reguladores de l'expressió dels miRNAs, o bé a la regió promotora del gen hoste. Hem investigat, en els miRNA associats, l'expressió en cervell, els gens diana, les vies diana i els homòlegs en altres espècies. La majoria dels gens diana validats per als miRNAs identificats s'havien relacionat prèviament amb d'altres malalties neurològiques. Les anàlisis d'anotació funcional apunten als gens miR-7-1 i miR-3666 com a candidats prometedors per al trastorn. Les dianes conegudes de miR-7-1 inclouen gens prèviament implicats en l'empatia cognitiva (*EIF4E*) i el trastorn bipolar (*EGFR*), així com gens (*SLC17A7*) que s'expressen exclusivament en teixits cerebrals. El miR-3666 regula *TAC1* i *MEOX2*, el primer associat amb comportaments de risc i nivells elevats de nerviosisme, i el darrer relacionat amb el volum intracranial i de diverses regions cerebrals. L'anàlisi de vies funcionals ha assenyalat la regulació mediada per miRNAs de gens que codifiquen receptors de serotonina, crucials en la regulació de funcions neurològiques i implicats en moltes malalties del sistema nerviós. En aquesta mateixa línia, miR-4271 i miR-5193, assenyalats en l'anàlisi basada en SNPs, inhibeixen els gens *HTR1D* i *HTR4*, respectivament. Els resultats que presentem obren noves vies per a l'estudi del paper dels miRNAs en el TDAH.

Reference

Shivalikanjli A*, Reinbold C*, Cabana-Domínguez J, Stangier S, Demontis D, Borglum A, Faraone SV, Nöthen MM, Mattheisen M, Fernàndez-Castillo N, Forstner A, Cormand B. Exploring the impact of common variation in microRNA genes in attention-deficit/hyperactivity disorder. *Manuscript in preparation*.

*Equally contributed

EXPLORING THE IMPACT OF COMMON VARIATION IN MICRO-RNA GENES IN ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

Running title: ADHD and miRNAs

Anu Shivalikanjli[1,2,3,4,*], Céline Reinbold[5,6,*], Judit Cabana-Domínguez[1,2,3,4], Susann Stangier[7,8], Ditte Demontis[9,10,11], Anders Borglum[9,10,11], Stephen V. Faraone[12], Markus M. Nöthen[7,8], Manuel Mattheisen[9,10,11], Noèlia Fernàndez-Castillo[1,2,3,4], Andreas J. Forstner[7,8,&], Bru Cormand[1,2,3,4,&]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Catalonia, Spain

[2] Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Spain

[3] Institut de Biomedicina de la Universitat de Barcelona (IBUB), Catalonia, Spain

[4] Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Barcelona, Catalonia Spain

[5] Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland

[6] Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland

[7] Institute of Human Genetics, University of Bonn, Bonn, Germany

[8] Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany.

[9] The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

[10] Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

[11] Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark

[12] Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA.

[*]Joint first authors that equally contributed

[&]Senior authors that equally contributed

Corresponding authors:

Bru Cormand: Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Catalonia, Spain, Avinguda Diagonal 643, edifici Prevosti, 3ª planta, 08028, Barcelona, Catalonia, Spain. Email: bcormand@ub.edu

Andreas Forstner: Institute of Human Genetics, University of Bonn, Sigmund-Freud-Strasse 25, Bonn 53127, Germany. Email: forstner@uni-bonn.de

ABSTRACT

Attention-deficit/hyperactivity disorder (ADHD) is a common multifactorial neuropsychiatric disorder which aetiology is still largely unknown. Both genetic and environmental factors contribute to the disorder, with epigenetic mechanisms playing a role, too. Epigenetic modulation by microRNAs (miRNAs), a class of non-coding RNAs, has emerged as a key process in neuropsychiatric disorders. MiRNAs act as regulators of gene expression in the nervous system, where they exert their influence on brain development and neuroplasticity. In the present study, we conducted SNP-based and gene-based association studies across all known autosomal miRNAs (~1700) using the largest genome-wide association dataset of ADHD to date (20,000 cases and 35,000 controls). We focused on common variants located in miRNA genes and in their putative regulatory elements. Nineteen SNPs in 12 intragenic miRNAs showed a significant association with ADHD in the SNP-based analysis. The associated variants are located in the putative regulatory regions of the miRNA genes or in the promoter region of the host protein-coding gene. We inspected the brain expression, target genes, target pathways and homologs of the associated miRNAs. Most of the validated target genes for the highlighted miRNAs have previously been related to other neurological diseases. Our functional annotation results point at *miR-7-1* and *miR-3666* as promising candidates for the disorder. Known targets of *miR-7-1* include genes previously implicated in cognitive empathy (*EIF4E)* and bipolar disorder (*EGFR)* as well as genes (*SLC17A7)* that are expressed exclusively in brain tissues. *MiR-3666* targets *TAC1* and *MEOX2,* with the former associated with risk-taking and feeling nervous traits, and the latter linked to several brain regions and intracranial volumes. Pathway analysis revealed a miRNA-mediated regulation of serotonin receptor genes, well-known contributors to neurological functions and diseases. In this line, *miR-*

4271 and *miR-5193,* highlighted by our SNP-based analysis, were shown to inhibit *HTR1D* and *HTR4,* respectively. The present results direct further research to elucidate the precise involvement of miRNAs in ADHD.

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a multifactorial neurodevelopmental disorder with a high estimated heritability (76%)[1–3]. Most often characterized by the presence of inattentive and/or hyperactivity symptoms, the disorder affects 5% of the children and adolescents worldwide and persists in 2.5% of the adult population[3]. Both common[3] and rare variants[4] contribute to the impaired neurobiological mechanisms that underlie this complex phenotype, in combination with environmental risk factors[2]. We know that common variation accounts for a substantial fraction of ADHD liability, with single-nucleotide polymorphisms (SNPs) contributing to 22% of the phenotype variance[3]. However, the specific molecular underpinnings of the disorder remain still largely unknown, with only 12 susceptibility loci in the largest genome-wide association study (GWAS) to date that attain genome-wide significance and a few genes showing an increased burden of rare variants identified through whole-exome sequencing (WES). Thus, the pivotal functional networks in ADHD pathology are not evident yet, probably due to the high degree of polygenicity of the disorder, the small effect sizes of individual variants, and multiple regulatory and signaling mechanisms.

MicroRNAs (miRNAs), the most abundant type of small non-coding RNAs[5], are epigenetic modulators of the expression of up to 80% of the genome[6], including those genes that drive the development and function of the central nervous system[7,8]. The genes encoding miRNAs are thought to be evolutionarily conserved[9] and are organized as separate units or in clusters. Approximately 57% of miRNA-coding genes are embedded within protein-coding genes, also termed host genes[6]. MiRNA-mediated regulatory networks operate post-transcriptionally and are complex due to the ‘multivalent’ or one-to-many and synergic

relationships between miRNAs and their targeted transcripts. The expression levels of miRNAs are particularly high in brain tissues[10] and these molecules are often involved in neurogenesis, synaptic plasticity, neuronal survival, differentiation, neurite projection and memory formation[8,11,12].

The role of miRNAs in the development of major psychiatric disorders has extensively been addressed, beyond the traditional landscape of protein-coding genes. Individuals and groups of miRNAs have been investigated in schizophrenia, bipolar disorder, MDD, ASD and ADHD by employing high-throughput sequencing, gene expression analyses of blood and post-mortem brain tissues and genetic association case-control study designs[7]. Their role is most extensively documented in schizophrenia[6] where hundreds of miRNAs have been reported to be dysregulated[13], and the condition has recently been associated with a global increase in miRNA biogenesis and expression in the cerebral cortex[14]. A genome-wide examination of miRNA genes in bipolar disorder has implicated nine miRNAs[15]. In MDD, miRNAs let-7b, let-7c[16] and miR-124-3p[17] have been identified as potential biomarkers, as their target genes involve those previously implicated in the disorder. Differentially expressed miRNAs are reported in the cerebellar cortex of autistic group, whose targets include autism-risk genes like *SHANK3* and *NRXN1*[18].

The first genome-wide integrative study of miRNA and mRNA profiles in peripheral blood mononuclear cells of medication-naive individuals with ADHD identified 79 microRNAs that showed aberrant expression levels as compared to controls, with three of them, miR-26b-5p, miR-185-5p, and miR-191-5p, being highly predictive for diagnostic status in an independent dataset of ADHD cases[19]. Expression studies have further revealed significantly altered circulation levels of miR-let-7d[20], miR-18a-5p, miR-22-3p, miR-24-

3p, miR-106b-5p, miR-155a-5p and miR-107[21] in the serum and blood of patients. The function of let-7d was further investigated in the spontaneously hypertensive rat (SHR), a model for ADHD, where it was suggested to modulate the tyrosine hydroxylase (TH) production critical to dopamine pathway, by downregulating TH's upstream effector galectin-3 in the brain prefrontal cortex[22]. A recent study identified 13 potential miRNA biomarkers that show differential expression in patients[23]. Interestingly, many of the pinpointed miRNAs can be dysregulated in multiple disorders[24], perhaps lending further support to the existence of shared genetics or biological pathways among neurodevelopmental disorders[25]. Candidate gene association studies have linked sequence variants in the miR-34b/c locus[25] and the miR-183–96–182 cluster[26] to ADHD. Finally, two of the top 12 loci revealed by the first GWAS on ADHD with genome-wide significant signals included miRNA genes (miR3666 and miR9-2). As the potential of sub-threshold variants being true risk loci has lately been recognized[27,28], the plausibility that many of the influential miRNA loci might not be revealed in a genome-wide approach cannot be ignored.

We hypothesize that common variants located in miRNA genes or in their putative regulatory elements may play a crucial role in the disorder by altering the expression of the corresponding protein-coding genes. We hereby present the first study that systematically captures common SNP variation in miRNAs, a crucial epigenetic component, at genome-wide scale, utilizing the largest ADHD GWAS meta-analysis available to date.

MATERIALS AND METHODS

Selection of miRNA genes

We retrieved the genomic locations of all autosomal miRNA precursor sequences from miRBase release 21[29]. The miRBase search tool[30] was used to fetch miRNA sequences distributed as clusters that were defined with an inter-miRNA distance of less than 10kb. The genomic coordinates of the resultant miRNA clusters were determined by the two furthest positions of the miRNA transcripts at the extreme ends of each cluster. All coordinates primarily in GRCh38 were converted to their equivalents in GRCh37 using the UCSC liftOver tool[31].

We assigned miRNAs to one of the following categories, which were also used to establish the genomic windows for the inclusion of the putative regulatory elements of these miRNAs: (i) each 'singleton' non-clustered miRNA gene - a flank of 10 kb upstream and 5 kb downstream; (ii) clustered miRNAs sequences that are transcribed in the same direction - 10 kb upstream and 5 kb downstream from the cluster; (iii) clustered miRNA sequences transcribed in different directions - a 10 kb flank at either end of the cluster (Figure 1). We then used information from the miRIAD database[32] to identify miRNAs located within protein-coding genes (from now on, host genes) and to define the start position of the host genes. In those cases where intragenic miRNAs were transcribed in the same orientation as their host genes, we included a 10 kb region upstream from the transcription start site (TSS) of the host gene in our analyses.

Selection of tag variants and case-control association analysis

Bi-allelic variants with a minimum allele frequency (MAF) of 5% in the EUR subpopulation of the 1000 Genomes Phase 3 Project[33] were extracted for the selected genomic regions (including miRNAs and their potential regulatory sequences) by employing the VCFtools package[34]. Tagging variants were selected from all the obtained

variants on the basis of their pairwise linkage disequilibrium (LD) in the 1000 Genomes Phase 3 EUR reference panel and with an $r^2 > 0.85$ using Haploview 4.2 software[35]. The tags were tested for their association with ADHD using the summary statistics of the GWAS meta-analysis carried out by the PGC and iPSYCH on 19,099 cases and 34,194 controls of European ancestry[3]. The statistically significant associations were corrected for multiple testing considering a 5% False Discovery Rate (FDR) calculated using R package *qvalue*[36].

Functional annotation of the highlighted miRNAs

The regional association plots for the 500 kb region centered on the associated variants were generated using LocusZoom[37]. Any signal present within the miRNA locus if in a high or moderate LD (r^2>0.6) with the index SNP of the region was considered to be a miRNA-associated signal. We explored BrainSpan Atlas[38,39], miRmine[40], miRIAD[32] and early human brain development spatio-temporal assessment of microRNA expression from Ziats and Rennert[41] to assess the expression levels of miRNAs in brain tissues. The experimentally validated target mRNAs and putative biological pathways were deciphered using the Ingenuity Pathway Analysis 8.8 software (IPA) (http://www.ingenuity.com/products/ipa; Ingenuity Systems, Redwood city, CA, USA) where we used all the highlighted miRNAs as input. We used the human pre-miRNAs SNPs reported in miRNASNP2 database[42] to identify any patterns of LD with the significantly associated variants. GTEx data[43] was utilized to investigate expression quantitative trait loci (eQTLs) information. Orthologs for the miRNA genes in other model species were searched for in the Alliance of Genome Resources web portal[44], and the level of conservation of the miRNA genes across model species was retrieved using microRNA

Viewer[45]. The degree of conservation was calculated as the proportion of identical bases between the two sequences[45].

Functional annotation of miRNA target genes

We retrieved the expression levels of the target genes for the highlighted miRNAs in the brain tissues from the GTEx database. NHGRI-EBI GWAS Catalog was used to decipher whether the target genes had previously been a hit in any psychiatric GWAS[46]. The interactive visualization of association p-value results and linkage disequilibrium patterns for a genomic region of interest was done using the LDassoc tool[47].

RESULTS

Genome-wide miRNA association analysis

Our analysis involved 1,761 autosomal miRNA genes of 1,881 published miRNAs (miRBase v21), of which 1,355 miRNA sequences are encoded individually and the remaining 406 organized into 135 gene clusters. A total of 1,754 miRNA sequences were successfully retained post assembly lift over, 879 of those intragenic, i.e. located within protein-coding genes. These 1,754 miRNAs were covered by 22,423 tag variants. We inspected these variants for association with ADHD in the summary statistics of the largest ADHD GWAS meta-analysis of 8,094,094 markers, which contained 76.3% of our tag variants. We identified 19 significant associations with ADHD (5% FDR, $p \leq 4.77e\text{-}05$) highlighting 12 miRNAs (Table 1). All these miRNAs are located within introns of host protein-coding genes. The associated variants are located in the putative regulatory regions of the miRNA genes or in the promoter regions of the host genes (Table 1). Two of the

Table 1. miRNA variants associated with ADHD overcoming 5% FDR

			Alleles			Freq A1					
Variant	**Chr**	**Pos**	**A1**	**A2**	**OR (A1)**	**Cases**	**Controls**	**meta-analysis p**	**miRNA gene**	**Location in miRNA region**	**Gene**
rs839764	**1**	**43833374**	**A**	**T**	**0.92802**	**0.368**	**0.387**	**8.71E-08**	**miR-6734**	**5'**	***ELOVL1***
rs2251802	**1**	**43917637**	**A**	**G**	**0.93072**	**0.407**	**0.425**	**1.68E-07**	**miR-6735**	**3'**	***SZT2***
rs56319043	**1**	**44171211**	**T**	**C**	**0.89987**	**0.248**	**0.266**	**1.37E-11**	**miR-6079**	**phg**	***ST3GAL3***
rs37453	**1**	**44295047**	**A**	**G**	**1.07821**	**0.659**	**0.643**	**1.10E-07**	**miR-6079**	**5'**	***ST3GAL3***
rs3011216	**1**	**44296439**	**C**	**G**	**1.07724**	**0.463**	**0.448**	**3.53E-07**	**miR-6079**	**5'**	***ST3GAL3***
rs3011217	**1**	**44303266**	**A**	**G**	**1.08893**	**0.282**	**0.266**	**1.51E-08**	**miR-6079**	**5'**	***ST3GAL3***
rs11708763	3	20172672	T	C	1.06109	0.701	0.692	5.46E-05	miR-3135a	5'	*KAT2B*
rs58936320	3	49305140	A	T	0.93342	0.519	0.534	2.53E-05	miR-4271	5'	*C3orf62*
rs1799844	3	49847642	A	G	1.07907	0.192	0.181	9.03E-06	miR-5193	5'	*UBA7*
chr3:50310286	3	50310286	A	AAATAATAATAAT	0.93165	0.665	0.678	5.59E-05	miR-6872	5'	*SEMA3B*
rs10250550	7	1889599	T	C	1.05739	0.434	0.418	4.07E-05	miR-4655	5'	*MAD1L1*
rs2045292	**7**	**114287116**	**A**	**T**	**0.92839**	**0.314**	**0.33**	**6.06E-07**	**miR-3666**	**5'**	***FOXP2***
rs7782412	7	114290415	T	C	1.06695	0.595	0.581	4.07E-05	miR-3666	5'	*FOXP2*
rs7799269	7	114290491	A	C	1.07358	0.512	0.497	4.37E-06	miR-3666	5'	*FOXP2*
rs76100764	9	86582925	T	TAATA	1.06599	0.746	0.737	3.28E-05	miR-7-1	3'	*HNRNPK*
rs296886	9	86592026	A	G	1.06919	0.783	0.774	3.73E-05	miR-7-1	5'	*HNRNPK*
rs296894	9	86598444	T	G	0.93463	0.201	0.209	5.28E-05	miR-7-1	phg	*HNRNPK*
rs143942298	16	15701372	T	C	1.10794	0.0985	0.0915	1.24E-05	miR-6506	3'	*KIAA0430*
rs605921	16	24204897	A	G	1.06173	0.501	0.488	4.77E-05	miR-1273h	5'	*PRKCB*

Variant: Single nucleotide polymorphism, indel and structural variant; Chr: Chromosome; Pos: Position (build hg19); A1: Allele 1; A2: Allele 2; Underlined allele: risk allele for ADHD; OR: Odds Ratio (calculated on A1); Freq A1: Frequency of allele 1; meta-analysis p: p-value obtained from the PGC+iPSYCH meta-analysis; phg: Promoter of host gene; Gene: Host genes for intragenic miRNAs - all miRNAs are intragenic; In bold: Overcome Bonferroni correction.

highlighted loci, on chromosomes 1 and 7, have been reported as among the top ADHD risk loci in the largest ADHD GWAS meta-analysis published to date[3].

Follow-up of miRNA associations—regional association plots and miRNA brain expression

A visual inspection of the regional association plots detected miRNA-associated signals for six of the nine miRNAs (miR-6079, miR- 6734, miR-6735, miR- 5193, miR-4655, miR-7-1) (Supplementary Fig 1). Eleven out of the 12 highlighted miRNAs were found to be expressed in brain according to various expression databases (Supplementary Tables 1-3). For instance, miRIAD shows that six of the 12 highlighted miRNAs are expressed in different brain tissues, especially in cerebellum (Supplementary Table 1)), while for the rest, this information was not available. In this tissue-wise expression data across the brain, cerebellum, heart, testis and kidney, miR-6734 and miR-7-1 were more expressed in the brain and cerebellum than in other reported tissues (Supplementary Table 1). Of interest is miR-4655, which was shown to be exclusively expressed in the brain (Supplementary Table 1). According to miRmine, one mature transcript each of miR-6735 (miR-6735-5p) and miR-1273h (miR-1273h-5p), and both mature sequences of miR-7-1 (hsa-miR-7-1-3p and hsa-miR-7-5p), are brain-expressed (Supplementary Table 2). The two mature transcripts of miR-7-1 are found in nearly equal amounts in the brain (Supplementary Table 2). In addition, the expression profiles of sixteen cortical and subcortical structures of human brain revealed the presence of miRNAs in cerebellar cortex (miR-7-1, miR-3135a), primary somatosensory cortex (miR-3666, miR-4271, miR-4655-3p), primary visual cortex (miR-4655-5p) and ventral parietal cortex (miR-5193) (Supplementary Table 3).

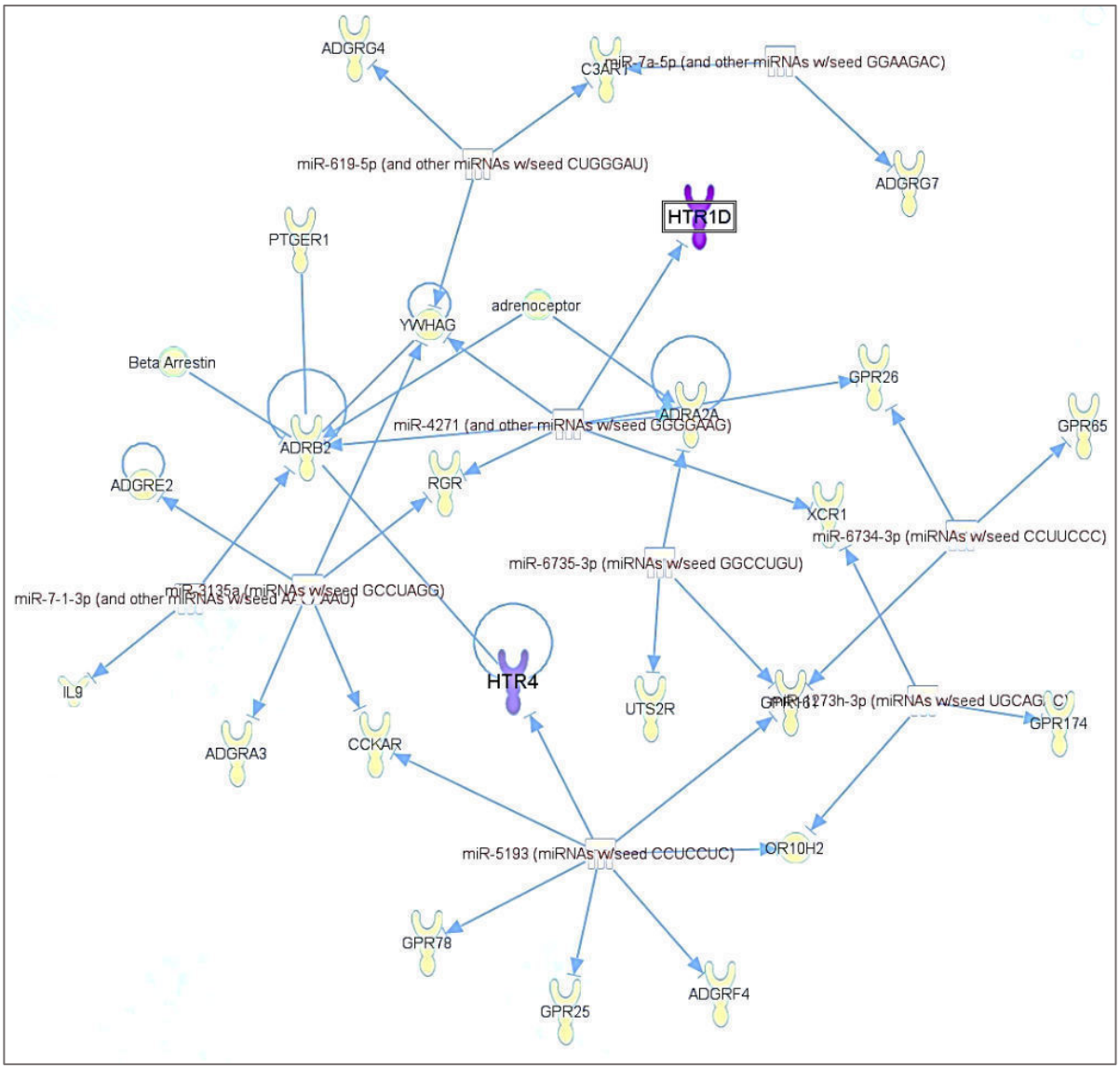


Fig 1. IPA pathway analysis for the highlighted miRNAs. All miRNAs were given as input. Nine focal miRNAs are depicted in the network. The pathway is redicted to be involved in neurological disorder. Serotonin receptors are colored dark.

Targets for associated miRNAs

Three of the highlighted miRNAs - miR-7-1, miR-3666, and miR-1273h have validated target sites according to IPA analysis (18, 9 and 1 mRNA targets, respectively) (Table 2). The target genes of miR-7-1 include *SLC17A7, SNCA, NEFM, SEPT3, RAF1* and *MKNK2,* which are expressed in more than one brain tissue. It is noteworthy that *SLC17A7, SNCA, NEFM* and *SEPT3* are expressed exclusively in the brain and not in the other tissues reported in GTEx. Of all the target genes, *SLC17A7* shows the highest expression in cerebellum, cerebellar hemisphere, hippocampus, amygdala and cortex. MiR-3666 targeted *TAC1* is moderately expressed in the caudate, nucleus accumbens and putamen. MiR-1273h has only one validated target, androgen receptor (AR), which shows a low regional specificity in human brain.

eQTL analysis of associated variants using GTEx data

We subsequently explored the impact of the identified variants on the expression of the nearby miRNA genes. This analysis is particularly relevant, as all the miRNAs identified by us lie within protein-coding genes. Since eQTL data for miRNAs are largely underrepresented in gene expression datasets, we could not establish miR-eQTL connections for our highlighted miRNAs. Nonetheless, we inspected whether the ADHD-associated variants at our miRNA loci are eQTLs for the host protein-coding genes in brain tissues (Supplementary Table 4). Of the 19 miRNA variants significantly associated with ADHD, rs3011217 and rs11708763 are eQTLs in brain regions for their host protein-coding genes *ST3GAL3* and *KAT2B*, respectively, eight variants are eQTLs for protein-coding genes in the region other than the host gene, while the remaining nine variants were not found to be brain eQTLs for any protein-coding gene (Supplementary Table 4). A

Targets for associated miRNAs

Three of the highlighted miRNAs - miR-7-1, miR-3666, and miR-1273h have validated target sites according to IPA analysis (18, 9 and 1 mRNA targets, respectively) (Table 2). The target genes of miR-7-1 include *SLC17A7, SNCA, NEFM, SEPT3, RAF1* and *MKNK2,* which are expressed in more than one brain tissue. It is noteworthy that *SLC17A7, SNCA, NEFM* and *SEPT3* are expressed exclusively in the brain and not in the other tissues reported in GTEx. Of all the target genes, *SLC17A7* shows the highest expression in cerebellum, cerebellar hemisphere, hippocampus, amygdala and cortex. MiR-3666 targeted *TAC1* is moderately expressed in the caudate, nucleus accumbens and putamen. MiR-1273h has only one validated target, androgen receptor (AR), which shows a low regional specificity in human brain.

eQTL analysis of associated variants using GTEx data

We subsequently explored the impact of the identified variants on the expression of the nearby miRNA genes. This analysis is particularly relevant, as all the miRNAs identified by us lie within protein-coding genes. Since eQTL data for miRNAs are largely underrepresented in gene expression datasets, we could not establish miR-eQTL connections for our highlighted miRNAs. Nonetheless, we inspected whether the ADHD-associated variants at our miRNA loci are eQTLs for the host protein-coding genes in brain tissues (Supplementary Table 4). Of the 19 miRNA variants significantly associated with ADHD, rs3011217 and rs11708763 are eQTLs in brain regions for their host protein-coding genes *ST3GAL3* and *KAT2B*, respectively, eight variants are eQTLs for protein-coding genes in the region other than the host gene, while the remaining nine variants were not found to be brain eQTLs for any protein-coding gene (Supplementary Table 4). A

single variant appeared as eQTL for multiple genes. For instance –rs3011217, located in the *ST3GAL3* on chr1 is an eQTL for *ST3GAL3* in cortex but also for *ARTN* in brain regions of cerebellum, hippocampus, putamen, caudate, amygdala, cortex and nucleus accumbens. Another variant, rs1799844, located within *UBA7* showed up as an eQTL in brain regions for eleven genes (Supplementary Table 4).

Orthologs and conservation in model species

An orthologous gene with a high degree of evolutionary conservation (97%) has been identified in mouse for miR-3666, species in which homologs for only two of the highlighted miRNAs (miR-3666 and miR-4271) have been reported. Orthologs for human miR-7-1 are present in mouse and rat as mmu-miR-7-1 and rno-miR-7a-1, respectively. miR-137 has a known ortholog in mouse as mmu-mir-137.

Table 2: Experimentally validated targets of miRNAs

miRNA	Symbol	Source database	Target gene
hsa-miR-3666	miR-130a-3p (and other miRNAs w/seed AGUGCAA)	TarBase, TargetScan Human	*TAC1*
		TargetScan Human, miRecords	*ZFPM2*
		Ingenuity Expert Findings, TargetScan Human	*ATG2B*
		TarBase, TargetScan Human, miRecords	*CSF1*
		Ingenuity Expert Findings	*DICER1*
		TarBase, TargetScan Human, miRecords	*HOXA5*
		TarBase, TargetScan Human, miRecords	*MAFB*
		TarBase, TargetScan Human, miRecords	*MEOX2*
		Ingenuity Expert Findings	*SMAD4*
hsa-miR-1273h-5p	miR-1273h-5p (and other miRNAs w/seed UGGGAGG)	Ingenuity Expert Findings, TargetScan Human	*AR*
hsa-miR-7-5p	miR-7a-5p (and other miRNAs w/seed GGAAGAC)	TargetScan Human, miRecords	*EGFR*
		Ingenuity Expert Findings, TargetScan Human	*EIF4E*
		TarBase, miRecords	*FOS*
		TargetScan Human, miRecords	*IRS1*
		TargetScan Human, miRecords	*IRS2*
		Ingenuity Expert Findings, TargetScan Human	*KMT5A*
		Ingenuity Expert Findings	*MAPKAP1*
		Ingenuity Expert Findings, TargetScan Human	*MKNK1*
		Ingenuity Expert Findings	*MKNK2*
		Ingenuity Expert Findings	*NEFM*
		Ingenuity Expert Findings	*p70 S6k*
		Ingenuity Expert Findings, TargetScan Human, miRecords	*PAK1*
		TargetScan Human, miRecords	*RAF1*
		miRecords	*SEPT3*

	miRecords	*SLC17A7*
	miRecords	*SLC3A2*
	TargetScan Human, miRecords	*SNCA*
	miRecords	*SYNE1*
Experimentally targets found for three of the highlighted miRNAs using IPA. Symbol: miRNAs whose seed sequence match the highlighted miRNAs.		

DISCUSSION

Expression of the miRNAs

The reported genetic associations for miRNA-coding genes indicate that miRNAs may have a role in to the development of ADHD. All 12 significantly associated miRNAs are located in host genes and 11 of these are brain-expressed (Supplementary Tables 1-3). Multiple studies have identified miRNA-host gene pairs showing cross-species conserved co-location, co-expression, and that the intronic miRNAs are derived in parallel from the same primary transcripts as their host genes[48,49]. Thus, miRNA host genes' expression profiles might serve as a possible proxy for those of the resident miRNAs[50], and can regulate protein-encoding mRNAs in a synergistic pattern to fine-tune the protein output[51]. The more recent miRNA gene families in mammals have an inclination to largely express themselves in two nervous system tissues, cortex and cerebellum[52]. Half of the highlighted miRNAs in this study are expressed in cerebellum and hippocampus, which are of known importance in ADHD development. MiRNAs show differential expression both within and between brain regions[41]. For instance - miR-7-1, one of the highlighted genes in our study, is differentially expressed between PFC and cerebellum during late childhood development. The greatest shifts in miRNA expression can occur soon after birth, during the transition from infancy to early childhood[41].

Pathway analysis

The involvement of miRNAs in ADHD it is grounded on the downstream effects of their target genes that were highlighted in the pathway analysis. MiRNA network analysis shows that miR-3135a mediates the expression of *ADGRE2 (EMR2)*, SNPs in which have been

associated with major depression[53].. MiR-4271 regulates the *YWHAG* gene, previously found associated with schizophrenia[46,54] and encoding a protein that mediates signal transduction by binding to phosphoserine-containing proteins. MiR-7 targets *C3AR1*, encoding a central protein in the complement system, which has a role in synapse loss in psychiatric illnesses beyond the canonical immune functions[55]. C3aR deletion in mouse models markedly increased physiological and behavioral responses to innate anxiety-provoking stimuli[56]. The gene for the cholecystokinin A receptor (*CCKAR*), involved in dopamine release in the CNS, is targeted by miR-3135a and miR-5193. A SNP in this gene was found to modulate language laterization, and the schizophrenia risk allele of the polymorphism was related to reduced functional asymmetry[57]. Variation in the *CCKAR* locus can also affect superior frontal gyrus grey matter volume[46,58]. The orphan receptor GPR78, targeted by miR-5193, lies within a region which showed linkage to bipolar affective disorder (BPAD) and association with schizophrenia in the Scottish population. The GPR78 mRNA also has a potential role in the functioning of the hypothalamic-pituitary-adrenal (HPA) axis and in pregnancy, thus possibly connecting prenatal insults to the pathogenesis of psychiatric illness[59]. A paralog of GPR78 is GPR26, another gene shown in the network analysis, targeted by miR-4271 and miR-6734. GPR26 encodes a protein distantly related to the serotonin receptors and is expressed exclusively in brain[59,60]. This polypeptide is important for emotion regulation in mice, a function probably mediated by the phosphorylation of CREB (cAMP responsive element-binding protein (CREB) - neuropeptide Y (NPY) signaling) in the central amygdala[60]. In general, critical functions for members of this GPCR family have been demonstrated in neurodevelopment. The identified pathway further depicts how miRNAs can interact with neurotransmitter,

psychiatric risk and immune gene systems in interwoven networks, and this further supports the neuroimmune crosstalk.

Genes targeted by the associated miRNAs

MiRNAs that show differential spatio-temporal expression in early human brain development have been shown to be highly enriched in genes associated with child-onset psychiatric conditions including autism, schizophrenia, bipolar disorder, and depression[41]. Overall, a high number of validated targets of miRNAs pinpointed in this study have previously been related to psychiatric and neurological diseases. MiR-7-1 targeted *EGFR* is located in one of the top regions of a GWAS for lithium-responsive bipolar disorder[61] and associated with brain thalamus volume. *EIF4E,* targeted by miR-7-1, is associated with cognitive empathy[62] and depressive episodes in bipolar disorder[63]. *SNCA* harbors common and rare variants implicated in the risk to Parkinson disease [64]. *MKNK1* has been associated with schizophrenia in Ashkenazi Jews population[65]. *SLC17A7* (*VGLUT1*) is expressed in a brain-specific manner, and encodes the vesicular glutamate transporter 1 which mediates the uptake of glutamate into synaptic vesicles at presynaptic nerve terminals of excitatory neural cells. Altered levels of SLC17A7 have been consistently reported for cognitive decline, schizophrenia, MDD and bipolar disorder[66–68]. MiR-3666 targets the *TAC1* gene, which encodes several peptide hormones and neuropeptides thought to function as neurotransmitters and to induce behavioral responses. SNPs in the *TAC1* gene have been associated with risk-taking[69] and feeling nervous traits[70] in previous large scale GWASs. Another miR-3666 target, *MEOX2,* is associated with brain regions volume, total intracranial and subcortical volumes[71]. Much of the previous literature of miRNAs is in the context of greater studied psychiatric disorders, but the extensive pleiotropic mechanisms

acting in the human genome under the influence of environment may implicate additional disorders, including ADHD. Therefore, supported by brain-expression, validated brain-expressed targets, and presence of homologs in model species, mirR-7-1 and miR-3666 arise as strong candidates for further analyses.

Also, one of the significantly highlighted genes, miR-6079, is located in a region on chromosome 1 that showed genome-wide association in the largest GWAS meta-analyses of ADHD[3]. As this is a high-LD region, with several protein-coding genes (*ST3GAL3*, *ARTN*, *KDM4A*), but also the miRNA, the locus requires further refinement to infer whether the association signal is related to the protein-coding genes and/or also the miRNA gene. It is worth mentioning that while around 40% miRNAs are known to be intergenic, all the significant signals reported in our genome-wide study arise for the miRNAs located within protein-coding genes. How the miRNA expression is regulated continues to be largely unknown. Recent studies have identified significant miRNA eQTLs in the mouse brain. An enrichment of brain-related pathways has also been shown among miRNA targets with significant miRNA-eQTLs[72].

The present study investigated common variants at the miRNA loci and flanking sequences to capture possible regulatory regions. Our study has to major strengths: (i) The work includes the highest set of miRNA genes investigated so far in ADHD and (ii) we comprehended the genomic organization of miRNA genes as intragenic, singletons and clusters, in addition to the transcriptional orientation, and established the genomic coordinates for the gene and its putative regulatory regions. However, one of the weaknesses of the approach is that it did not allow for the inquiry of SNPs with *trans*-eQTL

effects on miRNAs and it has been previously suggested that ~50% of the identified miRNA eQTLs are *trans*-eQTLs[73].

TABLES AND FIGURES

SUPPLEMENTARY MATERIAL

REFERENCES

1. Schachar, R. Genetics of Attention Deficit Hyperactivity Disorder (ADHD): Recent Updates and Future Prospects. *Curr. Dev. Disord. Reports* **1**, 41–49 (2014).

2. Grimm, O., Kittel-Schneider, S. & Reif, A. Recent developments in the genetics of ADHD. *Psychiatry Clin. Neurosci.* (2018). doi:10.1111/pcn.12673

3. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

4. Demontis, D. *et al.* Whole-Exome Sequencing Reveals Increased Burden of Rare Functional and Disruptive Variants in Candidate Risk Genes in Individuals With Persistent Attention-Deficit/Hyperactivity Disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 521–523 (2016).

5. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 1910–1922 (2014).

6. Caputo, V., Ciolfi, A., Macri, S. & Pizzuti, A. The Emerging Role of MicroRNA in Schizophrenia. *CNS Neurol. Disord. - Drug Targets* **14**, 208–221 (2015).

7. O'Connor, R. M., Gururajan, A., Dinan, T. G., Kenny, P. J. & Cryan, J. F. All Roads Lead to the miRNome: miRNAs Have a Central Role in the Molecular Pathophysiology of Psychiatric Disorders. *Trends Pharmacol. Sci.* **37**, 1029–1044 (2016).

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

9. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–R29 (2006).

10. Tang, Y., Liu, D., Zhang, L., Ingvarsson, S. & Chen, H. Quantitative analysis of miRNA expression in seven human foetal and adult organs. *PLoS One* **6**, e28730 (2011).

11. Xu, B., Karayiorgou, M. & Gogos, J. A. MicroRNAs in psychiatric and neurodevelopmental disorders. *Brain Res.* **1338**, 78–88 (2010).

12. Lu, J. & Clark, A. G. Impact of microRNA regulation on variation in human gene expression. *Genome Res.* **22**, 1243–54 (2012).

13. Wang, J., Wang, Y., Yang, J. & Huang, Y. microRNAs as novel biomarkers of schizophrenia (Review). *Exp. Ther. Med.* **8**, 1671–1676 (2014).

14. Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A. & Cairns, M. J.

Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Mol. Psychiatry* **15**, 1176–89 (2010).

15. Forstner, A. J. *et al.* Genome-wide analysis implicates microRNAs and their target genes in the development of bipolar disorder. *Transl. Psychiatry* **5**, e678–e678 (2015).

16. Gururajan, A. *et al.* MicroRNAs as biomarkers for major depression: a role for let-7b and let-7c. *Transl. Psychiatry* **6**, e862–e862 (2016).

17. Roy, B., Dunbar, M., Shelton, R. C. & Dwivedi, Y. Identification of MicroRNA-124-3p as a Putative Epigenetic Signature of Major Depressive Disorder. *Neuropsychopharmacology* **42**, 864–875 (2017).

18. Abu-Elneel, K. *et al.* Heterogeneous dysregulation of microRNAs across the autism spectrum. *Neurogenetics* **9**, 153–161 (2008).

19. Sánchez-Mora, C. *et al.* Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. *Neuropsychopharmacology* 1 (2018). doi:10.1038/s41386-018-0297-0

20. Wu, L. H. *et al.* Circulating MicroRNA Let-7d in Attention-Deficit/Hyperactivity Disorder. *NeuroMolecular Med.* **17**, 137–146 (2015).

21. Kandemir, H. *et al.* Evaluation of several micro RNA (miRNA) levels in children and adolescents with attention deficit hyperactivity disorder. *Neurosci. Lett.* **580**, 158–162 (2014).

22. Wu, L. *et al.* A novel function of MicroRNA Let-7d in regulation of galectin-3 expression in attention deficit hyperactivity disorder rat brain. *Brain Pathol.* **20**, 1042–1054 (2010).

23. Wang, L.-J. *et al.* Blood-Bourne MicroRNA Biomarker Evaluation in Attention-Deficit/Hyperactivity Disorder of Han Chinese Individuals: An Exploratory Study. *Front. Psychiatry* **9**, 227 (2018).

24. Mellios, N. & Sur, M. The Emerging Role of microRNAs in Schizophrenia and Autism Spectrum Disorders. *Front. psychiatry* **3**, 39 (2012).

25. Garcia-Martínez, I. *et al.* Preliminary evidence for association of genetic variants in pri-miR-34b/c and abnormal miR-34c expression with attention deficit and hyperactivity disorder. *Transl. Psychiatry* **6**, e879–e879 (2016).

26. Sánchez-Mora, C. *et al.* Evaluation of single nucleotide polymorphisms in the miR-183–96–182 cluster in adulthood attention-deficit and hyperactivity disorder (ADHD) and substance use disorders (SUDs). *Eur. Neuropsychopharmacol.* **23**, 1463–1473 (2013).

27. Wang, X. *et al.* Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* **5**, (2016).

28. Gagliano, S. A. *et al.* Allele-Skewed DNA Modification in the Brain: Relevance to a Schizophrenia GWAS. *Am. J. Hum. Genet.* **98**, 956–962 (2016).

29. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).

30. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2007).

31. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590-8 (2006).

32. Hinske, L. C. *et al.* miRIAD-integrating microRNA inter- and intragenic data. *Database (Oxford).* **2014**, (2014).

33. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

34. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

35. Barrett, J. C. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.* **4**, pdb.ip71-pdb.ip71 (2009).

36. John D. Storey with contributions from Andrew J. Bass and Alan Dabney and David Robinson. qvalue: Q-value estimation for false discovery rate control. (2015). Available at: http://github.com/jdstorey/qvalue. (Accessed: 24th November 2017)

37. Pruim, R. J. *et al.* LocusZoom: Regional visualization of genome-wide association scan results. in *Bioinformatics* **27**, 2336–2337 (2011).

38. BrainSpan: Atlas of the Developing Human Brain. Available at: https://www.brainspan.org/static/home. (Accessed: 25th July 2020)

39. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).

40. Panwar, B., Omenn, G. S. & Guan, Y. miRmine: A Database of Human miRNA Expression Profiles. *Bioinformatics* **33**, btx019 (2017).

41. Ziats, M. N. & Rennert, O. M. Identification of differentially expressed microRNAs across the developing human brain. *Mol. Psychiatry* **19**, 848–852 (2014).

42. Gong, J. *et al.* An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database (Oxford).* **2015**, bav029 (2015).

43. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

44. Agapite, J. *et al.* Alliance of Genome Resources Portal: Unified model organism research platform. *Nucleic Acids Res.* **48**, D650–D658 (2020).

45. Kiezun, A. *et al.* MiRviewer: A multispecies microRNA homologous viewer. *BMC Res. Notes* **5**, 92 (2012).

46. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

47. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6030966/. (Accessed: 25th July 2020)

48. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241–247 (2005).

49. Godnic, I. *et al.* Genome-Wide and Species-Wide In Silico Screening for Intragenic MicroRNAs in Human, Mouse and Chicken. *PLoS One* **8**, (2013).

50. Biasiolo, M. *et al.* Impact of Host Genes and Strand Selection on miRNA and miRNA* Expression. *PLoS One* **6**, e23854 (2011).

51. Chen, X. *et al.* MicroRNAs tend to synergistically control expression of genes encoding extensively-expressed proteins in humans. *PeerJ* **2017**, (2017).

52. Meunier, J. *et al.* Birth and expression evolution of mammalian microRNA genes. *Genome Res.* **23**, 34–45 (2013).

53. Wong, M. L. *et al.* The PHF21B gene is associated with major depression and modulates the stress response. *Mol. Psychiatry* **22**, 1015–1025 (2017).

54. Ikeda, M. *et al.* Genome-wide association study detected novel susceptibility genes for schizophrenia and shared trans-populations/diseases genetic effect. *Schizophr. Bull.* **45**, 824–834 (2019).

55. Druart, M. & Le Magueresse, C. Emerging Roles of Complement in Psychiatric Disorders. *Frontiers in Psychiatry* **10**, 573 (2019).

56. Westacott, L. J. *et al.* Manipulating immune pathways differentially impact emotionality: Dissociable effects of complement C3 and C3aR on learned fear and innate anxiety. *bioRxiv* 685537 (2019). doi:10.1101/685537

57. Ocklenburg, S. *et al.* Cholecystokinin A Receptor (CCKAR) Gene Variation Is

Associated with Language Lateralization. *PLoS One* **8**, e53643 (2013).

58. Hashimoto, R. *et al.* Common variants at 1p36 are associated with superior frontal gyrus volume. *Transl. Psychiatry* **4**, e472 (2014).

59. Underwood, S. L. *et al.* Association analysis of the chromosome 4p-located G protein-coupled receptor 78 (GPR78) gene in bipolar affective disorder and schizophrenia. *Mol. Psychiatry* **11**, 384–394 (2006).

60. Zhang, L. L. *et al.* GPR26-deficient mice display increased anxiety- and depression-like behaviors accompanied by reduced phosphorylated cyclic AMP responsive element-binding protein level in central amygdala. *Neuroscience* **196**, 203–214 (2011).

61. Song, J. *et al.* Genome-wide association study identifies SESTD1 as a novel risk gene for lithium-responsive bipolar disorder. *Mol. Psychiatry* **21**, 1290–1297 (2016).

62. Warrier, V. *et al.* Genome-wide meta-analysis of cognitive empathy: Heritability, and correlates with sex, neuropsychiatric conditions and cognition. *Mol. Psychiatry* **23**, 1402–1409 (2018).

63. Fabbri, C. & Serretti, A. Genetics of long-term treatment outcome in bipolar disorder. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **65**, 17–24 (2016).

64. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

65. Goes, F. S. *et al.* Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168**, 649–659 (2015).

66. Fung, S. J., Webster, M. J. & Weicker, C. S. Expression of VGluT1 and VGAT mRNAs in human dorsolateral prefrontal cortex during development and in schizophrenia. *Brain Res.* **1388**, 22–31 (2011).

67. Oni-Orisan, A., Kristiansen, L. V., Haroutunian, V., Meador-Woodruff, J. H. & McCullumsmith, R. E. Altered Vesicular Glutamate Transporter Expression in the Anterior Cingulate Cortex in Schizophrenia. *Biol. Psychiatry* **63**, 766–775 (2008).

68. Uezato, A., Meador-Woodruff, J. H. & McCullumsmith, R. E. Vesicular glutamate transporter mRNA expression in the medial temporal lobe in major depressive disorder, bipolar disorder, and schizophrenia. *Bipolar Disord.* **11**, 711–725 (2009).

69. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).

70. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & Van Der Sluis, S. Item-level

Supplementary Table 1: miRIAD (v. 2018) expression levels for significantly associated miRNAs

miRNAs	Expression Data Available	Expressed in Brain
miR-6734	Yes	Yes
miR-6735	Yes	Yes
miR-3135a	Yes	Yes
miR-4655	Yes	Yes
miR-7-1	Yes	Yes
miR-6506	Yes	Yes
miR-6079	No	_
miR-4271	No	_
miR-5193	No	_
miR-3666	No	_
miR-1273h	No	_
miR-6872	No	_

Availability of expression data of the highlighted miRNAs in brain.

Supplementary Table 2: miRmine (v. Jan2017) expression levels for significantly associated miRNAs

Mature miRNA ID	Precursor miRNA ID	MAX observed expression level	Tissue with the highest expression level	SRX375448 (Brain)
hsa-miR-1273h-5p	hsa-mir-1273h	13079.7	ERX358449 (Hair follicle)	0.6
hsa-miR-6735-5p	hsa-mir-6735	190.1	SRX513283 (Breast)	0.6
hsa-miR-7-1-3p	hsa-mir-7-1	974	SRX513286 (Breast)	59.1
hsa-miR-7-5p	hsa-mir-7-1	50782.4	SRX290620 (Pancreas)	60.4
hsa-miR-1273h-3p	hsa-mir-1273h	522.1	SRX513284 (Breast)	0
hsa-miR-3135a	hsa-mir-3135a	2.7	SRX386680 (Blood)	0
hsa-miR-3666	hsa-mir-3666	0	SRX375448 (Brain)	0
hsa-miR-4271	hsa-mir-4271	0	SRX375448 (Brain)	0
hsa-miR-4655-3p	hsa-mir-4655	0.9	SRX262193 (Plasma)	0
hsa-miR-4655-5p	hsa-mir-4655	1.1	SRX362362 (Plasma)	0
hsa-miR-5193	hsa-mir-5193	87	SRX513284 (Breast)	0
hsa-miR-6079	hsa-mir-6079	0	SRX375448 (Brain)	0
hsa-miR-6506-3p	hsa-mir-6506	0.1	SRX386681 (Blood)	0
hsa-miR-6506-5p	hsa-mir-6506	13079.7	ERX358449 (Hair follicle)	0
hsa-miR-6734-3p	hsa-mir-6734	1.6	SRX666579 (Blood)	0
hsa-miR-6734-5p	hsa-mir-6734	432.1	SRX349069 (Serum)	0
hsa-miR-6735-3p	hsa-mir-6735	10301.7	ERX358443 (Hair follicle)	0
hsa-miR-6872-3p	hsa-mir-6872	0.1	DRX011927 (Bladder)	0
hsa-miR-6872-5p	hsa-mir-6872	1.2	SRX262196 (Placenta)	0

MAX observed expression level: Expression counts for miRNA transcripts in transcripts per million. Tissue IDs corresponds to those in miRmine.

Supplementary Table 3: BrainSpan atlas expression levels for significantly associated miRNAs

miRNA	Max observed RNA-seq count	Structure with Max RNA-seq count	Structure
hsa-miR-3135a	7	H376_VI_50-CBC-L	Cerebellar cortex
hsa-miR-3666	2	H376_VI_50-S1C-L	somatosensory cortex
hsa-miR-4271	1	H376_VIII_51_S1C-R	somatosensory cortex
hsa-miR-4655-3p	1	H376_X_53-S1C-L	somatosensory cortex
hsa-miR-4655-5p	2	H376_X_50_V1C-L	(striate cortex, area
hsa-miR-5193	5	H376_VI_52_IPC-L	(ventral) parietal cortex
hsa-miR-7-1-3p	1769	H376_IX_52-CBC-L	Cerebellar cortex

Raw RNA-seq read-count data for microRNAs contained for the isolated small RNA samples from the indicated structures and specimens.

Supplementary Table 4. GTEx (v8) eQTL information for significantly associated miRNA variants in brain tissues

Variant	Chr	meta-analysis p	Host gene	eQTL for gene	Gene Chr	Brain Tissue	P-Value
rs839764	1	8.711E-08	*ELOVL1*	*TIE1*	*1*	Hippocampus	3.80E-06
				MED8	*1*	Spinal cord (cervical c-1)	2.20E-05
rs56319043	1	1.37E-11	*ST3GAL3*	*ARTN*	*1*	Cerebellar Hemisphere	1.30E-10
				RP11-7O11.3	*1*	Caudate (basal ganglia)	3.40E-05
rs3011216	1	3.53E-07	*ST3GAL3*	*ARTN*	*1*	Cerebellar Hemisphere	5.50E-14
						Cerebellum	5.50E-13
						Hippocampus	1.20E-05
				RP11-7O11.3	*1*	Caudate (basal ganglia)	1.50E-05
rs3011217	1	1.51E-08	*ST3GAL3*	*ARTN*	*1*	Cerebellum	1.80E-25
				ARTN		Cerebellar Hemisphere	2.60E-21
				ARTN		Hippocampus	6.20E-08
				ARTN		Putamen (basal ganglia)	2.30E-06

				ARTN		Spinal cord (cervical c-1)	3.70E-06
				ST3GAL3		Cortex	6.60E-06
				ARTN		Caudate (basal ganglia)	7.80E-06
				ARTN		Amygdala	2.50E-05
				ARTN		Anterior cingulate cortex (BA24)	3.50E-05
				ARTN		Cortex	4.00E-05
				ARTN		Nucleus accumbens (basal ganglia)	6.00E-05
rs11708763	3	5.46E-05	*KAT2B*	*KAT2B*	*3*	Putamen (basal ganglia)	3.00E-06
				KAT2B		Cerebellum	2.10E-05
rs1799844	3	9.03E-06	*UBA7*	*GMPPB*	*3*	Cerebellum	6.10E-34
				GMPPB		Cerebellar Hemisphere	1.60E-23
				GMPPB		Cortex	2.80E-19
				GMPPB		Spinal cord (cervical c-1)	2.70E-18
				GMPPB		Putamen (basal ganglia)	5.80E-18
				GMPPB		Nucleus accumbens (basal ganglia)	9.80E-18
				GMPPB		Hippocampus	1.00E-16
				GMPPB		Caudate (basal ganglia)	1.80E-12
				GMPPB		Hypothalamus	7.50E-12
				AMT		Cerebellum	6.80E-11
				GMPPB		Substantia nigra	7.00E-11
				AMT		Cortex	7.80E-11
				GMPPB		Anterior cingulate cortex (BA24)	9.00E-10
				GMPPB		Frontal Cortex (BA9)	9.10E-10
				GMPPB		Amygdala	6.80E-09
				RNF123		Cerebellum	7.30E-09
				AMT		Hippocampus	3.10E-08
				GPX1		Cortex	6.60E-08
				GPX1		Nucleus accumbens (basal ganglia)	9.90E-08
				AMT		Caudate (basal ganglia)	9.00E-07

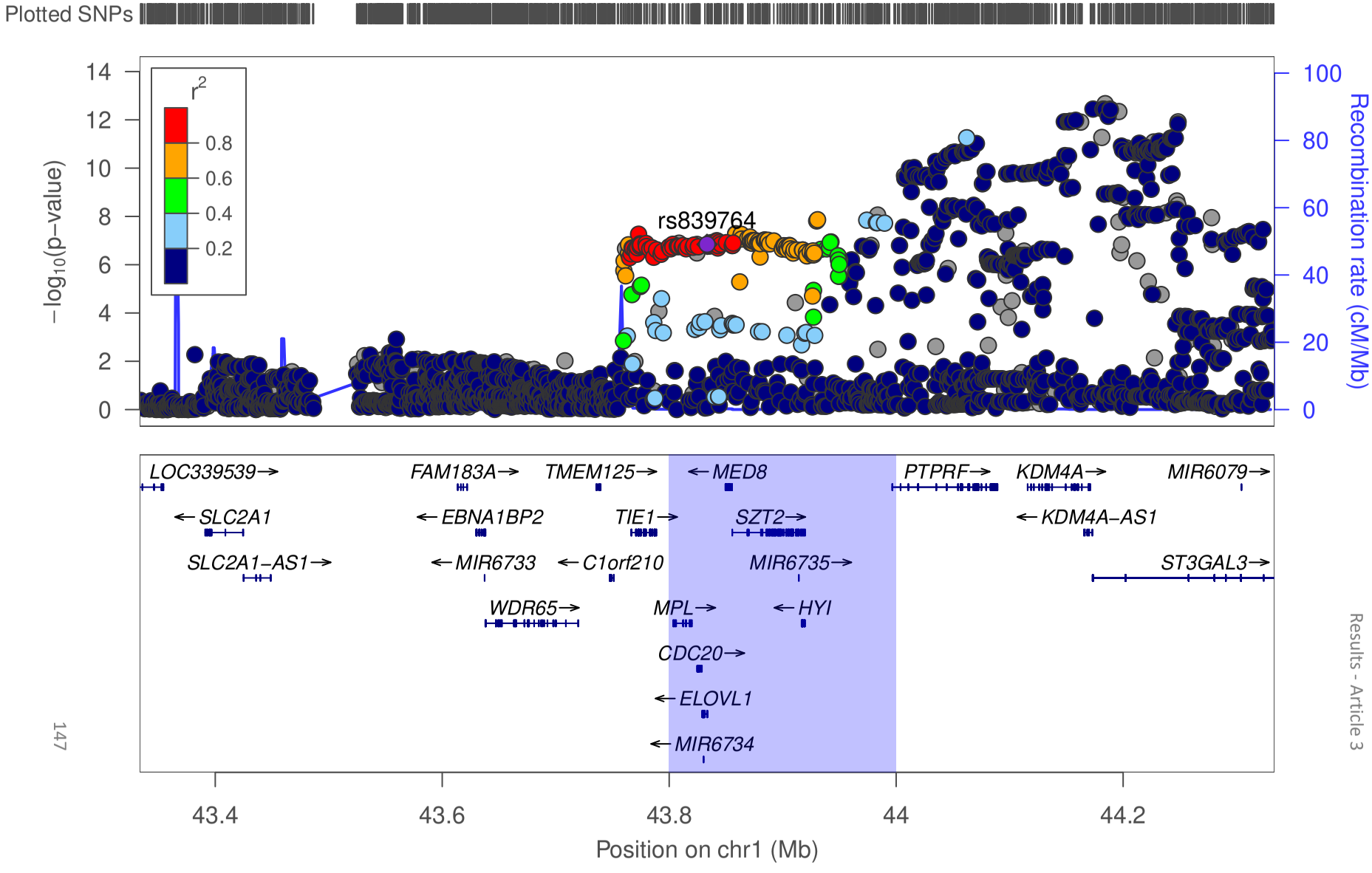
				GPX1		Caudate (basal ganglia)	1.10E-06
				GPX1		Frontal Cortex (BA9)	1.30E-06
				GPX1		Cerebellum	1.80E-06
				P4HTM		Cortex	3.20E-06
				AMT		Cerebellar Hemisphere	4.70E-06
				FAM212A		Cerebellum	6.20E-06
				AMT		Frontal Cortex (BA9)	8.20E-06
				AMT		Anterior cingulate cortex (BA24)	1.80E-05
				AMT		Substantia nigra	1.90E-05
				RNF123		Nucleus accumbens (basal ganglia)	4.40E-05
				RBM6		Cortex	4.60E-05
				HYAL3		Spinal cord (cervical c-1)	5.60E-05
				CCDC71		Amygdala	8.20E-05
				AMT		Putamen (basal ganglia)	9.60E-05
				RBM6		Cerebellar Hemisphere	1.10E-04
				RNF123		Cerebellar Hemisphere	1.40E-04
				RP11-694I15.7		Cerebellum	1.40E-04
				AMT		Nucleus accumbens (basal ganglia)	1.50E-04
				DALRD3		Cerebellum	1.70E-04
				RBM6		Frontal Cortex (BA9)	2.00E-04
				RBM6		Nucleus accumbens (basal ganglia)	2.30E-04
				P4HTM		Cerebellum	2.60E-04
				RNF123		Cortex	3.00E-04
rs10250550	7	4.07E-05	*MAD1L1*	*AC110781.3*	7	Nucleus accumbens (basal ganglia)	8.30E-10
				AC110781.3		Caudate (basal ganglia)	1.10E-07
				AC110781.3		Frontal Cortex (BA9)	1.10E-05
				AC110781.3		Hypothalamus	1.90E-05

rs296886	9	3.73E-05	*HNRNPK*	*GKAP1* *GKAP1* *RMI1* *RMI1* *RMI1*	*9*	Cerebellar Hemisphere Cerebellum Cortex Spinal cord (cervical c-1) Cerebellum	6.50E-08 1.70E-07 4.90E-06 1.80E-05 2.00E-05
rs296894	9	5.28E-05	*HNRNPK*	*GKAP1* *GKAP1* *RMI1*	*9*	Cerebellar Hemisphere Cerebellum Cortex	3.40E-07 3.80E-07 0.000062
rs143942298	16	1.24E-05	*KIAA0430*	*RP11-680G24.6* *RP11-680G24.4* *RP11-680G24.4* *RP11-680G24.4*	*16*	Cerebellum Cerebellum Putamen (basal ganglia) Nucleus accumbens (basal ganglia)	1.70E-08 1.20E-06 1.20E-05 8.70E-05
rs2251802	1	1.68E-07	*SZT2*	*None*			
rs37453	1	1.10E-07	*ST3GAL3*	*None*			
chr3:50310286	3	5.59E-05	*SEMA3B*	*None*			
rs58936320	3	2.53E-05	*C3orf62*	*None*			
rs2045292	7	6.06E-07	*FOXP2*	*None*			
rs7782412	7	4.07E-05	*FOXP2*	*None*			
rs7799269	7	4.37E-06	*FOXP2*	*None*			
rs76100764	9	3.28E-05	*HNRNPK*	*None*			
rs605921	16	4.77E-05	*PRKCB*	*None*			

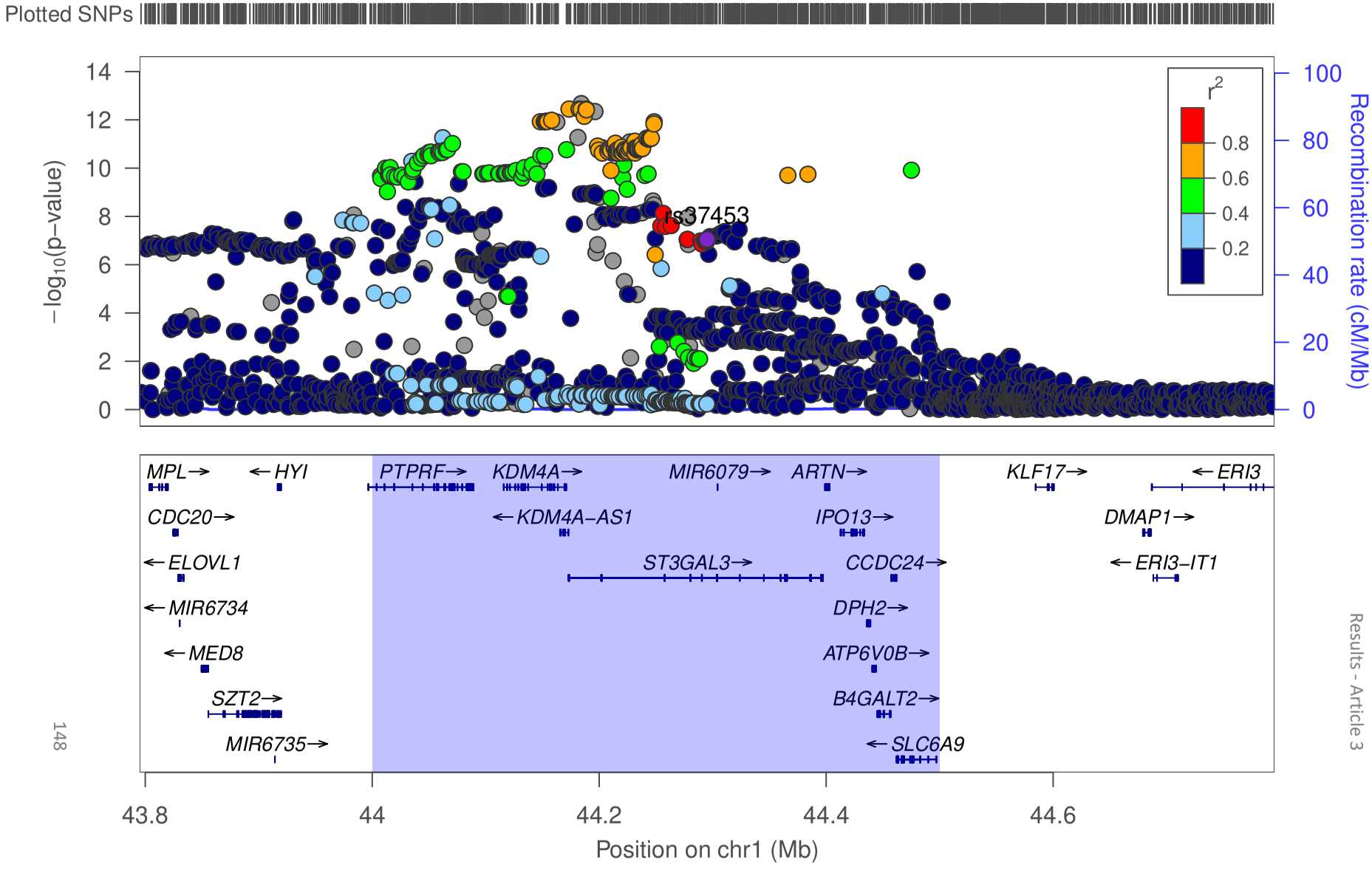
eQTLs for the significantly associated variants. Host gene: Host gene for the miRNA highlighted by the variant. P-value and normalised effect size provided by GTEx.

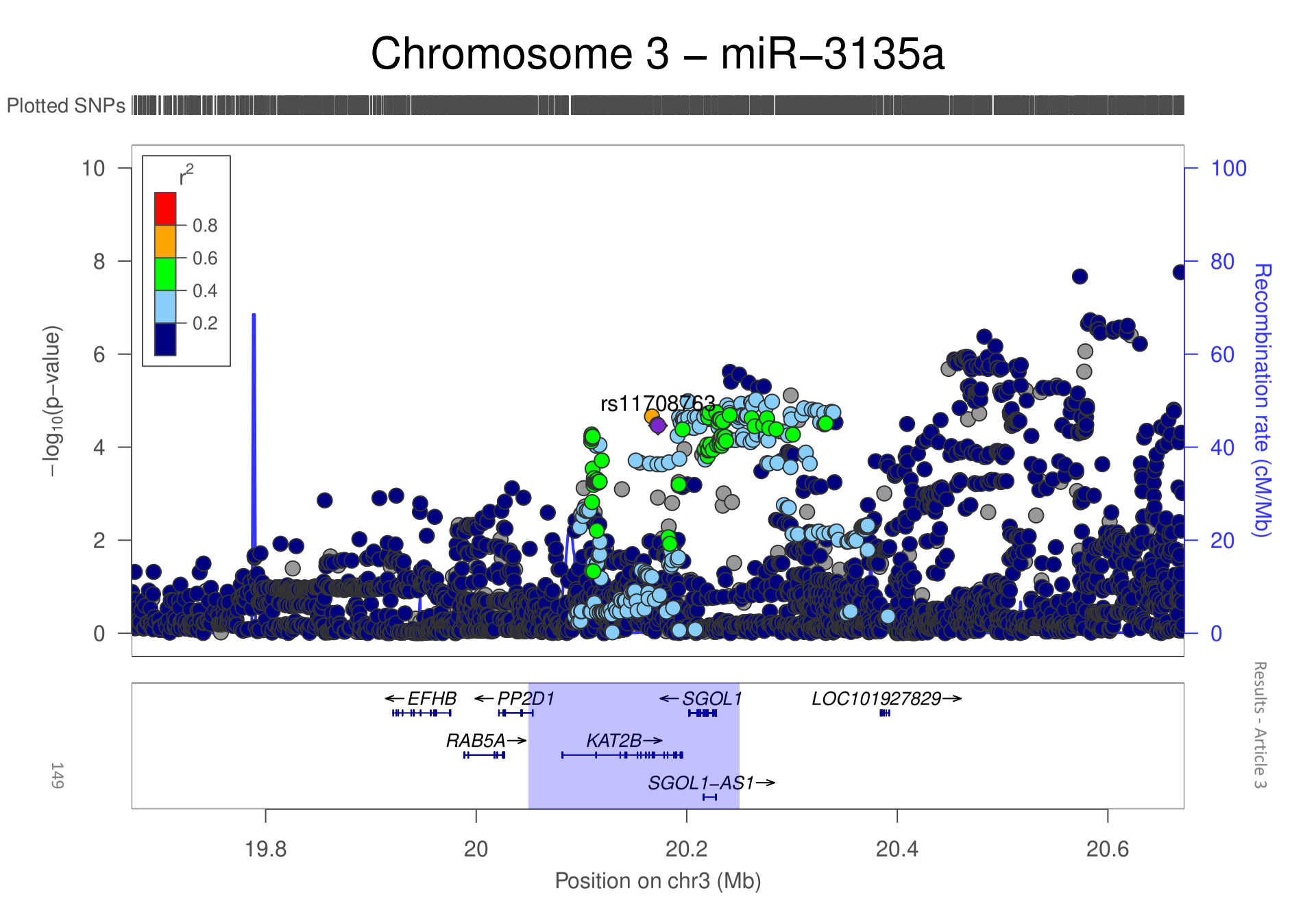
Supplementary Fig 1. Regional association plots for the highlighted miRNAs

Chromosome 1 – miR–6734 and miR–6735

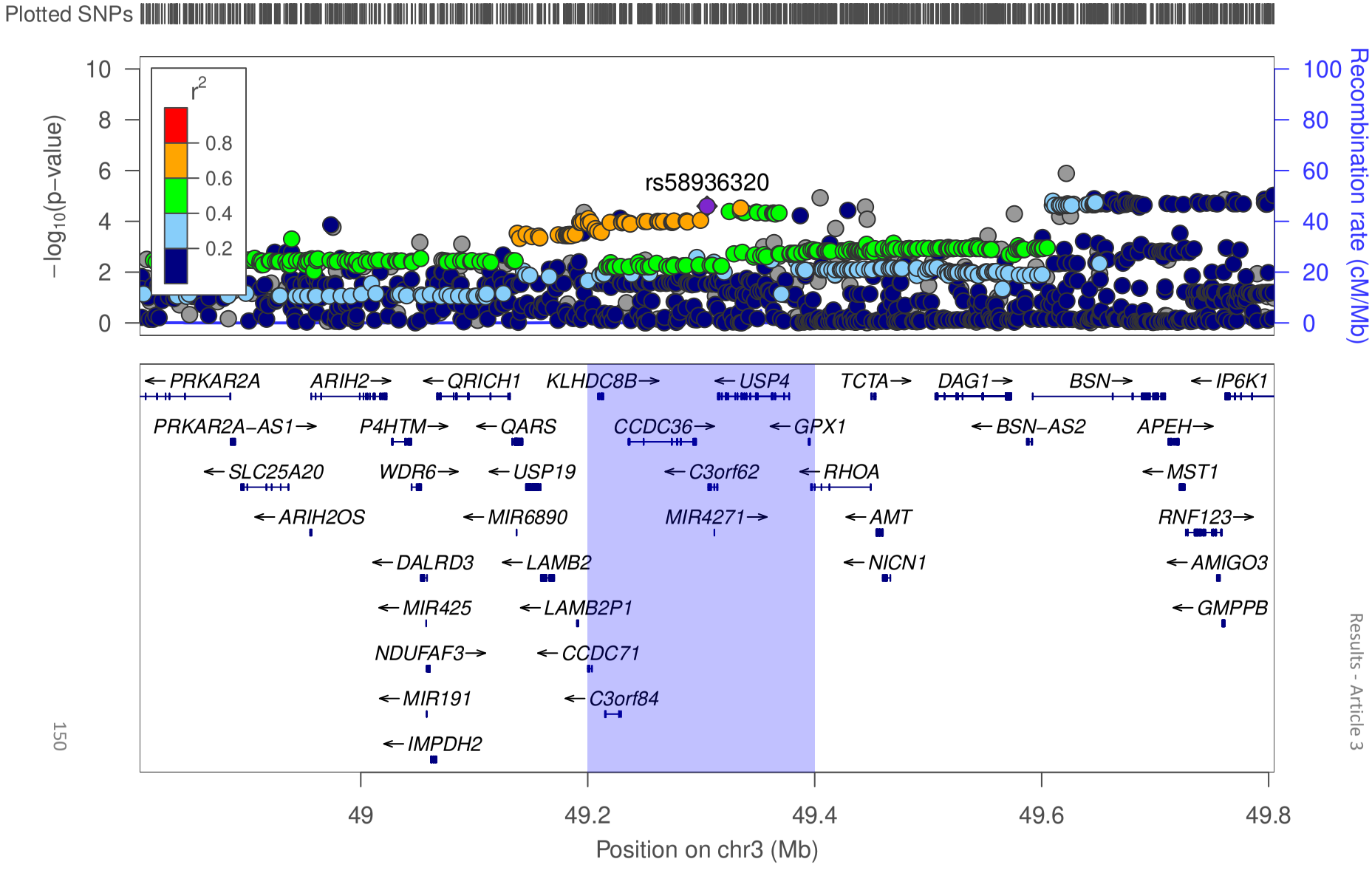


Chromosome 1 – miR–6079

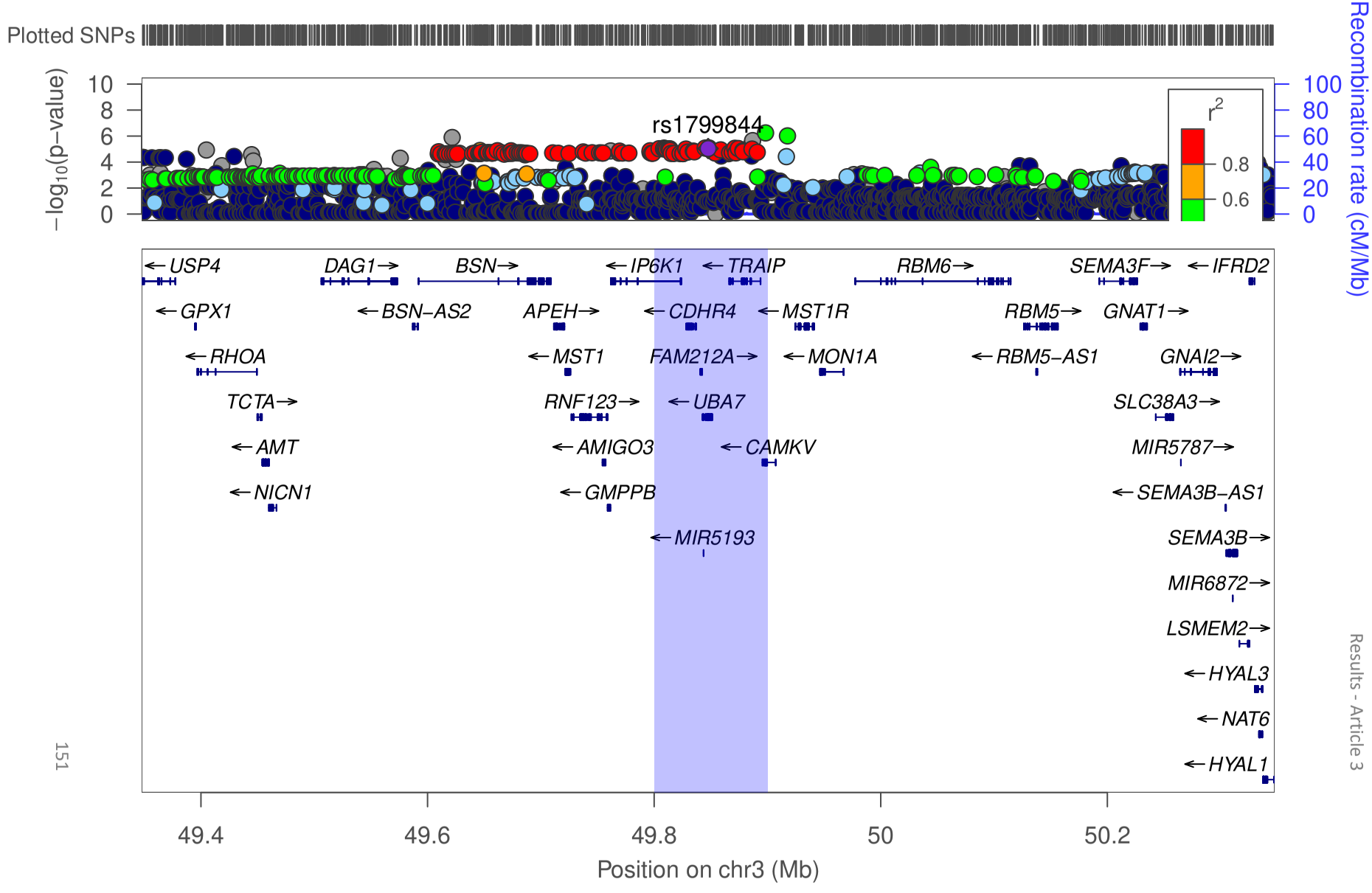



Chromosome 3 – miR–3135a
Plotted SNPs
r2
0.8
0.6
0.4
0.2
rs11708763
−log10(p−value)
Recombination rate (cM/Mb)
EFHB
PP2D1
SGOL1
LOC101927829
RAB5A
KAT2B
SGOL1–AS1
Position on chr3 (Mb)
19.8
20
20.2
20.4
20.6

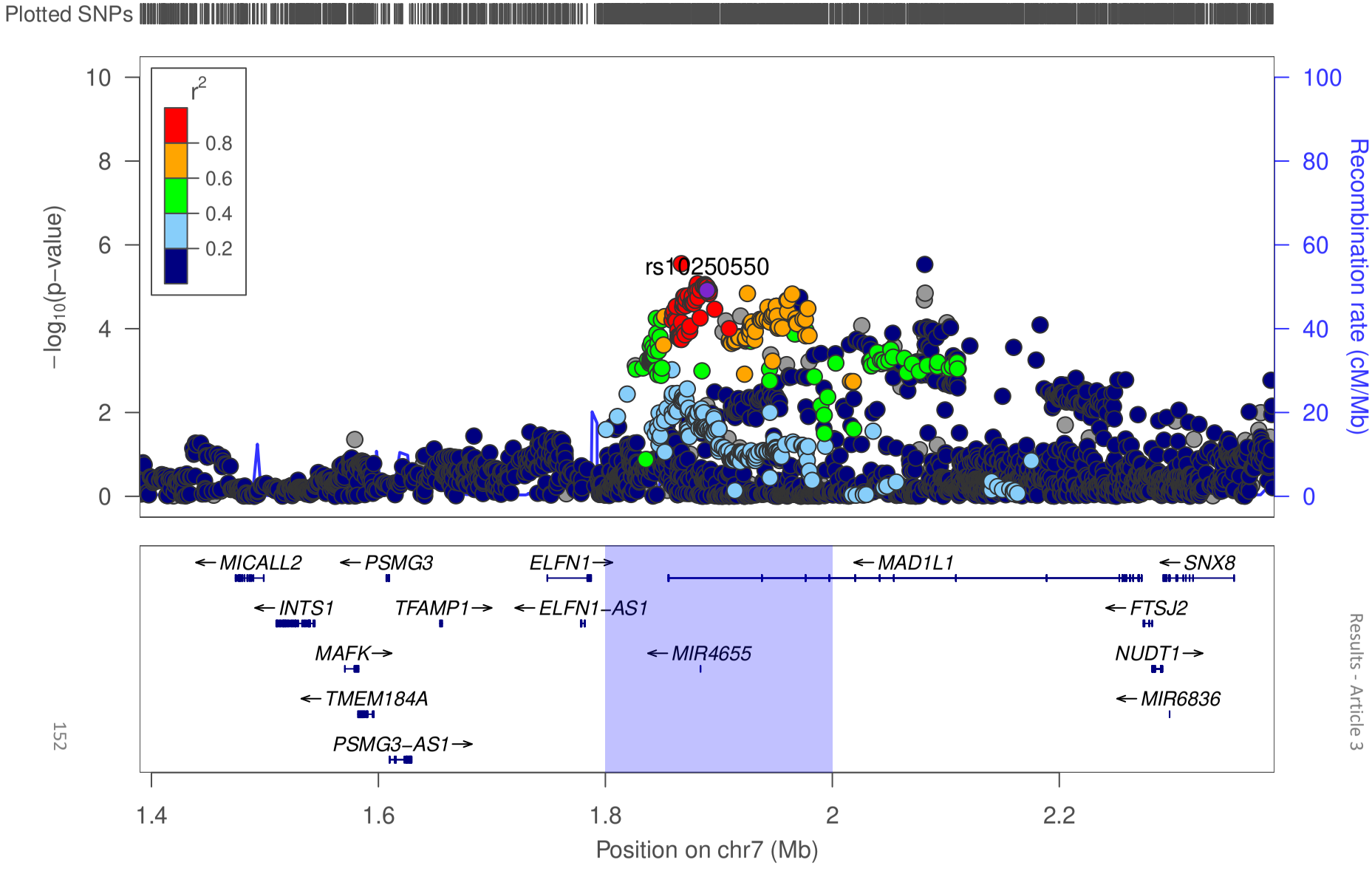
Chromosome 3 – miR–4271

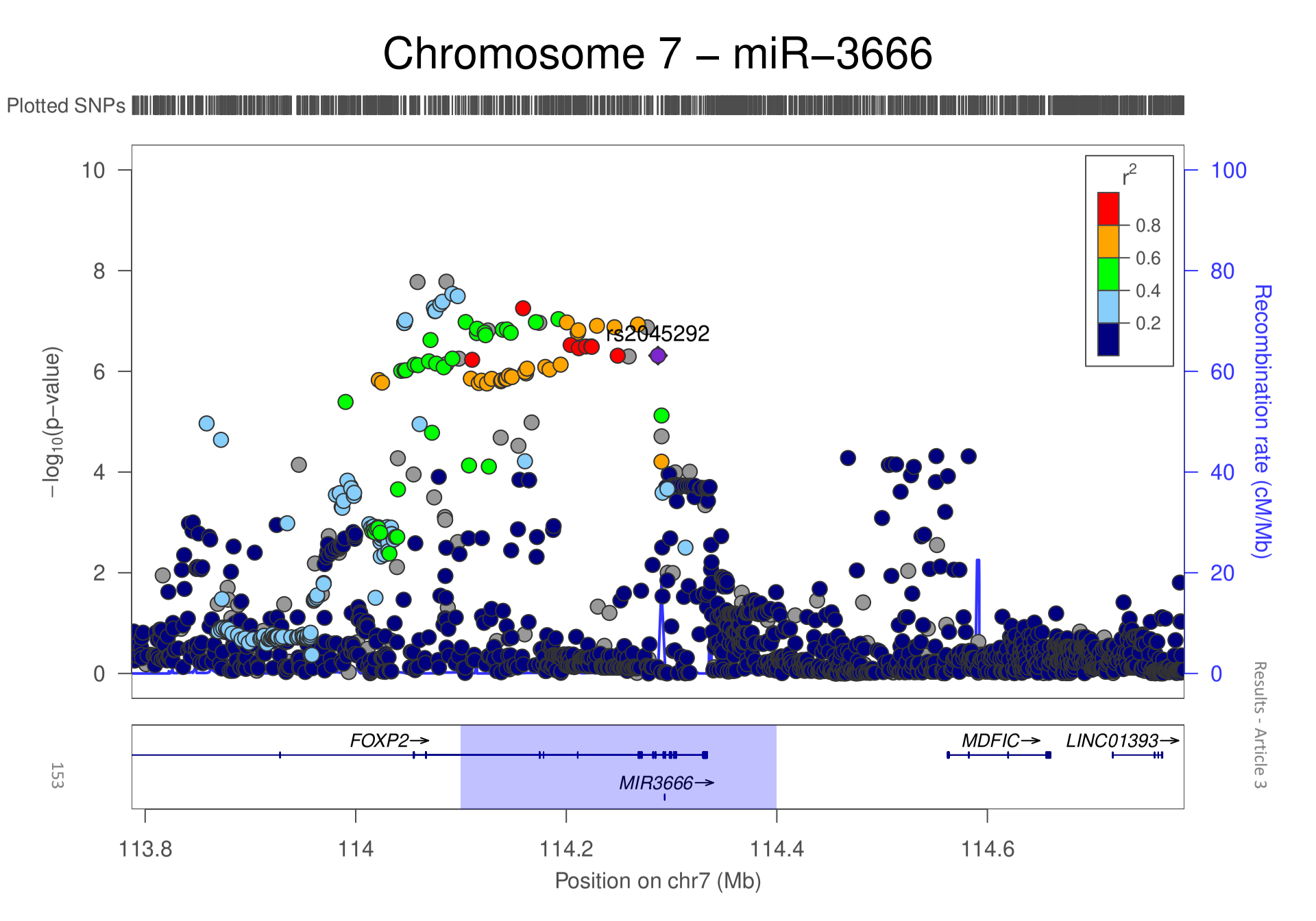


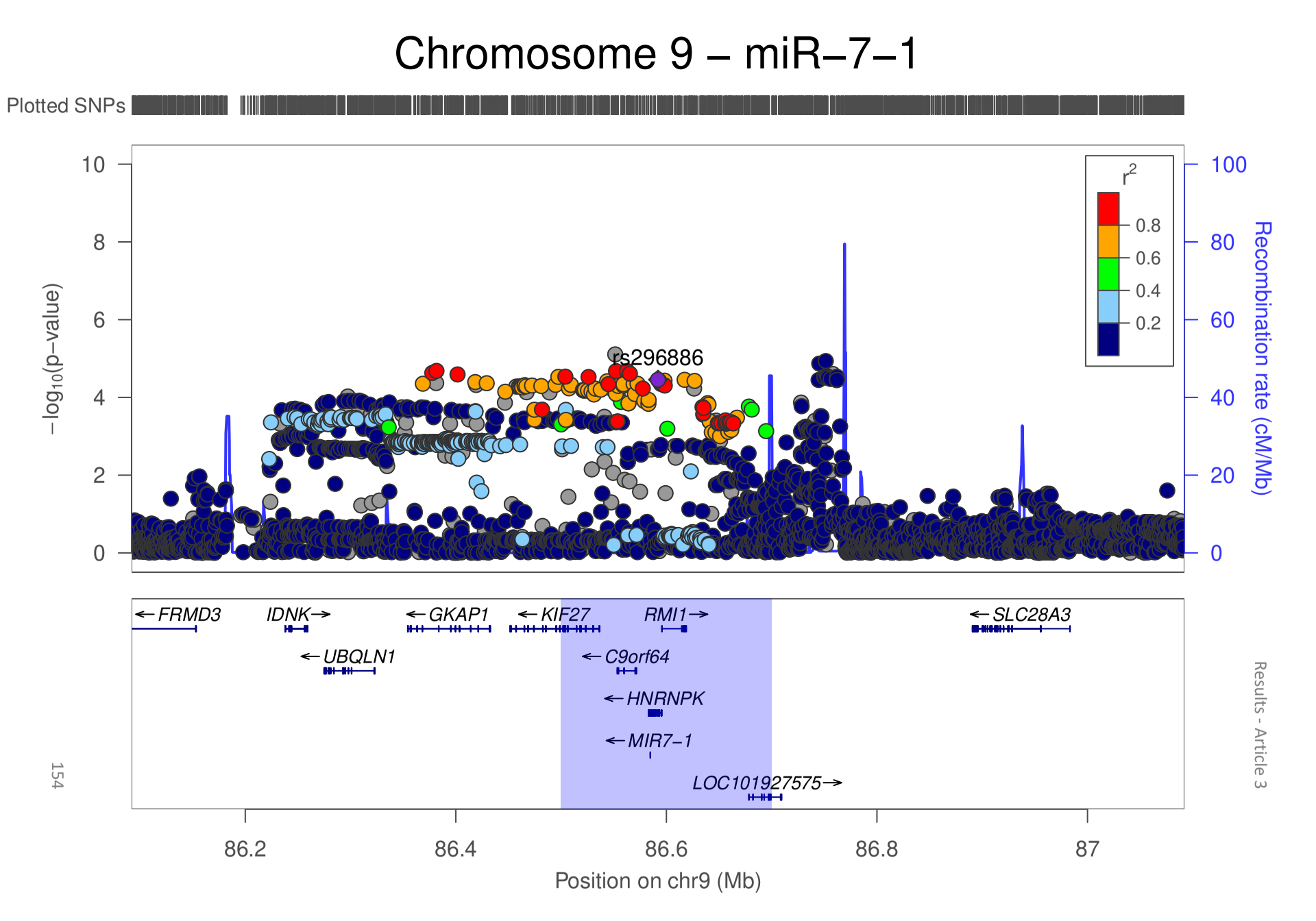
Chromosome 3 – miR–5193



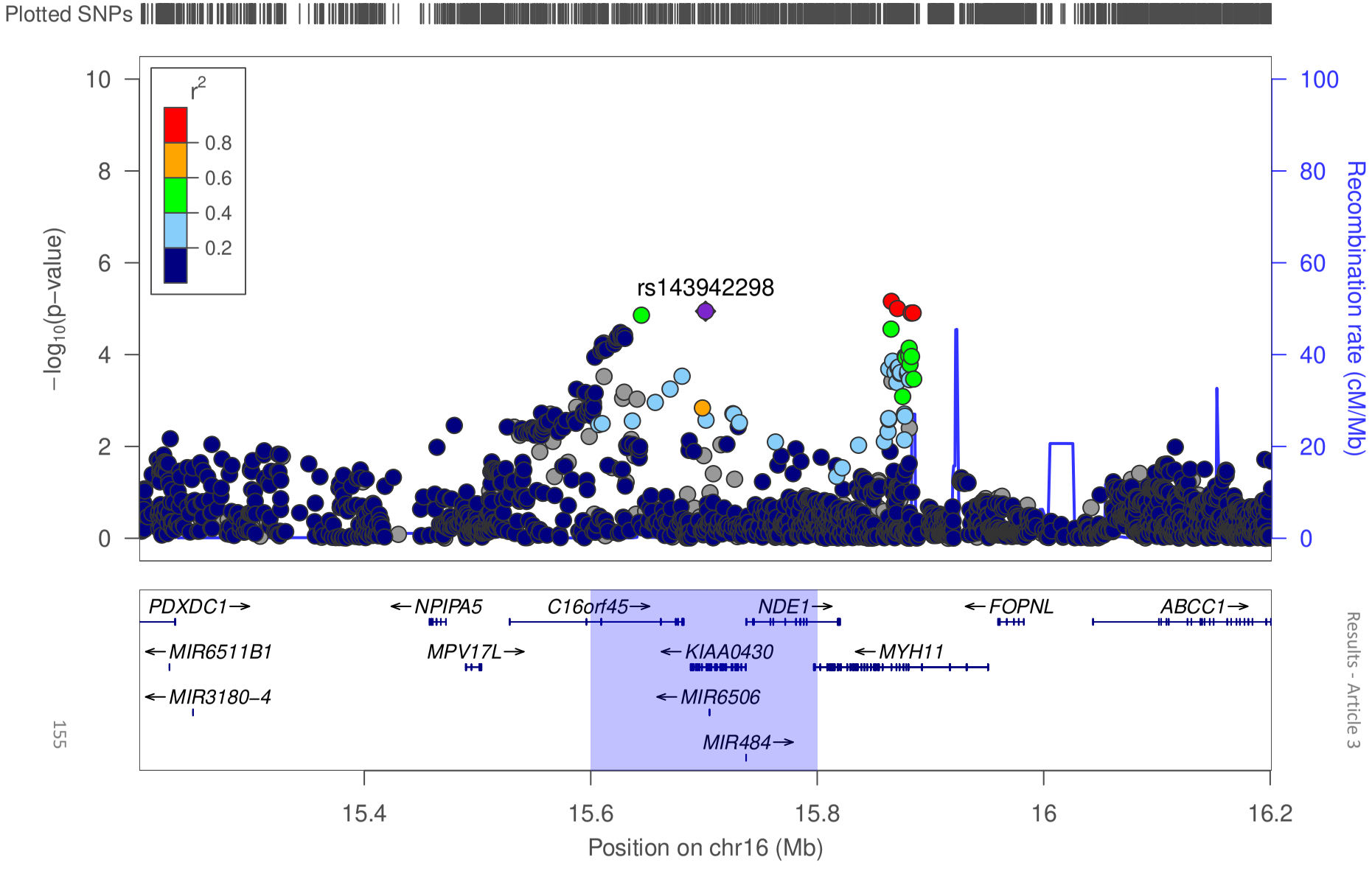
Chromosome 7 – miR–4655



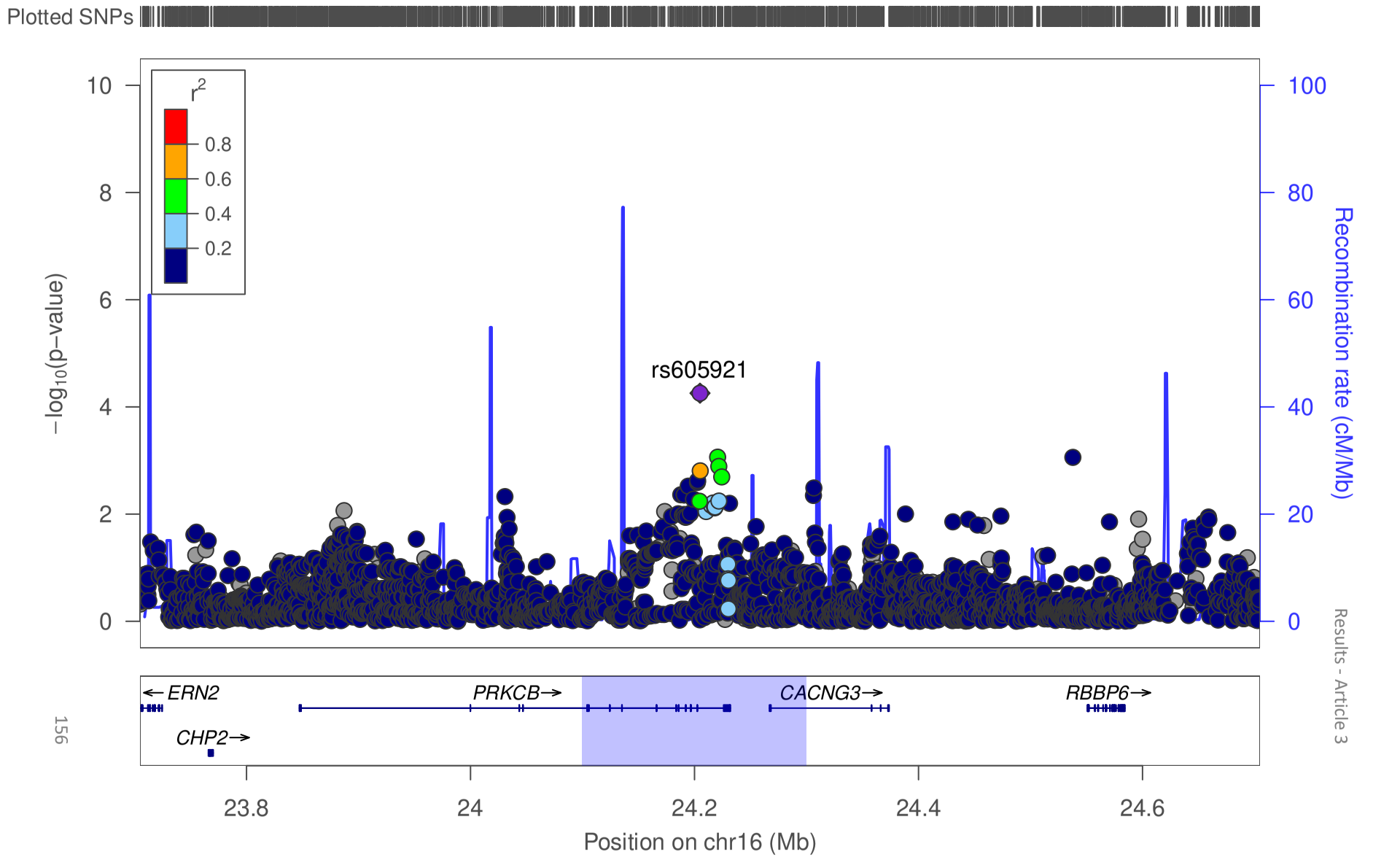

Chromosome 7 – miR–3666
Plotted SNPs
rs2045292
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FOXP2
MIR3666
MDFIC
LINC01393
113.8
114
114.2
114.4
114.6
Position on chr7 (Mb)


Chromosome 9 – miR–7–1
Plotted SNPs
rs296886
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FRMD3
IDNK
GKAP1
KIF27
RMI1
SLC28A3
UBQLN1
C9orf64
HNRNPK
MIR7−1
LOC101927575
Position on chr9 (Mb)
154
Results - Article 3

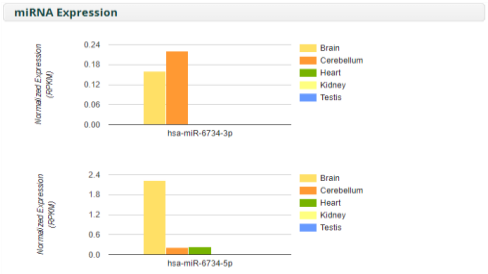
Chromosome 16 – miR–6506



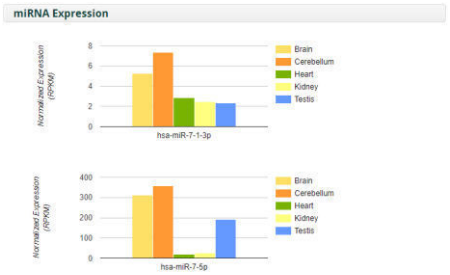
Chromosome 16 – miR–1273h



Supplementary Figure 2. miRIAD (v. 2018) expression levels for significantly associated miRNAs

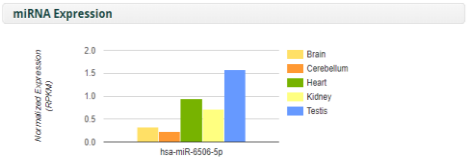


miRNA	miRNA	tissue	expression
hsa-mir-6734	hsa-miR-6734-3p	Brain	0.16
hsa-mir-6734	hsa-miR-6734-3p	Cerebellum	0.22
hsa-mir-6734	hsa-miR-6734-3p	Heart	0.0
hsa-mir-6734	hsa-miR-6734-3p	Kidney	0.0
hsa-mir-6734	hsa-miR-6734-3p	Testis	0.0
hsa-mir-6734	hsa-miR-6734-5p	Brain	2.22
hsa-mir-6734	hsa-miR-6734-5p	Cerebellum	0.22
hsa-mir-6734	hsa-miR-6734-5p	Heart	0.24
hsa-mir-6734	hsa-miR-6734-5p	Kidney	0.0
hsa-mir-6734	hsa-miR-6734-5p	Testis	0.0

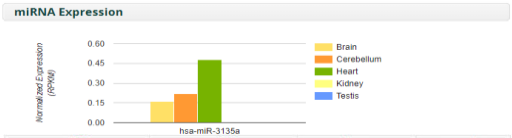


miRNA	miRNA	tissue	expression
hsa-mir-7-1	hsa-miR-7-1-3p	Brain	5.24
hsa-mir-7-1	hsa-miR-7-1-3p	Cerebellum	7.35
hsa-mir-7-1	hsa-miR-7-1-3p	Heart	2.85
hsa-mir-7-1	hsa-miR-7-1-3p	Kidney	2.5
hsa-mir-7-1	hsa-miR-7-1-3p	Testis	2.36
hsa-mir-7-1	hsa-miR-7-5p	Brain	310.79
hsa-mir-7-1	hsa-miR-7-5p	Cerebellum	359.34
hsa-mir-7-1	hsa-miR-7-5p	Heart	20.46
hsa-mir-7-1	hsa-miR-7-5p	Kidney	25.52
hsa-mir-7-1	hsa-miR-7-5p	Testis	191.93

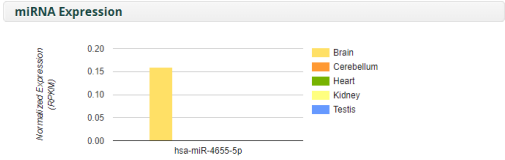
miRNA	miRNA	tissue	expression
hsa-mir-6735	hsa-miR-6735-3p	Brain	0.16
hsa-mir-6735	hsa-miR-6735-3p	Cerebellum	0.22
hsa-mir-6735	hsa-miR-6735-3p	Heart	0.24
hsa-mir-6735	hsa-miR-6735-3p	Kidney	0.0
hsa-mir-6735	hsa-miR-6735-3p	Testis	0.0
hsa-mir-6735	hsa-miR-6735-5p	Brain	1.11
hsa-mir-6735	hsa-miR-6735-5p	Cerebellum	1.11
hsa-mir-6735	hsa-miR-6735-5p	Heart	0.0
hsa-mir-6735	hsa-miR-6735-5p	Kidney	0.0
hsa-mir-6735	hsa-miR-6735-5p	Testis	0.0



miRNA	miRNA	tissue	expression
hsa-mir-6506	hsa-miR-6506-5p	Brain	0.32
hsa-mir-6506	hsa-miR-6506-5p	Cerebellum	0.22
hsa-mir-6506	hsa-miR-6506-5p	Heart	0.95
hsa-mir-6506	hsa-miR-6506-5p	Kidney	0.71
hsa-mir-6506	hsa-miR-6506-5p	Testis	1.58



miRNA	miRNA	tissue	expression
hsa-mir-3135a	hsa-miR-3135a	Brain	0.16
hsa-mir-3135a	hsa-miR-3135a	Cerebellum	0.22
hsa-mir-3135a	hsa-miR-3135a	Heart	0.48
hsa-mir-3135a	hsa-miR-3135a	Kidney	0.0
hsa-mir-3135a	hsa-miR-3135a	Testis	0.0



miRNA	miRNA	tissue	expression
hsa-mir-4655	hsa-miR-4655-5p	Brain	0.16
hsa-mir-4655	hsa-miR-4655-5p	Cerebellum	0.0
hsa-mir-4655	hsa-miR-4655-5p	Heart	0.0
hsa-mir-4655	hsa-miR-4655-5p	Kidney	0.0
hsa-mir-4655	hsa-miR-4655-5p	Testis	0.0

DISCUSSION

In this work, we studied:

(i) ADHD from an epigenetic point of view by targeting the systems of miRNAs and ASM applying hypothesis free and association studies respectively

(ii) Genetic basis of cocaine dependence using genome-wide association meta-analysis and shared genetics with comorbid conditions.

The selected methodologies have been discussed in detail.

CHAPTER 1. QUESTIONS UNDERLYING THE GENETIC ARCHITECTURE OF PSYCHIATRIC DISORDERS

Psychiatric disorders are complex multifactorial genetic disorders that exhibit no clear-cut pattern of Mendelian inheritance. The genetic architecture underlying the pathogenic mechanisms of psychiatric disorders like ADHD and cocaine dependence remains still elusive, although research performed in recent years starts to shed some light on it. For instance, some open questions are: Can we quantify the number of susceptibility and protective genetic variants? What are their frequencies and effect sizes? What about unaccounted heritability (h^2)? How do the variants exert their effects? How do these variants interact with each other and with the environmental risk factors? How to construe the genes and biological pathways disturbed by these variants? Why do comorbidities appear? Are there shared genetic bases underlying comorbid disorders? Does epigenetics play a significant role? Can we identify biomarkers for psychiatric disorders[141]? We aimed to contribute to these issues in the context of ADHD and cocaine dependence.

1.1 Genetic models for psychiatric disorders

Much of our fundamental knowledge in computational psychiatric genetics has developed from the case studies on schizophrenia which has accumulated huge amounts of data for both common and rare variation. Lack of appearance of single causal genes in most individuals affected by psychiatric phenotypes gave rise to the theory of polygenicity. This model was originally investigated in schizophrenia[142], but it is also extensible to all psychiatric phenotypes studied so far. As per polygenic model, multiple genetic and non-genetic effects act in combination to contribute to disease susceptibility, and they appear to be normally distributed. All individuals in a population are assumed to carry genetic risk variants and are exposed to non-genetic risk factors; however, it is when the cumulative load exceeds a burden of risk threshold that the symptoms of disease develop.

To break down the questions about the number, frequency, and effect sizes of individual causal variants, and the additive or non-additive action of the causal loci, we debate on the following models for psychiatric disorders:

1) Common disease/common variant (CDCV) model – The phenotype is due to the cumulative impact of hundreds or thousands of common genetic variants where each variant exerts an individual small effect. Here, a drawback is that the functional consequences of the common

variants in protein-coding regions can be imprecise or difficult to detect; or much of the highlighted variation is located within non-protein coding genomic regions, for which prior knowledge is often limited.

2) Common disease/rare variant (CDRV) model – The phenotype appears as a consequence of the effect of various rare genetic variants of strong effect. The causal variants should cluster in a set of vital neurodevelopmental and/or neurofunctional genes[143].

Although not so frequently seen, there are also some few cases in psychiatry where a single genetic variant is the cause for the disorder. These rare penetrant and damaging mutations have been reported for example in ID, ASDs and in schizophrenia, sometimes in single genes and others in the form of CNVs that may span more than one gene[144–148]. Besides, *de novo* mutations in single genes occur in rare number of affected cases[145]. These *de novo* mutations are the genetic variants present in the DNA of a child but not of parents. Their contribution to the risk of disease in the form of effect sizes can be both small and large. WES reveals that *de novo* mutations can have an important role in Mendelian diseases and in also some complex conditions like autism and mental retardation. Increased burden of *de novo* mutations has been identified in patients, although in general it is uncertain which specific *de novo* mutations are causal. As de novo mutations are not shared among members of a family, they are unlikely to contribute to the heritability.

Our work on both ADHD and cocaine dependence has focused on the common variant model in view of several arguments favouring this paradigm: 1) Most genetic variants are common; 2) in many cases, causal common variants associated with a continuous endophenotype have been associated with disease, and in some cases, these have been confirmed by in vitro biochemical assays for structural and regulatory effects[149]; and 3) expression quantitative trait loci (eQTL) analyses have shown that gene expression and splicing are heavily influenced by common variants, perhaps for the majority of transcripts[149]. We therefore studied (i) common variants in miRNA genes for their contribution to ADHD, (ii) common variants that display ASM in ADHD and (iii) common variants contributing to risk to cocaine dependence.

1.2 Heritability and missing heritability

"Unaccounted h^2" is the unexplained difference between SNP-based heritability (h^2_{SNP}) estimates and twin-based or pedigree-based heritability estimates (twin-h^2 or pedigree-h^2). SNP-based heritability estimated from genome-wide SNP data rely largely on the sample size

of the GWAS. However, even with sufficiently sized genomic studies (e.g. in schizophrenia GWAS), SNP-h^2 usually remains half of the twin-h^2 at most [141]. However, recently the polygenic analyses, have been useful in delineating the "hidden heritability", i.e., the increase from h^2_{GWS} to h^2_{SNP}. It is presumed that with sufficiently large sample sizes, h^2_{GWS} should equal h^2_{SNP} (Figure 4). The heritability estimates may differ between populations, across ages and when non-genetic factors are counted in the analysis[150].

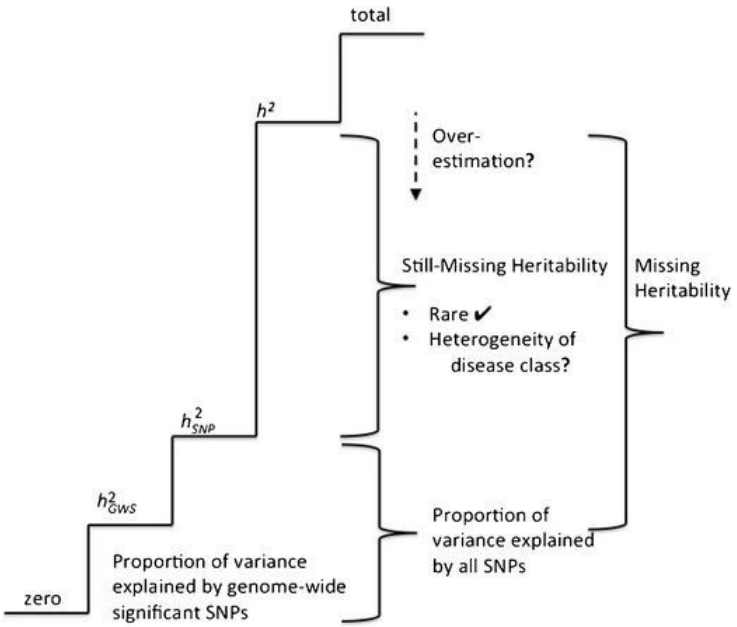


Figure 4. A hierarchy of heritability. Adapted from Wray *et al.*, 2014.

1.2.1 Overestimation of heritability from family studies

Classical heritability estimates have been achieved using full siblings and twins in families, who also share non-additive gene combinations and a common environment. These are the confounders that can be difficult to adjust for and thus the missing heritability may also reflect an overestimation of h^2. This difference between estimates of h^2 from family data and the "true" h^2 is termed "phantom heritability".

There is a dearth of gene-environment interaction (GxE) studies in family and twin studies which could consistently record both interaction and/or correlation between G and E. This obscures the impact the environmental risk factors on disease risk, especially in the already susceptible individuals because of which the SNP effect sizes in cases may not be accurately calculated.

1.2.2 Disease heterogeneity

As the heritability estimated from large population samples is lower than that estimated from twin studies, the presence of greater diagnostic heterogeneity in larger cohorts is hinted at. It further leads us to speculate if the larger cohorts may be more representative of the cases currently brought together for analysis in genetic studies. Heterogeneity in psychiatric disorders often corroborates the subtypes, which may be independent, correlated or overlapping. It is correspondingly arguable that a currently recognized disease category may turn out to be a diagnostic aggregation of subtypes. Acknowledging the phenotype heterogeneity in psychiatric disorders challenges our existing estimates for the proportions of phenotype variance due to genetic factors and may cover a part of the missing heritability.

1.2.3 Variants not tagged by common SNPs

Missing heritability can be due to the presence of other genomic risk variants not well tagged by common SNPs on the SNP chips used to genotype cases and controls. This includes CNVs or other rare variants detected through WES/WGS studies. Rare variants require relatively larger effect sizes in order to be detected. However, as they have lower frequencies, their contribution towards increasing the risk the entire set of population is small. Hence, a massive number of rare variants will be needed to lower the percentage of missing heritability.

1.2.4 Imputation panels

The significance of small structural variants is currently underrepresented in the genomic context; however, they may as well be important, e.g. tandem repeat polymorphisms with an impact on biological functions. The SNP-based 1000 Genomes reference panel used for imputation may not finely represent the small structural variation. Estimation of h^2_{SNP} based on haplotypes constructed from SNPs and not the SNPs alone may provide a way to tag uncommon structural variants missing from the imputation reference panels, although this approach can be highly sensitive to genotyping errors.

Heritable variation can be present within genomic features outside the coding sequence in the form of epigenetic modification. This includes promoter methylation, histone tail modifications and altered expression of non-coding RNAs which mediate gene regulation in normal development. Advanced technologies like methylation profiling and miRNA-seq can assist the quest for decoding the non-coding heritable variation. For instance, abundant QTLs for DNA CpG methylation across the genome have been reported for brain tissues.

In our work we computed the SNP heritability for cocaine dependence using GWAS data and these methodologies: LD Score (LDSC) regression analysis and the Genome-based REstricted Maximum Likelihood analysis implemented in the tool Genome-wide Complex Trait Analysis (GCTA-GREML). In both analyses, a population prevalence for cocaine dependence of 1.1% was considered[151]. To inspect if some functional categories of the genome contribute disproportionately to the heritability of cocaine dependence, we examined partitioned heritability for cocaine dependence (a polygenic condition with high SNP heritability) using LDSC based on 24 functional overlapping annotations[152]. Enrichment in the heritability of a functional category was defined as the proportion of SNP heritability explained divided by the proportion of SNPs in that category, and the issue of multiple testing was addressed using the Bonferroni correction. Our LDSC estimated a SNP heritability in liability scale of $h^2_{SNP} = 0.30$ and $h^2_{SNP} = 0.27$ for GCTA-GREML. Studies with comparable sample sizes have returned similar h^2_{SNP} for cocaine dependence, alcohol dependence, ADHD and schizophrenia (h^2_{SNP}=0.25-0.33)[153–155]. Increasing sample size can sometimes lower the SNP-heritability estimates e.g. in alcohol dependence (h^2_{SNP}= 0.09)[156] and MDD ($h^2_{SNP} = 0.09$)[157], but h^2_{SNP} remained the same in schizophrenia and ADHD[1,158]. This calls for a larger number of samples to confirm our results. The partitioned heritability analysis yielded a significant enrichment in the heritability by SNPs located in intronic regions, and a nominal result for conserved genomic regions. These results are not uncommon in the analyses of complex disorders, where the disease-associated variation is more often seen out of exonic regions than in Mendelian disorders.

1.3 Psychiatric epigenetics and epigenomics

Epigenetics encompasses regulation of DNA sequences without variation in their actual base composition, and the epigenetic marks can be both stable and plastic. The operating molecular epigenetic mechanisms are multifaceted, much more dynamic than the genetic code, and therefore they can be greatly intertwined to decode. Epigenetic dysregulations in CNS are associated with both monogenic and polygenic neuropsychiatric illnesses such as Fragile X, Rett syndrome, MDD, ASD and schizophrenia. Many characteristics of psychiatric diseases can be explained by epigenetics. For example: Epigenetic downregulation of genes is thought to underlie the GABAergic neuronal dysfunction observed in schizophrenia[159]. It is also noticeable that histone modifications are present during the development of schizophrenia although the precise mechanism of action is ambiguous.

The epigenetic model of complex diseases is analogous to the polygenic model. Small epigenetic mis-regulations in an individual may add up over time periods until a certain threshold is met, after which the disease manifests. The pre-epimutation disturbances that occur during the maturation of the germline may initially be tolerated, however over time, accumulation of these may increase the risk of attaining a disease. Due to the reversible nature of epimutations, their severity may well fluctuate at time points, and can show "remission" and "relapse". The age of disease onset may be subjected to the effects of tissue differentiation, stochastic features, hormones, and external environmental factors like nutrition, infections, medications or addictions[159]. Epigenetic changes often produce gene expression changes, which may in turn affect neural circuitry and eventually behavior.

Epigenetic adaptions are inquisitive as they appear distinct between developmental stages and adult life (Figure 5). It is speculated that developmental events may mark broader impressions on epigenetic states and neural functions than similar events at later stages in life. Thus, looking into the epigenetic mechanisms of gene regulation can help in understanding how prenatal or initial lifetime exposures to stimuli (stress, trauma, toxins, viral infections, nutritional deficits) shape neuropsychiatric effects for lifetime[160]. It is now feasible to perform true genome-wide assessments of epigenetic marks, such as DNA methylation (methylomes) or chromatin modifications (chromatinomes), to interrogate into the proposed questions.

In this Thesis we targeted two epigenetic mechanisms, DNA methylation and miRNAs, for their involvement in ADHD (detailed in Chapter 4).

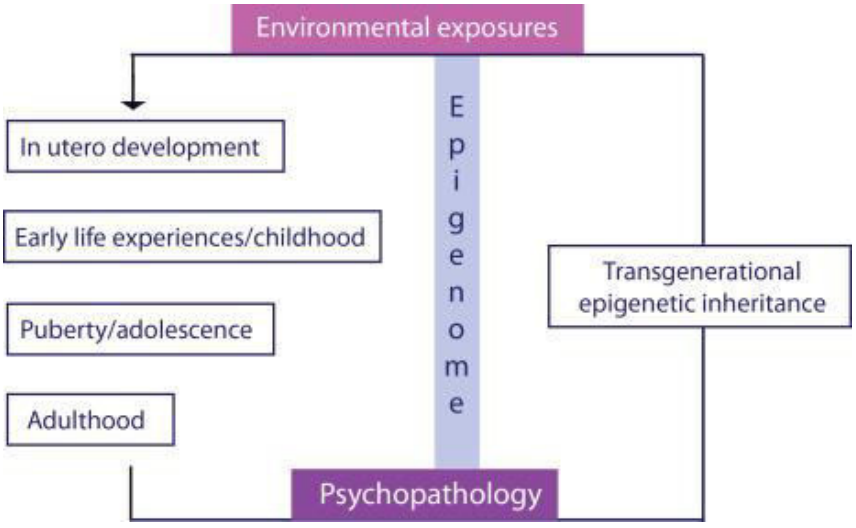


Figure 5. Possible contribution of epigenetic mechanisms to psychopathology.
Adapted from Chapter 19 - Epigenetics of Psychiatric Disorders, 2016 by M.Kundakovic.

1.4 From association to function - Mapping variants to genes and/or pathways

As psychiatric GWASs have identified multiple genetic signals associated with disorders, it has become clear that nominating the associated SNPs does not immediately lead to the identification of causal variants or to those genes and proteins which alteration is responsible for the phenotype. To be able to infer the real picture of deregulated genes and pathways to pin potential drug targets, it is necessary to map functional connections of the GWS and sub-threshold GWS variants from GWASs. In this Thesis we used and integrated the following established methods to aid translation of the detected associations into functions (detailed in Chapter 4).

1.4.1 eQTL mapping

An approach to distinguish molecular mechanisms underlying complex traits is to intersect GWAS hits with quantitative trait loci for molecular-level traits, like gene expression or methylation, which paves the way for linking genotype to disease. Cis-regulatory QTLs are particularly informative, as the vast majority of GWAS hits are found in noncoding regions.

With this perspective in mind, we used eQTL mapping to assess the contribution to ADHD of genetic variants altering cis-methylation levels in the brain. Starting from a number of differentially methylated CpG sites identified (influenced by SNPs that are associated with ADHD), we filtered those mapping into potential promoter regions of genes for which they were also eQTLs. This analysis resulted in several genetic variants affecting the methylation of six CpG sites, which in turn affected the regulation of three genes, *ARTN*, *C2orf82* and *PIDD1,* in multiple brain regions.

In addition to methylation, we investigated the possible contribution of miRNA dysregulation in ADHD. Our search revealed ADHD-associated variants located in the putative regulatory regions of miRNAs or in the promoter regions of their host protein-coding genes. We mined public databases for eQTL evidences on the ADHD-associated miRNA variants, although the effect of genetic variations on ncRNAs, including miRNAs, has been poorly explored. Therefore, given the eQTL annotation bias towards protein-coding genes, it is still difficult to functionally annotate ADHD-associated variants based on the expression levels of miRNAs. Recently, some studies have started to catalogue genetic variants associated with miRNAs whose genotypes affect gene expression in the human genome[161]. Interestingly, nearly half of the *cis*-miR-eQTLs are located 300–500 kb upstream from their associated intergenic

microRNAs. These *cis*-miR-eQTLs are highly enriched for *cis*-mRNA-eQTLs and regulatory SNPs and some of these have been associated with complex traits in prior GWAS[162].

1.4.2 Genome-Wide Pathway Analysis (GWPA)

GWPA uses GWAS or WGS data to aggregate all the individual SNPs into genes and gene-sets or pathways for the appearance of any over-represented/significant functional groups. Using annotation databases, any pathway analysis program maps individual variants to their respective genomic location, and the genes are positioned into gene sets or pathways. A cumulative p-value is calculated for each gene group or pathway from the p-values of the input SNPs and this cumulative p-value, if overcoming the significance threshold, will represent an enrichment of the corresponding functional group (Figure 6)[163].

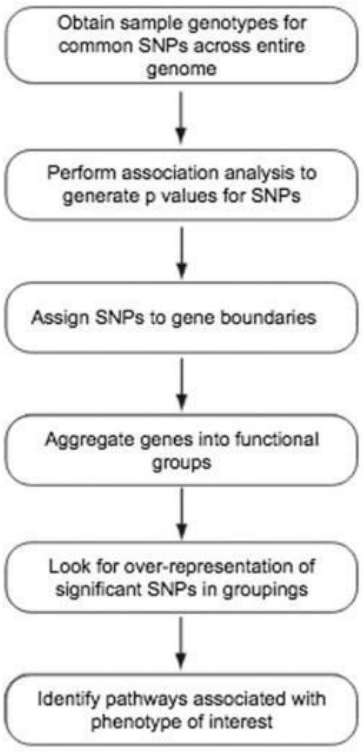


Figure 6. Overall protocol for pathway based GWAS/WGS analysis.

Adapted from White *et al.*, 2018.

1.4.3 Candidate and Hypothesis-Free Pathway Analysis

Candidate pathway analysis is a hypothesis-driven approach to investigate the enrichment of specific pathways, or gene sets of interest that are selected based on prior knowledge. The output from the candidate input pathways is a gene set or a list of gene sets that overcome the significance threshold in an association study with the disorder. However, factors such as the number of genes contained in each pathway can have an impact on the results. Also, hypothesis-free approaches can be used in pathway analysis that test all pathways in a given

database (e.g. GO or Biocarta) for association with a given phenotype. Enrichment scores for the known pathways are calculated, which are used to investigate whether there is an over-representation of genes in a certain pathway that are associated with the phenotype. These methods still do not necessarily prioritize the enriched sets for further analyses, as the p-value may not reflect the strength of association or significance of these gene sets.

Methods such as FORGE or SETSCREEN highlight the pathways comprising genes with multiple, independent association signals (both strong and weak), while others as ALIGATOR or INRICH pick up the single most significant SNP in a gene to assign significance to genes and will thus use those genes showing individually, stronger associations to detect enriched pathways. The generated pathway rankings from different methods are generally in accordance with each other.

In psychiatric genetics, pathway analyses carried out using GWAS data have reported significant associations of several biological processes to the disorders and show that the risk variants for psychiatric disorders aggregate in particular biological pathways. The underlying pathways are often common across these disorders. It is however debatable whether this observation can be due to annotation bias or if genuine pleiotropy exists across the underlying mechanisms as there are only tens of thousands genes for millions of traits[164].

In our study on the genetic basis of ADHD we constructed networks for the 12 miRNAs highlighted for ADHD. Two biological pathways were predicted arising from direct targets of the ADHD-associated miRNAs. One of the pathways showed a miRNA-mediated regulation of serotonin receptor genes (*HTR1D* and *HTR4*) and was suggested to be involved in neurological diseases and functions. This seems reasonable in the context of ADHD, given the reward deficiency mechanisms prominent in ADHD etiology.

In our work on the genetics of cocaine dependence we used MAGMA to evaluate both gene-based and gene-set associations using the summary statistics from our GWAS meta-analysis. For gene-based analysis, the *p*-values for SNPs located within the transcribed region were considered for the statistic test (SNP-wise mean model). A threshold of 10% false discovery rate (FDR) was applied to correct for multiple testing. In the gene-set analysis, we employed a competitive test procedure using: “All Canonical Pathways” (1329 gene sets), “GO” (4436 gene sets) and “BioCarta” (217 gene sets) provided by MsigDB and each gene set was individually corrected for multiple testing using permutation based empirical correction built

in MAGMA. We refrained from using Bonferroni correction in gene-set analysis as the categorised gene sets are strongly overlapping and Bonferroni test prove to be quite conservative in such cases. Our gene-based analysis mapped approximately three million SNPs from the GWAS meta-analysis to around 18,000 protein-coding genes, and a histone H2B type 1-D protein-coding gene (*HIST1H2BD)* showed a significant gene-wise association with cocaine dependence. One of the BioCarta immunity pathways "BIOCARTA TNFR2 PATHWAY" showed a trend (uncorrected $P = 5.38e\text{-}04$, corrected $P = 0.09$), being the most significantly associated among the canonical pathway gene sets. Seven of the ten GO gene sets with lower *p*-values, seven were reported from the processes of synapse organization, glutamatergic neurotransmission and brain functions innately relevant to the psychiatric abnormalities.

1.5 Shared genetics underlie psychiatric disorders

Multiple studies have attempted to find pairwise genetic correlations for psychiatric diseases[141]. The genetic correlation between pairs of disorders can be estimated using SNP data from GWAS, and these estimates determine if the disorders share genetic risk factors. For example, in Figure 7, the genetic correlation between bipolar disorder and schizophrenia is calculated as 0.6. For two disorders that are genetically correlated, some shared risk SNPs may appear for both disorders that may not appear for the individual disorders. Thus, by combining the polygenic risk score (PRS) analyses and GWAS, both coinciding and disease-specific genetic risk factors can be highlighted. In our study we calculated genetic correlation between cocaine dependence and six selected comorbid traits using LDSC and PRS (Detailed in Chapter 2), showing that comorbidity between disorders that is seen at the clinical level may be due, at least in part, to shared genetics.

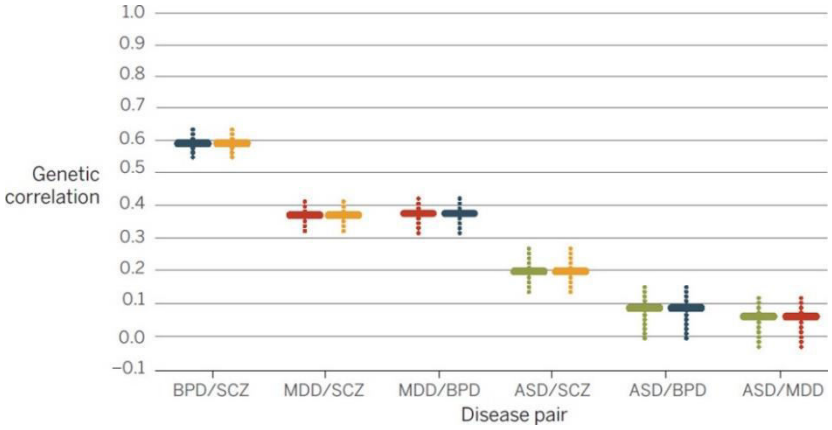


Figure 7. Pairwise genetic correlations for four psychiatric disorders.

Adapted from Geschwind and Flint, 2015.

1.6 Endophenotypes

The idea that some phenotypes (the so-called endophenotypes) bear a closer relationship to the biological processes that give rise to psychiatric illness than diagnostic categories is appealing. Endophenotypes include e.g. electroencephalographic variances or heritable behaviours that are recorded from laboratory standards such as neurocognitive execution deficits or impaired facial emotion recognition. Thus, endophenotypes may improve the odds of detecting genetic variants that predispose individuals to an illness. They are measured in both affected and unaffected subjects and as these are specific to disorders, they are perceived to be diagnostic biomarkers.

Endophenotypes in psychiatric illnesses are being increasingly researched on because (i) these are easier to work with than the psychiatric illness itself, (ii) they can enable the detection of genetic risk variants and genes using smaller samples, since the effect sizes of genetic loci contributing to individual endophenotypes are larger than those contributing to disease susceptibility.

Lookup

The current scenario is conducive for decoding the mechanisms of onset and progression of psychiatric disorders thanks to the advances in genomic technology, bio-computational methods, and to the initiatives of international consortia for building large clinical cohorts and for data sharing. In brief, to dissect a psychiatric disorder for both polygenicity (the small effects of individual loci), and large-effect rare loci, one of the needs has been to gather sufficiently large clinical cohorts. In practice, this implies that depth of phenotyping is likely to be lowered to be able to achieve huge cohorts[143]. However, methods like GCTA-GREML, network construction, eQTL mapping, targeting epigenetic systems are functional even if sample sizes are lesser, and these functional studies remain essential in recovering the “missing biology” and “missing heritability”. As the prevalence of neuropsychiatric disorders is on the rise, the quest for genetic or epigenetic biomarkers can assist in the development of novel therapeutics.

CHAPTER 2. IS THE COMORBIDITY BETWEEN ADHD AND COCAINE DEPENDENCE EVIDENT AND GENETICALLY DETERMINED?

2.1 Comorbidity between ADHD and cocaine dependence

Over the past decade, the clinical association between ADHD and SUD has become an increasing focus of investigation. Individuals reaching out to seek treatment for SUD commonly demonstrate the symptoms of ADHD. The very earlier community-based studies surveying the disorders occurring in population did not include adult ADHD. However, after the inclusion of ADHD in community-based surveys, the National Comorbidity Survey Replication (NCS-R) estimated the prevalence of adult ADHD to be 4.4%. The comorbidity survey uncovered that ~15% of individuals with adult ADHD also met DSM-IV criteria for a SUD as compared to ~5% of individuals without ADHD. This difference turns to be significant with an odds ratio of 3.0. It was also revealed that ~10% of the individuals with SUD meet criteria for adult ADHD, in contrast to the 4% individuals without SUD[127]. It is curious that while ADHD has long been documented to be a childhood onset condition and the symptoms of which can be discerned before the age of 12, SUDs are more noticeable only during adolescence and early adulthood. This observation is mostly attributed to the underlying ADHD symptoms of impulsivity, emotional dysregulation and so accordingly the poor social interaction and academic performances, which in turn enhance the overall setup for developing SUDs.

Findings on whether cocaine in particular is the preferred substance of choice in individuals with ADHD, and so the rates of cocaine abuse surpasses that of other SUDs in the ADHD group, remain inconclusive[165]. It was speculated that cocaine being a powerful psychostimulant may be used more frequently to self-medicate the symptoms of ADHD (knowingly or unknowingly) than alcohol, nicotine or cannabis. However, this is difficult to conclude as numerous factors like the legal availability of a substance, financial means to access it, individual's awareness about the harmful effects of the drug and individual response to the pharmacological action of drugs causing pleasant or unpleasant effects also dictate the extent of abuse of substances.

What is ambiguous is whether this relationship between ADHD and cocaine dependence is causal by nature possibly in these ways: (i) ADHD leads to self-medication with substances

including cocaine, (ii) ADHD subjects are impulsive and this favors the first contact with the drug, or (iii) substance use leads to ADHD through dysregulation of neurotransmitter systems. If not causal by nature, could this co-occurrence be the result of overlapping risk factors such as shared genetics and/or environment?

2.2 Is the comorbidity between ADHD and cocaine dependence genetically determined?

We now know that many psychiatric disorders that are comorbid or share some symptomatology do share common genetic risk factors as shown by several studies including heritable conditions like autism, ADHD, bipolar disorder, major depression, schizophrenia, anorexia, OCD, Tourette syndrome, anxiety disorders and post-traumatic stress disorder (PTSD)[155,166]. Similar to these most-widely studied disorders, substance use disorders are also heritable (h^2 = 40–70%), and highly comorbid with other psychopathologies[167,168]. Family studies have been equivocal about the co-occurrence of SUDs and severe psychiatric conditions like schizophrenia and bipolar disorder[167,169]. On the contrary, data from twin studies[170] suggest that the genetic factors entailing common psychopathologies also augment the general risk for substance use.

So far, only a few studies have explored the role of shared genetic influences on the comorbidity between substance use disorders and psychiatric conditions, and the causality of this correlation remains even less articulated. To be able to test whether shared genetic risk factors underlie the lifetime co-occurrence of psychiatric disorders, many bioinformatic approaches have been applied: PRS, genetic correlation, Mendelian randomization (MR) and multi-trait analysis of GWAS (MTAG). Using these methods, significant genetic overlaps have emerged between cocaine dependence and schizophrenia or MDD[171–173] and also between substance use disorders and other psychiatric disorders[174,175]. PRS generated from the Psychiatric Genomics Consortium (PGC) cross-disorder meta-analysis explained about 1% of the variance in general substance involvement factor in the target SAGE sample that is enriched for substance use[171].

PRS analysis reveals that the genetic variation underlying risk for clinically diagnosed ADHD also contributes to higher risk taking, and substance use (alcohol and nicotine; cocaine samples not included)[174]. PRS analyses now substantiate that genetic liability to ADHD is associated with a higher risk of SUD in individuals with ADHD. Although other risk factors like comorbid ODD/CD, male sex, parental factors (SUD, low paternal income, low maternal education, etc.)

also mediate the appearance of SUD in ADHD, the association between the common genetic liability to ADHD and SUD exists above what could be explained by other risk factors for SUD. Altogether, higher PRS-ADHD heightened the risk of any SUD, multiple SUD types and their severities[64]. In a GWAS meta-analysis of eight psychiatric disorders, it was found that 75% of the LD-independent associated regions were associated with more than one disorder[176]. These findings support that the co-occurrence of several traits with ADHD is explained, at least in part, by shared genetic risk factors.

In this Thesis we have investigated whether the phenotypic associations between cocaine dependence and six comorbid psychiatric/behavioral traits (ADHD, schizophrenia, MDD, risk-taking behavior, antisocial behavior and children's aggressive behavior) are genetically mirrored by performing genetic correlation analyses using two approaches: LDSC and PRS. For the first time, we found a significant genetic correlation of cocaine dependence with ADHD, MDD and risk-taking behavior, although these results should be taken with caution and need to be followed up in a larger sample of cocaine-dependent individuals. The PRS analysis included the individual-level SNP data, resulting in higher statistical power and allowed for direct testing of interaction effects. According to our results, all the tested comorbid conditions are associated with cocaine dependence status, suggesting that cocaine dependence is more likely in individuals with many risk alleles for the tested conditions than in those with fewer risk alleles.

So far in this section, we have elaborated on the shared genetics between disorders at genome-wide level, using methods that provide general figures of overlapping but do not point at specific biological functions. It has been hypothesized that because of the involvement of dopamine in both cocaine use disorders and ADHD, there might be some shared genetic bases underlying the co-occurrences of these disorders. Hypothesis-driven case-control association studies help to identify such overlapping genetic risk factors. An association exists between cocaine dependence symptoms and dopamine-related genes at the biological system level according to a genetic risk score based on SNPs from selected dopaminergic genes[177]. In this study, a cocaine dopaminergic genetic risk score accounts for variance in cocaine dependence symptoms that is largely independent of the variance coupled to other substance dependencies.

Dopamine is one of the key neurotransmitter systems in generating the rewarding effects of cocaine use and candidate gene studies support that specific variants underlying dopaminergic genes affect risk for cocaine dependence (Figure 8)[177]. For instance, polymorphisms in

SLC6A3, encoding the dopamine transporter (DAT/DAT1), have been repeatedly examined for response to cocaine and have been shown to confer risk of cocaine dependence and also overdoses/fatalities[177,178]. The differential expression of the same gene, *SLC6A3*, due to the presence of polymorphic variants, influences self-regulation skills and ADHD symptoms [179–182]. This receptor remains a favored target for pharmacogenetic drugs like methylphenidate for ADHD and disulfiram treatment for cocaine addiction[182,183].

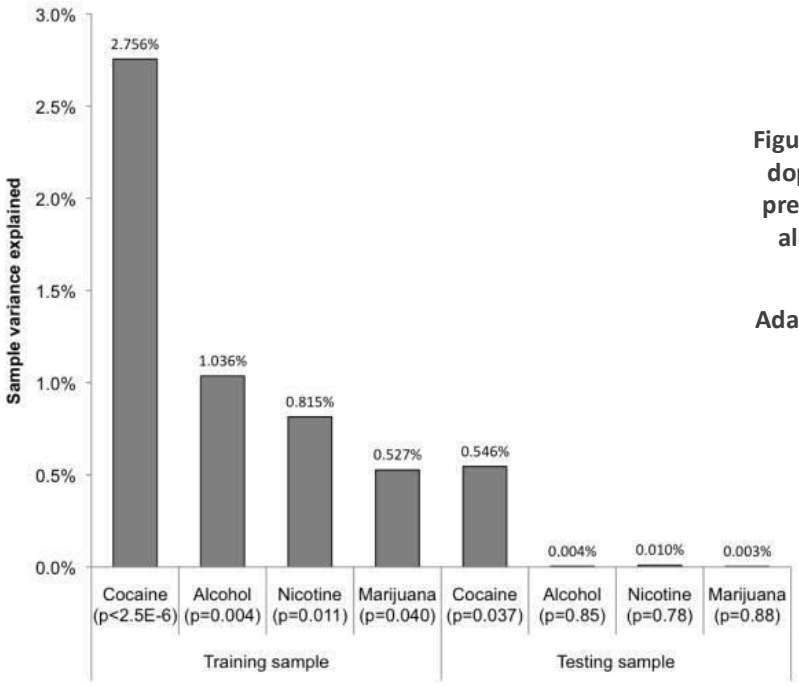


Figure 8. Ability of cocaine-identified dopaminergic genetic risk score to predict sample variance in cocaine, alcohol, nicotine, and marijuana dependence symptoms.

Adapted from Derringer *et al.*, 2012.

Similarly, variants in another dopaminergic gene encoding the dopamine D2 receptor (*DRD2*) have been associated with multiple addictions like drug dependency, alcoholism, smoking, pathological gambling and also ADHD and compulsive behaviors[184]. Besides adding to disease susceptibility individually, the dopamine receptor genes may show significant gene-gene interactions to form heteromers (like DRD2-DRD4, DRD2-DAT1) which upsurge impulsivity, novelty-seeking, addiction susceptibility, and ADHD-like symptoms[178,185–187].

Another noteworthy involvement relates to Human Immunodeficiency Virus Type I Enhancer Binding Protein 2 (HIVEP2), a dopaminergic transcriptional regulator in DA neurons. HIVEP2 can activate *SLC6A3* by targeting its intronic sequence and significant HIVEP2-SLC6A3 interactions were observed for SUD in male rat models and male clinical subjects[188]. The fact that HIVEP2 is functionally related to an ADHD gene, and that it is involved in the regulation of diverse neurodevelopmental pathways makes it pertinent to the genetics of ADHD. *De novo* likely damaging variants occurring in *HIVEP2* have been associated with intellectual disability

and developmental delay in humans, and *HIVEP2*-knockout mice likewise exhibit several working memory deficits, increased anxiety, and hyperactivity[189–191].

The shared genetic bases between these two conditions may well extend beyond the dopaminergic system. Evidences are now emerging for serotonergic neurotransmission (e.g. *HTR2A*) in the genetic elements that underlie the predisposition to cocaine addiction[192,193], and the genes modelling the serotonergic system have long been highlighted as important risk factors for the development of ADHD (e.g. *HTR2A*, *5-HT1B*)[194–197]. Even the polymorphisms in dopaminergic genes can alter serotonergic signaling as seen in the case of the rare DAT p.Ala559Val variant, found in ADHD cases and associated to drug-related behaviors. DAT Val559 mice models lack a locomotor response to cocaine and this arises from SERT blockade and an enhanced 5-HT signaling relative to the cocaine actions in wildtype mice[198]. More recently, *ADGRL3* (*LPHN3*), a brain-specific member of the latrophilin subfamily of G-protein-coupled receptors has been found to confer ADHD susceptibility and mediate methylphenidate pharmacogenetics. The gene is also most strongly expressed in brain regions implicated in the neurophysiological basis of ADHD. Interactions of *ADGRL3* variants with variants located on chromosome 11q improve the prediction of ADHD development and medication response[199]. An added possible functional role for *ADGRL3* has surfaced in modulating drug-seeking behavior, and *ADGRL3* is identified as a risk gene for SUD in different populations regardless of the type of abused substance. The variants in *ADGRL3* may also mediate individual susceptibility to the long-term protective effects of cocaine medication treatment[200].

In our meta-analysis of cocaine dependence, no genome-wide association was identified with any SNP, due to limited sample size (2,100 cases and 4,300 controls). However, we investigated the suggestive associations (p-value for association<1e-05) for functional relevance and found 22 genomic risk loci containing 112 genes. One of these risk loci is a genomic region on chromosome 6 (6p22.1) enriched in immune system and histone-related genes that also pops up in several schizophrenia GWAS. This observation supports the presence of shared genetic risk factors in these two comorbid disorders, although this should be further investigated in larger samples. The shared genomic region is defined by two lead SNPs and encompasses 77 genes and 458 nominally associated SNPs. All the SNPs in this region emerge as brain eQTLs for a small group of genes including *BTN3A2*, *HIST1H2AK*, *ZSCAN31*, *PRSS16* and *ZNF184*.

2.3 SUDs and late-onset ADHD

What is also curious is that ADHD is now being thought of more than a pediatric condition, with reports claiming the onset of ADHD in adults or late-onset of ADHD[201]. However, this is a controversial issue, as 'adult-onset ADHD' may reflect individuals not properly diagnosed in childhood or subthreshold subjects showing many of the ADHD symptoms before their formal clinical categorization[4]. In light of this, new dimensions surface for exploration. For example, whether individuals with SUDs and with no history of childhood ADHD can present ADHD symptoms during adolescence or adulthood. If so, what are the odds that this late-onset of ADHD in adulthood will contribute to the development of SUDs? With the availability of a longitudinal data for individuals, studying together the comorbid behaviors may also help in identifying the genetic, epigenetic, and environmental mediators.

Lookup

ADHD goes together with an expansive comorbid spectrum, which anticipates a worse lifetime trajectory; hence, a true diagnosis and treatment of ADHD is dependent on the accurate identification of the symptoms and subtypes. Understanding the presence of any comorbid phenotype enables the determination of a most debilitating disorder in an individual during the clinical screening, so that the associated behavioral symptoms are treated early. It is also not unlikely that various conditions might follow during the later course of ADHD or post diagnosis; hence a precise understanding of the comorbidities can help anticipating the ones to appear together with ADHD symptoms. What is important to remember is that ADHD might not be fatal by itself but be accompanied by behaviors/disorders that can be so in multiple ways. Being able to understand the genetic bases of the comorbidities is the first step towards devising predictors of the risk of comorbidities in an individual. For instance - A late diagnosis of ADHD (after age 13 years) is a novel risk factor identified for SUD[175]. This will help an individual to undertake any precautions and measures like CBT or pharmacological interventions to cope up with developing the condition or altogether circumvent the comorbidity.

CHAPTER 3. VALIDITY OF METHODS USED

There are multiple key issues that have to be considered in any genetic association analysis to attain robust results. Thus, importance of proper selection of patients and controls, sample size, accurate definition of disease phenotype, consideration of linkage disequilibrium, correction for multiple comparisons, and the need for functional assessment of disease-associated polymorphisms should guide a basic association design. While many of the methods in conducting a genetic association study are uniformly acknowledged, others can be specific to the study, or debatable and thus need to be tackled. We address here the methodologies used in our work, and the strategies to deal with challenging points.

3.1 Association studies in complex trait genetics

Genetic association studies compare the frequencies of alleles or genotypes at common variants in the affected group and controls. Both large-scale and small-scale association studies are widely used to determine variants contributing to genetic susceptibility in complex diseases. The studies can be carried out on families or on unrelated individuals; and the control individuals can be selected from unaffected family members, or community or hospital-based sources (Figure 9)[202].

Setting up family-based design can be challenging in terms of recruitment of related affected individuals and potential ascertainment bias; however, it is not affected by population admixture. On the contrary, recruiting unrelated groups is easier, but can contain population admixture. A cohort (prospective) design recruits individuals from a pre-defined population and independent of disease status. All these recruited individuals are followed longitudinally for the development of the disease. On the other hand, a case-control design is reflective (retrospective) in nature where individuals are ascertained by disease status[202]. The latter is the experimental design chosen for the association studies performed in this Thesis: A case-control GWAS in cocaine dependence (detailed in Chapter 4) and two case-control association studies with focus on SNPs with a potential impact on epigenetic variation (miRNAs and methylation) (Chapter 4). In both cases, the samples have been recruited in a clinical setting.

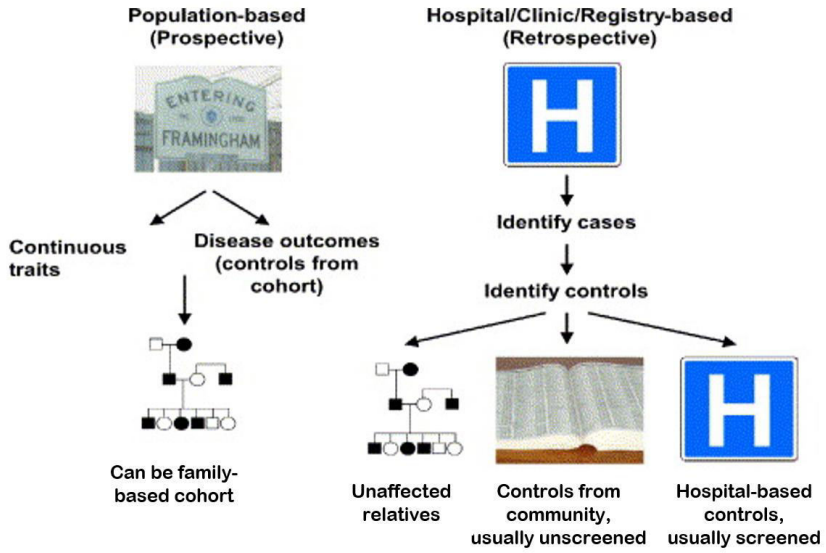


Figure 9. Methods to select individuals for a genetic association study.

Adapted from Newton-Cheh *et al.*, 2005.

3.1.1 Genome-wide association studies (GWAS)

GWAS highlights new risk loci or genes by inspecting a massive number of variants across the genome in a hypothesis-free manner. Over the last decade, GWAS has remarkably contributed to the detection of reproducible genomic loci associated with common traits not limited to somatic disorder like breast and ovarian cancers, coronary artery disease or type 2 diabetes[203]. In psychiatric genetics, more than 80 loci associated with depression have been successfully replicated in GWASs, which seems to end the debate on the lack of replicability of GWAS. Increasing sample sizes - despite introducing more phenotype heterogeneity - has led to the identification of large number of genetic variants. Thus, GWAS for depression used more lenient - "minimal" phenotyping to facilitate attaining of larger sample sizes to identify additional risk variants[157].

The GWAS landscape has appropriately expanded to employ complex approaches of large and cross-disorder meta-analysis, pleiotropy and MR. Many of these new methodologies involve re-analysis of summary statistics results from GWAS. Full *P*-value summary statistics are defined as the aggregate *P*-values and association data for every variant analyzed in an independent GWAS. GWS variants are not necessarily causal and may tag the real causal variant(s).

Our work on ADHD is founded on the association strategy that benefits from the summary statistics of the largest and well-powered available ADHD meta-analysis, performed on 12 case-control samples from different populations. We used this dataset to interrogate SNPs that

influence methylation, or SNPs with potential to alter miRNA expression. The strength and direction of association were deciphered from the original ADHD meta-analysis, but the top findings among those disease-associated SNPs potentially related to epigenetic changes were determined post corrections for our numbers of analyzed SNPs (i.e. SNPs with potential epigenetic effects). On the other hand, the association study with cocaine dependence required performing a GWAS meta-analysis ourselves, using previous data from four independent GWAS datasets. One of the previously GWASed samples (SAGE) included both cases and controls, which were non-dependent unrelated subjects. The other three samples, all accessed via the public repository dbGAP, included only cases, and therefore we used independent control samples, which included unscreened unrelated individuals with the same ethnic origin.

3.1.2 Meta-analysis of GWAS

Meta-analysis is the method to combine the results of multiple studies that perform genome-wide genotyping to improve power for finding true associations. Meta-analyses have led to the identification of thousands of genotype-trait associations. For instance - meta-analyses coalescing primary datasets have led to a better understanding of complex traits including height, body mass, Crohn's disease and Type 2 diabetes mellitus, and the method has been extended by the PGC for illuminating the genetic architecture of schizophrenia, MDD, bipolar disorder, ADHD and autism, among other psychiatric conditions. Therefore, meta-analysis has become essential in human complex trait genetics[204]. We performed meta-analysis using four reported GWAS on cocaine dependence since individual genome-wide genotype data can often be underpowered. A study-specific GWA QC for each SNP was carried out, and genome-wide SNP array data was imputed. We then computed association statistics for each SNP, including effect size estimates, allele frequencies, and p-values. In most GWAS meta-analyses, the unavailability of individual participant data in accordance with data sharing guidelines creates unique analytical challenges for QC, requiring specific statistical and graphical tools to track errors in the study-specific analysis from the available aggregated data[205]. This was not the case for our study, as we gained access to individual genotype data from all individuals in our cocaine dependence genetic association study, through the dbGAP platform.

3.2 Hypothesis-free versus hypothesis-driven analyses

The two main approaches in SNP selection for association studies are: (i) hypothesis-free association study (GWAS), in which a huge number of SNPs in the range of several hundred

thousand to over a million are studied for association with a phenotype in cases versus controls; and (ii) a hypothesis-based approach, in which SNPs from relevant, hypothesis-driven pathways are examined. These approaches are however not as dualistic as they may seem. For instance - A GWAS in schizoaffective disorder, bipolar type pointed to involvement of gamma-aminobutyric acid $(GABA)_A$ receptor β1 subunit, GABRB1, which was then a starting point for a hypothesis-driven examination of variants in genes encoding GABA receptor subunits in several bipolar phenotypes. This genetic association was found to be remarkably specific to a precise sub-phenotype in the studied diagnostic category[206].

GWAS is independent of the prior bias of traditional biology and thus it improves the odds of an all-inclusive description of the genetic causes of complex diseases. Although tagged "hypothesis-free", GWAS are reliant on underlying or *a priori* hypotheses and dictated by the design of genotyping platforms and analysis methodologies. The implicit hypotheses in a GWAS are: (i) common disease/common variant hypothesis, (ii) genotyped SNPs are the genetic variants responsible for the disorder or are proxies for the causal variants, and (iii) genetic predisposition to complex disorders is convened by independent effects of SNPs. Therefore, the outcomes of any psychiatric GWAS are principally determined by the extent to which these hypotheses hold true[207]. Thus, our work to discern the genomic regions that underlie cocaine dependence using GWAS relies on the existence of a significant SNP-based heritability for the studied phenotype, i.e. on the assumption that a relevant fraction of the heritability of the disorders is due to SNP variation.

In the present Thesis, we also use a mixed hypothesis-driven/hypothesis-free approach for the association study of ADHD with SNPs in miRNA genes: Hypothesis-driven because we targeted a group of functional elements of the genome (miRNAs), that we suspect may be involved in ADHD etiology because of their regulatory role; and hypothesis-free because we targeted all miRNAs in the genome.. Thus, we systematically interrogated those SNPs or small indels that tag the genomic regions encompassing the miRNA genes.

Similarly, our exploration of ASM in ADHD revolved around the hypothesis that SNPs influencing DNA methylation *in cis* may be involved in the etiology of the disorder based on their potential impact on gene regulation. We built our work on two existing studies, each of which identified SNPs at genome-wide scale that correlated with differential levels of methylation in brain tissues. Here, we consider that changes observed in brain tissues will

directly bridge the ADHD biology. Again, our rationale is based on a mixed hypothesis-driven/hypothesis-free approach.

3.3 Selection of genes and polymorphisms in candidate systems

In hypothesis-driven association studies, the common genetic variants are chosen based on revelations or speculations from clinical, genetic, pharmacological or animal model studies, among other inputs.

Our investigation of the ASM system in ADHD was hypothesis-driven in the sense that the SNPs selected for the analyses were based on evidences from two previous studies. These studies identified abundant QTL for DNA CpG methylation across the genome, i.e. SNPs influencing methylation in multiple brain regions of post-mortem human samples. We mined our set of candidate variants from these two published integrated datasets that defined mQTLs or eQTLs as correlation between SNP genotypes and DNA methylation or expression.

Our work on the role of miRNA variation in ADHD started with a systematic selection of SNP variants that encompass all reported miRNA genes. We did that in a hypothesis-free manner (we explored all miRNAs) in view of these factors: (i) Until recently, only specific miRNAs have been explored in psychiatric disorders and a complete investigation of the miRNA common variation is yet lacking; (ii) there is a scarcity of studies documenting SNPs that dysregulate miRNA expression. Moreover, we systematically addressed miRNA-specific intricacies like (i) the presence of clustered miRNA genes versus singleton miRNAs; (ii) the intragenic versus intergenic location of miRNAs; and (iii) a majority -but not all- of intragenic miRNAs are transcribed in the same direction as the host gene, and in this situation they can share regulatory elements. Integrating all these conditioning factors, we selected small SNP and indel variants that tag miRNA genomic regions.

3.4 Technical decisions in the association workflows

3.4.1 Phenotype heterogeneity

In psychiatric genetics, one of the most crucial issues while performing a case-control study is how to define the cases and controls. Studying complex phenotypes necessitates an adequate selection of the sample. To facilitate the identification of genetic risk factors, the patient group needs to be as homogeneous as possible. However, while selecting patients with addiction

disorders, it becomes challenging to control the variability because of the practice of multi-drug abuse by the substance users[208]. For our study, subjects who received a diagnosis for cocaine dependence per DSM-IV guidelines, involving the standardized Structured Clinical Interview (SCID), form the case samples.

Although a large proportion of these patients had also received diagnosis for other drug abuses or dependencies: 41.1% alcohol, 35.6% cannabis, 22.7% opiate and 6.8% benzodiazepines, they all have in common the addition to cocaine. In any case, comorbidity is the rule rather than the exception in all psychiatric illnesses. Approximately three-quarters (73.4%) of patients with cocaine abuse or dependence also present comorbidities, such as MDD, schizophrenia, ADHD, anxiety or personality disorders[209]. Such comorbidities may interfere with the true associations, and a potential solution to uncover these true associations is to compare the results from multiple individual disorder-based case studies and inspect the replicated hits. However, as of now, there is a scarcity of studies that focus on a single drug use due to limited availability of samples and therefore, selecting the multi-drug abusers as cases seem to be the sole choice in studying illicit SUDs like cocaine-use disorder.

3.4.2 Selection of the controls

Another debated issue for association studies in substance use disorders is the selection of controls as in terms of exposed and non-exposed controls. Some experts in the field argue that an appropriate set of controls for drug dependence studies are the individuals who have been exposed to the drug of interest at least once in their lives and have not developed a dependency to that drug[210,211]. In such a scenario, the association study would capture the predisposing genetic component involved in the transition from use to addiction, but it would exclude all possibilities for examination of key risk factors for drug dependence i.e. impulsivity and risk-taking behavior. These risk factors are the primary compelling drivers for individuals' first contact with drugs and show a high genetic component[212]. Other experts favor the practice of using control individuals who do not show dependency to any drug of abuse, irrespective of their exposure status to the drug[213,214]. Most published genetic studies in addiction tend to utilize unexposed controls which is suitable for assessing dependencies to drugs but may lead to a reduced power when analyzing intermediate or later stages of addiction[215]. GWASs of alcohol and nicotine dependence typically use exposed controls. However, for studying illicit drug dependency, this can severely reduce the sample size of the control group.

In order not to miss those risk factors that explain the initial steps that lead to drug addiction (e.g. risk-taking behaviors), we used unscreened controls from the general population accessed from blood donors from the Blood and Tissue Bank of Barcelona. The blood donation protocol excludes individuals who have ever injected non-prescribed drugs. We also estimated that any probable contamination of cocaine-dependent individuals in our control sample from the general population is less than 1% (given that 3.4% of adults in Spain consume cocaine at least once in their lives[216], and some 15-16% of these individuals will develop dependence within 10 years of first cocaine use[217]). This approach, due to the presence of some cases in the control sample, may possibly cause us to miss a few true association (false negatives), but should not generate false positives. Therefore, the likelihood of obtaining altered results due to this selection bias if any would be negligible.

3.4.3 Confounding factors

The samples selected for association studies may be a mix of individuals belonging to different groups because of ethnicity differences, or due to technical disparities in genotyping. The presence of subgroups in the study sample may reflect a population stratification bias. In an association design, both cases and controls should be represented in an equivalent manner in terms of existing subgroups, else the population stratification may cause false positive associations: i.e. the differences observed in the allelic or genotypic frequencies between cases and controls would be due to the factor that differentiates the subgroups rather than the phenotype investigated for[218]. It is therefore imperative to construct genetically homogeneous samples of cases and controls with individuals from the same region and/or same ethnic group. We therefore limited our cases and controls to individuals of European ancestry.

More confounding variables can exist besides population stratification, again leading to false positive associations if not corrected for. For instance, sex of an individual is known to be a confounding variable in association studies on cocaine dependence as this disorder shows a higher prevalence in males. To control for this confounder, we maintained the same gender proportion in both cases and controls. Yet another possible confounder is age, with a differential distribution in cases and controls, and this was included as a covariate in our analysis.

In GWAS studies confounding factors can be readily detected due to the availability of large numbers of genotypes. Statistical method like multidimensional scaling (MDS) and principal

component analysis (PCA) allow for the identification of any underlying population stratification and confounding variables (like age, sex and related individuals)[219]. We used PCA to identify the principal components (PCs) to be used as covariates in our case-control association design[219]. As a general rule, the first 10 or 20 PCs are considered covariates, to eliminate the need of figuring out the possible confounders, and this approach reduces the bias in the downstream analyses. The four GWASs used for our meta-analysis on cocaine dependence had also addressed the issue of population stratification and other confounding variables through these approaches.

3.4.4 Genotyping errors

Genome-wide association studies start with automated genotyping of a massive number of evenly distributed polymorphisms, typically SNPs, in a large number of samples, and erroneous allocation of genotypes may arise when the observed genotype for an individual does not correspond to the true genotype. It is therefore usual to include some sample duplicates in the genotyping plates, samples with known genotypes and negative controls to test the technical consistency of the results. Poor quality or low DNA concentration are the main contributors to errors in genotyping[220]. To restrict spurious results arising from genotyping errors, the genotyping rate is determined per variant and per individual and the markers not meeting a certain threshold (80-90%) were eliminated from our analysis. Genotyping errors harbored in the assay can be spotted in the form of SNPs that show deviations from Hardy-Weinberg equilibrium (HWE)[221]. We applied the HWE checks and a high genotyping rate to resolve any low-quality genotyped markers prior to imputation.

3.4.5 Genotype imputation and controls from other studies

Genotype imputation is a technique that allows for an accurate evaluation of the evidence for association of a phenotype with genetic markers that are not directly genotyped, based on the patterns of LD among these markers in a reference sample (Figure 10). Imputation is an essential tool in GWAS design and increases the power of GWAS. It is particularly useful for standardizing and combining the association results across studies that rely on different genotyping platforms[222], also an issue encountered in our work. The genotype data across different studies is required to be grouped prior to the imputation[222] and in an ideal scenario, these data should be generated using the same genotyping platforms. We observed that grouping cases and controls from the same study but genotyped on different chips would still

show evidences of population stratification. To resolve this known erroneous source, we selected the controls from other studies but genotyped on the same platform and restated the analysis using this matched set of controls.

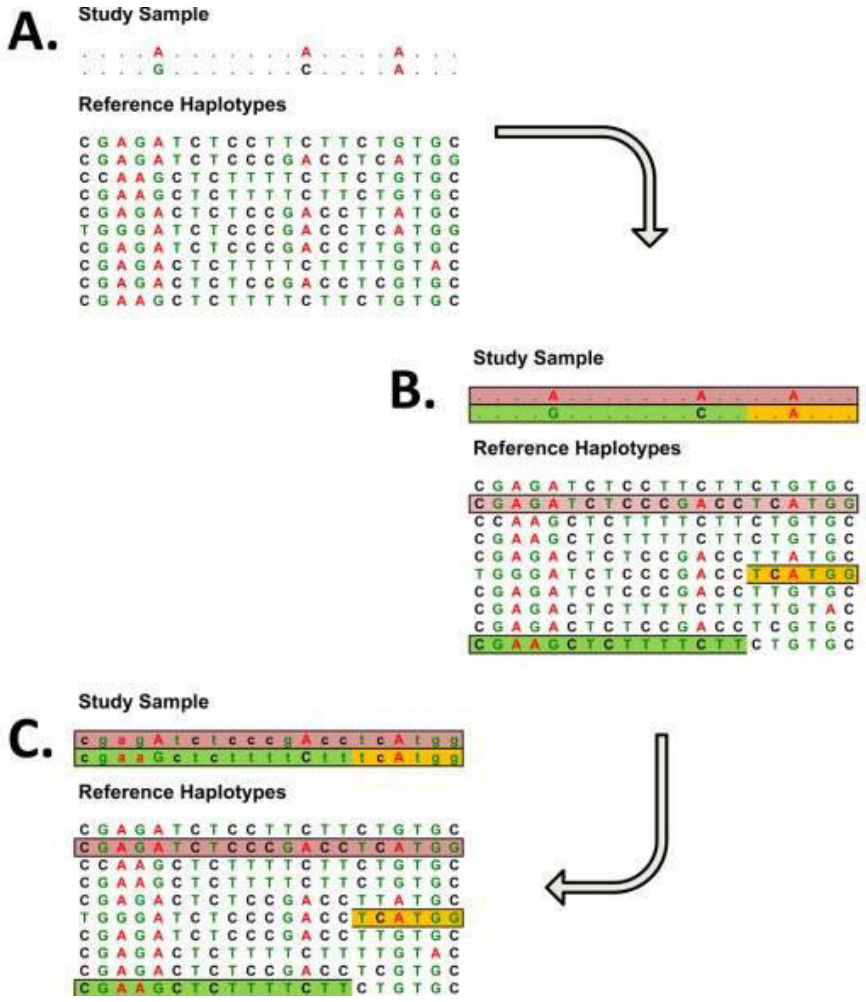


Figure 10. Genotype imputation within a sample of apparently unrelated individuals in three consecutive steps (A, B, C).

Adapted from Li *et al.*, 2010.

As a good quality imputation relies on the reference panel employed, the panel must comprehend as many individuals as possible and must be genetically/ethnically similar to the target sample[223]. A key issue in imputation hereafter is deciding which markers are taken forward for analysis. Not all markers can be well imputed, and different measures have been suggested to help identify satisfactorily imputed markers. For example – including genotypes that are imputed with >90% certainty (the average probability that an imputed genotype call is correct) or utilize the r^2 coefficient that captures the correlation between imputed genotype calls and the true underlying genotypes[222]. We incorporated a set of only high-quality markers where autosomal SNPs were found in all datasets, MAF > 0.05, HWE *P* value > 1e-03, SNP call rate > 0.98, and which were finally pruned for LD (with $r^2 < 0.2$ in a 200Kb SNPs window).

3.4.6 Sample size and statistical power

The statistical power in any association study refers to the probability of detecting a real association and is directly dependent on the sample size used. Multiple factors can influence the estimation of statistical power, such as disease prevalence, sample size, the allelic frequencies of markers associated with the disease, odds-ratio of these markers (i.e. the risk they confer) and the inheritance model[224]. In our association study on cocaine dependence we have aggregated cases and controls from four previous works, together with additional control individuals. As a result, the overall sample size has increased and so has the statistical power to detect true associations than in either of the input studies when taken individually. However, although our GWAS meta-analysis comprises the largest sample of cocaine-dependent cases of European descent, the statistical power is still insufficient to detect significant associations at genome-wide scale.

3.4.7 Multiple testing correction

As a massive number of polymorphic variants are tested for in GWASs, the number of statistical comparisons performed to detect associations increases, and concurrently increases the likelihood of detecting random associations or false positives. Different rigorous thresholds for correcting multiple comparisons are therefore applied to control the false positives, although there is no single universal method. The most rigorous of the methods curtails the false positives but may increase the false negative associations; and vice versa. We implemented the following two methods to control for multiple tests:

(i) Bonferroni correction: Bonferroni determined threshold restricts the probability of finding false positives to less than 5%. A new threshold of significance (α') is established based on the number of independent tests performed (n) and is calculated as $\alpha' = \alpha / n$, where $\alpha = 0.05$ and the null hypothesis is rejected only if the p-value is less than the α'. The method is highly conservative and assumes tests independence (which is often not the case), in a way that the probability of false negatives increases considerably, when a high number of tests are performed.

(ii) FDR: It estimates the proportion of false positive associations from all the associations obtained. It can be expressed as $f_p/f_p + t_p$ where f_p is the number of false-positive associations and t_p is the number of true-positive associations. The FDR correction is much less likely to

eliminate true associations (false negatives) at the expense of having an acceptable proportion of false-positive associations.

While the Bonferroni false positive rate of 0.05 means that 5% of all results will be truly negative, the FDR value of 0.05 means that 5% of declared positive results are truly negative. Some authors argue that the most preferable approach is FDR[225]. For our association studies of ADHD with miRNA variation and ASM-variants, we applied FDR corrections; although in both these works, we also obtain results that attain Bonferroni significance. On the other hand, GWAS use a universal genome-wide significance Bonferroni corrected threshold of $5x10^{-8}$ ($0.05/10^6$), considering one million independent tests assuming that each SNP is independent of each other. However, this is not the case given the dependent nature of genetic data, where SNPs in LD are correlated to some degree[226]. Therefore, a second threshold has been established at $1x10^{-5}$ to discover variants showing a suggestive association, as meaningful associations can lie 'hidden' below current thresholds and these 'sub-threshold' signals may represent novel loci[227].

Lookup

In an attempt to identify reliable disease-associated signals through our studies, we applied methodologies adapted to the psychiatric association studies. Both hypothesis-free and hypothesis-driven approaches were incorporated to test multiple hypotheses underlying the disease etiology of ADHD and cocaine dependence and shared risk factors. We investigated potentially functional SNP sets with impact on epigenetic variation (methylation or miRNAs), and also SNPs on a genome-wide scale. We applied stringent statistical measures to control for false positive signals and possible confounders (like population stratification, gender or genotyping errors). While we used the largest pre-existing summary statistics for ADHD to detect associations in two epigenetic systems, the sample size of the GWAS performed on cocaine dependence is still limited and needs further amplification.

CHAPTER 4. INTERPRETING THE UTILITY OF DETECTED GENETIC ASSOCIATIONS – AND THE CHALLENGES

4.1 Utility of associations identified in three systems

The three systems in question - miRNA genes and DNA methylation in ADHD and protein-coding genes in cocaine dependence - essentially vary in nature and so do the methods that substantiate the highlighted genetic loci in these systems. We used a number of bioinformatic tools together with annotation resources to connect the contributions of associated variants to gene expression or phenotype. Below we elaborate on the prioritization of association signals.

4.1.1 System 1: miRNA genes in ADHD

Our analysis involved a total of 1,761 autosomal miRNA genes out of 1,881 published miRNAs (miRBase v21) that were flagged by approximately 22,000 tagSNPs. Inspection of tagSNPs in the summary statistics of ADHD meta analyses (that contained 76.3% of the tag variants) revealed 19 significant associations with ADHD and highlighted 12 miRNAs. All these miRNAs are located within introns of host protein-coding genes. The associated variants lie in the putative regulatory regions of the miRNA genes or in the promoter regions of the host protein-coding genes; however this is not that rare since the actual miRNA gene is only 19-21 nucleotides long, and the likelihood of a tagSNP actually falling within few base pairs can be ultra-low. Also, about 51% of the miRNAs in the genome are located within the sequence of a protein-coding gene. In any case, our results raise the question whether the disorder associates with the host protein-coding gene or with the miRNA contained in it. Notably, two of the highlighted loci in this analysis, on chromosomes 1 and 7, have been reported as among the top ADHD risk loci in the source ADHD GWAS meta-analysis.

We annotated the highlighted miRNAs using (i) brain-expression data (ii) target gene analysis, and (iii) pathway analysis.

(i) Brain-expression data

Around 75% of annotated miRNAs are detectable in human brain[228] and 70% in the mouse brain[229]. There are many miRNAs that are specifically or highly expressed in the mammalian brain compared to other organs, and they are differentially distributed between distinct brain areas[228,229]. For instance - miR-128 and miR-124 are brain-enriched miRNAs and miR-9-1 is brain-specific[229]. Even among the closely related cells of the developing brain, miRNA

abundance exhibits cell type-specific patterns and highly dynamic changes in the expression (Figure 11)[230].

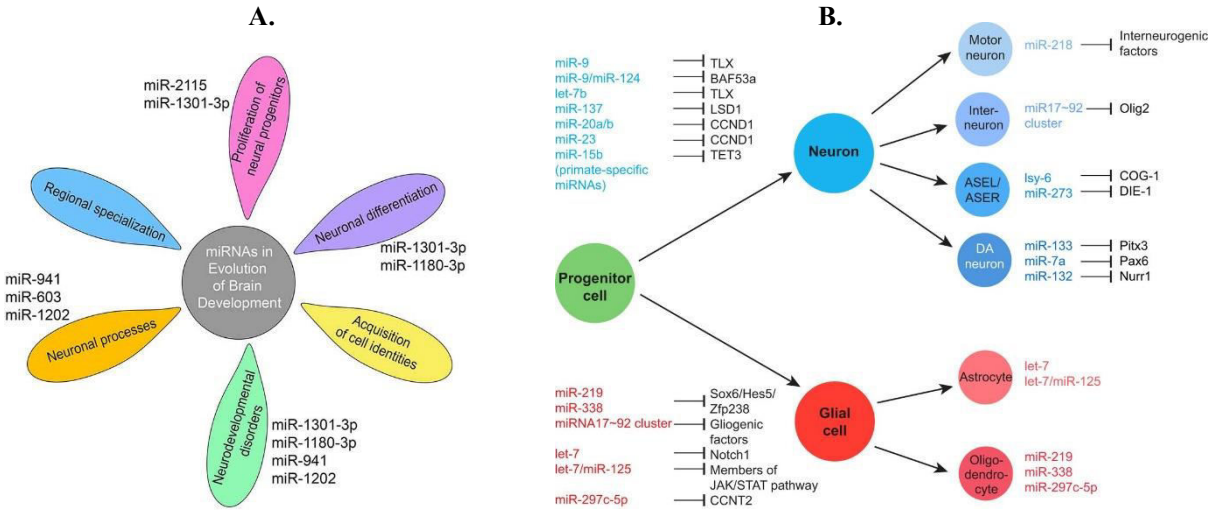


Figure 11. A: Examples of miRNAs that shape gene networks during the evolution of human and non-human primate brain development. B: miRNA function in neuronal and glial cell fate determination.

Adapted from Prodromidou and Matsas, 2019; and Rajman and Schratt, 2017.

Therefore, looking into the miRNAs' expression profiles across brain tissues is essential to prioritize the relevant miRNAs. All but one of the 12 highlighted miRNAs are brain-expressed according to the information from different expression databases. In a tissue-wise expression dataset across the brain, cerebellum, heart, testis and kidney (https://bmi.ana.med.uni-muenchen.de/miriad/), miR-6734 and miR-7-1 were more expressed in the brain and cerebellum than in other reported tissues. Of interest is miR-4655, which is seen to be expressed solely in the brain. The expression profiles from the cortical and subcortical structures of human brain revealed the presence of miRNAs in cerebellar cortex (miR-7-1, miR-3135a), primary somatosensory cortex (miR-3666, miR-4271, miR-4655-3p), primary visual cortex (miR-4655-5p) and ventral parietal cortex (miR-5193). MiR-7-1 was also found to be differentially expressed between PFC and cerebellum during late childhood development. PFC is critical for 'high-level' executive functions, including working memory, sustained attention, decision-making, and emotional control[231] (Figure 12). ADHD is believed to result from weaker structure and function of PFC circuits, especially in the right hemisphere. In adolescent SHR, a model for ADHD, the diminished function of glutamate receptor (AMPARs) is observed in the PFC, which can be restored by the administration of a clinically relevant dose

of methylphenidate[231]. Likewise, all effective pharmacologic treatments for ADHD enhance catecholamine signaling in the PFC and strengthen its regulation of attention and behavior (Figure 13)[232].

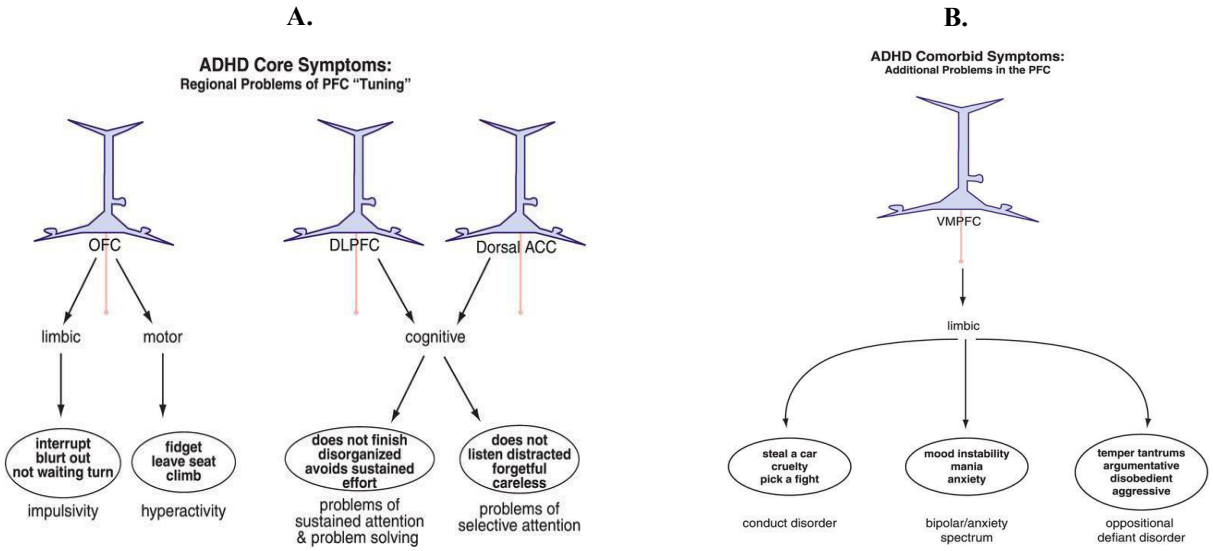


Figure 12. A: Out-of-tune PFC in ADHD. B: Additional dysfunctions within the PFC–limbic network that result in comorbidities associated with ADHD.

Adapted from basicmedicalkey.com.

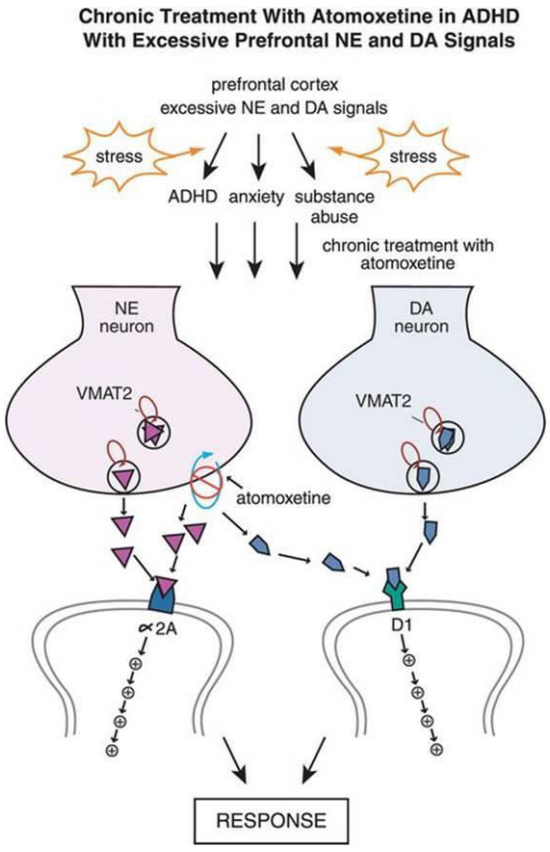


Figure 13. Chronic treatment with atomoxetine in ADHD. NE: norepinephrine; DA: dopamine; VMAT2: Vesicular transporter.

Adapted from basicmedicalkey.com.

(ii) Target gene analysis

Analysis of miRNA targets has gained momentum in psychiatric phenotypes and is useful in connecting the underlying regulatory mechanisms. Investigations on schizophrenia have identified the role of putative miR-137 target genes through over-expression experiments of miR-137 *in vitro*. Among the several loci significantly associated with schizophrenia is the SNP within an intron of a host gene that encodes a long non-coding RNA (lncRNA) and contains the miR-137 gene. Interestingly, four of the other identified GWS hits in schizophrenia were predicted to be targets of miR-137, which tied the functionality of the pri-miR-137 SNP in schizophrenia. MiR-137-mediated regulation of the four genes was further confirmed *in vitro* by luciferase assays[233]. A gene, *RORa*, associated with both autism and schizophrenia, was found to be directly repressed by miR-137[234]. Likewise, in MDD, downregulation of primate specific miR-1202 is observed in the PFC. Bioinformatic and *in vitro* studies supported a glutamate receptor gene, *GRM4*, as a target of miR-1202, which was subsequently implicated in anxiety-related behaviors and forms an attractive drug target[235].

Along these lines, we investigated connections of ADHD-associated miRNAs with genes known to be associated with traits that underlie ADHD. Three of the highlighted miRNAs - miR-3666, miR-7-1 and miR-1273h have validated target sites - 9, 18 and 1 mRNA, respectively. Some of these target genes have previously been reported as associated with psychiatric traits in the NHGRI-EBI Catalog of human genome-wide association studies (https://www.ebi.ac.uk/gwas/). The *EGFR* gene, targeted by miR-7-1, is located in one of the top regions for lithium-responsive bipolar disorder. Another gene targeted by miR-7-1 is *EIF4E,* which is associated with cognitive empathy and depressive episodes in bipolar disorder. Additionally, SNPs in the miR-3666-targeted *TAC1* gene have been found associated with general risk-taking and feeling nervous traits in previous large-scale GWASs. *MEOX2*, targeted by miR-3666, is associated with brain region and intracranial volumes. We observe using the Genotype-Tissue Expression (GTEx) data that the targets of miR-7-1 (e.g. *SLC17A7*, *SNCA*) are highly expressed in brain relative to other tissues, while the targets of miR-3666 show low to moderate expression in brain. It is likely that, as additional genes will be identified in even larger ADHD GWASs, expression of some brain-expressed genes found associated with ADHD may be directly shaped by the variation contained in miRNAs, as observed in the case of schizophrenia.

(iii) Pathway analysis

Pathway analysis helps to characterize the biological functions controlled by the miRNAs and interpret those which would be dysregulated due to altered miRNA-mRNA binding or to altered levels of the miRNA (Figure 14). The rationale is the following: i) Every single miRNA has multiple targets, ii) several microRNAs may contribute to ADHD susceptibility, iii) not all the genes targeted by these miRNAs do contribute to ADHD susceptibility, and iv) finding that several of the targeted genes belong to a functional pathway that is relevant to brain function would help to identify the relevant ones. We conducted pathway analyses using Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis), and included only experimentally validated gene targets to avoid potentially erroneous miRNA targeting.

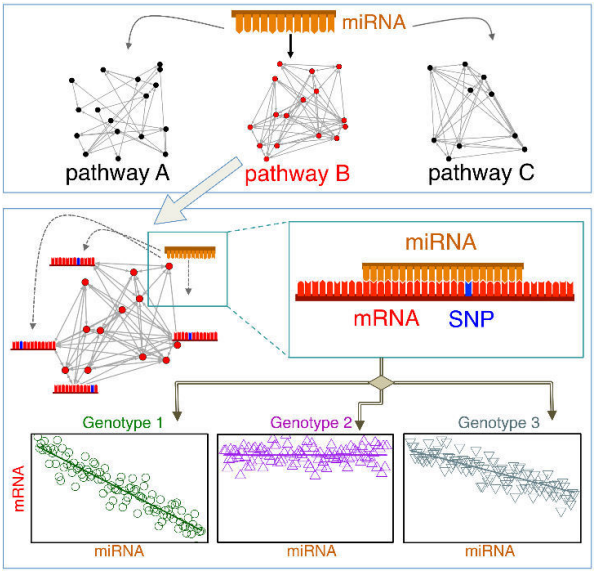


Figure 14. Dysregulated pathways due to the genetic variants that modify the miRNA-mRNA relationship.

Adapted from Wilk and Braun, 2018.

We found two biological pathways that arise from the direct regulation of ten and nine focal miRNAs respectively. One of the pathways is involved in neurological diseases and arises from the miRNA-mediated regulation of two serotonin receptor genes - *HTR1D* and *HTR4*. In this pathway, miR-4271 and miR-5193 inhibit *HTR1D* and *HTR4*, respectively, including several other genes. Another gene targeted by miR-4271 in the network is *YWHAG,* previously found associated with schizophrenia and encoding a protein that mediates signal transduction by binding to phosphoserine-containing proteins.

Aggregating these functional annotations, we propose at least miR-7-1 and miR-3666 as promising candidates since both are brain-expressed miRNAs, have validated brain-expressed targets, and homologs in model species.

4.1.2 System 2: Genetic variation that influences brain methylation in ADHD

We examined the possibility of a connection between ADHD and genetic variants that have been reported to alter methylation in the brain. In the process, we identified a total of 60 variants from eight LD blocks that are associated with ADHD and which were then found to correlate with differential levels of methylation at six different CpG sites.

We carried out a four-fold functional annotation of the significant ASM-SNPs:

(i) HaploReg and ENCODE data

One of the popular methods to annotate non-coding variants is the HaploReg tool, which annotates variants with respect to ENCODE data. HaploReg has effectively characterized SNPs associated with cardiovascular disease, autoimmune disorders, cancer, diabetes, and neurological disorders[236]. It can be used to find if the SNP of interest or nearby loci are positioned in defined promoters, enhancers, or protein binding sites. Active promoters are enriched for H3K4me3 and histone H3 or H4 acetylation. Primed enhancers are marked by H3K4me1 together with the depletion of H3K4me3, whereas active enhancers are enriched for H3K4me1 and H3K27ac[237] (Figure 15). An increasing number of methylation-related functions of specific sites (like H3K4, H3K9, H3K27 etc.) are implicated in major psychiatric diseases. For instance, increased levels of H3K4 methylation mark are found in the hippocampus during memory formation, and modifiers of H3K4 methylation are mutated in cognitive impairments[238]. Similarly, genetic risk variants for seven major psychiatric traits (including ADHD) are found to be enriched in cortical H3K27ac domains[239].

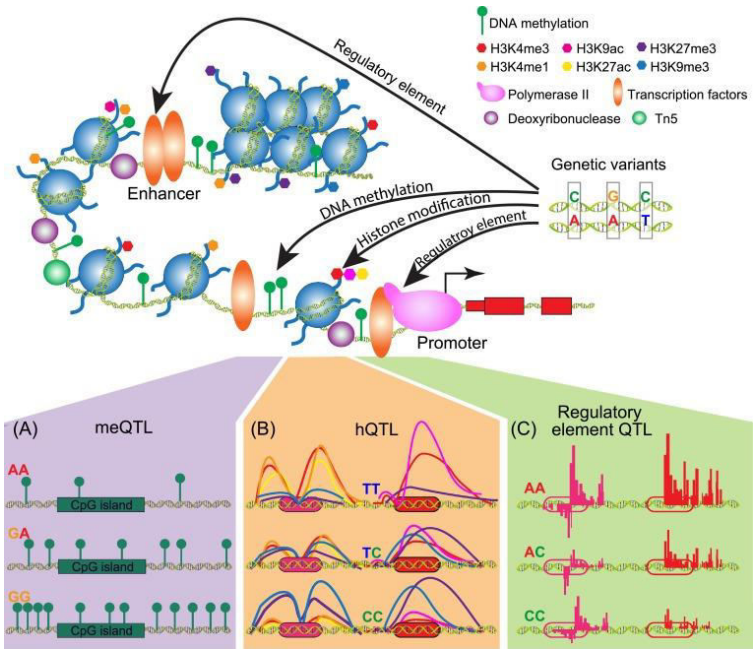


Figure 15. Genetic variants are associated with epigenetic regulation, including DNA methylation (A), histone modification (B), and regulatory elements (C).

Adapted from Ye *et al.*, 2020.

We therefore annotated the regions containing those ASM-SNPs found associated with ADHD for the presence of histone modifications related to enhancer regions (H3K4me1 and H3K27ac) and promoters (H3K4me3 and H3K9ac) in 10 different brain regions. We observed that 85% of the 60 ASM-SNPs are located within a region with an enhancer or promoter histone mark in at least one brain area. Correspondingly, all the SNPs in the LD blocks of ASM-SNPs lie within regions with histone marks, ranging from 3 to 17 in enhancer regions and from 4 to 16 in promoter regions.

(ii) eQTL analysis

One of the first methods developed to map the functional effects of non-coding variants is the inspection of the eQTLs, that is, SNPs at which the genotype correlates with expression of one or more genes. Mapping of eQTLs within haplotype blocks drawn in from the ASM association results can point genes whose genetically regulated expression is implicated in the phenotype. To date, inquiries into *cis*-acting eQTLs are more common than those that operate in *trans* since resolving *trans*-acting eQTLs can present computational challenges[99].

We analyzed eQTLs through the GTEx portal for all available brain tissues: amygdala, anterior cingulate cortex (BA24), caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens basal ganglia, putamen basal ganglia, spinal cord cervical c-1, and substantia nigra. Seven out of the eight putative causal SNPs are eQTLs for a minimum of one gene in the brain. Fifty-two additional SNPs marked by the tagSNPs are also eQTLs for different genes in brain regions. We focused on methylation occurring in promoter regions, which is well established to alter gene expression. The eQTLs for *ARTN*, *C2orf82*, and *PIDD1* correlated with methylation of CpG sites lying in their possible promoter regions and presented opposite directions for methylation and gene expression levels.

It is well known that DNA methylation in promoter regions inversely correlates with the levels of gene expression[18], and the observed effects of all the ASM variants associated with ADHD in our study are in concordance with this statement. The ADHD risk alleles are associated with increased expressions of *ARTN* (in cerebellum and subcortical region), *PIDD1* (in cerebellum and cortex), and with a decreased expression of *C2orf82* (in cortical, subcortical, and cerebellar regions). The eQTL analysis helps to get insight on the functions that are altered in the disorder, as they connect the 'aseptic' genetic variants to actual genes.

(iii) Transcriptome imputation

A recently developed approach called transcriptome imputation integrates genotype data and publicly accessible expression data to predict altered gene expression in traits. Most of the predicted expression-trait associations by transcriptome imputation overlay the GWAS risk loci, so this method can capably identify potential causal genes within established risk loci[240] (Figure 16). We used a generalized framework called MetaXcan that can incorporate the results of multiple transcriptome-wide association studies (TWAS) and colocalization methods (eQTL and GWAS signals) to investigate the gene to phenotype relationship across more than 100 phenotypes with greater power and fewer false positives[241]. FUSION is another such software for transcriptome imputation.

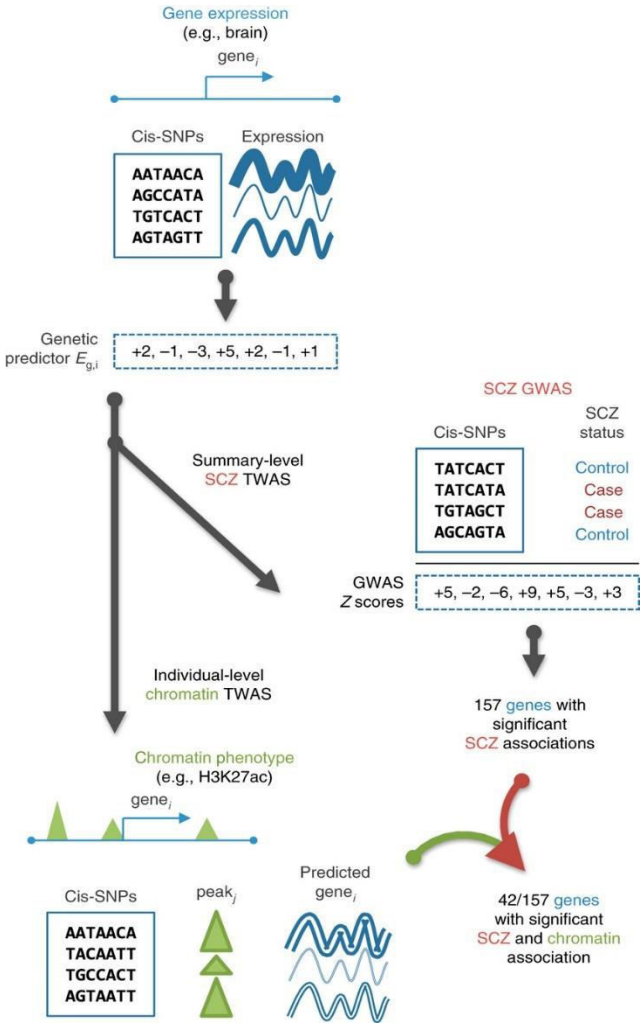


Figure 16. Schematic illustration of the TWAS approach in the case of Schizophrenia.

Adapted from Gusev *et al.*, 2018.

For the MetaXcan run, the input was the summary statistics of ADHD GWAS meta-analysis, and prediction models were trained with RNA-Seq data of 10 GTEx brain tissues and CommonMind dorsolateral prefrontal cortex. The SNP covariance matrices were generated

using the 1000 Genomes Project Phase 3 EUR genotypes of the prediction model SNPs. We imputed all the SNPs, together with the ASMs (to account for LD) located within ±1 Mb from the transcription start site (TSS) of each gene. The imputation results can help infer if the overall genetically determined expression of the genes (using the input SNPs) correlated with ADHD. We found significant associations between predicted expression levels of genes and ADHD. *ARTN* and *PIDD1* again showed statistically significant increased expression in three and four brain tissues respectively, while a decreased expression of *C2orf82* was observed in eleven brain tissues.

(iv) Influence of SNPs on subcortical brain structures

Structural MRI data has established that patients with ADHD have altered brains and the reported brain differences are independent of the symptom severity, comorbid disorders, or medication effects, and robustly related to the ADHD diagnosis itself. Accumbens, amygdala, caudate, hippocampus, and putamen are reported to have smaller volumes in ADHD patients. The largest effect was found in the amygdala and is of particular importance since this region links ADHD to problems in emotional regulation[242]. We therefore obtained the summary statistics of the GWAS meta-analysis of eight MRI volumetric measures of nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen, and thalamus produced by the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium[243]. We investigated in this summary statistics if the associated ASM-SNPs can as well influence the subcortical brain structures (Figure 17). We found that three of the putative causal SNPs that correlate with methylation of three different CpGs also correlate with volumes of nucleus accumbens, caudate nucleus and thalamus. Moreover, all the ASM-SNPs in the LD block for *C2orf82* nominally correlate with increased volumes of nucleus accumbens and caudate nucleus subcortical regions.

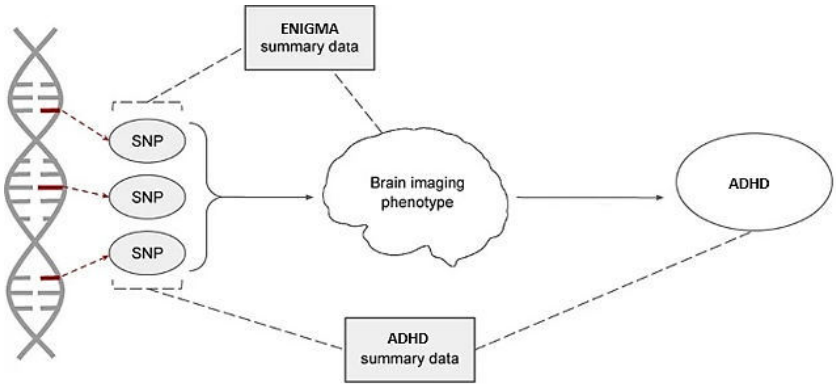


Figure 17. A model of integration of brain imaging data and GWAS summary data.

Adapted from Knuston and Pan, 2020.

The stated four-fold functional evidences (Figure 18) emphasize the candidacy of *ARTN* and *C2orf82* in ADHD development and lends more confidence for performing downstream functional analyses, e.g. by generating genetically modified animal models.

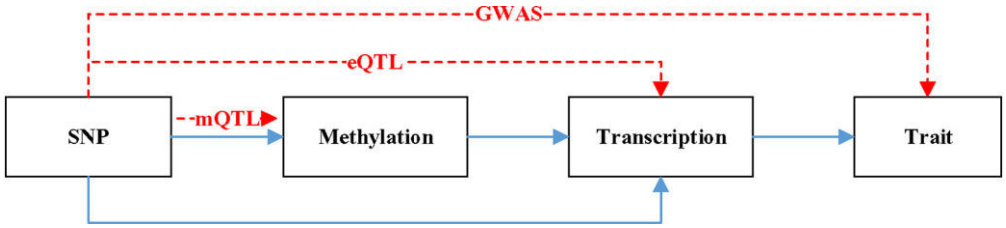


Figure 18. A model of integration of mQTL, eQTL and GWAS data.

Adapted from Zhao et al., 2019.

4.1.3 System 3: Protein-coding genes in cocaine-dependence

We conducted a GWAS meta-analysis of cocaine dependence using datasets from the dbGaP repository and identified 22 independent regions that contain at least one variant with a suggestive association (P < 1e-05). No genome-wide significant finding was identified in our study.

Results from GWAS do not directly translate into causal variants in general, as the majority of hits are within non-coding regions, and the LD present among the variants causes the effects to statistically spread out across multiple variants[244]. GWAS SNPs are enriched for functional annotations, with 81% of GWAS LD regions containing at least one functional SNP[245]. Earlier developed resources and tools did annotate SNPs only across coding regions, but newer methods feature also the non-coding regions. However, the interpretation of the extracted biological information from various available repositories is not always straightforward or error-free. Two methods that have been specifically developed for the analysis of GWAS data are FUMA GWAS (Functional Mapping and Annotation of Genome-Wide Association Studies) and INFERNO (INFERring the molecular mechanisms of Noncoding genetic variants), and they successfully integrate many forms of functional genomics annotations[236]. FUMA GWAS represents a statistical framework that functionally annotates GWAS findings and prioritizes the most likely causal SNPs and genes by accumulating positional, eQTL and chromatin interaction mappings from 18 of the publicly available datasets[246] (Figure 19). FUMA has successfully annotated GWAS variants associated with schizophrenia, depression and volumetric variations of human brain. INFERNO is another method to annotate GWAS

summary statistics by identifying nearby SNPs that are likely causal using similar datasets as FUMA (Figure 19).

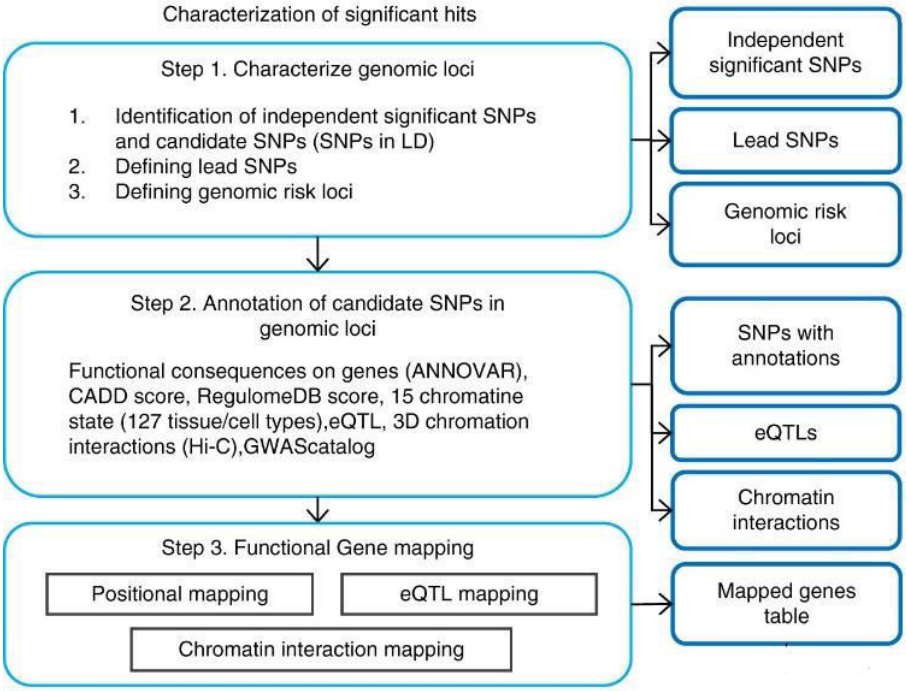


Figure 19. FUMA pipeline to annotate and prioritize SNPs and genes from GWAS summary statistics.

Adapted from Watanabe *et al.*, 2017.

To identify potentially interesting regions with FUMA, we considered SNPs that showed a suggestive level of association ($P < 1e\text{-}05$), in absence of any GWS hit ($P < 5e\text{-}08$) in our SNP-based analysis. We identified 23 lead SNPs which correspond to 22 genomic risk loci containing 112 genes. Interestingly, the risk locus located on chromosome 6 (6p22.1) contains the maximum number of suggestive associations and overlaps with a region associated with schizophrenia. This region is defined by two lead SNPs (rs806973, $P = 3.1e\text{-}06$ and rs56401801 with P=3.4e-06) and includes 77 genes and 458 nominally associated SNPs. The genomic region is highly enriched for genes that encode histones and proteins of the immune system, two functional groups known to be associated with psychiatric illnesses[247]. Moreover, most of the SNPs in this region (447) are brain eQTLs for at least one member of a small group of 12 genes, including *BTN3A2*, *HIST1H2AK*, *ZSCAN31*, *PRSS16* and *ZNF184*.

These functional results coupled with the genetic correlations between phenotypes point towards a shared genetic basis to the clinical co-occurrence of schizophrenia and cocaine dependence.

4.2 Challenges in functional mapping of associations

Our analysis on these candidate systems highlighted some gaps that persist in the omics knowledge and can mask the relevance of newly discovered associations. Below we review some of these obstacles.

4.2.1 Current deficiencies in functional annotations

GWASs have found that a majority of the associated SNPs do not lie within genes, which can imply that the non-coding regions are clinically pertinent too. International consortiums like ENCODE developed methods to annotate transcription factor binding sites, chromatin states, DNA methylation, RNA-protein interactions, and three-dimensional chromatin interactions (Figure 20). Additional programs like the NIH Roadmap characterized chromatin marks throughout the genome, and modENCODE performed ENCODE framework in model systems such as yeast, worms, and flies. This information now benefits tools like HaploReg and RegulomeDB that annotate variants for regulatory roles. Despite these continued efforts, genome-wide exploration of non-coding regions for their contribution to disease phenotypes still lacks insight into functional aspects. And merely a handful of studies have addressed how the genetic variation can influence the non-coding genome mediated regulation.

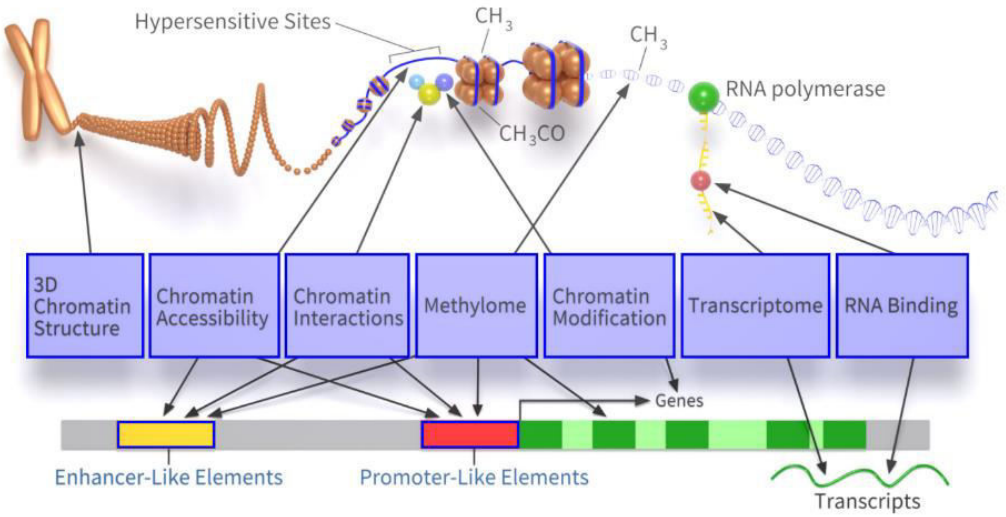


Figure 20. Data production by ENCODE project for the identification of functional elements. Adapted from encodeproject.org.

4.2.1.1 SNP-miRNA-mRNA connections or miRNA-eQTLs

The follow-up of any SNP-based association results includes the inspection of the influence of a SNP on the expression of nearby protein-coding genes (eQTL analysis). However, connections between SNPs and the expression of miRNAs is still in its infancy (Figure 21).

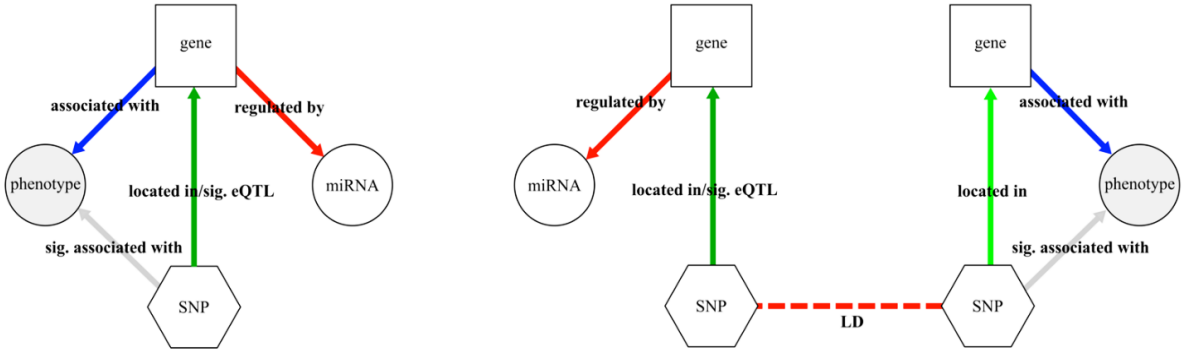


Figure 21. A model for miRNA eQTL association networks. sig.: Significantly associated

Adapted from Branco *et al.*, 2018.

Most studies have reported miRNA-eQTLs using whole blood, a straightforwardly available tissue. A whole blood miRNA-eQTL mapping discovered that *cis*-miRNA-eQTLs are enriched for *cis*-mRNA-eQTLs and regulatory SNPs and >50% of *cis*-miRNA-eQTLs are located upstream of mature/primary miRNAs. Notably, 11 mature miRNAs from intragenic miRNAs shared *cis* eQTLs with their host protein-coding genes, and numerous *cis*-miRNA-eQTLs were found associated with complex diseases/traits in GWAS[248]. A key question here is whether the genetic control of gene expression is similar in brain and blood tissues, and if whole blood is a useful 'proxy' for investigating brain eQTLs. An overlap has been demonstrated between blood and brain eQTLs from various studies[249], but it sometimes happens that the observed co-expression is due to the contamination of brain tissue with blood during extraction. Whole blood investigations may extend well to a set of genes that show tissue-independent expression unlike miRNAs that can be brain-specific[250]. The best approach remains to infer eQTLs in a tissue-specific manner wherever feasible[249,250].

Only recently, studies are emerging to fill this deficiency in the miRNA-eQTL literature by providing resources specific to brain tissue. For instance, in mouse brain, expression levels of 881 miRNAs and 1,416 genomic locations were studied to identify miRNA-eQTLs. Of the 38 significant miRNA-eQTLs identified, 10 miRNAs had target genes enriched for brain-related pathways and mapped to four miRNA-eQTL hotspots[251]. On the other hand, brain tissue from humans has not been the focus of any genome-wide miRNA-eQTL analyses despite the known

importance of miRNAs in brain-related diseases (Figure 22). This limits us in studying the impact of the identified variants on the *cis*-miRNA gene expression in brain.

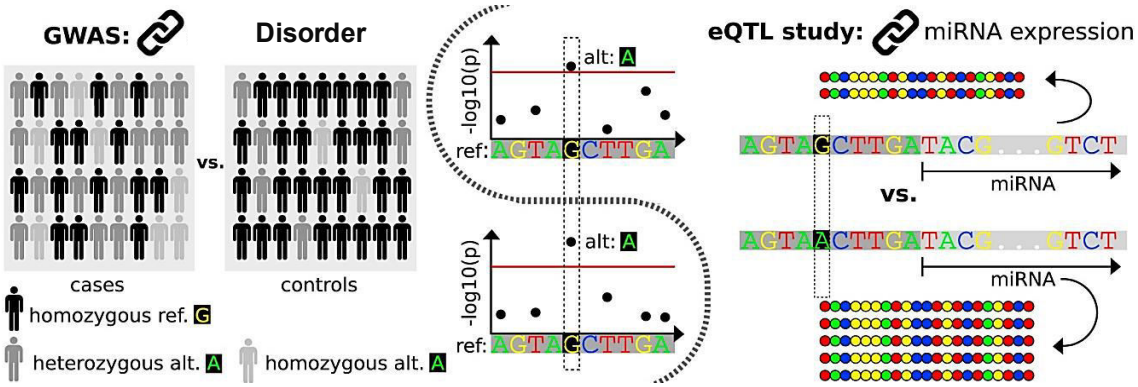


Figure 22. MiRNA-eQTL analyses across GWAS regions associated with a disorder. Adapted from Branco *et al.*, 2018.

4.2.1.2 Brain expression of miRNAs

There are now over 2700 miRNAs discovered in humans of which 75% are expressed in the brain[228]. Curiously, until now, merely a handful of miRNAs are expressed in a brain-specific or brain-enriched manner[252]. Fewer human brain samples have been mined for the expression levels of miRNAs relative to other more available tissue types, and most studies do not focus on all brain structures. Even with a paucity of data of brain expression of miRNAs, a number of brain-specific miRNAs have been linked with the shaping of human cognition and neuropsychiatric disorders so far[81].

Resources like human miRNA Tissue Atlas (https://ccb-web.cs.uni-saarland.de/tissueatlas/) have quantified the abundance of miRNAs in 61 tissue biopsies of different organs from two individuals. The project focused on ~2000 miRNAs, but it lacks expression data for the newly referenced miRNAs and for individuals showing disordered phenotypes. A more widespread resource like BrainSpan, Atlas of the Developing Human Brain (https://www.brainspan.org) provides annotations for over 16 structures across 13 developmental stages. Other resources like miRIAD or miRmine have also annotated a proportion of the miRNAs. However, there exist different annotations for the same miRNA across multiple platforms, and it is unclear which annotation is more reliable. For instance, miR-4655 and miR-5193 are brain-expressed according to BrainSpan, but not so by miRmine. The inconsistencies pose questions on the study methods used for data generation, and the need to have systematized protocols.

Many studies use microarray technology, which limits the number of miRNAs quantified and the discovery of novel miRNAs. To address this limitation, novel methods like small RNA-seq have been developed that uncovered 99 putative novel miRNAs from 93 post-mortem human prefrontal cortex samples[253]. This yields the possibility that there are brain-specific miRNAs yet to be discovered. Also, for cross-species miRNA comparability, similar pipelines are in progress that estimate miRNA expression in mouse across cell types within nervous system tissues[254].

4.2.1.3 MiRNA targets

MiRNA targets are identified by three general approaches: bioinformatic target prediction, biochemical isolation of miRNA/mRNA complexes, and transcriptomic/proteomic analysis. The essential basis of target binding used by bioinformatic methods is the 6-nucleotides long seed sequence of the miRNA to which mRNAs can bind. But complementary base-pairing rule applied to this small seed sequence usually yields a large number of target genes, many of which are likely to be false negatives. To improve the accuracy of target prediction, additional factors including sequence conservation, flanking sequence determinants, and compensatory pairing outside the seed region are incorporated by some tools (Figure 23A). miRNA targeting can occur anywhere along the entire mRNA (Figure 23B); however, many algorithms limit the predicted targets to mRNA 3′ UTRs as this area is assumed to be the most frequently targetable by miRNAs[255].

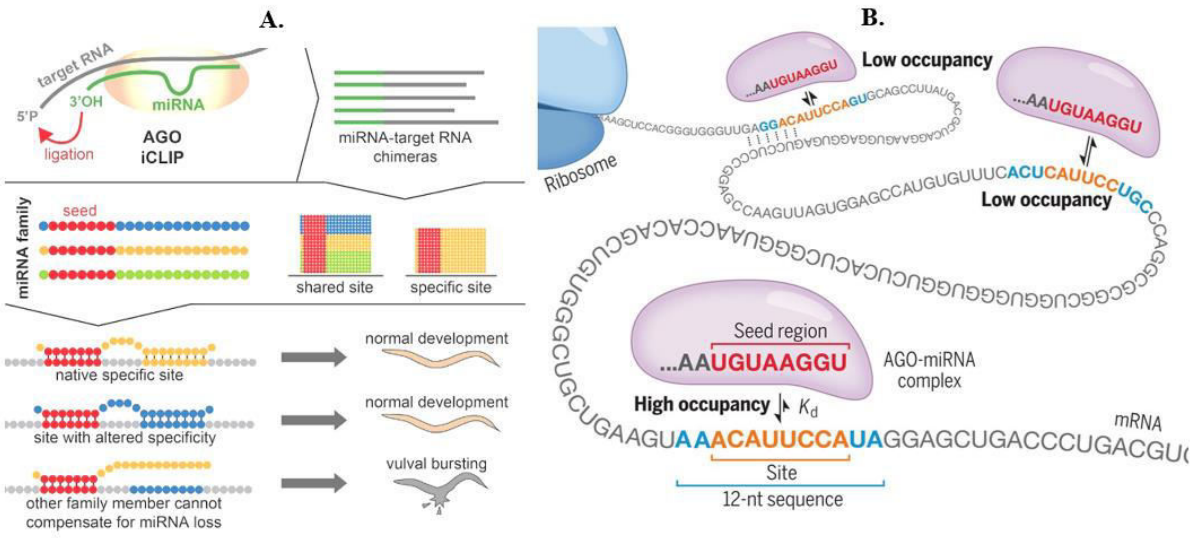


Figure 23. A: Additional base-pairing beyond the seed sequences with the 3' end of miRNA in miRNA targeting. B: miRNA-mRNA binding outside of 3'UTR of target mRNA.

Adapted from Broughton *et al.*, 2016, and McGeary *et al.*, 2019.

As it stands, a confident prediction of miRNA targets remains disguised given a lack of validation of the predictions. It will be particularly useful to integrate more datasets from the biochemical and transcriptomic approaches that isolate co-expressed miRNA-mRNAs and quantify the protein outputs, so as to refine the existing bioinformatic target prediction models.

4.2.1.4 MiRNA annotation in model species

Analysis of miRBase shows a large difference between the number of miRNAs referenced in human (2656 mature miRNAs, 1917 precursors), mouse (1978 mature miRNAs, 1234 precursors), and in rat (764 mature miRNAs, 496 precursors)[75]. Thus, it seems reasonable to state that a large percentage of the miRNAs expressed in the two model species have not been uncovered yet. Elucidation of miRNAs in the model species is a key factor in devising functional experiments for brain-related disorders, as a high percentage of miRNAs can be expressed in rat or mouse brain. For instance, 365 of 495 known rat miRNAs were found to be expressed in five CNS structures (Figure 24). In addition, 90 novel miRNAs that regulate the functions of neurons were discovered in rats with some of them having orthologs in mouse or human[256].

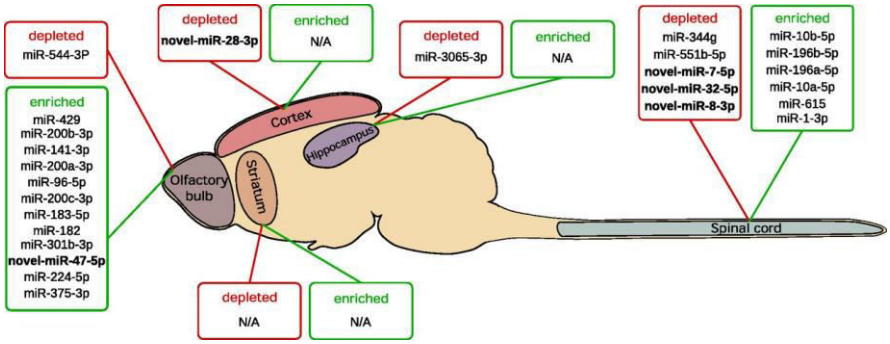


Figure 24. Enrichment/depletion of miRNAs in structures of rat CNS.

Adapted from Soula *et al.*, 2018.

4.2.1.5 Gender-specific annotations

Gender differences in the prevalence of psychiatric disorders, including ADHD and addiction, is among the established findings in psychiatry. Sex hormone regulated miRNAs have been implicated in diseases, including psychiatric, autoimmune or metabolic phenotypes. Initial attempts have been made to identify the sex-specific eQTLs (ss-eQTLs) and understand how the gender of an individual interacts with genotypes to produce different phenotypes. Co-localization of ss-eQTLs and variants that correlate with complex traits may imply a participation of the ss-eQTL loci in the appearance of sexually dimorphic traits. Further well-powered tissue-specific studies are needed to uncover sex-specific eQTLs and genes differentially expressed between the genders[257].

4.2.2 Lack of data integration in bioinformatics

The need to fill-in the missing epigenetics has accelerated the creation of novel resources. Unfortunately, the enigma expands when it comes to the versatility of these methods in annotating one's own data. Current methods like Ensembl Variant Effect Predictor (VEP), RegulomeDB and FunciSNP can annotate the effect of variants on miRNA and regulatory regions using only functional genomics information. There also exist more dedicated tools that determine the probability of a variant to interfere with miRNA-mediated gene regulation in view of base pairing, thermodynamics, sequence conservation, number of targets sites per transcript, and miRNA expression level.

Very often, a new tool is created to address a missing but complementary functionality rather than integrating it to the existing tool (Figure 25A). Hence, we see so many methods but not a gold standard tool for assured functional results (Figure 25B). For example, it has been reported that 52% of SNPs in the dbSNP could generate novel miRNA binding sites[258]. Even though bioinformatic tools allow for this estimation of the effect of a variant on the miRNA binding site, the overlap between prediction results can be very low and ranges from 5% to 70%[259].

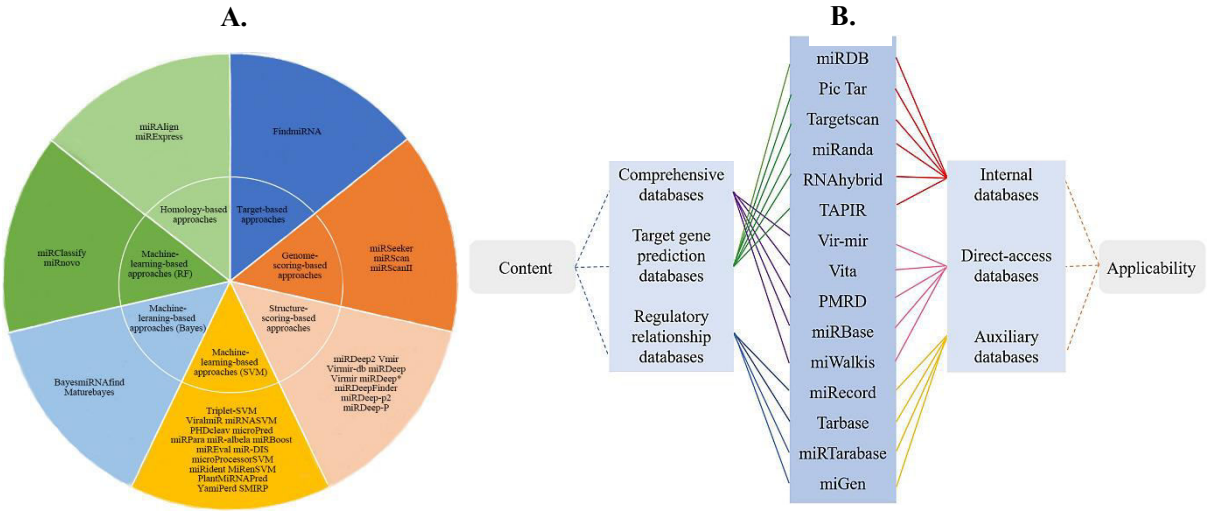


Figure 25. A: miRNA prediction methods and software. B: Broad classification of miRNA databases and applications for target prediction or regulatory networks construction.

Adapted from Yu *et al.*, 2020.

Thus, for miRNA target prediction, there is no consensual target prediction tool, and most of the predicted targets are admittedly false positives (Figure 26). These inconsistencies confound the selection of SNPs and miRNAs for functional testing. To resolve the redundancy and yield

more confidence in predictions, perhaps a practical solution would be to integrate the independent predictions and provide an overall confidence score to a SNP or miRNA target. While inferring any predictions, it should be kept in mind that all prediction-based methods employ training sets that incorporate functional annotations from databases. These methods are prone to biases present in the underlying annotations such as enrichments of variants near genes, gaps in functional annotations, or insufficient training data. Thus, all methods will miss functional elements that do not coincide with known annotation co-occurrence patterns[260]. The power of these predictions will increase with an increase in the experimental validation of putative causal SNPs and genes. Efforts are underway to generate reliable, curated functional information for mammalian miRNAs. For example, two new bioinformatic data sets deliver Gene Ontology annotations associated with over 500 miRNAs and over 2400 experimentally validated miRNA-target interactions[261].

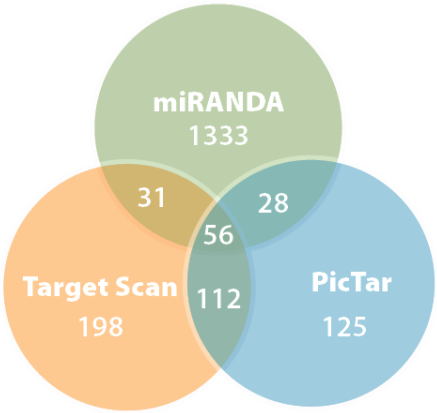


Figure 26. Overlap of results from different miRNA target prediction programs.

Adapted from info.abmgood.com.

We also observe that many variant annotation tools report the annotation anticipated from a single queried variant only, overlooking the variants in LD with the lead SNP (e.g., RegulomeDB, CADD/DANN, deltaSVM). To be able to accurately use these methods, one must use additional tools that return all SNPs in LD with the lead one (e.g. LDlink). Nonetheless, other methods like FunciSNP, HaploReg, GWAS3D integrate LD SNPs on their own[260].

Given these heterogeneous data sources, it is tempting to try out all available methods on our own data. The need to try more resources is even exacerbated when we must compare or reproduce the results from our study to the existing ones. It happens that a part of the available information originates from the specific needs of a certain group; hence it may not be suitable for obtaining a generic solution for varied questions. Not all tools are regularly enhanced, and it is not uncommon that many tools are quickly abandoned after publications. This causes more time being spent in trying new methods than analyzing data. Annotation resources require

continual updating with the exploding data growth. A few long-term well-maintained tools that allow users to run several annotation models in parallel, will be more effective than searching through an array of computational platforms (Figure 27).

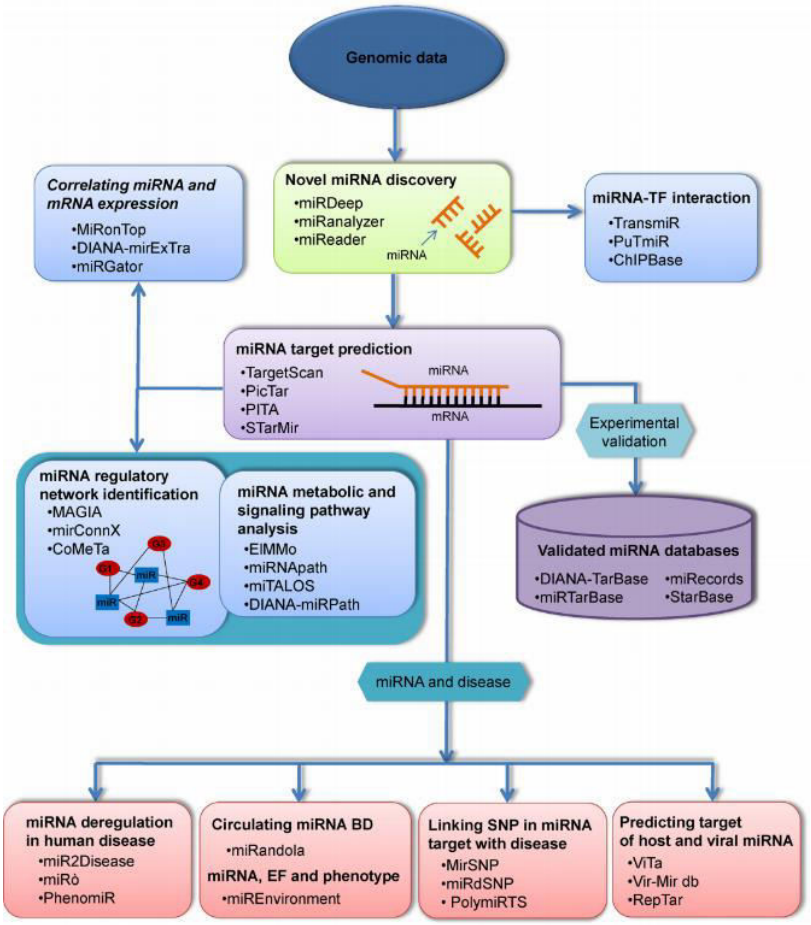


Figure 27. Bioinformatic tools classified according to their main functionality, and sample tools for each category.

Adapted from Akhtar *et al.*, 2016.

4.2.3 Methods to replicate the identified associations

The replication of association results is an important facet as (i) it provides reliable validation of the scientific discoveries, by confirming their true-positive status; and (ii) it allows for an assessment of any sources of bias in case of a lack of replication. One way to assess replication of association results is to use the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/), which includes associations with a suggestive statistical evidence ($P<10^{-5}$) from 3567 publications that describe a total of 71,693 associations for different traits[262]. In Figure 28, we observe that that most of the top hits included in the GWAS Catalog had already been reported in previous publications, and hence correspond to replications of known SNP-trait associations[263]. A positive association at variants in strong LD with the lead variants also evidences replication for the original marker-disease association.

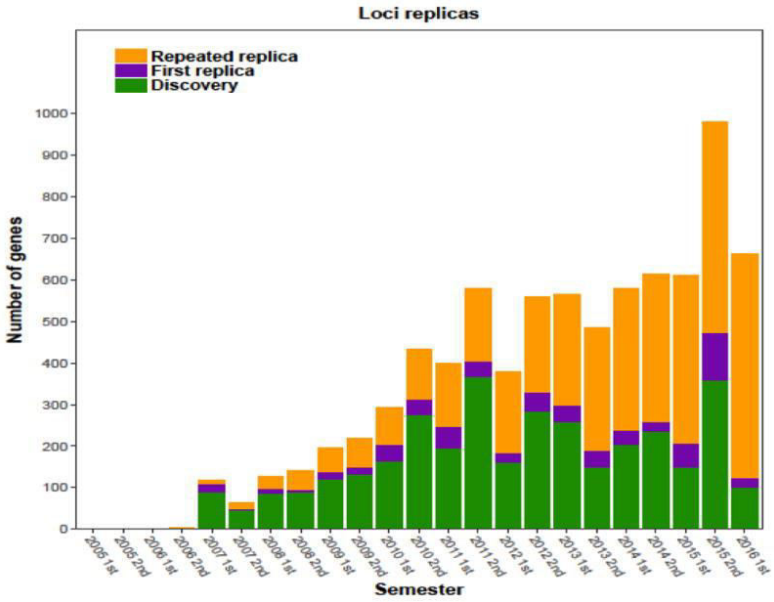


Figure 28. Discoveries of new loci and re-discovery of previously discovered loci for 60 diseases included in the GWAS Catalog with >15 discovered SNPs.

Adapted from Marigorta *et al.*, 2018.

Our association results from global analyses of miRNA loci is the first for ADHD. We observe that the NHGRI-EBI GWAS catalog till date documents only ~180 miRNA genes for association with miscellaneous traits. This reveals the lack of case-control genome-wide association analyses on miRNA genes which deters any attempts for replication for this system. Additional genome-wide ADHD studies will remain crucial to establish the patterns of emergence of associations.

For cocaine-dependence, while many case-control association studies have been performed on candidate genes, only a few risk variants have been identified and replicated; such as rs16969968 in the *CHRNA5* gene, encoding the cholinergic receptor nicotinic alpha 5 subunit, and rs806368 in *CNR1*, coding for the cannabinoid receptor 1[264]. The only existing GWAS of cocaine-dependence has been performed on the African-American (AA) and European-American subjects (EA)[265]. The strongest (and GWS) finding in the combined AA and EA sample using the Sympcount model is an intronic SNP rs2629540 at the *FAM53B* locus on chromosome 10. Although the association signal was stronger in AAs, it was supported in both populations. This hit could not be replicated in a subsequent study with a Spanish Caucasian sample of substance dependence where 59% of the subjects were cocaine-dependent[266].

Similarly, three GWAS of illicit drug use collectively reported four significant associations: two in EA population and one each in EA and AA populations[213,267,268]. One of these significant associations (rs2952621) was replicated under the dominant model, the risk allele (T) being the same as identified in its originating study[266]. The SNP is located near an uncharacterized gene

and a long intergenic non-protein coding RNA gene, which makes it difficult to determine its role in the predisposition to dependence.

The lack of replication of results between studies can be attributed to multiple factors, such as the sample features (ethnicity, sample size, delineation of cases and controls) and association workflow (control for covariates, multiple testing, genetic model, insufficient number of genetic markers). A factor quite likely to interfere with the replicability of associations from our cocaine dependence GWAS with the GWAS by Gelernter *et al.* would be the experimental design. (i) In the previous GWAS, the AA and EA samples were combined, and the former showed a greater weight in the association signal. Our GWAS design outlined individuals of European ancestry only. (ii) Additionally, the proportion of cases and controls in the previous study was skewed, with cases being four times the controls. We therefore increased the number of controls in our analysis to achieve a balance of cases and controls. (iii) The controls used by Gelernter *et al.*, were individuals that had been exposed to cocaine, at least once in their lives, but had not developed addictions. In contrast, we considered controls from the general population, irrespective of their drug exposure status.

While replication can effectively benchmark the disease associations for follow-up experiments, the replication constraint may cloud the true genetic effects especially when samples originate from distinct geographic or ethnic base[269]. The patterns of replicability of variants or lack of, thereof, can be a function of disease heterogeneity and can provide clues into it[263]. For instance, racial differences can cause large heterogeneity in odds ratios in a way that an allele exerting protective effect in some samples can turn out to be a risk factor in others[269]. Thus, current practice of replication of a GWAS result can be thought of as the replication of a specific statistical design[270].

The essential unit of replication for a GWAS should be the genomic region within which the variants in strong LD with the lead signal must undergo assessment[270]. Subsequently, evidences surfacing from gene-specific and pathway-specific information should be included for the genetic associations in question (Figure 29)[269]. This is valuable since effects from multiple variants can aggregate to reach a certain threshold at which the symptoms for a disorder will appear. All in all, digging into the functional hierarchy will indeed be more insightful into the mechanisms of disorder progression than achieving a statistical model dependent variant replication.

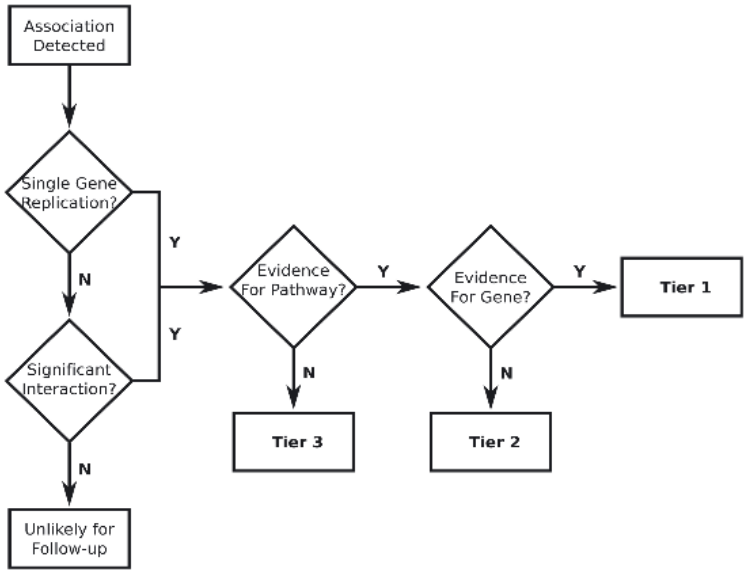


Figure 29. Methods to divide candidate SNPs into tiers for evaluation based on statistical results and biological information.

Adapted from Greene *et al.*, 2009.

Lookup

The detailed association analyses substantiate the involvement of the three systems in two psychiatric phenotypes investigated for. An important aim of this work is to understand if and how the epigenetic modules can be successfully bridged to genetics by means of the available resources. As shown in the case of ADHD-associated ASM SNPs, epigenetics research can be key to understanding the means of gene dysregulation in disorders. The work also underlines the importance of structural and functional evidences from human brain tissues in connecting the sequence variation to the brain alterations established in psychiatric traits. We also inform the methods and obstacles for genetic and epigenetic association analyses. By and large, the foremost direction for any association study is a reliable selection of the most promising signals. Bioinformatics backed by the functional genomics data integration remains the leading approach for the prioritization process. Both coding and non-coding variation can now be effectively mapped for alterations in the protein-coding genome. That said, in the pipeline is the regulatory genome. Challenges in evaluating the effects of genomic variation on regulatory gene expression are manifold. Present day research is at the point of attempting to untangle the complexities, however it necessitates more bioinformatics development. The inflation in computational tools and data has not been synonymous with the presence of meaning as yet. We reckon that more high-throughput omics will expand our knowledge of the noncoding genomic landscape which can help overcome the obstacles in bioinformatic analyses and will empower further decryption of the (epi)genetic control in psychiatric phenotypes.

CHAPTER 5. NEW WAYS FORWARD

The fundamentals of psychiatric disorders are now known more than ever before. However, there is still a long way to go. The exciting discoveries so far have put more research questions on the agenda, with the final aim of designing interventions that can minimize the damaging effects of these disorders. I hereby list some of the prospects in psychiatry research for the upcoming years.

Objective 1. Accelerating epigenetic research and its integration with genetics

Objective 2. Finding more common variants and also rare

Objective 3. Prioritizing likely causal genes for functional follow up

Objective 4. Establishing the genetic architecture of the disorders

Objective 5. Linking genetics to intermediate phenotypes

Objective 6. Understanding genetic pleiotropy and unraveling causal relationships among traits

Objective 7. Deep phenotyping and big data projects

Objective 8. Modelling psychiatric disorders

Objective 9. Psychiatric therapy and precision medicine

Objective 1. Accelerating epigenetic research and its integration with genetics

The last couple of decades have established that non-coding genome is not equivalent to a non-functional genome. DNA methylation, histone modifications, non-coding RNAs, and others regulate the expression of genes. Psychiatric epigenetic research needs to be pursued driven by the evidences that epigenetic mechanisms fine-tune physiological processes in real-time for cell's response to intrinsic and/or extrinsic conditions (Figure 30)[271]. Studies on model organisms emphasize that some genes, more than others are susceptible to environmental modulations and that this may confer evolutionary advantages[272].

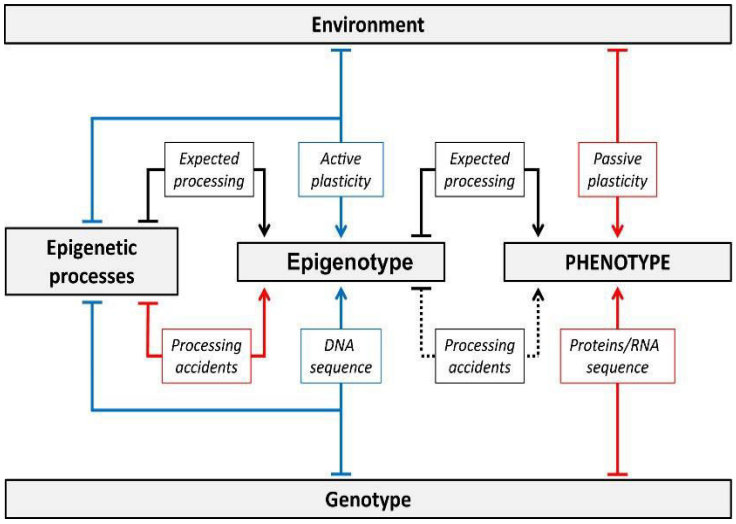


Figure 30. Sources of epigenetic and phenotypic variation. The formation of any phenotype results from a series of processes starting with gene expression.

Adapted from Angers *et al.*, 2020.

In order for the epigenetic data to be useful, accounting for several factors like tissue-specificity, age, gender, comorbidities, medication, environmental events and multiple time-points is needed because of the dynamic nature of epigenetic processes. Genome-wide epigenetics need efforts from large collaborative consortium projects. In the absence of organized efforts, most epigenetic study designs are likely to be case-control, cross-sectional and can be small with insufficient statistical power, and epigenotyping costs can as well be a limiting factor. To control for possible data inconsistencies and enable comparability in data integration, consensus designs with stringent controls should be adopted by individual studies[273]. It is arguable that the epigenetic variation may appear post the development of a trait and may not be the causal to the disorder. However, the value of biomarkers lies not only in predicting the source of disease, but also in the ability to track disease progression and the treatment[274]. Moreover, the peripheral epigenetic biomarkers can be afterwards linked to the changes in brain using *postmortem* brain samples. The PsychENCODE initiative

(www.psychencode.org) provides both depth and breadth of brain-omics data from a large number of individuals with SCZ, ASD, and bipolar disorder while collecting multifarious types of data from the same individuals and can be beneficial to prioritize epigenetic factors that induce brain-specific outcomes[273].

A handful of studies have endeavored integration of epigenetic and genetic methods to trace altered genetic regulation. As an example, integrating epigenetic, genetic and gene-set enrichment analyses allowed to connect *NOTCH* to PTSD's etiology, which would not have been possible through a mere candidate gene association study[275]. mQTLs form another useful tool to refine functional/regulatory GWAS loci as both adult and fetal brains mQTLs are found to be enriched amongst schizophrenia-associated risk loci[276–278]. The method can as well be extended to other disorders including ADHD and cocaine-dependence whose risk loci should be evaluated for the presence of mQTLs.

Objective 2. Finding more common variants and also rare

GWASs in ADHD, SUDs and other less studied disorders will expand in terms of sample size, and additional GWS loci will be identified as seen with major psychiatric disorders. However, cracking any psychiatric disorder will require aggregation of sub-threshold common variants in addition to the GWS loci, that can explain a much larger portion of the genetic variation. For example - in the case of schizophrenia, hundreds or thousands of sub-threshold variants with low effect sizes are now believed to confer risk. Take the case of the first 12 discovered independent GWS loci for ADHD, where only one gene - *FOXP2* - emerged that had previously been implicated in ADHD and in a phenotype, severe speech and language problems, observed in ADHD individuals[52]. Surprisingly, none of the GWS loci contained any candidates clearly belonging to gene systems like dopaminergic neurotransmission, classically implicated in ADHD. It is possible that larger GWAS are needed to implicate some of the classical candidate genes, as the variants located within these genes can exert small effect sizes for which the current sample sizes remain underpowered; or the other possibility is that the variants from dopaminergic genes aggregate in the group of sub-threshold variants, if not finally emerging as GWS loci (Figure 31). But even with the current absence of previous candidate genes in genome-wide findings, these remain valuable as targets of the most common pharmacological treatments for ADHD. Thus, classical candidate genes might still be relevant to the neurobiology of a disorder[279]. The etiology of ADHD and SUDs is likely to unfold with uncovering of more common variation.

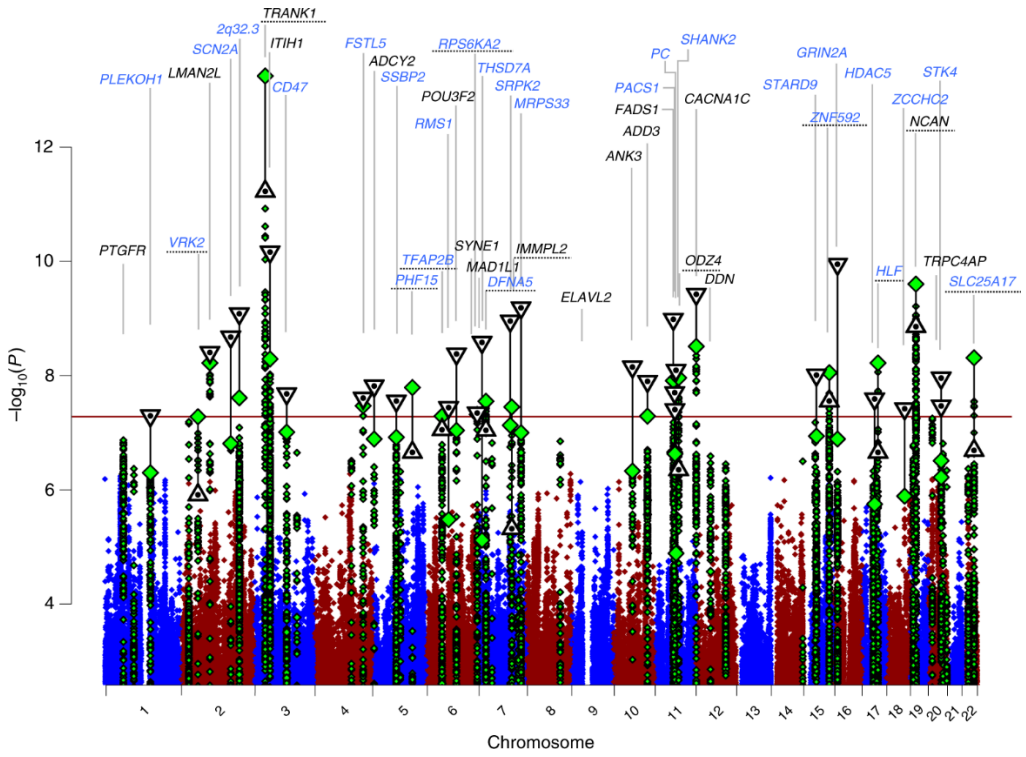


Figure 31. BD GWAS have reported a total of 20 GWS loci. 12 of the previously reported loci are not GWS in this GWAS meta-analysis, but all have P ≤ 1.3 × 10−5. Labels correspond to gene symbols previously reported for published loci (black) and the nearest genes for newly identified loci (blue)

Adapted from Stahl *et al.*, 2019.

Very importantly, it has now been realized that individuals with ADHD and individuals with ASD share a similar burden of rare protein-truncating variants in evolutionarily constrained genes[271]. However, investigation of rare mutations in ADHD trail behind ASD, even though they are highly comorbid conditions. The role of ultra-rare deleterious variants too was limited to autism and schizophrenia, and only recently, highly evolutionary conserved gene sets have been found to carry ultra-rare deleterious variants in ADHD[280]. Similarly, the extensive genetic pleiotropy seen across psychiatric disorders includes CNVs. Thus genomic regions spanned by CNVs associated with an increased risk of ADHD are also associated with autism and schizophrenia[279]. Prioritization of CNV genes for ADHD by an integration of CNV studies spotted *POLR3C* and *RBFOX1* of a several hundred high-priority ADHD candidates, and the two genes also map within ADHD-GWS regions[281]. Hence, this study establishes a possible convergence of rare and common variants in ADHD.

Objective 3. Prioritizing likely causal genes for functional follow up

Traditionally, GWASs have provided variants that merely flag genomic regions without necessarily explaining the connection of the variant/loci to underlying biological mechanisms.

The organization, difficulties and costs associated with conducting functional studies, makes it important to prioritize the likely causal genes. As the sample size of GWASs has increased, more genes have been identified with high confidence. Which genes of all should be pursued for follow-up experiments rely on numerous factors, some of which being the replication status of the gene, involvement in other brain disorders, possible effect of the genetic change on the protein or on the regulation of gene expression, overlap with pathogenic CNVs or other variants, the effect size of the genetic variant, expression of the gene in brain tissues, being part of a disease-implicated protein network, or conservation across species, among others. The success of GWAS has conventionally been weighted by the number of genes discovered for a phenotype, however this is simply one aspect of what the method has to offer. Even a single GWAS finding that can lead to an effectual treatment is rationally an indubitable success[282].

One way to test the possible causality of genes within highlighted loci is to apply the Mendelian Randomization (MR) framework to integrate evidences from GWAS with eQTL data. The method allows to unravel the causal relations isolated from the effects of confounders and modifiable risk factors (Figure 32), such as the environmental influences or stress that impact the manifestation or severity of nearly all psychiatric disorders. MR is less likely to be affected by confounding bias or reverse causation present in the traditional observational studies. However, an effective MR requires well assessed assumptions and an adequate number of genetic markers linked with the exposure to increase the strength of the detection of causal relationship[157]. A recent bioinformatic tool called Integrated MEntal-disorder GEnome Score (iMEGES) can employ personal genomes (whole-genome sequencing data of individual patients) for the prioritization of variants and genes that influence each patient's susceptibility to mental disorders (Figure 33) in a patient-specific manner[283].

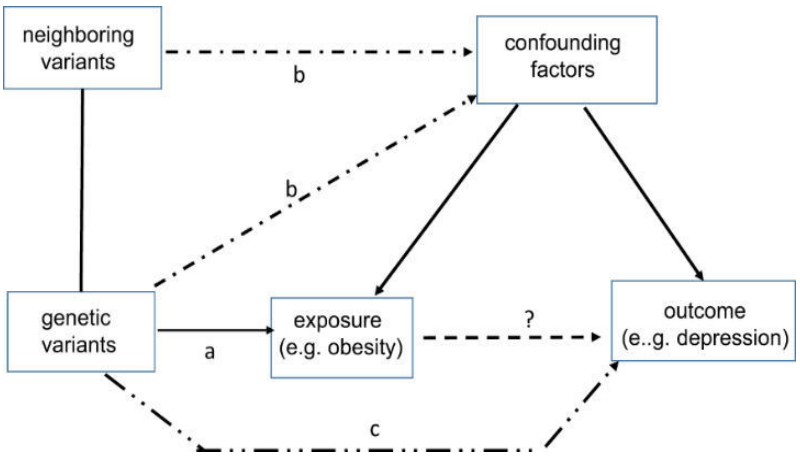


Figure 32. Assumptions in Mendel Randomization. (a) Robust association between genetic variant and exposure, (b) absence of (direct/indirect) association between genetic variant and confounding factors and (c) absence of other pathways between genetic variants.

Adapted from Verduijn *et al.*, 2010.

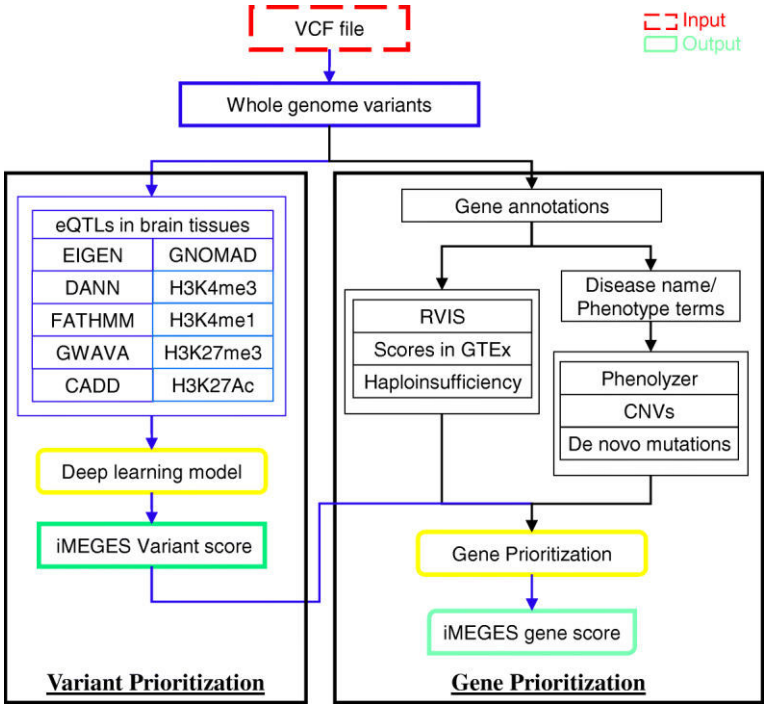


Figure 33. Schematic overview of iMEGES.

Adapted from Khan *et al.*, 2018.

Objective 4. Establishing the genetic architecture of the disorders

The common disease–common variant (CDCV) model (<100 of common variants with small-to- moderate effects) has been the foundation of most GWAS. But of late, alternate models of genetic architecture for psychiatric traits have been projected that accommodate an increasing number of discovered common variation and that also reserve a place for contributions from rare variation and environmental effects (Figure 34).

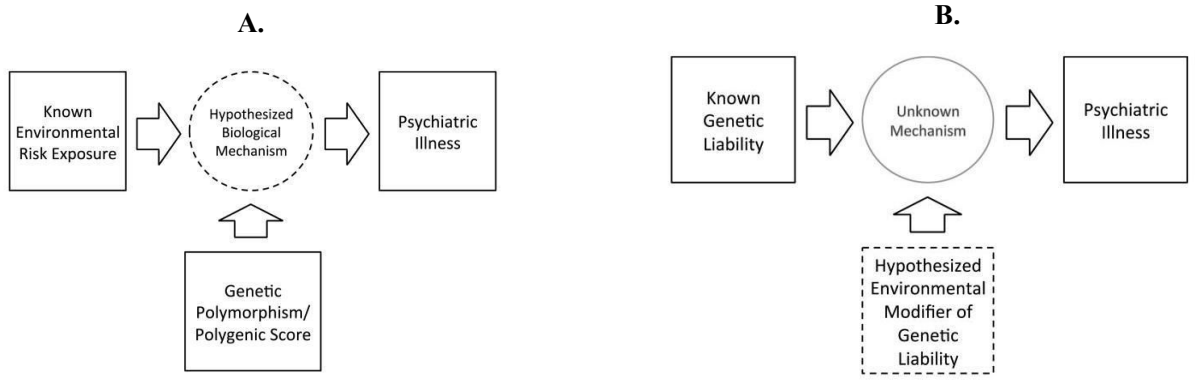


Figure 34. A: Type 1 GxE questions about the biology through which an environmental exposure contributes to the psychiatric illness. B: Type 2 GxE questions about the environmental conditions under which a genetic liability to a psychiatric illness is discerned.

Adapted from Belsky *et al.*, 2015.

These newer models to explain heritability are (i) the *infinitesimal* model – a large number (»100) of small-effect common variants operate, (ii) the *rare allele* model - large number of rare variants, including CNVs with relatively large effects act, (iii) the *broad-sense heritability* model - besides the additive effects of common variants, heritability is due to rare variants, non-additive GxG (dominance, epistasis) and GxE interactions as well as epigenetic effects, and (iv) the *omnigenic* model (Figure 35) - the genetic architecture of complex traits is exemplified by a huge number of peripheral, more general genes and a lesser number of "core" or disease-specific genes[157].

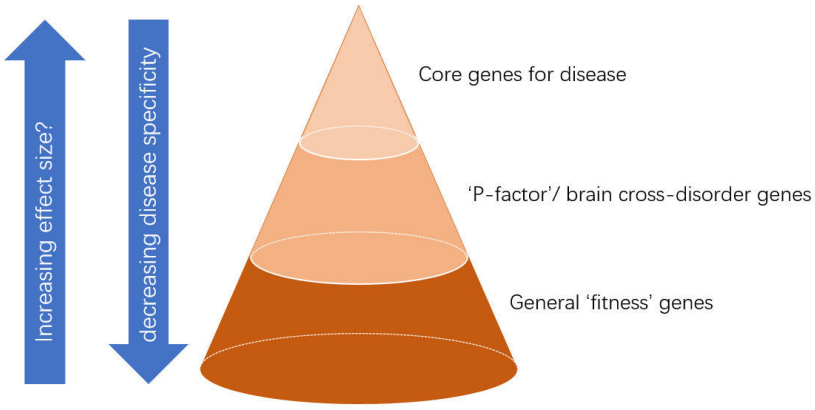


Figure 35. Omnigenic model. Adapted from Franke *et al.*, 2017.

Any psychiatric disorder, irrespective of the underlying genetic architecture will rarely manifest in isolation and defining the complete genetic architecture therefore rests on how complete the comorbid spectrum is in the study designs. ADHD is strongly comorbid with ID and with lower IQ, and shared heritability explains much of the correlation of comorbid conditions with ADHD symptomatology. Unfortunately, studies on ADHD do not share a unanimous criterion for the cut-off of IQs in cases. While some studies include IQs of 70 and above, others cut-off is an IQ of 80, and no IQ is considered in others. Despite the known fact that individuals with ADHD are more likely to have lower IQs, studies on ADHD often exclude participants based on lower than average IQ (i.e., between 70 and 85). Exclusion of ADHD participants with lower IQ limits the cognitive spectrum being studied and thus the applicability of the results. At the same time it masks the true genetic relationship of comorbidities and can lead to ineffective treatment or have negative repercussions[284]. It should be also be noted that presence of ID has not been a condition for exclusion in the case studies on autism and schizophrenia even though ID can co-occur with both disorders[279]. Only recently, DSM-V established that ASD is no longer an exclusion criterion in ADHD studies, in contrast to DSM-

IV[285]. Research specifically focused on comorbid conditions like ADHD, ASD, ID and lower IQ and ADHD and SUDs should be attempted to resolve the underlying architecture with a greater power arising from phenotypic and genetic similarities among psychiatric disorders[285].

The eventual goal of delineating the genetic architecture of complex disorders is to understand how the discoveries from large populations will contribute to illness in an individual. Correct genetic inferences need target datasets that are from different ancestries. A critical limitation of psychiatric genetics is the lack of inclusion of population diversity. The majority of GWASs have been conducted on individuals of European ancestry or African-American ancestry. Only a small proportion of studies have employed Asian populations like Han Chinese, Japanese, Korean, Indian, Pakistani, or populations originating from Africa or South America. From a genetics point of view, including lesser studied ancestral groups in GWAS (transethnic GWAS) can help narrow blocks of LD and fine-map the genetic architecture of disorders[157,273].

Objective 5: Linking genetics to intermediate phenotypes

The trajectory of manifestation of a psychiatric phenotype involves a number of intermediate disruptions of neural circuits starting from the genetic variation[286]. Thus, the intermediate traits or phenotypes are scattered between genetic variation and the end phenotype (Figure 36). Personality traits, brain activity and brain structural variation are some known intermediate phenotypes in psychiatric genetics[287]. For instance - motion tracking-based hyperactivity factor and a reduced eye movement control can be candidate intermediate phenotypes in ADHD, as the inability to sustain attention or ocular fixation will lead to a reduction in optimal performance in everyday cognitive and behavioral activities[288,289]. Altered brain structure with decreased gray matter volumes in the right inferior frontal gyrus and an increased white matter volume in the posterior right inferior fronto-occipital fasciculus are observed in ADHD individuals[290]. Exploratory locomotion, a behavioral phenotype is a predictor of vulnerability to addiction, and inversely correlates with spontaneous anxiety and depression-like behaviors[291].

Each intermediate phenotype is structured by relatively fewer risk alleles than the global phenotype[286]. Thus, they form a more precise connection to the underlying genetics than to the disorder and are attractive for gene discovery[273,287]. Although not clearly evidenced, the genetic variants can exert greater effect sizes on intermediate phenotypes than on the disorder traits, so GWASs of intermediate phenotypes could perhaps be more powerful in portraying the genetics of the disorder. Somehow, the idea is to divide the phenotype in pieces that are more heritable,

and investigate each of them separately, to end up by joining the pieces again. This would be important as the pattern of risk alleles that shape the overall disease phenotype is turning out to be far more complex than initially anticipated.

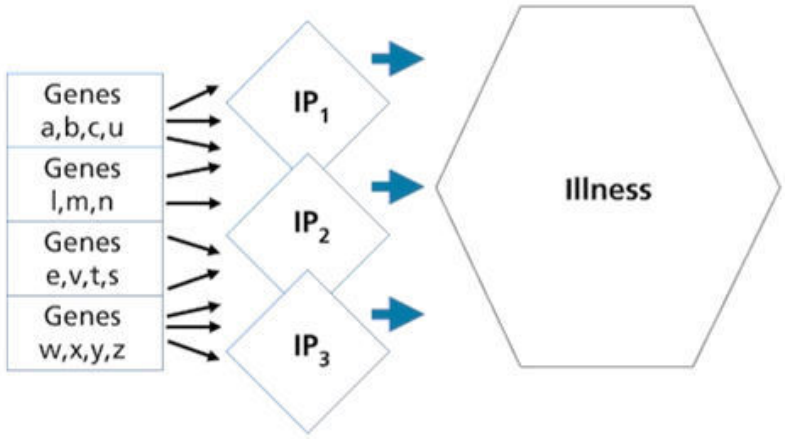


Figure 36. The relationship between genes, intermediate phenotypes and psychiatric illness.

Adapted from Preston and Weinberger, 2005.

Neuroimaging can reveal neuronal mechanisms that underlie emotion, reward, and craving that drive both ADHD and cocaine addiction. For example - positive reward in reward circuits is accompanied by an activation of ventral striatum and other brain areas that can be seen through neuroimaging techniques. Emotional and stressful stimuli images activate amygdala mediated by the serotonin transporter SLC6A4 and monoamine oxidase (MAOA). Similarly, PFC is activated during cognitive tasks and the impairment of PFC is linked to catechol-O-methyl transferase (*COMT*) and *MAOA* variation. It is already known that the biology of addiction, anxiety, impulsivity and reward is influenced by SLC6A4, MAOA and COMT (that metabolizes dopamine, norepinephrine, and catecholamines). The role of genetic variation identified in these genes might fit in better for one of these intermediate phenotype than the end symptomalogy[292].

As with shared genetic factors, intermediate phenotypes can as well be shared across comorbid conditions. For instance, stress resiliency and externalizing behaviors of disinhibition, aggression, and impulsivity, are thought to underlie the comorbidity between addictions and other psychiatric diseases[293]. An interesting connection can be made in the co-occurrence of cocaine-dependence and ADHD as different routes of cocaine administration can lead to equally different neurocognitive impairment profiles. The smoked cocaine dependence group (in contrast to insufflated cocaine dependence) was specifically associated with deficits in attention and executive functions. The differential profiles may not (only) be due direct effects of cocaine but also because of cognitive and biological differences in key executive functioning

and attention areas[294]. Externalizing problems can also dictate both intermediate and independent phenotypes in ADHD and ASD[295]. Attentional control is also a likely intermediate phenotype shared by both autism and ADHD, where atypical development of attentional control restricts adaptive functioning in later stages like education, life and social skills in an autism and ADHD-like manner[296].

In psychiatric illnesses, the nature of human brain can confound the discovery for genetic origins as individual neurons can display distinct transcriptomes and phenotypes. The genetic variants that appear weak at the full-disease level may be highly robust for the intermediate level. Thus, the variants that impact the brain function and structure should be followed up to reveal the underlying neurobiological mechanisms even if those are not the causative genetic factors[286].

Objective 6. Understanding genetic pleiotropy and unraveling causal relationships among traits

Pleiotropy and redundancy are key aspects of the biological pathways that underlie psychiatric disorders (Figure 37). A genetic variant can influence two or more phenotypes through the phenomenon of pleiotropy, and the same function can be controlled by multiple systems through redundancy. The genetic overlap in psychiatric disorders is not a novel notion anymore and is a rule rather than the exception[279,297].

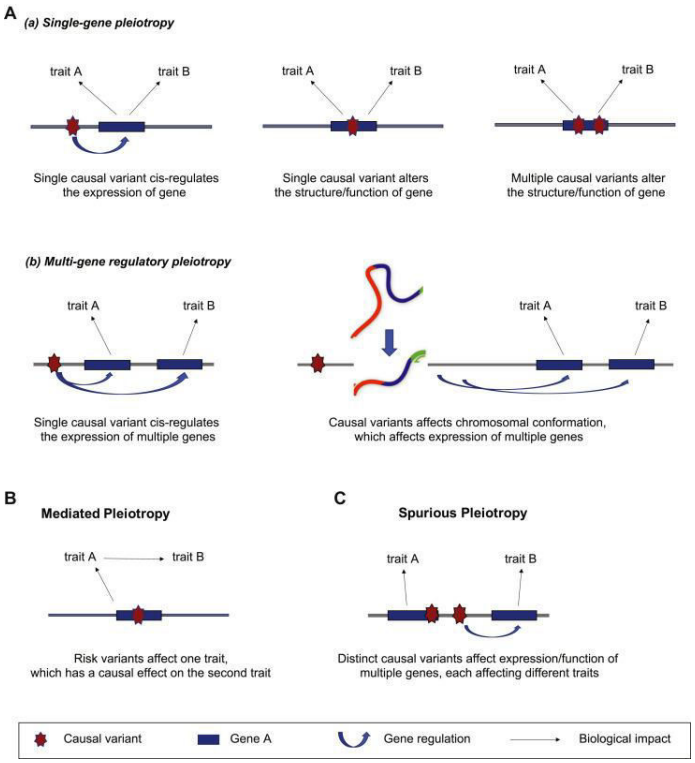


Figure 37. Pleiotropic mechanisms underlying cross-phenotype associations.

Adapted from *Lee et al.*, 2020.

Genetic risk variants discovered for specific conditions may also have broader pleiotropic effects; for example, genetic risk variants for ADHD reflect a genetic liability toward broad childhood psychopathology in the general population[298]. Finding genes that can trigger more than one disorder might enlighten the shared pathogenesis among different psychiatric disorders.

Will alterations in some genes then make us more susceptible to multiple disorders? In a study across eight disorders, 109 pleiotropic loci where found to contribute to at least two psychiatric disorders, and 23 loci were pleiotropic for four disorders or more. One of most highly pleiotropic loci related to all eight disorders corresponded to the *DCC* gene, a master regulator of white matter projections that mediates axonal guidance during neuronal development and the adolescent expansion of mesocorticolimbic dopaminergic connections to prefrontal cortex (Figure 38)[299]. Interestingly, this work also showed that certain loci can exert antagonistic effects on multiple disorders, which can partially explain that psychiatric drugs designed to treat one condition might worsen another inadvertently. Indeed, in psychiatric genetics, it is difficult to find genetic variants that would confer risk to solely one phenotype. To be able to

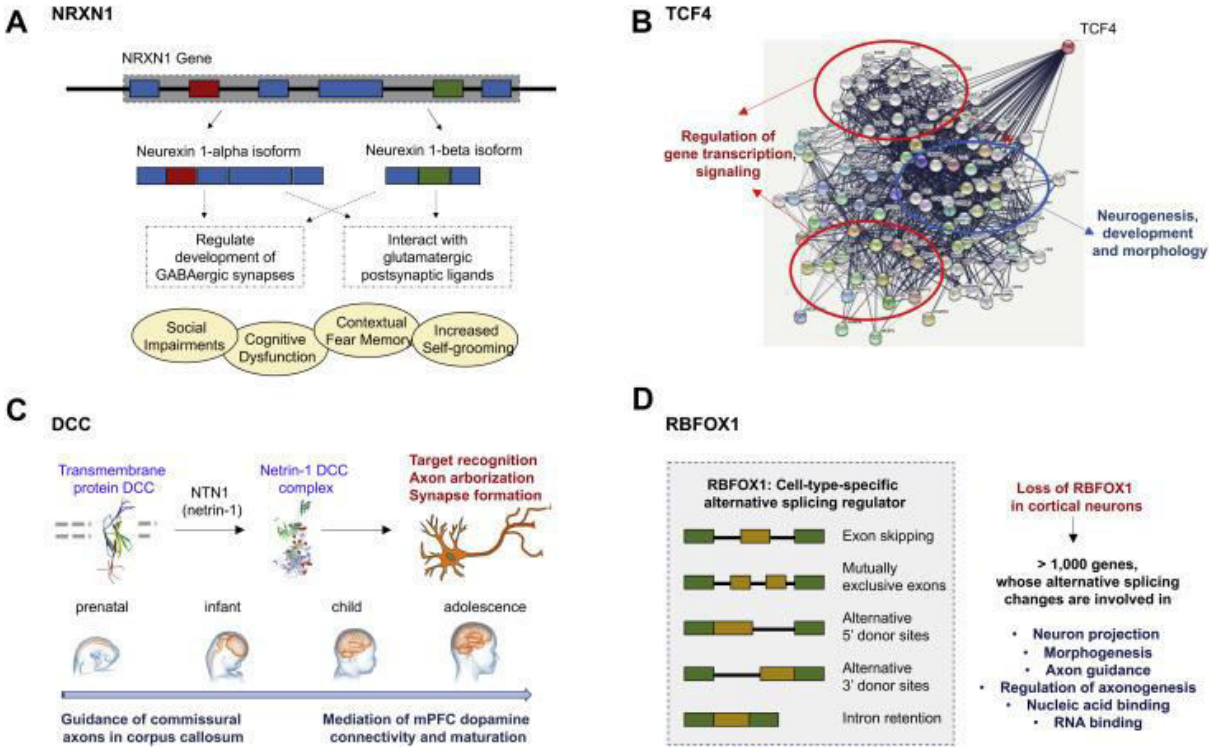


Figure 38. Examples of single-gene pleiotropy associated with psychiatric disorders. (A) Cell adhesion protein NRXN1. (B) Transcription factor TCF4. (C) DCC, a master regulator that governs axon guidance during early neurodevelopment and mediation of media.

Adapted from *Lee et al.*, 2020.

tell apart the horizontal (a genetic variation can affect two or more phenotypes independently) and vertical pleiotropy (when genetic variation affects one phenotype, the expression of which mediates a second phenotype) would have implications on causal theories of psychopathologies[300]. Future analysis should embrace subjects with ADHD only (if any), with ADHD and other highly comorbid conditions like ID or SUDs, and with the comorbid traits only. It is important that the number of subjects recruited for each phenotype is proportional, as a lower number of subjects from lesser studied disorders might mask true pleiotropic effects.

In addition to the shared genetic effects among psychiatric disorders, pervasive pleiotropy exists between psychiatric and immune-mediated disorders. Pleiotropic loci are known to cluster into histone methylation, synaptic biology, immune and neurotrophic pathways, neurodevelopment, glutamate receptor signalling and voltage-gated calcium channel signalling[301]. In fact, the chromosome 6 locus identified in our work that exhibits pleiotropic effects for cocaine dependence and schizophrenia is enriched in immune function and histone-related genes. Previously, the MHC region on chromosome 6 emerged important, along with regions like cytoband 1p13.2 in mediating pleiotropic effects on two systems[302]. How psychiatric disorders and immune system disorders are related genetically is yet another intriguing dimension for exploration.

Pleiotropy can further transform our view of individual psychiatric disorders into interrelated components of a syndrome. The present diagnostic categories can be subtypes under one umbrella term. Autism was once considered as childhood schizophrenia, until the 1970s. If pleiotropic mechanisms underlie the manifestation of comorbid conditions, can the same genomic patterns of pleiotropy recognize disorders, perhaps as categories that can be more meaningful than the existing symptomatic approach? The meta-analysis across eight disorders identified that ADHD shares 37% and 27% of its common genetic variation of risk with ASD and Tourette syndrome respectively, and the three disorders were so clustered as ‘early-onset neurodevelopmental disorders’. Similarly, the underlying genetics of schizophrenia, MDD and bipolar disorder clustered these into one group - Mood and psychotic disorders -, and a third group ‘Disorders with compulsive behaviours’ clustered together OCD, anorexia nervosa and Tourette syndrome again. Thus, genetic correlations revealed three subgroups of highly genetically related disorders among eight psychiatric disorders, and this grouping based on shared genomics (pleiotropic effects) is starkly different from that of DSM. However, the discovery of shared biology does not simply imply that clinical categories are entirely

collapsible, mainly because the genetics can be continuous rather than categorical and clear genetic signatures or boundaries that can define neuropsychiatric syndromes will be a rarity[301].

Objective 7. Deep phenotyping and big data projects

Identification of the factors that predict disease course and outcome requires suitably powered clinical data routinely collected over extended time periods. Attaining sufficient sample sizes lie beyond the reach of a single group, and several consortia were born that coalesce data from research groups across countries and continents to make it accessible for researchers all over. 'Big data' is required on a longitudinal perspective and not only a cross-sectional one, and is more meaningful when achieved through deep phenotyping. Some consortia (like the PGC) accumulate population cohorts for GWAS, while others provide neuroimaging for genetics (like ENIGMA). There are more consortium projects that gather phenotype data for traits like personality, sleep, smoking, brain volume, cognitive functions (e.g. PsychENCODE), and others for deep phenotyping (e.g., Philadelphia Neurodevelopmental Cohort)[273].

The so-called 'big data' can dramatically change our ways to do research, especially when we are starting to assess the developmental trajectory of mental disorders. However, in order to have truly informative data, the study methods can be refined further. For instance - Psychiatric neuroimaging has largely followed the case-control model and include individuals that meet a clinical construct for a particular diagnosis. A minority of newer studies have examined clinical cohorts using continuous symptom measures rather than the case-control setup. Studies perform individual-level repeated neuroimaging to record highly stable signal estimates, or individual-level neuroimaging, using repeated assessments of a single individual and record session-to-session variations in signals. However, the two approaches are hardly integrated and there is still a lack of large-scale attempts for "deep neural phenotyping" in psychopathology. Also, findings from current MRI-based studies should be treated with caution since MRI signals are sensitive to frequent confounders like weight, stress, mental state, alcohol, substance use, and even head movement and breathing[303].

The boom in genomic data due to GWAS and NGS and backing of Electronic Health Records (EHR) in medicine have opened the prospects to unify phenotype and genotype data into medical records. Biobanks (like the UK Biobank, Mayo Clinic) paired with EHR are a straightforward way to build large datasets. EHRs store longitudinal data that can be used to structure phenotypes of patients and support personalized medicine. The PsycheMERGE

leverages the resources and existing infrastructure of the Electronic Medical Records and Genomics (eMERGE) network, the PGC, and local EHR and biobanks for psychiatric disorders (including Vanderbilt, MGH, Philadelphia Neurodevelopmental Cohort) to build powerful opportunity for psychiatric genetics[304]. Otherwise, the present EHR-biobank projects lack a major emphasis on psychiatric disorders yet[273].

Still, even comprehensive EHRs can lack high-quality data due to issues like incompleteness, inaccuracy, complexity and bias. The EHR is not a perfect image of the individual and physiology but is rather a reflection of the health care recording procedure with noise and feedback loops. Therefore, variability and imprecision in the clinical documentation challenges the phenotyping in the EHR[305]. To synchronize phenotype studies, a universal terminology and ontology should be used while labelling phenotypes. The Human Phenotype Ontology project is which is also linked to the Online Mendelian Inheritance in Man (OMIM) resource is one way to achieve consistent labelling. Ontologies can be used to calculate phenotype similarity metrics between patients and 'high-throughput phenotyping' can thus be built that will reduce the knowledge engineering effort (Figure 39)[306].

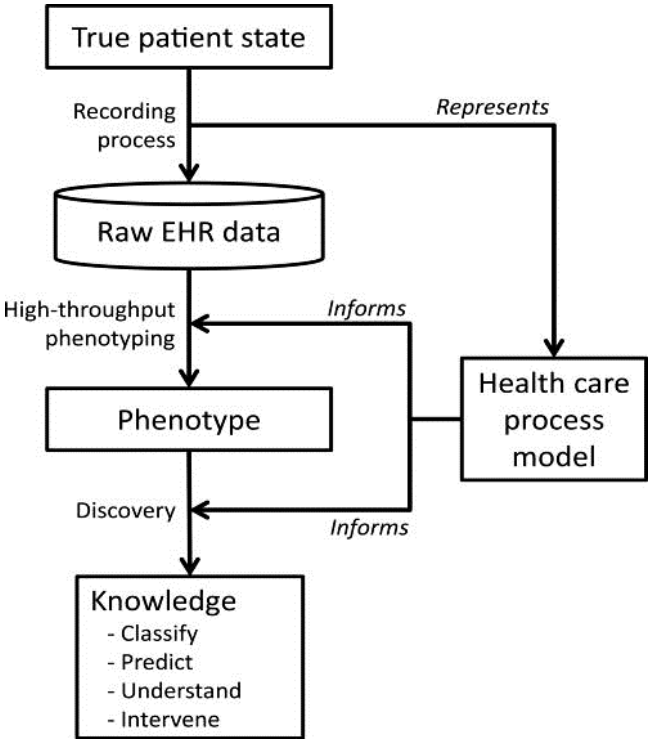


Figure 39. Phenotyping and discovery.

Adapted from Hripcsak and Albers, 2012.

Longitudinal deep phenotyping efforts can as well support a much-desired N-of-1 study design. In this design, individuals may receive multiple acute interventions, along with control

interventions, so that the effect of the intervention can be statistically tested for that individual. N-of-1 studies are considered to be at the top of the hierarchy of evidence-based medicine methods[307]. In-depth phenotyping is also being applied to refine the taxonomy of psychiatric disorders. The symptom-based diagnostic categories often poorly align with the findings on dysfunctions in brain circuits or neurobiological pathways, where the identified dysfunctions are shared by different DSM/ICD diagnoses and are transdiagnostic. The Biological Classification of Mental Disorders (BeCOME) study aims to identify classes of mental disorders that are based on the biology gleaned from in-depth phenotyping procedures and several levels of omics, cellular, imaging and psychophysiological assessments[308]. PsyCourse, another transdiagnostic study on affective-to-psychotic continuum combines longitudinal deep phenotyping of positive, depressive, and manic symptoms. The study advocates the introduction of a psychosis spectrum disorder as in day-to-day reality the psychotic symptoms between schizophrenia and bipolar disorder show a major overlap[309].

Environmental exposures are known drivers of psychiatric behaviors and disease outcomes. 'Exposome' studies that pursue the effects of exposures on behaviors and disease risk across the life course require high-quality environmental exposure data. In the Netherlands a platform has been setup by the Geoscience and hEalth Cohort COnsortium (GECCO) and the Global Geo Health Data Center (GGHDC) that centralizes and shares environmental variables as personal exposures. This includes a range of environmental data including high spatial and temporal resolution information on urban infrastructure, physicochemical exposures, availability of community services; but this data was scattered until it was centralized by GECCO. The resource now supplements 23 cohort studies and is a blueprint to set up similar big data projects[310]. Data from different regions will be shaped by different environmental risk factors. Indeed, if correctly interpreted, it might be a unique opportunity to assess cultural and regional differences in social and environmental factors pertaining to mental health.

If all is achieved, research using big data will still have some inherent challenges. Data security and privacy remains a major issue when it comes to ambulatory assessments, which necessitates geolocation and information relating to behaviors such as substance use. There is also high variability in ethical requirements across countries for research, making it difficult to uniformize research practices. Patients reaching institutional care settings may feel undue pressure to participate in research and may harbor fears that outcomes emerging from study participation might negatively affect their status[311].

Objective 8. Modeling psychiatric disorders

Animal models for psychiatric illnesses have been long established to elucidate their pathophysiology and to test novel treatment strategies. Unfortunately, but unsurprisingly, the utility of current animal models is limited in terms of translatability of the findings, probably because the models do not simulate the complex disorders wholly. The function of individual genes prioritized by GWAS can be assessed in gene knockout models (without reproducing the actual genetic variation), but the feasibility of incorporating the true nature of disease-associated genetic variants is questionable. How can one model the common variants of small effects leading or not to gene and/or protein changes in mouse orthologues of human genes? And also, how to deal with polygenicity in an animal model?

Take the case of *DISC1*, a candidate gene for schizophrenia and affective disorders - Several animal models have investigated mutations in the DISC1 gene that were identified in families, but only one, found in a particular family from Scotland, showed a comparable phenotype expression as the human allele. Thus, achieving a valid construct requires need for attention for each associated mutation. Another example is the neuroligin 3 (NLGN3) p.R451C missense mutation identified in ASD patients. $NLGN3^{R451C}$ mouse displayed social impairment but this phenotype was not seen in the NLGN3 knockout mouse. This suggests that a gain-of-function cannot be modelled with traditional approaches[312]. High-penetrance disease-associated rare variants can be modelled in rodents through CRISPR-based technologies but even so, the variant-endophenotype-behaviour inferences drawn from rodent models must be treated with caution before translation because of the extreme pleiotropy and redundancy observed in human[313].

Besides individual genetic manipulations, how can we create animal models that do not simply mimic the phenotypic appearance but epitomize the psychiatric etiology[314]? Projecting the progression of these disorders from initiatory states to full-blown stages in animal models is the key to understand the development of abnormal networks scattered over time points in order to target apt treatment strategies. Much of the animal modelling has overlooked the incorporation of spatial and temporal components to generate valid constructs. Models that incorporate environmental components trail behind genetic constructs, while models that aim to combine genetics and environment are even rare[312].

Lately, phenotypic assessment in animal models, particularly mice is in progress and can be particularly useful to choose those mice that will serve as better models. Some strains more than others due to their genetic background will match the behaviors that are observed in psychiatric diseases. For instance, in contrast to the commonly used inbred mouse strain C57/BL6, another strain, BTBR T+tf/J, displays less reciprocal social interaction, more grooming and different ultrasonic vocalizations. Thus, the phenotypic attributes of BTBR mice makes it a better model to study genetics of autism and antisocial-like behaviors[314].

There is a scope to continue refining animal models utilizing the phenotype-matched animals and optogenetics models (Figure 40) which are very much advanced than the traditional gene knock-out and knock-in mice[315]. However, given the intrinsic caveats in animal models - 80–90 million years of evolutionary divergence, different selection pressures due to the evolutionary niches, significant differences in the neural cells types and neural circuits, especially, but not limited to prefrontal cortex and its projection, none of the models can replicate the relapsing–remitting nature of the disease, and it seems increasingly unlikely that

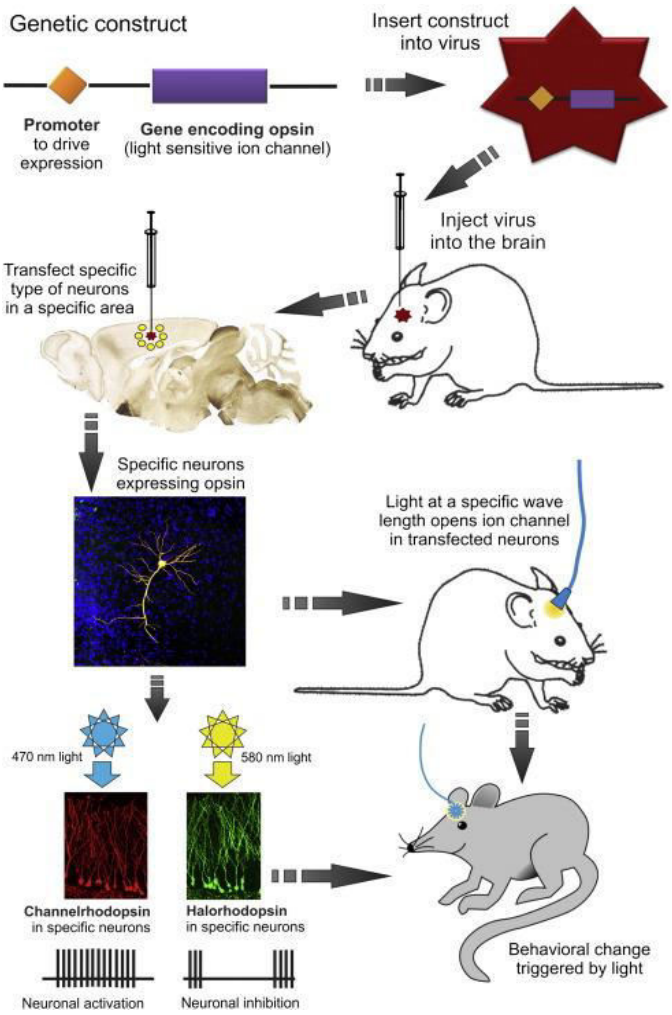


Figure 40. Major steps creating a behavioral experiment involving optogenetic tools.

Adapted from Czéh *et al.*, 2015.

modelling these diseases in animals will depict the 'model' of a psychiatric disorder and not a 'disorder-like' behavior[145,314–316].

Perhaps more sophisticated approaches to overcome these hurdles include cellular models engendered by stem cell technology, and which can contain the entire genetic background of the donor. Readily available human cells such as skin fibroblasts can be reprogrammed into neurons and glial cells *in vitro* either directly or through an intermediate cell (Figure 41). Induced pluripotent stem cell (iPSC) models that can generate patient-derived neurons are under testing for psychiatric disorders including schizophrenia, bipolar disorder, and Rett syndrome; and in some cases, they have uncovered gene-associated phenotypes in processes of progenitor cell proliferation, synaptic maturation, neuronal morphology and activity. iPSC-derived neurons may have downsides like limited epigenetic memory, immaturity of neurons, but it is possible to recapitulate the epigenetic modifications post neuronal maturation *in vitro*[317]. The original epigenetic landscape can also be preserved using transdifferentiation *i.e.* directly inducing a somatic cell like fibroblast into functional-induced neurons (iNs) without a transient stage of pluripotent stem cells (Figure 41)[318]. Therefore, iNs are a powerful valid alternative for the modeling of neuropsychiatric diseases, as it will retain much of epigenetic changes present in the somatic cells of the patients.

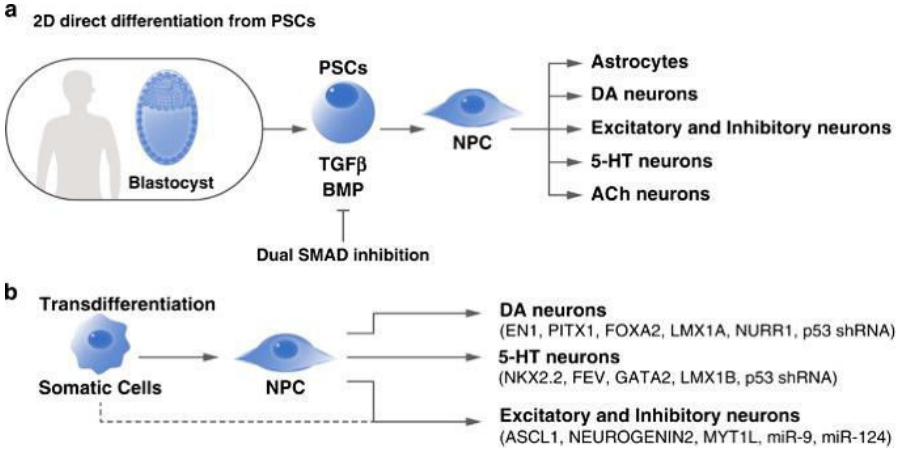


Figure 41. Neuronal differentiation.

Adapted from Parr *et al.*, 2017.

Even superior models that can mimic brain development and disorders are three-dimensional iPSC-derived brain organoids[313]. Brain organoids, with its complex tissue structure have the potential to recapitulate developing neural circuits and circuit level interactions, and are responsive to drug treatment too. Brain organoids generated from ASD patients captured maturation-related signatures including an increased production of GABAergic neurons, and

have been used to model Timothy syndrome, a severe neurodevelopmental disease characterized by ASD and epilepsy[317,319]. The organoid field is young and currently lacks some of the brain cell types but is highly promising, with much room for improvement.

Objective 9. Psychiatric therapy and precision medicine

Pharmacogenetics research into psychiatric diseases is in its early stages. Each drug is tested in fewer than 5000 samples and 75% of psychiatric drugs fail to effectively treat the patient's condition. The low efficacy of psychiatric drugs arises from a dearth of insights into the pathogenesis of psychiatric traits, together with indistinct diagnoses based on symptom counts.

How soon will today's genetics research help clinicians and patients? Can the risk variants in absence of the knowledge of causal genes contribute to treatment? Indeed, the identified variants from the GWASs so far are valuable for genomic risk prediction and to identify druggable targets[157]. While GWASs aim to detect reproducible disease-associated loci, genomic risk prediction doesn't necessarily require a causal interpretation. This means that as a greater number of identified variants are discovered, more robust risk prediction models can be developed. The variants should be evaluated for an association with clinical features of disorders including early-onset, recurrence, severity, and anatomical and functional differences[157]. Multiple data types from neuroimaging, genetic and clinical predictors improve the performance of the models and so can the pervasive pleiotropy among psychiatric phenotypes[320]. One example is that of 22q11 deletions, which is a strong susceptibility factor for a range of psychiatric disorders including ASD, ADHD, ID, anxiety, mood disorders, as well as subsyndromal cognitive and psychiatric impairments[109]. Estimation of disease risk is applicable for individuals that carry 22q11 deletion to assist with genetic counseling.

Not all risk variants will be present in all individuals, as population and inter-individual genetic differences play a part. Risk allele weights derived from the major ancestral populations are needed to predict genetic risk scores (GRS). The failure to include diverse, multi-ethnic populations in genetic studies exacerbates health disparities as the therapeutic models would mostly work on the Europeans but can be potentially dangerous if extended to other ethnic groups[273]. Thus, a lack of ancestral diversity in current GWAS can impede the clinical use of GRS globally. A personalized analysis of variants and genes (through iMEGES) can therefore aid the identification of potential risk factors in patients, so that the risk predictions and treatment would be more effective[283].

For an effective drug therapy, affected individuals with similar neurophysiological abnormalities should be clustered or grouped, and novel interventions that would work with a degree of specificity for groups should be developed and tested (Figure 42)[313]. Well powered GWASs designed to investigate the efficacy and side effects of treatment can predict appropriate drug target choices than current candidate genes (like *CYP2D6* and *CYP2C19*)[273]. Phenome-wide association studies (PheWAS) is another promising way to aid drug development using data from electronic medical records, disease-agnostic cohorts and GWAS, but is largely untapped for prioritization of drug targets. With PheWAS, associations between a specific genetic variant and a range of physiological, clinical, and phenotypic outcomes can be tested, while accounting for the pleiotropic effects. PheWAS can elucidate mechanisms of drug action, identify alternative indications, and adverse drug events (ADEs)[321]. Targeting loci that have pleiotropic effects on disorders can provide broad-spectrum therapeutic effects.

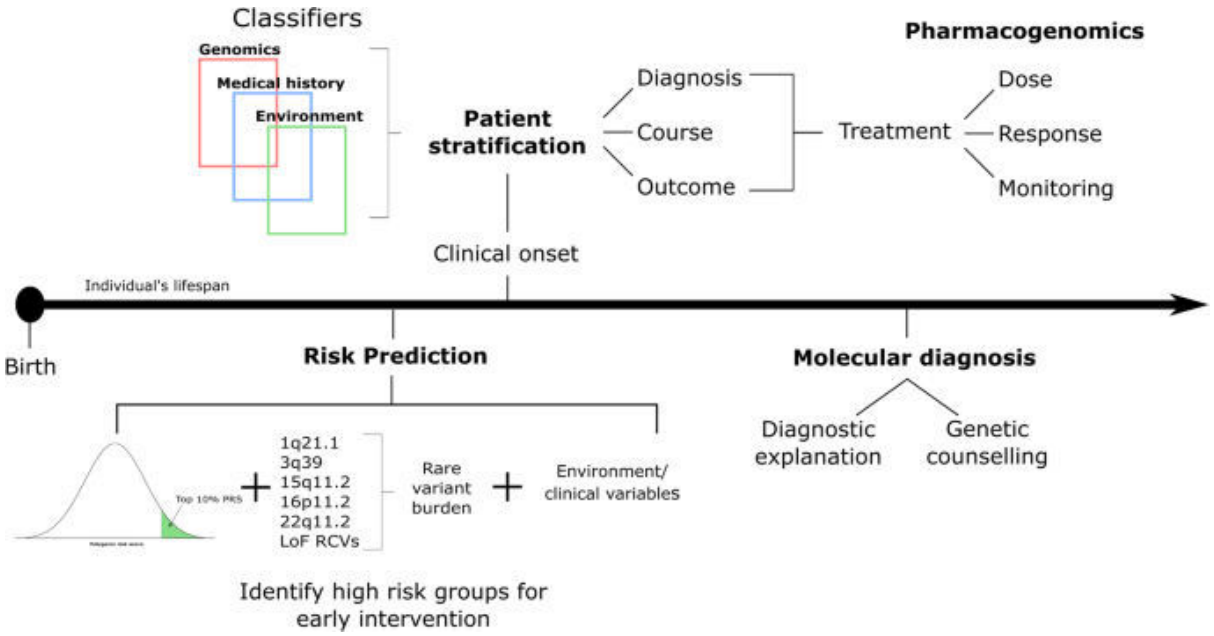


Figure 42. Precision medicine to tailor healthcare to individual patients four key areas (risk prediction, patient stratification, pharmacogenomic and molecular diagnostics).

Adapted from Rees and Owen, 2020.

In addition to the conventional 'genetic' or 'protein-based' drug targets, epigenetic modifications due to their reversible nature might represent novel targets for therapeutic improvement (Figure 43). A new class of medication, 'epidrugs', has emerged to modulate epigenetic signaling. In the earliest stages, three classes of molecules are potentially efficacious

including histone demethylase inhibitors (HMT), histone deacetylase inhibitors (HDAC), and DNA methyltransferase (DNMT) inhibitors acting on DNA methylation. The role of DNMT inhibitors as anti-cancer drugs in cancer treatment has been widely recognized but their therapeutic potential in the treatment of psychiatric disorders is still at the preclinical stage. Targeted epigenetic strategies can be aimed to correct only the putative pathogenic marks and leave the homeostatic 'beneficial' marks unchanged[322].

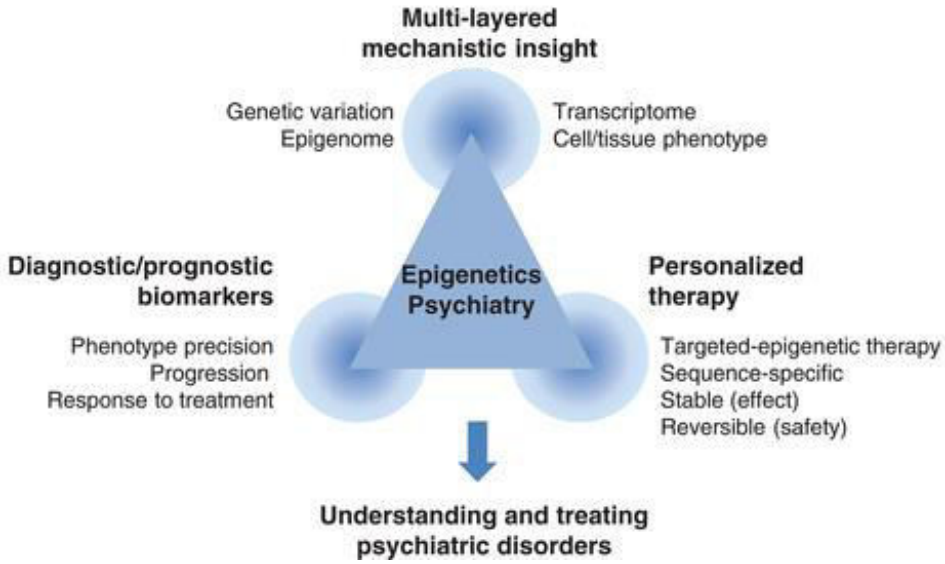


Figure 43. Clinical opportunities brought by epigenetic studies in psychiatry.

Adapted from Kular and Kular., 2018.

The success of medications in psychiatric disorders can be greatly enhanced by personalized biopsychosocial interventions that sometimes are undervalued. The critical developmental windows are instrumental to long-term deleterious epigenetic effects, and so the correction of early epigenetic disruptions is proving to be a key influence in psychiatric therapies. For instance - prevention of early psycho-trauma in rodents has been shown to alleviate long-term detrimental impacts and is one such successful environmental/epigenetic intervention. Implementing a supportive enriched environment (EE) can also convalesce depression more effectively than antidepressant drugs, alleviate anxiety and cognitive impairments, as seen in young adult mice. The positive effects in EE intervention result from an epigenetic reprogramming of genes encoding receptors like BDNF or CRH, the key factors in stress response, and these positive effects can persist across generations. Thus it is hinted that initial interventions can go a long way in improving genetic predispositions to impairments induced by early life stress[322].

Lookup

The genetics of psychiatric disorders is evolving exponentially, and continues to reveal the incredible layers of complexity, heterogeneity and pleiotropy. Given the marked progress in psychiatric illnesses to date because of GWAS, WGS, WES, and new analytical approaches, these methods will continue to expand, covering more disorders, phenotypes, diverse populations, drug responses in the gene and drug discovery efforts. The clustering and splitting of psychiatric disorders are likely to continue, on newly found grounds of biological insights steering a more bottom-up classification of psychiatric disorders. There remain critical glitches to be solved before psychiatry can finally support each and every patient - many risk variants have not yet been uncovered, inclusion of more ethnicities, we are yet to fully appreciate the ability of studying the comorbidities together, the mechanisms of action of pleiotropic loci, improving model systems and more. Nonetheless, the current omics, informatics and state-of-the-art technologies altogether have built an empirical platform upon which psychiatry can now progress these issues. What is important to remember is that the current genetic findings have the power to make an immediate impact on psychiatry; by testing rare pathogenic CNVs in patients, detection of brain abnormalities, screening for comorbidities, genetic risk assessment, counselling and biopsychosocial interventions. The decades of psychiatric research have gradually diminished the clinical and genetic boundaries and have established that psychiatric genetics is no longer peripheral but is a more natural driver for clinical psychiatry.

CHAPTER 6: FUTURE WORK

6.1. Expanding research on the miRNA system

So far, we have explored the role of common variation in miRNA genes in the predisposition to ADHD. As a next step, we aim to investigate the impact of genetic variation located in the miRNA target sites in the 3'UTR of mRNAs, known as poly-miRTSs. Approximately 11% of known SNPs are located in the 3'UTR regions of 16,810 genes[323]. A variation in a 3′-UTR can either introduce or remove miRNA target sequences or change the binding efficiency, which in turn can alter the gene expression.

6.2. Studying epigenetic risk factors in other comorbid disorders

We plan to investigate the role of miRNAs and ASM also in ASD, highly comorbid with ADHD. The most recent genome-wide association meta-analysis of ASD, including 18,381 cases and 27,969 controls, published in 2019, revealed five GWS loci. It would be interesting to investigate the epigenetic underpinnings in ASD and find any possible overlap with ADHD biology. Joint analysis for comorbid disorders will be pursued to find genetic specificity but also shared risk factors.

6.3. Attempting larger cocaine-dependence GWAS

We aim to increase the sample size for cocaine-dependence GWAS so as to establish higher levels of significance for confident identification of common susceptibility variants. The different GWASs performed so far in a number of psychiatric disorders indicate that only after reaching around 10,000 cases the first GWS findings emerge. Keeping this figure in mind, we are in touch with different research groups from different centers (e.g. Aarhus University, Vall d'Hebron Institut de Recerca or Universidade Federal do Rio Grande do Sul, Porto Alegre, among others) that are currently genotyping several thousands of additional patients. In parallel, the PGC also intends to genotype a larger number of subjects that are dependent on different drugs, including cocaine, opiates, cannabis and alcohol, which may be also useful.

6.4. Trans-ancestry studies

Individuals from East Asian and African American descent are now being increasingly studied as this will help in both achieving a finer resolution in GWAS and eliminating health disparities. Thus, replicating the epigenetic and genetic signals that we have identified in

European samples in samples from other ancestries remains a challenge and a commitment that the scientific community should take.

6.5. Functional follow-up of candidate genes

We are following up the *C2orf82* gene, one of our prioritized candidates in ADHD, using a gene-knockout mouse model. Our studies involve extensive behavioral testing that have already revealed abnormalities like hypolocomotion and impulsivity. Interestingly, the depletion of the gene also causes decreased motivation for cocaine, indicating that it may have pleiotropic effects. In parallel, the group investigates the effect of genes pointed in studies that focus on cocaine dependence, such as *PLCB1,* encoding the signaling molecule phospholipase C beta 1, also through mouse modelling. Zebrafish is also used by us to study genes involved in several psychiatric conditions or traits, including aggressive behavior or ASD. Although we are aware that we are modelling conditions that are polygenic and in general explained by the effect of common variants, investigating single-gene knockouts provides valuable information on the impact on the brain associated with the malfunction of such genes.

CONCLUSIONS

Exploring genetic variation that influences brain methylation in attention-deficit/hyperactivity disorder (ADHD)

1. Common genetic risk variants for ADHD identified in a previous genome-wide association study (GWAS) that included 20,000 cases and 35,000 controls are enriched in SNPs that correlate with levels of DNA methylation.
2. Eight Allele-Specific Methylation tagSNPs are significantly associated with ADHD and correlate with differential methylation at six CpG sites *in cis* in different brain areas.
3. These six CpG sites are located at possible promoter regions of six genes expressed in brain: *ARTN*, *C2orf82*, *NEUROD6*, *PIDD1*, *RPLP2* and *GAL*.
4. For three of these six genes, SNPs associated with ADHD and correlating with methylation levels are eQTLs in brain. Consistently, methylation and gene expression show opposite directions: *ARTN* and *PIDD1* (reduced methylation, increased expression), *C2orf82* (increased methylation, reduced expression).
5. ADHD risk alleles are associated with increased brain expression of *ARTN* and *PIDD1* and with decreased brain expression of *C2orf82*.
6. SNPs in *C2orf82* correlate with changes in brain volumes.
7. In summary, our study highlights the *ARTN*, *C2orf82* and *PIDD1* genes as potential contributors to ADHD susceptibility.

Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions

1. We have performed the largest cocaine dependence GWAS meta-analysis in individuals of European ancestry, including 2,100 cases and 4,300 controls.
2. Although the SNP-based analysis revealed no genome-wide significant associations with cocaine dependence, probably due to limited sample size, the gene-based analysis identified the *HIST1H2BD* gene, previously associated with schizophrenia.
3. The estimated SNP-based heritability of cocaine dependence is approximately 30%.
4. A significant genetic correlation has been observed between cocaine dependence and ADHD, schizophrenia, major depressive disorder and risk-taking behaviour, suggesting a shared genetic basis across pathologies and traits.
5. Polygenic risk score (PRS) analysis shows that all the comorbid features analysed (ADHD, schizophrenia, major depressive disorder, aggressiveness, antisocial personality or risk-taking behaviour) can predict cocaine dependence.

Exploring the impact of common variation in miRNA genes on attention-deficit/hyperactivity disorder

1. We have performed a case-control association study to investigate the contribution to ADHD of common genetic variation in 1,761 autosomal miRNAs using pre-existing GWAS data from 20,000 cases and 35,000 controls.
2. We have identified 19 significant associations of SNPs with ADHD that highlight 12 miRNA genes, all located within protein-coding genes.
3. The associated variants are located in the putative regulatory regions of the miRNA genes or in the promoter region of the host protein-coding gene.
4. Two of the 12 highlighted miRNA loci, miR-6079 and miR-3666, overlap with genome-wide association findings from the pre-existing ADHD meta-analysis that was used as a starting dataset for the present study.
5. Six of the 12 highlighted miRNAs are expressed in different brain tissues, specifically in cerebellum, whereas for the rest this information is not yet available.
6. Three of the highlighted miRNAs - miR-3666, miR-7-1 and miR-1273h - have validated target mRNAs.
7. Pathway analysis of ADHD-associated miRNAs revealed two biological pathways. One of the pathways involves miRNA-mediated regulation of serotonin receptor genes and it is likely to be involved in neurological functions and diseases.

In summary, our studies have contributed to identify common genetic and epigenetic risk factors that underlie the susceptibility to ADHD and to cocaine dependence. The results reinforce the idea that epigenetic mechanisms dictate the differential expression of genes that may be causal to ADHD. Cocaine dependence, which has been widely believed to occur under environmental and epigenetic influences, is also in part genetically determined. Finally, ADHD and cocaine dependence are comorbid disorders, and the observed genetic correlation between these conditions can reflect biological pleiotropy.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).

2. Thomas, R., Sanders, S., Doust, J., Beller, E. & Glasziou, P. Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics* **135**, e994–e1001 (2015).

3. Grimm, O., Kittel-Schneider, S. & Reif, A. Recent developments in the genetics of ADHD. *Psychiatry Clin. Neurosci.* (2018). doi:10.1111/pcn.12673

4. Franke, B. *et al.* Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. *Eur. Neuropsychopharmacol.* **28**, 1059–1088 (2018).

5. American Psychiatric Association. & American Psychiatric Association. DSM-5 Task Force. *Diagnostic and statistical manual of mental disorders : DSM-5.* (2014).

6. Bell, C. C. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA J. Am. Med. Assoc.* **272**, 828 (1994).

7. Epstein, J. N. & Loren, R. E. A. Changes in the Definition of ADHD in DSM-5: Subtle but Important. *Neuropsychiatry (London).* **3**, 455–458 (2013).

8. Curatolo, P., D'Agati, E. & Moavero, R. The neurobiological basis of ADHD. *Ital. J. Pediatr.* **36**, 79 (2010).

9. Katzman, M. A., Bilkey, T. S., Chokka, P. R., Fallu, A. & Klassen, L. J. Adult ADHD and comorbid disorders: clinical implications of a dimensional approach. *BMC Psychiatry* **17**, 302 (2017).

10. Weissenberger, S. *et al.* ADHD, Lifestyles and Comorbidities: A Call for an Holistic Perspective – from Medical to Societal Intervening Factors. *Front. Psychol.* **8**, 454 (2017).

11. Schmidt, S. & Petermann, F. Developmental psychopathology: Attention Deficit Hyperactivity Disorder (ADHD). *BMC Psychiatry* **9**, 58 (2009).

12. Mrug, S. *et al.* Peer rejection and friendships in children with Attention-Deficit/Hyperactivity Disorder: contributions to long-term outcomes. *J. Abnorm. Child Psychol.* **40**, 1013–26 (2012).

13. Saylor, K. E. & Amann, B. H. Impulsive Aggression as a Comorbidity of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *J. Child Adolesc. Psychopharmacol.* **26**, 19–25 (2016).

14. Wehmeier, P. M., Schacht, A. & Barkley, R. A. Social and Emotional Impairment in Children and Adolescents with ADHD and the Impact on Quality of Life. *J. Adolesc. Heal.* **46**, 209–217 (2010).

15. Cherkasova, M., Sulla, E. M., Dalena, K. L., Pondé, M. P. & Hechtman, L. Developmental course of attention deficit hyperactivity disorder and its predictors. *J. Can. Acad. Child Adolesc. Psychiatry* **22**, 47–54 (2013).

16. Stefanatos, G. A. & Baron, I. S. Attention-Deficit/Hyperactivity Disorder: A Neuropsychological Perspective Towards DSM-V. *Neuropsychol. Rev.* **17**, 5–38 (2007).

17. Stevenson, R. D. & Wolraich, M. L. Stimulant medication therapy in the treatment of children with attention deficit hyperactivity disorder. *Pediatr. Clin. North Am.* **36**, 1183–97 (1989).

18. Stevens, J. R., Wilens, T. E. & Stern, T. A. Using stimulants for attention-deficit/hyperactivity disorder: clinical approaches and challenges. *Prim. care companion CNS Disord.* **15**, (2013).

19. Wilens, T. E., Morrison, N. R. & Prince, J. An update on the pharmacotherapy of attention-deficit/hyperactivity disorder in adults. *Expert Rev. Neurother.* **11**, 1443–65 (2011).

20. Lensing, M. B., Zeiner, P., Sandvik, L. & Opjordsmoen, S. Four-Year Outcome in Psychopharmacologically Treated Adults With Attention-Deficit/Hyperactivity Disorder. *J. Clin. Psychiatry* **74**, e87–e93 (2013).

21. Shaw, M. *et al.* A systematic review and analysis of long-term outcomes in attention deficit hyperactivity disorder: Effects of treatment and non-treatment. *BMC Med.* **10**, 99 (2012).

22. Powers, R. L., Marks, D. J., Miller, C. J., Newcorn, J. H. & Halperin, J. M. Stimulant treatment in children with attention-deficit/hyperactivity disorder moderates adolescent academic outcome. *J. Child Adolesc. Psychopharmacol.* **18**, 449–59 (2008).

23. Shier, A. C., Reichenbacher, T., Ghuman, H. S. & Ghuman, J. K. Pharmacological treatment of attention deficit hyperactivity disorder in children and adolescents: clinical strategies. *J. Cent. Nerv. Syst. Dis.* **5**, 1–17 (2013).

24. Sonuga-Barke, E. J. S. *et al.* Nonpharmacological Interventions for ADHD: Systematic Review and Meta-Analyses of Randomized Controlled Trials of Dietary and Psychological Treatments. *Am. J. Psychiatry* **170**, 275–289 (2013).

25. Froelich, J., Doepfner, M. & Lehmkuhl, G. EFFECTS OF COMBINED COGNITIVE BEHAVIOURAL TREATMENT WITH PARENT MANAGEMENT TRAINING IN ADHD. *Behav. Cogn. Psychother.* **30**, 111–115 (2002).

26. Vidal, R. *et al.* Group Therapy for Adolescents With Attention-Deficit/Hyperactivity Disorder: A Randomized Controlled Trial. *J. Am. Acad. Child Adolesc. Psychiatry* **54**, 275–282 (2015).

27. Antshel, K. M., Faraone, S. V. & Gordon, M. Cognitive Behavioral Treatment Outcomes in Adolescent

ADHD. *J. Atten. Disord.* **18**, 483–495 (2014).

28. Faraone, S. V. *et al.* Molecular Genetics of Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* **57**, 1313–1323 (2005).

29. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol. Psychiatry* **24**, 562–575 (2019).

30. Roza, S. J. *et al.* Effects of maternal smoking in pregnancy on prenatal brain development. The Generation R Study. *Eur. J. Neurosci.* **25**, 611–617 (2007).

31. de Zeeuw, P., Zwart, F., Schrama, R., van Engeland, H. & Durston, S. Prenatal exposure to cigarette smoke or alcohol and cerebellum volume in attention-deficit/hyperactivity disorder and typical development. *Transl. Psychiatry* **2**, e84 (2012).

32. Langley, K., Rice, F., van den Bree, M. B. M. & Thapar, A. Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour. A review. *Minerva Pediatr.* **57**, 359–71 (2005).

33. Kotimaa, A. J. *et al.* Maternal smoking and hyperactivity in 8-year-old children. *J. Am. Acad. Child Adolesc. Psychiatry* **42**, 826–33 (2003).

34. Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M. & Anand, K. J. S. Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born Preterm. *JAMA* **288**, 728 (2002).

35. Franz, A. P. *et al.* Attention-Deficit/Hyperactivity Disorder and Very Preterm/Very Low Birth Weight: A Meta-analysis. *Pediatrics* **141**, e20171645 (2018).

36. Ben Amor, L. *et al.* Perinatal complications in children with attention-deficit hyperactivity disorder and their unaffected siblings. *J. Psychiatry Neurosci.* **30**, 120–6 (2005).

37. Getahun, D. *et al.* In Utero Exposure to Ischemic-Hypoxic Conditions and Attention-Deficit/Hyperactivity Disorder. *Pediatrics* **131**, e53–e61 (2013).

38. Wang, Y., Huang, L., Zhang, L., Qu, Y. & Mu, D. Iron Status in Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-Analysis. *PLoS One* **12**, e0169145 (2017).

39. Bener, A., Kamal, M., Bener, H. & Bhugra, D. Higher prevalence of iron deficiency as strong predictor of attention deficit hyperactivity disorder in children. *Ann. Med. Health Sci. Res.* **4**, S291-7 (2014).

40. LaChance, L., McKenzie, K., Taylor, V. H. & Vigod, S. N. Omega-6 to Omega-3 Fatty Acid Ratio in Patients with ADHD: A Meta-Analysis. *J. Can. Acad. Child Adolesc. Psychiatry* **25**, 87–96 (2016).

41. Froehlich, T. E. *et al.* Update on environmental risk factors for attention-deficit/hyperactivity disorder. *Curr. Psychiatry Rep.* **13**, 333–44 (2011).

42. Avale, M. E. *et al.* Interplay of beta2* nicotinic receptors and dopamine pathways in the control of spontaneous locomotion. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15991–6 (2008).

43. Neuman, R. J. *et al.* Prenatal Smoking Exposure and Dopaminergic Genotypes Interact to Cause a Severe ADHD Subtype. *Biol. Psychiatry* **61**, 1320–1328 (2007).

44. Laucht, M. *et al.* Interacting Effects of the Dopamine Transporter Gene and Psychosocial Adversity on Attention-Deficit/Hyperactivity Disorder Symptoms Among 15-Year-Olds From a High-Risk Community Sample. *Arch. Gen. Psychiatry* **64**, 585 (2007).

45. Asherson, P. *et al.* A high-density SNP linkage scan with 142 combined subtype ADHD sib pairs identifies linkage regions on chromosomes 9 and 16. *Mol. Psychiatry* **13**, 514–521 (2008).

46. Arcos-Burgos, M. *et al.* Attention-Deficit/Hyperactivity Disorder in a Population Isolate: Linkage to Loci at 4q13.2, 5q33.3, 11q22, and 17p11. *Am. J. Hum. Genet.* **75**, 998–1014 (2004).

47. Smalley, S. L. *et al.* Genetic Linkage of Attention-Deficit/Hyperactivity Disorder on Chromosome 16p13, in a Region Implicated in Autism. *Am. J. Hum. Genet.* **71**, 959–963 (2002).

48. Zhou, K. *et al.* Meta-analysis of genome-wide linkage scans of attention deficit hyperactivity disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B**, 1392–8 (2008).

49. Reif, A. *et al.* DIRAS2 is associated with adult ADHD, related traits, and co-morbid disorders. *Neuropsychopharmacology* **36**, 2318–27 (2011).

50. Arcos-Burgos, M. *et al.* A common variant of the latrophilin 3 gene, LPHN3, confers susceptibility to ADHD and predicts effectiveness of stimulant medication. *Mol. Psychiatry* **15**, 1053–1066 (2010).

51. Gizer, I. R., Ficks, C. & Waldman, I. D. Candidate gene studies of ADHD: a meta-analytic review. *Hum. Genet.* **126**, 51–90 (2009).

52. Ribasés, M. *et al.* An association study of sequence variants in the forkhead box P2 (FOXP2) gene and adulthood attention-deficit/hyperactivity disorder in two European samples. *Psychiatr. Genet.* **22**, 155–160 (2012).

53. Lesch, K.-P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J. Neural Transm.* **115**, 1573–1585 (2008).

54. Zayats, T. *et al.* Genome-Wide Analysis of Attention Deficit Hyperactivity Disorder in Norway. *PLoS One* **10**, e0122501 (2015).

55. Sánchez-Mora, C. *et al.* Case–Control Genome-Wide Association Study of Persistent Attention-Deficit Hyperactivity Disorder Identifies FBXO33 as a Novel Susceptibility Gene for the Disorder. *Neuropsychopharmacology* **40**, 915–926 (2015).

56. Anney, R. J. L. *et al.* Conduct disorder and ADHD: Evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **147B**, 1369–1378 (2008).

57. Neale, B. M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–97 (2010).

58. van Hulzen, K. J. E. *et al.* Genetic Overlap Between Attention-Deficit/Hyperactivity Disorder and Bipolar Disorder: Evidence From Genome-wide Association Study Meta-analysis. *Biol. Psychiatry* **82**, 634–641 (2017).

59. Martin, J., O'Donovan, M. C., Thapar, A., Langley, K. & Williams, N. The relative contribution of common and rare genetic variants to ADHD. *Transl. Psychiatry* **5**, e506 (2015).

60. Ramos-Quiroga, J.-A. *et al.* Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder. *J. Psychiatr. Res.* **49**, 60–67 (2014).

61. Williams, N. M. *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet (London, England)* **376**, 1401–8 (2010).

62. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–21 (2014).

63. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

64. Satterstrom, F. K. *et al.* ASD and ADHD have a similar burden of rare protein-truncating variants. *bioRxiv* 277707 (2018). doi:10.1101/277707

65. Lipman, V. 2 Valuable Tips To Help Manage Employees With ADHD. *Forbes* (2017). doi:2017-05-19

66. Corominas, J. *et al.* Identification of ADHD risk genes in extended pedigrees by combining linkage analysis and whole-exome sequencing. *Mol. Psychiatry* (2018). doi:10.1038/s41380-018-0210-6

67. Zayats, T. *et al.* Exome chip analyses in adult attention deficit hyperactivity disorder. *Transl. Psychiatry* **6**, e923 (2016).

68. Tovo-Rodrigues, L. *et al.* DRD4 rare variants in Attention-Deficit/Hyperactivity Disorder (ADHD): further evidence from a birth cohort study. *PLoS One* **8**, e85164 (2013).

69. Hawi, Z. *et al.* Rare DNA variants in the brain-derived neurotrophic factor gene increase risk for attention-deficit hyperactivity disorder: a next-generation sequencing study. *Mol. Psychiatry* **22**, 580–584 (2017).

70. Kim, D. S. *et al.* Sequencing of sporadic Attention-Deficit Hyperactivity Disorder (ADHD) identifies novel and potentially pathogenic de novo variants and excludes overlap with genes associated with autism spectrum disorder. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **174**, 381–389 (2017).

71. Slatkin, M. Epigenetic inheritance and the missing heritability problem. *Genetics* **182**, 845–50 (2009).

72. Trerotola, M., Relli, V., Simeone, P. & Alberti, S. Epigenetic inheritance and the missing heritability. *Hum. Genomics* **9**, 17 (2015).

73. O'Connor, R. M., Gururajan, A., Dinan, T. G., Kenny, P. J. & Cryan, J. F. All Roads Lead to the miRNome: miRNAs Have a Central Role in the Molecular Pathophysiology of Psychiatric Disorders. *Trends Pharmacol. Sci.* **37**, 1029–1044 (2016).

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

75. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).

76. Hicks, S. D. & Middleton, F. A. A comparative review of microRNA expression patterns in autism spectrum disorder. *Front. Psychiatry* **7**, 176 (2016).

77. Cao, T. & Zhen, X.-C. Dysregulation of miRNA and its potential therapeutic application in schizophrenia. *CNS Neurosci. Ther.* **24**, 586–597 (2018).

78. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin. Chem.* **56**, 1733–41 (2010).

79. Wang, H., Peng, R., Wang, J., Qin, Z. & Xue, L. Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage. *Clin. Epigenetics* **10**, 59 (2018).

80. Luoni, A. & Riva, M. A. MicroRNAs and psychiatric disorders: From aetiology to treatment. *Pharmacol. Ther.* **167**, 13–27 (2016).

81. Xu, B., Karayiorgou, M. & Gogos, J. A. MicroRNAs in psychiatric and neurodevelopmental disorders. *Brain Res.* **1338**, 78–88 (2010).

82. Srivastav, S., Walitza, S. & Grünblatt, E. Emerging role of miRNA in attention deficit hyperactivity disorder: a systematic review. *ADHD Atten. Deficit Hyperact. Disord.* **10**, 49–63 (2018).

83. Zadehbagheri, F., Hosseini, E., Bagheri-Hosseinabadi, Z., Rekabdarkolaee, H. M. & Sadeghi, I. Profiling

of miRNAs in serum of children with attention-deficit hyperactivity disorder shows significant alterations. *J. Psychiatr. Res.* **109**, 185–192 (2019).

84. Wu, L. H. *et al.* Circulating MicroRNA Let-7d in Attention-Deficit/Hyperactivity Disorder. *NeuroMolecular Med.* **17**, 137–146 (2015).

85. Wu, L. *et al.* A novel function of MicroRNA Let-7d in regulation of galectin-3 expression in attention deficit hyperactivity disorder rat brain. *Brain Pathol.* **20**, 1042–1054 (2010).

86. Wu, L. H. *et al.* Nr3C1-Bhlhb2 Axis Dysregulation Is Involved in the Development of Attention Deficit Hyperactivity. *Mol. Neurobiol.* **54**, 1196–1212 (2017).

87. Wang, L.-J. *et al.* Blood-Bourne MicroRNA Biomarker Evaluation in Attention-Deficit/Hyperactivity Disorder of Han Chinese Individuals: An Exploratory Study. *Front. Psychiatry* **9**, 227 (2018).

88. Sánchez-Mora, C. *et al.* Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. *Neuropsychopharmacology* 1 (2018). doi:10.1038/s41386-018-0297-0

89. Chen, K. *et al.* Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. *Carcinogenesis* **29**, 1306–1311 (2008).

90. Németh, N., Kovács-Nagy, R., Székely, A., Sasvári-Székely, M. & Rónai, Z. Association of Impulsivity and Polymorphic MicroRNA-641 Target Sites in the SNAP-25 Gene. *PLoS One* **8**, e84207 (2013).

91. Ye, C. *et al.* Two SNAP-25 genetic variants in the binding site of multiple microRNAs and susceptibility of ADHD: A meta-analysis. *J. Psychiatr. Res.* **81**, 56–62 (2016).

92. Garcia-Martínez, I. *et al.* Preliminary evidence for association of genetic variants in pri-miR-34b/c and abnormal miR-34c expression with attention deficit and hyperactivity disorder. *Transl. Psychiatry* **6**, e879–e879 (2016).

93. Šerý, O. *et al.* A 40-bp VNTR polymorphism in the 3'-untranslated region of DAT1/SLC6A3 is associated with ADHD but not with alcoholism. *Behav. Brain Funct.* **11**, 21 (2015).

94. Sánchez-Mora, C. *et al.* Evaluation of single nucleotide polymorphisms in the miR-183–96–182 cluster in adulthood attention-deficit and hyperactivity disorder (ADHD) and substance use disorders (SUDs). *Eur. Neuropsychopharmacol.* **23**, 1463–1473 (2013).

95. Shi, Z., Shen, T., Liu, Y., Huang, Y. & Jiao, J. Retinoic acid receptor γ (Rarg) and nuclear receptor subfamily 5, group A, member 2 (Nr5a2) promote conversion of fibroblasts to functional neurons. *J. Biol. Chem.* **289**, 6415–28 (2014).

96. Karakas, U. *et al.* Regulating the Regulators in Attention-Deficit/Hyperactivity Disorder: A Genetic Association Study of microRNA Biogenesis Pathways. *Omi. A J. Integr. Biol.* **21**, 352–358 (2017).

97. Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res.* **20**, 883–9 (2010).

98. van Mil, N. H. *et al.* DNA methylation profiles at birth and child ADHD symptoms. *J. Psychiatr. Res.* **49**, 51–59 (2014).

99. Do, C. *et al.* Genetic–epigenetic interactions in cis: a major focus in the post-GWAS era. *Genome Biol.* **18**, 120 (2017).

100. Tak, Y. G. & Farnham, P. J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57 (2015).

101. Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J. & Fallin, M. D. Epigenetic Research in Neuropsychiatric Disorders: the "Tissue Issue". *Curr. Behav. Neurosci. reports* **3**, 264–274 (2016).

102. Qiu, A. *et al.* Basal Ganglia Volume and Shape in Children With Attention Deficit Hyperactivity Disorder. *Am. J. Psychiatry* **166**, 74–82 (2009).

103. Aylward, E. H. *et al.* Basal Ganglia Volumes in Children With Attention-Deficit Hyperactivity Disorder. *J. Child Neurol.* **11**, 112–115 (1996).

104. Shaw, P. *et al.* Trajectories of cerebral cortical development in childhood and adolescence and adult attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **74**, 599–606 (2013).

105. Wilens, T. E. *et al.* Presenting ADHD symptoms, subtypes, and comorbid disorders in clinically referred adults with ADHD. *J. Clin. Psychiatry* **70**, 1557–1562 (2009).

106. Ottosen, C. *et al.* Sex Differences in Comorbidity Patterns of Attention-Deficit/Hyperactivity Disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **58**, 412-422.e3 (2019).

107. Leitner, Y. The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know? *Front. Hum. Neurosci.* **8**, 268 (2014).

108. Rosello, B., Berenguer, C., Baixauli, I., Colomer, C. & Miranda, A. ADHD symptoms and learning behaviors in children with ASD without intellectual disability. A mediation analysis of executive functions. *PLoS One* **13**, e0207286 (2018).

109. Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion

syndrome. *Res. Dev. Disabil.* **30**, 763–773 (2009).

110. Ahuja, A., Martin, J., Langley, K. & Thapar, A. Intellectual disability in children with attention deficit hyperactivity disorder. *J. Pediatr.* **163**, 890–5.e1 (2013).

111. DuPaul, G. J., Gormley, M. J. & Laracy, S. D. Comorbidity of LD and ADHD. *J. Learn. Disabil.* **46**, 43–51 (2013).

112. McGillivray, J. A. & Baker, K. L. Effects of Comorbid ADHD with Learning Disabilities on Anxiety, Depression, and Aggression in Adults. *J. Atten. Disord.* **12**, 525–531 (2009).

113. Lin, Y.-J., Lai, M.-C. & Gau, S. S.-F. Youths with ADHD with and without tic disorders: Comorbid psychopathology, executive function and social adjustment. *Res. Dev. Disabil.* **33**, 951–963 (2012).

114. Roessner, V. *et al.* Familiality of Co-existing ADHD and Tic Disorders: Evidence from a Large Sibling Study. *Front. Psychol.* **7**, 1060 (2016).

115. Poh, W., Payne, J. M., Gulenc, A. & Efron, D. Chronic tic disorders in children with ADHD. *Arch. Dis. Child.* **103**, 847–852 (2018).

116. King, S. & Waschbusch, D. A. Aggression in children with attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics* (2010). doi:10.1586/ern.10.146

117. Bartels, M. *et al.* Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3-16 years from multiple raters in six cohorts in the EU-ACTION project. *Eur. Child Adolesc. Psychiatry* **27**, 1105–1121 (2018).

118. Philipp-Wiegmann, F. *et al.* ADHD modulates the course of delinquency: a 15-year follow-up study of young incarcerated man. *Eur. Arch. Psychiatry Clin. Neurosci.* **268**, 391–399 (2018).

119. Bunford, N., Evans, S. W. & Langberg, J. M. Emotion Dysregulation Is Associated With Social Impairment Among Young Adolescents With ADHD. *J. Atten. Disord.* (2018). doi:10.1177/1087054714527793

120. Nierenberg, A. A. *et al.* Clinical and diagnostic implications of lifetime attention-deficit/ hyperactivity disorder comorbidity in adults with bipolar disorder: Data from the first 1000 STEP-BD participants. *Biol. Psychiatry* (2005). doi:10.1016/j.biopsych.2005.01.036

121. Duffy, A. The nature of the association between childhood ADHD and the development of bipolar disorder: A review of prospective high-risk studies. *American Journal of Psychiatry* (2012). doi:10.1176/appi.ajp.2012.11111725

122. Wingo, A. P. & Ghaemi, S. N. A systematic review of rates and diagnostic validity of comorbid adult

attention-deficit/hyperactivity disorder and bipolar disorder. *J. Clin. Psychiatry* **68**, 1776–84 (2007).

123. McIntyre, R. S. *et al.* Attention-deficit/hyperactivity disorder in adults with bipolar disorder or major depressive disorder: Results from the international mood disorders collaborative project. *Prim. Care Companion J. Clin. Psychiatry* (2010). doi:10.4088/PCC.09m00861gry

124. LeBlanc, N. & Morin, D. Depressive symptoms and associated factors in children with attention deficit hyperactivity disorder. *J. Child Adolesc. Psychiatr. Nurs.* (2004). doi:10.1111/j.1744-6171.2004.00049.x

125. Ilbegi, S. *et al.* Substance use and nicotine dependence in persistent, remittent, and late-onset ADHD: a 10-year longitudinal study from childhood to young adulthood. *J. Neurodev. Disord.* **10**, 42 (2018).

126. Wilens, T. E. & Kaminski, T. A. The Co-Occurrence of ADHD and Substance Use Disorders. *Psychiatr. Ann.* **47**, 328–332 (2018).

127. Mariani, J. J. & Levin, F. R. Treatment strategies for co-occurring ADHD and substance use disorders. *Am. J. Addict.* **16 Suppl 1**, 45–54; quiz 55–6 (2007).

128. Wajszilber, D., Santiseban, J. A. & Gruber, R. Sleep disorders in patients with ADHD: Impact and management challenges. *Nature and Science of Sleep* (2018). doi:10.2147/NSS.S163074

129. Hanć, T. & Cortese, S. Attention deficit/hyperactivity-disorder and obesity: A review and model of current hypotheses explaining their comorbidity. *Neurosci. Biobehav. Rev.* **92**, 16–28 (2018).

130. Altfas, J. R. Prevalence of attention deficit/hyperactivity disorder among adults in obesity treatment. *BMC Psychiatry* **2**, 9 (2002).

131. Cortese, S. & Tessari, L. Attention-Deficit/Hyperactivity Disorder (ADHD) and Obesity: Update 2016. *Curr. Psychiatry Rep.* **19**, 4 (2017).

132. Cortese, S. & Peñalver, C. M. Comorbidity between ADHD and obesity: Exploring shared mechanisms and clinical implications. *Postgraduate Medicine* (2010). doi:10.3810/pgm.2010.09.2205

133. Brunault, P. *et al.* Adulthood and childhood ADHD in patients consulting for obesity is associated with food addiction and binge eating, but not sleep apnea syndrome. *Appetite* (2019). doi:10.1016/j.appet.2019.01.013

134. Blum, K. *et al.* Reward deficiency syndrome: a biogenetic model for the diagnosis and treatment of impulsive, addictive, and compulsive behaviors. *J. Psychoactive Drugs* **32 Suppl**, i–iv, 1–112 (2000).

135. Williams, E. D. *et al.* Personality disorder in ADHD Part 1: Assessment of personality disorder in adult ADHD using data from a clinical trial of OROS methylphenidate. *Ann. Clin. Psychiatry* **22**, 84–93 (2010).

136. Olsen, J. L., Reimherr, F. W., Marchant, B. K., Wender, P. H. & Robison, R. J. The effect of personality disorder symptoms on response to treatment with methylphenidate transdermal system in adults with attention-deficit/hyperactivity disorder. *Prim. care companion CNS Disord.* **14**, (2012).

137. Korsgaard, H. O., Torgersen, S., Wentzel-Larsen, T. & Ulberg, R. Personality disorders and Axis I comorbidity in adolescent outpatients with ADHD. *BMC Psychiatry* **16**, 175 (2016).

138. World Drug Report 2018. Available at: https://www.unodc.org/wdr2018/. (Accessed: 7th August 2019)

139. Rösler, M., Retz, W., Yaqoobi, K., Burg, E. & Retz-Junginger, P. Attention deficit/hyperactivity disorder in female offenders: prevalence, psychiatric comorbidity and psychosocial implications. *Eur. Arch. Psychiatry Clin. Neurosci.* **259**, 98–105 (2009).

140. Moukhtarian, T. R., Mintah, R. S., Moran, P. & Asherson, P. Emotion dysregulation in attention-deficit/hyperactivity disorder and borderline personality disorder. *Borderline Personal. Disord. Emot. dysregulation* **5**, 9 (2018).

141. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).

142. Gottesman, I. I. & Shields, J. A polygenic theory of schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 199–205 (1967).

143. Geschwind, D. H. & Flint, J. Genetics and genomics of psychiatric disease. *Science* **349**, 1489–94 (2015).

144. Mefford, H. C., Batshaw, M. L. & Hoffman, E. P. Genomics, Intellectual Disability, and Autism. *N. Engl. J. Med.* **366**, 733–743 (2012).

145. Hyman, S. E. The daunting polygenicity of mental illness: making a new map. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, (2018).

146. Escudero, I. & Johnstone, M. Genetics of Schizophrenia. *Current Psychiatry Reports* **16**, 502 (2014).

147. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–996 (2009).

148. Gejman, P. V., Sanders, A. R. & Duan, J. The role of genetics in the etiology of schizophrenia. *Psychiatric Clinics of North America* **33**, 35–66 (2010).

149. Gibson, G. Rare and common variants: Twenty arguments. *Nature Reviews Genetics* **13**, 135–145 (2012).

150. Wray, N. R. & Maier, R. Genetic Basis of Complex Genetic Disease: The Contribution of Disease Heterogeneity to Missing Heritability. *Curr. Epidemiol. Reports* **1**, 220–227 (2014).

151. Compton, W. M., Thomas, Y. F., Stinson, F. S. & Grant, B. F. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Arch. Gen. Psychiatry* **64**, 566–576 (2007).

152. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

153. Huggett, S. B. & Stallings, M. C. Cocaine'omics: Genome-wide and transcriptome-wide analyses provide biological insight into cocaine use and dependence. *Addict. Biol.* **25**, (2020).

154. Mbarek, H. *et al.* The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168**, 739–748 (2015).

155. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).

156. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).

157. Ormel, J., Hartman, C. A. & Snieder, H. The genetics of depression: successful genome-wide association studies introduce new challenges. *Transl. Psychiatry* **9**, 1–10 (2019).

158. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

159. Ptak, C. & Petronis, A. Epigenetic approaches to psychiatric disorders. *Dialogues in clinical neuroscience* **12**, 25–35 (2010).

160. Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A. & Akbarian, S. Epigenetic Basis of Mental Illness. *Neuroscientist* **22**, 447–463 (2016).

161. Branco, P. R., de Araújo, G. S., Barrera, J., Suarez-Kurtz, G. & de Souza, S. J. Uncovering association networks through an eQTL analysis involving human miRNAs and lincRNAs. *Sci. Rep.* **8**, (2018).

162. Huan, T. *et al.* Genome-wide identification of microRNA expression quantitative trait loci. *Nat. Commun.* **6**, (2015).

163. White, M. J. *et al.* Strategies for Pathway Analysis Using GWAS and WGS Data. *Curr. Protoc. Hum. Genet.* **100**, e79 (2019).

164. J. Rowe, S. & Tenesa, A. Human Complex Trait Genetics: Lifting the Lid of the Genomics Toolbox - from Pathways to Prediction. *Curr. Genomics* **13**, 213–224 (2012).

165. Levin, F. R. *et al.* How treatment improvement in ADHD and cocaine dependence are related to one another: A secondary analysis. *Drug Alcohol Depend.* **188**, 135–140 (2018).

166. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science (80-.).* **360**, 8757 (2018).

167. Swendsen, J. *et al.* Mental disorders as risk factors for substance use, abuse and dependence: Results from the 10-year follow-up of the National Comorbidity Survey. *Addiction* **105**, 1117–1128 (2010).

168. Wang, J.-C., Kapoor, M. & Goate, A. M. The genetics of substance dependence. *Annu. Rev. Genomics Hum. Genet.* **13**, 241–61 (2012).

169. Steinhausen, H.-C., Jakobsen, H. & Munk-Jørgensen, P. Family aggregation and risk factors in substance use disorders over three generations in a nation-wide study. *PLoS One* **12**, e0177700 (2017).

170. Kendler, K. S., Prescott, C. A., Myers, J. & Neale, M. C. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* **60**, 929–937 (2003).

171. Carey, C. E. *et al.* Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front. Genet.* **7**, (2016).

172. Hartz, S. M. *et al.* Association Between Substance Use Disorder and Polygenic Liability to Schizophrenia. *Biol. Psychiatry* **82**, 709–715 (2017).

173. Reginsson, G. W. *et al.* Polygenic risk scores for schizophrenia and bipolar disorder associate with addiction. *Addict. Biol.* **23**, 485–492 (2018).

174. Du Rietz, E. *et al.* Association of Polygenic Risk for Attention-Deficit/Hyperactivity Disorder With Co-occurring Traits and Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 635–643 (2018).

175. Gurriarán, X. *et al.* Relationships between substance abuse/dependence and psychiatric disorders based on polygenic scores. *Genes, Brain Behav.* **18**, e12504 (2019).

176. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genome wide meta-analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *bioRxiv* 528117 (2019). doi:10.1101/528117

177. Derringer, J. *et al.* The aggregate effect of dopamine genes on dependence symptoms among cocaine users: Cross-validation of a candidate system scoring approach. *Behav. Genet.* **42**, 626–635 (2012).

178. Sullivan, D. *et al.* Dopamine transporter DAT and receptor DRD2 variants affect risk of lethal cocaine abuse: a gene–gene–environment interaction. *Transl. Psychiatry* **3**, e222–e222 (2013).

179. Cómbita, L. M., Voelker, P., Abundis-Gutiérrez, A., Pozuelos, J. P. & Rueda, M. R. Influence of the SLC6A3-DAT1 Gene on Multifaceted Measures of Self-regulation in Preschool Children. *Front. Psychol.* **8**, (2017).

180. Tong, J. H. S. *et al.* An association between a dopamine transporter gene (*SLC6A3*) haplotype and ADHD symptom measures in nonclinical adults. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168**, 89–96 (2015).

181. Waldie, K. E. *et al.* Dopamine transporter (DAT1/SLC6A3) polymorphism and the association between being born small for gestational age and symptoms of ADHD. *Behav. Brain Res.* **333**, 90–97 (2017).

182. Franke, B. *et al.* Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD. *Neuropsychopharmacology* **35**, 656–664 (2010).

183. Kampangkaew, J. P. *et al.* Pharmacogenetic Role of *Dopamine Transporter* (*SLC6A3*) Variation on Response to Disulfiram Treatment for Cocaine Addiction. *Am. J. Addict.* **28**, 311–317 (2019).

184. Blum, K. *et al.* Dopamine D2 receptor gene variants: Association and linkage studies in impulsive-addictive-compulsive behaviour. *Pharmacogenetics* **5**, 121–141 (1995).

185. Stolf, A. R. *et al.* Effects of DRD2 splicing-regulatory polymorphism and DRD4 48 bp VNTR on crack cocaine addiction. *J. Neural Transm.* **126**, 193–199 (2019).

186. Beaver, K. M. *et al.* A gene × gene interaction between DRD2 and DRD4 is associated with conduct disorder and antisocial behavior in males. *Behav. Brain Funct.* **3**, (2007).

187. Ribasés, M. *et al.* Candidate system analysis in ADHD: Evaluation of nine genes involved in dopaminergic neurotransmission identifies association with DRD1. *World J. Biol. Psychiatry* **13**, 281–292 (2012).

188. Zhao, J., Chen, C., Bell, R. L., Qing, H. & Lin, Z. Identification of HIVEP2 as a dopaminergic transcription factor related to substance use disorders in rats and humans. *Transl. Psychiatry* **9**, (2019).

189. Srivastava, S. *et al.* Loss-of-function variants in HIVEP2 are a cause of intellectual disability. *Eur. J. Hum. Genet.* **24**, 556–561 (2016).

190. Steinfeld, H. *et al.* Mutations in HIVEP2 are associated with developmental delay, intellectual disability, and dysmorphic features. *Neurogenetics* **17**, 159–164 (2016).

191. Pereira, A. S., Rocha, J. C., Pfundt, R. & Sá, M. J. N. A novel pathogenic mutation in the HIVEP2 gene in a patient with Angelman-like syndrome and hyperphagia. *Eur. J. Paediatr. Neurol.* **21**, e47 (2017).

192. Fernàndez-Castillo, N. *et al.* Association study of 37 genes related to serotonin and dopamine neurotransmission and neurotrophic factors in cocaine dependence. *Genes, Brain Behav.* **12**, 39–46 (2013).

193. Araos, P. *et al.* Serotonin is the main tryptophan metabolite associated with psychiatric comorbidity in abstinent cocaine-addicted patients. *Sci. Rep.* **9**, (2019).

194. Quist, J. F. *et al.* The serotonin 5-HT1B receptor gene and attention deficit hyperactivity disorder. *Mol. Psychiatry* **8**, 98–102 (2003).

195. Quist, J. F. *et al.* Evidence for the serotonin HTR2A receptor gene as a susceptibility factor in attention deficit hyperactivity disorder (ADHD). *Mol. Psychiatry* **5**, 537–541 (2000).

196. Panzenhagen, A. C. *et al.* GENE-SET ANALYSIS OF SEROTONERGIC SYNAPSE GENES IN ADULTHOOD ADHD. *Eur. Neuropsychopharmacol.* **29**, S823 (2019).

197. Ribasés, M. *et al.* Exploration of 19 serotoninergic candidate genes in adults and children with attention-deficit/hyperactivity disorder identifies association for 5HT2A, DDC and MAOB. *Mol. Psychiatry* **14**, 71–85 (2009).

198. Stewart, A. *et al.* Serotonin transporter inhibition and 5-HT 2C receptor activation drive loss of cocaine-induced locomotor activation in DAT Val559 mice. *Neuropsychopharmacology* **44**, 994–1006 (2019).

199. Bruxel, E. M. *et al.* LPHN3 and attention-deficit/hyperactivity disorder: A susceptibility and pharmacogenetic study. *Genes, Brain Behav.* **14**, 419–427 (2015).

200. Arcos-Burgos, M. *et al.* ADGRL3 (LPHN3) variants predict substance use disorder. *Transl. Psychiatry* **9**, (2019).

201. Sibley, M. H. *et al.* Late-onset ADHD reconsidered with comprehensive repeated assessments between ages 10 and 25. *Am. J. Psychiatry* **175**, 140–149 (2018).

202. Newton-Cheh, C. & Hirschhorn, J. N. Genetic association studies of complex traits: Design and analysis issues. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 54–69 (2005).

203. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356–369 (2008).

204. Sullivan, P. F. *et al.* A mega-Analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).

205. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).

206. Licinio, J. Editorial: Messages from hypothesis-driven genotyping: The case of schizoaffective disorder, bipolar type. *Molecular Psychiatry* **15**, 113–114 (2010).

207. Kitsios, G. D. & Zintzaras, E. Genome-wide association studies: hypothesis-"free" or 'engaged'? *Translational Research* **154**, 161–164 (2009).

208. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). *European Monitoring Centre for Drugs and Drug Addiction: European Drug Report 2018: Trends and Developments*. *EMCDDA, Lisbon, June 2018* (2018). doi:10.2810/800331

209. Arias, F. *et al.* Cocaine abuse or dependency and other psychiatric disorders. Madrid study on dual pathology. *Rev. Psiquiatr. y Salud Ment. (English Ed.* **6**, 121–128 (2013).

210. Gelernter, J. *et al.* Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci. *Mol. Psychiatry* **19**, 41–49 (2014).

211. Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5082–5087 (2010).

212. Bevilacqua, L. & Goldman, D. Genetics of impulsive behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, (2013).

213. Johnson, E. O. *et al.* KAT2B polymorphism identified for drug abuse in African Americans with regulatory links to drug abuse pathways in human prefrontal cortex. *Addict. Biol.* **21**, 1217–1232 (2016).

214. Ikeda, M. *et al.* Evidence for shared genetic risk between methamphetamine-induced psychosis and schizophrenia. *Neuropsychopharmacology* **38**, 1864–1870 (2013).

215. Kalsi, G. *et al.* Genome-wide association of heroin dependence in Han Chinese. *PLoS One* **11**, (2016).

216. European Monitoring Centre for Drugs and Drug Addiction. *Country drug report 2017 : Spain*.

217. Wagner, F. A. & Anthony, J. C. From first drug use to drug dependence: Developmental periods of risk for dependence upon marijuana, cocaine, and alcohol. *Neuropsychopharmacology* **26**, 479–488 (2002).

218. Freedman, M. L. *et al.* Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* **36**, 388–393 (2004).

219. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* **11**, 459–463 (2010).

220. Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P. Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics* **6**, 847–859 (2005).

221. Hosking, L. *et al.* Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur. J. Hum. Genet.* **12**, 395–399 (2004).

222. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype Imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).

223. Roshyara, N. R. & Scholz, M. Impact of genetic similarity on imputation accuracy. *BMC Genet.* **16**, 90 (2015).

224. Hong, E. P. & Park, J. W. Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics Inform.* **10**, 117 (2012).

225. Jafari, M. & Ansari-Pour, N. Why, when and how to adjust your P values? *Cell J.* **20**, 604–607 (2019).

226. Johnson, R. C. *et al.* Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* **11**, 724 (2010).

227. Wang, X. *et al.* Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* **5**, (2016).

228. Hommers, L. G., Domschke, K. & Deckert, J. Heterogeneity and Individuality: microRNAs in Mental Disorders. *Journal of Neural Transmission* **122**, 79–97 (2015).

229. Ji, F., Lv, X. & Jiao, J. The Role of microRNAs in Neural Stem Cells and Neurogenesis. *Journal of Genetics and Genomics* **40**, 61–66 (2013).

230. Nowakowski, T. J. *et al.* Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat. Neurosci.* **21**, 1784–1792 (2018).

231. Cheng, J., Liu, A., Shi, M. Y. & Yan, Z. Disrupted glutamatergic transmission in prefrontal cortex contributes to behavioral abnormality in an animal model of ADHD. *Neuropsychopharmacology* **42**, 2096–2104 (2017).

232. Arnsten, A. F. T. The Emerging Neurobiology of Attention Deficit Hyperactivity Disorder: The Key Role of the Prefrontal Association Cortex. *J. Pediatr.* **154**, I-S43 (2009).

233. Wright, C., Turner, J. A., Calhoun, V. D. & Perrone-Bizzozero, N. Potential Impact of miR-137 and Its Targets in Schizophrenia. *Front. Genet.* **4**, 58 (2013).

234. Devanna, P. & Vernes, S. C. A direct molecular link between the autism candidate gene RORa and the schizophrenia candidate MIR137. *Sci. Rep.* **4**, (2014).

235. Lopez, J. P. *et al.* MiR-1202 is a primate-specific and brain-enriched microRNA involved in major

depression and antidepressant treatment. *Nat. Med.* **20**, 764–768 (2014).

236. Miller, J. E., Veturi, Y. & Ritchie, M. D. Innovative strategies for annotating the 'relationSNP' between variants and molecular phenotypes. *BioData Mining* **12**, (2019).

237. Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat. Genet.* **50**, 73–82 (2018).

238. Collins, B. E., Greer, C. B., Coleman, B. C. & Sweatt, J. D. Histone H3 lysine K4 methylation and its role in learning and memory. *Epigenetics Chromatin* **12**, 7 (2019).

239. Hannon, E., Marzi, S. J., Schalkwyk, L. S. & Mill, J. Genetic risk variants for brain disorders are enriched in cortical H3K27ac domains. *Mol. Brain* **12**, 7 (2019).

240. Fryett, J. J., Inshaw, J., Morris, A. P. & Cordell, H. J. Comparison of methods for transcriptome imputation through application to two common complex diseases. *Eur. J. Hum. Genet.* **26**, 1658–1667 (2018).

241. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9**, 1–20 (2018).

242. Hoogman, M. *et al.* Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The Lancet Psychiatry* **4**, 310–319 (2017).

243. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).

244. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

245. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).

246. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1–11 (2017).

247. O'dushlaine, C. *et al.* Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).

248. Huan, T. *et al.* Genome-wide identification of microRNA expression quantitative trait loci. *Nat. Commun.* **6**, 1–9 (2015).

249. McKenzie, M., Henders, A. K., Caracella, A., Wray, N. R. & Powell, J. E. Overlap of expression

Quantitative Trait Loci (eQTL) in human brain and blood. *BMC Med. Genomics* **7**, 31 (2014).

250. Sullivan, P. F., Fan, C. & Perou, C. M. Evaluating the comparability of gene expression in blood and brain. *Am. J. Med. Genet. - Neuropsychiatr. Genet.* **141 B**, 261–268 (2006).

251. Kordas, G., Rudra, P., Hendricks, A., Saba, L. & Kechris, K. Insight into genetic regulation of miRNA in mouse brain. *BMC Genomics* **20**, 849 (2019).

252. Adlakha, Y. K. & Saini, N. Brain microRNAs and insights into biological functions and therapeutic potential of brain enriched miRNA-128. *Molecular Cancer* **13**, 33 (2014).

253. Wake, C. *et al.* Novel microRNA discovery using small RNA sequencing in post-mortem human brain. *BMC Genomics* **17**, 776 (2016).

254. Pomper, N., Liu, Y., Hoye, M. L., Dougherty, J. D. & Miller, T. M. CNS microRNA profiles: a database for cell type enriched microRNA expression across the mouse central nervous system. *Sci. Rep.* **10**, 1–8 (2020).

255. Hammond, S. M. An overview of microRNAs. *Advanced Drug Delivery Reviews* **87**, 3–14 (2015).

256. Soula, A. *et al.* Small RNA-Seq reveals novel miRNAs shaping the transcriptomic identity of rat brain structures. *Life Sci. Alliance* **1**, (2018).

257. Shen, J. J., Wang, Y.-F. & Yang, W. Sex-Interacting mRNA- and miRNA-eQTLs and Their Implications in Gene Expression Regulation and Disease. *Front. Genet.* **10**, 313 (2019).

258. Gong, J. *et al.* Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. *Hum. Mutat.* **33**, 254–263 (2012).

259. Afonso-Grunz, F. & Müller, S. Principles of miRNA-mRNA interactions: Beyond sequence complementarity. *Cellular and Molecular Life Sciences* **72**, 3127–3141 (2015).

260. Nishizaki, S. S. & Boyle, A. P. Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms. *Trends in Genetics* **33**, 34–45 (2017).

261. Huntley, R. P. *et al.* Expanding the horizons of microRNA bioinformatics. *RNA* **24**, 1005–1017 (2018).

262. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

263. Marigorta, U. M., Rodríguez, J. A., Gibson, G. & Navarro, A. Replicability and Prediction: Lessons and Challenges from GWAS. *Trends in Genetics* **34**, 504–517 (2018).

264. Bühler, K. M. *et al.* Common single nucleotide variants underlying drug addiction: More than a decade of research. *Addiction Biology* **20**, 845–871 (2015).

265. Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol. Psychiatry* **19**, 717–723 (2014).

266. Pineda-Cirera, L. *et al.* Evaluation of previous substance dependence genome-wide significant findings in a Spanish sample. *Drug Alcohol Depend.* **187**, 358–362 (2018).

267. Wetherill, L. *et al.* Association of substance dependence phenotypes in the COGA sample. *Addict. Biol.* **20**, 617–627 (2015).

268. McGue, M. *et al.* A genome-wide association study of behavioral disinhibition. *Behav. Genet.* **43**, 363–373 (2013).

269. Greene, C. S., Penrod, N. M., Williams, S. M. & Moore, J. H. Failure to Replicate a Genetic Association May Provide Important Clues About Genetic Architecture. *PLoS One* **4**, e5639 (2009).

270. Bush, W. S. & Moore, J. H. Chapter 11: Genome-Wide Association Studies. *PLoS Comput. Biol.* **8**, (2012).

271. Satterstrom, F. K. *et al.* Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat. Neurosci.* **22**, 1961–1965 (2019).

272. Burrows, E. L. & Hannan, A. J. Decanalization mediating gene-environment interactions in schizophrenia and other psychiatric disorders with neurodevelopmental etiology. *Front. Behav. Neurosci.* **7**, (2013).

273. Liu, C., Faraone, S. V. & Glatt, S. J. Psychiatric Genetics, Epigenetics, and Cellular Models in Coming Years. *J. Psychiatry Brain Sci.* **4**, (2019).

274. Mirkovic, B., Chagraoui, A., Gerardin, P. & Cohen, D. Epigenetics and Attention-Deficit/Hyperactivity Disorder: New Perspectives? *Front. Psychiatry* **11**, 579 (2020).

275. Conrad, D. *et al.* Integrated genetic, epigenetic, and gene set enrichment analyses identify NOTCH as a potential mediator for PTSD risk after trauma: Results from two independent African cohorts. *Psychophysiology* **57**, (2020).

276. Tenenbaum, J. D. *et al.* Translational bioinformatics in mental health: Open access data sources and computational biomarker discovery. *Brief. Bioinform.* **20**, 842–856 (2017).

277. Hannon, E. *et al.* Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* **19**, 48–54 (2015).

278. Hannon, E. *et al.* An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* **17**, 176 (2016).

279. Thapar, A. Discoveries on the genetics of ADHD in the 21st century: New findings and their implications. *American Journal of Psychiatry* **175**, 943–950 (2018).

280. Demontis, D. *et al.* The Role of Ultra-Rare Coding Variants In ADHD. *Eur. Neuropsychopharmacol.* **29**, S724–S725 (2019).

281. Harich, B. *et al.* From Rare Copy Number Variants to Biological Processes in ADHD. *Am. J. Psychiatry* **177**, 855–866 (2020).

282. Bergen, S. E. & Petryshen, T. L. Genome-wide association studies of schizophrenia: Does bigger lead to better results? *Curr. Opin. Psychiatry* **25**, 76–82 (2012).

283. Khan, A., Liu, Q. & Wang, K. iMEGES: Integrated mental-disorder GEnome score by deep neural network for prioritizing the susceptibility genes for mental disorders in personal genomes. *BMC Bioinformatics* **19**, 501 (2018).

284. Mackenzie, G. B. & Wonders, E. Rethinking intelligence quotient exclusion criteria practices in the study of attention deficit hyperactivity disorder. *Frontiers in Psychology* **7**, (2016).

285. Young, S. *et al.* Guidance for identification and treatment of individuals with attention deficit/hyperactivity disorder and autism spectrum disorder based upon expert consensus. *BMC Med.* **18**, 146 (2020).

286. Preston, G. A. & Weinberger, D. R. Intermediate phenotypes in schizophrenia: A selective review. *Dialogues in Clinical Neuroscience* **7**, 165–179 (2005).

287. Flint, J., Timpson, N. & Munafò, M. Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. *Trends in Neurosciences* **37**, 733–741 (2014).

288. Falck-Ytter, T. *et al.* Difficulties maintaining prolonged fixation and attention-deficit/hyperactivity symptoms share genetic influences in childhood. *Psychiatry Res.* **293**, 113384 (2020).

289. Reh, V., Schmidt, M., Rief, W. & Christiansen, H. Preliminary evidence for altered motion tracking-based hyperactivity in ADHD siblings. *Behav. Brain Funct.* **10**, 7 (2014).

290. Pironti, V. A. *et al.* Neuroanatomical abnormalities and cognitive impairments are shared by adults with attention-deficit/hyperactivity disorder and their unaffected first-degree relatives. *Biol. Psychiatry* **76**, 639–647 (2014).

291. Zhou, Z. *et al.* Exploratory locomotion, a predictor of addiction vulnerability, is oligogenic in rats selected

for this phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 13107–13115 (2019).

292. Ducci, F. & Goldman, D. The Genetic Basis of Addictive Disorders. *Psychiatric Clinics of North America* **35**, 495–519 (2012).

293. Bevilacqua, L. & Goldman, D. Genes and addictions. *Clinical Pharmacology and Therapeutics* **85**, 359–361 (2009).

294. de la Fuente, L. *et al.* Structural and functional signatures of executive deficits after early use of cocaine depend upon the route of administration. *medRxiv* 2020.06.12.20129577 (2020). doi:10.1101/2020.06.12.20129577

295. Carta, A. *et al.* Characterization of Clinical Manifestations in the Co-occurring Phenotype of Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder. *Front. Psychol.* **11**, (2020).

296. Hendry, A. *et al.* Atypical Development of Attentional Control Associates with Later Adaptive Functioning, Autism and ADHD Traits. *J. Autism Dev. Disord.* **50**, 4085–4105 (2020).

297. Caspi, A. & Moffitt, T. E. All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry* **175**, 831–844 (2018).

298. Brikell, I. *et al.* The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *Mol. Psychiatry* **25**, 1809–1821 (2020).

299. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).

300. Jang, S. K. *et al.* Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychol. Med.* 1–11 (2020). doi:10.1017/S003329172000272X

301. Lee, P. H., Feng, Y. C. A. & Smoller, J. W. Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. *Biological Psychiatry* **89**, 20–31 (2020).

302. Wang, Q., Yang, C., Gelernter, J. & Zhao, H. Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Hum. Genet.* **134**, 1195–1209 (2015).

303. Palk, A., Illes, J., Thompson, P. M. & Stein, D. J. Ethical issues in global neuroimaging genetics collaborations. *Neuroimage* **221**, 117208 (2020).

304. Davis, L. Mental health and chronic disease: A psycheMERGE investigation into the shared biology underlying psychiatric disorders and their physical comorbidities.

305. Frey, L. J., Lenert, L. & Lopez-Campos, G. EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group. *Yearb. Med. Inform.* **9**, 206–211 (2014).

306. Hripcsak, G. & Albers, D. J. Next-generation phenotyping of electronic health records. *J. Am. Med. Informatics Assoc.* **20**, 117–121 (2013).

307. Etkin, A. A Reckoning and Research Agenda for Neuroimaging in Psychiatry. *Am. J. Psychiatry* **176**, 507–511 (2019).

308. Brückl, T. M. *et al.* The biological classification of mental disorders (BeCOME) study: A protocol for an observational deep-phenotyping study for the identification of biological subtypes. *BMC Psychiatry* **20**, 1–25 (2020).

309. Budde, M. *et al.* A longitudinal approach to biological psychiatric research: The PsyCourse study. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **180**, 89–102 (2019).

310. Lakerveld, J. *et al.* Deep phenotyping meets big data: the Geoscience and hEalth Cohort COnsortium (GECCO) data to enable exposome studies in The Netherlands. *Int. J. Health Geogr.* **19**, 49 (2020).

311. Liu, S. *et al.* Opportunities, risks and challenges in global mental health and population neuroscience: a case of Sino-German cooperation. *European Archives of Psychiatry and Clinical Neuroscience* **1**, 3 (2020).

312. McOmish, C. E., Burrows, E. L. & Hannan, A. J. Identifying novel interventional strategies for psychiatric disorders: integrating genomics, 'enviromics' and gene-environment interactions in valid preclinical models. *British journal of pharmacology* **171**, 4719–4728 (2014).

313. Rees, E. & Owen, M. J. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. *Genome Medicine* **12**, 1–16 (2020).

314. Kaiser, T. & Feng, G. Modeling psychiatric disorders for developing effective treatments. *Nature Medicine* **21**, 979–988 (2015).

315. Czéh, B., Fuchs, E., Wiborg, O. & Simon, M. Animal models of major depression and their clinical implications. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **64**, 293–310 (2016).

316. Stanford, S. C. Some Reasons Why Preclinical Studies of Psychiatric Disorders Fail to Translate: What Can Be Rescued from the Misunderstanding and Misuse of Animal 'Models'? *Altern. Lab. Anim.* **48**, 106–115 (2020).

317. Vadodaria, K. C., Amatya, D. N., Marchetto, M. C. & Gage, F. H. Modeling psychiatric disorders using patient stem cell-derived neurons: A way forward. *Genome Medicine* **10**, 1 (2018).

318. Soliman, M. A., Aboharb, F., Zeltner, N. & Studer, L. Pluripotent stem cells in neuropsychiatric disorders. *Molecular Psychiatry* **22**, 1241–1249 (2017).

319. Di Lullo, E. & Kriegstein, A. R. The use of brain organoids to investigate neural development and disease. *Nature Reviews Neuroscience* **18**, 573–584 (2017).

320. Manchia, M., Pisanu, C., Squassina, A. & Carpiniello, B. Challenges and Future Prospects of Precision Medicine in Psychiatry. *Pharmgenomics. Pers. Med.* **Volume 13**, 127–140 (2020).

321. Diogo, D. *et al.* Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.* **9**, (2018).

322. Kular, L. & Kular, S. Epigenetics applied to psychiatry: Clinical opportunities and future challenges. *Psychiatry Clin. Neurosci.* **72**, 195–211 (2018).

323. Moszyńska, A., Gebert, M., Collawn, J. F. & Bartoszewski, R. SNPs in microRNA target sites and their potential role in human disease. *Open Biology* **7**, (2017).