

# An evolutionary epigenetics approach to schizophrenia

Niladri Banerjee

Thesis for the Degree of Philosophiae Doctor (PhD)  
University of Bergen, Norway  
2018

UNIVERSITY OF BERGEN



# An evolutionary epigenetics approach to schizophrenia

Niladri Banerjee



Thesis for the Degree of Philosophiae Doctor (PhD)  
at the University of Bergen

2018

Date of defence: 28.09.2018

© Copyright Niladri Banerjee

The material in this publication is covered by the provisions of the Copyright Act.

Year: 2018

Title: An evolutionary epigenetics approach to schizophrenia

Name: Niladri Banerjee

Print: Skipnes Kommunikasjon / University of Bergen

# **An evolutionary epigenetics approach to schizophrenia**

**Niladri Banerjee**



Thesis for the degree of Philosophiae Doctor (PhD)  
at the University of Bergen

2018

Date of defence:

□ Copyright Niladri Banerjee

The material in this publication is protected by copyright law

Year: 2018

Title: An evolutionary epigenetics approach to schizophrenia

Author: Niladri Banerjee

Print: Skipnes Kommunikasjon / Universitetet i Bergen

---

## **Scientific Environment**

The present work has been conducted in the Department of Clinical Science, Faculty of Medicine & Dentistry, University of Bergen, Norway where I have been privileged to be part of the Dr Einar Martens group for Biological Psychiatry. As part of the group, I feel deeply thankful to have had the chance to expand my knowledge and understanding of the several aspects of schizophrenia through a multitude of methodologies including clinical, statistical and genetic approaches. Professor Stéphanie Le Hellard acted as my main supervisor with senior researcher Timothy Hughes, PhD at the Oslo University Hospital acting as the first co-supervisor and Tatiana Polushina, PhD at the Department of Clinical Science, University of Bergen as second co-supervisor.

The Martens group is part of the K.G Jebsen Centre for Psychosis, a joint endeavour between the University of Oslo and University of Bergen with the Centre of Excellence(CoE)- Norwegian Centre for Mental Disorders Research (NORMENT): a joint collaboration between University of Bergen, University of Oslo and Oslo University Hospital funded by the Research Council of Norway (RCN). This allowed me national and international level access to knowledge, expertise and networks all involved in dissecting the complex aetiology of mental disorders.

## **Acknowledgement**

This thesis owes a heartfelt thanks to the people and places that shaped everything that occurred both within and outside the lab environment. Nothing in life is isolated, and this thesis is no exception.

First and foremost, I would like to thank my supervisor Professor Stéphanie Le Hellard, who nurtured my inquisitiveness from day one and gave a sense of direction to my curiosity. Words alone are not enough to express my deepest gratitude for her. I may even go in so far as to say that I could not have hoped for a better supervisor to guide me these past four years. She has been truly exemplary in her approach, and for the faith she put in me and my abilities, I can only hope that I have fulfilled at least some of her expectations.

My sincerest thanks also to my co-supervisor, Timothy Hughes, PhD who with his extensive background in bioinformatics methodologies helped out at critical junctures with astute observations and suggestions that resolved the conundrums I faced. I am forever grateful for the direction and focus areas suggested by him that helped me develop skills instrumental in seeing the projects through.

I am also, heavily indebted to my second co-supervisor, Tatiana Polushina, PhD whose skills in statistics and knowledge are akin to the surgeon who is a master at his craft. Thank you for the many hours of boiling down complex statistical techniques into basic ideas I could follow and implement, and for suggesting various statistical approaches to resolve the peculiar bottlenecks I faced.

A warm thanks to Professor Vidar M. Steen and everyone at the Martens group who through the weekly meetings and the annual Martens dinner reminded me about the importance of having good humour in life and in work. I thank you all for being wonderful colleagues who provided a kind and supportive professional environment

---

that allowed me to grow through the years and expand my horizons on the multi-faceted nature of neuropsychiatric disorders.

I reserve special words of thanks to my senior labmate, Sudheer Giddarluru, PhD, who I am thankful for quite literally showing me the ropes when I landed fresh in a strange country that had no English written anywhere and helping me tide over the initial culture shock. I am indebted to your patience and for the many memorable moments discussing work and life in general. A sincere word of thanks also goes out to Anne-Kristin Stavrum, PhD who helped demystify methylation analyses techniques and Carla P.D Fernandes, PhD who showed it is possible to finish a PhD and still remain sane. Acknowledgement and thanks are also owed to the funding support received from the University of Bergen for the PhD position as part of SFF-NORMENT.

The thesis, as alluded to earlier, owes its successful completion to people and events outside the lab as well. Over the course of these last four years, I met and experienced several novel situations, one of the most memorable and potentially life-changing of which was meeting my girlfriend, Tora. You have been an absolute joy and an excellent source of support for me. I know for sure that life here has been much better with your presence in it and I hope life will continue to bless us as we move forward.

To my friends here Torgeir, Øystein, Cathrine and Vilde, thank you for making a stranger from a strange land a small part of your lives. It means a lot to have company and social circle here in Norway, a place where it is not particularly easy to make friends and acquaintances. Special thanks also go out to Dhana, Sam, Divya, Shahin, Irene and Waqas whom I chanced upon accidentally in the final months of my PhD here. It was a welcome surprise and a chance to connect with things one yearns for when one is in a foreign land. To my friends back home, thank you for remembering me in your hearts and reaching out to me. It means the world.



---

I would be remiss to forget my athletics club Fana IL and my coaches Gisle Ellingsen and Endre Raa who provided a supportive training environment and a chance to meaningfully distract myself from the rigours of regular PhD work. My successful completion of the thesis depended as much on the opportunities provided by them and the work-life balance in Norway, as it did on the individual effort that went into this body of work. Thank you for giving me the chance to grow and develop as an athlete and an individual. The focus, discipline and determination I have learnt from the track has had great carryover in seeing this PhD through to the end.

Last but not least, my parents and my grandparents, you are the reason for my existence and for that I can never thank enough nor repay in action nor deed. I will count my blessings to have had the support and encouragement from all of you in taking a leap into the unknown. My only hope is that when the time comes, I am there for you all as you have been for me.

Bergen, June 2018

Niladri Banerjee

कर्मण्येवाधकारस्ते मा फलेषु कदाचन।  
मा कर्मफलहेतुर्भूर्मा ते सङ्गोऽस्त्वकर्मणि २-४७

Karmanye vadhikaraste Ma Phaleshu Kadachana,  
Ma Karmaphalaheturbhurma Te Sangostvakarmani  
Bhagwad Gita, Chapter 2: Verse 47

**Translated,**

- You have the right only to perform the work, but never to its fruits.

Let neither the fruits of action be your motive, nor let your attachment be to  
inaction.

## Summary

Schizophrenia is a psychotic disorder with an estimated lifetime prevalence of ~1% worldwide. Despite reduced fecundity of patients, this rate is stable across population groups separated by geography and time. There is also evidence from the Mesopotamian culture ca 5000 years ago of symptoms that today would be classified as schizophrenia. To explain this stable occurrence of the disorder, the so-called '*Evolutionary hypothesis of schizophrenia*' has been gaining ground. The most well known of which was propositioned by T.J. Crow in 1998, though others including Huxley (1964) and Essen-Möller (1959) have also argued about selective advantages of the disorder. A key assertion made by Crow was the emergence of schizophrenia as a by-product of human evolution, as 'the price humans pay for language'.

In the past two decades, the emergence of genomic technologies and resources have made it possible to test the evolutionary aspect of Crow's hypothesis. There is a growing body of evidence from the field of genomics that suggests human evolution may have played a role in the susceptibility to schizophrenia.

Developments within the last five years have allowed researchers to trace the evolution of epigenomes giving an unprecedented window on gene-environment (GxE) interactions of the past several thousand to millions of years. In the present thesis, I undertake a body of work that investigates the evolutionary question of schizophrenia from this new field of evolutionary epigenetics. We first test whether human-specific methylated regions, determined in comparison to Neanderthals and Denisovans are enriched for schizophrenia markers. These methylated regions represent at least 750,000 years of evolution since the last common ancestor diverged from Neanderthals and Denisovans ( **Paper I** ). This was followed up by investigating primate-methylated regions that represent at least 13 MYA of epigenomic evolution ( **Paper II** ). Finally, we investigate whether human-specific methylated regions, as

---

defined in the first study are amenable to methylation variation in patients with schizophrenia ( **Paper III** ).

We find evidence that recent evolution denoted by human-specific methylated regions tracing ~750,000 years of methylation development are enriched for schizophrenia markers. Primate methylation markers are not enriched for schizophrenia variants with the exception of the extended Major-Histocompatibility Region ( MHC ) region. Finally, we find evidence of methylation disruption in brain samples of patients with schizophrenia in regions that underwent human-specific methylation evolution.

Our results provide support that recent evolution, denoted by methylation changes since the divergence of the common ancestor of humans, Neanderthals and Denisovans, harbour more schizophrenia associated markers than expected by chance, and thus, may have played a role in susceptibility to schizophrenia at a group-level.

---

## Publication List

### Paper I

Banerjee N, Polushina T, Bettella F, Giddaluru S, Steen VM, Andreassen OA, Le Hellard S. **Recently evolved human specific methylated regions are enriched in schizophrenia signals**, *BMC Evolutionary Biology*, 2018, 18:63, <https://doi.org/10.1186/s12862-018-1177-2>

### Paper II

Banerjee N, Polushina T, Bettella F, Giddaluru S, Steen VM, Andreassen OA, Le Hellard S. **Analysis of differentially methylated regions in primates and non-primates provides support for the evolutionary hypothesis of schizophrenia**, *Schizophrenia Research* ( Submitted March 2018, Under review )

### Paper III

Banerjee N, Polushina T, Stavrum AK, Steen VM, Le Hellard S. **Functional analyses of evolutionary human-specific methylated regions in schizophrenia patients** ( Manuscript )

---

## Abbreviations

<i>Word</i>	<i>Full form</i>
450k array	Illumina Infinium HumanMethylation450 BeadChip array
A	Adenine
AA	Arachidonic Acid
ADHD	Attention Deficit Hyperactivity Disorder
<i>BDNF</i>	Brain Derived Neurotrophic Factor
BMI	Body Mass Index
bp	base pairs
BCE	Before Common Era
BPD	Bipolar Disorder
ca	circa
C	Cytosine
CE	Common Era
<i>CERS3</i>	Ceramide Synthase 3
CNV	Copy Number Variation
CpG	Cytosine-phosphate Guanine di-nucleotide
DBP	Diastolic blood pressure
<i>DDX43</i>	DEAD-Box Helicase 43
DHA	Docosahexaenoic acid
<i>DPPA5</i>	Developmental Pluripotency Associated 5
<i>DISC1</i>	Disrupted-in-schizophrenia 1
DMP	Differentially methylated position
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
<i>DNMT1</i>	DNA Methyltransferase 1
DSM	Diagnostic & and Statistical Manual of Mental Disorders
ecdf	Empirical cumulative distribution function
e.g.	exempli gratia (for example)
etc.	et cetera
EWAS	Epigenome-wide association study

---

---

FDR	False Discovery Rate
<i>FOSB</i>	Finkel-Biskis-Jinkins murine osteosarcoma viral oncogene homolog B
<i>FGFR1</i>	Fibroblast growth factor receptor 1
G	Guanine
GABA	Gamma-amino-butyric acid
GEO	Gene expression omnibus
GWAS	Genome-wide association study
GxE	Gene by environment interaction
HAR	Human accelerated region
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
hg18	Human genome build 18
hg19	Human genome build 19
HDL	High-density lipoprotein
ICD	International Statistical Classification of Diseases and Related Health Problems
i.e.	id est (that is)
INRICH	Interval enRICHment Analysis
IPA	Ingenuity pathway analyses
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
<i>LINC00606</i>	Long Intergenic Non-Protein Coding RNA 606
<i>LY6G5C</i>	Lymphocyte antigen 6 family member G5C
MAF	Minor allele frequency
<i>MAOA</i>	Monoamine oxidase A
<i>MARK3</i>	Microtubule Affinity Regulating Kinase 3
<i>MEK1</i>	Mitogen-activated protein kinase kinase 1
MHC	Major histocompatibility complex

---

---

MYA	Million years ago
NIMH	National Institute of Mental Health
NSS	Neanderthal Selective Sweep
<i>PRDM9</i>	PR-domain containing protein 9
PUFA	Polyunsaturated Fatty Acid
QQ	Quantile-quantile
$r^2$	Squared correlation coefficient (linkage disequilibrium)
RA	Rheumatoid arthritis
RNA	Ribonucleic acid
<i>REC8</i>	REC8 meiotic recombination protein
<i>RELN</i>	Reelin
RDC	Research Diagnostic Criteria
RDoC	Research Domain Criteria
SBP	Systolic blood pressure
SCZ	Schizophrenia
SNP	Single nucleotide polymorphism
<i>SOX10</i>	Sex-determining Y-box 10
T	Thymine
TC	Total cholesterol
TG	Triglycerides
UCSC	University of California, Santa Cruz
USA	United States of America
WHO	World Health Organization
YA	Years Ago
<i>ZSCAN12P1</i>	Zinc finger and SCAN domain containing 12 pseudogene 1

---



---

**Table of Contents**

<b>SCIENTIFIC ENVIRONMENT .....</b>	<b>III</b>
<b>ACKNOWLEDGEMENT.....</b>	<b>IV</b>
<b>SUMMARY .....</b>	<b>VIII</b>
<b>PUBLICATION LIST .....</b>	<b>X</b>
PAPER I.....	X
PAPER II.....	X
PAPER III .....	X
<b>ABBREVIATIONS.....</b>	<b>XI</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 SCHIZOPHRENIA .....	1
1.1.1 <i>Historical perspectives</i> .....	1
1.1.2 <i>From Dementia praecox to Schizophrenia</i> .....	2
1.1.3 <i>The modern diagnostic system(s)</i> .....	3
1.2 BRIEF PRIMER ON GENETIC STUDIES OF SCHIZOPHRENIA .....	6
1.2.1 <i>Genetic studies on schizophrenia</i> .....	6
1.2.2 <i>Genome-Wide Association Studies(GWAS) on schizophrenia</i> .....	7
1.3 EPIGENETICS IN SCHIZOPHRENIA .....	8
1.3.1 <i>General Introduction to epigenetics</i> .....	8
1.3.2 <i>DNA methylation and schizophrenia</i> .....	12
1.3.3 <i>Gene x Environment (GxE) interactions</i> .....	16
1.4 EVOLUTIONARY HYPOTHESIS OF SCHIZOPHRENIA .....	17
1.5 WHY ANALYSE THROUGH THE PRISM OF EVOLUTIONARY EPIGENETICS .....	19
<b>2. STUDY AIMS .....</b>	<b>23</b>
<b>3. METHOD SUMMARIES .....</b>	<b>24</b>
3.1 DATA ACQUISITION .....	24
3.1.1 <i>Differentially Methylated Region (DMR) data</i> .....	24

---

3.1.2 Human Accelerated Region (HAR) data .....	25
3.1.3 Neanderthal Selective Sweep (NSS) data .....	26
3.1.4 GWAS data .....	26
3.1.5 Methylation datasets.....	27
3.2 DATA ANALYSES .....	27
3.2.1 SNP assignment to DMRs.....	27
3.2.2 Conditional Quantile-Quantile (QQ) plots .....	27
3.2.3 INRICH.....	28
3.2.4 Pathway analysis.....	29
3.2.5 Selection of DMRs for patient analyses .....	29
3.2.6 Methylation analyses.....	29
3.2.7 Statistical significance of DMRs .....	31
3.2.8 Visualisation of DMRs.....	32
3.2.9 Annotation of evolutionarily enriched DMRs with Roadmap Epigenomics Data.....	32
<b>4. PAPER SUMMARIES .....</b>	<b>33</b>
4.1 PAPER I .....	33
4.2 PAPER II.....	34
4.3 PAPER III.....	35
<b>5. DISCUSSION &amp; CONCLUSION.....</b>	<b>36</b>
5.1 METHODOLOGICAL CONSIDERATIONS.....	36
5.1.1 Mapping of DMRs and SNPs with LDsnpR.....	36
5.1.2 Enrichment analyses with conditional Quantile-Quantile (QQ) plots .....	38
5.1.3 Enrichment analyses with INRICH .....	39
5.1.4 Evolutionary DMR analyses in patient samples.....	41
5.1.5 Methylation analyses in brain and blood samples .....	43
5.1.6 Limitations of the study .....	50
5.2 BIOLOGICAL SIGNIFICANCE OF OUR RESULTS .....	52
5.2.1 Discussion of results in Papers I-III.....	52

---

5.2.2 <i>An epigenetic perspective of the results in the broader context of recent human evolution</i> .....	55
5.3 CONCLUSION & FUTURE PERSPECTIVES .....	60
<b>6. REFERENCES</b> .....	<b>62</b>
<b>7. PAPERS I-III</b> .....	<b>89</b>

## **1. Introduction**

### **1.1 Schizophrenia**

Schizophrenia is a debilitating mental disorder that causes considerable loss in quality of life to the patient affected [1–3] as well as significant costs to society [4–8]. The term ‘schizophrenia’ was first coined by Eugen Bleuler in 1908 to classify a constellation of symptoms observed in the disorder using the Greek roots *schizen* (‘to split’) and *phren* (‘mind, spirit, soul’) [9–11].

It can be diagnosed as per the Diagnostic and Statistical Manual for Mental Disorders 5 (DSM 5) by the presence of characteristic symptoms such as delusions, hallucinations or disorganised speech in addition to catatonia or avolition. Additionally, there is a marked decrease in social and occupational function such as work, self-care and interpersonal relations. For a diagnosis, symptoms must be present for six months, including at least one month of characteristic symptoms. It is also necessary that schizoaffective or bipolar disorder with psychoses is ruled out, and that the patient is not abusing any drugs. Furthermore, if autism is present, then it needs to be accompanied by hallucinations for at least one month [12].

#### *1.1.1 Historical perspectives*

The modern definition of schizophrenia owes a lot to the work of Eugen Bleuler and Emil Kraepelin from late 19<sup>th</sup>-early 20<sup>th</sup> century. However, they were neither the first to observe the disorder, nor the first to describe its symptoms. The disorder has accompanied mankind at least since humans started keeping written records [13,14].

---

Cuneiform tablets of Mesopotamia from the third millennium BCE [13,14] describe some of the earliest examples of symptoms that today would be classified as schizophrenia. The ancient Hindu texts of Vedas (ca 1400 BCE) provide descriptions of psychoses that is distinct from manic-depressive illness [14,15]. Other descriptions lie in the satires of Horace from the first century BCE and the writings of Roman physicians Caelius Aurelianus and Alexander of Tralles from the fifth and sixth century CE. The first century CE Hindu treatise on Ayurvedic medicine, ‘*Charaka Samhita*’ also describes symptoms akin to schizophrenia [16]. Reviewing the literature from the Middle Ages in Europe led Nigel M. Bark to conclude the existence of schizophrenia for at least 2000 years [17–19].

### *1.1.2 From Dementia praecox to Schizophrenia*

Industrialization in Europe in the 18<sup>th</sup>-19<sup>th</sup> centuries led to an increased medical observation of patients with mental illnesses [18]. Several physicians from across Europe, therefore, took a keen interest in the development, diagnoses and nomenclature of the disorder. In 1801, Phillipe Pinel from France laid the groundwork for what Emil Kraepelin later termed ‘dementia praecox’ and Eugen Bleuler, ‘schizophrenia’. Pinel used the word ‘*démence*’, meaning ‘loss of mind’ to describe the degradation of mental faculties of chronically ill, hospitalised mental patients. The word itself had been used in French literature since 1381 to describe the condition of mental deterioration [14]. In 1847, Millingen emphasised the post-pubertal onset of the disease [20]. To highlight this aspect of the disorder, Benedict Augustin Morel coined the French term ‘*démence précoce*’ in 1852 that literally means ‘loss of mind at a young age’ that differentiated it from ‘*démence senilis*’ or ‘loss of mind due to old age’.

In 1893, Emil Kraepelin introduced the term ‘dementia praecox’ in the 4<sup>th</sup> edition of his book [21]. Amongst the major contributions of Kraepelin were his division of psychotic illnesses into two categories: dementia praecox and manic-depressive insanity [14]. He made the observation that age of onset, familial history and premorbid personalities could

---

distinguish between the two kinds of illnesses and emphasised the presence of hereditary factors in his patients with dementia praecox. Furthermore, he was the first to challenge the notion of schizophrenia as a European phenomenon caused by industrialisation. He observed the remarkable similarity of symptoms in cases as far away as Singapore and in populations of Chinese, Malay, Tamils and Japanese [22]. He stated the necessity of finding the real cause of a disorder that did not depend on race, climate or food [14,22].

The Swiss psychiatrist Eugen Bleuler, however, differed with Kraepelinian concepts of the disorder [14]. He asserted that not all patients displayed symptoms at a young age, nor were they all equally likely to be predisposed to deterioration of their mental faculties [11]. Furthermore, Bleuler was influenced by the works of Wundt, Freud and Jung in his approach to mental illnesses [11,23]. He, therefore, brought forward the psychological aspects of the disorder in his work with Carl Jung [24] and introduced the concept of ‘schizophrenia’ from ‘*schizen*’ meaning ‘to split’ and *phren* meaning the ‘mind’ [11]. While expanding the scope of the disorder, Bleuler’s work also helped characterise the disorder not as a single entity but rather as a heterogeneous mixture of several disorders- ‘a genus, not a species’ [14,25].

Finally, Kurt Schneider [26], a German psychiatrist, paved the way for a systematic diagnostic system that was replicable between different psychiatrists (inter-rater reliability) [27] and whose principles were incorporated into current psychiatric diagnostic criteria [14].

### *1.1.3 The modern diagnostic system(s)*

The modern understanding of the disorder as described in Diagnostic and Statistical Manual for Mental Disorders (DSM), the International Statistical Classification of Diseases and Related Health Problems (ICD), as well as the Research Diagnostic Criteria (RDC), incorporates several concepts not only from Kraepelin and Bleuler, but also from Schneider [14].

---

The RDC system was established in the 1970s [28–31] and had strict diagnostic criteria that excluded labelling patients as schizophrenics if they exhibited major affective disorder or other borderline disorders. Furthermore, the age of onset and duration of symptoms was also limited to forty years and six months respectively. These features helped improve the reliability of the psychiatric diagnostic systems in existence at the time [32,33]. This system was instrumental in the development and refinement of the diagnostic system from the American Psychiatric Association- the ‘Diagnostic and Statistical Manual of Mental Disorders (DSM)’. Before the RDC, the DSM-II had vague descriptions of the symptoms and diagnostic criteria [32,33]. Subsequent development of DSM-III [34] incorporated several elements from RDC and eventually became the *de-facto* standard for diagnoses of psychiatric illnesses [32].

The subsequent revision to DSM-IV [35] established one of the most widely used diagnostic criteria with high reliability and fair validity [36,37] that was also used in one of the largest genetic investigations of the disorder [38]. The system emphasises the positive (hallucinations, delusions) [39,40] and negative symptoms of the disorder ( avolition, anhedonia, affective flattening, alogia) while age restrictions have been dropped [14]. In this diagnostic system, the symptoms must have persisted over the previous six months with at least one month of hallucinations and delusions [14,35]. The latest iteration is the DSM-V [12,41]. It retains much of the diagnostic criteria used to define schizophrenia in DSM-IV but removes the classic schizophrenia subtypes, defines in further detail the relationship of schizophrenia to catatonia, expands upon the differences between schizophrenia and schizoaffective disorder and eliminates the prioritisation of Schneiderian ‘first-rank symptoms’ [12].

Parallel to the definitions and diagnostic criteria of the American Psychiatric Association, the World Health Organization (WHO) developed its own set of diagnostic criteria- the International Statistical Classification of Diseases and Related Health Problems (ICD), currently in its tenth iteration, ICD-10 [42,43]. This is a system that was entrusted to the

---

WHO, upon its creation in 1948 and has undergone several revisions [44]. The eleventh revision, ICD-11 is due for release in 2018 [44]. The ICD system is much broader in scope, includes criteria for the classification of other diseases and is widely used around the world for reporting mortality and morbidity statistics in contrast to the DSM that is limited only to mental disorders [44]. The system has good descriptive validity for schizophrenia [45] and has been more widely used in the clinical setting than DSM-IV that finds more use in research environments [46,47].

Research Domain Criteria (RDoC) is the third initiative for diagnosis of mental disorders like schizophrenia. It is published by the National Institute of Mental Health (NIMH), USA and has a significantly different approach compared to either DSM or ICD [48–50]. It is an initiative that seeks to incorporate new findings from the field of genetics and neuroscience into the classification of mental disorders. The initiative was launched due to a disconnect between clinically diagnosed schizophrenia cases and the pathophysiology being discovered by genetic and neuroscience approaches. The ultimate aim of the RDoC initiative is to significantly influence the diagnostic system of the future by integrating precise biological markers into the definition, classification and treatment of mental disorders [50].

In summary, considerable effort is being put into diagnostic methods for psychiatric illnesses including schizophrenia. Many of these diagnostic methods share similarities, yet each system has its unique strengths and perspectives. The inter-rater reliability of the diagnostic systems also varies considerably [51,52]. This is a measure of how consistently a person would get diagnosed with a particular illness by different psychiatrists. For example, Cheniaux et al. [52] found that schizophrenia is diagnosed twice as frequently using ICD-10 compared to DSM-IV. Furthermore, the inter-rater reliability from a theoretical maximum of 1 was only 0.59 and 0.56 for DSM-IV and ICD-10 respectively. Thus, the same person might get a different diagnosis when analysed by a different psychiatrist or diagnostic system. The DSM-V fares even worse than DSM-IV with an inter-rater reliability of only 0.46 for schizophrenia [51]. This



---

makes the genetic and epigenetic study of schizophrenia challenging- genetic and epigenetic studies focus on discrete, quantifiable and measurable units (genes and chemical changes to the genome), unlike the diagnostic system used to categorise the patients in the first place. The findings in this thesis must be understood with this caveat in mind.

## **1.2 Brief primer on genetic studies of schizophrenia**

### *1.2.1 Genetic studies on schizophrenia*

Initial efforts to understand the genetics of schizophrenia relied on twin [53–56], family [15,57–59] and adoption studies [60,61]. These studies revealed a heritability between 60-90% [56,62,63] indicating a strong genetic component to the aetiology of the disorder. However, twin studies revealed a concordance rate of 40-60% for monozygotic twins [56,64,65] suggesting genetic factors did not entirely drive the disorder.

Additionally, the advent of cytogenetic techniques, such as karyotyping enabled researchers to perform linkage analyses that investigates global chromosomal abnormalities amongst patients with schizophrenia. Chromosomal abnormalities have frequently been associated with the disorder [66] and include copy-number variations (CNVs) [66–69].

Prior to the sequencing of the whole human genome [70,71] research in the genetics of schizophrenia was often informed by the neurobiology of the disorder [72,73]. As such the research was hypothesis-driven with a focus on candidate genes [74] and regions involved in the regulation of pathways implicated in the aetiology of schizophrenia such as dopamine [75,76], serotonin [77,78], glutamate [79,80] and GABA [81].

---

Although potential candidate genes involved in the aetiology of the disorder have been identified [82], such as *DISC1* (Disrupted in schizophrenia 1) [83–85], most of the genes identified in these studies have not been robust to subsequent replication studies [82,86,87]. This lack of replication at a population level may suggest that the genes implicated affect only the particular families studied. However, as recently reported, the evidence against the use of candidate gene approaches for schizophrenia genetics is conclusive due to the chance nature of the association of the top candidate genes [88,89].

Partly due to lack of success in determining consistent genes in the aetiology of the disorder and partly due to improvements in technology [90,91], a new wave of studies analysing the entire genome, namely genome-wide association study (GWAS) was performed [38,92–95].

### *1.2.2 Genome-Wide Association Studies(GWAS) on schizophrenia*

After the sequencing of the human genome [70,71,96] and mapping of variation in human genomes at a population level [97], efforts were made to identify genome-wide candidates in the aetiology of schizophrenia. These GWAS have identified more than hundred different genetic loci associated with the disorder, giving support to the notion of schizophrenia as a complex disorder [38,92–95].

The biggest difference between the GWAS approach and candidate-gene studies is that GWAS is hypothesis-free [98,99]. The methodology employed involves genotyping millions of single-nucleotide polymorphisms (SNPs) on genotyping microarrays. Frequency differences of specific SNPs between patients and controls help associate variants with the disease [100,101]. Thus, compared to candidate-gene approaches where hypothesis-driven loci are analysed, GWAS approaches allow one to interrogate the variation across the entire human genome including genic and non-coding regions. This helps derive new insights into the biology of the disorder [102].

---

Results of most GWAS do not point to direct, causal associations between SNPs and disease. Instead, the majority of associations are non-causal, due to the phenomenon of linkage-disequilibrium (LD) between SNPs [100]. LD is the phenomenon where groups of markers tend to be inherited together more often than expected by chance due to common population ancestries [100]. LD is usually quantified by  $r^2$  that denotes the correlation between two markers [100]. Furthermore, most of the significant markers in GWAS results are found in non-coding regions of the genome [103–107].

The latest GWAS on schizophrenia revealed 108 loci associated with the disorder from which more than eighty were previously unreported [38]. These loci together contain 128 SNPs that cross the genome-wide significance level ( $p < 5 \times 10^{-8}$ ). The loci provide support for known neurotransmitters implicated in the disorder such as those involving the dopaminergic and glutamatergic systems. Furthermore, the loci are not randomly distributed but show enrichment in enhancers of brain tissues and immune systems. The authors note the convergence of pathways from those obtained with rare-variant analyses and genome-wide approaches. Additionally, as observed in other GWAS, only a limited number of protein-coding variants were found [38].

One way to gain insight into the role of non-coding regions in the aetiology of diseases is to utilise epigenetic approaches [103,108,109] that are discussed in the subsequent sections.

## **1.3 Epigenetics in schizophrenia**

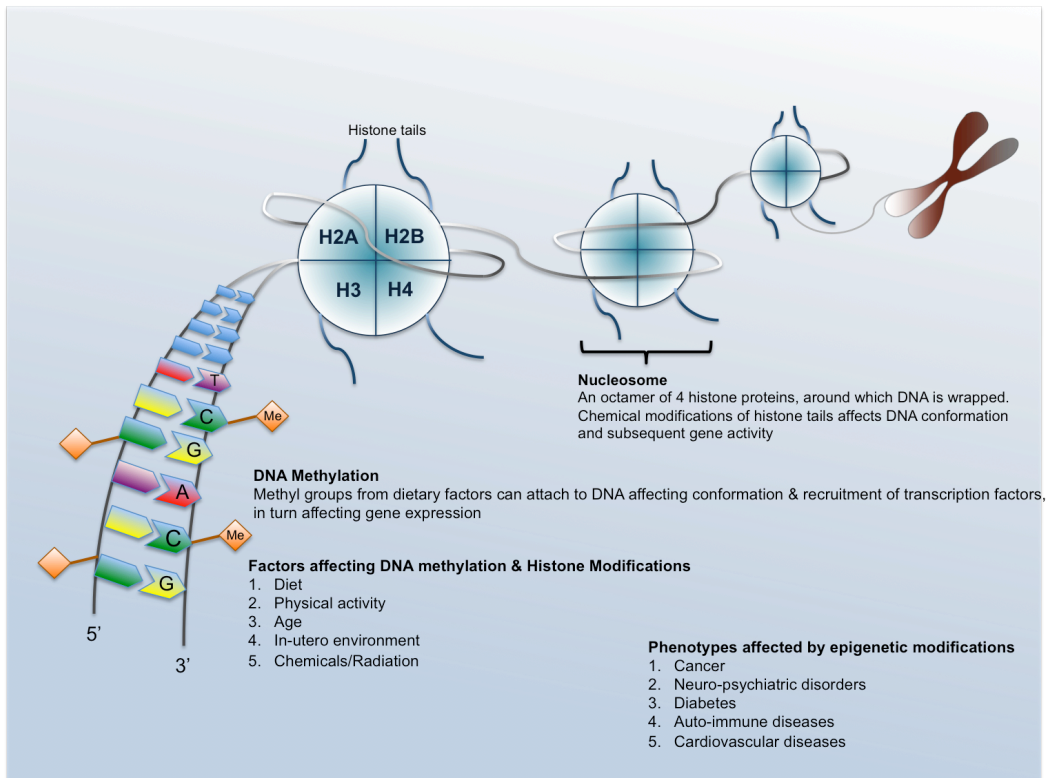
### *1.3.1 General Introduction to epigenetics*

Much work is going on to understand the genome [70,71,96,110,111], transcriptome [112–114] and proteome [113,115–118]. However, another layer of cellular information, stored in the form of epigenome has begun to attract attention and focused research [109].

---

This field has grown in importance in the last decade, especially after the sequencing of the human genome [70,71,119] and development of new sequencing technologies [120–122]. Epigenetics is the study of heritable changes in gene regulation that occurs independently of the DNA sequence [123,124]. ‘Heritable’ in epigenetics implies stable transmission of chemical marks from one cell division to another. Epigenetic phenomena include molecular processes such as DNA methylation, histone modifications and chromatin re-organisation. These modifications are preserved when cells divide [125]. The study of these processes on a genome-wide scale is referred to as epigenomics.

The main effect of epigenomic modification is to affect the organisation of the chromatin, which in turn affects gene expression or repression [126]. Indeed, researchers have begun looking at gene expression from the perspective of higher order DNA organisation, i.e. the chromatin. The chromatin is a protein-DNA complex comprised of DNA wrapped around a group of packaging proteins called histones, forming bead-like structures called nucleosomes. These nucleosomes subsequently help condense the DNA sufficiently to fit within the dimensions of a nucleus.



*Figure 1: Depiction of various epigenetic factors affecting chromatin organisation.*

Shown here is an expanded view of condensed chromatin found in a chromosome. Nucleosomes comprise of histone proteins whose tails are susceptible to chemical modifications. Individual CpG dinucleotides on the DNA strand can undergo methylation that influences DNA conformation that subsequently can regulate gene expression (Adapted from Brown WM, 2015 [127]).

There are five different histone proteins named H1, H2A, H2B, H3 and H4. H2A-H2B form a dimer while H3 and H4 form a tetramer giving rise to an octamer around which 146bp of DNA are wrapped. This forms a single nucleosome. H1 acts as a linker histone which holds the spool of DNA in place and locks it together with the rest of the nucleosome. These histone proteins have an amino acid tail, and the N-terminal ends of these amino acid tails are subject to chemical modifications which are added via post-translational modifications [128]. These modifications act as unique signatures determining gene regulation [109] and understanding these chemical signatures is one of

---

the key research areas in epigenetics. Thus, it is possible to predict which genes are going to be activated or repressed based on the presence or absence of these chemical groups.

In addition, DNA methylation also provides another layer of control over chromosomal conformation [129]. DNA methylation was the first epigenetic modification to be characterised [130]. In this particular epigenetic modification, methyl groups are added to the 5' ends of cytosine residues in Cytosine-phosphate Guanine (CpG) dinucleotides to form 5-methylcytosine [130]. The methylated CpG binding proteins such as meCP1 and meCP2 subsequently bind to such methylated CpG dinucleotides [131]. These proteins have DNA binding domains and a transcription repression domain and can additionally recruit other factors that condense the chromatin resulting in the formation of heterochromatin that represses gene activity [132–134].

DNA methylation is a highly stable and heritable epigenetic mark [123,131,135]. Post-implantation, *de-novo* methyltransferases DNMT3a and DNMT3b add methyl marks to the embryo [136]. Subsequent maintenance of methylation is carried out by another group of methyltransferases called DNMT1 with specificity to *hemi*-methylated DNA strands, i.e. the daughter DNA strands which are produced after each round of DNA replication [131].

DNA methylation is implicated in a whole range of functions such as transcriptional silencing of genes [137], genome imprinting [138,139], regulation of tumour suppressor genes [140–142], X-chromosome inactivation [131,139], maintaining genomic integrity through silencing of transposons [143] and repetitive regions [131,144–146]. The promoter regions of many genes have CpG dinucleotides clustered into what is known as CpG islands [131]. Such CpG islands are usually protected from DNA methylation [131,144]. In healthy human cells, tumour suppressor genes are often *hypo*-methylated, and intergenic regions and repetitive elements are *hyper*-methylated [131,137]. In cancerous cells, the reverse is usually observed, i.e. genome-wide *hypo*-methylation and CpG island *hyper*-methylation [137–140]. This brings to attention the increasing

---

relevance of studying DNA methylation in the context of diseases [137,146]. Among the unique characteristics of DNA methylation is the variation that occurs in the regions methylated between tissues [147], individuals [148] and related species [149]. Such regions are usually abbreviated as DMRs for differentially methylated regions. These variations are thought to contribute to phenotypic variations between people and may therefore potentially provide a way to describe susceptibility to complex diseases. Additionally, these regions have been found in or proximal to regions of recent evolution, indicating their role in the evolution of modern-day humans [150].

### *1.3.2 DNA methylation and schizophrenia*

Efforts to understand the epigenetic machinery and especially DNA methylation in schizophrenia has primarily followed two approaches: candidate genes and genome-wide methodologies. Amongst candidate genes, aberrant DNA methylation has been implicated at the reelin gene (*RELN*) [151–154], brain-derived neurotrophic factor (*BDNF*) [155], DNA Methyltransferase 1 (*DNMT1*) [156] and Sex-determining Y-box containing gene 10 (*SOX10*) [157].

Reelin is produced in GABA-ergic neurons and is important for correct neuronal positioning during brain development [158]. It has been found to be hyper-methylated in the brain [153] as well as blood samples [152] of patients with schizophrenia. Mouse models of hyper-methylation of this gene have found a downregulation in Reelin gene expression levels [159,160]. A hypothesis involving increased activation of *DNMT1* leading to higher DNA methyltransferase activity has been proposed to explain the aberrant methylation at *RELN* [154,158]. Indeed, it has been observed that *DNMT1* also has aberrant methylation in patients with schizophrenia. Specifically, DNA-methyltransferase 1 is found to be overexpressed in the cortex of patient samples [156]. This increased activation of DNMT1 mirrors the subsequent downregulation of reelin observed in GABA-ergic neurons in patients [161].

---

Downregulation of BDNF has also been observed in schizophrenia patients [155,162,163]. BDNF is important for nerve cell survival, neural differentiation and synaptic plasticity [164]. Significant hyper-methylation was found in peripheral blood cells from Japanese patients [155]. However, an earlier study found an opposite direction of effect with patients from Iran showing hypo-methylation compared to controls [165]. This discrepancy could be due to different methods employed or due to population differences.

Abnormal DNA methylation has also been observed for *SOX10* gene. It is an oligodendrocyte-specific transcription factor [166], and oligodendrocyte dysfunction has been observed in patients with schizophrenia [157]. Hyper-methylation of CpG island of *SOX10* is associated with concomitant downregulation of the gene [157].

However, these candidate gene studies investigating the role of methylation in the disorder are not without their drawbacks. For example, different methods have been used to measure the methylation in candidate gene studies. This, however, may reflect the lack of a standardised technology in the then-nascent field, instead of poor study design. Furthermore, most of the candidate gene methylation studies were conducted on an insufficient number of sample sizes. Early studies on the *RELN* gene were conducted with only 5 [153] to 10 samples per group [151]. Furthermore, different studies analysed different regions of the brain making it challenging to make broad generalisations of methylation patterns in patient samples.

Improvements in technology and cost helped initiate epigenome-wide analyses (EWAS) of methylation patterns in patients with schizophrenia. Unlike the variant association findings in candidate-gene studies of schizophrenia which could not be replicated in genome-wide studies, a majority of findings from methylation studies using candidate-genes have been replicated in EWAS [166]. In addition, genome-wide scans of methylation have revealed novel pathways and genes involved in the aetiology of schizophrenia [166].



---

As an example, the first EWAS for patients with schizophrenia by Mill and colleagues, [167] confirmed the role of GABAergic pathways in the aetiology of schizophrenia. Additionally, they found variation in regions associated with glutamatergic pathways as well as antipsychotic mediated effects on DNA methylation levels in the promoter region of Mitogen-activated protein kinase kinase 1 (*MEK1*) gene. Compared to candidate gene studies investigating the role of methylation, the sample size was considerably higher (n = 105, 35 cases with schizophrenia, 35 cases with bipolar disorder and 35 controls). However, the technology used at the time was not standardised and has now been superseded by microarrays from Illumina.

Another paper by Xiao and colleagues [168] undertook a comprehensive look at the methylome and transcriptome of 2 brain regions: frontal cortex and anterior cingulate in patients with schizophrenia and those with bipolar disorders. They found significant differences in the methylation levels of these two regions between controls and patients with schizophrenia/bipolar disorder. Most interesting perhaps was their observation that the differentially methylated regions (DMRs) could accurately distinguish between cases and controls whereas differentially expressed genes could not.

Similarly, Wockner and colleagues [169] performed EWAS in the post-mortem brains from 24 schizophrenia samples and 24 healthy controls. Using Illumina 450k array, they found significant methylation differences in more than 4000 probes across nearly 3000 genes. It is known that antipsychotics can affect DNA methylation [170–172]. In this study all but two of the patients were consuming antipsychotics, so the effect of medication on the DNA methylation levels cannot be ruled out.

Subsequently, another group also published EWAS for patients with schizophrenia from two different brain sample datasets [173]. Combined analyses by Wockner and colleagues [174] across the three different datasets [169,173] revealed differentially methylated regions (DMRs) in *CERS3*, *DPPA5*, *PRDM9*, *DDX43*, *REC8*, *LY6G5C* that

---

were consistently differentially methylated in patients with schizophrenia across the three different brain datasets.

A recent review by Teroganova and colleagues [175] summarised the available evidence for DNA methylation alterations in schizophrenia from both candidate-gene and epigenome-wide approaches. They report that good agreement exists between candidate genes analysed for aberrant methylation and results from EWAS.

However, a significant challenge in the field is the tissue heterogeneity of brain. Epigenetic marks such as DNA methylation are cell and tissue-specific [135]. However, in the studies described previously, the methylation signal is obtained from various regions of the brain that makes a direct comparison between studies challenging. Another challenge when performing methylation studies to study schizophrenia is the lack of information on antipsychotic medications. This is because they are known to influence DNA methylation [176]. Additionally, it is useful to know the demographic background of the patient samples since this can influence DNA methylation as well [177].

To account for some of these limitations, two analyses were recently published with a large sample size that also included information on the cellular composition and patient demographics - one using post-mortem brain tissue [178] and the other using whole blood [179]. Jaffe and colleagues [178] analysed post-mortem prefrontal cortex brains from 191 patient samples and 240 controls and found small but significant hypomethylation across more than 2000 probes in patients with schizophrenia. Furthermore, these probes were found to be enriched amongst a set of probes differentially methylated in fetal life [178] supporting a neurodevelopmental model of schizophrenia that suggests aetiological origins before birth [173].

Hannon and colleagues [179] conducted the largest ( $n = 1714$ ) EWAS of methylation differences in patients with schizophrenia performed to date using whole blood samples. The study focused on finding differentially methylated positions (DMPs) between

---

patients with schizophrenia and controls across three different cohorts – a ‘discovery cohort’ with  $n=675$  (patients = 353, controls = 322); a ‘replication cohort’ with  $n = 847$  (patients = 414, controls = 433) and 97 monozygotic twin pairs discordant for schizophrenia. In the discovery cohort, they identified over 1000 probes that were differentially methylated at a  $p$ -value  $< 10e-5$  that were robust to methylomic variation from covariates such as cell composition, age, sex and smoking. Results from the replication and twin cohort showed high consistency in number and direction of methylation of top DMPs. A meta-analysis of the three datasets revealed a total of 343 DMPs at an experiment wide-threshold of  $p$ -value  $< 10e-7$  [179].

To identify differentially methylated regions (DMRs), they used two different methods that resulted in 12 to 76 DMRs. Some of these DMRs occurred in known regions of association such as the major histocompatibility locus (MHC) on chromosome 6 [179]. Furthermore, they observed that several of these DMRs had a combined  $p$ -value that was more significant than the individual probes contained within the DMRs highlighting, the importance of finding extended regions that are differentially methylated in patients with schizophrenia [179].

The current consensus in the field is that patients with schizophrenia show global hypo-methylation with local hyper-methylation of specific genes and that administration of antipsychotics partially reverses the global hypo-methylation [166]. In our studies, we focused on DNA methylation because it is the most widely studied of all epigenetic changes [175]. Furthermore, as DNA methylation patterns can be inherited [123,132,180], this is the epigenetic modification that has been most thoroughly investigated from an evolutionary standpoint [149,150,181–184].

### *1.3.3 Gene x Environment (GxE) interactions*

Central to the premise of investigating the effect of methylation variation on the susceptibility to schizophrenia is the concept of gene by environment interactions (GxE).

---

Complex phenotypes, such as schizophrenia, can be influenced by both genetic and environmental factors. GxE analyses aim at understanding how the interaction between variation at the genotypic level and exposure to environmental risk can influence a phenotype. Different genotypes can cause different phenotypes even in the exposure to the same environmental stressors [185]. One of the first reports to look at such GxE interactions found evidence of the role of genotype in altering the phenotype in the presence of the same external stimuli of childhood violence. The study found that a polymorphism in the promoter region of the monoamine oxidase A enzyme gene (*MAOA*) modulates the susceptibility of developing aggressive tendencies in adulthood: those producing more MAOA were less susceptible to the detrimental effects of maltreatment [186]. However, investigating the role of GxE in a complex neuropsychological disorder like schizophrenia has been difficult because of the polygenic nature of the disorder. Existing literature points to the influence of environmental factors such as childhood trauma, cannabis consumption, migration and urban living environment as environmental risk factors for schizophrenia [187]. How exactly these environmental risks are influencing the development of the disorder remains unknown. Some studies have shown that environmental factors influence gene expression through epigenomic modifications. For instance, it is well known that physical exercise changes metabolism via epigenetic modifications [188,189]. However, the influence of environmental risk factors on the risk for schizophrenia through epigenetic mechanisms warrants future research.

#### **1.4 Evolutionary Hypothesis of Schizophrenia**

One of the most intriguing aspects of the disorder has been the remarkable persistence throughout recorded human history [13]. The incidence is roughly 1% [190,191] and is stable across populations separated by geography and time [192,193]. There are written records from the Mesopotamian culture ca 5000 years ago (YA) [13] that in today's time would be classified as schizophrenia symptoms.

---

Thus, the disorder has been maintained in the human population despite reductions in fecundity of patients with schizophrenia [194,195]. The analyses of Swedish birth registry data with more than 2 million individuals born between 1950 to 1970 revealed that male schizophrenia patients had on average a nearly 80% reduction in the number of children [194] compared to the general population while female schizophrenia patients had half as many as the general population [194]. To explain the persistence of a disorder with such a substantial effect on fitness, the evolutionary hypothesis of schizophrenia was put forward. While some have argued for the existence of selective advantages of the disorder [190,196], the version that is most well known today is attributed to Timothy J. Crow [192,193,197,198].

T.J. Crow argued the disorder to be a by-product of human evolution through language [192,193]. To explain this, he proposed that lateralisation is fundamentally involved in the pathophysiology of schizophrenia [192,193]. Lateralization and cerebral asymmetry of the brain is the concept wherein specific cognitive tasks tend to be more dominant in one hemisphere over the other. These cognitive tasks, for example, include language [199]. According to Crow and the so-called *bi-hemispheric theory* [193,200–203], the asymmetry of brain development permitted the separation of hemispheric function that enabled the evolution of language [192]. While lateralisation is observed in many species [204], it is the assertion of lateralisation of language that forms a key component of his evolutionary hypothesis.

He makes the observation that patients with schizophrenia often have abnormal brain structure especially reduced lateralisation [192,193]. He asserts that it is this abnormal lateralisation that subsequently affects both language development and language perception in patients with the disorder such that they suffer from a dissociation between thought and speech that ultimately leads to the auditory hallucinations [192,193,205]. Thus, the involvement of language is seen by him as a key feature of schizophrenia, indeed identifying it as a disorder of language [192]. Crow, therefore, lays down the

---

fundamentals of the evolutionary hypothesis by arguing schizophrenia to be the ‘price humans pay for language’ [198].

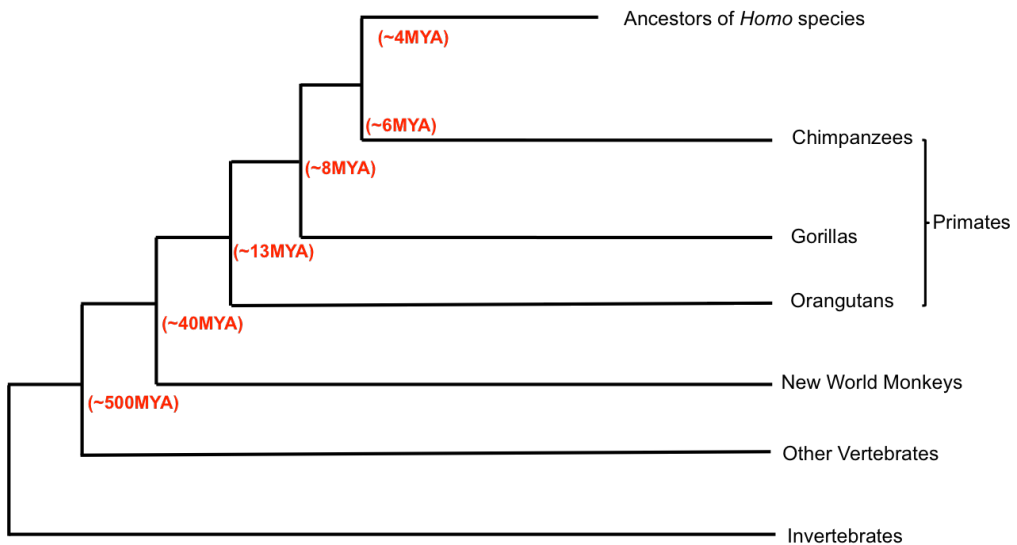
### **1.5 Why analyse through the prism of evolutionary epigenetics**

It is a challenge to design experiments that can test T.J. Crow’s assertion of the emergence of language during human evolution as a risk factor for schizophrenia. The technologies and resources available allow the possibility to investigate regions of the genome that have evolved [206–208] and test for enrichment of schizophrenia markers in regions specific to human evolution [209–211]. This does not directly test the language component of Crow’s hypothesis.

Genomic methods allow determination of genetic level changes on evolutionary timelines, but an important mediator of genetic changes is the environment which is impossible to investigate using genomic approaches alone [212,213]. In present-day species, the environmental effects on the genome are often investigated via *epi*-genomic approaches [214,215]. This is because the environment interacts with the genome through chemical modifications of the genome that affect gene regulation and gene expression [214,215]. Such environmental effects often leave long-lasting changes [216] that in certain species have been observed to persist for as many as 14 generations [217]. It is, therefore, now possible to trace the development and maintenance of epigenomic (and potentially environmental changes) for multiple generations in present-day species [216,217].

However, if one were to be interested in tracing the development of epigenome across historical time (and as a proxy the influence of environmental factors), it becomes imperative to use evolutionary epigenetic approaches that allow comparison of epigenomes from different time points [149,150,181]. Recently, efforts have been made to trace the evolution of epigenomes from primates to humans [149], humans with other extinct hominids [150], and even between ancient and modern humans [182] using

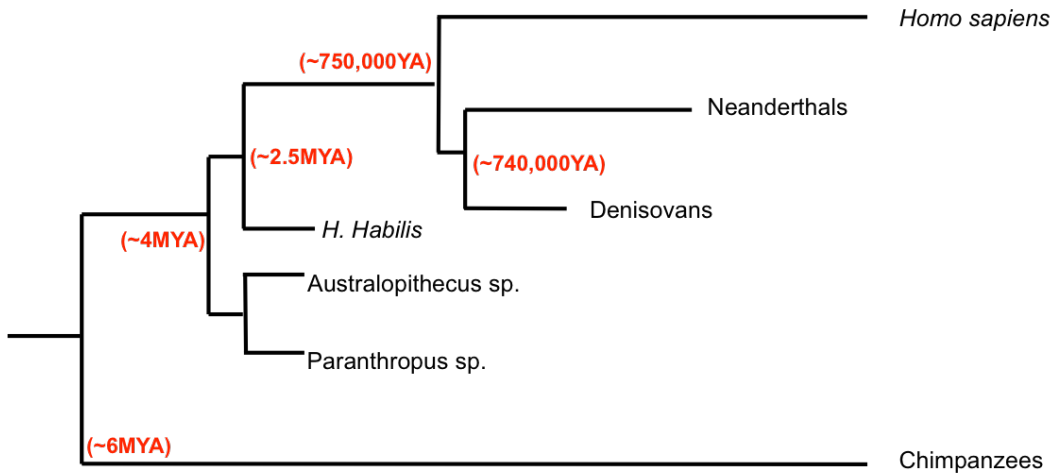
mainly methylation data [149,150,182]. So, while the first anatomically modern humans emerged at least 300,000 years ago [218,219] and can be inferred to be human, it is known that modern human capabilities such as abstract thinking and organised societies did not develop until at least 70,000 years ago [220–227]. If we assume that the genome of the 300,000-year-old human fossil is similar to a modern-day human genome, then it is possible that GxE interactions mediated through epigenetic changes [215] could be one of the driving factors responsible for the emergence of intellectual abilities, language and possibly psychosis.



(Timeline estimates obtained from Smithsonian, Hasegawa et al 1985, Srinivasan et al 2017)

Figure 2: Depiction of a phylogenetic tree with estimated times of divergence

*A phylogenetic tree providing a general view of evolution since the emergence of invertebrates. Except for the ancestors of the Homo lineage, all other species have survived to the present age. The tree of ancestors of Homo lineage is expanded in the next figure.*



(Timeline estimates obtained from Smithsonian, Rogers et al 2017)

*Figure 3: Depiction of the expanded phylogenetic tree of the ancestors of Homo species.*

*All ancestral species of Homo sapiens are extinct except for chimpanzees that survive to the present age. Chimpanzees, as such, are our closest living relatives.*

In this thesis, to test the evolutionary hypothesis of schizophrenia, we primarily use two different kinds of evolutionary methylation data and determine if they are enriched for schizophrenia markers. These methylation data contain regions of the human genome whose methylation levels have changed over evolutionary time frames. By checking for the enrichment of these methylated regions for schizophrenia markers, we can then determine if such regions where methylation levels have changed over evolutionary time in humans are localised in regions implicated in the aetiology of schizophrenia. We broadly categorise the two different methylation datasets as primate and non-primate differentially methylated regions (DMRs).

The non-primate DMRs were obtained from the dataset of Gokhman et al., 2014 [150]. These are a group of methylated regions whose levels have changed over evolutionary time between modern humans, Neanderthals and Denisovans. These regions depict



---

changes in the methylation pattern since the divergence from the last common ancestor between the three hominids, ~750,000 YA [150].

The primate DMRs are another set of DMRs obtained from Hernando-Herraez et al. [149]. These trace the evolution of methylation through primate history from orangutans, through gorillas, chimpanzees and finally humans. Since the orangutans diverged from the human lineage ca 13 MYA and chimpanzees ca 6 MYA, these regions allow observation of methylation developments that took place in the lineage between 6 to 13 MYA. While these datasets would not allow one to test whether language and schizophrenia are linked with evolution as per T.J. Crow, they would, however, allow one to investigate whether epigenomic evolution has taken place in genomic regions implicated in the aetiology of schizophrenia.

Thus, using evolutionary epigenomic approaches allows one to not only investigate how methylomes evolved as species diverged from one another but also opens a window of research to investigate potential environmental changes that would accompany such epigenetic changes [182]. The latter aspect can be especially crucial in determining when humans developed superior intellectual capacities such as language and, according to T.J. Crow, the vulnerability to psychosis that came along with it [198].

## 2. Study Aims

The overarching goal of our analyses was to uncover evolutionary epigenetic loci that are enriched for association with schizophrenia.

GWAS studies have already implicated several genomic loci associated with schizophrenia while the various DMRs point to regions of the genome important from an evolutionary perspective. Methylation can influence the underlying genome sequence and conversely, the genome sequence may influence methylation. Therefore, we used both GWAS and DMR data, to test whether regions of the genome where methylation changes occurred over evolutionary time (DMRs) localise with markers associated with schizophrenia.

In order of progression, we sought to address this in the following way:

1. Determine whether evolutionary human-specific DMRs are enriched for human traits and phenotypes including schizophrenia. **(Paper I)**
2. Determine whether primate DMRs show enrichment of association with schizophrenia markers. **(Paper II)**
3. Determine whether evolutionary DMRs as defined in **Paper I** and containing significantly associated schizophrenia markers, shows a different methylation pattern in patients with schizophrenia, compared to healthy controls. **(Paper III)**

---

### 3. Method Summaries

A consolidated summary of all the methods and datasets are given below in two broad categories (data acquisition and data analyses). The individual **Papers I-III** provide the details.

#### 3.1 Data Acquisition

All three papers extensively used publicly available datasets that are described below.

##### 3.1.1 *Differentially Methylated Region (DMR) data*

Coordinates of differentially methylated regions (DMRs) were obtained from two primary sources. For analyses presented in **Paper I**, the data was obtained from Supplementary Table S2 of Gokhman et al., 2014 [150]. This file contained DMRs inferred by comparing genome sequence of fossilised Neanderthal and Denisovan limb samples with methylation data from osteoblasts of modern humans. From the genomes of the Neanderthal and Denisovan samples, Gokhman and colleagues inferred methylation by utilising the natural degradation of methylated cytosine (C) to thymine (T). Unmethylated cytosines degrade to uracils, which were removed in the sequencing reads of Neanderthals and Denisovans [150]. Using the published high coverage genomes of Neanderthal [228] and Denisovan [229], the pre versus post-mortem Cs and Ts were determined to create a C/T ratio [150]. This was used to infer methylation in Neanderthals and Denisovans. The methylation information, in the form of C/T ratio, was then compared with each of the three species and classified according to the hominid in which the methylation change occurred, i.e. human-specific, Neanderthal-specific and Denisovan-specific DMRs. These DMRs do not represent tissue-specific methylation but species-specific methylation [150]. DMRs that could not be classified reliably (unclassified DMRs) [150] were not used.

---

The second source of DMRs was from Hernando-Herraez et al. [230] that was used in **Paper II**. The DMRs were determined by comparing the methylation profile from peripheral blood samples of orangutans, chimpanzees and gorillas with humans. This dataset contained hypo-methylated and hyper-methylated DMRs for each of the species. A pairwise comparison was performed to determine the DMRs, i.e. for a group of methylated regions in species A, regions that had lower (hypo-methylated) or higher methylation levels (hyper-methylated) in other species were found. If a particular region in species-A was consistently hypo-methylated while it was hyper-methylated in other species, that region was termed as ‘species-A hypo-methylated region’. Similarly, hyper-methylated regions for species-A was determined. Since our interest is in human evolution, we analysed human hypo and hyper-methylated DMRs for the enrichment of schizophrenia and other traits (**Paper II**). In comparison to DMRs from **Paper I** that comprised of a comparison between modern humans and extinct hominids, the DMRs in this paper are all from extant species. These DMRs can be inferred to represent methylation changes since the point of divergence of the common ancestors of primates from humans. Thus, they represent an evolutionary course of history spanning at least 13 MYA from orangutans [231,232] to 6 MYA when the common ancestors of chimpanzees and humans diverged from one another [231,232]. For distinguishing these DMRs from those used in **Paper I**, we refer to them as primate DMRs.

### *3.1.2 Human Accelerated Region (HAR) data*

These are a group of regions in the human genome that are selectively accelerated. They were first described by Pollard et al. [206] by comparing human genomes with chimpanzee genomes to determine regions that remained conserved in chimpanzees but evolved rapidly in humans. Subsequent research by other groups has significantly expanded this list [208].

Genomic coordinates were obtained from publicly available data (docpollard.com/2x) for three classes of human accelerated region: HARs, in which regions conserved in

---

mammals are accelerated in humans; PARs, in which regions conserved in mammals are accelerated in primates; and pHARs, in which regions conserved in primates (but not other mammals) are accelerated in humans.

### *3.1.3 Neanderthal Selective Sweep (NSS) data*

Just like HARs describe the selective evolution of human genome in comparison with primates and other mammals, the Neanderthal selective sweep (NSS) describes the particular selection of specific nucleotides in the human genome over the Neanderthal genome. It thus offers evolutionary insight at a genomic level on regions of the human genome that likely diverged since splitting up from the last common ancestor of humans and Neanderthals [233,234].

NSS data was obtained from Srinivasan et al. [209] as a list of markers with corresponding NSS values. Markers with negative values, indicating positive selection in humans, were selected and used for analysis.

### *3.1.4 GWAS data*

In our analyses, we made extensive use of publicly available results from genome-wide association studies (GWAS). These studies analyse at a population level the frequency differences between groups of markers (single-nucleotide polymorphisms, SNPs) for specific phenotypes using a case-control study design. The final results from such studies usually provide information on SNPs with an association value in the form of the  $p$ -value. Summary statistics from thirteen different phenotypes were obtained from their respective published GWAS studies : schizophrenia (SCZ) [38], bipolar disorder (BPD) [235], attention deficit hyperactivity disorder (ADHD) [236], rheumatoid arthritis (RA) [237], blood lipid markers (high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides (TG), total cholesterol (TC)) [238], blood pressure (systolic blood

---

pressure (SBP), diastolic blood pressure (DBP) [239], body mass index (BMI) [240], height [241] and intelligence [242].

### *3.1.5 Methylation datasets*

Publicly available methylation datasets from GEO for brain at accession numbers GSE61107 [169], GSE6143 [173], GSE61380 [173], GSE74193 [178] and for blood at GSE80417 [179] were used. The analysis of methylation datasets requires a formatted text file, commonly called ‘Sample Sheet’ that details the information about the samples such as age, gender, specific well in the probe used for analysis etc. This was created with in-house scripts.

## **3.2 Data Analyses**

### *3.2.1 SNP assignment to DMRs*

SNPs were assigned to DMRs with LDsnpr [243] using positional binning and LD (linkage disequilibrium)-based binning in R [244]. We used both methods because DMR-localized SNPs that were not genotyped in a specific GWAS would be missed if we used positional binning alone [243].

### *3.2.2 Conditional Quantile-Quantile (QQ) plots*

QQ plots are a useful tool to visualise the spread of data and any deviations from expected null distributions. They are frequently utilised in GWAS to depict enrichment of true signals. When the observed distribution matches the expected distribution, a line of equality is obtained that depicts the null hypothesis. If the observed and expected distributions differ, there will be a deviation from this null line. Due to the extremely low

---

$p$ -values in GWAS, it is common to depict  $p$ -values by converting them to negative  $\log_{10}$  values so that with smaller  $p$ -values, higher negative logarithmic values are obtained. We plotted the negative  $\log_{10}$  of observed  $p$ -values against the expected negative  $\log_{10}$  of a normal distribution. Leftwards deflections from the null line represent enrichment because this suggests a higher than expected chance of the observed data possessing true signals [245]. Conditional Q-Q plots are similar to quantile-quantile (QQ) plots but depict distributions of data ‘conditioned’ on particular criteria such as genomic annotation. Thus, a specific stratum of data, e.g.  $p$ -values from a specific category of SNPs is compared against the whole dataset. In our case, we were interested to investigate the strata of data containing schizophrenia markers classified as being present within and in linkage disequilibrium (LD) with the DMR regions, compared to all the schizophrenia markers from the full GWAS. Enrichment occurred if the data distribution in the SNPs ‘conditioned’ on DMRs showed greater leftward deflection than what is observed for all GWAS markers.

### 3.2.3 INRICH

INterval EnRICHment Analysis (INRICH) is a robust bioinformatics pipeline to determine enrichment of genomic intervals implicated by LD with predefined or custom gene sets [246]. It takes into account several potential biases that can otherwise lead to false positives. It is well suited for testing GWAS-implicated SNPs for association with gene sets as it controls for variable gene size, SNP density, LD within and between genes, and overlapping genes with similar annotations. We followed the procedure described previously [211], with the extended MHC region (chr6:25-35Mb) masked and SNPs with minor allele frequency (MAF) <0.05 excluded in the schizophrenia GWAS.

---

### 3.2.4 Pathway analysis

Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) from QIAGEN ([www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity), last accessed 26<sup>th</sup> August 2016). The reference set was Ingenuity Knowledge Base (Genes). Both direct and indirect relationships were analysed. All data sources were included with the confidence parameter set to experimentally observed and highly predicted pathways for humans. Under the category of ‘Tissues & Cell Lines’, we performed the analysis once with all organ systems and once with only the nervous system. 5338 enriched DMR SNPs were mapped to 349 unique RefSeq genes and 446 RefSeq genes in LD using the method of Schork et al. [245]. Genes in LD blocks containing enriched NSS markers were determined similarly. 4276 enriched NSS markers mapped to 648 overlapping RefSeq genes and 1363 RefSeq genes in LD. IPA was performed on these gene-lists.

### 3.2.5 Selection of DMRs for patient analyses

In **Paper III**, we use a total of 9 DMRs from more than 800 analysed in **Paper I** to test if they show disrupted methylation in patient samples. These 9 DMRs were selected for containing SNPs associated with schizophrenia based on a study-wide significance threshold using the method of Moskvina et al. [247]. The study-wide threshold with Bonferroni correction was determined to be  $2.4 \times 10^{-5}$  at  $\alpha = 0.05$ . SNPs were found at or above this threshold spread across 9 DMRs: *DMR10*, *DMR127*, *DMR203*, *DMR204*, *DMR236*, *DMR237*, *DMR291*, *DMR526* and *DMR527*.

### 3.2.6 Methylation analyses

The techniques for analysing methylation data from Illumina BeadChip platforms such as the 450k array [248,249] and the latest EPIC array [250] is an evolving field with several methods proposed for optimal pre-processing and normalisation of data [251–253].



---

In **Paper III**, altogether five different publicly available methylation datasets for schizophrenia patients were analysed. Four of these datasets were obtained from the prefrontal cortex of deceased schizophrenia patients while the fifth dataset was from peripheral blood samples.

Raw IDAT files were available for the brain methylation dataset from Wockner et al. [169]. The preprocessing step was similar to the one reported by them [169]. Samples with missing age information were removed along with probes that failed at a detection  $p$ -value < 0.05 in 50% of the samples. The original analyses [169] also removed the probes on the sex chromosomes but, this was not done for the analyses in **Paper III** as none of the evolutionary DMRs selected were present on them. Additionally, due to the extremely sparse probe coverage in the DMRs, we performed the analysis without removing probes with SNPs. Subsequent normalisation was done using functional normalisation [251].

For the datasets from Pidsley et al. [173], only the normalised beta matrices were available. These were used directly in downstream analysis. However, two of the DMRs lacked probes: *DMR526* and *DMR527*. These DMRs have only one probe each in the 450k array. It is possible that Pidsley et al. [173] removed them during the pre-processing step for their analyses.

The dataset from Jaffe et al. [178] initially comprised of more than 650 samples including prenatal samples. Since our interest in **Paper III** was to determine methylation variation in patients schizophrenia compared to a control population of similar age distribution, all pre-natal samples were excluded. The final analysis was performed on 191 schizophrenia samples and 231 control samples.

For the analysis from blood samples using the discovery cohort from Hannon et al. [179] with 675 samples, the normalised beta matrix was directly used.

---

To determine the statistical significance of differentially methylated positions (DMPs), linear regression modelling with the *limma* [254] package in R [244] along with empirical Bayes [255] in the *minfi* package [256] was implemented on the M values of individual probes. M values were obtained from the beta values as follows:

$$M \text{ value} = \log_2\left(\frac{\text{Beta}}{1 - \text{Beta}}\right)$$

### 3.2.7 Statistical significance of DMRs

After determining the statistical significance of the individual probes, it was essential to find a way to determine if the evolutionarily enriched DMRs show statistically significant methylation variation in patient samples compared to controls. Traditional methods of determining large regions of genomes with methylation variation rely on software such as *bumphunter* [257].

However, in our case, since the DMRs were already predefined, we utilised the *comb-p* algorithm [179,258] that can determine the statistical significance of custom regions. The algorithm is especially useful for array data due to the irregularly spaced nature of the probes. It calculates the correlation between probes and combines adjacent *p*-values using the Stouffer-Liptak-Kechris corrections [258]. Briefly, this correction weighs down a given *p*-value based on the correlation with nearby probes. We used the default parameters in the *region\_p* programme [258] in this algorithm, to determine the statistical significance of methylation variation within the evolutionary DMRs. This programme also performs a Sidak one-step correction for multiple testing. We use the results from the Sidak correction to ascertain whether a given DMR shows statistically significant methylation difference in patient samples.

---

### 3.2.8 Visualisation of DMRs

To visualise the methylation variation in the evolutionary DMRs showing statistically significant difference between patients and controls, we utilize the *Gviz* [259] package in R. Amongst the many useful features of the package is the ability to visualise custom data tracks with trendlines and annotation from most data tracks in the UCSC table browser [260]. Annotation information on the probes in the 450k array was obtained from Zhou et al. [253]. CpG Island information was downloaded from the AnnotationHub package [261]. SNP data and RefSeq data in hg19 was downloaded from UCSC Table Browser [260].

### 3.2.9 Annotation of evolutionarily enriched DMRs with Roadmap Epigenomics Data

The evolutionary DMRs selected for containing schizophrenia-associated SNPs above the study-wide threshold were also annotated using the data available from the Roadmap Epigenomics consortia [262]. Prefrontal cortex data was available from sample E073 that was obtained from healthy controls aged 75 and 81 years old. The 18-state chromatin data was loaded in R and intersected with the coordinates of the 9 DMRs with the GenomicRanges package [263]. Information about the 18-states was downloaded from the Roadmap Epigenomics data portal.

## 4. Paper Summaries

### 4.1 Paper I

In this paper, we show that regions whose methylation levels changed specifically in the human branch compared to Neanderthals or Denisovans are enriched for markers of schizophrenia.

We come to this conclusion by analysing various phenotypes and testing for enrichment both visually and statistically.

For visual confirmation of enrichment, we implement the method from Schork et al. [245] that was also used successfully to depict enrichment in the Neanderthal selective sweep by Srinivasan et al. [209].

We did not find any evidence for enrichment of Neanderthal and Denisovan specific methylated regions for schizophrenia markers. However, the method may be suboptimal as the GWAS markers are based on the structure of human genome that may not be identical to Neanderthals and Denisovans.

Our paper lends evidence to the evolutionary hypothesis of schizophrenia and suggests recent human evolution at play.

---

## 4.2 Paper II

In this paper, we wanted to analyse if primate DMRs that denote methylation changes from ca 13 MYA could be enriched for schizophrenia markers. We undertook this study because we were interested in checking how far back the evolutionary lineage do methylation changes potentially remain enriched for schizophrenia (SCZ) markers. This could give clues to whether schizophrenia markers predated the Homo lineage.

We separately undertook analyses in hypo- and hyper-methylated human regions, i.e. regions in the human genome that were either a) hypo-methylated in humans and hyper-methylated in primates or b) hyper-methylated in humans and hypo-methylated in primates.

We found evidence of enrichment for hypo-methylated human DMRs driven by the extended MHC region and a considerable reduction in enrichment when the MHC region was masked. This suggests that the risk loci for SCZ in the extended MHC region are of an ancient origin and that they tend to have lost methylation over evolutionary time. When we remove the extended MHC region from the analyses, we lose the hypo-methylated DMRs and the markers contained therein.

In contrast, regions, which became hyper-methylated in humans compared to primates showed no evidence for enrichment in MHC region. The enrichment effect of hypo-methylated DMRs was confirmed with INRICH for LD-implicated SCZ intervals below the genome-wide threshold.

Furthermore, there was a negligible overlap of the primate DMRs with other evolutionary annotations including the human-specific DMRs from **Paper I**, HARs and NSS. This may suggest that different regions of the human genome underwent evolution and that these changes occurred at different evolutionary time points.

---

### 4.3 Paper III

In this study, we investigated whether the evolutionary DMRs (from **Paper I**) that contain significantly associated schizophrenia markers above a study-wide threshold, also show variable methylation in patients with schizophrenia. We undertook the analyses using publicly available methylation datasets in brain and blood samples.

We find statistically significant evidence of variable methylation in evolutionary DMRs in patient samples from the brain. The findings, however, could not be replicated in blood samples. This may be because schizophrenia is considered a brain disorder rather than a blood disorder.

The DMRs showing variable methylation in patient samples occur in genes previously implicated in schizophrenia such as *ZSCAN12P1*, *LINC00606*, and *FGFR1*. These genes may be relevant from an evolutionary standpoint since they contain methylation regions important in human evolution. Finally, we find a major limitation of the Illumina 450k BeadChip platform that has inadequate coverage of several of the evolutionary DMRs.

---

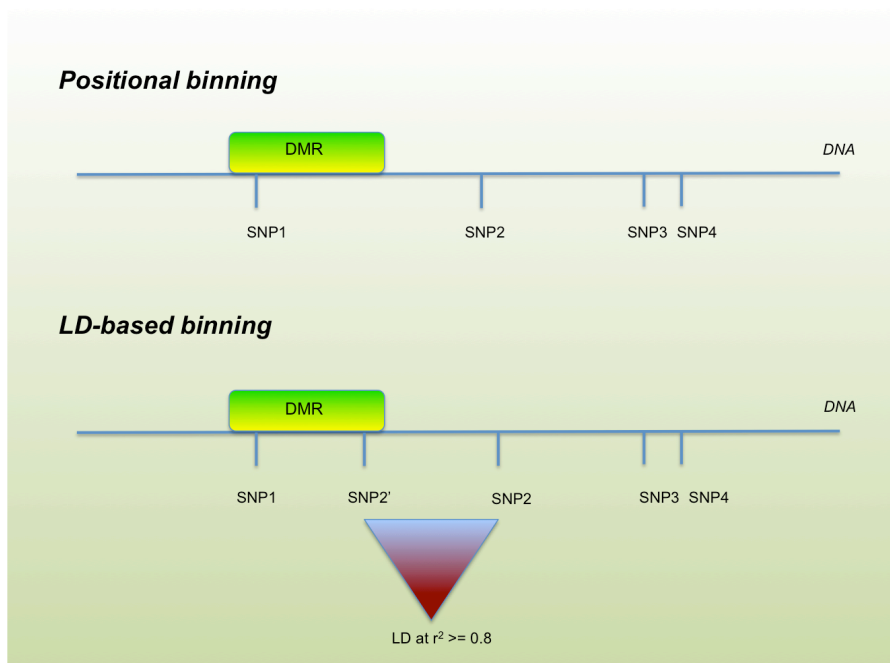
## 5. Discussion & Conclusion

### 5.1 Methodological Considerations

#### 5.1.1 Mapping of DMRs and SNPs with LDsnpR

A fundamental aspect of the present thesis has been the annotation of the evolutionary DMRs from Gokhman et al. [150] with SNPs from schizophrenia GWAS [38]. As presented in **Paper I**, this annotation was performed using the software LDsnpR [243] that was previously developed in the group [243]. Unlike other mapping tools such as Galaxy [264] and Genomic Ranges [263] that can provide information on overlaps, LDsnpR provides LD-based information as well [243]. This is useful because the individual nucleotides in the human genome are not independent of each other. There are places in the genome where groups of nucleotides tend to be inherited together and have low recombination. This biological phenomenon is exploited in designing GWAS chips where only a subset of all nucleotides is incorporated. This allows the chips to probe the entire genome by 'tagging' regions with a few select SNPs.

However, this advantage comes at the cost of resolution, i.e. any SNP associated with a trait actually specifies a particular region, but seldom the causative variant for the trait under investigation. Furthermore, since GWAS chips are designed using LD, utilising only overlapping SNPs to annotate the DMRs could potentially lead to an underestimation of DMRs containing schizophrenia-associated SNPs.



*Figure 4: Depiction of DMR binning strategies.*

Using ‘Positional binning’, only SNPs overlapping a DMR are tagged. In this case, only 1 SNP is tagged to the DMR. This strategy would work if all SNPs were typed in a genotyping chip. However, due to the nature of LD in the human genome, there may be more SNPs within the DMR that are not directly genotyped in the GWAS chip. This is depicted in the case of ‘LD-based binning’ where an additional SNP that is in LD with SNP2 is tagged to the DMR as SNP2’. Thus using ‘LD-based binning’, one can ‘capture’ more SNPs tagging to a particular DMR than using only ‘Positional-binning’ alone.

Therefore, LDsnpR is an indispensable tool to circumvent the shortcoming of using only positional binning. The software was used for annotating all the GWAS SNPs to the DMRs through both LD-based and positional binning. A potential shortcoming of using LD-based information is that human populations have different LD-structures. However, this shortcoming was overcome using specific population background from the 1000 Genomes project [110] for the European population. This was done keeping in view that the markers used in our study are obtained from the schizophrenia GWAS in which 49



---

out of 52 cohorts were of European origin [38]. Similarly, the majority of samples from other GWAS were of European ancestry.

### 5.1.2 Enrichment analyses with conditional Quantile-Quantile (QQ) plots

The method of performing enrichment analysis with conditional quantile-quantile (QQ) plots is relatively new [245]. The regular QQ plots used in GWAS typically show the distribution of the observed  $p$ -values of SNPs from the expected distribution under the null hypothesis. Schork and colleagues [245] were the first to extend this method to depict distributions of SNPs selected (conditioned) on specific criteria and compare it to the overall distribution of SNPs. They found that trait association signals could be significantly enhanced with a concomitant improvement in false discovery rate (FDR) when SNPs were selected based on specific genomic annotations such as 5'UTRs, 3'UTRs, Exons, Introns etc. [245]. We, therefore, used the same method and applied it to SNPs selected for annotating to DMRs.

Traditionally, the deviations of test-statistics at the tail-end of QQ plots were assumed to occur due to population stratification or cryptic relatedness of cases [265]. However, as reported by another paper by Schork and colleagues [266], it is the polygenic architecture of complex traits that is displayed by the extreme deviations, and such deviations are not due to spurious association.

Schork and colleagues [245] also provide a way to quantify the enrichment observed using a two-sample *Kolmogorov-Smirnov* test that analyses the empirical cumulative distribution frequencies (*ecdf*) of specific strata of SNPs with that of intergenic SNPs. They also provide the so-called enrichment score given by normalised mean ( $z\text{-score}^2 - 1$ ) where  $z\text{-score}$  denotes the effect size of the SNP under consideration: the number of standard deviations away from the mean of a normal (0,1) distribution.

However, we could not quantify the enrichment observed in DMRs with these methods. We hypothesize this could be partly due to an inadequate number of SNPs tagging DMRs

---

compared to the overall number of SNPs in the GWAS. The two-sample *Kolmogorov-Smirnov* test is dependent on sample sizes. In our case, we had a disproportionately smaller number of SNPs tagged to DMRs (~27,000) compared to the overall number of SNPs in the GWAS (~9.4 million full schizophrenia GWAS). Furthermore, Schork and colleagues [245], implement the test by comparing the specific category of SNPs against intergenic SNPs. In our opinion, intergenic SNPs would not be an appropriate background since many DMRs are themselves at intergenic regions, while others are in the gene bodies or promoters. This made the use of any other category of SNPs such as those tagging 5'UTR, exons, introns or 3'UTRs extremely challenging.

Similarly, the enrichment score of the mean ( $z\text{-score}^2 - 1$ ) was implemented by normalising the various SNP categories against the category of SNPs possessing the highest  $z$ -scores. In the study by Schork et al. [245], this was the 5'UTR region, and therefore the  $z$ -scores of all other SNP categories (exons, introns etc.) were normalised against it. The method, therefore, could not be implemented for our dataset as our genomic category of DMR could not be normalised against any specific background set of known genomic annotation of exon, intron etc. There are several DMRs that are present in gene bodies and promoters, and since we wanted to test for enrichment of the entire DMRs, we could not directly implement the enrichment score as described by Schork and colleagues.

It is due to these limitations that we chose to implement the INRICH pipeline [246] to quantify the enrichment observed in conditional QQ-plots [267].

### 5.1.3 Enrichment analyses with INRICH

The INterval enRICHment analysis tool [246] implements permutation and bootstrapping procedures, to test LD-implicated genome intervals for the enrichment of specific gene-sets. It is extremely valuable for testing regions implicated by GWAS studies as it corrects for several confounding factors such as variable gene size, overlapping genes, overlapping intervals, SNP density, LD within and between genes. The default

---

parameters involve performing 10,000 rounds of permutation to generate empirical significance values for each gene-set. This is followed by 5000 rounds of bootstrapping to correct for testing several different gene-sets [246]. Xu and colleagues [211] used this method to demonstrate enrichment of genes flanking HARs for schizophrenia. We implemented a similar procedure as them to demonstrate enrichment of genes flanking DMRs for schizophrenia and also independently verified enrichment of genes tagged by NSS for schizophrenia [267].

However, we failed to replicate the original findings of Xu and colleagues [267]. This might have been due to a different gene annotation or version. Additionally, they used specific background gene-sets in their INRICH analysis that we did not implement. During the INRICH analysis, it is possible to compare the enrichment of test gene-sets against a background gene-set. As Xu and colleagues test for enrichment the primate HARs, i.e. regions conserved in non-human primates but evolved in humans, they make the argument of using a background set of primate orthologous gene-set instead of all human genes.

However, Xu and colleagues fail to mention if any enrichment was observed when using all human genes as background. This is important because the schizophrenia regions tested by them are generated on a genome-wide GWAS. If only primate orthologs are used as background, then it becomes crucial also to limit the schizophrenia-associated LD intervals to the orthologous regions. It is unclear whether this correction was performed. Hence, in our study, we utilized all human genes as background because a) the LD-implicated intervals are derived from a genome-wide schizophrenia GWAS, b) NSS was generated on genome-wide scale, and c) we tested a set of genes flanking DMRs, NSS and HARs that together covered ~50% of known human genes. It is for these reasons we chose to argue against using specific background gene-sets in our INRICH analyses.

An important caveat of the INRICH analyses is that the final  $p$ -values generated after permutation and bootstrapping procedures are dependent on the number of gene-sets

---

being tested. Testing only the DMR gene set will give a different  $p$ -value compared to testing a combined gene set flanking DMRs, NSS and all the different HARs. Furthermore, due to a random seed generator in the computational pipeline of INRICH, repeating the analysis will never give the same  $p$ -values as before, although the trend and direction remain same. It is, however, possible to obtain the same  $p$ -values if the same random seed generator is used in the new analyses.

Lastly, as INRICH tests for gene sets, it is not directly comparable to polygenic enrichment methods such as conditional QQ plots that test enrichment at a SNP level. It is therefore important to consider how the gene sets are constructed. We used the identical procedure as described by Xu and colleagues of a 100kb flanking window for HARs [211]. However, this may or may not be an ideal window to tag genes for DMRs simply because we do not yet know the full functionality of the DMRs. An ideal gene set for DMRs would be genes that have been experimentally verified to be affected by the variable methylation brought about when the ancestors of modern humans diverged from those of Neanderthals and Denisovans ~750,000 years ago [268]. This would require future work.

#### *5.1.4 Evolutionary DMR analyses in patient samples*

From the results of **Paper I** that described the enrichment of human-specific DMRs for schizophrenia associated markers, the next task was to determine which of the DMRs should be analysed for functional analysis in patient samples. Since we wanted to focus on the DMRs that contained significantly associated schizophrenia markers, it was important to determine an appropriate threshold for the significance cutoff of the SNPs. Traditionally, for GWAS data containing millions of SNPs, the significance threshold is assigned using the Bonferroni correction for multiple testing [100]. This is an arbitrary figure based on the assumption of a million independent SNPs [269] on the GWAS chip that gives a genome-wide significance threshold of  $5 \times 10^{-8}$  at  $\alpha = 0.05$ .

---

However, unlike in a GWAS, the genome-wide threshold could not be used to determine DMRs containing significantly associated SNPs. In our case, we had ~27,000 SNPs tagging DMRs in LD. These SNPs are thus not independent of one another. It was, therefore, necessary to determine the appropriate number of independent SNPs to correct for in multiple testing.

The method of Moskvina [247] was implemented in R [244]. The effective number of SNPs is estimated based on pairwise correlation, which accounts for underlying LD and provides more realistic estimates of independent SNPs. This was estimated to be ~2048 SNPs that was used to determine the study-wide significance threshold at  $\alpha = 0.05$  as follows:

$$\text{Studywide threshold} = \frac{0.05}{\text{actual number of independent tests (2048)}}$$

This gave a value of  $2.4 \times 10^{-5}$ . Subsequently, 9 DMRs were found to possess SNPs that crossed this threshold.

Thus, in the analysis presented in **Paper III**, we assume that it is the DMRs possessing significantly associated schizophrenia SNPs that are relevant for an investigation. It is possible to speculate that there may be other evolutionary DMRs from Gokhman et al. [150] that do show methylation variation in patients with schizophrenia, despite not possessing significant schizophrenia-associated SNPs. This should be investigated in future studies, ideally with an upgraded Illumina methylation array that has improved coverage for these DMRs. This aspect of coverage of the evolutionary DMRs is discussed in the following section.

---

### *5.1.5 Methylation analyses in brain and blood samples*

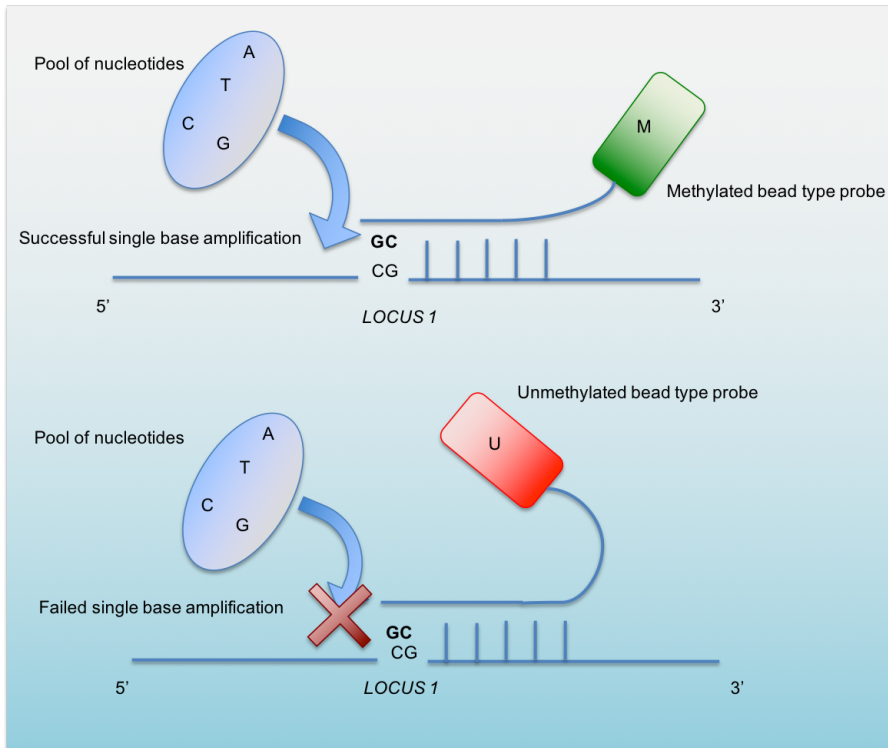
The datasets utilised for carrying out the methylation analyses were performed in Illumina's HumanMethylation450 BeadChips also known as the 450k array [248,249]. Briefly, this is a microarray-based technology that measures methylation variation in the samples with the help of probes. In the 450k array, there is a little more than 450,000 probes (hence the name ) spread across the genome. These probes are used for measuring methylation variation at a specific CpG site in the genome and together cover 99% of all RefSeq genes, more than 19,000 unique CpG islands as well as 3000 probes at non-CpG sites [248].

When using this platform, each plate can analyse 96 samples at a time. These samples are distributed across eight slides, where each slide can handle 12 samples. The 12 samples are arranged in a grid of 2x6 cells, i.e. two columns and six rows. Each cell in this grid contains probes with two different types of chemistries: Infinium Type I probe and Infinium Type II probe.

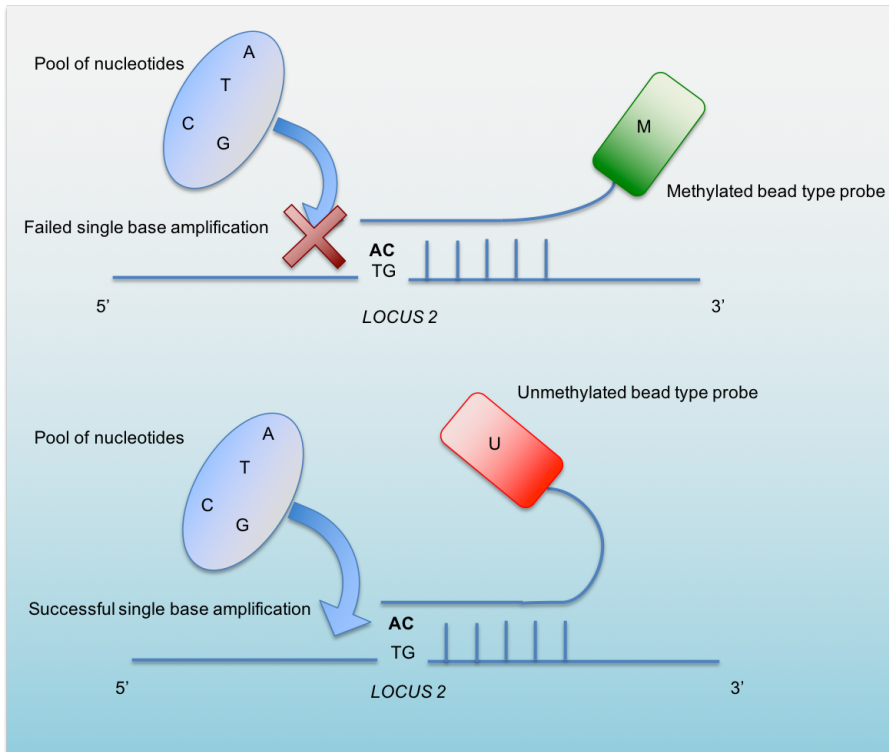
Type I probe was initially designed for the predecessor of the 450k array namely the Infinium 27k array [270]. This probe design utilizes two bead-bound probes per CpG location - one probe measures the methylated signal, the other measures the unmethylated signal. After bisulphite conversion of sample DNA that converts unmethylated cytosines to uracil, for an unmethylated CpG, the probe for detecting unmethylated position hybridises to the DNA, followed by a successful single base pair extension of a fluorescent labelled nucleotide. The bead-bound probe for detecting methylation fails at the single base extension step and therefore does not emit a colour signal.

For the methylated CpG, the other bead-bound probe for detecting methylated position hybridises, followed by a successful allele-specific single base extension. The methylation level at each CpG locus is then determined as the ratio of the signal from the

methyated probe to the total signal intensity ( $\beta$ ). This value has a range from 0 for a completely unmethylated locus to 1 for a completely methylated locus.



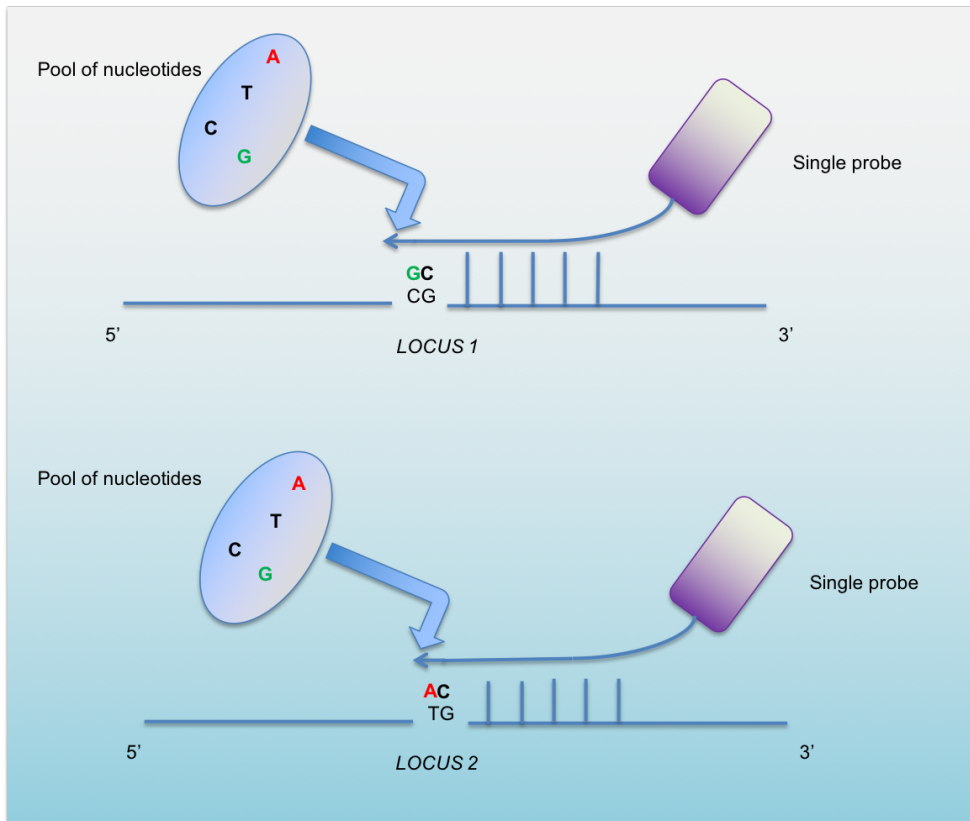
*Figure 5: Depiction of the Type I probe design in the 450k array on methylated locus (Locus 1). Bisulphite treatment preserves the methylated cytosine base. The probe for detecting methylation (green), hybridises and successfully performs a single base amplification while the same results in failure for the probe detecting unmethylated loci (red) (Adapted from Illumina 450k array technical note at [https://www.illumina.com/documents/products/technotes/technote\\_hm450\\_data\\_analysis\\_optimization.pdf](https://www.illumina.com/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf))*



*Figure 6: Depiction of the Type I probe design in the 450k array on an unmethylated locus (Locus 2).*

*Bisulphite treatment converts a non-methylated cytosine to thymine. The probe (red) for detecting an unmethylated locus successfully hybridises and performs single base extension while the same results in failure for the methylation detecting probe (green) (Adapted from Illumina 450k array technical note at [https://www.illumina.com/documents/products/technotes/technote\\_hm450\\_data\\_analysis\\_optimization.pdf](https://www.illumina.com/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf))*





*Figure 7: Depiction of the Type II probe design in the 450k array.*

*Locus 1 is a methylated CpG while Locus 2 is an unmethylated CpG. Bisulphite treatment converts all unmethylated cytosines to thymines (locus 2). A fluorescent-labelled nucleotide gets hybridised indicating presence or absence of methylation at a specific locus (Adapted from Illumina 450k array technical note at [https://www.illumina.com/documents/products/technotes/technote\\_hm450\\_data\\_analysis\\_optimization.pdf](https://www.illumina.com/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf))*

About 70% of the probes in the 450k array are of Type 2 design [248]. This probe design utilises only one probe per CpG locus. This enables more probes to be assimilated in the array design, thereby expanding the coverage of CpG sites across the genome. In the Infinium Type II probe, after the probe hybridises to the bisulphite treated DNA, single base extension with a differentially labelled fluorescent nucleotide indicates the

methylation status. If the CpG site initially was methylated, then the 3' end of the probe incorporates a complementary nucleotide of G that has a green fluorescent label while a red fluorescent labelled A is incorporated if the original CpG site was unmethylated.

Probes in the 450k array are 50bp in length

([https://www.illumina.com/documents/products/technotes/technote\\_hm450\\_data\\_analysis\\_optimization.pdf](https://www.illumina.com/documents/products/technotes/technote_hm450_data_analysis_optimization.pdf)). For the evolutionarily enriched DMRs, as depicted in the table below, the coverage is weak except for *DMR204* where more than half of the DMR can be analysed with the probes. The results in **Paper III**, therefore, need to be viewed with this caveat in mind.

<b>DMR name</b>	<b>Chr</b>	<b>Start</b>	<b>End</b>	<b>Length</b>	<b>No. of probes overlapping</b>	<b>Percentage covered of DMR</b>	<b>Probes needed for 100% coverage</b>
DMR10	1	2378997	2379771	774	1	6.4%	15
DMR127	3	10805466	10809820	4354	11	12.6%	871
DMR203	6	28058845	28060341	1496	6	20.0%	30
DMR204	6	28833654	28834570	916	10	54.6%	18
DMR236	7	1952455	1953065	610	4	32.8%	12
DMR237	7	2048069	2049043	974	3	15.4%	19
DMR291	8	38287269	38289920	2651	1	1.8%	53
DMR526	14	103745950	103746209	259	1	19.3%	51
DMR527	14	104006646	104008448	1802	1	2.7%	16

*Table 1: Depiction of probe coverage of the 450k array in DMRs selected for patient analyses.*

*The table summarises information on the 9 DMRs analysed in Paper III with genomic coordinates provided in hg19 and probe coverage in these DMRs from the 450k array.*

---

Assuming that 100% probe coverage in these DMRs is needed for a more accurate understanding of methylation variation in patients with schizophrenia, the 450k platform is under-equipped to paint a full picture. Since CpG sites are correlated to at least 50bp, one may not need as many probes as depicted in the table above. However, for a more accurate representation of the methylation variation in these evolutionary DMRs, the probe coverage needs to increase substantially.

The poor probe coverage is especially relevant when considering the statistical tests in the *comb-p* algorithm [258]. As described previously under the Methods section, this algorithm is highly suitable for analysing methylation arrays where the probes, and thus their respective  $p$ -values, are unevenly spread across the genome. *Comb-p* determines the correlation between adjacent probes and helps combine adjacent  $p$ -values to determine true significance of probes or regions. In **Paper III**, the *region\_p* programme of *comb-p* was used for determining the statistical significance of the methylation variation in the evolutionarily enriched DMRs. This programme initially computes the auto-correlation of probes up to the maximum length of the regions to be tested. It then performs the Stouffer-Liptak-Kechris correction where the  $p$ -value of each probe within a particular DMR is adjusted according to adjacent  $p$ -values as weighted according to the auto-correlation. When neighbouring  $p$ -values are low, a given  $p$ -value gets pulled lower. Finally, a Sidak one-step correction for multiple testing is performed. For a given DMR, the number of tests to correct for is determined by the total length of bases covered by probes divided by the size of the region. Since this correction takes into account the number of bases covered by the probes in the DMRs, it may vary if a) probe coverage in the region is inconsistent, and, b) probes are dropped during pre-processing steps.

An additional confounder in the analyses presented in **Paper III** relates to the variable cell composition of all the brain datasets analysed [174]. Although all the studies utilised brain samples from the prefrontal cortex, there is considerable variation in the ratio of neuronal to non-neuronal cells between the different datasets. In fact, when comparing the prefrontal cortex dataset from Wockner and colleagues [169] to the two datasets from

---

Pidsley et al. [173], the Pidsley datasets contained twice the proportion of neuronal cells in the schizophrenia samples as the dataset from Wockner et al. [174]. This may partly explain the opposite direction of effect in methylation observed in schizophrenia patients for *DMR127*. Furthermore, the data from Jaffe et al. [178] estimates cell composition using three additional cell populations apart from the neuronal and non-neuronal cell types.

Finally, an interesting point to consider is the fact that the methylation in brain samples are all post-mortem while the methylation in blood samples is antemortem. We, therefore, do not know for sure, if the methylation difference observed in the brain samples from schizophrenia patients has more to do with the fact that the subjects had passed away or the fact that brain and blood do not share the same methylation profile in these evolutionarily enriched DMRs. One way to unravel this conundrum could be to test post-mortem blood samples with post-mortem brain samples. Additionally, it has been previously reported that the vast majority of probes do not show similar methylation profiles between brain and blood [271]. Thus, using blood to infer methylation variation in these evolutionary DMRs for patients with schizophrenia may not be optimal.

Furthermore, antipsychotics are known to influence DNA methylation levels and as such their effect cannot be ruled out [176,272,273]. As reported previously by Mill and colleagues [167], DNA methylation of a CpG island located upstream of *MEK1* is correlated with lifetime antipsychotic usage in the brain. Similarly, *FOSB* is activated by administration of antipsychotic medications and was found to be hypo-methylated in female patients with major psychoses [167].

However, as reported previously [274], DNA methylation differences can be observed in antipsychotic-free patients using peripheral leukocyte samples. But the study may have potential limitations such as the relatively young age of patients ( mean age:  $31.9 \pm 9.7$  years) that suggests they were unlikely to be stable long-term patients with schizophrenia. Furthermore, the results may also be confounded due to the fact there are several

---

different cell types in leukocytes with their own specific DNA methylation signatures [275].

It has been suggested that certain antipsychotics like clozapine are superior to haloperidol due to the ability of clozapine to induce hypo-methylation in GABA and glutamatergic promoters [273]. A recent study [276] used peripheral blood samples and 450k array to look into the effects of clozapine on treatment-resistant patients with schizophrenia. They found significant methylation differences at more than 29,000 positions after one year of clozapine treatment. However, the study did not take into account the peripheral blood cell composition.

#### *5.1.6 Limitations of the study*

The epigenetic change of DNA methylation of CpG dinucleotides has been of prime focus in **Papers I-III**. This is, however, not the only epigenetic modification in the human genome. There are reports of variation in DNA methylation termed 5-hydroxymethylation that we did not consider for investigation. Furthermore, research has revealed that 5-hydroxymethylation is a more common methylation modification in the brain [277,278] than in other organs. This makes the investigation of this epigenetic mark very relevant for brain disorders such as schizophrenia. We were however limited by the fact that the evolutionary epigenetic changes were investigated only in CpG (5-methylcytosine) methylation and not 5-hydroxymethylation. We, therefore, limited ourselves to investigating only 5-methylcytosine and not 5-hydroxymethylation.

Additionally, we did not investigate histone marks for similar reasons. To the best of our knowledge, they have not been investigated from an evolutionary point of view. RNA is also a vital component of epigenetic machinery [123,279], not considered in the present study. Although some forms of RNA molecules have been shown to be responsible for intergenerational inheritance [280], they have yet to be investigated from an evolutionary

standpoint. So far, it is only CpG methylation in the form of 5-methylcytosine that has been investigated from an evolutionary perspective [181–183].

In the context of psychiatric disorders, bipolar disorder is known to share genetic similarity with schizophrenia [281]. We failed to find evidence of enrichment of evolutionary human-specific DMRs for bipolar disorder associated SNPs. This could partly be due to the fact that bipolar disorder GWAS was underpowered [235] because of a moderate sample size ( $n = 7,481$  cases and 9,250 controls). There is a new bipolar disorder study currently available as a pre-print on bioRxiv that we did not use [282]. Furthermore, as a limitation for the polygenic enrichment method employed, a future GWAS study on schizophrenia with more power could shift the  $p$ -value distribution entirely. This could cause more SNPs to cross the significance threshold and as such more evolutionary DMRs may be relevant for an investigation.

## 5.2 Biological significance of our results

### 5.2.1 Discussion of results in Papers I-III

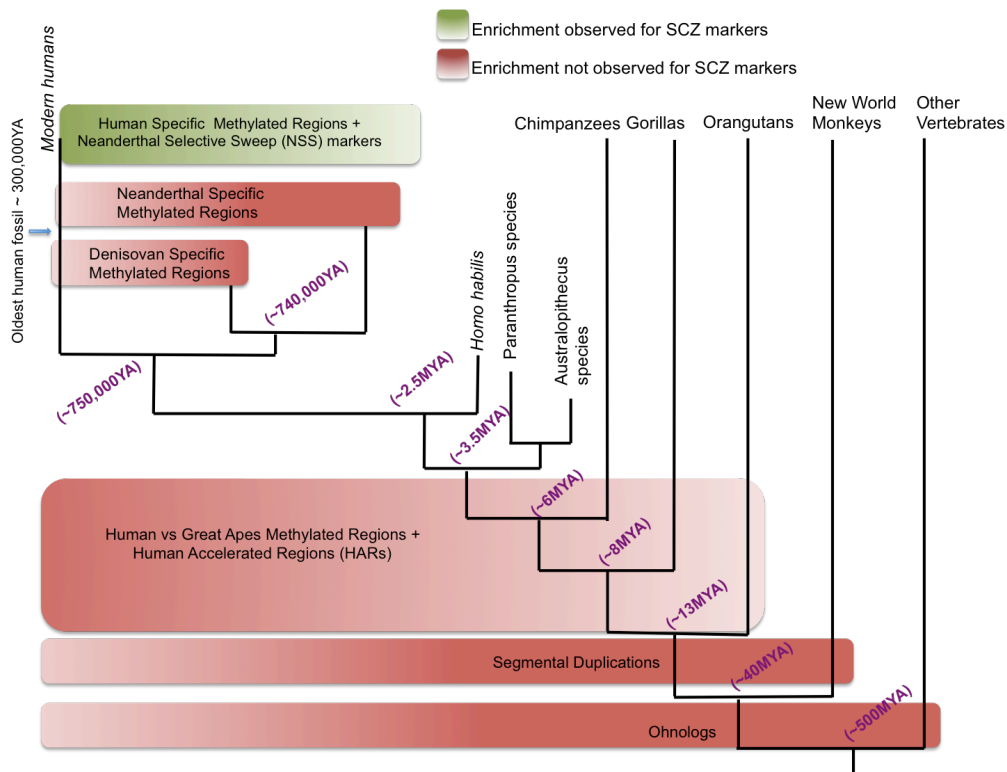


Figure 8: Depiction of all evolutionary datasets tested in literature for the enrichment of schizophrenia markers

The diagram depicts various evolutionary annotations, superimposed on a phylogenetic tree, which have been tested in literature for the enrichment of schizophrenia markers. Red denotes lack of enrichment while green depicts positive enrichment for markers associated with schizophrenia. (Timeline estimates from the Smithsonian (<http://humanorigins.si.edu/>), Hasegawa et al 1985 [283], Srinivasan et al 2017 [210], Rogers et al 2017 [268])

Taken altogether, our results hint at epigenomic evolution through methylation, having taken place in genomic regions implicated in the aetiology of schizophrenia (**Paper I**).

---

These regions harbour markers that are involved in the regulation of various neuro-developmental pathways. The fact that methylation changes also took place in the very same regions suggest a complex gene by environment interaction in the evolution of humans, especially pathways that led to the development of our brain. It is known that various factors from the environment can make long-lasting changes in the DNA methylation patterns that can be subsequently inherited at a population level [284–286]. Thus, the true significance of our findings from an evolutionary standpoint suggests, the interaction of the epigenetic machinery through DNA methylation with environmental factors over the past 300,000 years when the first anatomically modern humans appeared on Earth [218,219]. These gene by environment interactions may in part have been responsible for changes in neuro-developmental pathways, some of which also affected neuro-developmental pathways implicated in the aetiology of schizophrenia.

We also show that DNA methylation variation that occurred in *Homo sapiens* since the divergence of the common ancestor from the great apes are not enriched for markers of schizophrenia, with the exception of the extended MHC region (**Paper II**). This suggests that the evolutionary events of the past 300,000 years but not ca 6-7 MYA, may have influenced the DNA methylation of genomic regions associated in the aetiology of schizophrenia at a genome-wide level. In Paper II, we call the DMRs observed in humans by comparison with chimpanzees, gorillas and orangutans as primate DMRs. The observation of enrichment of primate DMRs for schizophrenia risk markers only in the extended MHC region might suggest that the immune system markers associated in schizophrenia have been undergoing evolution from a very long time. Indeed it has been shown that the MHC region bears signatures of both recent and ancient natural selection [287].

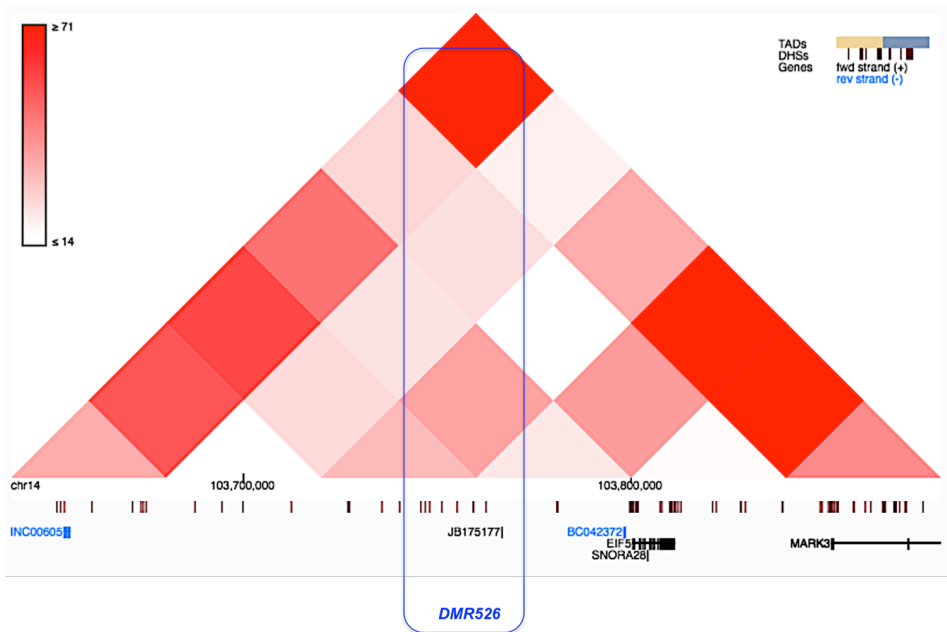
The exact biological consequences of this enrichment of primate DMRs in the extended MHC region are however not easy to infer. This is because the extended MHC region is the most gene-dense region of the human genome containing over 250 genes at an average gene density of one gene every 16 kb along with extended linkage-



---

disequilibrium [287]. Thus it is challenging to infer which genes might have had their regulation affected by variation in DNA methylation. When we remove the extended MHC region from the analysis, any evidence of enrichment of primate DMRs disappears suggesting a highly localised evolution of the immune system markers in schizophrenia since the divergence from great apes.

Finally, in **Paper III**, we investigate whether the DMRs from Paper I that contain schizophrenia markers above a study-wide threshold show disrupted methylation in patients with schizophrenia. We observe that some of these DMRs do show variation in patients samples. The variation is observed however only in brain tissue and not blood samples. This is the first attempt at investigating if regions whose methylation has evolved in *Homo sapiens* shows disrupted patterns of methylation in patients with schizophrenia. The results suggest *LINC00606*, *ZSCAN12P1* and *FGFR1* to be relevant from an evolutionary perspective and also important in the aetiology of schizophrenia as they contain SNPs associated with schizophrenia at or above the study-wide threshold of  $p = 2.4 \times 10^{-5}$ . Additionally, we find statistically significant variation in a DMR that does not overlap any particular gene. Data from the Roadmap Epigenomics Project [262] in the prefrontal cortex tissue suggests it to be present in a region demarcated as a 'bivalent enhancer'. Using Hi-C data in the hippocampus [288], we may speculate that this particular DMR could be important in the regulation of *MARK3*, a gene that has been implicated in disorders of astrocytes.



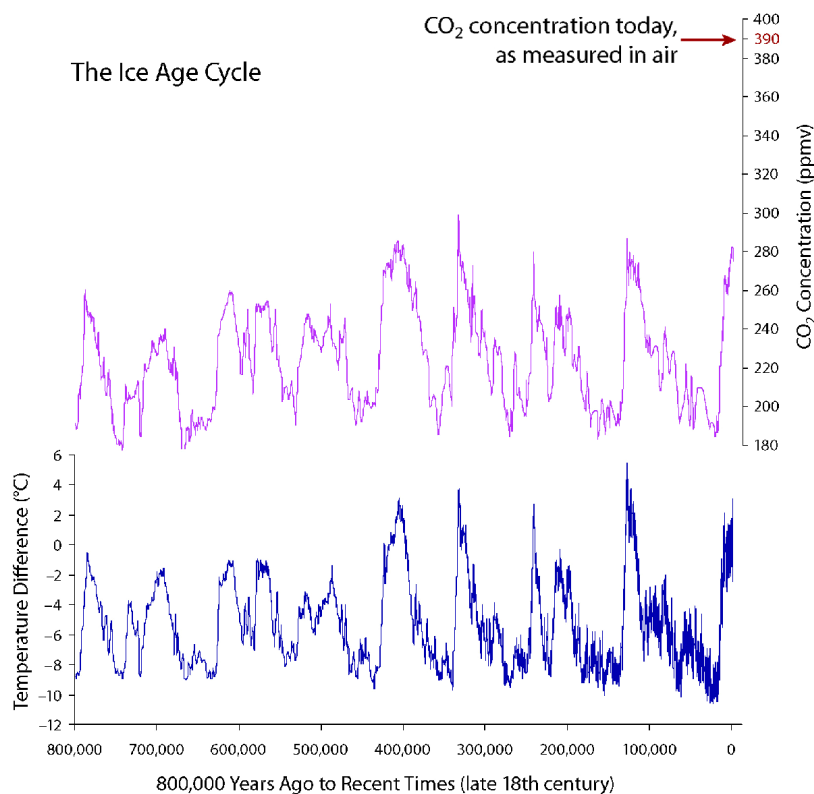
*Figure 9: The potential regulatory role of DMR526 on MARK3*

*The depiction of a Hi-C contact matrix with the approximate position of DMR 526 at chr14: 103,745,950-103,746,209. Figure derived from Hi-C results on Hippocampus published by Schmitt et al., 2016 [288] and visualised in 3D Genome Browser at <http://promoter.bx.psu.edu/hi-c/view.php>*

### *5.2.2 An epigenetic perspective of the results in the broader context of recent human evolution*

It is important to place the results of the present body of work in the broader understanding of human evolution. A variety of factors could have driven the human-specific methylation changes. One could speculate about the potential environmental agents that might have influenced the methylation patterns over the past 300,000 years in *Homo sapiens*.

Climate fluctuations have often occurred over the past 800,000 years [289]. The period of rapid expansion of the human intellectual capacity and colonisation of the world co-occurs with the last ice-age in which average temperatures in the South Pole reduced as much by 10 °C along with a concomitant reduction in global CO<sub>2</sub> levels [290]. These climatic changes could potentially have affected the food sources available to our ancestors. Food and dietary patterns are well known to affect DNA methylation patterns [291–293].



*Figure 10: Depiction of the global climatic shifts over the past 800,000 years.*

*The figure depicts variations in temperature and carbon dioxide (CO<sub>2</sub>) levels as observed at the South Pole. The emergence of modern Homo sapiens ca 70,000-40,000 YA occurs during the last ice age, while the oldest fossil of an anatomically modern H. sapiens is dated to ca 300,000 YA. (Source: National Research Council, 2010a, Lüthi, D et al., 2008 [290], Jouzel J et al. 2007 [289], Reproduced with permission from National Academies Press).*

---

Long-chain polyunsaturated fatty acids (PUFA) are important components of brain phospholipids [294]. These include arachidonic acid (AA) from meat and docosahexaenoic acid (DHA) from fish. Abnormal brain phospholipid metabolism has been observed in patients with schizophrenia [294]. It has been suggested that the diet of our ancestors, especially of those living along the sea-coast and inland lakes, was rich in consumption of fish [295]. A diet rich in fish would provide DHA to babies via their mother's milk [295,296]. Long-term availability of PUFA from such food sources may have influenced the development of the brain [294] through epigenetic mechanisms [297]. This may represent yet another environmental means of altering DNA methylation in our ancestors. It remains unknown, however, the extent to which the consumption of fish and meat can influence methylation in the human-specific DMRs analysed in the present body of work.

Some researchers have proposed that methylation changes can be driven by sequence changes [298] and while that may be true, there also exists the phenomenon of differential methylation between identical twins called 'metastable epialleles' [182]. Thus, not all methylation changes may be driven by the underlying genomic sequence. Gokhman and colleagues observed that only some but not all the human-specific methylation changes were influenced by the underlying genomic sequence [150].

Anatomically modern humans appeared on earth at least 300,000 years ago [218,219]. However, it is challenging to put an exact date on the emergence of superior mental abilities that define the modern *Homo sapiens*. In particular, anthropologists and archaeologists studying human evolution often cite the emergence of the superior intellectual abilities of *Homo sapiens* from about 70,000 years onwards [221]. This is the period from which there is irrefutable evidence of the emergence of art, religion [222,299] and possibly spoken language [227]. From an evolutionary perspective, it suggests a massive leap in the animal kingdom because *Homo sapiens* became the first to not only develop the capacity to think and imagine things that do not exist [222,299] but also to communicate the same ideas to other members of the species [227]. This aspect

would have been critical in the ability to coordinate and cooperate effectively with large groups numbering more than 150 individuals and even needed in keeping a group together [220,300].

This ability to cooperate and form bonds with other members of their species, united by a shared belief system may have allowed larger groups of humans to co-exist. This aspect gains significance from the recent work by Rogers et al. [268] that suggests Neanderthals, although spread far and wide as evidenced by the rich collection of their fossils, existed in small bands of individuals. One could thus speculate that although for most of the time in existence, humans were not able to displace the Neanderthal from their natural habitats, the ‘intellectually-superior’ humans, characterised by larger, well-coordinated groups may have driven the smaller bands of Neanderthals to extinction ca 40,000 years ago [301].

Using the data from Gokhman et al. [150], we gain an insight into the methylation variation that has potentially occurred since the last common ancestor of humans, Neanderthals and Denisovans diverged ca 750,000 years ago [268]. There thus remains a gap in our knowledge of the evolutionary events occurring in the human lineage from the period of ca 750,000 to 300,000 years when anatomically modern humans appeared. Assuming the superior intellectual abilities appeared ca 70,000 years, then there remains another gap in our knowledge of evolutionary events that allowed the massive leap in intellectual capacity.

Using the present study to determine when exactly in the period of the last 300,000 years [218,219] *Homo sapiens* developed their superior intellectual abilities or psychosis is therefore challenging. If the method used in the reconstruction of genomes of Neanderthals and Denisovans [150,302] could be implemented on samples of ancient *Homo sapiens* from different time periods [303–310] then theoretically it should be possible to reconstruct the methylomes of ancient humans [311]. Subsequently, a more ‘time-series’ analyses of change in methylation patterns of regions controlling neuro-

---

developmental processes implicated in schizophrenia and psychosis should be feasible. This idea has been further expanded upon by recent work from Gokhman et al. [182] and Hanghøj et al. [312].

Amongst the defining hallmarks of schizophrenia are hallucinations [313,314]- a phenomenon of imagining and experiencing things that do not exist. Also amongst the defining features of our species is the ability to think and imagine things that do not necessarily have a physical form of existence such as belief systems [222,223,299]. Given that abstract thinking including art and religion appeared at least 40,000 years ago [222,299], it may very well be the case then that the earliest psychotic patients may have belonged in this period and not in the *Homo sapiens* of ca 300,000 year period [218,219]. The *Homo sapiens* from ca 300,000 years, though physically similar to modern humans, may, in fact, have had sub-developed intellectual skills in working memory [225] that in turn is critical in creativity [315–317]. Indeed, the rapid expansion of *Homo sapiens* from ca 70,000 years onwards might have been spurred by a distinct advantage over earlier *Sapiens* through the origin of language or belief systems [220,224,227] that may have allowed larger bands of individuals to cooperate, migrate and colonise the world [220,225,300]. One could speculate that within this ability to bring together massive numbers of individuals through religion [220] and language [227] may be where psychosis originated. The answer therefore to the origins of psychosis may thus ultimately lie in the origins of creativity, language and religion that may be answered through evolutionary epigenomics of ancient *Homo sapiens* [182,312].



---

SNPs. Our results in principle support the hypothesis that evolution may in part have contributed to schizophrenia risk. However, at present, we cannot ascertain whether the recent methylation changes harbouring schizophrenia markers are causative of the disorder or piggyback due to some hitherto unknown stochastic phenomenon. Furthermore, schizophrenia is a highly polygenic trait [266] that has shared pleiotropy with several traits [318–320]. Although we do not find enrichment of other trait-associated markers with evolutionary DMRs (with the possible exception of height), we cannot rule out that the enriched DMRs affect the pathophysiology of other traits.

Additionally, the hypothesis put forward by T.J. Crow stipulates a dysregulation of the language process being central in the aetiology of schizophrenia. Our current results neither confirm nor contradict this aspect of the hypothesis. This is partly due to the fact that the methods employed cannot test for language origin, proficiency or dysregulation. Furthermore, to utilise evolutionary methylation changes for language analysis, one would need much recent methylation data from a time where there is incontrovertible evidence of human civilisation and language flourishing [13].

In conclusion, while we find evidence to support the hypothesis that recent evolution may have played a role in modulating methylation in genomic regions associated in schizophrenia, the present methods lack the resolution to determine language dysregulation as a key component of human evolution and psychosis. As a consequence, there is much work that remains to be done to either prove or disprove T.J. Crow's evolutionary hypothesis that schizophrenia is the price humans pay for language. Future work should also investigate the human-specific DMRs using either an array technology with better probe resolution or whole-genome sequencing methods for measuring methylation variation in patients with schizophrenia.



---

## 6. References

- [1] Huppert JD, Weiss KA, Lim R, Pratt S, Smith TE. Quality of life in schizophrenia: contributions of anxiety and depression. *Schizophrenia Research* 2001;51:171–80.
- [2] Saarni SI, Viertiö S, Perälä J, Koskinen S, Lönnqvist J, Suvisaari J. Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. *The British Journal of Psychiatry* 2010;197:386–94.
- [3] Kolotkin RL, Corey-Lisle PK, Crosby RD, Swanson JM, Tuomari AV, L'italien GJ, et al. Impact of obesity on health-related quality of life in schizophrenia and bipolar disorder. *Obesity (Silver Spring)* 2008;16:749–54.
- [4] Chong HY, Teoh SL, Wu DB-C, Kotirum S, Chiou C-F, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatric Disease and Treatment* 2016;12:357.
- [5] Weiden PJ, Olfson M. Cost of relapse in schizophrenia. *Schizophrenia Bulletin* 1995;21:419.
- [6] Wu EQ, Birnbaum HG, Shi L, Ball DE, Kessler RC, Moulis M, et al. The economic burden of schizophrenia in the United States in 2002. *J Clin Psychiatry* 2005;66:1122–9.
- [7] Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull* 2004;30:279–93.
- [8] Knapp M. Costs of schizophrenia. *The British Journal of Psychiatry* 1997;171:509–18.
- [9] Fusar-Poli P, Politi P. Paul Eugen Bleuler and the birth of schizophrenia (1908). *American Journal of Psychiatry* 2008;165:1407–1407.
- [10] Ashok AH, Baugh J, Yeragani VK. Paul Eugen Bleuler and the origin of the term schizophrenia (SCHIZOPRENIEGRUPPE). *Indian Journal of Psychiatry* 2012;54:95.
- [11] Bleuler E. *Dementia praecox oder Gruppe der Schizophrenien*. Handbuch Der Psychiatrie 1911.
- [12] Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, et al. Definition and description of schizophrenia in the DSM-5. *Schizophr Res* 2013;150:3–10.

- 
- [13] Jeste DV, Del Carmen R, Lohr JB, Wyatt RJ. Did schizophrenia exist before the eighteenth century? *Compr Psychiatry* 1985;26:493–503.
- [14] Aderibigbe YA, Theodoridis D, Vieweg WVR. Dementia praecox to schizophrenia: the first 100 years. *Psychiatry and Clinical Neurosciences* 1999;53:437–48.
- [15] Gottesman II. *Schizophrenia genesis: The origins of madness*. WH Freeman, Times Books Henry Holt & Co; 1991.
- [16] Haldipur C. Madness in ancient India: Concept of insanity in Charaka Samhita (1st century AD). *Comprehensive Psychiatry* 1984;25:335–44.
- [17] Bark N, Varadi R, Sorrentino A. Schizophrenia in history. *Schizophrenia Research* 2008;98:25.
- [18] Bark NM. On the history of schizophrenia. *New York State Journal of Medicine* 1988;88:374–83.
- [19] Bark NM. Did Shakespeare know schizophrenia? The case of Poor Mad Tom in King Lear. *Br J Psychiatry* 1985;146:436–8.
- [20] Millingen JG. *Mind and Matter: Illustrated by Considerations on Hereditary Insanity, and the Influence of Temperament in the Development of the Passions*. Hurst; 1847.
- [21] Kraepelin E. *Psychiatrie. Ein lehrbuch für studirende und ärzte [Psychiatry 4th Ed: A Textbook for Students and Physicians]* 1893.
- [22] Kraepelin E. *Dementia Praecox and Paraphrenia* (translated by RM Barclay, facsimile edition published in 1971) 1919.
- [23] Jung CG. *The Psychology of Dementia Praecox*, 1906. Translated by AA Brill Nervous & Mental Disease Publishing Co, New York 1936.
- [24] Bleuler E, Jung CG. Komplexe und Krankheitsursachen bei Dementia praecox. *Zentralblatt Fur Nervenheilkunde Und Psychiatrie* 1908:220–7.
- [25] Odegård O. Changes in the prognosis of functional psychoses since the days of Kraepelin. The forty-first Maudsley Lecture, delivered before the Royal Medico-Psychological Association, 18th November, 1966. *The British Journal of Psychiatry: The Journal of Mental Science* 1967;113:813.

- 
- [26] Schneider K. Psychischer befund und psychiatrische diagnose. G. Thieme verlag; 1939.
- [27] Hoenig J. The concept of Schizophrenia. Kraepelin-Bleuler-Schneider. The British Journal of Psychiatry 1983;142:547–56.
- [28] Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. Archives of General Psychiatry 1978;35:773–82.
- [29] Spitzer R, Endicott J, Robins E. Research diagnostic criteria. Psychopharmacology Bulletin 1975;11:22–5.
- [30] Spitzer RL. Research diagnostic criteria (RDC) for a selected group of functional disorders. Biometric Research 1978.
- [31] Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R. Diagnostic criteria for use in psychiatric research. Archives of General Psychiatry 1972;26:57–63.
- [32] Spitzer RL. The development of diagnostic criteria in psychiatry. Current Contents Social & Behavioural Science 1989:20–20.
- [33] Spitzer RL, Endicott J, Robins E, Kurlansky J, Gurland B. Preliminary report of the reliability of research diagnostic criteria (RDC) applied to psychiatric case records. Prediction in Psychopharmacology: Preclinical and Clinical Correlates, A Sudilovsky, S Gershon, R Deer, Eds 1975:231–45.
- [34] Spitzer RL, Williams JB. Diagnostic and statistical manual of mental disorders. American Psychiatric Association, 1980.
- [35] Frances A, others. Diagnostic and statistical manual of mental disorders: DSM-IV. American Psychiatric Association; 1994.
- [36] Tandon R, Nasrallah HA, Keshavan MS. Schizophrenia, “just the facts” 4. Clinical features and conceptualization. Schizophr Res 2009;110:1–23.
- [37] Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. Am J Psychiatry 2003;160:4–12.
- [38] Ripke S, Neale BM, Corvin A, Walters JT, Farh K-H, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511:421.
- [39] Andreasen NC, Olsen SA, Dennert JW, Smith MR. Ventricular enlargement in schizophrenia: Relationship to positive and negative symptoms. Am J Psychiatry

---

1982;139:297–302.

- [40] Andreasen NC, Olsen S. Negative v positive schizophrenia: Definition and validation. *Archives of General Psychiatry* 1982;39:789–94.
- [41] American Psychiatric Association and others. *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub; 2013.
- [42] WHO. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. vol. 1. World Health Organization; 1992.
- [43] WHO. *International statistical classification of diseases and related health problems*. vol. 1. World Health Organization; 2004.
- [44] WHO. *International Classification of Diseases (ICD) information sheet*. Accessed April 2013;15.
- [45] Jäger M, Bottlender R, Strauss A, Möller H-J. On the descriptive validity of ICD-10 schizophrenia: empirical analyses in the spectrum of non-affective functional psychoses. *Psychopathology* 2003;36:152–9.
- [46] Tyrer P. A comparison of DSM and ICD classifications of mental disorder. *Advances in Psychiatric Treatment* 2014;20:280–5.
- [47] Sørensen MJ, Mors O, Thomsen PH. DSM-IV or ICD-10-DCR diagnoses in child and adolescent psychiatry: does it matter? *European Child & Adolescent Psychiatry* 2005;14:335–40.
- [48] Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry* 2014;171:395–7.
- [49] Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013;11:126.
- [50] Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders 2010.
- [51] Freedman R, Lewis DA, Michels R, Pine DS, Schultz SK, Tamminga CA, et al. The initial field trials of DSM-5: new blooms and old thorns. *American Journal of Psychiatry* 2013;170(1):1-5
- [52] Cheniaux E, Landeira-Fernandez J, Versiani M. The diagnoses of schizophrenia, schizoaffective disorder, bipolar disorder and unipolar depression: interrater reliability and congruence between DSM-IV and ICD-10. *Psychopathology*

---

2009;42:293–8.

- [53] Gottesman II. Severity/concordance and diagnostic refinement in the Maudsley-Bethlem schizophrenic twin study. *Journal of Psychiatric Research* 1968;6:37–48.
- [54] Skre I, Onstad S, Torgersen S, Lygren S, Kringlen E. A twin study of DSM-III-R anxiety disorders. *Acta Psychiatr Scand* 1993;88:85–92.
- [55] Shih RA, Belmonte PL, Zandi PP. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* 2004;16:260–83.
- [56] Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* 1999;56:162–8.
- [57] Mortensen PB, Pedersen CB, Westergaard T, Wohlfahrt J, Ewald H, Mors O, et al. Effects of family history and place and season of birth on the risk of schizophrenia. *N Engl J Med* 1999;340:603–8.
- [58] Kendler KS, McGuire M, Gruenberg AM, O'hare A, Spellman M, Walsh D. The Roscommon family study: I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. *Archives of General Psychiatry* 1993;50:527–40.
- [59] Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon Family Study: III. Schizophrenia-related personality disorders in relatives. *Archives of General Psychiatry* 1993;50:781–8.
- [60] Tienari P, Wynne LC, Moring J, Lahti I, others. The Finnish adoptive family study of schizophrenia: Implications for family research. *The British Journal of Psychiatry* 1994.
- [61] Tienari P, Sorri A, Lahti I, Naarala M, Wahlberg K, Rönkkö T, et al. The Finnish adoptive family study of schizophrenia. *The Yale Journal of Biology and Medicine* 1985;58:227.
- [62] Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003;60:1187–92.
- [63] Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 2009;373:234–9.

- 
- [64] Onstad S, Skre I, Torgersen S, Kringlen E. Twin concordance for DSM-III-R schizophrenia. *Acta Psychiatrica Scandinavica* 1991;83:395–401.
- [65] McGuffin P, Owen MJ, Farmer AE. Genetic basis of schizophrenia. *Lancet* 1995;346:678–82.
- [66] Bassett AS, Chow EW, Weksberg R. Chromosomal abnormalities and schizophrenia. *American Journal of Medical Genetics Part A* 2000;97:45–51.
- [67] International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008;455:237.
- [68] Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–43.
- [69] Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014;506:185.
- [70] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304–51.
- [71] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921
- [72] Crow TJ. Molecular pathology of schizophrenia: more than one disease process? *British Medical Journal* 1980;280:66.
- [73] Bowers MB. Central dopamine turnover in schizophrenic syndromes. *Archives of General Psychiatry* 1974;31:50–4.
- [74] Kwon JM, Goate AM. The candidate gene approach. *Alcohol Res Health* 2000;24:164–8.
- [75] Arinami T, Gao M, Hamaguchi H, Toru M. A functional polymorphism in the promoter region of the dopamine D2 receptor gene is associated with schizophrenia. *Hum Mol Genet* 1997;6:577–82.
- [76] Crocq MA, Mant R, Asherson P, Williams J, Hode Y, Mayerova A, et al. Association between schizophrenia and homozygosity at the dopamine D3 receptor gene. *J Med Genet* 1992;29:858–60.
- [77] Inayama Y, Yoneda H, Sakai T, Ishida T, Nonomura Y, Kono Y, et al. Positive association between a DNA sequence variant in the serotonin 2A receptor gene and

---

schizophrenia. *Am J Med Genet* 1996;67:103–5.

- [78] Williams J, Spurlock G, McGuffin P, Owen M, Mallet J, Nöthen M, et al. Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. *The Lancet* 1996;347:1294–6.
- [79] Harrison P, McLaughlin D, Kerwin R. Decreased hippocampal expression of a glutamate receptor gene in schizophrenia. *The Lancet* 1991;337:450–2.
- [80] Olney JW, Farber NB. Glutamate receptor dysfunction and schizophrenia. *Archives of General Psychiatry* 1995;52:998–1007.
- [81] Akbarian S, Kim JJ, Potkin SG, Hagman JO, Tafazzoli A, Bunney WE, et al. Gene expression for glutamic acid decarboxylase is reduced without loss of neurons in prefrontal cortex of schizophrenics. *Archives of General Psychiatry* 1995;52:258–66.
- [82] Riley B, Kendler KS. Molecular genetic studies of schizophrenia. *Eur J Hum Genet* 2006;14:669–80.
- [83] Hodgkinson CA, Goldman D, Jaeger J, Persaud S, Kane JM, Lipsky RH, et al. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *The American Journal of Human Genetics* 2004;75:862–72.
- [84] Callicott JH, Straub RE, Pezawas L, Egan MF, Mattay VS, Hariri AR, et al. Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:8627–32.
- [85] Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR, et al. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* 2005;310:1187–91.
- [86] Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, et al. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *American Journal of Psychiatry* 2008;165:497–506.
- [87] Shi J, Gershon ES, Liu C. Genetic associations with schizophrenia: meta-analyses of 12 candidate genes. *Schizophr Res* 2008;104:96–107.
- [88] Johnson EC, Border R, Melroy-Greif WE, de Leeuw CA, Ehringer MA, Keller MC. No evidence that schizophrenia candidate genes are more associated with

---

schizophrenia than noncandidate genes. *Biological Psychiatry* 2017;82:702–8.

- [89] Sullivan PF. How Good Were Candidate Gene Guesses in Schizophrenia Genetics? *Biological Psychiatry* 2017;82:696–7.
- [90] Farrell M, Werge T, Sklar P, Owen MJ, Ophoff R, O'Donovan MC, et al. Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry* 2015;20:555.
- [91] Collins AL, Kim Y, Sklar P, O'Donovan MC, Sullivan PF. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med* 2012;42:607–16.
- [92] O'donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics* 2008;40:1053.
- [93] Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. *Nature* 2009;460:744.
- [94] Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* 2011;43:969–76.
- [95] Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet* 2013;45:1150–9.
- [96] Collins FS, Lander ES, Rogers J, Waterston RH, Conso I. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–45.
- [97] Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. *Nature* 2003;426:789–96.
- [98] Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–76.
- [99] Pearson TA, Manolio TA. How to interpret a genome-wide association study. *Jama* 2008;299:1335–44.
- [100] Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011;6:121–33.



- 
- [101] Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012;8:e1002822.
- [102] Koenen KC, Duncan LE, Liberzon I, Ressler KJ. From candidate genes to genome-wide association: the challenges and promise of posttraumatic stress disorder genetic studies. *Biol Psychiatry* 2013;74:634–6.
- [103] Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012;30:1095–106.
- [104] Lowe WL, Reddy TE. Genomic approaches for understanding the genetics of complex disease. *Genome Research* 2015;25:1432–41.
- [105] Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 2014;1842:1910–22.
- [106] Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;106:9362–7.
- [107] Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Human Molecular Genetics* 2015;24:R102–R110.
- [108] Tak YG, Farnham PJ. 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* 2015;8:57.
- [109] Dunham I, Birney E, Lajoie BR, Sanyal A, Dong X, Greven M, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012.
- [110] Durbin R, Abecasis G, Bentley D, Chakravarti A, Clark A, Donnelly P, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491.
- [111] Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research* 2007;35:406–13.
- [112] Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, et al. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Research* 2003;13:1998–2004.
- [113] Akan P, Alexeyenko A, Costea PI, Hedberg L, Solnestam BW, Lundin S, et al. Comprehensive analysis of the genome transcriptome and proteome landscapes of

---

three tumor cell lines. *Genome Biology* 2012;13:R106.

- [114] Caron H, Schaik B van, Mee M van der, Baas F, Riggins G, Sluis P van, et al. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science Signaling* 2001;291:1289.
- [115] Cottingham K. First draft of the human ductal salivary proteome. *Journal of Proteome Research* 2008;7:2592–2592.
- [116] Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. *Nature Biotechnology* 2010;28:1248–50.
- [117] Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA-K, et al. A genecentric Human Protein Atlas for expression profiles based on antibodies. *Molecular & Cellular Proteomics* 2008;7:2019–27.
- [118] Uhlén M, Björling E, Agaton C, Szigyarto CA-K, Amini B, Andersen E, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & Cellular Proteomics* 2005;4:1920–32.
- [119] Collins F, Lander E, Rogers J, Waterston R, Conso I. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–45.
- [120] Ji H. Computational Analysis of ChIP-seq Data. *Methods in Molecular Biology* (Clifton, NJ) 2010;674:143.
- [121] Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology* 2008;26:1351–9.
- [122] Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* 2012;22:1813–31.
- [123] Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007;128:635–8.
- [124] Bird A. Perceptions of epigenetics. *Nature* 2007;447:396–8.
- [125] Jaenisch R, Young R, others. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 2008;132:567.
- [126] Cavalli G, Misteli T. Functional implications of genome topology. *Nature Structural & Molecular Biology* 2013;20:290–9.

- 
- [127] Brown WM. Exercise-associated DNA methylation change in skeletal muscle and the importance of imprinted genes: a bioinformatics meta-analysis. *Br J Sports Med* 2015;bjsports-2014.
- [128] Kouzarides T, others. Chromatin modifications and their function. *Cell* 2007;128:693.
- [129] Keshet I, Lieman-Hurwitz J, Cedar H. DNA methylation affects the formation of active chromatin. *Cell* 1986;44:535–43.
- [130] Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. *Science* 1975;187:226–32.
- [131] Bird A. DNA methylation patterns and epigenetic memory. *Genes & Development* 2002;16:6–21.
- [132] Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. *Science* 2001;293:1068–70.
- [133] Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998;393:386.
- [134] Jones PL, Veenstra GCJ, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics* 1998;19:187.
- [135] Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007;447:425–32.
- [136] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247–57.
- [137] Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* 2008;9:465.
- [138] Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature* 1993;366:362.
- [139] Paulsen M, Ferguson-Smith AC. DNA methylation in genomic imprinting, development, and disease. *J Pathol* 2001;195:97–110.

- 
- [140] Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nature Reviews Genetics* 2006;7:21.
- [141] Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics* 2002;3:415.
- [142] Jones PA, Laird PW. Cancer epigenetics comes of age. *Nat Genet* 1999;21:163–7.
- [143] Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics* 1997;13:335–40.
- [144] Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* 2005;33:6823–36.
- [145] Ehrlich M. DNA methylation in cancer: too much, but also too little. *Oncogene* 2002;21:5400–13.
- [146] Robertson KD. DNA methylation and human disease. *Nature Reviews Genetics* 2005;6:597–610.
- [147] Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 2013;500:477–81.
- [148] Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, et al. Genetic control of individual differences in gene-specific methylation in human brain. *The American Journal of Human Genetics* 2010;86:411–9.
- [149] Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsum C, et al. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* 2013;9:e1003763.
- [150] Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, et al. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* 2014;344:523–7.
- [151] Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A, et al. Reelin promoter hypermethylation in schizophrenia. *Proc Natl Acad Sci USA* 2005;102:9341–6.
- [152] Fikri RMN, Norlelawati AT, El-Huda ARN, Hanisah MN, Kartini A, Norsidah K, et al. Reelin (RELN) DNA methylation in the peripheral blood of schizophrenia. *Journal of Psychiatric Research* 2017;88:28–37.

- 
- [153] Abdolmaleky HM, Cheng K, Russo A, Smith CL, Faraone SV, Wilcox M, et al. Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005;134B:60–6.
- [154] Grayson DR, Chen Y, Costa E, Dong E, Guidotti A, Kundakovic M, et al. The human reelin gene: transcription factors (+), repressors (-) and the methylation switch (+/-) in schizophrenia. *Pharmacology & Therapeutics* 2006;111:272–86.
- [155] Ikegame T, Bundo M, Sunaga F, Asai T, Nishimura F, Yoshikawa A, et al. DNA methylation analysis of BDNF gene promoters in peripheral blood cells of schizophrenia patients. *Neurosci Res* 2013;77:208–14.
- [156] Veldic M, Guidotti A, Maloku E, Davis JM, Costa E. In psychosis, cortical interneurons overexpress DNA-methyltransferase 1. *Proc Natl Acad Sci USA* 2005;102:2152–7.
- [157] Iwamoto K, Bundo M, Yamada K, Takao H, Iwayama-Shigeno Y, Yoshikawa T, et al. DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia. *Journal of Neuroscience* 2005;25:5376–81.
- [158] Tsankova N, Renthal W, Kumar A, Nestler EJ. Epigenetic regulation in psychiatric disorders. *Nature Reviews Neuroscience* 2007;8:355–67.
- [159] Tremolizzo L, Carboni G, Ruzicka WB, Mitchell CP, Sugaya I, Tueting P, et al. An epigenetic mouse model for molecular and behavioral neuropathologies related to schizophrenia vulnerability. *Proc Natl Acad Sci USA* 2002;99:17095–100.
- [160] Matrisciano F, Tueting P, Dalal I, Kadriu B, Grayson DR, Davis JM, et al. Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. *Neuropharmacology* 2013;68:184–94.
- [161] Dong E, Ruzicka W, Grayson D, Guidotti A. DNA-methyltransferase 1 (DNMT1) binding to CpG rich GABAergic and BDNF promoters is increased in the brain of schizophrenia and bipolar disorder patients. *Schizophrenia Research* 2015;167:35–41.
- [162] Lee AH, Lange C, Ricken R, Hellweg R, Lang UE. Reduced brain-derived neurotrophic factor serum concentrations in acute schizophrenic patients increase during antipsychotic treatment. *Journal of Clinical Psychopharmacology* 2011;31:334–6.

- 
- [163] Hashimoto T, Bergen SE, Nguyen QL, Xu B, Monteggia LM, Pierri JN, et al. Relationship of brain-derived neurotrophic factor and its receptor TrkB to altered inhibitory prefrontal circuitry in schizophrenia. *Journal of Neuroscience* 2005;25:372–83.
- [164] Huang EJ, Reichardt LF. Neurotrophins: roles in neuronal development and function. *Annual Review of Neuroscience* 2001;24:677–736.
- [165] Kordi-Tamandani DM, Sahranavard R, Torkamanzahi A. DNA methylation and expression profiles of the brain-derived neurotrophic factor (BDNF) and dopamine transporter (DAT1) genes in patients with schizophrenia. *Molecular Biology Reports* 2012;39:10889–93.
- [166] Yasui D, Peedicayil J, Grayson DR. *Neuropsychiatric Disorders and Epigenetics*. Academic Press; 2016.
- [167] Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, et al. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *The American Journal of Human Genetics* 2008;82:696–711.
- [168] Xiao Y, Camarillo C, Ping Y, Arana TB, Zhao H, Thompson PM, et al. The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. *PLoS ONE* 2014;9:e95875.
- [169] Wockner L, Noble E, Lawford B, Young RM, Morris C, Whitehall V, et al. Genome-wide DNA methylation analysis of human brain tissue from schizophrenia patients. *Translational Psychiatry* 2014;4:e339.
- [170] Detich N, Bovenzi V, Szyf M. Valproate induces replication-independent active DNA demethylation. *Journal of Biological Chemistry* 2003;278:27586–92.
- [171] Melas PA, Rogdaki M, Ösby U, Schalling M, Lavebratt C, Ekström TJ. Epigenetic aberrations in leukocytes of patients with schizophrenia: association of global DNA methylation with antipsychotic drug treatment and disease onset. *The FASEB Journal* 2012;26:2712–8.
- [172] Dong E, Nelson M, Grayson DR, Costa E, Guidotti A. Clozapine and sulpiride but not haloperidol or olanzapine activate brain DNA demethylation. *Proc Natl Acad Sci USA* 2008;105:13614–9.
- [173] Pidsley R, Viana J, Hannon E, Spiers H, Troakes C, Al-Saraj S, et al. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biology* 2014;15:483.

- 
- [174] Wockner L, Morris C, Noble E, Lawford B, Whitehall V, Young R, et al. Brain-specific epigenetic markers of schizophrenia. *Translational Psychiatry* 2015;5:e680.
- [175] Teroganova N, Girshkin L, Suter CM, Green MJ. DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. *BMC Genetics* 2016;17:27.
- [176] Connor CM, Akbarian S. DNA methylation changes in schizophrenia and bipolar disorder. *Epigenetics* 2008;3:55–8.
- [177] Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. *Genome Biol* 2012;13:R8.
- [178] Jaffe AE, Gao Y, Deep-Soboslay A, Tao R, Hyde TM, Weinberger DR, et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nature Neuroscience* 2016;19:40.
- [179] Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology* 2016;17:176.
- [180] Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology* 2013;38:23–38.
- [181] Mendizabal I, Keller T, Zeng J, Soojin VY. Epigenetics and evolution. *Integr Comp Biol* 2014;54:31–42.
- [182] Gokhman D, Malul A, Carmel L. Inferring Past Environments from Ancient Epigenomes. *Molecular Biology and Evolution* 2017;34:2429–38.
- [183] Gokhman D, Meshorer E, Carmel L. Epigenetics: it's getting old. Past meets future in paleoepigenetics. *Trends in Ecology & Evolution* 2016;31:290–300.
- [184] Gokhman D, Tamir LA, Housman G, Rafinia MN, Colón MN, Gu H, et al. Recent Regulatory Changes Shaped the Human Facial and Vocal Anatomy. *bioRxiv* 2017:106955.
- [185] Caspi A, Moffitt TE. Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nature Reviews Neuroscience* 2006;7:583–90.
- [186] Caspi A, McClay J, Moffitt TE, Mill J, Martin J, Craig IW, et al. Role of genotype in the cycle of violence in maltreated children. *Science* 2002;297:851–4.

- 
- [187] Van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature* 2010;468:203–12.
- [188] Barres R, Yan J, Egan B, Treebak JT, Rasmussen M, Fritz T, et al. Acute exercise remodels promoter methylation in human skeletal muscle. *Cell Metabolism* 2012;15:405–11.
- [189] Nitert MD, Dayeh T, Volkov P, Elgzyri T, Hall E, Nilsson E, et al. Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. *Diabetes* 2012;61:3322–32.
- [190] Huxley J, Mayr E, Osmond H, Hoffer A. Schizophrenia as a genetic morphism. *Nature* 1964;204:220–1.
- [191] Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, et al. Schizophrenia: manifestations, incidence and course in different cultures A World Health Organization Ten-Country Study. *Psychological Medicine Monograph Supplement* 1992;20:1–97.
- [192] Crow TJ. The ‘big bang’ theory of the origin of psychosis and the faculty of language. *Schizophrenia Research* 2008;102:31–52.
- [193] Crow TJ. Schizophrenia as the price that Homo sapiens pays for language: a resolution of the central paradox in the origin of the species. *Brain Research Reviews* 2000;31:118–29.
- [194] Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M, et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* 2013;70:22–30.
- [195] Brown S. Excess mortality of schizophrenia. A meta-analysis. *Br J Psychiatry* 1997;171:502–8.
- [196] Burns JK. An evolutionary theory of schizophrenia: cortical connectivity, metarepresentation, and the social brain. *Behav Brain Sci* 2004;27:831–55; discussion 855–85.
- [197] Crow TJ. A Darwinian approach to the origins of psychosis. *Br J Psychiatry* 1995;167:12–25.
- [198] Crow TJ. Is schizophrenia the price that Homo sapiens pays for language? *Schizophr Res* 1997;28:127–41.



- 
- [199] Pujol J, Deus J, Losilla JM, Capdevila A. Cerebral lateralization of language in normal left-handed people studied by functional MRI. *Neurology* 1999;52:1038–1038.
- [200] Beeman MJ, Chiarello C. Complementary right-and left-hemisphere language comprehension. *Current Directions in Psychological Science* 1998;7:2–8.
- [201] Cook ND. *The brain code: Mechanisms of information transfer and the role of the corpus callosum*. Methuen; 1986.
- [202] Cook ND. Bihemispheric language: How the two hemispheres collaborate in the processing of language. *The Speciation of Modern Homo Sapiens* 2002;106.
- [203] Coney J, Evans KD. Hemispheric asymmetries in the resolution of lexical ambiguity. *Neuropsychologia* 2000;38:272–82.
- [204] Halpern ME, Güntürkün O, Hopkins WD, Rogers LJ. Lateralization of the vertebrate brain: taking the side of model systems. *Journal of Neuroscience* 2005;25:10351–7.
- [205] Crow TJ, Ball J, Bloom SR, Brown R, Bruton CJ, Colter N, et al. Schizophrenia as an anomaly of development of cerebral asymmetry: a postmortem study and a proposal concerning the genetic basis of the disease. *Archives of General Psychiatry* 1989;46:1145–50.
- [206] Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2006;2:e168.
- [207] Gittelman RM, Hun E, Ay F, Madeoy J, Pennacchio L, Noble WS, et al. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Research* 2015;25:1245–55.
- [208] Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 2011;478:476–82.
- [209] Srinivasan S, Bettella F, Mattingsdal M, Wang Y, Witoelar A, Schork AJ, et al. Genetic markers of human evolution are enriched in schizophrenia. *Biological Psychiatry* 2015.
- [210] Srinivasan S, Bettella F, Hassani S, Wang Y, Witoelar A, Schork AJ, et al. Probing the association between early evolutionary markers and schizophrenia. *PloS One* 2017;12:e0169227.

- 
- [211] Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol Biol Evol* 2015;32:1148–60.
- [212] Olden K, Freudenberg N, Dowd J, Shields AE. Discovering how environmental exposures alter genes could lead to new treatments for chronic illnesses. *Health Aff (Millwood)* 2011;30:833–41.
- [213] Choi JK, Kim SC. Environmental effects on gene expression phenotype have regional biases in the human genome. *Genetics* 2007;175:1607–13.
- [214] Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics* 2003;33 Suppl:245–54.
- [215] Meaney MJ, Szyf M. Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience* 2005;7:103.
- [216] Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nature Neuroscience* 2014;17:89–96.
- [217] Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. Transgenerational transmission of environmental information in *C. elegans*. *Science* 2017;356:320–3.
- [218] Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, et al. The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature* 2017;546:293–6.
- [219] Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature* 2017;546:289–92.
- [220] Norenzayan A, Shariff AF. The origin and evolution of religious prosociality. *Science* 2008;322:58–62.
- [221] Mellars P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proceedings of the National Academy of Sciences* 2006;103:9381–6.
- [222] Kind C-J, Ebinger-Rist N, Wolf S, Beutelspacher T, Wehrberger K. The smile of the Lion Man. Recent excavations in Stadel Cave (Baden-Württemberg, southwestern Germany) and the restoration of the famous Upper Palaeolithic figurine. *Quartär* 2014;61:129–45.

- 
- [223] Culotta E. On the origin of religion. *Science* 2009;326:784–7.
- [224] Whitley DS. *Cave paintings and the human spirit: the origin of creativity and belief*. Prometheus Books; 2009.
- [225] Balter M. Did working memory spark creative culture? *Science* 2010;328:160–3.
- [226] D’Errico F, Henshilwood C, Nilssen P. An engraved bone fragment from c. 70,000-year-old Middle Stone Age levels at Blombos Cave, South Africa: implications for the origin of symbolism and language. *Antiquity* 2001;75:309–18.
- [227] Atkinson QD. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 2011;332:346–9.
- [228] Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 2014;505:43–9.
- [229] Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 2012;338:222–6.
- [230] Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, et al. The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Research* 2015;43:8204–14.
- [231] Glazko GV, Nei M. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol* 2003;20:424–34.
- [232] Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, et al. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci USA* 2012;109:15716–21.
- [233] Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science* 2010;328:710–22.
- [234] Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, et al. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 2010;328:723–5.
- [235] Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genet* 2011;43:977–83.

- 
- [236] 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. *bioRxiv* 2017:145581.
- [237] Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet* 2010;42:508–14.
- [238] Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;466:707–13.
- [239] Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011;478:103.
- [240] Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015;518:197–206.
- [241] Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet* 2014;46:1173–86.
- [242] Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JR, Krapohl E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics* 2017.
- [243] Christoforou A, Dondrup M, Mattingsdal M, Mattheisen M, Giddaluru S, Nöthen MM, et al. Linkage-disequilibrium-based binning affects the interpretation of GWASs. *Am J Hum Genet* 2012;90:727–33.
- [244] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: 2017.
- [245] Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, et al. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet* 2013;9:e1003449.
- [246] Lee PH, O’Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 2012;28:1797–9.

- 
- [247] Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology* 2008;32:567–73.
- [248] Morris TJ, Beck S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods* 2015;72:3–8.
- [249] Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011;98:288–95.
- [250] Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology* 2016;17:208.
- [251] Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology* 2014;15:503.
- [252] Liu J, Siegmund KD. An evaluation of processing methods for HumanMethylation450 BeadChip data. *BMC Genomics* 2016;17:469.
- [253] Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Research* 2017;45:e22–e22.
- [254] Smyth G. *limma: Linear Models for Microarray Data*. In: Gentleman, Robert and Carey, Vincent J. and Huber, Wolfgang and Irizarry, Rafael A. and Dudoit, Sandrine, editor. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer New York; 2005, p. 397–420.
- [255] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.
- [256] Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363–9.
- [257] Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;41:200–9.
- [258] Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* 2012;28:2986–8.

- 
- [259] Hahne F, Ivanek R. Visualizing genomic data using Gviz and bioconductor. *Statistical Genomics: Methods and Protocols* 2016:335–51.
- [260] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Research* 2002;12:996–1006.
- [261] Martin M. AnnotationHub: Client to access AnnotationHub resources. R package version 2.12.0 2018.
- [262] Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.
- [263] Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;9:e1003118.
- [264] Hillman-Jackson J, Clements D, Blankenberg D, Taylor J, Nekrutenko A. Using Galaxy to perform large-scale interactive data analyses. *Curr Protoc Bioinformatics* 2012;Chapter 10:Unit10.5.
- [265] Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997–1004.
- [266] Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA. New statistical approaches exploit the polygenic architecture of schizophrenia—implications for the underlying neurobiology. *Current Opinion in Neurobiology* 2016;36:89–98.
- [267] Banerjee N, Polushina T, Bettella F, Giddaluru S, Steen VM, Andreassen OA, et al. Recently evolved human-specific methylated regions are enriched in schizophrenia signals. *BMC Evolutionary Biology* 2018;18.
- [268] Rogers AR, Bohlender RJ, Huff CD. Early history of Neanderthals and Denisovans. *Proceedings of the National Academy of Sciences* 2017;114:9859–63.
- [269] Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516–7.
- [270] Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* 2009;1:177–200.
- [271] Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. *Schizophrenia Bulletin* 2015;42:406–14.

- 
- [272] Melka MG, Laufer BI, McDonald P, Castellani CA, Rajakumar N, O'Reilly R, et al. The effects of olanzapine on genome-wide DNA methylation in the hippocampus and cerebellum. *Clinical Epigenetics* 2014;6:1.
- [273] Guidotti A, Grayson DR. DNA methylation and demethylation as targets for antipsychotic therapy. *Dialogues in Clinical Neuroscience* 2014;16:419.
- [274] Kinoshita M, Numata S, Tajima A, Shimodera S, Ono S, Imamura A, et al. DNA methylation signatures of peripheral leukocytes in schizophrenia. *Neuromolecular Med* 2013;15:95–101.
- [275] Accomando WP, Wiencke JK, Houseman EA, Nelson HH, Kelsey KT. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biology* 2014;15:R50.
- [276] Kinoshita M, Numata S, Tajima A, Yamamori H, Yasuda Y, Fujimoto M, et al. Effect of clozapine on DNA methylation in peripheral leukocytes from patients with treatment-resistant schizophrenia. *International Journal of Molecular Sciences* 2017;18:632.
- [277] Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* 2013;23:555–67.
- [278] Kato T, Iwamoto K. Comprehensive DNA methylation and hydroxymethylation analysis in the human brain and its implication in mental disorders. *Neuropharmacology* 2014;80:133–9.
- [279] Iorio MV, Piovan C, Croce CM. Interplay between microRNAs and the epigenetic machinery: an intricate network. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 2010;1799:694–701.
- [280] Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* 2015:aad6780.
- [281] Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* 2018;359:693–7.
- [282] Stahl E, Forstner A, McQuillin A, Ripke S, Ophoff R, Scott L, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv* 2017:173062.

- 
- [283] Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–74.
- [284] Giuliani C, Sazzini M, Bacalini MG, Pirazzini C, Marasco E, Fontanesi E, et al. Epigenetic variability across human populations: a focus on DNA methylation profiles of the KRTCAP3, MAD1L1 and BRSK2 genes. *Genome Biology and Evolution* 2016;8:2760–73.
- [285] Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. *Genome Research* 2013;23:1363–72.
- [286] Grönninger E, Weber B, Heil O, Peters N, Stäb F, Wenck H, et al. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet* 2010;6:e1000971.
- [287] Meyer D, Aguiar VR, Bitarello BD, Brandt DY, Nunes K. A genomic perspective on HLA evolution. *Immunogenetics* 2017:1–23.
- [288] Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* 2016;17:2042–59.
- [289] Jouzel J, Masson-Delmotte V, Cattani O, Dreyfus G, Falourd S, Hoffmann G, et al. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* 2007;317:793–6.
- [290] Lüthi D, Le Floch M, Bereiter B, Blunier T, Barnola J-M, Siegenthaler U, et al. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* 2008;453:379.
- [291] Zhang FF, Morabia A, Carroll J, Gonzalez K, Fulda K, Kaur M, et al. Dietary patterns are associated with levels of global genomic DNA methylation in a cancer-free population. *The Journal of Nutrition* 2011;141:1165–71.
- [292] Delgado-Cruzata L, Zhang W, McDonald JA, Tsai WY, Valdovinos C, Falci L, et al. Dietary Modifications, Weight Loss, and Changes in Metabolic Markers Affect Global DNA Methylation in Hispanic, African American, and Afro-Caribbean Breast Cancer Survivors, 2. *The Journal of Nutrition* 2015;145:783–90.
- [293] Ross SA. Epigenetics of Mediterranean diet: altering disease risk. *Mediterranean Diet*, Springer; 2016, p. 203–16.



- 
- [294] Muskiet FA, Kemperman RF. Folate and long-chain polyunsaturated fatty acids in psychiatric disease. *The Journal of Nutritional Biochemistry* 2006;17:717–27.
- [295] Kuipers RS, Fokkema MR, Smit EN, van der Meulen J, Boersma ER, Muskiet FA. High contents of both docosahexaenoic and arachidonic acids in milk of women consuming fish from lake Kitangiri (Tanzania): targets for infant formulae close to our ancient diet? *Prostaglandins, Leukotrienes and Essential Fatty Acids* 2005;72:279–88.
- [296] Quinn EA, Kuzawa CW. A dose-response relationship between fish consumption and human milk DHA content among Filipino women in Cebu City, Philippines. *Acta Paediatr* 2012;101:e439–45.
- [297] Hussey B, Lindley MR, Mastana SS. Omega 3 fatty acids, inflammation and DNA methylation: an overview. *Clinical Lipidology* 2017;12:24–32.
- [298] XXV World Congress of Psychiatric Genetics (WCPG): Poster Abstracts. *European Neuropsychopharmacology* 2017, *in press*.
- [299] Aubert M, Brumm A, Ramli M, Sutikna T, Saptomo EW, Hakim B, et al. Pleistocene cave art from Sulawesi, Indonesia. *Nature* 2014;514:223–7.
- [300] Harari YN. *Sapiens: A brief history of humankind*. Harvill Secker London; 2014.
- [301] Gilpin W, Feldman MW, Aoki K. An ecocultural model predicts Neanderthal extinction through competition with modern humans. *Proceedings of the National Academy of Sciences* 2016;113:2134–9.
- [302] Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 2010;38:e87.
- [303] Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 2014;514:445–9.
- [304] Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci USA* 2013;110:2223–7.
- [305] Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas A-S, Manica A, Moltke I, et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* 2014;346:1113–8.

- 
- [306] Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 2010;463:757–62.
- [307] Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 2014;505:87–91.
- [308] Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 2015;524:216–9.
- [309] Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015;528:499–503.
- [310] Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 2014;513:409–13.
- [311] Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, et al. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research* 2014;24:454–66.
- [312] Hanghøj K, Seguin-Orlando A, Schubert M, Madsen T, Pedersen JS, Willerslev E, et al. Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. *Molecular Biology and Evolution* 2016;33:3284–98.
- [313] Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootenck S, et al. A functional neuroanatomy of hallucinations in schizophrenia. *Nature* 1995;378:176–9.
- [314] Shergill SS, Brammer MJ, Williams SC, Murray RM, McGuire PK. Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. *Archives of General Psychiatry* 2000;57:1033–8.
- [315] Lee CS, Theriault DJ. The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence* 2013;41:306–20.
- [316] Vandervort LR, Schimpf PH, Liu H. How working memory and the cerebellum collaborate to produce creativity and innovation. *Creativity Research Journal* 2007;19:1–18.

- 
- [317] De Dreu CK, Nijstad BA, Baas M, Wolsink I, Roskes M. Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin* 2012;38:656–69.
- [318] Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics* 2013;92:197–209.
- [319] Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet* 2013;9:e1003455.
- [320] Andreassen OA, Harbo HF, Wang Y, Thompson W, Schork A, Mattingsdal M, et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Molecular Psychiatry* 2015;20:207.

---

## 7. Papers I-III







RESEARCH ARTICLE

Open Access



# Recently evolved human-specific methylated regions are enriched in schizophrenia signals

Niladri Banerjee<sup>1,2</sup>, Tatiana Polushina<sup>1,2</sup>, Francesco Bettella<sup>3,4</sup>, Sudheer Giddaluru<sup>1,2</sup>, Vidar M. Steen<sup>1,2</sup>, Ole A. Andreassen<sup>3,4</sup> and Stephanie Le Hellard<sup>1,2,5\*</sup> 

## Abstract

**Background:** One explanation for the persistence of schizophrenia despite the reduced fertility of patients is that it is a by-product of recent human evolution. This hypothesis is supported by evidence suggesting that recently-evolved genomic regions in humans are involved in the genetic risk for schizophrenia. Using summary statistics from genome-wide association studies (GWAS) of schizophrenia and 11 other phenotypes, we tested for enrichment of association with GWAS traits in regions that have undergone methylation changes in the human lineage compared to Neanderthals and Denisovans, i.e. human-specific differentially methylated regions (DMRs). We used analytical tools that evaluate polygenic enrichment of a subset of genomic variants against all variants.

**Results:** Schizophrenia was the only trait in which DMR SNPs showed clear enrichment of association that passed the genome-wide significance threshold. The enrichment was not observed for Neanderthal or Denisovan DMRs. The enrichment seen in human DMRs is comparable to that for genomic regions tagged by Neanderthal Selective Sweep markers, and stronger than that for Human Accelerated Regions. The enrichment survives multiple testing performed through permutation ( $n = 10,000$ ) and bootstrapping ( $n = 5000$ ) in INRICH ( $p < 0.01$ ). Some enrichment of association with height was observed at the gene level.

**Conclusions:** Regions where DNA methylation modifications have changed during recent human evolution show enrichment of association with schizophrenia and possibly with height. Our study further supports the hypothesis that genetic variants conferring risk of schizophrenia co-occur in genomic regions that have changed as the human species evolved. Since methylation is an epigenetic mark, potentially mediated by environmental changes, our results also suggest that interaction with the environment might have contributed to that association.

**Keywords:** Differentially methylated regions, Schizophrenia, Evolution, Epigenetics, Height, Neanderthal selective sweep score, Human accelerated regions

## Background

Schizophrenia is a psychiatric disorder that has been reported throughout human history, possibly as far back as 5000 years [1, 2]. Family, twin and adoption studies estimate that schizophrenia has a high heritability of 60–90% [3–6]. Today, schizophrenia is estimated to have a prevalence of

1%. It is associated with reduced fertility and increased mortality [7–11], and its persistence despite this heavy burden is paradoxical. Power et al. [11] leveraged Swedish registry data to demonstrate the reduced fecundity of patients with schizophrenia, despite the novel finding that sisters of individuals with schizophrenia had higher fitness than controls. They henceforth suggested hitherto unknown mechanisms for persistence of the disease. One explanation for this persistence is that evolution has indirectly selected the disease instead of eliminating it - the disease may co-segregate with creativity and intellectual prowess, providing selective advantages to the kin of affected individuals [9, 12]. Crow first

\* Correspondence: [stephanie.hellard@uib.no](mailto:stephanie.hellard@uib.no)

<sup>1</sup>NORMENT - K.G. Jebsen Center for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway

<sup>2</sup>Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.



argued that language and psychosis may have common origins, which could explain the persistence of schizophrenia in human populations [12, 13]. This evolutionary hypothesis of the origins of schizophrenia can now be tested, thanks to the identification of genetic factors implicated in schizophrenia [14–16] and the availability of datasets that reflect recent genomic evolution in humans [17–19].

Large genome-wide association studies (GWAS) have identified thousands of variants that are associated with schizophrenia [14–16] but our mechanistic understanding of the candidate variants is poor. One approach to investigating the function of schizophrenia-associated variants is comparative genomics, which investigates the evolutionary relevance of certain genomic regions [20]. This field has introduced new datasets to test disease origins in humans, including Human Accelerated Regions (HARs) and Neanderthal Selective Sweep (NSS) scores [18, 19]. HARs are genomic regions that are highly conserved in non-human species, but have undergone rapid sequence change in the human lineage [20–24]. Xu et al. [18] showed that genes near HARs are enriched for association with schizophrenia. Neanderthals were hominids that co-existed and even bred with modern humans [25, 26]. Comparison of Neanderthal and human genome sequences [27, 28] has revealed genomic regions that have experienced a selective sweep in modern humans, presumably following a favorable mutation [28]. Negative NSS scores can be used to pinpoint mutations (usually single nucleotide changes) that were positively selected in humans as they diverged from Neanderthals. Srinivasan et al. [19] found that genomic regions tagged by negative NSS scores show enrichment of association with schizophrenia.

Using specific interpretation of genome sequencing in two recently extinct hominids, Neanderthals and Denisovans, Gokhman et al. [29] mapped genome-wide methylation levels (i.e. the methylome) and compared them to modern humans. While 99% of the methylation maps were identical in the three hominids, nearly 2000 differentially methylated regions (DMRs) were identified, which give the first clues about the role of epigenomic evolution in generating anthropometric differences between modern humans and their ancient cousins [29]. These DMRs provide a dataset of evolutionary annotations complementary to pre-existing datasets. Unlike HARs and NSS scores, which are based on DNA sequence changes, DMRs provide information on the evolution of epigenomes. Since epigenomes can act as an interface with the environment [30, 31], these datasets provide the opportunity to investigate environmentally driven evolutionary changes. Keeping in mind the evolutionary hypothesis for schizophrenia proposed by Crow, we thus examined if these evolutionary DMRs are enriched for

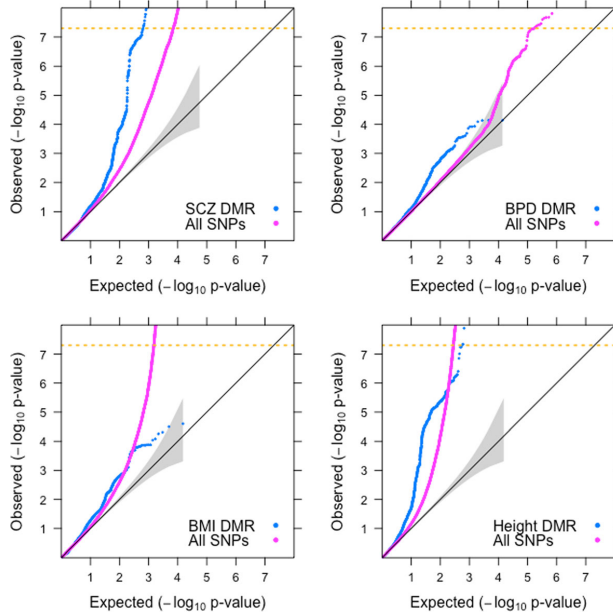
association with schizophrenia. We also examined a range of human traits to compare the possible enrichment in other traits. Using previously published methodologies [19, 32, 33] and publicly available GWAS datasets we systematically analyzed twelve diverse phenotypes to investigate the potential role of regions susceptible to epigenetic variation in the emergence of specific traits in the human lineage.

## Results

### SNPs in human-specific DMRs are enriched for association with schizophrenia.

The genomic locations of human-specific DMRs were obtained from data published by Gokhman et al. [29] (see [Methods](#) for full details). GWAS summary statistics for 12 common traits were obtained from published datasets: schizophrenia [14], bipolar disorder (BPD) [34], attention deficit hyperactivity disorder (ADHD) [35], rheumatoid arthritis [36], high density lipoprotein [37], low density lipoprotein [37], triglycerides [37], total cholesterol [37], systolic blood pressure [38], diastolic blood pressure [38], body mass index [39], and height [40]. The GWAS datasets are summarized in Additional file 1, Table S1. For each trait, we generated a list of single nucleotide polymorphisms (SNPs) within DMRs (positional annotation) and a list of SNPs in linkage disequilibrium (LD-based annotation) with markers within DMRs (Additional file 1, Table S1).

We used quantile-quantile (QQ) plots as described by Schork et al. [32] to test whether the DMR SNPs are enriched for association with the GWAS trait compared to the complete set of SNPs (see [Methods](#) for additional details). In such plots the baseline is the null line of no difference between expected distribution of  $p$ -values and observed  $p$ -values. Deviation of the observed data distributions from the expected data distribution indicates the presence of true associations. When the  $p$ -values for a set of selected markers show greater leftwards deflection, they are enriched for association compared to the overall GWAS set. For the schizophrenia GWAS, enrichment was observed both for SNPs in LD with markers in DMRs (Fig. 1, Additional file 1: Fig. S1) and for SNPs located within DMRs (Fig. 2). Although there was a slight leftward deflection in the higher  $p$ -values (smaller negative  $\log_{10}$  of  $p$ -values) in some other traits (e.g. height; Fig. 1, Additional file 2: Figure S1), the observed enrichment only crosses the genome-wide significance level of  $5 \times 10^{-8}$  for the schizophrenia SNPs. The enrichment of disease-associated markers in DMRs is thus specific to schizophrenia and is independent of LD.



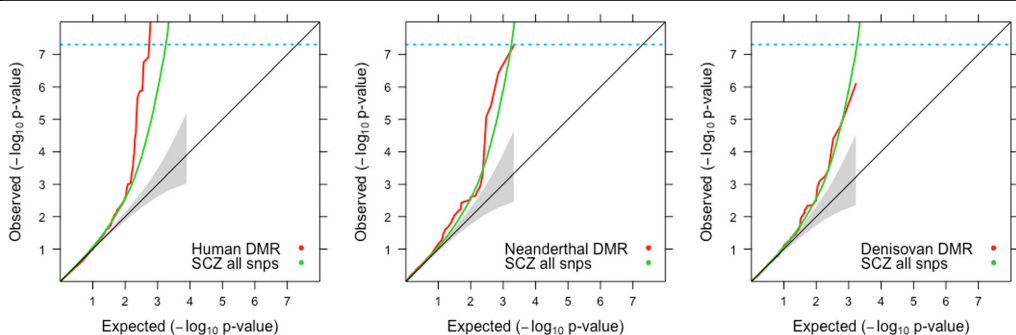
**Fig. 1** Enrichment of DMR SNPs across SCZ, BPD, BMI and Height. Quantile-Quantile (QQ) plots of GWAS SNPs for Schizophrenia (SCZ) with the extended MHC region masked (chr6: 25-35 Mb), Bipolar Disorder (BPD), Body Mass Index (BMI) and Height. The X-axis shows expected  $-\log_{10} p$ -values under the null hypothesis. The Y-axis shows actual observed  $-\log_{10} p$ -values. The values for all GWAS SNPs are plotted in pink while the values for SNPs in linkage disequilibrium (LD) with DMRs are plotted in blue. Leftwards deflections from the null line (grey diagonal line) indicate enrichment of true signals - the greater the leftward deflection, the stronger the enrichment. Genomic correction was performed on all SNPs with global lambda

**Human-specific DMR enrichment in schizophrenia is independent of the MHC region, other genomic annotations and total markers genotyped**

The Major histocompatibility complex (MHC) region harbors several significant schizophrenia markers and

could potentially bias our results because of long-range LD. The QQ plots show that the enrichment remains when the MHC is excluded (Fig. 1) or included (Fig. 2).

The schizophrenia GWAS had the highest density of markers genotyped (~ 9.4 million) and thus had the most



**Fig. 2** Comparison of enrichment of association with schizophrenia for SNPs within Human, Neanderthal and Denisovan DMRs. The figure shows QQ plots for all schizophrenia (SCZ) GWAS SNPs in green while SNPs within the species-specific DMRs are plotted in red. The location of the MHC is unknown in Neanderthal and Denisovan genomes, so the MHC region was not masked in the human genome

SNPs in DMR regions (Additional file 1: Table S1), which could artificially inflate the enrichment. We normalized the total number of DMR SNPs with the total number of SNPs genotyped in each GWAS and found that the proportion of SNPs in DMRs is nearly identical for all traits (Additional file 1: Figure S3). To further eliminate the possibility that the enrichment is due to variation in the number of markers analyzed, we extracted ~2.4 million SNPs that were common across the twelve GWAS. Although not as strong as with the full set, the deflection observed for the schizophrenia GWAS remains higher than any other trait (Additional file 1: Figure S1), indicating the presence of significant disease markers in DMRs. These validations point to a true enrichment of association of the DMR SNPs with schizophrenia that is independent of the number of markers in a GWAS. It should be noted that we cannot rule out enrichment in the ADHD and BPD GWAS, because they are lacking in power (Additional file 1: Figure S1).

Additionally, we considered the distribution of schizophrenia-associated SNPs based on genomic annotations of 5' untranslated regions (5'UTRs), Exons, Introns and 3' untranslated regions (3'UTRs) [32]. Contrary to previously published findings [32], the enrichment was highest for intronic SNPs and lowest for 5'UTR SNPs (Additional file 1: Figure S4).

#### Only human-specific DMRs are enriched for association with schizophrenia

Next, we used QQ plots to test whether markers located in the Neanderthal- and Denisovan-specific DMRs are enriched for association with schizophrenia. Coordinates for these DMRs were obtained from data published by Gokhman et al., 2014 [29] (see [Methods](#) for details). Since we do not know the precise coordinates of the MHC for Neanderthals and Denisovans, the analysis for human DMRs included the MHC region. No enrichment was observed for Neanderthal or Denisovan DMRs (Fig. 2). It should be noted that this approach may not be appropriate for testing Neanderthal- and Denisovan-specific DMRs since (a) the schizophrenia GWAS was conducted in humans; (b) SNP and LD information is available only for humans; (c) the three hominids had variable number of DMRs, which affected the number of SNPs captured via positional annotation.

#### Comparison of human DMRs with other evolutionary annotations

We compared the enrichment observed for the human DMRs with the enrichment previously reported for NSS markers and HARs [18, 19] (see [Methods](#) for details). We first compared the enrichment via QQ plots and find that the enrichment of human DMRs in schizophrenia is

comparable to that observed for NSS markers and far greater than that observed for HARs (Fig. 3).

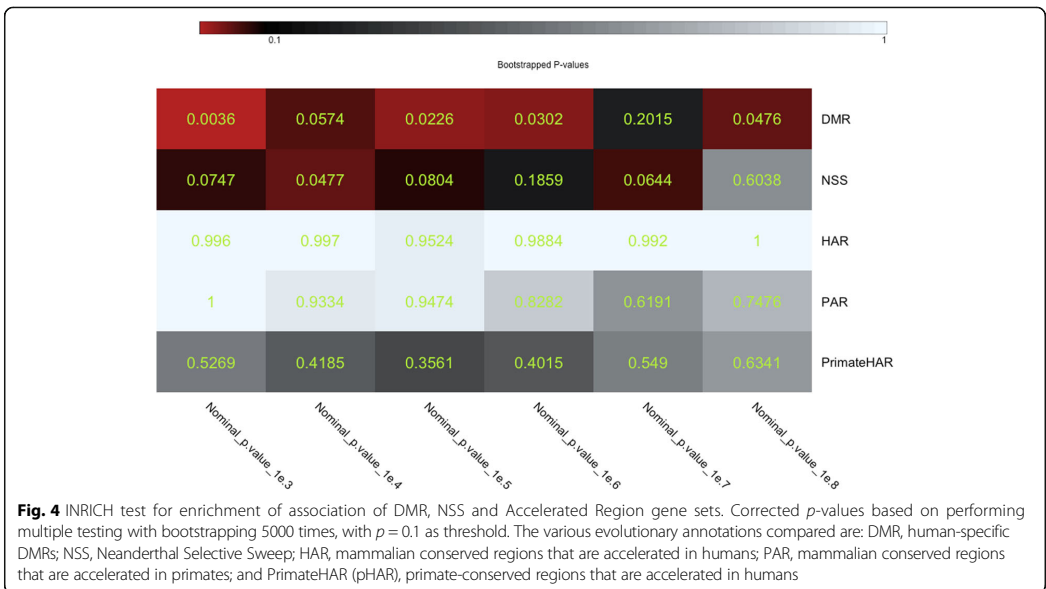
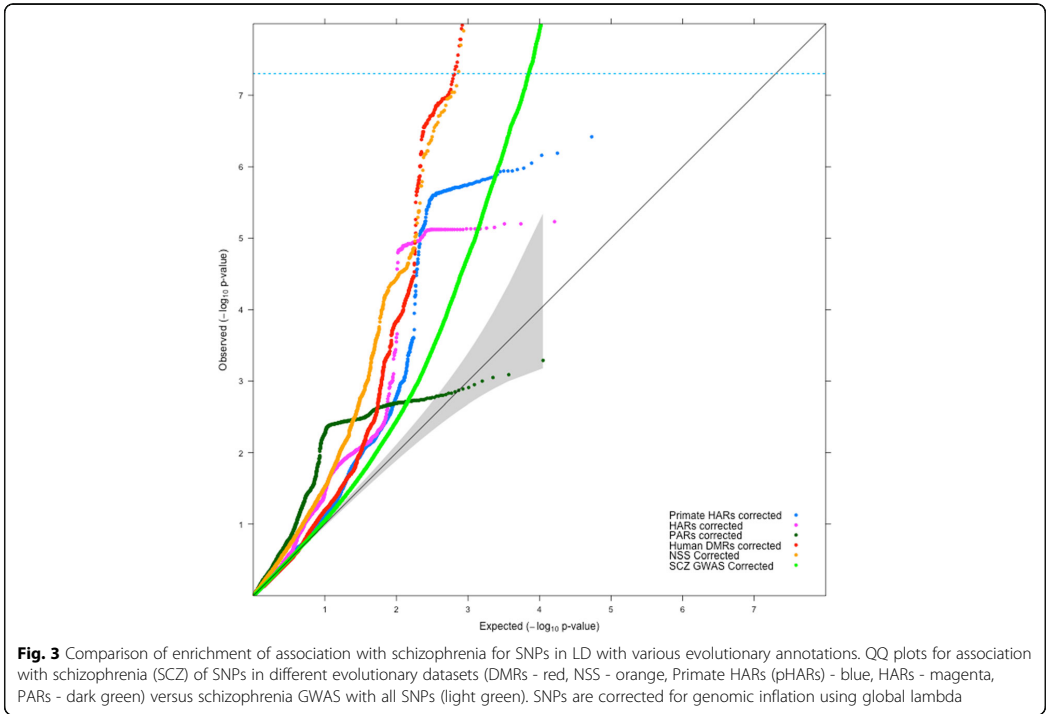
In these analyses, it was important to check the extent of overlap of markers (SNPs) annotated to various genomic regions of DMRs, NSS markers and HARs. Reassuringly, the various evolutionary annotations do not share the same group of markers, indicating that we did not test the same regions or SNPs (Additional file 1: Figure S2). The overlap between NSS markers and DMR markers involved less than 0.5% of all NSS markers and less than 0.2% of all DMR markers (Additional file 1: Figure S2). The SNPs in the DMRs thus represent a different group of markers that have not been annotated or analyzed previously from an evolutionary standpoint (Additional file 2, Additional file 3).

#### Statistically-significant enrichment exists for human DMRs

To determine the statistical significance of the DMR enrichment in schizophrenia, we utilized the INRICH software pipeline. INRICH is a pathway analysis tool for GWAS, designed for detecting enriched association signals of LD-independent genomic regions within biologically relevant gene sets (in our case genes which contain DMRs). It performs permutation and bootstrapping procedures to determine the significance of association of markers in LD intervals while maintaining the SNP density and gene density of the original intervals [33]. INRICH confirmed significant ( $p < 0.05$ ) enrichment of association for human DMRs with schizophrenia after correcting for multiple testing through bootstrapping at most  $p$ -value thresholds of LD intervals. Additionally, INRICH independently verified the previously reported enrichment of NSS markers with schizophrenia [19] (Fig. 4). Furthermore, INRICH identified gene-level enrichment of association for DMRs with height (Additional file 1: Figure S5), while at the SNP level the enrichment in height was seen only for smaller effects, i.e. the enrichment did not remain below  $p < 10^{-8}$ .

#### Pathway analysis

We utilized Ingenuity Pathway Analysis (IPA) to analyze DMR SNPs that show enrichment of association with schizophrenia (for details of the genes analyzed, please refer to the 'Pathway analysis' section in the [Methods](#)). We found 'CREB signaling in neurons' and 'Synaptic long term potentiation' amongst the top canonical pathways when analyzing pathways overrepresented in nervous system. Additionally, under physiological systems, 'Nervous system development and function' is also enriched (Additional file 1: Table S2). We repeated the same analysis for NSS markers as they also show enrichment of association with schizophrenia. 'CREB signaling in neurons' was also amongst the top canonical pathways



for enriched NSS markers (Additional file 1: Table S4). Additionally, we repeated the analyses with all organ systems and even then, 'CREB signaling in neurons' and 'Synaptic long term potentiation' emerged amongst the top canonical pathways for both enriched DMRs (Additional file 1: Table S3) and enriched NSS (Additional file 1: Table S5). This is an interesting result since there is very little marker overlap between the DMR and NSS SNPs (Additional file 1: Figure S2). Interestingly, genes containing enriched DMRs are also overrepresented in 'Hair and skin development' when considering all organ systems (Additional file 1: Table S3). This may suggest potential gene-by-environment interactions, modulated by methylation variation over human evolution (see [Discussion](#) below).

## Discussion

Our results suggest that SNPs in regions of the human genome that have undergone recent changes in DNA methylation status are enriched for association with schizophrenia, and to a lesser extent with height. Amongst all the traits analyzed, the enrichment observed in QQ plots was strongest for schizophrenia and passed the genome-wide significance threshold of  $5 \times 10^{-8}$  when the MHC was both excluded (Fig. 1) and included (Fig. 2). INRICH analysis confirms significant enrichment ( $p < 0.01$ ) in human DMRs that survived multiple testing through bootstrapping (Fig. 4) for association with schizophrenia, and also suggests a possible effect on height (Additional file 1: Figure S5).

Xu et al. [18] and Srinivasan et al. [19] respectively demonstrated that variants located in HARs and in regions containing NSS markers were enriched for association with schizophrenia. In our study, we compared the evolutionary enrichments of schizophrenia risk variants in DMRs, NSS markers and HARs. We validate the results of Srinivasan et al. [19] (Fig. 3 and Fig. 4). HARs do not show enrichment of disease markers by QQ plots and INRICH, unlike NSS markers and DMRs (Fig. 3 and Fig. 4). This difference with the report of Xu et al. could be due to a different freeze of the gene database used; it could also be because Xu et al. used a more stringent Hardy-Weinberg equilibrium (HWE) threshold to filter out markers from the schizophrenia GWAS [14], a step we could not replicate as the genotype data are not publicly available. We used the publicly available schizophrenia dataset that has a HWE  $p$ -value  $> 10^{-6}$  in controls and  $p$ -value  $> 10^{-10}$  in cases [14]. Interestingly, all the evolutionary annotations (DMRs, NSS markers and HARs) cover different sections of the genome with very little overlap between them (Additional file 1: Figure S2). Between the three evolutionary annotations, nearly 70,000 SNPs occur around regions with evolutionary significance (Additional file 1: Figure S2). Our results

supply a wealth of information on genomic regions that are important for the evolution of humans and are also enriched for schizophrenia risk variants (NSS markers and DMRs, Additional file 3). In addition, our study provides genetic support from two independent datasets that regions which differ between modern and ancient hominids could be implicated in the development of schizophrenia. An interesting hypothesis to consider is the possibility that methylation patterns are potentially driven by the genomic sequence underneath. This hypothesis is supported by preliminary findings presented at the recently concluded World Psychiatric Genetics Congress [41]. As such it is possible that the human specific DMRs analyzed here represent regions of the human genome where the underlying sequence might have diverged from Neanderthals and Denisovans. This hypothesis may be partially true as Gokhman et al. [29] observed that some, but not all of the methylation changes were indeed driven by sequence changes. On the other hand, there also exist metastable epialleles where there are methylation differences in genetically identical individuals [42]. As such, this would suggest that not all methylation differences are driven by the underlying genomic sequence alone. We did not test whether the schizophrenia markers are human-specific or not and therefore should be investigated in future research.

Neanderthal- or Denisovan-specific DMRs showed no enrichment of association (Fig. 2). This suggests that SNPs conferring vulnerability to schizophrenia occur in genomic regions whose methylation levels were altered in the modern human lineage but not in the ancestral lineages. It is possible that the evolutionary changes driving the variation in methylation status could also have made the human lineage more vulnerable to schizophrenia. A caveat to this result is that the LD structure in archaic genomes is unknown, so we cannot test LD-based enrichment in Neanderthal or Denisovan genomes. Our inter-lineage analyses with enrichment plots were thus restricted to SNPs occurring exclusively *within* DMRs. The other limitation to this comparative approach is that the GWAS data is specific to modern humans.

In previous studies [32], it was reported that 5'UTRs are the functional annotation harboring the most association with a given trait. However, the DMRs enriched for association with schizophrenia tended to localize in intronic regions (Additional file 1: Figure S4), which is in agreement with the expectation that methylation regions should not be localized in exons and UTRs. This shows that using more information to label some genomic regions in greater detail, such as potential regulatory regions in introns, might give a more precise annotation of regions of association.

Despite the genetic overlap between bipolar disorder and schizophrenia, we do not find evidence of enrichment of association of DMRs with bipolar disorder either at the SNP level (Fig. 1, Additional file 1: Figure S1) or the gene level (data available on request). This could possibly be due to lack of sufficient power in the bipolar disorder GWAS [34]. Additionally, the gene-level approach utilized by INRICH depicts enrichment of association of human DMRs with height (Additional file 1: Figure S5). This evidence is lacking at the SNP level as depicted by QQ plots (Fig. 1, Additional file 1: Figure S1). We speculate that this could be due to the difference in testing DMR-localized SNPs compared to genes flanking human DMRs.

Although the DMRs utilized here were obtained from bone samples, Gokhman et al. [29] assert that the DMRs refer to species-specific methylation differences and not tissue-specific variations [43]. Similarly, Hernando-Herraez et al. [44] noted that species-specific DMRs tend to be conserved across tissues and as such should not represent tissue-specific variations. Other studies also showed that neurological systems were enriched for methylation differences even when the tissue samples analyzed were not neurological [45–47]. Therefore, we believe that our results are valid for a ‘brain’ phenotype even though the DMRs were derived from non-brain tissues. The enrichment seen for schizophrenia also corroborates the results of Gokhman et al. [29] who reported that DMRs were more enriched around genes implicated in the nervous system amongst all the organ systems tested for evolutionary changes in methylation patterns. Hernando-Herraez et al. [44] also found that methylation differences between humans and great apes were located around genes controlling neurological and developmental features. It is therefore possible that the methylation differences were mediated by evolution of genomic regions controlling neurodevelopmental processes. The results of pathway analysis are consistent with this. Both the DMR and NSS regions that are enriched for association with schizophrenia contain genes that are overrepresented in ‘CREB signaling in neurons’ and ‘Synaptic long term potentiation’.

Our results hint that epigenomic evolution has taken place in genomic regions implicated in the aetiology of schizophrenia. Furthermore, these regions harbor markers that are involved in the regulation of various neurodevelopmental pathways. The fact that methylation changes also took place in these very same regions suggests a complex gene-by-environment interaction in the evolution of humans, especially for pathways that led to the development of our brain. While it is known that various factors from the environment can make long-lasting changes in DNA methylation patterns that can be subsequently inherited at a population level [30, 31, 48],

the true significance of our findings from an evolutionary standpoint suggests that the superior mental abilities of our species may in part have been driven by environmental factors during the past 300,000 years [49, 50].

It is difficult to put an exact date on the emergence of the superior mental abilities that define the modern *Homo sapiens*. Anthropologists often date the onset of the advanced intellectual abilities of *Homo sapiens* from about 70,000 years onwards [51], a period which saw the emergence of art, religion [52, 53] and possibly spoken language [54]. From an evolutionary perspective, it suggests a massive leap in the animal kingdom because *Homo sapiens* became the first species not only to develop the capacity to think and imagine things that do not exist [52, 53], but also to communicate these ideas to other members of the species [54]. This ability would have been critical for effective coordination and cooperation within large groups and may even have been needed to keep a group together [55, 56]. The genomic approach to analyze mental disorders used in the present study and other studies can interrogate the effect of changes which appeared in the last 300,000 years [49, 50], but it will clearly be interesting to trace mores recent changes. If a similar method used in the reconstruction of Neanderthal and Denisovan genomes [29, 57] could be implemented on samples of ancient *Homo sapiens* from different time periods [58–65], then theoretically it should be possible to reconstruct the methylomes and regions of recent evolution from ancient humans [66]. Subsequently, more detailed ‘time-course’ analyses of changes in methylation patterns and in other regions of recent evolution and their implications in schizophrenia will surely result in more detailed elucidation of the evolutionary hypothesis of schizophrenia. This is the promise of the novel field of paleoepigenetics that seeks to infer past environmental cues that affected the epigenomes of ancient individuals [42, 43].

## Conclusions

In summary, we have demonstrated that human genomic regions whose methylation status was altered during evolution are enriched in markers that show association with schizophrenia. Our results concur with previous genomic studies demonstrating that methylation changes in *Homo sapiens* have had the greatest impact on the nervous system and provide evidence that epigenomic evolution plays a role in conferring a high risk of schizophrenia on humans. Future research should attempt to perform a finer temporal resolution of the origins of psychosis through the prism of evolutionary epigenomics. To explore the period of evolution before the *Homo* lineage, it would also be interesting to determine whether methylation signatures from primates are enriched for schizophrenia markers. Future research

should also investigate the influence of human-specific DMRs on height.

## Methods

### Differentially methylated region data

Coordinates for DMRs were obtained from data publicly available in Supplementary Table S2 of Gokhman et al., 2014 [29]. This file contained DMRs inferred by comparing genome sequence of fossilized Neanderthal and Denisovan limb samples with methylation data from osteoblasts of modern humans. From the genomes of the Neanderthal and Denisovan samples, Gokhman et al. inferred methylation by utilizing the natural degradation of methylated cytosine (C) to thymine (T) to create a C→T ratio [29]. The methylation information, in the form of C→T ratio, was then compared with each of the three species and classified according to the hominid in which the methylation change occurred, i.e. human-specific, Neanderthal-specific and Denisovan-specific DMRs. These DMRs do not represent tissue-specific methylation but species-specific methylation [29]. The human-specific DMRs comprise regions that have both gained and lost methylation in comparison to Neanderthal- and Denisovan-specific DMRs. DMRs that could not be classified reliably in any of the three species (unclassified DMRs) [29] were not used. Full methodological details for assigning DMRs are in the Additional file of the original paper [29].

### HAR data

Genomic coordinates were obtained from publicly available data ([docpollard.com/2x](http://docpollard.com/2x)) for three classes of human accelerated region: HARs, in which regions conserved in mammals are accelerated in humans; PARs, in which regions conserved in mammals are accelerated in primates; and pHARs, in which regions conserved in primates (but not other mammals) are accelerated in humans. Conversion to hg19 assembly was performed using the liftOver tool from the UCSC Genome Browser.

### NSS data

NSS data was obtained as a list of markers with corresponding NSS values from Srinivasan et al. [19]. Markers with negative values, indicating positive selection in humans, were filtered out and used for analysis.

### GWAS data

Summary statistics from GWAS of 12 common traits were obtained from published datasets: schizophrenia (SCZ) [14], bipolar disorder (BPD) [34], attention deficit hyperactivity disorder (ADHD) [35], rheumatoid arthritis (RA) [36], blood lipid markers (high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides (TG), total cholesterol (TC)) [37], blood pressure (systolic

blood pressure (SBP), diastolic blood pressure (DBP)) [38], body mass index (BMI) [39], and height [40]. For studies published with hg18 coordinates (BPD, SBP, DBP, HDL, LDL, TG, TC, ADHD, RA), conversion to hg19 was performed using the command line version of the liftOver tool from the UCSC Genome Browser ([http://hgdownload.cse.ucsc.edu/downloads.html#utilities\\_downloads](http://hgdownload.cse.ucsc.edu/downloads.html#utilities_downloads)). For BMI and height SNPs, the genomic coordinates were obtained by mapping them to the assembly of 1000 Genomes Project Phase 1 reference panel SNPs [67].

### SNP assignment to DMRs

SNPs were assigned to DMRs with LDsnpR [68] using positional binning and LD (linkage disequilibrium)-based binning in R [69]. We used both methods because DMR-localized SNPs that were not genotyped in a specific GWAS would be missed if we used positional binning alone [68] (Additional file 1: Table S1). The LD file utilized in HDF5 format was constructed on the European reference population of 1000 Genomes Project and can be publicly downloaded at: <http://services.cbu.uib.no/software/ldsnpr/Download>.

### Enrichment analyses with stratified quantile-quantile (QQ) plots

QQ plots are an effective tool to visualize the spread of data and any deviations from the expected null distributions. They are frequently utilized in GWAS to depict enrichment of true signals. When the observed distribution of data matches the expected distribution, there is a lack of enrichment and a line of equality is obtained that depicts the null hypothesis. A distribution such as this reflects no enrichment of observed over expected data distribution. However, if the observed and expected distributions differ, there will be deviation from this null line. As described in detail by Schork et al. [32], leftwards deflections from this null line represent enrichment. The higher the leftward deflection, the greater is the enrichment of true signals. In GWASs, due to the extremely low  $p$ -values of SNPs, it is common to depict  $p$ -values by converting them to negative  $\log_{10}$  values so that smaller  $p$ -values give higher negative logarithmic values. We plotted the negative  $\log_{10}$  of the observed  $p$ -values of SNPs against the expected negative  $\log_{10}$  of a normal distribution. The distributions were corrected for genomic inflation by  $\lambda_{GC}$ . This method of enrichment was used to show for example [32] that specific genomic regions are enriched for trait-associated SNPs and are much more likely to associate with a given trait than SNPs distributed across a genome. In other words, when SNPs are stratified according to specific genomic regions, there is a greater enrichment of true signals than what is observed in the GWAS. Using a similar

approach, we binned SNPs that fall in DMR regions and plotted the stratified  $p$ -value distribution.

### Enrichment analyses with INRICH

The stratified QQ plots are a useful visual tool for observing the presence or absence of enrichment of true signals in a given set of SNPs. However, to quantify the enrichment visually observed, we used the INterval ENRICHment Analysis (INRICH) tool. It is a robust bioinformatics pipeline to determine enrichment of genomic intervals implicated by LD with predefined or custom gene sets [33]. It takes into account several potential biases that can otherwise lead to false positives. It is well suited for testing GWAS-implicated SNPs for association with gene sets as it controls for variable gene size, SNP density, LD within and between genes, and overlapping genes with similar annotations. We followed the procedure described by Xu et al. [18], with the extended MHC region (chr6:25–35 Mb) masked and SNPs with minor allele frequency (MAF) < 0.05 excluded. Full details may be found in Additional file 1.

### Pathway analysis

Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) from QIAGEN (<https://www.qiagen.com/no/shop/analytics-software/biological-data-tools/ingenuity-pathway-analysis/>, last accessed 26th August 2016). The reference set was Ingenuity Knowledge Base (Genes). Both direct and indirect relationships were analyzed. All data sources were included with the confidence parameter set to experimentally observed and highly predicted pathways for Human. Under ‘Tissues & Cell Lines’, we performed the analysis once with all organ systems and once with only the nervous system. 5338 enriched DMR SNPs in 329 enriched DMRs (Additional file 3) were mapped to 349 unique RefSeq genes and 446 RefSeq genes in LD using the method of Schork et al. [32]. Genes in LD blocks containing enriched NSS markers were determined in a similar manner. 4276 enriched NSS markers mapped to 648 overlapping RefSeq genes and 1363 RefSeq genes in LD. IPA was performed on these gene lists.

### Additional Files

**Additional file 1:** Additional Method, Figures and Tables (DOCX 475 kb)

**Additional file 2:** Annotation of all DMRs with schizophrenia-associated SNPs. This file contains annotation of all the human-lineage specific DMRs that are associated with schizophrenia markers. Details of the various markers present within each DMR is provided, along with the marker with most significant  $p$ -value. (XLSX 263 kb)

**Additional file 3:** Annotation of enriched DMRs with genes, promoters, CpG islands and enhancers. This file contains detailed annotation of those human-lineage specific DMRs that are enriched for association with schizophrenia markers (except those in the MHC region). Compared to

Additional file 2, these DMRs represent those that are enriched and whether they are present in any genes, promoters, enhancers or CpG islands. (XLSX 26 kb)

### Abbreviations

3'UTR: 3' untranslated region; 5'UTR: 5' untranslated region; ADHD: Attention deficit hyperactivity disorder; BMI: Body mass index; BPD: Bipolar disorder; CpG: 5' Cytosine-phosphate-Guanine 3'; CREB: Cyclic adenosine monophosphate responsive element binding protein; DBP: Diastolic blood pressure; DMR: Differentially methylated region; DNA: Deoxyribonucleic acid; GWAS: Genome-wide association studies; HARs: Human Accelerated Regions; HDL: High density lipoprotein; HWE: Hardy-Weinberg equilibrium; INRICH: Interval enRICHment analysis tool; IPA: Ingenuity Pathway Analysis; LD: Linkage disequilibrium; LDL: Low density lipoprotein; MAF: Minor allele frequency; MHC: Major histocompatibility complex; NSS: Neanderthal Selective Sweep; QQ: Quantile-quantile; RA: Rheumatoid arthritis; SBP: Systolic blood pressure; SCZ: Schizophrenia; SNP: Single nucleotide polymorphism; TC: Total cholesterol; TG: Triglycerides

### Acknowledgements

We thank Profs. Anders Dale and Wesley Thompson, University of California, San Diego for helpful discussions and Isabel Hanson Scientific Writing for formatting and submission. The code used to generate the QQ plots was graciously made publicly available by Matthew Flickinger PhD, University of Michigan at [https://genome.sph.umich.edu/wiki/Code\\_Sample\\_Generating\\_QQ\\_Plots\\_in\\_R](https://genome.sph.umich.edu/wiki/Code_Sample_Generating_QQ_Plots_in_R). We thank Ke Xu from Icahn School of Medicine, Mount Sinai for providing the URL for downloading the HAR datasets. We thank the author of the book *Sapiens: A Brief History of Humankind*, Yuval Noah Harari, PhD, for some of the ideas discussed in the paper. Finally, we thank the reviewers for their time and their suggestions for improving the manuscript.

### Funding

This work was supported by the Research Council of Norway (NFR; NORMENT-Centre of Excellence #2 T23273) and the KG Jebsen Foundation (SKGJ-MED-008). The funding bodies had no role in the collection, analysis or interpretation of the data, or in preparing the manuscript for publication.

### Availability of data and materials

The code supporting the results of this article is available in the Zenodo repository at <https://doi.org/10.5281/zenodo.198451>. GWAS datasets, DMR data and HAR data are publicly available as described in the [Methods](#) section.

### Authors' contributions

NB carried out the bioinformatics analyses, contributed to the design of the study and drafted the manuscript. TP, SG, FB contributed to statistical and LD analyses. OAA and VMS critically revised the manuscript. SLH conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

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

### Author details

<sup>1</sup>NORMENT - K.G. Jebsen Center for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway. <sup>2</sup>Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. <sup>3</sup>NORMENT - K.G. Jebsen Center for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>4</sup>NORMENT - K.G. Jebsen Centre, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. <sup>5</sup>Department of Clinical Medicine, Laboratory Building, Haukeland University Hospital, N-5021 Bergen, Norway.



Received: 16 July 2017 Accepted: 11 April 2018  
Published online: 11 May 2018

## References

- Shih RA, Belmonte PL, Zandi PP. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry*. 2004;16:260–83.
- Jeste DV, Del Carmen R, Lohr JB, Wyatt RJ. Did schizophrenia exist before the eighteenth century? *Compr Psychiatry*. 1985;26:493–503.
- Skre I, Onstad S, Torgersen S, Lygren S, Kringlen E. A twin study of DSM-III-R anxiety disorders. *Acta Psychiatr Scand*. 1993;88:85–92.
- Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC. othersHeritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry*. 1999;56:162–8.
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–92.
- Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet*. 2009;373:234–9.
- Larson CA, Nymn GE. Differential fertility in schizophrenia. *Acta Psychiatr Scand*. 1973;49:272–80.
- Bassett AS, Bury A, Hodgkinson KA, Honer WG. Reproductive fitness in familial schizophrenia. *Schizophr Res*. 1996;21:151–60.
- Nichols C. Is there an evolutionary advantage of schizophrenia? *Pers Individ Diff*. 2009;46:832–8.
- Brown S. Excess mortality of schizophrenia. A meta-analysis. *Br J Psychiatry*. 1997;171:502–8.
- Power RA, Kyaga S, Uher R, MacCabe JH, Långström N, Landen M, et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA psychiatry*. 2013;70:22–30.
- Crow TJ. A Darwinian approach to the origins of psychosis. *Br J Psychiatry*. 1995;167:12–25.
- Crow TJ. Is schizophrenia the price that *Homo sapiens* pays for language? *Schizophr Res*. 1997;28:127–41.
- Consortium SWG of the PG. othersBiological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–7.
- Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al. othersGenome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet*. 2013;45:1150–9.
- Consortium SPG-WAS (GWAS). othersGenome-wide association study identifies five new schizophrenia loci. *Nature Genet*. 2011;43:969–76.
- Crespi B, Summers K, Dorus S. Adaptive evolution of genes underlying schizophrenia. *Proc Roy Soc Lond [Biol]*. 2007;274:2801–10.
- Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol Biol Evol*. 2015;32:1148–60.
- Srinivasan S, Bettella F, Mattingdal M, Wang Y, Witoelar A, Schork AJ, et al. others Genetic markers of human evolution are enriched in schizophrenia. *Biol Psychiatry*. 2015;80:284–292.
- Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, et al. others Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*. 2006;2:e168.
- Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011;478:476–82.
- Bush EC, Lahn BT. A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol Biol*. 2008;8:17.
- Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, et al. Fast-evolving noncoding sequences in the human genome. *Genome Biol*. 2007;8:R118.
- Prabhakar S, Noonan JP, Pääbo S, Rubin EM. Accelerated evolution of conserved noncoding sequences in humans. *Science*. 2006;314:786.
- Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*. 2014;507:354–7.
- Vernot B, Akey JM. Resurrecting surviving Neanderthal lineages from modern human genomes. *Science*. 2014;343:1017–21.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. others The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 2014;505:43–49.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. others A draft sequence of the Neanderthal genome. *Science*. 2010;328:710–722.
- Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, et al. Reconstructing the DNA methylation maps of the Neanderthal and the Denisovan. *Science*. 2014;344:523–7.
- Giuliani C, Sazzini M, Bacalini MG, Pirazzini C, Marasco E, Fontanesi E, et al. Epigenetic variability across human populations: a focus on DNA methylation profiles of the KRTPCAP3, MADD1L1 and BRSK2 genes. *Genome Biol Evol*. 2016;8:2760–73.
- Gröniger E, Weber B, Heil O, Peters N, Stüb F, Wenck H, et al. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet*. 2010;6:e1000971.
- Schork AJ, Thompson WK, Pham P, Torkamani A, Rodney JC, Sullivan PF, et al. others All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet*. 2013;9:e1003449.
- Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics*. 2012;28:1797–9.
- Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. others Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genet*. 2011;43:977–983.
- Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-P, et al. others Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010;49:884–897.
- Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. others Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet*. 2010;42:508–514.
- Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707–13.
- International Consortium for Blood Pressure Genome-Wide Association Studies and others. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103–9.
- Locke AE, Kahali B, Berndt SJ, Justice AE, Pers TH, Day FR, et al. others Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518:197–206.
- Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. others Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet*. 2014;46:1173–1186.
- XXV World Congress of Psychiatric Genetics (WCPG): Poster Abstracts. *European Neuropsychopharmacology*; 2017. [http://www.europeanneuropsychopharmacology.com/pb/assets/raw/Health%20Advance/journals/neupsy/2017\\_4\\_Poster%20Abstracts.pdf](http://www.europeanneuropsychopharmacology.com/pb/assets/raw/Health%20Advance/journals/neupsy/2017_4_Poster%20Abstracts.pdf). in press.
- Gokhman D, Malul A, Carmel L. Inferring past environments from ancient epigenomes. *Mol Biol Evol*. 2017;34:2429–38.
- Gokhman D, Meshorer E, Carmel L. Epigenetics: it's getting old. Past meets future in paleoepigenetics. *Trends Ecol Evol*. 2016;31:290–300.
- Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, et al. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet*. 2013;9:e1003763.
- Mendizabal I, Keller T, Zeng J, Soojin VY. Epigenetics and evolution. *Integr Comp Biol*. 2014;54:31–42.
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, et al. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell*. 2011;146:1029–41.
- Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, et al. The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res*. 2015; 43:8204–14.
- Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. others DNA methylation contributes to natural human variation. *Genome Res*. 2013;23:1363–72.
- Richter D, Grün R, Joannes-Boyau R, Steele TE, Amani F, Rué M, et al. others The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the middle stone age. *Nature*. 2017;546:293–296.
- Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. others New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*. 2017;546:289–292.

51. Mellars P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci.* 2006;103:9381–6.
52. Kind C-J, Ebinger-Rist N, Wolf S, Beutelspacher T, Wehrberger K. The smile of the lion man. Recent excavations in Stadel cave (Baden-Württemberg, southwestern Germany) and the restoration of the famous upper Palaeolithic figurine. *Quartär.* 2014;61:129–45.
53. Aubert M, Brumm A, Ramli M, Sutikna T, Saptomo EW, Hakim B, et al. Pleistocene cave art from Sulawesi, Indonesia. *Nature.* 2014;514:223–7.
54. Atkinson QD. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science.* 2011;332:346–9.
55. Harari YN. *Sapiens: a brief history of humankind.* London: Harvill Secker; 2014.
56. Norenzayan A, Shariff AF. The origin and evolution of religious prosociality. *Science.* 2008;322:58–62.
57. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 2010;38:e87.
58. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. others Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature.* 2014;514:445–449.
59. Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan cave, China. *Proc Natl Acad Sci U S A.* 2013;110:2223–7.
60. Seguin-Orlando A, Korneliusen TS, Sikora M, Malaspina A-S, Manica A, Moltke I, et al. others Genomic structure in Europeans dating back at least 36,200 years. *Science.* 2014;346:1113–1118.
61. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature.* 2010;463:757–62.
62. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of native Americans. *Nature.* 2014;505:87–91.
63. Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, et al. others An early modern human from Romania with a recent Neanderthal ancestor. *Nature.* 2015;524:216–219.
64. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. others Genome-wide patterns of selection in 230 ancient Eurasians. *Nature.* 2015;528:499–503.
65. Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature.* 2014;513:409–13.
66. Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, et al. others Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* 2014;24:454–466.
67. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1092 human genomes. *Nature.* 2012;491:56–65.
68. Christoforou A, Dondrup M, Mattingsdal M, Mattheisen M, Giddaluru S, Nöthen MM, Rietschel M, Cichon S, Djurovic S, Andreassen OA, et al. Linkage-disequilibrium-based binning affects the interpretation of GWASs. *Am J Hum Genet.* 2012;90:727–33.
69. R Core Team: *R: A Language and Environment for Statistical Computing.* Vienna, Austria: 2017 URL <https://www.R-project.org/>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



2  
3 **Recently evolved human-specific methylated regions are enriched in**  
4 **schizophrenia signals**

5 Banerjee et al  
6

7 **Contents**  
8

9  
0 **Additional Method** **2**  
1

2  
3 **Additional Figures S1-S5** **4**

4 FIGURE S1 4

5 FIGURE S2 5

6 FIGURE S3 6

7 FIGURE S4 7

8 FIGURE S5 8  
9

0 **Additional Tables S1-S5** **9**

1 TABLE S1 9

2 TABLE S2 10

3 TABLE S3 10

4 TABLE S4 11

5 TABLE S5 11  
6

7 **Additional References** **12**  
8  
9  
0  
1  
2

## Additional Method

### Enrichment Analysis with INRICH

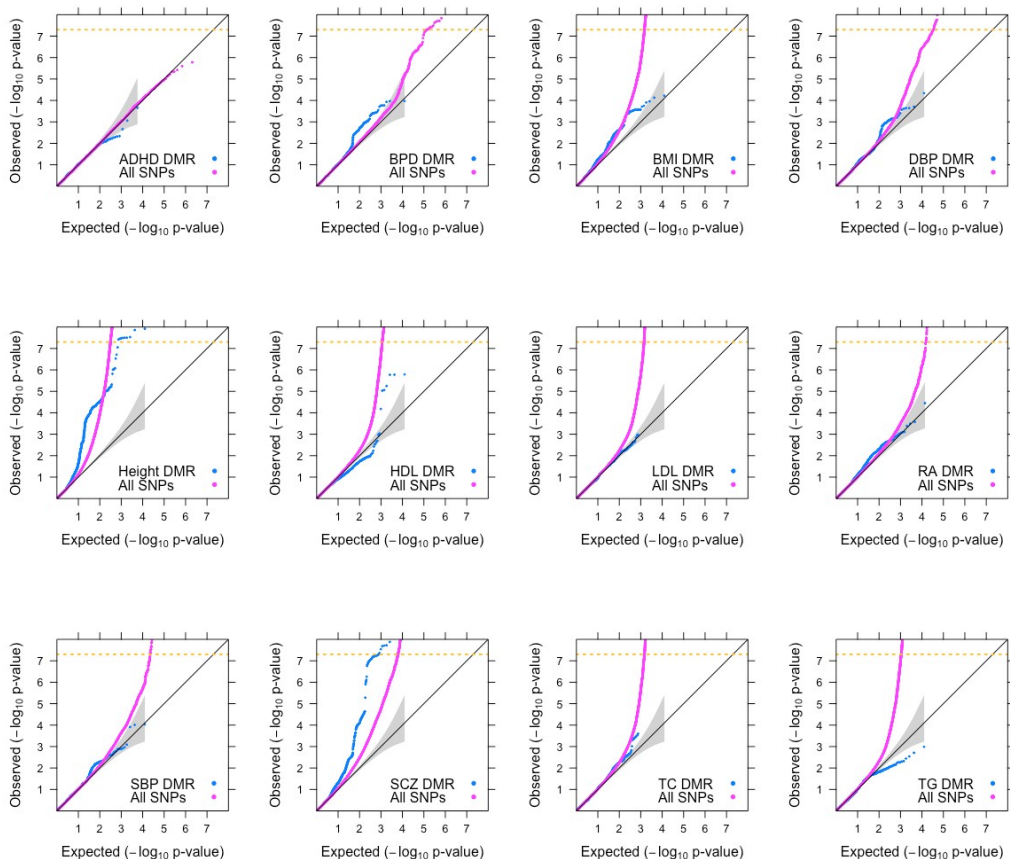
To assess enrichment for SNPs with different disease significance thresholds in various gene sets, we generated a range of LD-implicated regions through LD clumping in PLINK for index SNPs with  $p$ -values from  $1 \times 10^{-3}$  to  $1 \times 10^{-8}$ . LD Clumps were formed at  $r^2=0.5$  with the clump range limited to 250kb. INRICH was run on all the sets of LD intervals using default parameters described by Lee *et al* (2012).

All GENCODE V19 genes were used as the background set. The various gene sets tested included genes within a 100 kb flanking region of DMRs, HARs, PARs and NSS markers similar to the procedure described by Xu *et al*, (2015). For NSS markers, genes were assigned via LD blocks of  $r^2 \geq 0.8$  since NSS markers are single-base markers unlike the HARs and DMRs that are interval regions. GENCODE v19 gene database (last accessed 5<sup>th</sup> February 2016) was used to map the genes to DMRs, NSS markers and HARs. In their analysis, Xu *et al* had 893 genes within 100 kb of pHARs, 326 genes within 100 kb of mHARs (regions conserved in all mammals which are accelerated in humans) and 305 genes within 100 kb of PARs. In our study, using GENCODE v19, we had 3700, 1316 and 1268 genes within 100 kb of pHARs, mHARs and PARs respectively.

INRICH merges overlapping genes and overlapping LD-implicated intervals to prevent potentially inflated results due to multi-counting of the same genes/intervals. A total of 2510, 1015, 445, 207, 108 and 68 LD-implicated intervals were analyzed respectively for SNPs with  $p$ -values from  $1 \times 10^{-3}$  to  $1 \times 10^{-8}$  in the schizophrenia GWAS. Similarly, a total of 4321, 2498, 1596, 1130, 892 and 704 intervals were analyzed for the height GWAS. INRICH employs a two-stage procedure for assigning the statistical significance of enrichment of a given interval set with a given gene set. The first stage employs 10,000 permutations to empirically derive the null distribution of the overlap of intervals with gene sets. The second stage employs multiple testing correction via 5000 rounds of

0 bootstrapping. Enrichment of a given interval is then determined by the likelihood of a  
1 chance overlap with a gene set over the empirically observed distribution. The final output  
2 from INRICH lists gene sets with an empirical  $p$ -value at a default of  $P = 0.1$ . This is not  
3 the threshold at which INRICH performs the statistical tests. It is only the threshold to  
4 control which gene sets are displayed in the output. The corrected  $p$ -value obtained via  
5 bootstrapping is also displayed alongside. Since bootstrapping is a very robust procedure  
6 and causes many gene sets that would otherwise be significant at an empirical level to lose  
7 significance, an empirical  $p$ -value default of 0.1 adjusts what gene sets are visible at the end  
8 of the analyses. INRICH also outputs global enrichment of unique genes in gene sets at  
9 three thresholds that describes an excess of enriched genes at nominal gene-set  $P = 0.001$ ,  
0 0.01 and 0.05 (Lee *et al* (2012)).

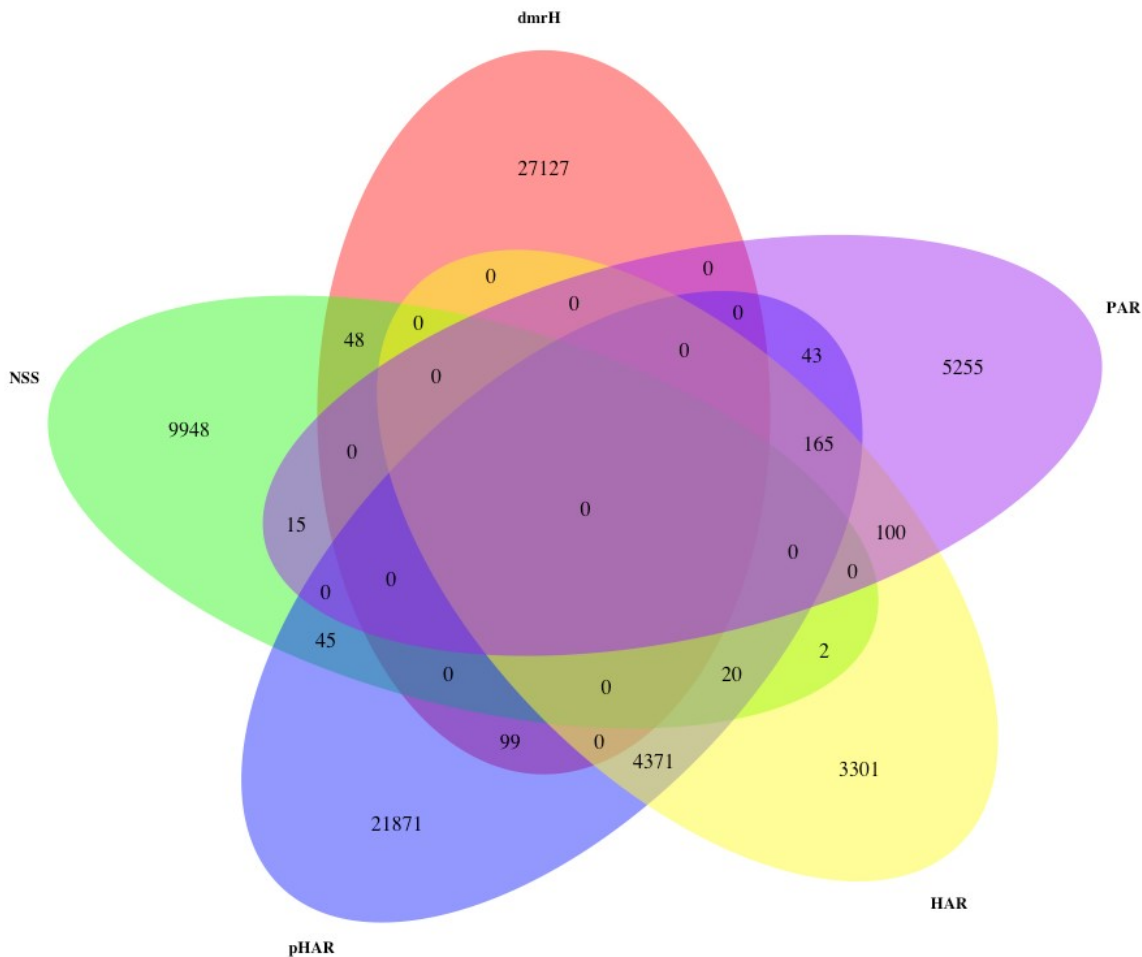
## Additional Figures



**Figure S1: Enrichment plots for common SNPs shared by all 12 GWAS**

Different numbers of SNPs were genotyped in the different GWAS, which could potentially bias our results. To test this, we generated a common set of  $\sim 2.4$  million SNPs that was determined by intersecting the SNP lists across all twelve GWAS including all SNPs from the ADHD GWAS ( $\sim 1.2$  million SNPs). ADHD, attention deficit hyperactivity disorder; BPD, bipolar disorder; BMI, body mass index; DBP, diastolic blood pressure; HDL, high density lipoprotein; LDL, low density lipoprotein; RA, rheumatoid arthritis; SBP, systolic blood pressure; SCZ, schizophrenia; TC, total cholesterol; TG, triglycerides. We find that the number of SNPs does not influence our results for SCZ, as maximum enrichment is seen for SCZ even when the common set of SNPs is used. The common set of SNPs did not include MHC SNPs, so the MHC was not represented in this analysis.

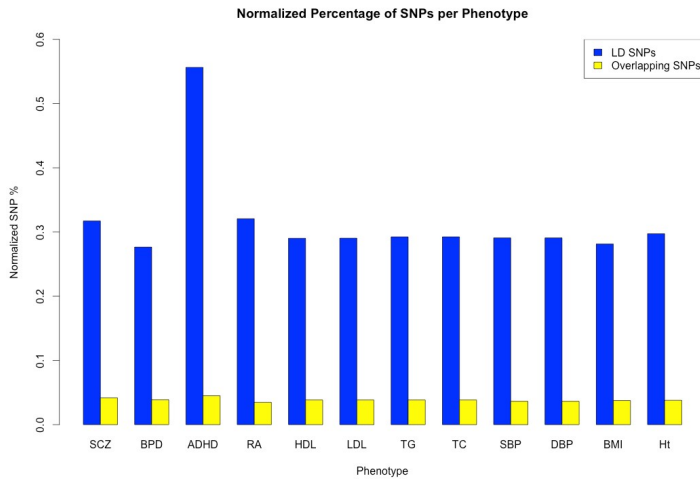
8  
9



0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
0

**Figure S2: Overlap of SNPs between various evolutionary annotations**

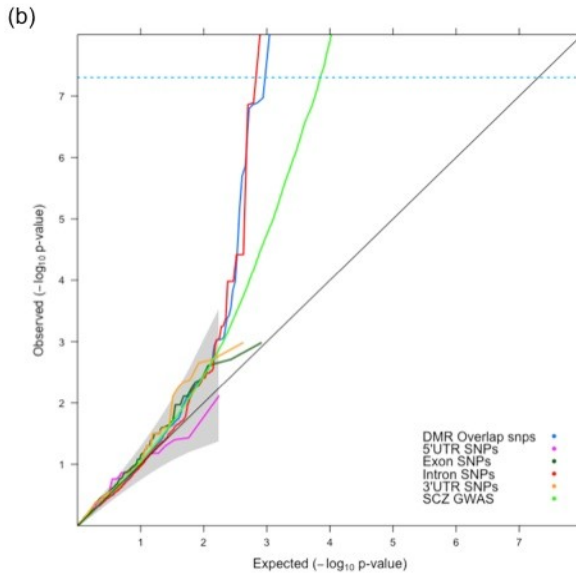
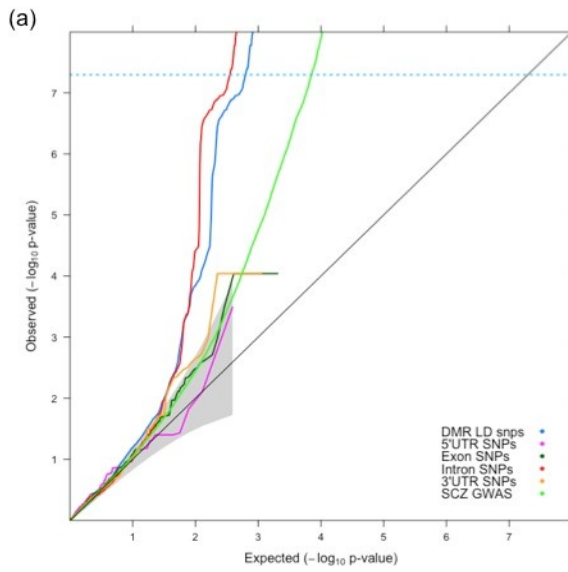
The figure shows that the overlap between the SNPs analysed for enrichment in various evolutionary annotations is very small. SNPs depicted here are those in LD with the respective regions at  $r^2 \geq 0.8$ . The biggest overlap is between the SNPs in the areas demarcated pHAR (regions conserved in primates that are accelerated in humans) and HAR (regions conserved in mammals that are accelerated in humans), where >50% of SNPs in LD with HARs are also in LD with pHARs. dmrH, human DMRs; NSS, NSS markers; PAR: regions conserved in mammals that are accelerated in primates.



**Figure S3: Proportion of DMR SNPs per GWAS**

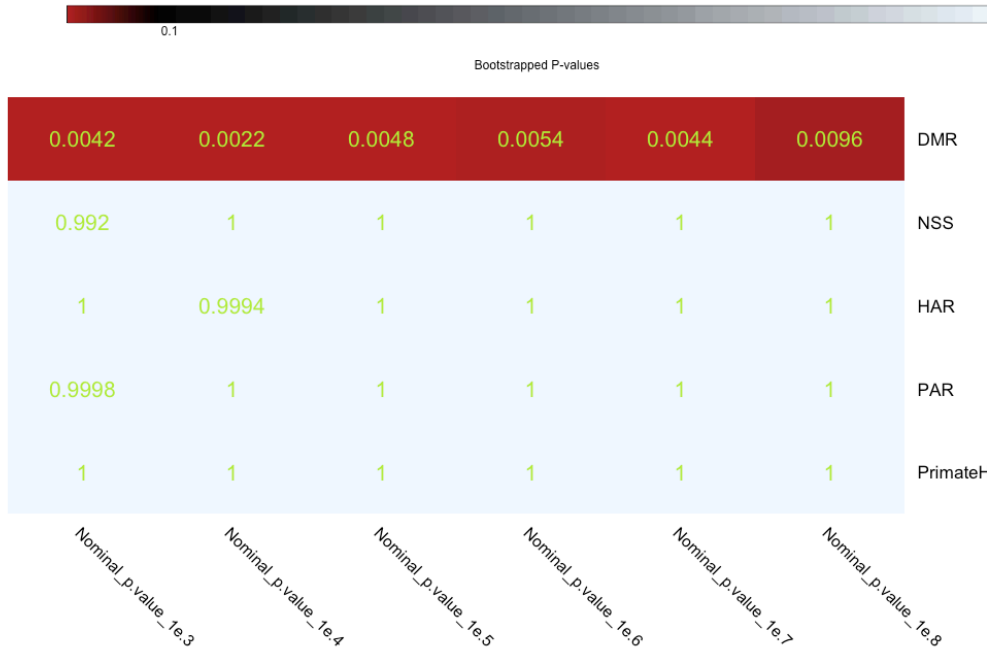
The figure depicts the normalization performed to determine if the varying number of markers genotyped in different GWAS influences the number of SNPs in DMRs. The total number of SNPs obtained for DMRs in each trait was divided by the total number of SNPs present in the respective GWAS and multiplied by 100 to obtain the normalized percentage of SNPs in DMRs. Blue bars depict the normalized percentage for SNPs in LD with DMRs while yellow bars depict the normalized percentage for SNPs that are physically located within DMRs. We observe that the total number of SNPs genotyped in a GWAS does not influence the proportion of SNPs that are physically within the DMR regions. The same largely holds true for SNPs in LD with DMR regions except for ADHD, possibly because the GWAS was underpowered.





**Figure S4: SCZ SNPs in DMRs stratified by genome annotation**

The figure depicts (a) SCZ SNPs in linkage disequilibrium (LD) with DMRs (blue) and (b) SCZ SNPs within DMRs (blue) stratified according to the following genomic annotations: 5'UTR (magenta), Exon (dark green), Intron (red), 3'UTR (orange). The light green line shows all SNPs from the SCZ GWAS.



6

7

### Figure S5: INRICH test for enrichment of association of DMRs, NSS and Accelerated Regions with height

8

9

Corrected  $p$ -values based on performing multiple testing with bootstrapping 5000 times and  $p=0.1$  as threshold. LD clumps of height markers from  $p$ -value  $1e^{-3} - 1e^{-8}$  were tested for enrichment in: DMR, human-specific DMRs; NSS, Neanderthal Selective Sweep; HAR, mammalian conserved regions that are accelerated in humans; PAR, mammalian conserved regions that are accelerated in primates; and PrimateHAR (pHAR), primate-conserved regions that are accelerated in humans.

2

3

4

5

6

7

8

9

0

1

2

3

4

5 **Additional Tables**

6

Phenotype	Total Study Size (n)	Number of SNPs (hg19)	Reference	SNPs within DMRs	SNPs in LD with DMRs
Schizophrenia	150,064	9,444,320	PGC2 (2014)	3930	29,954
Bipolar Disorder	63,766	2,426,991	Sklar <i>et al</i> (2011)	938	6711
Attention Deficit Hyperactivity Disorder	5,415	1,206,332	Neale <i>et al</i> (2010)	542	3970
Rheumatoid Arthritis	41,282	2,553,357	Stahl <i>et al</i> (2010)	884	8187
High Density Lipoprotein	99,900	2,620,435	Teslovich <i>et al</i> (2010)	1006	7603
Low Density Lipoprotein	95,454	2,620,568	Teslovich <i>et al</i> (2010)	1006	7608
Triglycerides	96,598	2,620,567	Teslovich <i>et al</i> (2010)	1006	7663
Total Cholesterol	100,184	2,620,450	Teslovich <i>et al</i> (2010)	1006	7663
Systolic Blood Pressure	200,000	2,461,102	ICBP GWAS (2011)	894	7157
Diastolic Blood Pressure	200,000	2,461,102	ICBP GWAS (2011)	894	7157
Body Mass Index	339,224	2,551,876	Locke <i>et al</i> (2015)	902	7364
Height	253,288	2,545,021	Wood <i>et al</i> (2011)	964	7567

7 **Table S1: Summary of GWAS and DMR SNPs**

8 For each GWAS, the table shows the sample size, marker density, the reference to the  
9 specific study, the number of SNPs located within DMRs, and the number of SNPs in LD  
0 with DMRs.  
1  
2  
3  
4

<b>Top Canonical Pathways</b>	<i>P</i> -value (Fisher's Exact test)
Wnt/Ca+ pathway	1.34E-05
P2Y Purigenic Receptor Signalling Pathway	1.04E-04
Thioredoxin Pathway	1.07E-04
CREB Signalling in Neurons	1.45E-04
Synaptic Long Term Potentiation	1.49E-04
<b>Top Physiological System Development and Function</b>	
Nervous System Development and Function	4.41E-02 - 2.23E-02
Tissue Morphology	2.23E-02 - 2.23E-02

**Table S2: Pathway analysis results for genes in LD with enriched SNPs in DMRs (Nervous System only).**

<b>Top Canonical Pathways</b>	<i>P</i> -value(Fisher's Exact test)
Wnt/Ca+ pathway	1.58E-05
P2Y Purigenic Receptor Signaling Pathway	1.14E-04
CREB Signaling in Neurons	1.49E-04
Synaptic Long Term Potentiation	2.05E-04
Thioredoxin Pathway	2.11E-04
<b>Top Physiological System Development and Function</b>	
Nervous System Development and Function	4.42E-02 - 1.41E-02
Cardiovascular System Development and Function	4.42E-02 - 2.24E-02
Connective Tissue Development and Function	4.42E-02 - 2.24E-02
Hair and Skin Development and Function	2.24E-02 - 2.24E-02
Hematological System Development and Function	2.24E-02 - 2.24E-02

**Table S3: Pathway analysis results for genes in LD with enriched SNPs in DMRs (All Organ Systems)**

<b>Top Canonical Pathways</b>	<i>P</i> -value(Fisher's Exact test)
CREB Signalling in Neurons	1.38E-04
IGF-1 Signalling	6.05E-04
T Cell Receptor Signalling	1.26E-03
Prolactin Signalling	1.27E-03
AMPK Signalling	1.31E-03
<b>Top Physiological System Development &amp; Function</b>	
Organismal Development	4.70E-03 - 4.70E-03
Tissue Development	4.70E-03 - 4.70E-03
Nervous System Development and Function	4.41E-02 - 1.35E-02
Tissue Morphology	4.41E-02 - 4.41E-02

**Table S4: Pathway analysis results for genes in LD with enriched NSS markers (Nervous System only)**

<b>Top Canonical Pathways</b>	<i>P</i> -value(Fisher's Exact test)
CREB Signalling in Neurons	1.53E-04
Protein Kinase A Signalling	3.02E-04
Synaptic Long Term Potentiation	8.90E-04
IGF-1 Signalling	9.67E-04
ERK/MAPK Signalling	1.57E-03
<b>Top Physiological System Development &amp; Function</b>	
Organismal Development	2.32E-02 - 4.23E-03
Tissue Development	2.32E-02 - 4.23E-03
Nervous System Development and Function	4.81E-02 - 1.21E-02
Connective Tissue Development and Function	2.32E-02 - 2.32E-02
Embryonic Development	2.32E-02 - 2.32E-02

**Table S5: Pathway analysis results for genes in LD with enriched NSS markers (All Organ Systems)**

5  
6  
7 **Additional References**  
8

9 ICBP GWAS: The International Consortium for Blood Pressure Genome-Wide Association Studies.  
0 Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*  
1 2011; **478**: 103–109.

2 Lee PH *et al.* INRICH: interval-based enrichment analysis for genome wide association studies.  
3 *Bioinformatics* 2012; **28**: 1797-1799.

4 Locke AE *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature*  
5 2015; **518**: 197–206.

6 Neale BM *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity  
7 disorder. *Journal of the American Academy of Child and Adolescent Psychiatry* 2010; **49**: 884–897.

8 PGC2: Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights  
9 from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.

0 Sklar P *et al.* Large-scale genome-wide association analysis of bipolar disorder identifies a new  
1 susceptibility locus near ODZ4. *Nature Genetics* 2011; **43**: 977.

2 Stahl EA *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid  
3 arthritis risk loci. *Nature genetics* 2010; **42**: 508–514.

4 Teslovich TM *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*  
5 2010; **466**: 707–713.

6 Wood AR *et al.* Defining the role of common variation in the genomic and biological architecture of  
7 adult human height. *Nature Genetics* 2014; **46**: 1173–1186.

8 Xu K *et al.* Genomic and network patterns of schizophrenia genetic variation in human evolutionary  
9 accelerated regions. *Molecular Biology and Evolution* 2015; **32**: 1148-1160.  
0









1 **Title:**

2

3 **Analysis of differentially methylated regions in primates and non-primates**  
4 **provides support for the evolutionary hypothesis of schizophrenia**

5

6 Niladri Banerjee<sup>a,b</sup>, Tatiana Polushina<sup>a,b</sup>, Francesco Bettella<sup>c,d</sup>,

7 Vidar M. Steen<sup>a,b</sup>, Ole A. Andreassen<sup>c,d</sup>, Stephanie Le Hellard<sup>a,b</sup>

8

9 a. NORMENT - K.G. Jebsen Center for Psychosis Research, Department of Clinical  
10 Science, University of Bergen, Bergen, Norway

11 b. Dr. Einar Martens Research Group for Biological Psychiatry, Department of  
12 Medical Genetics, Haukeland University Hospital, Bergen, Norway

13 c. NORMENT - K.G. Jebsen Center for Psychosis Research, Institute of Clinical  
14 Medicine, University of Oslo, Oslo, Norway

15 d. NORMENT - K.G. Jebsen Centre, Division of Mental Health and Addiction, Oslo  
16 University Hospital, Oslo, Norway

17

18 **Correspondence to:** Prof. Stéphanie Le Hellard, Department of Clinical Medicine,  
19 Laboratory Building, Haukeland University Hospital, N-5021 Bergen, Norway.

20 **Telephone:** +47-55 97 53 37

21 **Fax:** +47-55 97 54 79

22 **Email:** stephanie.hellard@uib.no

23

24 **Authors' email addresses**

25 Niladri Banerjee                      niladri.banerjee@uib.no

26	Tatiana Polushina	tatiana.polushina@uib.no
27	Francesco Bettella	francesco.bettella@medisin.uio.no
28	Vidar M. Steen	vidar.martin.steen@helse-bergen.no
29	Ole A. Andreassen	o.a.andreassen@medisin.uio.no
30	Stéphanie Le Hellard	stephanie.hellard@uib.no
31		
32		

33 **Abstract**

34

35 Introduction: The persistence of schizophrenia in human populations separated by  
36 geography and time led to the evolutionary hypothesis that proposes schizophrenia as  
37 a by-product of the higher cognitive abilities of modern humans. To explore this  
38 hypothesis, we used here an evolutionary epigenetics approach building on  
39 differentially methylated regions (DMRs) of the genome.

40 Methods: We implemented a polygenic enrichment testing pipeline using the  
41 summary statistics of genome-wide association studies (GWAS) of schizophrenia and  
42 12 other phenotypes. We investigated the enrichment of association of these traits  
43 across genomic regions with variable methylation between modern humans and great  
44 apes (orangutans, chimpanzees and gorillas; primate DMRs) and between modern  
45 humans and recently extinct hominids (Neanderthals and Denisovans; non-primate  
46 DMRs).

47 Results: Regions that are hypo-methylated in humans compared to great apes show  
48 enrichment of association with schizophrenia only if the major histocompatibility  
49 complex (MHC) region is included. With the MHC region removed from the analysis,  
50 only a modest enrichment for SNPs of low effect persists. The INRICH pipeline  
51 confirms this finding after rigorous permutation and bootstrapping procedures.

52

53 Conclusion: The analyses of regions with differential methylation changes in humans  
54 and great apes do not provide compelling evidence of enrichment of association with  
55 schizophrenia, in contrast to our previous findings on more recent methylation  
56 differences between modern humans, Neanderthals and Denisovans. Our results

57 further support the evolutionary hypothesis of schizophrenia and indicate that the  
58 origin of some of the genetic susceptibility factors of schizophrenia may lie in recent  
59 human evolution.

60

61

62 **Key Words:** schizophrenia; evolutionary hypothesis; epigenetics; differentially  
63 methylated regions; primates; Neanderthals.

64

65

66

67

68

69

70

71

72 **1. Introduction**

73

74 Schizophrenia is a psychiatric disorder with a prevalence rate of 2.7-8.3/1,000 persons  
75 (Messias et al., 2007) and heritability estimated between 60-90% (Cardno et al., 1999;  
76 Lichtenstein et al., 2009; Skre et al., 1993; Sullivan et al., 2003). It occurs at quite  
77 similar rates across populations worldwide (Ayuso-Mateos, 2002; Brüne, 2004;  
78 WHO, 1973) and written records describing its symptoms exist dating back 5,000  
79 years (Jeste et al., 1985). This consistent persistence of the disease despite reduced  
80 fecundity (Brüne, 2004; Nichols, 2009) and increased mortality is a paradox (Bassett  
81 et al., 1996; Brown, 1997; Larson and Nyman, 1973), since the reduced fecundity of  
82 patients afflicted with schizophrenia does not appear to eliminate the disease from the  
83 population (Power et al., 2013) Part of the reason may be due to afflicted individuals  
84 reproducing prior to the onset of the disease (Markow, 2012). Another contributing  
85 factor could be that schizophrenia risk variants may have provided an advantage to  
86 the kin of the affected by conferring superior creative and intellectual abilities upon  
87 them (Kyaga et al., 2011; Nichols, 2009). To explain the constant occurrence of the  
88 disease, TJ Crow (Crow, 1997, 1995) proposed the so-called evolutionary hypothesis  
89 of schizophrenia, which suggests that the disease is a consequence of human  
90 evolution: the higher cognitive abilities of modern-day humans, including language,  
91 may predispose to psychiatric illnesses such as schizophrenia (Crow, 2008, 2000,  
92 1997).

93

94 In the post-genomic era (Lander et al., 2001; Venter et al., 2001), emerging lines of  
95 evidence are lending support to this hypothesis. Crespi et al. (Crespi et al., 2007) were  
96 amongst the first to show that genes with evidence of recent positive selection in

97 humans are also implicated more frequently in schizophrenia. More evidence has  
98 been provided by studies based on comparative genomics (Pollard et al., 2006;  
99 Srinivasan et al., 2015; Xu et al., 2015), a field in which genomes of progressively  
100 older species are compared to identify substitutions and mutations that help estimate  
101 divergence between the species. For instance, a group of regions defined by negative  
102 Neanderthal selective sweep (NSS) scores describe the selective evolution of genomic  
103 regions in modern-day humans over Neanderthals (Burbano et al., 2010; Green et al.,  
104 2010). These regions were shown by Srinivasan et al. (2015) to be enriched for  
105 schizophrenia risk markers, in line with the evolutionary hypothesis of schizophrenia.  
106 Other regions known as human accelerated regions (HARs) (Gittelman et al., 2015;  
107 Pollard et al., 2006; Xu et al., 2015), first described by Pollard et al. (2006), show  
108 accelerated evolution in humans compared to primates or mammals. HARs have also  
109 provided some evidence of enrichment of association with schizophrenia (Xu et al.,  
110 2015), but these findings may have been driven by a few genes since they were not  
111 replicated using a polygenic approach (Srinivasan et al., 2017, 2015) .

112

113 While several studies have looked at the evolution of the genome (Bird et al., 2007;  
114 Bush and Lahn, 2008; Gittelman et al., 2015; Paaby and Rockman, 2014; Pollard et  
115 al., 2006), there are reports that the epigenome is evolving as well (Gokhman et al.,  
116 2014; Hernando-Herraez et al., 2015, 2013; Mendizabal et al., 2014; Molaro et al.,  
117 2011). This provides new insights into events leading to the speciation and divergence  
118 of modern humans. The epigenome refers to the layer of chemical modifications, such  
119 as methylation and histone modifications, to the genome that regulate gene expression  
120 (Bernstein et al., 2007; Kundaje et al., 2015; Rivera and Ren, 2013). For instance,  
121 Gokhman et al. (2014) compared the methylomes of humans with Neanderthals and

122 Denisovans. They reported that while 97% of the methylome was comparable  
123 between humans, Neanderthals and Denisovans, some regions showed differential  
124 methylation between the three hominids. Previously (Banerjee et al., 2017), we  
125 analysed the differentially methylated regions (DMRs) identified for Neanderthals,  
126 Denisovans and modern humans by Gokhman et al. (2014), and found evidence that  
127 the regions of the genome with human-specific DMRs harbour relatively more genetic  
128 variants associated with schizophrenia than the rest of the genome, i.e. the DMRs  
129 were enriched for SCZ markers both at the single-nucleotide polymorphism (SNP)  
130 level and at the gene level. These human-specific DMRs thus provide evidence of  
131 enrichment of methylation changes in regions harbouring genetic variants associated  
132 with schizophrenia, at least since the divergence from Neanderthals and Denisovans  
133 (Banerjee et al., 2017).

134

135 Here, we sought to determine if evolutionarily older methylation differences can  
136 provide a further timeframe for the origin of schizophrenia risk markers in the human  
137 lineage. We asked whether we can find epigenetic evidence that the origin of  
138 schizophrenia risk markers predates the origins of the Homo genus, i.e. before the  
139 divergence of chimpanzees and humans around 6-8 million years ago (MYA) (Glazko  
140 and Nei, 2003; Langergraber et al., 2012). We tested this hypothesis by analysing  
141 primate DMRs that trace an evolutionary history of at least 13 million years (Glazko  
142 and Nei, 2003; Hasegawa et al., 1985; Rannala and Yang, 2003). We used the same  
143 statistical analyses as described by Lee et al. (2012), Schork et al. (2013), and  
144 Srinivasan et al. (2015) to test for polygenic enrichment of a set of markers from  
145 genome-wide association studies (GWAS). We interrogated regions of the human  
146 genome which are hypo- or hyper-methylated in comparison to the corresponding



147 ones in chimpanzees, gorillas and orangutans for enrichment of genetic variants

148 associated with schizophrenia or other human traits.

149

150 **2. Materials and methods**

151

152 *2.1. GWAS data*

153 Summary statistics for thirteen different phenotypes were obtained from their  
154 respective published GWAS studies: schizophrenia (SCZ) (Ripke et al., 2014), bipolar  
155 disorder (BPD) (Sklar et al., 2011), attention deficit hyperactivity disorder (ADHD)  
156 (Demontis et al., 2017), rheumatoid arthritis (RA) (Stahl et al., 2010), blood lipid  
157 markers (high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides  
158 (TG), total cholesterol (TC)) (Teslovich et al., 2010), blood pressure (systolic blood  
159 pressure (SBP), diastolic blood pressure (DBP)) (Ehret et al., 2011), body mass index  
160 (BMI) (Locke et al., 2015), height (Wood et al., 2014) and intelligence (Sniekers et  
161 al., 2017). For studies published with hg18 coordinates (BPD, SBP, DBP, HDL, LDL,  
162 TG, TC, RA), conversion to hg19 was performed using the command line version of  
163 the liftOver tool from the UCSC Genome Browser (Karolchik et al., 2014)  
164 ([http://hgdownload.cse.ucsc.edu/downloads.html#utilities\\_downloads](http://hgdownload.cse.ucsc.edu/downloads.html#utilities_downloads)). For BMI and  
165 height SNPs, the genomic coordinates were obtained by mapping them to the  
166 assembly of 1,000 Genomes Project (1KGP) Phase 1 reference panel SNPs (Durbin et  
167 al., 2012).

168

169 *2.2. Human hypo- and hyper-methylated regions from primate DMRs*

170 These methylated regions were retrieved from the study by Hernando-Herraez et al.  
171 (2015), who identified them by comparing the methylation profile of DNA from  
172 peripheral blood samples of orangutans, chimpanzees and gorillas to that of humans.  
173 Since the DMRs are determined by comparing humans with other primates, we refer  
174 to this set collectively as primate DMRs. Both hypo- and hyper-methylated DMRs

175 from humans were analysed. As these DMRs are identified in the same tissues in all  
176 samples, they are considered to represent species-specific methylation differences, not  
177 tissue-specific methylation differences (Gokhman et al., 2014). Altogether, the human  
178 hypo- and hyper-methylated DMRs can be used to represent an evolutionary course of  
179 history spanning from at least 13 MYA (Glazko and Nei, 2003; Langergraber et al.,  
180 2012), when orangutans diverged from the common ancestors, to 6 MYA, when the  
181 chimpanzees and humans diverged from each other (Glazko and Nei, 2003;  
182 Langergraber et al., 2012). Since our interest was in human-specific enrichment, we  
183 focused the analyses on human hypo- and hyper-methylated DMRs.

184

### 185 *2.3. Differentially methylated regions (DMRs) from Neanderthals, Denisovans and* 186 *modern humans*

187 As previously described (Gokhman et al., 2014), these methylated regions have been  
188 identified by comparing the methylomes of osteoblasts from modern-day humans  
189 with those from Neanderthals and Denisovans. We refer to them in this paper as non-  
190 primate DMRs. Gokhman et al. (2014) devised a strategy utilizing information in the  
191 form of cytosine (C) to thymine (T) ratios to decipher the ancient methylomes of  
192 Neanderthals and Denisovans. Subsequently, they compared the methylomes of  
193 Neanderthals, Denisovans and modern humans and inferred the species in which the  
194 methylation variation likely took place; this information was used to classify the  
195 DMRs as Neanderthal-specific, Denisovan-specific and human-specific. These DMRs  
196 represent species-specific methylation (Gokhman et al., 2014).

197

### 198 *2.4. Neanderthal selective sweep (NSS) data*

199 We obtained NSS marker data from Srinivasan et al. (2015). Negative scores for NSS  
200 markers indicate positive selection in humans. Markers with such scores were used in  
201 the downstream analyses.

202

### 203 *2.5. SNP assignment with LDsnpr*

204 The previously published R-based software package LDsnpr (Christoforou et al.,  
205 2012) was utilized for assigning SNPs to the respective DMRs using LD (linkage  
206 disequilibrium)-based binning at  $r^2 \geq 0.8$  in R (R Core Team, 2017). LD-based binning  
207 makes it possible to determine whether SNPs from a specific GWAS are in LD with  
208 the DMR of interest. Using LD allows the capture of a greater number of relevant  
209 SNPs in comparison to an approach where only physically overlapping SNPs are  
210 considered. The LD file utilized was in HDF5 format and was constructed from the  
211 European reference population of 1KGP and can be publicly downloaded at:  
212 <http://services.cbu.uib.no/software/ldsnpr/Download>.

213

### 214 *2.6. Enrichment analyses based on stratified quantile-quantile (QQ) plots*

215 QQ plots are an essential method used in GWASs to depict the presence of true  
216 signals. They help to visually observe the spread of data and deviations from the null  
217 distribution. Under the null hypothesis, no difference is expected between the  
218 observed and expected distributions of data. As such, a line of no difference or null  
219 line is obtained that is equidistant from both X and Y axes. However, if the null  
220 hypothesis were to be false, there would be a deviation of the observed data  
221 distribution from the expected data distribution. As described in depth by Schork et al.  
222 (2013), a leftward deflection of the observed distribution from the null line represents  
223 enrichment – the greater the leftward deflection, the stronger the enrichment of true

224 signals. This method has been used recently not only to show how specific genomic  
225 annotation affects the distribution of disease SNPs with true signals (Schork et al.,  
226 2013), but also to demonstrate that regions of recent evolution are enriched for  
227 schizophrenia markers (Banerjee et al., 2017; Srinivasan et al., 2015). We took the  
228 SNPs that are in LD with the DMR regions and plotted their  $p$ -value distributions  
229 from various GWASs. The observed  $p$ -value distributions were then determined to be  
230 enriched or not using conditional Q-Q plots as described by Schork et al. (2013).  
231 Genomic inflation was corrected by  $\lambda_{GC}$ .

232

### 233 *2.7. INRICH-based enrichment analysis*

234 The stratified QQ plots provide a visual depiction of data distributions and enrichment  
235 of true signals within a stratum of data, but they do not quantify this enrichment.  
236 Therefore, we used the INterval EnRICHment (INRICH) analysis tool to statistically  
237 quantify the enrichment observed. This pipeline performs permutation and  
238 bootstrapping procedures to determine with statistical confidence whether LD-  
239 implicated genomic intervals are enriched in specific gene sets (Lee et al., 2012). The  
240 INRICH analysis takes into account several potential biases that can otherwise lead to  
241 false positives, such as variable gene size, SNP density within genes, LD between and  
242 within genes, and overlapping genes in the gene sets. We used the same procedure  
243 reported previously (Banerjee et al., 2017; Xu et al., 2015) with SNPs in the extended  
244 MHC region and SNPs with MAF <0.05 excluded from the analysis. Additional  
245 details can be found in the Supplementary Information.

246

## 247 **3. Results**

248

249 *3.1. Co-localisation of human hypo-methylated regions and genetic variants*

250 *associated with schizophrenia in the MHC.*

251 We ascertained whether there is any enrichment of human hypo- and hyper-  
252 methylated regions in schizophrenia-associated SNPs. Using previously published  
253 methodology (Christoforou et al., 2012), we mapped schizophrenia markers to human  
254 hypo-methylated regions (hypo-DMRs) and hyper-methylated regions (hyper-DMRs).  
255 Out of a total of ~9.4 million SCZ markers obtained from the GWAS, 10,165 markers  
256 tagged hypo-DMRs and 4,503 tagged hyper-DMRs.

257

258 Figure 1A shows the conditional QQ plots for schizophrenia markers (all markers, the  
259 hypo-DMR set and the hyper-DMR set) including those in the MHC region. For  
260 hypo-DMR markers (Supplementary Dataset 1), we observed a significant enrichment  
261 as depicted by the leftward deviation. No enrichment was observed for hyper-DMR  
262 markers. Since the MHC region is a region of extended linkage disequilibrium, which  
263 can bias the enrichment estimates, and since it is the main region of association with  
264 schizophrenia, we also tested the enrichment with the MHC region removed (Figure  
265 1B). Under these conditions there is a trend for enrichment of hypo-DMR markers at  
266 higher  $p$ -value thresholds, but this enrichment is substantially less than when the  
267 MHC is included (Figure 1A).

268

269 *3.2. Enrichment of markers is not seen for other human traits*

270 Next, we tested if the human hypo- and hyper-methylated regions are enriched for  
271 other human traits and phenotypes. We tested a total of thirteen different phenotypes,  
272 full details of which can be found in section 2.1. Each GWAS had been performed  
273 with a different number of genotyped SNPs, and this difference could potentially bias

274 our results. To circumvent this, we created a list of ~2.4 million common SNPs that  
275 were genotyped across all the phenotypes investigated in the present study. Only  
276 SNPs on this list were used for enrichment analysis.

277

278 As can be seen in Fig. 2, no enrichment was observed in any of the traits, with the  
279 possible exception of height at higher  $p$ -value threshold markers. The common list of  
280 markers did not contain the MHC region and as such no enrichment is observed for  
281 schizophrenia either.

282

283 *3.3. Evidence of enrichment for hypo-methylated regions with SNPs at high  $p$ -values*

284 The enrichment plots allowed us to visually ascertain enrichment in the datasets.

285 However, they did not give any indication of the statistical robustness of the

286 enrichment. To ascertain if the human hypo- and hyper-methylated regions are

287 statistically enriched for schizophrenia and height markers, we implemented the

288 INRICH pipeline, which performs 10,000 permutations and 5,000 bootstrapping

289 calculations, to determine with statistical confidence the enrichment observed (Lee et  
290 al., 2012).

291

292 The INRICH analysis confirmed a significant ( $p < 0.05$ ) enrichment of association for

293 human hypo-DMRs, but not hyper-DMRs, with schizophrenia at SNPs of higher  $p$ -  
294 value thresholds ( $p < 10e-3$  to  $p < 10e-4$ ) (Fig. 3). This enrichment was at the gene level,

295 and complemented the enrichment observed at the SNP level for higher  $p$ -value

296 thresholds (Fig. 1B). Importantly, this enrichment persisted upon testing a pruned

297 schizophrenia dataset (Supplementary Fig. 1). The enrichment was however not

298 significant at the genome-wide threshold ( $p < 5 \times 10e-8$ ) and was much weaker than that

299 observed for non-primate DMRs (Fig. 3). We also observed a similar trend for height  
300 where there was enrichment at SNPs of higher but not lower  $p$ -value thresholds. This  
301 enrichment was similarly less pronounced than for non-primate DMRs  
302 (Supplementary Fig. 2).

303

304

#### 305 **4. Discussion**

306

307 In our study, we investigated if regions of the human genome whose methylation has  
308 evolved since the divergence of modern humans from great apes are enriched for  
309 markers of schizophrenia. We found evidence that there is enrichment for hypo-  
310 methylated DMRs driven by the MHC locus, a known risk region that harbours the  
311 most significant schizophrenia GWAS markers (Ripke et al., 2014). When the MHC  
312 region was excluded from the analysis, there remained a trend towards enrichment of  
313 hypo-DMRs driven by SNPs of higher  $p$ -value thresholds. This finding was  
314 complemented by the INRICH analyses that indicated significant enrichment among  
315 SNPs of higher  $p$ -value thresholds. When analysing a global SNP list common to  
316 GWAS of several traits, we failed to find evidence of enrichment of any trait with the  
317 possible exception of height at higher SNP  $p$ -value thresholds. We tested this further  
318 with the INRICH pipeline, which revealed gene-level enrichment of LD intervals for  
319 height markers below the genome-wide threshold ( $p < 5 \times 10^{-8}$ ). Compared to our  
320 previous study, in which we demonstrated enrichment of association with  
321 schizophrenia for non-primate DMRs that were derived by comparing human,  
322 Neanderthal and Denisovan methylomes (Banerjee et al., 2017), the primate DMRs  
323 tested here show far less enrichment. The primate and non-primate DMRs have very



324 little overlap, which suggests that the methylation changes that took place since the  
325 divergence of modern humans from Neanderthals and Denisovans occurred in  
326 different regions of the genome compared to those that took place since divergence  
327 from great apes.

328

329 The central role of the MHC region in the enrichment of human hypo-methylated  
330 regions poses interesting questions. The MHC region is known for its complex LD  
331 architecture, which renders the interpretation of genetic signals very challenging.  
332 Other groups have previously reported that the MHC region is one of the fastest  
333 evolving regions of the human genome (Meyer et al., 2017) and have implicated it in  
334 mate preference (Bernatchez and Landry, 2003; Kromer et al., 2016; Potts and  
335 Wakeland, 1990; Roberts et al., 2008; Wintemitz et al., 2017), odour perception  
336 (Roberts et al., 2008; Santos et al., 2005) and immune response (Benacerraf, 1981;  
337 Horton et al., 2004). Recently it was shown that a large proportion of the association  
338 of the region with schizophrenia can be explained by complement C4 haplotypes that  
339 include C4 copy number variation (Sekar et al., 2016). Nevertheless, there remains a  
340 part of the association in this region that is unexplained (Gejman et al., 2011) and will  
341 need further investigation. It is interesting to consider the possibility that the MHC  
342 region and the immune system in general play a central role in evolution at the  
343 epigenomic as well as at the genomic level (Meyer et al., 2017; Potts and Wakeland,  
344 1990; Sommer, 2005; Traherne, 2008). The mechanisms by which hypo-methylation  
345 could influence the aforementioned processes are open to speculation since the MHC  
346 region has more than 200 genes in close physical proximity and LD with one another  
347 (Beck et al., 1999). This makes it hard to interpret the exact biological consequences  
348 of our findings.

349

350 Interestingly, the gene-level analysis via INRICH seems to suggest enrichment of  
351 SNPs of higher  $p$ -value thresholds in primate DMRs for both schizophrenia and  
352 height. This enrichment is far lower than what we found for non-primate DMRs for  
353 both schizophrenia and height (Banerjee et al., 2017) and which persisted for  
354 schizophrenia even with pruned datasets.

355

356 The very small overlap between primate and non-primate DMRs might suggest that  
357 the divergence from Neanderthals and Denisovans brought about more significant  
358 methylation changes in regions implicated in the aetiology of schizophrenia and  
359 height than the divergence from great apes. In other words, our results might suggest  
360 that the evolutionary factors that regulate methylation variation acted on different  
361 segments of the genome at different time points. So while the methylation variation  
362 since the divergence from Neanderthals and Denisovans may mark a genome-wide  
363 increase of schizophrenia susceptibility (Banerjee et al., 2017), the methylation  
364 variation from the time period between 13 and 6 MYA appears not to have  
365 significantly increased the risk for schizophrenia (except possibly for some markers in  
366 the MHC region),

367

368 Our results are also in line with the findings of Srinivasan et al. (2017), who failed to  
369 find evidence of enrichment of schizophrenia using genomic markers of evolution  
370 dating back to 200 MYA. The same authors also reported enrichment of association  
371 for regions of more recent evolution in modern humans (Srinivasan et al., 2015).  
372 Interestingly, one of the evolutionary proxies used by Srinivasan and colleagues  
373 (2017), namely HARs, also showed enrichment for height, similar to our recent study

374 (Banerjee et al., 2017). This suggests that regions controlling both genomic and  
375 epigenomic variation in height may also be driven by recent evolution. Finally, our  
376 results agree well with the observation by Srinivasan et al. (2017) of some  
377 involvement of the MHC in an early evolutionary context.

378

379 Although our results are in line with several findings in the field, the current methods  
380 have some limitations. Highly polygenic traits such as schizophrenia have a large  
381 number of genetic loci contributing to the aetiology of a disease (Bulik-Sullivan et al.,  
382 2015; Schork et al., 2016). The ability to detect these large numbers of genetic loci is  
383 dependent on the sample size and adequate statistical power (Schork et al., 2016).  
384 Consequently, the polygenic enrichment methods may be limited by the statistical  
385 power of the respective GWAS and trait polygenicity. Furthermore, in the INRICH  
386 analysis that uses LD-clumping of SNPs at  $p < 10e-3$  to  $p < 10e-8$ , higher  $p$ -value  
387 thresholds (e.g.  $p < 10e-3$ ) still include SNPs of lower  $p$ -values, even though they  
388 become progressively smaller minorities. Thus, although higher  $p$ -values increase the  
389 number of LD-clumps tested, we do not expect this to increase the Type I error rate  
390 (Lee et al., 2012).

391

392 In conclusion, our results suggest that methylation markers tracing an evolutionary  
393 period dating back to 13 MYA (primate DMRs) are not enriched for schizophrenia  
394 markers, unlike methylation markers from a recent timeframe (non-primate DMRs)  
395 (Banerjee et al., 2017). Taken in consideration with previous studies of genomic  
396 markers of evolution dating back 200 MYA (Srinivasan et al., 2017), our results  
397 support the hypothesis that the origins of schizophrenia lie in more recent

398 evolutionary events, possibly after the divergence of modern-day humans from

399 Neanderthals and Denisovans.

400

401

402 **Appendix A. Supplementary data**

403 Supplementary Information: Additional Methods, Figures and Tables

404 Supplementary Dataset 1: Annotation of human hypo-methylated regions with

405 markers of schizophrenia

406

407 **References**

408

409 Ayuso-Mateos, J.L., 2002. Global Burden of schizophrenia in the year 2000-Version  
410 1 Estimates. World Health Organization.

411

412 Banerjee, N., Polushina, T., Bettella, F., Giddaluru, S., Steen, V.M., Andreassen,  
413 O.A., Le Hellard, S., 2017. Human-specific methylated regions are enriched in  
414 schizophrenia. bioRxiv 113175.

415

416 Bassett, A.S., Bury, A., Hodgkinson, K.A., Honer, W.G., 1996. Reproductive fitness  
417 in familial schizophrenia. *Schizophr. Res.* 21, 151–60.

418

419 Beck, S., Geraghty, D., Inoko, H., Rowen, L., 1999. Complete sequence and gene  
420 map of a human major histocompatibility complex. *Nature* 401, 921–923.

421

422 Benacerraf, B., 1981. Role of MHC gene products in immune regulation. *Science*  
423 212, 1229–1238.

424

425 Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: what have  
426 we learned about natural selection in 15 years? *J. Evol. Biol.* 16, 363–77.

427

428 Bernstein, B.E., Meissner, A., Lander, E.S., 2007. The mammalian epigenome. *Cell*  
429 128, 669 – 681.

430

431 Bird, C.P., Stranger, B.E., Liu, M., Thomas, D.J., Ingle, C.E., Beazley, C., Miller, W.,  
432 Hurles, M.E., Dermitzakis, E.T., 2007. Fast-evolving noncoding sequences in the  
433 human genome. *Genome Biol.* 8, R118.

434

435 Brown, S., 1997. Excess mortality of schizophrenia. A meta-analysis. *Br. J.*  
436 *Psychiatry* 171, 502–508.

437

438 Brüne, M., 2004. Schizophrenia—an evolutionary enigma? *Neurosci. Biobehav. Rev.*  
439 28, 41–53.

440

441 Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N.,  
442 Daly, M.J., Price, A.L., Neale, B.M., Schizophrenia Working Group of the  
443 Psychiatric Genomics Consortium, 2015. LD Score regression distinguishes  
444 confounding from polygenicity in genome-wide association studies. *Nature Genet.*  
445 47, 291.

446

447 Burbano, H.A., Hodges, E., Green, R.E., Briggs, A.W., Krause, J., Meyer, M., Good,  
448 J.M., Maricic, T., Johnson, P.L.F., Xuan, Z., Rooks, M., Bhattacharjee, A.,  
449 Brizuela, L., Albert, F.W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M.,  
450 Hannon, G.J., Pääbo, S., 2010. Targeted investigation of the Neandertal genome by  
451 array-based sequence capture. *Science* 328, 723–5.

452

- 453 Bush, E.C., Lahn, B.T., 2008. A genome-wide screen for noncoding elements  
454 important in primate evolution. *BMC Evol. Biol.* 8, 17.  
455
- 456 Cardno, A.G., Marshall, E.J., Coid, B., Macdonald, A.M., Ribchester, T.R., Davies,  
457 N.J., Venturi, P., Jones, L.A., Lewis, S.W., Sham, P.C., et al., 1999. Heritability  
458 estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch. Gen.*  
459 *Psychiatry* 56, 162–168.  
460
- 461 Christoforou, A., Dondrup, M., Mattingsdal, M., Mattheisen, M., Giddaluru, S.,  
462 Nthen, M.M., Rietschel, M., Cichon, S., Djurovic, S., Andreassen, O.A., et al.,  
463 2012. Linkage-disequilibrium-based binning affects the interpretation of GWASs.  
464 *Am. J. Hum. Genet.* 90, 727–733.  
465
- 466 Crespi, B., Summers, K., Dorus, S., 2007. Adaptive evolution of genes underlying  
467 schizophrenia. *Proc. Roy. Soc. Lond. [Biol]* 274, 2801–2810.  
468
- 469 Crow, T.J., 1995. A Darwinian approach to the origins of psychosis. *Br. J. Psychiatry*  
470 167, 12–25.  
471
- 472 Crow, T.J., 1997. Is schizophrenia the price that *Homo sapiens* pays for language?  
473 *Schizophr. Res.* 28, 127–41.  
474
- 475 Crow, T.J., 2000. Schizophrenia as the price that *Homo sapiens* pays for language: a  
476 resolution of the central paradox in the origin of the species. *Brain Research*  
477 *Reviews* 31, 118–129.  
478
- 479 Crow, T.J., 2008. The “big bang” theory of the origin of psychosis and the faculty of  
480 language. *Schizophr. Res.* 102, 31–52.  
481
- 482 Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E.,  
483 Belliveau, R., Bybjerg-Grauholm, J., Bækved-Hansen, M., Cerrato, F., et al., 2017.  
484 Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv*  
485 145581.  
486
- 487 Durbin, R., Abecasis, G., Bentley, D., Chakravarti, A., Clark, A., Donnelly, P.,  
488 Eichler, E., Flicek, P., Gabriel, S., Gibbs, R., et al., 2012. An integrated map of  
489 genetic variation from 1,092 human genomes. *Nature* 497, 56–65.  
490
- 491 Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I.,  
492 Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.-J., et al., 2011. Genetic  
493 variants in novel pathways influence blood pressure and cardiovascular disease  
494 risk. *Nature* 478, 103.  
495
- 496 Gejman, P.V., Sanders, A.R., Kendler, K.S., 2011. Genetics of schizophrenia: new  
497 findings and challenges. *Annu. Rev. Genomics Hum. Genet.* 12, 121–44.  
498
- 499 Gittelman, R.M., Hun, E., Ay, F., Madeoy, J., Pennacchio, L., Noble, W.S., Hawkins,  
500 R.D., Akey, J.M., 2015. Comprehensive identification and analysis of human  
501 accelerated regulatory DNA. *Genome Res.* 25, 1245–1255.  
502

- 503 Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of  
504 primate species. *Mol. Biol. Evol.* 20, 424–34.  
505
- 506 Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S.,  
507 Meshorer, E., Carmel, L., 2014. Reconstructing the DNA methylation maps of the  
508 Neandertal and the Denisovan. *Science* 344, 523–527.  
509
- 510 Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson,  
511 N., Li, H., Zhai, W., Fritz, M.H.-Y., et al., 2010. A draft sequence of the  
512 Neandertal genome. *Science* 328, 710–722.  
513
- 514 Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a  
515 molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22, 160–74.  
516
- 517 Hernando-Herraez, I., Heyn, H., Fernandez-Callejo, M., Vidal, E., Fernandez-Bellon,  
518 H., Prado-Martinez, J., Sharp, A.J., Esteller, M., Marques-Bonet, T., 2015. The  
519 interplay between DNA methylation and sequence divergence in recent human  
520 evolution. *Nucleic Acids Res.* 43, 8204–8214.  
521
- 522 Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H.,  
523 Hvilsom, C., Navarro, A., Esteller, M., Sharp, A.J., Marques-Bonet, T., 2013.  
524 Dynamics of DNA methylation in recent human and great ape evolution. *PLoS*  
525 *Genet* 9, e1003763.  
526
- 527 Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K.,  
528 Lush, M.J., Povey, S., Talbot Jr, C.C., Wright, M.W., et al., 2004. Gene map of the  
529 extended human MHC. *Nat. Rev. Genet.* 5, 889-899.  
530
- 531 Jeste, D.V., Del Carmen, R., Lohr, J.B., Wyatt, R.J., 1985. Did schizophrenia exist  
532 before the eighteenth century? *Compr. Psychiatry* 26, 493–503.  
533
- 534 Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M.,  
535 Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S.,  
536 Hinrichs, A.S., Learned, K., Lee, B.T., Li, C.H., Raney, B.J., Rhead, B.,  
537 Rosenbloom, K.R., Sloan, C.A., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn,  
538 R.M., Kent, W.J., 2014. The UCSC Genome Browser database: 2014 update.  
539 *Nucleic Acids Res.* 42, D764–70.  
540
- 541 Kromer, J., Hummel, T., Pietrowski, D., Giani, A., Sauter, J., Ehninger, G., Schmidt,  
542 A., Croy, I., 2016. Influence of HLA on human partnership and sexual satisfaction.  
543 *Scientific Reports* 6, 32550.  
544
- 545 Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A.,  
546 Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al., 2015. Integrative analysis  
547 of 111 reference human epigenomes. *Nature* 518, 317–330.  
548
- 549 Kyaga, S., Lichtenstein, P., Boman, M., Hultman, C., Långström, N., Landén, M.,  
550 2011. Creativity and mental disorder: family study of 300 000 people with severe  
551 mental disorder. *Brit. J. Psych.* 199, 373–379.  
552

- 553 Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon,  
554 K., Dewar, K., Doyle, M., FitzHugh, W., et al., 2001. Initial sequencing and  
555 analysis of the human genome. *Nature* 409, 860-921.  
556
- 557 Langergraber, K.E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K.,  
558 Inoue, E., Inoue-Muruyama, M., Mitani, J.C., Muller, M.N., Robbins, M.M.,  
559 Schubert, G., Stoinski, T.S., Viola, B., Watts, D., Wittig, R.M., Wrangham, R.W.,  
560 Zuberbühler, K., Pääbo, S., Vigilant, L., 2012. Generation times in wild  
561 chimpanzees and gorillas suggest earlier divergence times in great ape and human  
562 evolution. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15716–15721.  
563
- 564 Larson, C.A., Nyman, G.E., 1973. Differential fertility in schizophrenia. *Acta*  
565 *Psychiatr Scand* 49, 272–80.  
566
- 567 Lee, P.H., O’Dushlaine, C., Thomas, B., Purcell, S.M., 2012. INRICH: interval-based  
568 enrichment analysis for genome-wide association studies. *Bioinformatics* 28,  
569 1797–9.  
570
- 571 Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F.,  
572 Hultman, C.M., 2009. Common genetic determinants of schizophrenia and bipolar  
573 disorder in Swedish families: a population-based study. *Lancet* 373, 234–239.  
574
- 575 Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S.,  
576 Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., Ward, L.D., Lowe, C.B.,  
577 Holloway, A.K., Clamp, M., Gnerre, S., Alford, J., Beal, K., Chang, J., Clawson,  
578 H., Cuff, J., Di Palma, F., Fitzgerald, S., Flicek, P., Guttman, M., Hubisz, M.J.,  
579 Jaffe, D.B., Jungreis, I., Kent, W.J., Kostka, D., Lara, M., Martins, A.L.,  
580 Massingham, T., Moltke, I., Raney, B.J., Rasmussen, M.D., Robinson, J., Stark, A.,  
581 Vilella, A.J., Wen, J., Xie, X., Zody, M.C., Baldwin, J., Bloom, T., Whye Chin, C.,  
582 Heiman, D., Nicol, R., Nusbaum, C., Young, S., Wilkinson, J., Worley, K.C.,  
583 Kovar, C.L., Muzny, D.M., Gibbs, R. a., Cree, A., Dihm, H.H., Fowler, G.,  
584 Jhangiani, S., Joshi, V., Lee, S., Lewis, L.R., Nazareth, L.V., Okwuonu, G.,  
585 Santibanez, J., Warren, W.C., Mardis, E.R., Weinstock, G.M., Wilson, R.K.,  
586 Delehaunty, K., Dooling, D., Fronik, C., Fulton, L., Fulton, B., Graves, T., Minx,  
587 P., Sodergren, E., Birney, E., Margulies, E.H., Herrero, J., Green, E.D., Haussler,  
588 D., Siepel, A., Goldman, N., Pollard, K.S., Pedersen, J.S., Lander, E.S., Kellis, M.,  
589 2011. A high-resolution map of human evolutionary constraint using 29 mammals.  
590 *Nature* 478, 476–482.  
591
- 592 Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C.,  
593 Vedantam, S., Buchkovich, M.L., Yang, J., et al., 2015. Genetic studies of body  
594 mass index yield new insights for obesity biology. *Nature* 518, 197–206.  
595
- 596 Markow, T.A., 2012. Developmental instability: its origins and evolutionary  
597 implications: proceedings of the International Conference on Developmental  
598 Instability: Its Origins and Evolutionary Implications, Tempe, Arizona, 14-15 June  
599 1993. Springer Science and Business Media.  
600



- 601 Mendizabal, I., Keller, T., Zeng, J., Soojin, V.Y., 2014. Epigenetics and evolution.  
602 Integr. Comp. Biol. 54, 31–42.  
603
- 604 Messias, E.L., Chen, C.-Y., Eaton, W.W., 2007. Epidemiology of schizophrenia:  
605 review of findings and myths. Psychiatr. Clin. North Am. 30, 323–38.  
606
- 607 Meyer, D., Aguiar, V.R., Bitarello, B.D., Brandt, D.Y., Nunes, K., 2017. A genomic  
608 perspective on HLA evolution. Immunogenetics 70, 1–23.  
609
- 610 Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W.R., Hannon, G.J., Smith,  
611 A.D., 2011. Sperm methylation profiles reveal features of epigenetic inheritance  
612 and evolution in primates. Cell 146, 1029–1041.  
613
- 614 Nichols, C., 2009. Is there an evolutionary advantage of schizophrenia? Pers. Individ.  
615 Diff. 46, 832–838.  
616
- 617 Paaby, A.B., Rockman, M.V., 2014. Cryptic genetic variation: evolution’s hidden  
618 substrate. Nat. Rev. Genet. 15, 247–258  
619
- 620 Pollard, K.S., Salama, S.R., King, B., Kern, A.D., Dreszer, T., Katzman, S., Siepel,  
621 A., Pedersen, J.S., Bejerano, G., Baertsch, R., et al., 2006. Forces shaping the  
622 fastest evolving regions in the human genome. PLoS Genet 2, e168.  
623
- 624 Potts, W.K., Wakeland, E.K., 1990. Evolution of diversity at the major  
625 histocompatibility complex. Trends Ecol. Evol. (Amst.) 5, 181–7.  
626
- 627 Power, R.A., Kyaga, S., Uher, R., MacCabe, J.H., Långström, N., Landen, M.,  
628 McGuffin, P., Lewis, C.M., Lichtenstein, P., Svensson, A.C., 2013. Fecundity of  
629 patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa,  
630 or substance abuse vs their unaffected siblings. JAMA Psychiatry 70, 22–30.  
631
- 632 Rannala, B., Yang, Z., 2003. Bayes estimation of species divergence times and  
633 ancestral population sizes using DNA sequences from multiple loci. Genetics 164,  
634 1645–1656.  
635
- 636 R Core Team, 2017. R: A Language and Environment for Statistical Computing.  
637 Vienna, Austria.  
638
- 639 Ripke, S., Neale, B.M., Corvin, A., Walters, J.T., Farh, K.-H., Holmans, P.A., Lee, P.,  
640 Bulik-Sullivan, B., Collier, D.A., Huang, H., et al., 2014. Biological insights from  
641 108 schizophrenia-associated genetic loci. Nature 511, 421–427.  
642
- 643 Rivera, C.M., Ren, B., 2013. Mapping human epigenomes. Cell 155, 39–55.  
644
- 645 Roberts, S.C., Gosling, L.M., Carter, V., Petrie, M., 2008. MHC-correlated odour  
646 preferences in humans and the use of oral contraceptives. Proc. Biol. Sci. 275,  
647 2715–2722.  
648
- 649 Santos, P.S.C., Schinemann, J.A., Gabardo, J., Bicalho, M. da G., 2005. New  
650 evidence that the MHC influences odor perception in humans: a study with 58

- 651 Southern Brazilian students. *Horm. Behav.* 47, 384–388.
- 652
- 653 Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan,  
654 P.F., Kelsoe, J.R., O’Donovan, M.C., Furberg, H., Schork, N.J., et al., 2013. All  
655 SNPs are not created equal: genome-wide association studies reveal a consistent  
656 pattern of enrichment among functionally annotated SNPs. *PLoS Genet* 9,  
657 e1003449.
- 658
- 659 Schork, A.J., Wang, Y., Thompson, W.K., Dale, A.M., Andreassen, O.A., 2016. New  
660 statistical approaches exploit the polygenic architecture of schizophrenia—  
661 implications for the underlying neurobiology. *Curr. Opin. Neurobiol.* 36, 89–98.
- 662
- 663 Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N.,  
664 Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al., 2016. Schizophrenia  
665 risk from complex variation of complement component 4. *Nature* 530, 177–183.
- 666
- 667 Sklar, P., Ripke, S., Scott, L.J., Andreassen, O.A., Cichon, S., Craddock, N.,  
668 Edenberg, H.J., Nurnberger, J.I., Rietschel, M., Blackwood, D., et al., 2011. Large-  
669 scale genome-wide association analysis of bipolar disorder identifies a new  
670 susceptibility locus near ODZ4. *Nature Genet.* 43, 977–983.
- 671
- 672 Skre, I., Onstad, S., Torgersen, S., Lygren, S., Kringlen, E., 1993. A twin study of  
673 DSM-III-R anxiety disorders. *Acta Psychiatr. Scand.* 88, 85–92.
- 674
- 675 Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R., Krapohl, E.,  
676 Taskesen, E., Hammerschlag, A.R., Okbay, A., Zabaneh, D., et al., 2017. Genome-  
677 wide association meta-analysis of 78,308 individuals identifies new loci and genes  
678 influencing human intelligence. *Nature Genet.* 49, 1107–1112.
- 679
- 680 Sommer, S., 2005. The importance of immune gene variability (MHC) in  
681 evolutionary ecology and conservation. *Front. Zool.* 2, 16.
- 682
- 683 Srinivasan, S., Bettella, F., Hassani, S., Wang, Y., Witoelar, A., Schork, A.J.,  
684 Thompson, W.K., Collier, D.A., Desikan, R.S., Melle, I., et al., 2017. Probing the  
685 association between early evolutionary markers and schizophrenia. *PloS One* 12,  
686 e0169227.
- 687
- 688 Srinivasan, S., Bettella, F., Mattingsdal, M., Wang, Y., Witoelar, A., Schork, A.J.,  
689 Thompson, W.K., Zuber, V., Winsvold, B.S., Zwart, J.-A., et al., 2015. Genetic  
690 markers of human evolution are enriched in schizophrenia. *Biol. Psychiatry* 80,  
691 284–292.
- 692
- 693 Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li,  
694 Y., Kurreeman, F.A., Zernakova, A., Hinks, A., et al., 2010. Genome-wide  
695 association study meta-analysis identifies seven new rheumatoid arthritis risk loci.  
696 *Nature Genet.* 42, 508–514.
- 697
- 698 Sullivan, P.F., Kendler, K.S., Neale, M.C., 2003. Schizophrenia as a complex trait:  
699 evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–

700 1192.

701

702 Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M.,  
703 Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., Johansen,  
704 C.T., Fouchier, S.W., Isaacs, A., Peloso, G.M., Barbalic, M., Ricketts, S.L., Bis,  
705 J.C., Aulchenko, Y.S., Thorleifsson, G., Feitosa, M.F., Chambers, J., Orho-  
706 Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin  
707 Go, M., Jin Kim, Y., Lee, J.-Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R.,  
708 Croteau-Chonka, D.C., Lange, L.A., Smith, J.D., Song, K., Hua Zhao, J., Yuan, X.,  
709 Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R.Y.L., Wright, A.F., Witteman,  
710 J.C.M., Wilson, J.F., Willemsen, G., Wichmann, H.-E., Whitfield, J.B.,  
711 Waterworth, D.M., Wareham, N.J., Waeber, G., Vollenweider, P., Voight, B.F.,  
712 Vitart, V., Uitterlinden, A.G., Uda, M., Tuomilehto, J., Thompson, J.R., Tanaka,  
713 T., Surakka, I., Stringham, H.M., Spector, T.D., Soranzo, N., Smit, J.H., Sinisalo,  
714 J., Silander, K., Sijbrands, E.J.G., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S.,  
715 Salomaa, V., Saharinen, J., Sabatti, C., Ruukonen, A., Rudan, I., Rose, L.M.,  
716 Roberts, R., Rieder, M., Psaty, B.M., Pramstaller, P.P., Pichler, I., Perola, M.,  
717 Penninx, B.W.J.H., Pedersen, N.L., Pattaro, C., Parker, A.N., Pare, G., Oostra,  
718 B.A., O'Donnell, C.J., Nieminen, M.S., Nickerson, D.A., Montgomery, G.W.,  
719 Meitinger, T., McPherson, R., McCarthy, M.I., McArdle, W., Masson, D., Martin,  
720 N.G., Marroni, F., Mangino, M., Magnusson, P.K.E., Lucas, G., Luben, R., Loos,  
721 R.J.F., Lokki, M.-L., Lettre, G., Langenberg, C., Launer, L.J., Lakatta, E.G.,  
722 Laaksonen, R., Kyvik, K.O., Kronenberg, F., König, I.R., Khaw, K.-T., Kaprio, J.,  
723 Kaplan, L.M., Johansson, A., Jarvelin, M.-R., Janssens, A.C.J.W., Ingelsson, E.,  
724 Igl, W., Kees Hovingh, G., Hottenga, J.-J., Hofman, A., Hicks, A.A.,  
725 Hengstenberg, C., Heid, I.M., Hayward, C., Havulinna, A.S., Hastie, N.D., Harris,  
726 T.B., Haritunians, T., Hall, A.S., Gyllensten, U., Guiducci, C., Groop, L.C.,  
727 Gonzalez, E., Gieger, C., Freimer, N.B., Ferrucci, L., Erdmann, J., Elliott, P.,  
728 Ejebe, K.G., Döring, A., Dominiczak, A.F., Demissie, S., Deloukas, P., de Geus,  
729 E.J.C., de Faire, U., Crawford, G., Collins, F.S., Chen, Y.I., Caulfield, M.J.,  
730 Campbell, H., Burt, N.P., Bonnycastle, L.L., Boomsma, D.I., Boekholdt, S.M.,  
731 Bergman, R.N., Barroso, I., Bandinelli, S., Ballantyne, C.M., Assimes, T.L.,  
732 Quertermous, T., Altshuler, D., Seielstad, M., Wong, T.Y., Tai, E.-S., Feranil,  
733 A.B., Kuzawa, C.W., Adair, L.S., Taylor, H.A., Borecki, I.B., Gabriel, S.B.,  
734 Wilson, J.G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R.M., Mohlke,  
735 K.L., Ordovas, J.M., Munroe, P.B., Kooner, J.S., Tall, A.R., Hegele, R.A.,  
736 Kastelein, J.J.P., Schadt, E.E., Rotter, J.I., Boerwinkle, E., Strachan, D.P., Mooser,  
737 V., Stefansson, K., Reilly, M.P., Samani, N.J., Schunkert, H., Cupples, L.A.,  
738 Sandhu, M.S., Ridker, P.M., Rader, D.J., van Duijn, C.M., Peltonen, L., Abecasis,  
739 G.R., Boehnke, M., Kathiresan, S., 2010. Biological, clinical and population  
740 relevance of 95 loci for blood lipids. *Nature* 466, 707–713.

741

742 Traherne, J., 2008. Human MHC architecture and evolution: implications for disease  
743 association studies. *Int. J. Immunogenet.* 35, 179–192.

744

745 Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith,  
746 H.O., Yandell, M., Evans, C.A., Holt, R.A., et al., 2001. The sequence of the  
747 human genome. *Science* 291, 1304–1351.

748

- 749 WHO, 1973. Report of the international pilot study of schizophrenia.  
750  
751 Winternitz, J., Abbate, J., Huchard, E., Havlíček, J., Garamszegi, L., 2017.  
752 Patterns of MHC-dependent mate selection in humans and nonhuman primates: a  
753 meta-analysis. *Molecular ecology* 26, 668–688.  
754  
755 Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y.,  
756 Estrada, K., Luan, J., Kutalik, Z., et al., 2014. Defining the role of common  
757 variation in the genomic and biological architecture of adult human height. *Nature*  
758 *Genet* 46, 1173–1186.  
759  
760 Xu, K., Schadt, E.E., Pollard, K.S., Roussos, P., Dudley, J.T., 2015. Genomic and  
761 network patterns of schizophrenia genetic variation in human evolutionary  
762 accelerated regions. *Mol Biol Evol* 32, 1148–1160.  
763  
764  
  
765  
  
766  
  
767

768 **Figure legends**

769 **Fig. 1: Enrichment plots of hypo-DMR and hyper-DMR SNPs in schizophrenia**

770 Quantile-quantile (QQ) plots of GWAS SNPs for Schizophrenia (SCZ) with the  
771 extended MHC region (chr6: 25-35Mb) unmasked (A) and masked (B). Expected -  
772  $\log_{10} p$ -values under the null hypothesis are shown on the X-axis. Observed  $-\log_{10} p$ -  
773 values are on the Y-axis. The values for all GWAS SNPs are plotted in dark green  
774 while the values for SNPs in linkage disequilibrium (LD) with hypo-methylated  
775 DMRs are plotted in blue and SNPs in LD with hyper-methylated DMRs are plotted  
776 in pink. A leftward deflection of the plotted  $p$ -values from the line for all GWAS  
777 SNPs indicates enrichment of true signals – the greater the leftward deflection, the  
778 stronger the enrichment. Genomic correction was performed on all SNPs with global  
779 lambda.

780

781 **Fig. 2: Enrichment plots of hypo-DMR and hyper-DMR SNPs across**  
782 **multiple traits**

783 Thirteen different GWASs were analysed using a common set of ~2.4 million  
784 SNPs. The  $p$ -values for the common set of GWAS SNPs are plotted in dark  
785 green;  $p$ -values for SNPs that tag hypo-methylated DMRs are plotted in blue;  
786 and  $p$ -values for SNPs that tag hyper-methylated DMRs are plotted in pink.  
787 ADHD, attention deficit hyperactivity disorder; BMI, body mass index; BPD,  
788 bipolar disorder; DBP, diastolic blood pressure; HDL, high density lipoprotein;  
789 LDL, low density lipoprotein; RA, rheumatoid arthritis; SBP, systolic blood  
790 pressure; SCZ, schizophrenia; TC, total cholesterol; TG, triglycerides. The  
791 MHC region was absent from the common set of SNPs.

792

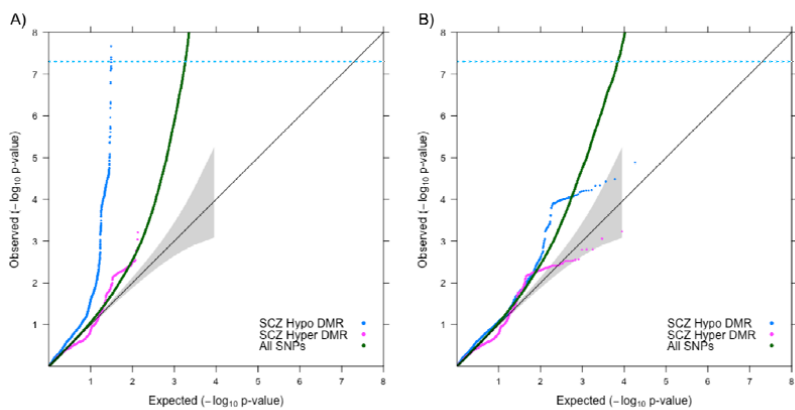
793 **Fig. 3: INRICH test for enrichment of association of DMR gene sets and NSS**  
794 **genes with SCZ, MHC masked**

795 A visual heatmap depicting  $p$ -values from bootstrapping with 5,000 iterations. The  
796 various evolutionary annotations compared are as follows. HypoDMR – human hypo-  
797 methylated DMRs; HyperDMR – human hyper-methylated DMRs. HypoDMR and  
798 HyperDMR were taken from the study by Hernando-Herraez et al. (2013). dmrH –  
799 human-specific DMRs (Gokhman et al, 2014), which are referred to as non-primate  
800 DMRs in this manuscript. NSS - Neanderthal selective sweep. Datasets marked with \*  
801 have been previously reported by Banerjee et al. (2017) and are presented here for  
802 comparison only.

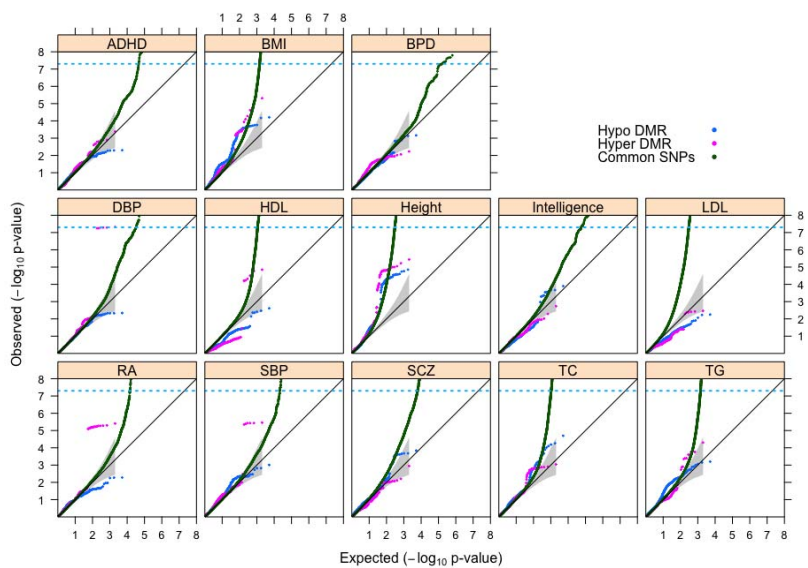
803

804

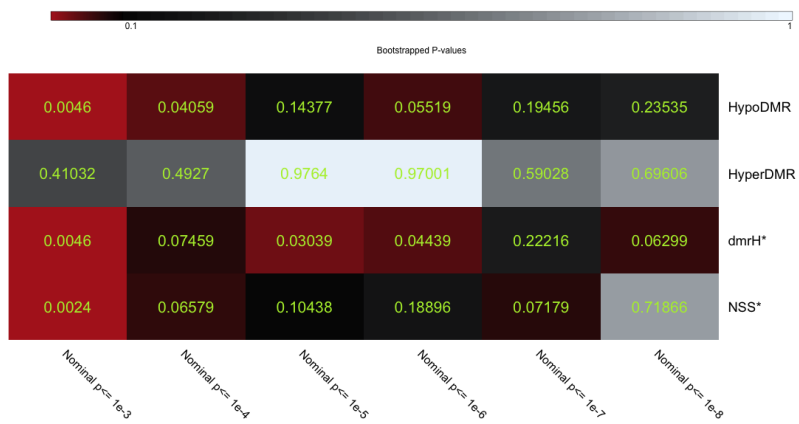
**Figure 1**



**Figure 2**



**Figure 3**





**Analysis of differentially methylated regions in primates and non-primates provides support for the evolutionary hypothesis of schizophrenia**

Niladri Banerjee, Tatiana Polushina, Francesco Bettella,

Vidar M. Steen, Ole A. Andreassen, Stephanie Le Hellard

**SUPPLEMENTARY MATERIAL**

<b>Supplementary Methods</b>	<b>page 2</b>
<b>Enrichment analysis with INRICH</b>	<b>page 2</b>
<b>LD Pruning</b>	<b>page 3</b>
<b>Supplementary Figures</b>	<b>page 4</b>
<b>Supplementary Tables</b>	<b>page 8</b>
<b>Supplementary References</b>	<b>page 10</b>

## Supplementary Methods

### Enrichment Analysis with INRICH

Before using INRICH to assess enrichment, it is necessary to generate genomic intervals to test for. For our purpose, we generated LD-implicated genomic intervals with PLINK (Purcell et al., 2007), using a procedure similar to that described by Lee et al. (2012). We generated intervals from SNPs at different disease significance thresholds through LD clumping in PLINK for index SNPs with  $p$ -values  $1 \times 10^{-3}$  to  $1 \times 10^{-8}$ . LD clumps were formed at  $r^2=0.5$  with the clump range limited to 250 kb. INRICH was run on all the sets of LD intervals using the default parameters described by Lee et al. (2012). The various gene sets tested were obtained using a procedure similar to that described by Xu et al. (2015), and included genes within a 100 kb flanking region of human-specific DMRs (dmrH), HARs, PARs, primate HARs, hypo-methylated DMRs, hyper-methylated DMRs, and NSS markers. For NSS markers, genes were assigned via LD blocks of  $r^2 \geq 0.8$  since NSS markers are single-base markers unlike the HARs and DMRs, which are interval regions. GENCODE v19 gene database (last accessed 5<sup>th</sup> February 2016) was used to map the genes to DMRs, NSS markers, HARs, hypo-methylated DMRs and hyper-methylated DMRs.

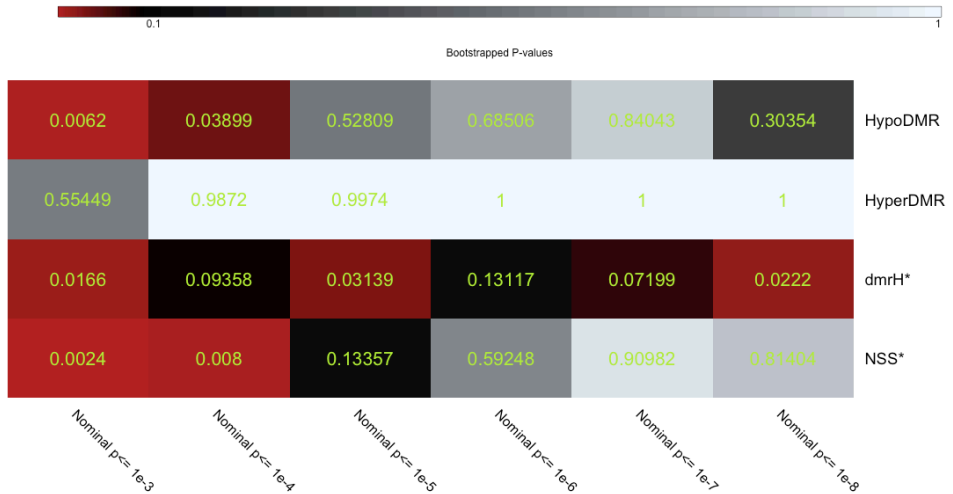
To avoid counting the same region of the genome more than once and potentially inflating the test statistics, INRICH combines overlapping genes and overlapping LD-implicated intervals. In this study, a total of 2480, 1010, 444, 206, 107 and 68 LD-implicated intervals were analysed respectively for SNPs with  $p$ -values  $\leq 1 \times 10^{-3}$ ,  $1 \times 10^{-4}$ ,  $1 \times 10^{-5}$ ,  $1 \times 10^{-6}$ ,  $1 \times 10^{-7}$  and  $1 \times 10^{-8}$  in the schizophrenia GWAS. For the same  $p$ -values, a total of 4097, 2395, 1536, 1089, 860 and 678 intervals were analysed for the height GWAS, and a total of 711, 254, 96, 56, 29 and 15 intervals were analysed for the

pruned SCZ GWAS (see below). Statistical significance in INRICH is determined via a two-stage procedure utilizing empirical null distributions and bootstrapping procedures. In the first step, a user-defined number of permutations is used to determine the background null distribution of chance overlap between intervals and gene sets. This is then followed by multiple testing procedures via bootstrapping in the second step. We used the default INRICH parameters of 10,000 permutations and 5,000 rounds of bootstrapping (Lee et al., 2012).

### **LD Pruning**

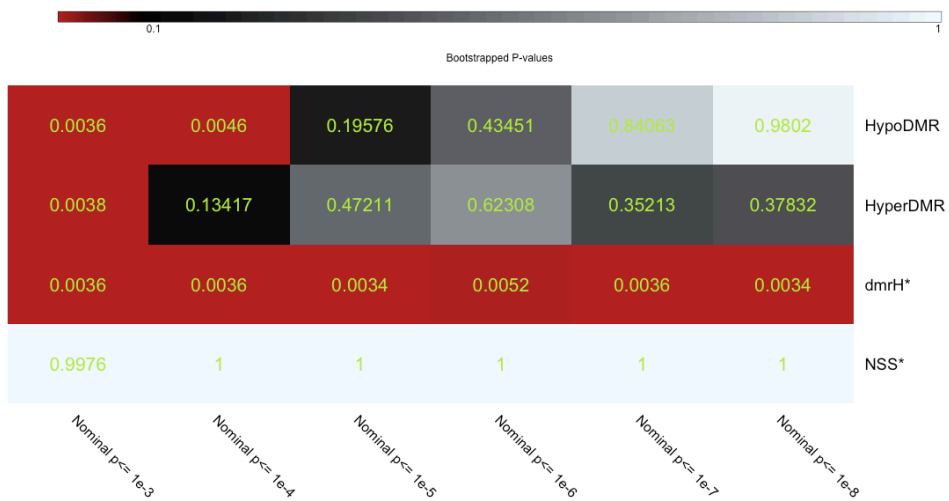
In GWAS studies, due to the phenomenon of LD, many SNPs end up providing similar information in terms of statistics. To avoid this redundancy and increase the robustness further, we also performed INRICH on the LD-implicated intervals of pruned sets of SNPs. We generated 100 randomly pruned sets for all SCZ GWAS SNPs and 100 randomly pruned sets for SCZ GWAS SNPs annotated to DMR regions. We mapped all SNPs to 1000 Genomes genotypes and computed pairwise correlations for all markers. For 8,572,136 GWAS SNPs and 26,975 DMR SNPs that were mapped to the 1000 Genomes genotypes, we repeated the LD pruning protocol from Schork et al. (2013). In that analysis, all non-SNP signals were excluded (for example, insertions and deletions). For further analysis, we masked the extended MHC region in the SCZ GWAS (chr6: 25Mbp-35Mbp, containing 52,903 markers). SNPs were pruned randomly to approximate independence ( $r^2 < 0.2$ ) 100 times. Finally, using PLINK, LD clumping was performed as described previously to generate LD intervals from ~180,000 pruned SNPs.

## Supplementary Figures



### Supplementary Fig. 1: INRICH test for enrichment of association of DMR gene sets and NSS gene sets with pruned sets of SCZ GWAS SNPs

A visual heatmap depicting the  $p$ -values obtained from 5,000 bootstrap iterations. The various evolutionary annotations compared are as follows. HypoDMR – human hypo-methylated DMRs; HyperDMR – human hyper-methylated DMRs. HypoDMR and HyperDMR are taken from the study by Hernando-Herraez et al. (2013). dmrH - human-specific DMRs, which are referred to as non-primate DMRs in the manuscript of Gokhman et al. (2014). NSS - Neanderthal Selective Sweep markers. Datasets marked with \* have been previously reported by Banerjee et al. (2017) and are shown here for comparison only.



## Supplementary Fig. 2: INRICH test for enrichment of association of DMR gene sets and NSS gene with height GWAS SNPs

A visual heatmap depicting the  $p$ -values obtained from 5,000 bootstrap iterations. The various evolutionary annotations compared are as follows. HypoDMR – human hypo-methylated DMRs; HyperDMR – human hyper-methylated DMRs. HypoDMR and HyperDMR are taken from the study by Hernando-Herraez et al. (2013). dmrH – human-specific DMRs, which are referred to as non-primate DMRs in the manuscript of Gokhman et al. (2014). NSS – Neanderthal Selective Sweep markers. HAR – mammalian conserved regions that are accelerated in humans. PAR - mammalian conserved regions that are accelerated in primates. PrHAR (primate HAR) – primate-conserved regions that are accelerated in humans. Datasets marked with \* have been previously reported by Banerjee et al. (2017) and are presented here for comparison only.

## Supplementary Tables

	<b>Gokhman DMR</b> (n = 891)	<b>Neanderthal DMR</b> (n = 307)	<b>Denisovan DMR</b> (n = 295)
<b>Human hypoDMR</b> (n = 360)	5	0	1
	<b>primate HAR</b> (n = 1928)	<b>mammalian HAR</b> (n = 556)	<b>PAR</b> (n = 577)
<b>Human hypoDMR</b> (n = 360)	1	0	0

**Supplementary Table 1** Table describing the overlaps of various previously defined evolutionary annotations with human hypo-methylated DMRs. The cells show the number of regions of the respective annotations that overlap with at least one human hypo-methylated DMR.

	<b>Gokhman DMR</b> (n = 891)	<b>Neanderthal DMR</b> (n = 307)	<b>Denisovan DMR</b> (n = 295)
<b>Human hyperDMR</b> (n = 210)	2	2	2
	<b>primate HAR</b> (n = 1928)	<b>mammalian HAR</b> (n = 556)	<b>PAR</b> (n = 577)
<b>Human hyperDMR</b> (n = 210)	0	0	0

**Supplementary Table 2** Table describing the overlaps of various previously defined evolutionary annotations with human hyper-methylated DMRs. The cells show the number of regions of the respective annotations that overlap with at least one human hyper-methylated DMR.

## Supplementary References

- Banerjee, N., Polushina, T., Bettella, F., Giddaluru, S., Steen, V.M., Andreassen, O.A., Le Hellard, S., 2017. Human-specific methylated regions are enriched in schizophrenia. *bioRxiv* 113175.
- Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S., Meshorer, E., Carmel, L., 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* 344, 523–527.
- Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H., Hvilsom, C., Navarro, A., Esteller, M., Sharp, A.J., Marques-Bonet, T., 2013. Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* 9, e1003763.
- Lee, P.H., O'Dushlaine, C., Thomas, B., Purcell, S.M., 2012. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 28, 1797–1799.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., others, 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559–575.
- Schork, A.J., Thompson, W.K., Pham, P., Torkamani, A., Roddey, J.C., Sullivan, P.F., Kelsoe, J.R., O'Donovan, M.C., Furberg, H., Schork, N.J., others, 2013. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet* 9, e1003449.
- Xu, K., Schadt, E.E., Pollard, K.S., Roussos, P., Dudley, J.T., 2015. Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol Biol Evol* 32, 1148–1160.



**Supplementary Dataset 1:** Annotation of human hypo-methylated regions with markers of schizophrenia

*This dataset may be downloaded for review at*

[https://www.dropbox.com/s/g48nubbem7v66nb/Banerjee\\_Supplementary\\_Dataset\\_forPhD\\_Dissertation.xlsx?dl=0](https://www.dropbox.com/s/g48nubbem7v66nb/Banerjee_Supplementary_Dataset_forPhD_Dissertation.xlsx?dl=0)





Graphic design: Communication Division, UIB / Print: Skjipes Kommunikasjon AS



[uib.no](http://uib.no)

ISBN: 978-82-308-3664-4