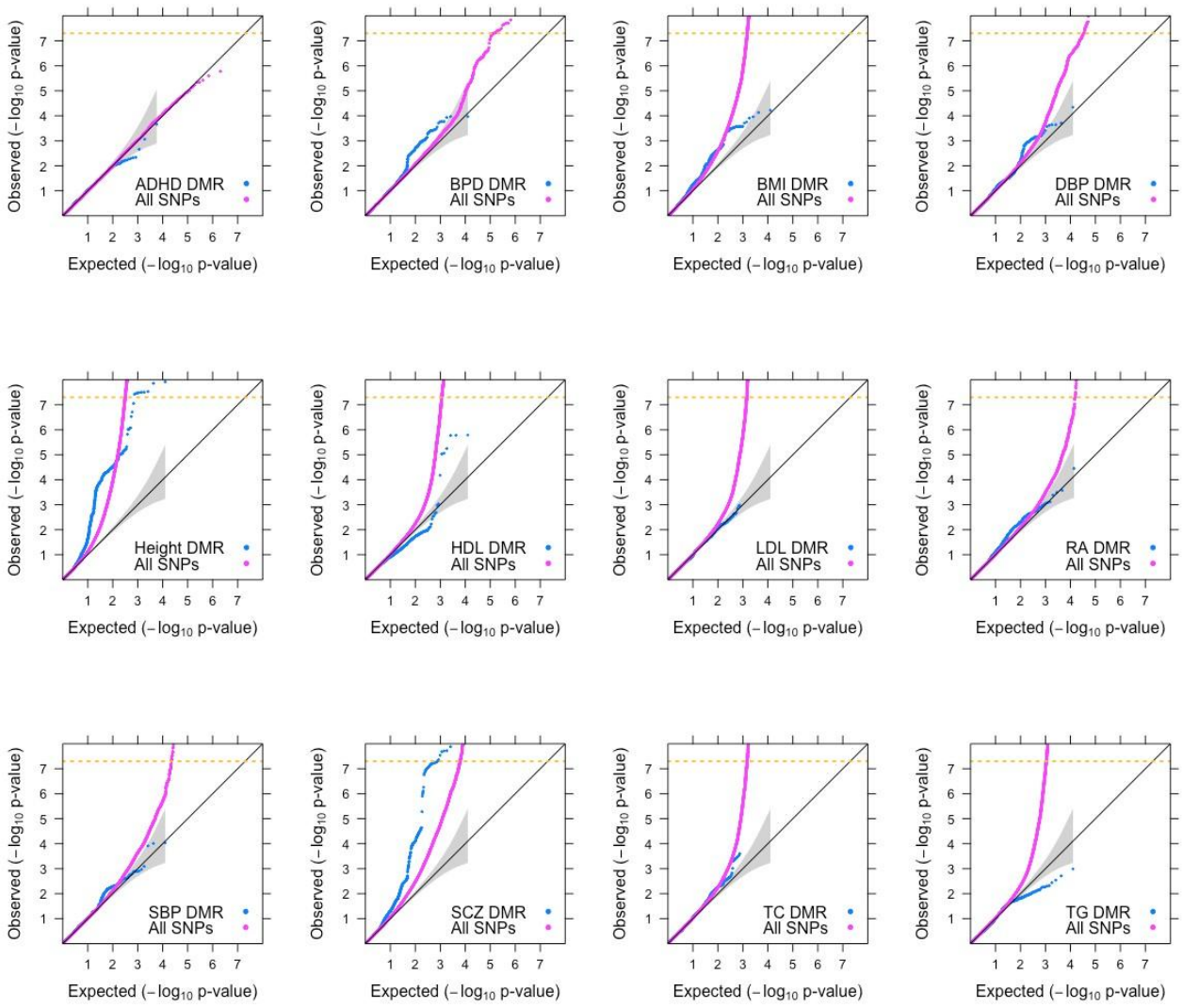
Supplemental Material

Human specific methylated regions are enriched in schizophrenia

Banerjee et al

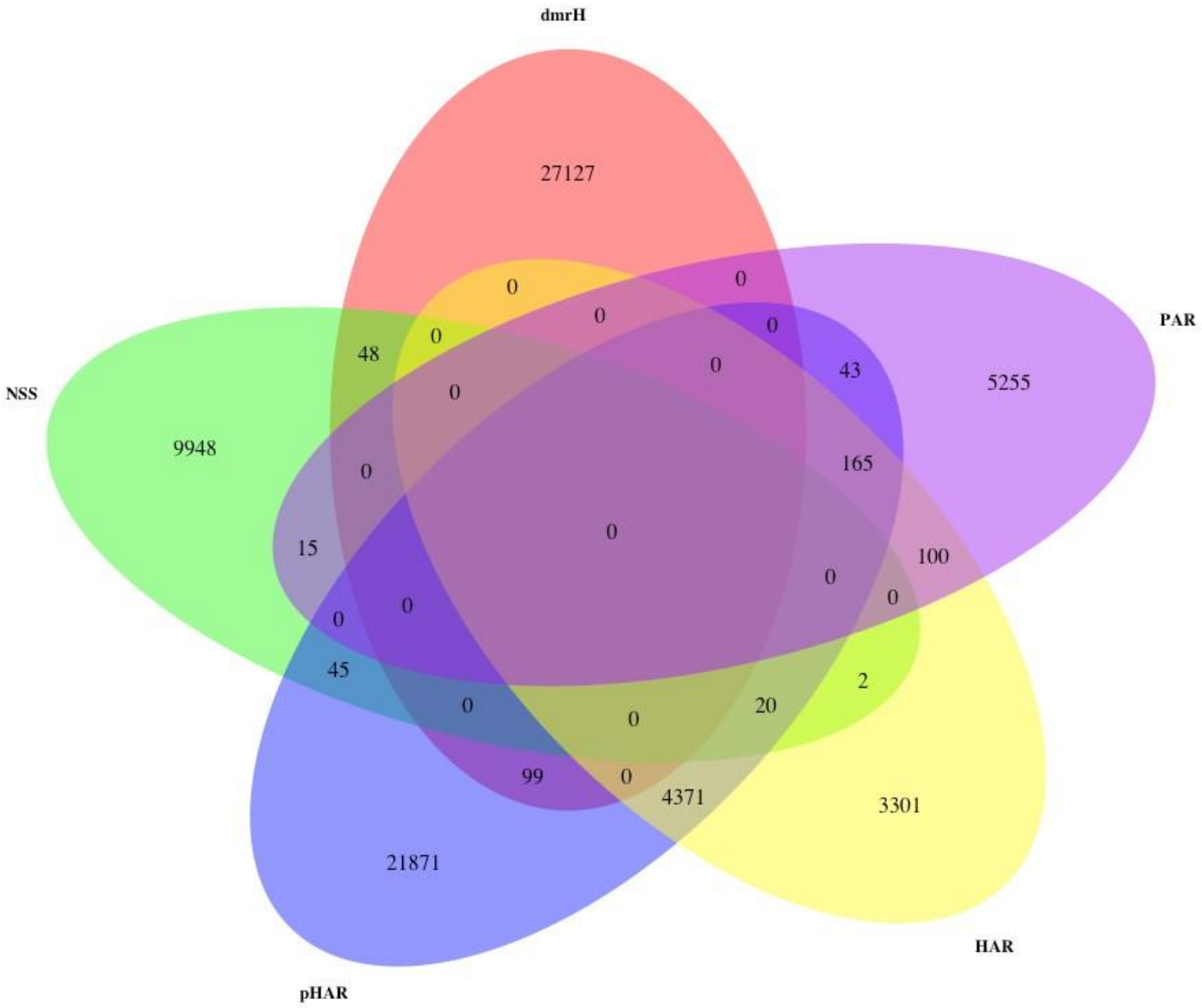
Contents

Supplemental Figures



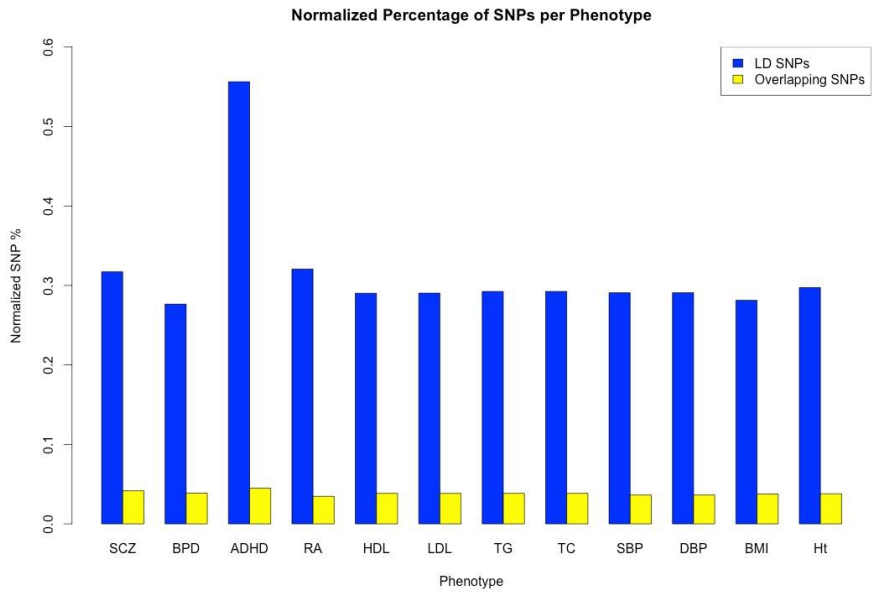
Supplemental Figure S1: **Enrichment plots for common SNPs shared by all 12 GWAS**

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 ~2.4 million SNPs that was determined by intersecting the SNP lists across all twelve GWAS including all SNPs from the ADHD GWAS (~1.2 million SNPs). We find that the number of SNPs does not influence our results for schizophrenia (SCZ), as maximum enrichment is seen for schizophrenia even when the common set of SNPs is used.



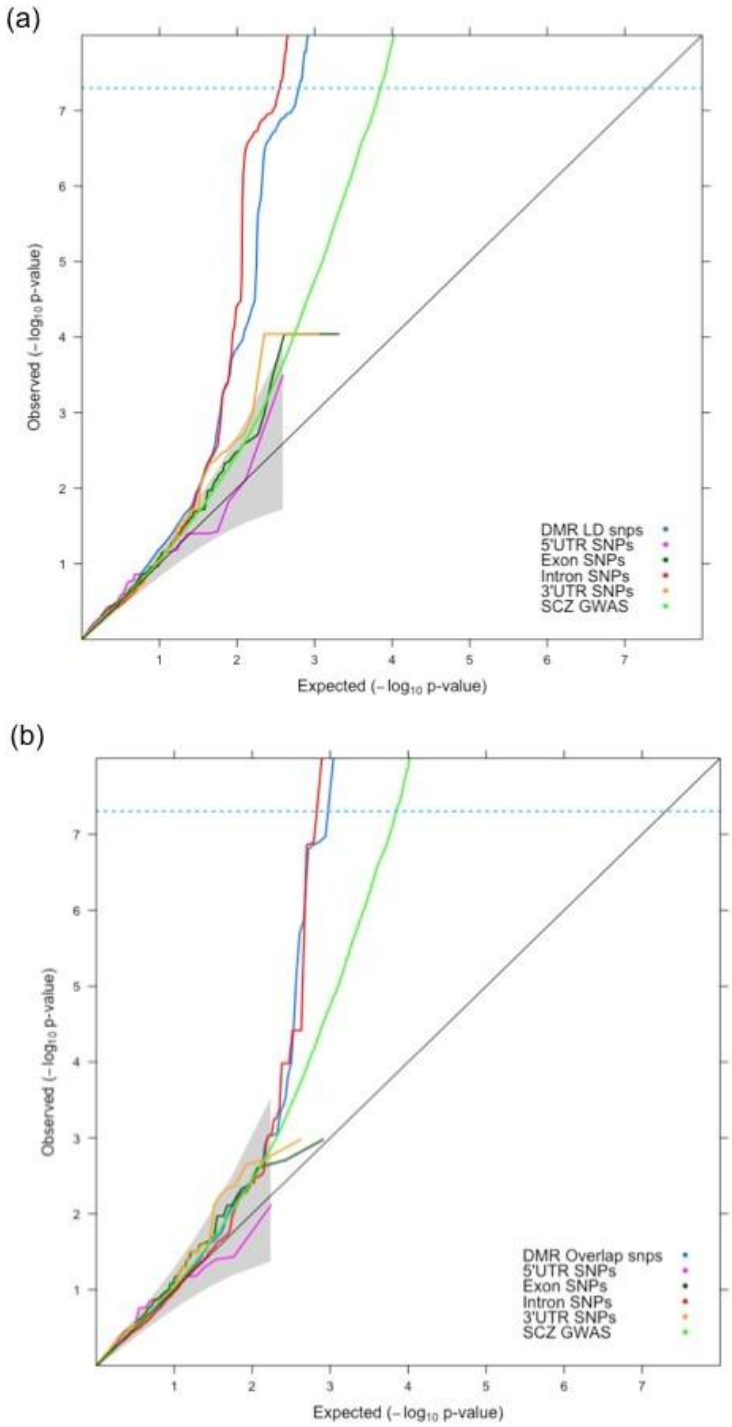
Supplemental Figure S2: Overlap of SNPs between various evolutionary annotations

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.



Supplemental 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.



Supplemental Figure S4: Stratified QQ Plots for SCZ SNPs in DMRs

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.

Supplemental Tables

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 *et al* (2011)	938	6711
Attention Deficit Hyperactivity Disorder	5,415	1,206,332	Neale *et al* (2010)	542	3970
Rheumatoid Arthritis	41,282	2,553,357	Stahl *et al* (2010)	884	8187
High Density Lipoprotein	99,900	2,620,435	Teslovich *et al* (2010)	1006	7603
Low Density Lipoprotein	95,454	2,620,568	Teslovich *et al* (2010)	1006	7608
Triglycerides	96,598	2,620,567	Teslovich *et al* (2010)	1006	7663
Total Cholesterol	100,184	2,620,450	Teslovich *et al* (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 *et al* (2015)	902	7364
Height	253,288	2,545,021	Wood *et al* (2011)	964	7567

Supplemental Table S1: Summary of GWAS and DMR SNPs

For each GWAS, the table shows the sample size, marker density, the reference to the specific study, the number of SNPs located within DMRs, and the numbed of SNPs in LD with DMRs.

Top Canonical Pathways	*P*-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
Top Physiological System Development and Function	
Nervous System Development and Function	4.41E-02 - 2.23E-02
Tissue Morphology	2.23E-02 - 2.23E-02

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

Top Canonical Pathways	*P*-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
Top Physiological System Development and Function	
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

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

Top Canonical Pathways	*P*-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
Top Physiological System Development & Function	
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

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

Top Canonical Pathways	*P*-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
Top Physiological System Development & Function	
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

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

Supplemental References

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

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

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

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

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

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

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

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