*Manuscript
Click here to view linked References

Human-specific methylated regions are enriched in schizophrenia

Niladri Banerjee[1,2], Tatiana Polushina PhD[1,2], Francesco Bettella PhD[3,4], Sudheer Giddaluru PhD[1,2], Vidar M. Steen MD PhD[1,2], Ole A. Andreassen MD PhD[3,4], Stephanie Le Hellard PhD[1,2]

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

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

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

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

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

Telephone: +47-55 97 53 37 **Fax:** +47-55 97 54 79 **Email:** stephanie.hellard@uib.no

Running Title: Human DMRs are enriched in schizophrenia

Key Words: differentially methylated regions; schizophrenia; evolution; epigenetics; Neanderthal Selective Sweep score; Human Accelerated Regions.

Number of words in Abstract: 250
Number of words in Article Body: 3472
Number of Figures: 4
Number of Tables: 0
Number of Supplemental Files: 1

Abstract

Background: Despite the reduced fertility of patients, schizophrenia has a prevalence of ~1% in the general population. One explanation for this persistence is that schizophrenia is a by-product of human evolution. This hypothesis is supported by evidence suggesting that recently-evolved genomic regions in humans are involved in the genetic risk for schizophrenia. We tested if genetic variants located within human-specific differentially methylated regions (DMRs) are enriched for association with schizophrenia.

Methods: We used analytical tools that evaluate the polygenic enrichment of a set of genomic variants against all genomic variants. SNPs identified in genome-wide association studies (GWAS) of schizophrenia and 11 other phenotypes were classified according to their location in previously defined human-specific DMRs. These DMR SNPs were then tested for enrichment of association with the GWAS trait.

Results: Schizophrenia was the only trait in which GWAS SNPs within human-specific DMRs 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.

Conclusions: Regions where DNA methylation modifications have changed during recent human evolution harbor more schizophrenia-associated genetic variants than expected. 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. Our results also suggest that interaction with the environment might have contributed to that association.

Introduction

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 (7–10), and its persistence despite this heavy burden is paradoxical. One explanation 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 (9,11). Crow first argued that language and psychosis may have common origins, which could explain the persistence of schizophrenia in human populations (11,12). This evolutionary hypothesis can now be tested, thanks to the identification of genetic factors implicated in schizophrenia (13–15) and the availability of datasets that reflect recent genomic evolution in humans (16–18).

Large genome-wide association studies (GWAS) have identified thousands of variants that are associated with schizophrenia (13–15) 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 evolutionarily relevance of certain genomic regions (19). This field has introduced new datasets to test disease origins in humans, including Human Accelerated Regions (HARs) and Neanderthal Selective Sweep (NSS) scores (17,18). HARs are genomic regions that are highly conserved in non-human species, but have undergone rapid sequence change in the human lineage (19–23). Xu et al (17) showed that genes near HARs are enriched for association with schizophrenia. Neanderthals were hominids that co-existed and even bred with modern humans (24,25). Comparison of Neanderthal and human genome sequences (26,27) has revealed genomic regions that have experienced a selective sweep in modern humans, presumably following a favorable mutation (27). 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 (18) 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 (28) 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 (28). 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 and hence an insight into how the environment could interact in different ways with the genome in the three hominids. We thus examined if these evolutionary DMRs are enriched for association with schizophrenia or other human traits. Using previously published methodologies (18,29,30) 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.

Methods and Materials

Differentially Methylated Region data

Coordinates for DMRs were obtained from data publicly available in Supplemental Table S2 of Gokhman et al, 2014 (28). This file contained DMRs inferred by comparing fossilized Neanderthal and Denisovan limb samples with osteoblasts from modern humans. DMRs are classified according to the hominid in which the change occurred, i.e. human-specific, Neanderthal-specific and Denisovan-specific DMRs. DMRs that could not be classified reliably (unclassified DMRs)(28) were not used. Full methodological details for assigning DMRs are in the Supplementary File of the original paper (28).

HAR data

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. 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 (18). Markers with negative values, indicating positive selection in humans, were filtered out and used for analysis.

GWAS data

Summary statistics (snpID, chr posn, bp, *p*-value) from GWAS of 12 common traits were obtained from published datasets: schizophrenia (SCZ) (13), bipolar disorder (BPD) (31), attention deficit

hyperactivity disorder (ADHD) (32), rheumatoid arthritis (RA) (33), blood lipid markers (high density lipoprotein (HDL), low density lipoprotein (LDL), triglycerides (TG), total cholesterol (TC)) (34), blood pressure (systolic blood pressure (SBP), diastolic blood pressure (DBP)) (35), body mass index (BMI) (36), and height (37). 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). For BMI and height SNPs, the genomic coordinates (chr posn, bp) were obtained by mapping them to the assembly of 1000 Genomes Project Phase 1 reference panel SNPs (38).

SNP assignment to DMRs

SNPs were assigned to DMRs with LDsnpR (39) using positional binning and LD (linkage disequilibrium)-based binning. 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 (39) (Supplemental Table S1). The LD file utilized in HDF5 format was constructed on the European reference population of 1KGP and can be publicly downloaded at: http://services.cbu.uib.no/software/ldsnpr/Download.

Quantile-Quantile (QQ) Plots

QQ plots are an effective tool to visualize the spread of data and any deviations from expected null distributions. They are frequently utilized 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 deviation from this null line. In GWASs, due to the extremely low *p*-values, 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 (29).

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 (30). 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 in ref 17, with the extended MHC region (chr6:25-35Mb) masked and SNPs with MAF <0.05 excluded. To assess enrichment for SNPs with different disease significance thresholds, we generated a range of LD-implicated regions through LD clumping in PLINK for index SNPs with *p*-values from $1x10^{-2}$ to $1x10^{-8}$. 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 (30). INRICH merges overlapping genes and overlapping intervals to prevent potentially inflated results due to multi-counting of the same genes/intervals. A total of 7252, 2510, 1015, 445, 207, 108 and 68 intervals were analyzed respectively for SNPs with *p*-values from $1x10^{-2}$ to $1x10^{-8}$. Enrichment was determined for genes located within 100kb of human DMRs, HARs, PARs and pHARs as described previously (17) and genes in LD blocks containing NSS markers (which are single-base markers, unlike HARs and DMRs). Since INRICH tests for enrichment of genes, we used an LD block with a strict threshold $r^2 \geq 0.8$ to include potential genes of interest near NSS markers. GENCODE v19 gene database (last accessed 5th February 2016) was used to map the genes to DMRs, NSS markers and HARs.

Pathway analysis

Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) from QIAGEN (www.qiagen.com/ingenuity, last accessed 26th August 2016). The DMR SNPs driving the enrichment were determined from QQ plots using the method of Schork et al (29). These SNPs were then mapped to genes using LDsnpR (39). 5338 enriched SNPs were mapped to 349 unique overlapping RefSeq genes and 446 RefSeq genes in LD. Genes in LD blocks containing NSS markers were determined in a similar manner. 4276 enriched SNPs mapped to 648 overlapping RefSeq genes and 1363 RefSeq genes in LD. IPA was run on all these gene sets separately. 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.

Results

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

We determined enrichment using QQ plots as described by Schork et al (29). 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. We tested the markers located in the human-specific DMRs for enrichment of association with schizophrenia (13), bipolar disorder (BPD) (31), attention deficit hyperactivity disorder (ADHD) (32), rheumatoid arthritis (33), high density lipoprotein (34), low density lipoprotein (34), triglycerides (34), total cholesterol (34), systolic blood pressure (35), diastolic blood pressure (35), body mass index (36), and height (37). The GWAS datasets are summarised in Supplemental Table S1. For each trait, we extracted the SNPs that occur in DMRs and compared them with the full list of SNPs. We generated a list of SNPs within DMRs (positional annotation) and a list of SNPs in linkage disequilibrium (LD-based annotation) with markers within DMRs (Supplemental Table S1).

For the schizophrenia GWAS, enrichment was observed for both SNPs in LD with markers in DMRs (Figure 1, Supplemental Figure S1) and for SNPs located within DMRs (Figure 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; Figure 1, Supplemental Figure S1), the observed enrichment only crosses the genome-wide significance level of 5×10^{-8} for the schizophrenia SNPs. The enrichment of disease-associated markers in DMRs is thus specific to schizophrenia and is independent of LD.

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

The MHC harbors several significant schizophrenia markers and could potentially bias our results because of long-range LD. We therefore repeated the analysis with the MHC region masked and found that the enrichment remains (Figure 1).

The schizophrenia GWAS had the highest density of markers genotyped (~9.4 million) and thus had the most SNPs in DMR regions (Supplemental Table S1), which could artificially inflate the enrichment. We normalized the total number of SNPs within DMRs 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 (Supplemental 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, indicating the presence of significant disease markers in DMRs. These validations point to a true enrichment of association of the human DMRs 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 (Supplemental Figure S1).

Additionally, we considered the distribution of significant SNPs based on genomic annotations of 5'UTRs, Exons, Introns and 3'UTRs (29). Contrary to previously published findings (29), the enrichment was highest for intronic SNPs and lowest for 5'UTR SNPs (Supplemental Figure S4). This makes sense biologically as one would not expect open chromatin regions, containing exons and UTRs, to possess methylation marks.

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. No enrichment was observed (Figure 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 (17,18). We first compared the enrichment via QQ plots. We 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 (Figure 3).

To ensure that we were not comparing the same groups of markers across the different evolutionary annotations we analyzed how much overlap exists between the various sets. Reassuringly, the various evolutionary annotations do not share the same group of markers, indicating that we did not test the same regions or SNPs (Supplemental 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.

Statistically-significant enrichment exists for Human DMRs

To determine the statistical significance of the DMR enrichment in schizophrenia, we utilized the INRICH software pipeline. This performs permutation and bootstrapping procedures to determine the significance of association of markers in LD intervals while maintaining SNP density and gene density of the original intervals (30). We utilized a similar procedure as reported by Xu et al (17) and tested LD intervals for SNPs with nominal *p*-values ranging from $1x10^{-2}$ to $1x10^{-8}$. In their analysis, Xu et al had 893 genes within 100kb of pHARs, 326 genes within 100kb of mHARs (regions conserved in all mammals which are accelerated in humans) and 305 genes within 100kb of PARs. In our study, using GENCODE V19, we had 3700, 1316 and 1268 genes within 100kb of pHARs, mHARs and PARs respectively. INRICH confirmed significant ($p<0.05$) enrichment of association for human DMRs with schizophrenia at all *p*-value thresholds of LD intervals. After

correcting for multiple testing through bootstrapping, most nominal *p*-value LD intervals maintained modest enrichment ($p<0.1$) for human DMRs. Additionally, INRICH independently verified the previously reported enrichment of NSS markers with schizophrenia (18) (Figure 4).

Pathway Analysis

We utilized Ingenuity Pathway Analysis (IPA) to markers that show enrichment of association with schizophrenia and found that 'CREB signalling in neurons' was also amongst the top canonical pathways (Supplemental Table S4). When we repeated the analyses with all organ systems, 'CREB signalling in neurons' and 'Synaptic long term potentiation' emerged amongst the top canonical pathways for genes in LD with enriched DMR SNPs (Supplemental Table S3) and for genes in LD with enriched NSS markers (Supplemental Table S5). determine which biological pathways are overrepresented by genes in LD with DMR SNPs that are enriched for association with schizophrenia. When analyzing pathways overrepresented in the nervous system, we find 'CREB Signalling in Neurons' and 'Synaptic Long Term Potentiation' amongst the top canonical pathways. Additionally, under physiological systems, 'Nervous system development and function' is enriched (Supplemental Table S2). We repeated the same analysis for genes in LD with NSS This is interesting since there is very little overlap between the DMR and NSS SNPs (Supplemental Figure S2). Considering all organs, not just the nervous system, genes in LD with enriched DMR SNPs are also overrepresented in 'Hair & Skin development' (Supplemental Table S3). This may suggest potential gene by environment interactions, modulated by methylation variation during human evolution.

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. Amongst all the traits analyzed, the enrichment observed in QQ plots was strongest for schizophrenia and passed the genome-wide significance threshold of $5x10^{-8}$ (Figure 1). INRICH analysis demonstrated significant enrichment ($p<0.05$) in human DMRs for all LD intervals tested (Figure 4a). Our results suggest that DMRs are enriched for genetic variation that confers risk of developing schizophrenia.

Neanderthal- or Denisovan-specific DMRs showed no enrichment of association (Figure 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 the 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.

Xu et al (17) and Srinivasan et al (18) 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 (18) (Figure 3, Figure 4). QQ plots for HARs do not show enrichment of disease markers unlike NSS markers and DMRs (Figure 3). The INRICH analysis showed enrichment for primate HARs only at LD intervals at a *p*-value threshold of $1x10^{-2}$ with borderline ($p>0.1$) enrichment at higher thresholds of LD intervals. 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 (13), a step we could not replicate as the HWE values 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 (13). Interestingly, all the evolutionary annotations (DMRs, NSS markers and HARs) cover different sections of the genome with very little overlap between them (Supplemental Figure S2). Between the three evolutionary annotations, nearly 70,000 SNPs occur around regions with evolutionary significance (Supplemental Figure S2). This provides 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).

Although we cannot rule out enrichment in the other phenotypes tested, schizophrenia was the only trait for which the enrichment of DMR-localized disease markers exceeded genome-wide significance (Supplemental Figure S1). Furthermore, while the DMRs utilized here were obtained from bone samples, the authors of the original study (28) referred to the report by Hernando-Herraez et al (40) which demonstrated 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 (41–43). 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 (28) 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 (40) 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 Signalling in Neurons' and 'Synaptic Long Term Potentiation'.

In summary, we have demonstrated that human genomic regions whose methylation status was altered during evolution are enriched for 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.

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 http://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. This work was supported by the Research Council of Norway (NFR; NORMENT-Centre of Excellence [#223273, #213837, #251134]) and the KG Jebsen Foundation (SKGJ-MED-008).

Financial Disclosure

The authors report no biomedical financial interests of potential conflicts of interest.

References

1. Shih RA, Belmonte PL, Zandi PP (2004): A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry* 16: 260–283.

2. Jeste DV, Del Carmen R, Lohr JB, Wyatt RJ (1985): Did schizophrenia exist before the eighteenth century? *Compr Psychiatry* 26: 493–503.

3. Skre I, Onstad S, Torgersen S, Lygren S, Kringlen E (1993): A twin study of DSM-III-R anxiety disorders. *Acta Psychiatr Scand* 88: 85–92.

4. Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, *et al*. (1999): Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* 56: 162–168.

5. Sullivan PF, Kendler KS, Neale MC (2003): Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 60: 1187–1192.

6. Lichtenstein P, Yip BH, Björk C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009): Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 373: 234–239.

7. Larson CA, Nyman GE (1973): Differential fertility in schizophrenia. *Acta Psychiatr Scand* 49: 272–280.

8. Bassett AS, Bury A, Hodgkinson KA, Honer WG (1996): Reproductive fitness in familial schizophrenia. *Schizophr Res* 21: 151–160.

9. Nichols C (2009): Is there an evolutionary advantage of schizophrenia? *Pers Individ Diff* 46: 832–838.

10. Brown S (1997): Excess mortality of schizophrenia. A meta-analysis. *Br J Psychiatry* 171: 502–508.

11. Crow TJ (1995): A Darwinian approach to the origins of psychosis. *Br J Psychiatry* 167: 12–25.

12. Crow TJ (1997): Is schizophrenia the price that Homo sapiens pays for language? *Schizophr Res* 28: 127–141.

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

14. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, *et al*. (2013): Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet* 45: 1150–1159.

15. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2011): Genome-wide association study identifies five new schizophrenia loci. *Nature Genet* 43: 969–976.

16. Crespi B, Summers K, Dorus S (2007): Adaptive evolution of genes underlying schizophrenia. *Proc Roy Soc Lond [Biol]* 274: 2801–2810.

17. Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT (2015): Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Mol Biol Evol* 32: 1148–1160.

18. Srinivasan S, Bettella F, Mattingsdal M, Wang Y, Witoelar A, Schork AJ, *et al*. (2015): Genetic markers of human evolution are enriched in schizophrenia. *Biol Psychiatry* 80: 284–292.

19. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, *et al*. (2006): Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2: e168.

20. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, *et al*. (2011): A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478: 476-482.

21. Bush EC, Lahn BT (2008): A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol Biol* 8: 17.

22. Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, *et al*. (2007): Fast-evolving noncoding sequences in the human genome. *Genome Biol* 8: R118.

23. Prabhakar S, Noonan JP, Pääbo S, Rubin EM (2006): Accelerated evolution of conserved noncoding sequences in humans. *Science* 314: 786.

24. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, *et al*. (2014): The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507: 354–357.

25. Vernot B, Akey JM (2014): Resurrecting surviving Neandertal lineages from modern human genomes. *Science* 343: 1017–1021.

26. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, *et al*. (2014): The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505: 43–49.

27. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, *et al*. (2010): A draft sequence of the Neandertal genome. *Science* 328: 710–722.

28. Gokhman D, Lavi E, Prüfer K, Fraga MF, Riancho JA, Kelso J, *et al*. (2014): Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* 344: 523–527.

29. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, *et al*. (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.

30. Lee PH, O'Dushlaine C, Thomas B, Purcell SM (2012): INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 28: 1797–1799.

31. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, *et al*. (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genet* 43: 977–983.

32. Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K-P, *et al*. (2010): Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 49: 884–897.

33. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, *et al*. (2010): Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet* 42: 508–514.

34. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, *et al*. (2010): Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707–713.

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

36. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, *et al*. (2015): Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518: 197–206.

37. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, *et al*. (2014): Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genet* 46: 1173–1186.

38. The 1000 Genomes Project Consortium (2012): An integrated map of genetic variation from 1,092 human genomes. 491: 56–65.

39. Christoforou A, Dondrup M, Mattingsdal M, Mattheisen M, Giddaluru S, Nöthen MM, *et al*. (2012): Linkage-disequilibrium-based binning affects the interpretation of GWASs. *Am J Hum Genet* 90: 727–733.

40. Hernando-Herraez I, Prado-Martinez J, Garg P, Fernandez-Callejo M, Heyn H, Hvilsom C, *et al*. (2013): Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* 9: e1003763.

41. Mendizabal I, Keller T, Zeng J, Soojin VY (2014): Epigenetics and evolution. *Integr Comp Biol* 54: 31–42.

42. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD (2011): Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146: 1029–1041.

43. Hernando-Herraez I, Heyn H, Fernandez-Callejo M, Vidal E, Fernandez-Bellon H, Prado-Martinez J, *et al*. (2015): The interplay between DNA methylation and sequence divergence in recent human evolution. *Nucleic Acids Res* 43: 8204–8214.

Figure legends

Figure 1: DMR enrichment across SCZ, BPD, BMI and Height

Quantile-Quantile (QQ) plots of GWAS SNPs for Schizophrenia (SCZ) with the extended MHC region masked (chr6: 25-35Mb), 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.

Figure 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.

Figure 3: Comparison of enrichment of association with schizophrenia for SNPs in LD with various evolutionary annotations

QQ plots for association with schizophrenia (SCZ) of SNPs in different evolutionary datasets (DMRs - red, NSS - orange, Primate HARs (pHARs) - blue, HARs - magenta, PARs - dark green) versus schizophrenia GWAS with all SNPs (light green). SNPs are corrected for genomic inflation using global lambda.

Figure 4: INRICH test for enrichment of association of DMR, NSS and Accelerated Region gene sets

(a) Empirical *p*-values obtained from the first round of 10000 replicates. **(b)** Corrected *p*-values based on bootstrapping 5000 samples. The same color gradation is used in both heatmaps to allow comparison with p=0.1 as threshold. The various evolutionary annotations compared are: 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.

Figure 1

Click here to download high resolution image

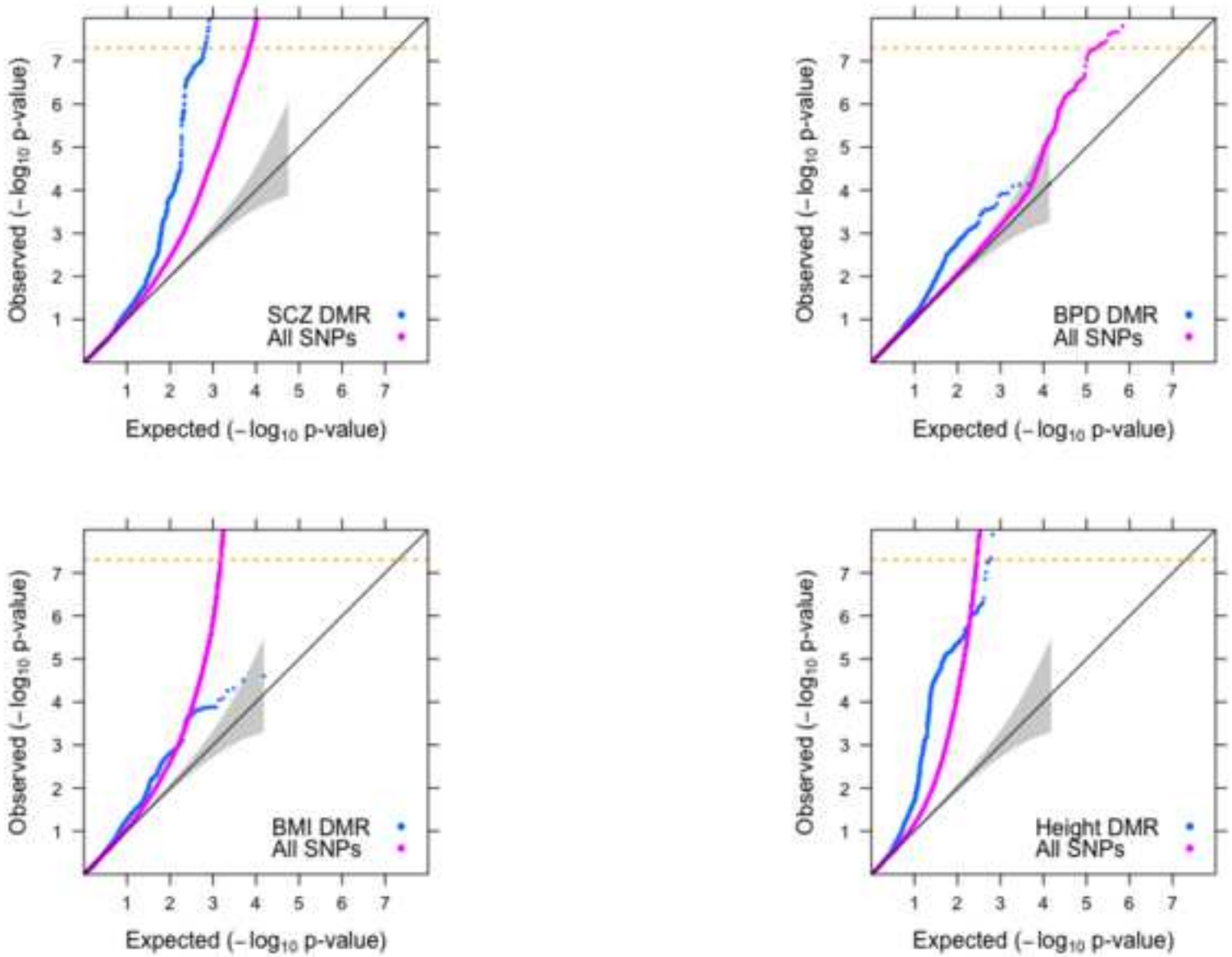


Figure 2

Click here to download high resolution image

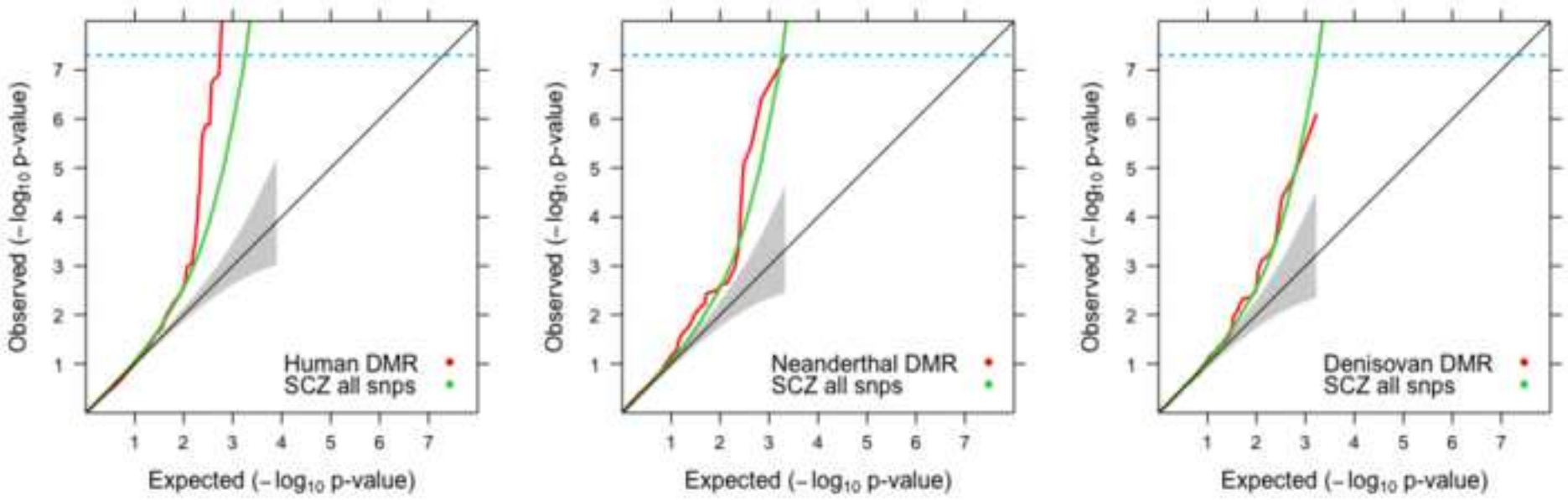


Figure 3

Click here to download high resolution image

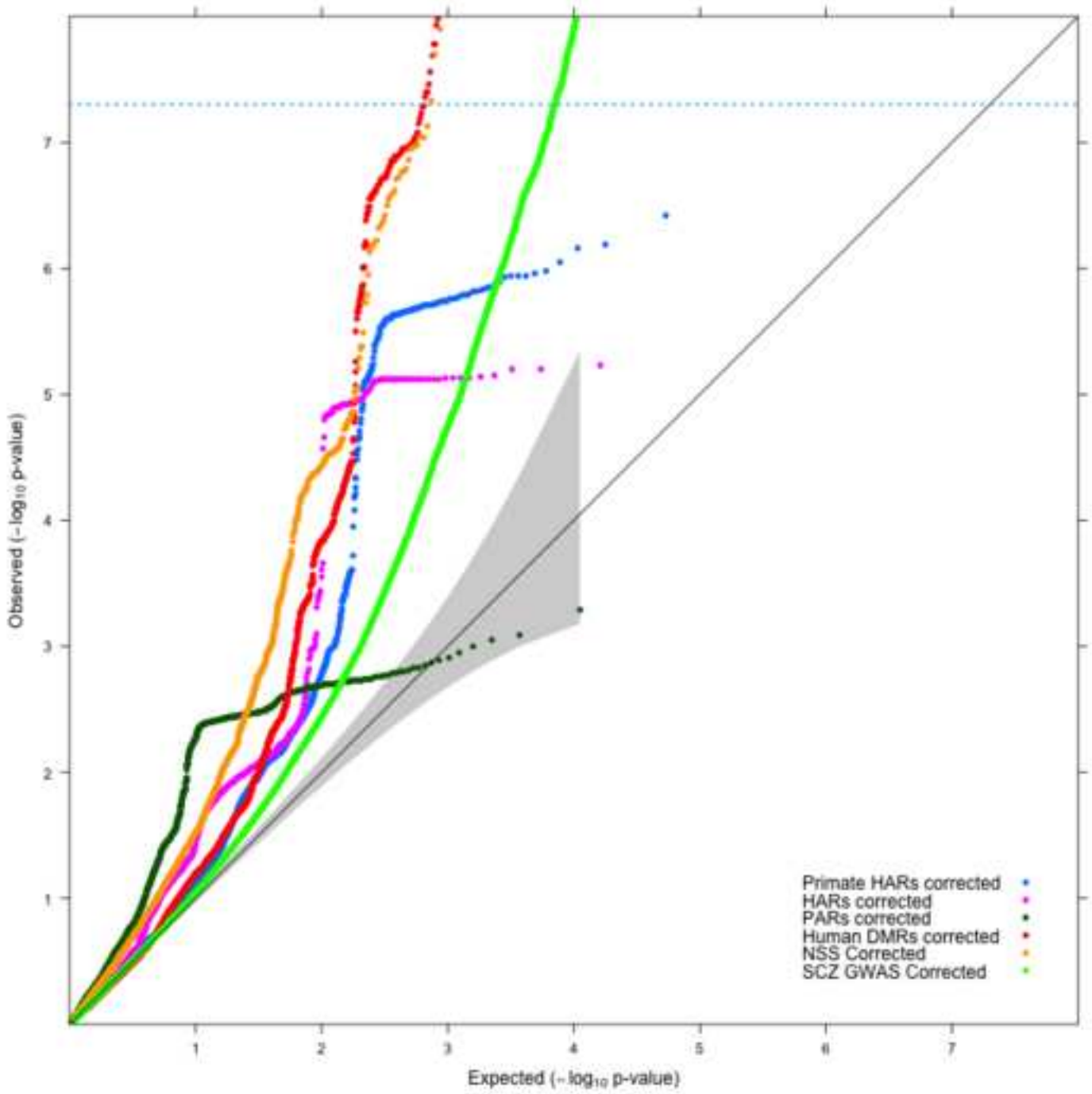
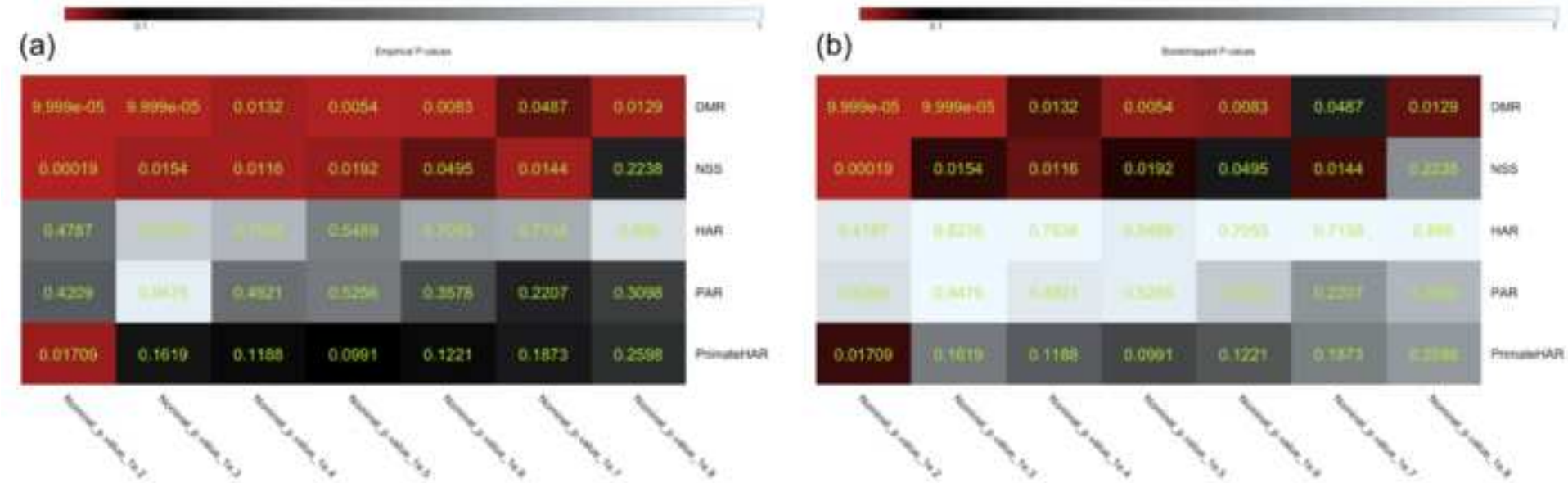


Figure 4

Click here to download high resolution image