

# UC San Diego

## UC San Diego Previously Published Works

### Title

Probing the Association between Early Evolutionary Markers and Schizophrenia

### Permalink

<https://escholarship.org/uc/item/6qb15983>

### Journal

PLOS ONE, 12(1)

### ISSN

1932-6203

### Authors

Srinivasan, Saurabh  
Bettella, Francesco  
Hassani, Sahar  
[et al.](#)

### Publication Date

2017

### DOI

10.1371/journal.pone.0169227

Peer reviewed

RESEARCH ARTICLE

# Probing the Association between Early Evolutionary Markers and Schizophrenia

Saurabh Srinivasan<sup>1,2</sup>, Francesco Bettella<sup>1,2</sup>, Sahar Hassani<sup>1,2</sup>, Yunpeng Wang<sup>1,2</sup>, Aree Witoelar<sup>1,2</sup>, Andrew J. Schork<sup>3,4,5</sup>, Wesley K. Thompson<sup>6,7</sup>, David A. Collier<sup>8</sup>, Rahul S. Desikan<sup>9</sup>, Ingrid Melle<sup>1,2</sup>, Anders M. Dale<sup>3,5,10,11</sup>, Srdjan Djurovic<sup>12,13</sup>, Ole A. Andreassen<sup>1,2,6\*</sup>

**1** NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, **2** Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, **3** Multimodal Imaging Laboratory, University of California at San Diego, La Jolla, CA, United States of America, **4** Cognitive Sciences Graduate Program, University of California, San Diego, La Jolla, CA, United States of America, **5** Center for Human Development, University of California at San Diego, La Jolla, CA, United States of America, **6** Institute of Biological Psychiatry, Mental Health Center St. Hans, Mental Health Services Copenhagen, Roskilde, Denmark, **7** The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark, **8** Eli Lilly & Co, Erl Wood Manor, Windlesham, Surrey, United Kingdom, **9** Neuroradiology Section, Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, United States of America, **10** Department of Neuroscience, University of California at San Diego, La Jolla, CA, United States of America, **11** Neuroradiology Section, Department of Radiology and Biomedical Imaging, University of California at San Francisco, San Francisco, CA, United States of America, **12** Department of Medical Genetics, Oslo University Hospital, Oslo, Norway, **13** NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway

\* [o.a.andreassen@medisin.uio.no](mailto:o.a.andreassen@medisin.uio.no)



**OPEN ACCESS**

**Citation:** Srinivasan S, Bettella F, Hassani S, Wang Y, Witoelar A, Schork AJ, et al. (2017) Probing the Association between Early Evolutionary Markers and Schizophrenia. PLoS ONE 12(1): e0169227. doi:10.1371/journal.pone.0169227

**Editor:** Peter John McKenna, SPAIN

**Received:** September 9, 2016

**Accepted:** December 13, 2016

**Published:** January 12, 2017

**Copyright:** © 2017 Srinivasan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data for psychiatric GWASs can be found at the psychiatric genomic consortium website <https://www.med.unc.edu/pgc/results-and-downloads>. The data for cardiovascular disease risk factors GWASs can be found at <http://csg.sph.umich.edu/abecasis/public/lipids2013/>. The data for anthropometric measurement GWASs can be found at [http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files). The data for immune mediated GWASs can be found at the International Inflammatory Bowel Disease Genetics Consortium website <https://www.ibdgenetics.org/downloads.html>. These are links to third party

## Abstract

Schizophrenia is suggested to be a by-product of the evolution in humans, a compromise for our language, creative thinking and cognitive abilities, and thus, essentially, a human disorder. The time of its origin during the course of human evolution remains unclear. Here we investigate several markers of early human evolution and their relationship to the genetic risk of schizophrenia. We tested the schizophrenia evolutionary hypothesis by analyzing genome-wide association studies of schizophrenia and other human phenotypes in a statistical framework suited for polygenic architectures. We analyzed evolutionary proxy measures: human accelerated regions, segmental duplications, and ohnologs, representing various time periods of human evolution for overlap with the human genomic loci associated with schizophrenia. Polygenic enrichment plots suggest a higher prevalence of schizophrenia associations in human accelerated regions, segmental duplications and ohnologs. However, the enrichment is mostly accounted for by linkage disequilibrium, especially with functional elements like introns and untranslated regions. Our results did not provide clear evidence that markers of early human evolution are more likely associated with schizophrenia. While SNPs associated with schizophrenia are enriched in HAR, Ohno and SD regions, the enrichment seems to be mediated by affiliation to known genomic enrichment categories. Taken together with previous results, these findings suggest that schizophrenia risk may have mainly developed more recently in human evolution.

websites and, to best of our knowledge, will remain accessible to any researcher interested in using the data in the future.

**Funding:** This study was supported by the Research Council of Norway (<http://www.forskingsradet.no/no/Forsiden/1173185591033>) (#213837, #223273, #225989) South-East Norway Health Authority (<http://www.helse-sorost.no/>) (#2013-123) and KG Jebsen Stiftelsen (<http://www.stiftkj.no/772/?lang=en>) (#SKGJ-Med-008). Eli Lilly & Co. UK provided support in the form of salary for the author [DAC], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript and has no commercial interests in this study. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing Interests:** David A. Collier is employed by Eli Lilly & Co. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

## Introduction

Schizophrenia has affected humans throughout history; it has heritability between 60–80% [1] and a global prevalence of around 1% [2, 3]. It is characterized by hallucinations and delusions, often involving language and thought disorders, and higher order cognitive dysfunction [4]. Hence, schizophrenia has been suggested to represent, in part, a by-product of adaptive changes during the hominization process [3, 5]. Archaeological and paleontological findings may provide us with evidence of the cultural and anatomical changes in skeletal structure, however, mental and psychiatric changes are hard to trace.

Humans differ from their ancestors on a range of skills including language, creativity, meta-cognition, executive function, and cooperation. [6, 7] These are important characteristics that involved genetic changes leading to functional advantages, with cultural and societal effects. These changes which set us apart from our ancestors could have also made us vulnerable to psychiatric disorders like schizophrenia, plausibly a human-specific disease. [5, 8]

It is still unknown at which stage of human evolution the risk factors for schizophrenia emerged. We have previously found evidence that genetic risk appeared during the divergence of modern *Homo sapiens* from *Homo Neanderthalensis*. [9] As every stage of evolution was driven by genetic changes, it is possible that gene variants associated with schizophrenia emerged even earlier, for example, when new world and old world monkeys split into different branches of the evolutionary tree [10], or even as early as when vertebrates appeared on earth. [11] Building on the recent progress in genome-wide association studies (GWAS) [12], we now have the opportunity to investigate the origin of psychosis further back in the evolution. The human genome consists of evolutionarily new and ancient regions. By investigating single nucleotide polymorphisms (SNPs) association with schizophrenia in these regions, it is possible to roughly estimate when the risk potential appeared during evolution.

Here, we investigated schizophrenia SNPs located in human accelerated regions, segmental duplications and ohnologs to determine to what extent SNPs tagging these regions, which are proxies for various periods in early evolutionary history, are associated with schizophrenia or other human traits and diseases. These represent early evolutionary markers older than 200,000 years when modern humans are regarded to have first appeared on earth. [13] Human accelerated regions (HAR) are DNA sequences that experienced rapid changes after the divergence of humans (including our hominin ancestors) from chimpanzees after remaining constant throughout primate evolution. [14] While most HARs are exclusively non-coding sequences, research suggests that many HARs are developmental gene regulatory elements and RNA genes, most of which evolved their uniquely human mutations through positive selection before divergence of archaic hominins and diversification of modern humans [15]. These regions harbor genes that have been shown to play important roles in neurocognitive development [14, 16]. It is known that various genomic regions show differential enrichment, [17] i.e. that certain genomic regions are more likely to harbor gene variants associated with human traits and diseases. Further, an enrichment of schizophrenia risk loci could be due to a general effect of brain related genes, [12] not necessarily of those implicated in evolution. Moreover, according to recent GWAS, gene variants at the major histocompatibility complex (MHC) region on human chromosome 6p22.1 are implicated in schizophrenia [12, 18, 19] and play a role in evolution. [20] Thus, it is important to disentangle the effect of these mediating factors from the effect of the evolutionary proxies.

Segmental duplications (SD), also known as low copy repeats, are a known source of genetic instability and evolution [10]. Most duplication in the hominid branch can be traced to events 35–40 million years ago [10], marking an early stage of hominid/primate evolution. Increased duplications are seen in regions that distinguish the hominid branch from other primates.

Eichler et al. [21] found that the primate ancestral branch leading to human and African great apes showed the most significant increase in duplication activity both in terms of base pairs and in terms of events. In light of the importance of SDs in contributing to copy-number changes associated with neurocognitive disease [22], it is suggested that this apparent acceleration had a profound impact on the reproductive success, adaptability and evolution of ancestral hominid populations [21].

Ohnologs (Ohno) are genes retained after whole genome duplication events, unlike segmental duplications which are duplications of smaller chunks of the genome. They are often over represented in copy number variations (CNVs) that cause complex neurodevelopmental disorders like schizophrenia, autism spectrum disorders, neurodevelopmental delay, intellectual disability and epilepsy [15, 23]. Whole genome duplications events are said to have occurred early in the vertebrate lineage around 500 million years ago [24] and the human genome contains many more duplications than would be expected by chance. These are evolutionarily important since gene duplication and divergence are the primary source of new genes in eukaryotes [25, 26].

We used a statistical framework suited for polygenic architectures [17, 27]. These methods have been applied, with success, to GWAS of complex human phenotypes to probe the overlap between phenotypes association and Neanderthal selective sweep [9]. Here, we use them to investigate if SNPs in regions of early human evolution are enriched of association with schizophrenia, while controlling for potential confounders.

## Methods and Materials

### Samples

We obtained summary statistics for single nucleotide polymorphisms (SNPs) from genome-wide association studies (GWAS) of schizophrenia (conducted by the Psychiatric Genomics Consortium (PGC)) [12] and other phenotypes representing a selection of morphological, cardiovascular, immunological and psychiatric phenotypes. They include anthropometric measures (body mass index (BMI) [28], height) [29], a cardiovascular disease risk factor (triglycerides (TG)) [30], an immune-mediated disease (Crohn's disease (CD)) [31] a neurological disorder (Alzheimer's disease (AD)) [32] as well as another psychiatric disorder (bipolar disorder (BD)) [33] (S1 Table). In order to make the data comparable all summary statistics were aligned to a set of about 2.5 million variants.

### Evolutionary proxies

We computed LD weighted HAR, SD and Ohno regional affiliation scores following the procedure detailed below.

**Human accelerated region (HAR) score.** The HAR score indicates a SNP's affiliation to HAR. All SNPs were first assigned a raw HAR indicator value of 1 or 0 according to whether they fell inside or outside any HAR, and subsequently an LD-weighted score. The list of HAR was obtained from <http://www.broadinstitute.org/scientific-community/science/projects/mammals-models/29-mammals-project-supplementary-info>.

**Human segmental duplication (SD) score.** The human SD score indicates a SNP's affiliation to SD regions in humans. All SNPs were first assigned a raw score of 1 or 0 and subsequently an LD-weighted score. The lists of segmental duplications were downloaded from the SD database at <http://humanparalogy.gs.washington.edu/>.

**Ohnolog (Ohno) score.** This score indicates a SNP's affiliation to Ohno regions. All SNPs were first assigned a raw score of 1 or 0 and subsequently an LD-weighted score. The lists of Ohno regions was obtained from McLysaght et al. [25].

## Confounding/mediating factors

We investigated if the following factors could affect the evolutionary enrichment of schizophrenia associations.

**Brain genes.** We used the NCBI resource (<http://www.ncbi.nlm.nih.gov/gene>) to select all genes with any relation to the brain. We identified a total of 2494 genes by filtering specifically for genes in *Homo sapiens* matching the query phrase “human brain”. All 1000 Genomes Project SNPs in these genes were assigned a “Brain” value of 1, the rest were assigned a “Brain” value of 0. All SNPs were subsequently assigned LD-weighted “Brain” scores.

**Major histocompatibility complex (MHC).** The MHC has been implicated in schizophrenia as well as a number of other phenotypes, particularly immune-mediated diseases. The evolution of MHC itself may have involved SD and other large scale genetic variations [34]. It is therefore reasonable to expect that SNPs in these regions might be confounding some of the evolutionary enrichment results. To test for the effect of MHC, we removed the SNPs in the MHC region (chromosome 6 region between genomic positions 25652429 and 33421466 in the hg19 assembly) and repeated the analyses.

**Annotation of genomic regions (LD based).** The SNPs that fall within certain regions of interest may capture only a limited portion of the association signal actually ascribable to that region. We used an LD-weighted scoring algorithm [17] in order to identify SNPs that tag specific DNA regions even if they are not situated within them. For each SNP a pairwise correlation coefficient approximation to LD ( $r^2$ ) was extracted for all 1KGP SNPs within a 1,000,000 base pairs (1Mb). All  $r^2$  values  $< 0.2$  were set to 0 and each SNP was assigned an  $r^2$  value of 1.0 with itself. LD-weighted region annotation scores for all DNA regions of interest were computed as the sum of LD  $r^2$  between the tag SNP and all 1KGP SNPs in those regions. Given SNP<sub>*i*</sub>, its LD-weighted region annotation score was computed as  $\text{LDscore}_i = \sum_j (\delta_j r_{ij}^2)$ , where  $r_{ij}^2$  is the LD  $r$ -squared between SNP<sub>*i*</sub> and SNP<sub>*j*</sub> and  $\delta_j$  takes values of 1 or 0 depending on whether the 1KGP SNP<sub>*j*</sub> is within the region of interest or not.

LD scores were also assigned to exons, introns, 3'UTR and 5'UTR, and the total LD score (TotLD) was computed following the same procedure but extending the tagging region to the whole 1Mb window.

**Intergenic correction.** Intergenic SNPs are defined as having LD-weighted annotation scores for exon, intron, 3'UTR and 5'UTR equal to zero and being in LD with no SNPs in the 1KGP reference panel located within 100,000 base pairs of a protein coding gene, within a non-coding RNA, within a transcription factor binding site or within a miRNA binding site. Those singled out in this way are expected to form a collection of non-genic SNPs not belonging to any (annotated) functional elements within the genome (including through LD) and therefore represent a collection of likely null associations. Intergenic SNPs were used to estimate the inflation of GWAS summary statistics due to cryptic relatedness. We used intergenic SNPs because their relative depletion of associations (17) suggests they provide a set of reliably null SNPs that is less contaminated by polygenic effects. The inflation factor,  $\lambda_{GC}$ , was estimated as the median squared  $z$ -score of independent sets of intergenic SNPs across one hundred LD-pruning iterations, divided by the expected median of a chi-square distribution with one degree of freedom.

## Analytical approach

We employed a genetic enrichment method recently developed to uncover more of the genetic architecture of complex traits [17, 35–38]. Specifically, we investigated the enrichment of associations concurrent with the evolutionary affiliations in a covariate-modulated statistical framework [37]. We investigated whether SNPs located in the evolutionarily salient regions

(HAR, SD, Ohno) or tagging other SNPs therein, are more likely associated with schizophrenia or other phenotypes using GWAS data from existing non-censored summary statistics.

All statistical analyses were carried out with a covariate-modulated enrichment analysis package developed on R ([www.r-project.org](http://www.r-project.org)) and MATLAB ([www.mathworks.se/products/matlab/](http://www.mathworks.se/products/matlab/)) programming platforms.

**Fold enrichment plots.** To visually assess genetic enrichment, we used conditional fold enrichment plots [39]. For this purpose the covariate of interest, i.e. the region affiliation score, is used to subdivide SNPs into two strata. For LD-weighted annotation scores, the choice of a threshold score is somewhat arbitrary. We chose 1 since this is the score an isolated SNP within a salient region would have. It has been shown elsewhere [17] that the method is robust to the choice of threshold.

The enrichment plots were obtained by computing the empirical cumulative distribution of  $-\log_{10}(p)$ -values for SNP association with a given phenotype for all SNPs, and for the two dichotomous SNPs strata determined by the region affiliation score. Then each stratum's fold enrichment was calculated as the ratio  $CDF_{\text{stratum}}/CDF_{\text{all}}$  between the  $-\log_{10}(p)$  cumulative distribution for that stratum and the  $-\log_{10}(p)$  cumulative distribution for all SNPs. The nominal  $-\log_{10}(p)$  values are plotted on the x-axis, the fold enrichment in the y-axis. To assess polygenic effects below the standard GWAS significance threshold, we focused the fold enrichment plots on SNPs with nominal  $-\log_{10}(p) < 7.3$  (corresponding to  $p > 5 \times 10^{-8}$ ). Enrichment is present if the line corresponding to the SNPs of interest has a positive deflection from the horizontal line through 1.

**Partial least squares regression (PLSR).** The fold enrichment plots give a visual impression of the different association propensities of SNPs affiliated to the evolutionarily salient regions. However, they do not give a quantitative measure of the eventual enrichment. One such measure is provided by squared association z-scores regression [9] which in addition allows controlling for covariates of no interest. Due to their nature, the effects of LD-weighted annotation scores can't be estimated by standard linear regression.

PLSR is a supervised subspace regression method that maximizes covariance between two data blocks: the so called descriptor data set (X) and response data set (Y) [40]. An important aspect of PLSR is that the regression model is statistically stabilized for data sets with highly inter-correlated variables [41]. The predictive PLSR model may be written as follows:

$$Y = XB_A + F_A$$

Where X and Y are descriptor and response data matrices respectively (both data sets are mean-centered and scaled prior to data modelling),  $B_A$  are the regression coefficients for a model including A latent variables (LVs) and  $F_A$  is the residual matrix for the corresponding model.

For all of the PLSR models in this study, we chose the optimal number of LVs (A) as the number of LVs that explained more than 99% of the variance in the descriptor data set.

**Statistical validation: Jackknife approximate t-tests of regression coefficients.** In order to assess the contribution of the descriptor variables to the PLSR model we carried out approximate t-tests of regression coefficients based on jackknife variance estimates [42]. For this purpose, we ran 50-fold cross-validation on the SNP samples and re-calculated the regression coefficients in every cross-validation round. The LD-matrix was used for partitioning the SNP samples into the cross-validation subsets. We then calculated jackknife estimates for the standard deviations of the regression coefficients and, thereafter, t-statistics and approximate p-values indicating the significance of the association with the corresponding descriptor variables in the PLSR model.

**Squared z-scores residuals versus TotLD stratified scatter plots.** The PLSR results are better visualized by plotting descriptor and response variables directly. We residualized the squared z-scores in PLSR models deprived of TotLD, binned the latter and plotted the average squared z-scores for all bins against the corresponding TotLD bin centers. To control for the effect of the evolutionary measures we residualized the squared z-scores in a second series of PLSR models deprived of TotLD as well as the evolutionary measure of interest and again plotted the average squared z-scores for all bins against the corresponding TotLD bin centers.

## Results

We assessed the influence exerted on schizophrenia association propensity by affiliation to HAR, SD and Ohno. The fold enrichment plots (Fig 1) suggest enrichment of schizophrenia association among SNPs in HAR and SD regions, and to some extent, Ohno.

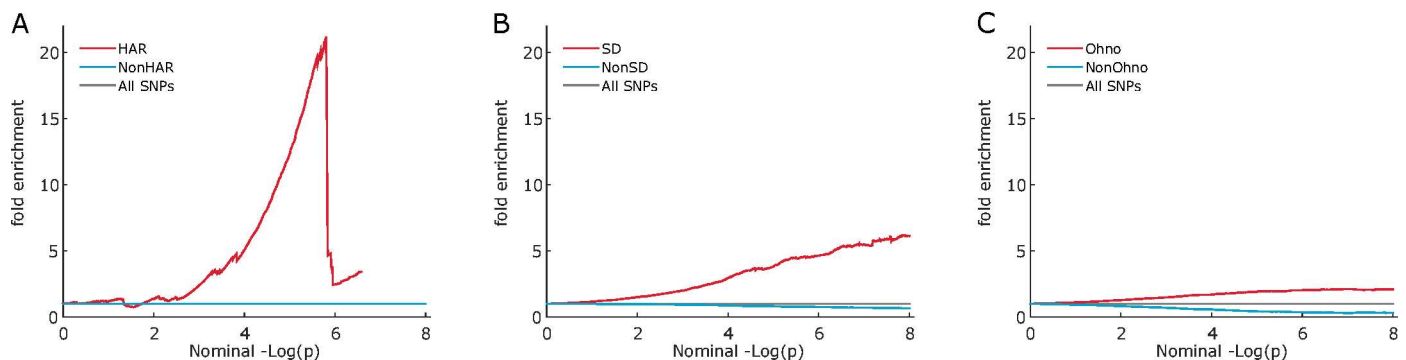
MHC has a known association with schizophrenia [18]. To assess the effect of immune-related genes, the analyses were repeated after exclusion of SNPs in the MHC region. The most significant reduction in fold enrichment occurred with SD but none was visible in HAR or Ohno (S1 Fig).

To test if the enrichment seen in schizophrenia genes was mediated by brain function, we investigated the fold enrichment of brain genes annotated to various combinations of regions of evolutionary interest (Fig 2). We observe a rather wider deflection from baseline for brain genes affiliated to HAR and SD, suggesting brain genes in these evolutionarily salient regions are more enriched compared to just any brain genes or just any SNPs in HAR, SD or Ohno.

We then ascertained the effect exerted on enrichment by other known factors, using PLSR. Regressing association squared z-scores in turn against HAR, SD and Ohno affiliation scores with exonic, intronic, 5'UTR, 3'UTR [17, 43], Brain and TotLD suggest that these co-factors explain nearly all of the enrichment (Tables 1–3A; HAR, non-centered standardized coefficients  $\beta = 0.002$ ,  $p = 0.57$ ; SD,  $\beta = -1.847$ ,  $p = 6.00 \times 10^{-6}$ ; Ohno,  $\beta = 0.496$ ,  $p = 0.694$ ).

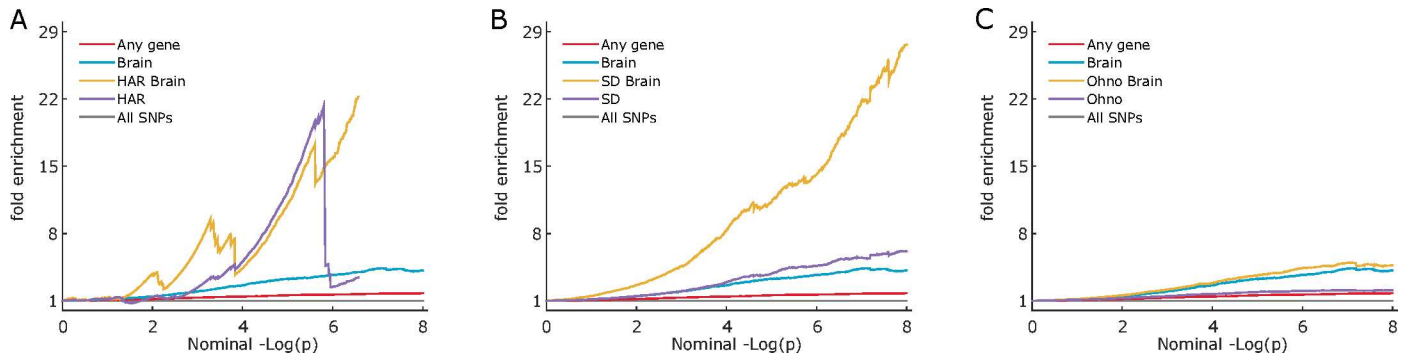
The regression results are reflected by the squared z-scores versus TotLD scatter plots (S2 Fig). These show the small but significant effect of TotLD alongside that of stratification. HAR has no apparent effect on the squared z-scores regression while SD and Ohno seem to dampen the effect of TotLD, suggesting a possible interaction effect.

Removal of MHC has no relevant effects on PLSR results apart from a loss of significance of the SD association (Tables 1–3B). In light of these results, we repeated the regression analyses



**Fig 1. ABC: Enrichment plots for schizophrenia, stratifying SNPs according to their affiliation to human accelerated regions segmental duplications and ohnologs.** Plot A shows all SNPs stratified by affiliation to human accelerated regions (HAR) and non (NonHAR). Plot B shows all SNPs stratified by affiliation to segmental duplications (SD) and non (NonSD). Plot C shows all SNPs stratified by affiliation to ohnologs (Ohno) and non (NonOhno). Some enrichment could be present in HAR but the drop at the leftmost end of the plot suggests low SNP count and consequent high error rate. A clearer enrichment is present in segmental duplications and a similar but weaker one in ohnologs. All annotations are LD-weighted.

doi:10.1371/journal.pone.0169227.g001



**Fig 2. ABC: Enrichment plots showing schizophrenia association enrichment of brain genes with various regional affiliation scores.** Enrichment plots for A) human accelerated regions (HAR), B) segmental duplications (SD) and C) ohnologs are shown for: SNPs annotated to genes with some relation to the brain, as established by an NCBI site search (“Brain”); SNPs affiliated to these regions of interest (HAR, SD or Ohno) and also annotated to genes with some relation to the brain (HAR Brain, SD Brain or Ohno Brain). In case of segmental duplications, the brain genes in the regions of interest (SD Brain) look more enriched (i.e. present a higher incidence of associations [lower p-values] with schizophrenia) compared to SD or just any Brain genes. In Ohnologs, the enrichment is way lower than in SD but Ohno Brain looks more enriched than other categories. HAR Brain and HAR show similar enrichment. All annotations are LD-weighted.

doi:10.1371/journal.pone.0169227.g002

using only SNPs in the MHC region (Tables 1–3C). We observe that the regression coefficient for HAR changes direction but remains non-significant. In case of SD it remains negative but loses significance. The Ohno affiliation coefficient turns significant suggesting that Ohno affiliation might be an enrichment factor specific to the MHC region.

To determine the effect of LD on the enrichment analyses, we generated random sets of SNPs matching the original HAR, SD and Ohno SNPs in total LD and minor allele frequency and repeated the analyses on these sets. The enrichment is somewhat reduced (S3 Fig) compared to the original set of SNPs but is still present despite the lack of evolutionary content.

The specificity of the schizophrenia results was tested by repeating the same analyses for GWAS summary statistics of other human phenotypes: neurological and psychiatric (Alzheimer’s disease and bipolar disorder), anthropometric (body mass index (BMI), height), cardiovascular (triglycerides (TG)), immune-mediated (Crohn’s disease (CD)). These GWAS were selected to be roughly comparable in sample size to the schizophrenia GWAS and therefore roughly equally powered. The fold enrichment plots show an overall high enrichment in HAR and lower in SD and Ohno across all phenotypes (S4 Fig). The PLSR shows no significant associations in HAR, negative associations between SD affiliation and BD (non-significant) and BMI (nominally significant), but positive association between Ohno and BD (significant). Other covariates like intron affiliation score are positively associated, and significantly so, across most phenotypes when regressed together with HAR or SD scores.

## Discussion

We used polygenic enrichment methods to investigate whether SNPs in the HAR, SD and Ohno regions of the human genome, related to early evolution are more likely to be associated with schizophrenia. Our results did not provide clear evidence that genetic variants in these regions are more likely to associate with schizophrenia. While association enrichment is present, it appears to be mediated by affiliation to known genomic enrichment categories like introns, 3’UTR, 5’UTR, and LD. Thus, the risk variants associated with schizophrenia do not seem to have an evolutionary origin prior to divergence between humans and chimpanzees.

While association enrichment is present, it is likely mediated by affiliation to known genomic enrichment categories like introns, 3’UTR, 5’UTR, and LD.



**Table 1. Human accelerated regions (HAR): Partial least square regression (PLSR) coefficients with multiple covariates.**

| <b>A</b>   |                |               |              |              |               |               |              |
|------------|----------------|---------------|--------------|--------------|---------------|---------------|--------------|
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR         | HAR           | Brain         | TotLD        |
| AD         | 2.187*         | 0.244*        | 0.317*       | 0.073        | 0.000         | 2.134         | -0.003       |
| BD         | 1.874          | 0.233         | 0.201        | 0.115        | 0.003         | 2.492         | 1.302*       |
| SCZ        | <b>2.583*</b>  | <b>0.189</b>  | <b>0.342</b> | <b>0.144</b> | <b>0.002</b>  | <b>1.568</b>  | <b>1.627</b> |
| BMI        | 2.196*         | -0.130        | 0.057        | 0.030        | 0.005         | 1.984         | 0.996        |
| Height     | 2.551**        | 0.153         | 0.228**      | 0.108*       | 0.006         | -0.598        | 0.644**      |
| TG         | 2.435**        | 0.336         | 0.581        | 0.157        | 0.004         | -0.870        | 0.377        |
| CD         | 2.594*         | 0.860         | 0.767        | 0.365        | -0.006        | 2.444*        | 1.433*       |
| <b>B</b>   |                |               |              |              |               |               |              |
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR         | HAR           | Brain         | TotLD        |
| AD         | 2.182*         | 0.251*        | 0.325*       | 0.067        | 0.000         | 2.141         | -0.003       |
| BD         | 1.953*         | 0.116         | 0.083        | 0.080*       | 0.003         | 2.374         | 1.044        |
| SCZ        | <b>2.551**</b> | <b>-0.059</b> | <b>0.163</b> | <b>0.043</b> | <b>0.002</b>  | <b>1.497</b>  | <b>1.125</b> |
| BMI        | 2.227*         | -0.131        | 0.065        | 0.030        | 0.005         | 1.948         | 1.006        |
| Height     | 2.560**        | 0.097         | 0.202**      | 0.083*       | 0.004         | -0.565        | 0.530**      |
| TG         | 2.482**        | 0.191         | 0.444        | 0.104        | -0.001        | -0.752        | 0.185        |
| CD         | 2.660**        | 0.608         | 0.583        | 0.264        | -0.007        | 2.172*        | 1.070*       |
| <b>C</b>   |                |               |              |              |               |               |              |
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR         | HAR           | Brain         | TotLD        |
| AD         | 2.511          | -0.254*       | -0.173       | 0.198        | 0.001         | 0.640         | -0.333       |
| BD         | -0.006         | 0.001*        | 0.001        | 0.000        | 0.000         | 0.000         | 0.004        |
| SCZ        | <b>-1.164</b>  | <b>0.660</b>  | <b>0.331</b> | <b>0.028</b> | <b>-0.004</b> | <b>-0.961</b> | <b>1.801</b> |
| BMI        | -2.363         | 0.112         | -0.089       | 0.049        | 0.009         | 0.477         | 0.608        |
| Height     | -2.198         | 0.171         | -0.085       | 0.070*       | 0.019         | -0.026        | 1.118        |
| TG         | -1.206         | 0.553         | 0.394        | 0.162        | 0.039         | 0.292         | 2.194        |
| CD         | -1.599         | 0.358         | 0.103        | 0.168*       | 0.018         | 1.583         | 1.094        |

\* nominally significant

\*\* significant after accounting for multiple testing

Phenotypes: psychiatric and neurological diseases (Alzheimer’s disease (AD), bipolar disorder (BD), schizophrenia (SCZ)), anthropometric measures (body mass index (BMI), height), cardiovascular risk factors (Triglycerides (TG)), and immune-mediated diseases (Crohn’s disease (CD)). **A**: regression on the complete set of SNPs, **B**: regression analysis after removing MHC SNPs and **C**: regression analysis keeping only MHC SNPs. The visual enrichment seen in schizophrenia (Fig 1A) is not supported by regression where no significant ( $p = 0.57$ ) association is detected. The results remain non-significant after removing MHC SNPs or using only them.

doi:10.1371/journal.pone.0169227.t001

HARs have the most recent origin among the three evolutionary proxies used in the current study [14]. HAR fold enrichment plots show some enrichment in schizophrenia, bipolar disorder and BMI, and more noteworthy enrichment in height, triglycerides and Crohn’s disease. However, regression analyses yielded no significant results in any phenotype, likely owing to the small number of SNPs in these regions. SD and Ohno show similar patterns of enrichment but generally less pronounced than in HAR. This suggests that the enrichment is not specific for schizophrenia but is influenced by the age and the extent of the evolutionary markers. We do observe significant and positive correlation between Ohno and BD but this is likely spurious since none of the GWAS from larger samples, including the schizophrenia GWAS, show any significant results.

Removal of MHC had the most visible effects on HAR followed by SD enrichments. HAR and SD, likely owing to their more recent origin, contain a higher proportion of MHC genes

**Table 2. Segmental duplication (SD): Partial least square regression (PLSR) coefficients with multiple covariates.**

| <b>A</b>   |                |               |              |               |                 |              |               |
|------------|----------------|---------------|--------------|---------------|-----------------|--------------|---------------|
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR          | SD              | Brain        | TotLD         |
| AD         | 2.577**        | 0.953         | 1.221        | 0.275         | 1.771           | 2.108        | -0.231        |
| BD         | 1.049          | 0.348         | 0.301        | 0.172         | -1.668          | 1.368        | 0.960*        |
| SCZ        | <b>1.190*</b>  | <b>0.271</b>  | <b>0.495</b> | <b>0.205</b>  | <b>-1.847**</b> | <b>0.541</b> | <b>1.006*</b> |
| BMI        | 1.065*         | -0.140        | 0.065        | 0.032         | -1.843*         | 1.010        | 0.739*        |
| Height     | 2.797**        | 0.997         | 1.478        | 0.725         | 1.127           | -0.502       | 0.555         |
| TG         | 0.206          | 1.420         | 2.470        | 0.665         | -0.445          | 0.027        | 0.092         |
| CD         | 0.024          | 2.253*        | 1.987**      | 0.962*        | -0.022          | 0.045        | 0.006         |
| <b>B</b>   |                |               |              |               |                 |              |               |
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR          | SD              | Brain        | TotLD         |
| AD         | 2.559**        | 0.971         | 1.226        | 0.252         | 1.807           | 2.088        | -0.234        |
| BD         | 1.120*         | 0.153         | 0.106        | 0.102         | -1.620          | 1.423        | 0.829*        |
| SCZ        | <b>1.277**</b> | <b>-0.073</b> | <b>0.197</b> | <b>0.049</b>  | <b>-1.816**</b> | <b>0.743</b> | <b>0.824*</b> |
| BMI        | 1.074          | -0.142        | 0.074        | 0.032         | -1.849*         | 0.986        | 0.740*        |
| Height     | 2.257**        | 1.095         | 2.365        | 1.004         | 0.743           | -0.446       | 0.358**       |
| TG         | 1.198*         | 0.840         | 1.983        | 0.456         | -1.173          | -0.080       | 0.253         |
| CD         | 0.781          | 1.978         | 1.879        | 0.855         | -0.993          | 0.488        | 0.432         |
| <b>C</b>   |                |               |              |               |                 |              |               |
| Phenotypes | Intron         | Exon          | 3UTR         | 5UTR          | SD              | Brain        | TotLD         |
| AD         | 1.785          | -0.876        | -0.631       | 0.584         | 0.005           | 1.343        | -0.240        |
| BD         | -1.336         | 0.563         | 0.826        | 0.046         | -1.339          | -0.203       | 1.179         |
| SCZ        | <b>0.113</b>   | <b>1.668</b>  | <b>0.868</b> | <b>-0.055</b> | <b>-0.112</b>   | <b>0.239</b> | <b>-1.600</b> |
| BMI        | -1.289         | 0.164         | -0.160       | 0.071         | 1.871           | 1.096        | -0.032        |
| Height     | -0.992         | 0.176         | -0.097       | 0.073         | 2.305           | 0.018        | 0.083         |
| TG         | -0.604         | 0.851         | 0.655        | 0.244         | 2.692           | 0.431        | 0.529         |
| CD         | -0.433*        | 0.260         | 0.102        | 0.129         | 2.474           | 1.213        | -0.110        |

\* nominally significant

\*\* significant after correction for multiple testing

Phenotypes: psychiatric and neurological diseases (Alzheimer’s disease (AD), bipolar disorder (BD), schizophrenia (SCZ)), anthropometric measures (body mass index (BMI), height), cardiovascular risk factors (Triglycerides (TG)), and immune-mediated diseases (Crohn’s disease (CD)). A: regression on the complete set of SNPs, B: regression analysis after removing MHC SNPs and C: regression analysis keeping only MHC SNPs. The enrichment seen in schizophrenia (Fig 1B) is not supported by regression where the association is significant ( $p = 6.00 \times 10^{-6}$ ) but the effect is negative. The association loses significance upon removal of MHC. No significant associations are found when regressing MHC SNPs only.

doi:10.1371/journal.pone.0169227.t002

compared to Ohno (S2 and S3 Tables). Ohnos are the oldest regions and cover large portions of the human genome [25]. As such, while still confounded by LD, their enrichment is largely determined by SNPs with low LD scores and hence less likely to tag any causal variants. The removal of MHC had no visible effect on Ohno enrichment but our analysis of SNPs in MHC regions interestingly yielded a significant positive association, suggesting the presence of some evolutionary effect specific to the MHC [44].

Since brain genes are more likely to be associated with schizophrenia, we carried out separate analyses targeting these. Our results suggest that brain SNPs in evolutionary salient regions show enrichment of associations with schizophrenia compared to just any brain genes or non-brain genes within the same regions. Interestingly, regions that are shared by Ohno and SD show greater enrichment of associations with schizophrenia compared to either of them separately. However, this could be due to higher LD in these shared regions. SNPs with

**Table 3. Ohnologs (Ohno): Partial least square regression (PLSR) coefficients with multiple covariates.**

| <b>A</b>   |               |               |              |              |                |               |              |
|------------|---------------|---------------|--------------|--------------|----------------|---------------|--------------|
| Phenotypes | Intron        | Exon          | 3UTR         | 5UTR         | Ohno           | Brain         | TotLD        |
| AD         | 2.126         | 0.664         | 0.846        | 0.192        | -0.034         | 0.007         | 2.406        |
| BD         | 0.437         | 0.678         | 0.534        | 0.314        | 3.001**        | 1.289         | 2.016*       |
| SCZ        | <b>0.214</b>  | <b>1.677</b>  | <b>2.984</b> | <b>1.277</b> | <b>0.496</b>   | <b>0.383</b>  | <b>0.443</b> |
| BMI        | 2.181         | -0.258        | 0.100        | 0.056        | 0.477          | 1.104*        | 2.119        |
| Height     | 2.443**       | 0.439         | 0.663        | 0.321        | -0.424         | 0.553**       | -0.645       |
| TG         | 0.968         | 1.347         | 2.297        | 0.621        | 1.042          | 0.274         | -0.950       |
| CD         | 0.902         | 2.772         | 2.454        | 1.182        | -0.039         | 0.661*        | 1.713*       |
| <b>B</b>   |               |               |              |              |                |               |              |
| Phenotypes | Intron        | Exon          | 3UTR         | 5UTR         | Ohno           | Brain         | TotLD        |
| AD         | 2.112         | 0.701         | 0.883        | 0.179        | -0.027         | 2.423         | 0.010        |
| BD         | 0.468         | 0.229         | 0.129        | 0.138        | 2.824**        | 1.595         | 0.962*       |
| SCZ        | <b>1.775</b>  | <b>-0.915</b> | <b>2.246</b> | <b>0.577</b> | <b>0.757</b>   | <b>0.953</b>  | <b>0.873</b> |
| BMI        | 2.214         | -0.256        | 0.112        | 0.055        | 0.399          | 2.068         | 1.098        |
| Height     | 2.484**       | 0.225         | 0.491        | 0.206        | -0.407         | -0.563        | 0.467**      |
| TG         | 0.970         | 1.068         | 2.427        | 0.562        | 1.255          | -0.809        | 0.169        |
| CD         | 1.310         | 2.816         | 2.674        | 1.221        | 0.122          | 2.008*        | 0.702*       |
| <b>C</b>   |               |               |              |              |                |               |              |
| Phenotypes | Intron        | Exon          | 3UTR         | 5UTR         | Ohno           | Brain         | TotLD        |
| AD         | 2.457         | -0.248*       | -0.168       | 0.196        | 1.194          | 0.629         | -0.265       |
| BD         | -0.718        | 0.960         | 1.493        | 0.062        | 2.074          | -0.328        | 0.463        |
| SCZ        | <b>-0.288</b> | <b>1.406</b>  | <b>0.721</b> | <b>0.027</b> | <b>0.620**</b> | <b>-1.735</b> | <b>0.499</b> |
| BMI        | -0.614        | 0.315*        | -0.355       | 0.134*       | -0.048         | 2.428         | 0.044        |
| Height     | -1.152        | 0.974         | -0.659       | 0.386        | 1.862*         | 0.426         | 0.551        |
| TG         | -0.289        | 2.496         | 1.882        | 0.700        | 1.918          | 1.358         | 0.277        |
| CD         | -0.166        | 0.388*        | 0.127        | 0.186        | -1.538         | 1.878         | 0.044        |

\* nominally significant

\*\* significant after correction for multiple testing

Phenotypes: psychiatric and neurological diseases (Alzheimer’s disease (AD), bipolar disorder (BD), schizophrenia (SCZ)), anthropometric measures (body mass index (BMI), height), cardiovascular risk factors (Triglycerides (TG)), and immune-mediated diseases (Crohn’s disease (CD)). A: regression on the complete set of SNPs, B: regression analysis after removing MHC SNPs and C: regression analysis keeping only MHC SNPs. The enrichment seen in schizophrenia (Fig 1C) is not supported by regression where no significant (p-value = 0.694) association is detected. The results remain non-significant after removing MHC SNPs but become significant when using only them.

doi:10.1371/journal.pone.0169227.t003

higher LD scores are likely to tag both Ohno and SD regions as much as they are likely to be associated with any causal variants.

Our analyses on random sets of SNPs matching the original ones suggest that most of the enrichment in the evolutionary salient regions is due to LD. However, a residual enrichment is clearly visible in SD and Ohno and can be guessed in HAR despite its larger variance (S3 Fig). Judging by the PLSR results, such residual enrichment should probably be attributed to Introns and UTRs (Tables 1–3). This is likely due to overall higher gene content in the case of SD and Ohno (S3 Table). The culprit in the case of HAR is likely to be found in LD with non-coding regulatory elements.

It has been proposed that schizophrenia is the price we paid for the development of a complex brain and our ability to have abstract reasoning and language [45]. The schizophrenia association enrichment in evolutionary markers appears to decrease as the age of the evolutionary

markers increases. However, the extent of the LD-weighted evolutionary markers studied here must also be considered. Limited enrichment sensitivity may also be a reason for the lack of findings in the present study, especially in the case of HAR. Other investigators found HAR to be enriched in schizophrenia [46]. However, they distinguished HAR that have accelerated in humans compared to non-human primates (pHAR) and compared to other mammals (mHAR). Further, their analyses were restricted to schizophrenia GWAS loci below given significance levels. They used LD ( $r^2 > 0.5$ ) to define their regions of interest while we treat LD as a confounder and control for it. These factors may be the reason for our failure to reproduce their findings. However, our polygenic techniques were able to show enrichment in regions under positive selection in humans after divergence from Neanderthals [9].

The modern human genome is the result of a complex process of evolution and adaptation. Using multiple evolutionary markers we investigated traces of human-specific pathology in the ancestral regions of the human genome. Our investigation did not reveal conclusive evidence supporting increased prevalence of schizophrenia association in the markers of early human evolution, HAR, SD or Ohno. Taken together with our previous report on markers of human-Neanderthal divergence [9], the current findings suggest that the increase in polygenic schizophrenia risk may have developed during more recent periods of human evolution.

## Supporting Information

**S1 Table. A: GWASs with available summary statistics.** List of genome wide association studies (GWAS) analyzed. The table shows the phenotypes, the sample size (i.e. the number of subjects, N); the total number of SNPs entering our analyses. PGC: Psychiatric Genomics Consortium, results from the second edition of the study.

(DOCX)

**S2 Table. Distribution of evolutionarily salient SNPs in each genomic category.** The table shows the number of SNPs that are affiliated to human accelerated regions (HAR), segmental duplications (SD), ohnologs (Ohno) and their distribution in various genomic categories.

(DOCX)

**S3 Table. Odds Ratio.** Table of odds ratios of HAR, SD and Ohno SNPs affiliation in each genomic category.

(DOCX)

**S1 Fig. Enrichment plots for schizophrenia, stratifying SNPs according to their affiliation to human accelerated regions segmental duplications and ohnologs after removing MHC SNPs.** Plot A shows all SNPs stratified by affiliation to human accelerated regions (HAR) and non (NonHAR). Plot B shows all SNPs stratified by affiliation to segmental duplications (SDLD) and non (NonSDLD). Plot C shows all SNPs stratified by affiliation to ohnologous (OhnoLD) and non ohnologous regions (NonOhno). We see some visible depletion of the enrichment in SD but none in HAR or ohnologs.

(TIF)

**S2 Fig. Stratified squared z-scores vs Total LD plots for schizophrenia.** 1A: human accelerated regions (HAR), 1B: Segmental duplications (SD), 1C: Ohnologs (Ohno). The SNPs are stratified based on their regional affiliation score: HAR, SD, Ohno vs non HAR, non SD, non Ohno, respectively. A scatter plot for all SNPs (None) is also reported in all figures. Mean squared z-scores are lower in SD and Ohno compared to non SD and non Ohno (Fig 1B and 1C). HAR has no apparent effect on squared z-scores. Non HAR, non SD and non Ohno

essentially overlap with the non-stratified set (None).  
(TIF)

**S3 Fig. Enrichment plots for schizophrenia, conditioning SNPs according to their affiliation to SNP sets matched to ohnologs, segmental duplications and human accelerated regions in minor allele frequency and total LD.** **Plot A** shows all SNPs stratified by affiliation to human accelerated regions (HAR) and non HAR. **Plot B** shows all SNPs stratified by affiliation to segmental duplication (SDLD) and non-segmental duplication regions (NonSD); **Plot C** shows all SNPs in ohnologous (Ohno) and non ohnologous regions (NonOhno). We observe some depletion of enrichment compared to the original set of SNPs but it is still present despite the lack of evolutionary content.  
(TIF)

**S4 Fig. Enrichment plots for some representative phenotypes, conditioning SNPs by their affiliation to human accelerated regions, segmental duplications and ohnologs.** The phenotypes are Alzheimer's disease (AD), bipolar disorder (BD), body mass index (BMI), height, triglycerides (TG) and Crohn's disease (CD). **Plot A** shows all SNPs stratified by affiliation to human accelerated regions (HAR) and non HAR; **Plot B** shows all SNPs stratified by affiliation to segmental duplication (SDLD) and non-segmental duplication regions (NonSD); **Plot C** shows all SNPs in ohnologous (Ohno) and non ohnologous regions (NonOhno). We observe some enrichment for HAR and more clear enrichment for segmental duplications but a weak enrichment for ohnologs.  
(TIF)

**S5 Fig. Enrichment plots for some representative phenotypes, conditioning SNPs by their affiliation to LD informed annotations based on human accelerated regions, segmental duplication and ohnologs after removing MHC SNPs.** The phenotypes are Alzheimer's disease (AD), bipolar disorder (BD), body mass index (BMI), Height, triglycerides (TG) and crohn's disease (CD) **Plot A** shows SNPs stratified by affiliation to human accelerated regions (HAR) and non HAR. **Plot B** shows SNPs stratified by affiliation to segmental duplication (SD) and non-segmental duplication regions (nonSD); **Plot C** shows SNPs in ohnologous (Ohno) and non ohnologous regions (NonOhno). We observe some depletion of enrichment in HAR and segmental duplications but none for ohnologs.  
(TIF)

**S1 File. Author notes.**  
(DOCX)

## Author Contributions

**Conceptualization:** FB SD OAA.

**Data curation:** SS.

**Formal analysis:** SS SH FB.

**Funding acquisition:** OAA.

**Investigation:** SS FB SH.

**Methodology:** SH AJS WKT AMD.

**Project administration:** OAA.

**Resources:** SD IM RD DAC OAA.

**Software:** YW SH AW AJS WKT AMD FB.

**Supervision:** OAA SD.

**Writing – original draft:** SS FB OAA.

**Writing – review & editing:** SS FB OAA.

## References

1. Lichtenstein P, Yip BH, Bjork 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(9659):234–9. Epub 2009/01/20. PubMed Central PMCID: PMC3879718. doi: [10.1016/S0140-6736\(09\)60072-6](https://doi.org/10.1016/S0140-6736(09)60072-6) PMID: [19150704](https://pubmed.ncbi.nlm.nih.gov/19150704/)
2. Burns JK. Psychosis: a costly by-product of social brain evolution in Homo sapiens. *Prog Neuropsychopharmacol Biol Psychiatry*. 2006; 30(5):797–814. doi: [10.1016/j.pnpbp.2006.01.006](https://doi.org/10.1016/j.pnpbp.2006.01.006) PMID: [16516365](https://pubmed.ncbi.nlm.nih.gov/16516365/)
3. Brune M. Schizophrenia—an evolutionary enigma? *Neurosci Biobehav Rev*. 2004; 28(1):41–53. doi: [10.1016/j.neubiorev.2003.10.002](https://doi.org/10.1016/j.neubiorev.2003.10.002) PMID: [15036932](https://pubmed.ncbi.nlm.nih.gov/15036932/)
4. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait—Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*. 2003; 60(12):1187–92. doi: [10.1001/archpsyc.60.12.1187](https://doi.org/10.1001/archpsyc.60.12.1187) PMID: [14662550](https://pubmed.ncbi.nlm.nih.gov/14662550/)
5. 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(2–3):118–29. PubMed Central PMCID: PMC10719140. PMID: [10719140](https://pubmed.ncbi.nlm.nih.gov/10719140/)
6. Hublin JJ, Spoor F, Braun M, Zonneveld F, Condemi S. A late Neanderthal associated with Upper Palaeolithic artefacts. *Nature*. 1996; 381(6579):224–6. doi: [10.1038/381224a0](https://doi.org/10.1038/381224a0) PMID: [8622762](https://pubmed.ncbi.nlm.nih.gov/8622762/)
7. Wynn T, Coolidge FL. The implications of the working memory model for the evolution of modern cognition. *Int J Evol Biol*. 2011; 2011:741357. PubMed Central PMCID: PMC3118292. doi: [10.4061/2011/741357](https://doi.org/10.4061/2011/741357) PMID: [21716664](https://pubmed.ncbi.nlm.nih.gov/21716664/)
8. Crow TJ. Is schizophrenia the price that Homo sapiens pays for language? *Schizophrenia Research*. 1997; 28(2–3):127–41. PMID: [9468348](https://pubmed.ncbi.nlm.nih.gov/9468348/)
9. Srinivasan S, Bettella F, Mattingsdal M, Wang Y, Witoelar A, Schork AJ, et al. Genetic Markers of Human Evolution Are Enriched in Schizophrenia. *Biol Psychiatry*. 2015.
10. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet*. 2006; 7(7):552–64. doi: [10.1038/nrg1895](https://doi.org/10.1038/nrg1895) PMID: [16770338](https://pubmed.ncbi.nlm.nih.gov/16770338/)
11. Dehal P, Boore JL. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLOS Biology*. 2005; 3(10):e314. doi: [10.1371/journal.pbio.0030314](https://doi.org/10.1371/journal.pbio.0030314) PMID: [16128622](https://pubmed.ncbi.nlm.nih.gov/16128622/)
12. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014; 511(7510):421–7. Epub 2014/07/25. PubMed Central PMCID: PMC4112379. doi: [10.1038/nature13595](https://doi.org/10.1038/nature13595) PMID: [25056061](https://pubmed.ncbi.nlm.nih.gov/25056061/)
13. Relethford JH. Genetic evidence and the modern human origins debate. *Heredity*. 2008; 100(6):555–63. doi: [10.1038/hdy.2008.14](https://doi.org/10.1038/hdy.2008.14) PMID: [18322457](https://pubmed.ncbi.nlm.nih.gov/18322457/)
14. 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(10):e168. PubMed Central PMCID: PMC1599772. doi: [10.1371/journal.pgen.0020168](https://doi.org/10.1371/journal.pgen.0020168) PMID: [17040131](https://pubmed.ncbi.nlm.nih.gov/17040131/)
15. Hubisz MJ, Pollard KS. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr Opin Genet Dev*. 2014; 29:15–21. doi: [10.1016/j.gde.2014.07.005](https://doi.org/10.1016/j.gde.2014.07.005) PMID: [25156517](https://pubmed.ncbi.nlm.nih.gov/25156517/)
16. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006; 443(7108):167–72. doi: [10.1038/nature05113](https://doi.org/10.1038/nature05113) PMID: [16915236](https://pubmed.ncbi.nlm.nih.gov/16915236/)
17. 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(4):e1003449. PubMed Central PMCID: PMC3636284. doi: [10.1371/journal.pgen.1003449](https://doi.org/10.1371/journal.pgen.1003449) PMID: [23637621](https://pubmed.ncbi.nlm.nih.gov/23637621/)
18. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. *Nature*. 2016; 530(7589):177–83. PubMed Central PMCID: PMC4752392. doi: [10.1038/nature16549](https://doi.org/10.1038/nature16549) PMID: [26814963](https://pubmed.ncbi.nlm.nih.gov/26814963/)

19. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009; 460(7256):744–7. PubMed Central PMCID: PMC3077530. doi: [10.1038/nature08186](https://doi.org/10.1038/nature08186) PMID: [19571808](https://pubmed.ncbi.nlm.nih.gov/19571808/)
20. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*. 2005; 2:16–. doi: [10.1186/1742-9994-2-16](https://doi.org/10.1186/1742-9994-2-16) PMID: [16242022](https://pubmed.ncbi.nlm.nih.gov/16242022/)
21. Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, et al. A burst of segmental duplications in the genome of the African great ape ancestor. *Nature*. 2009; 457(7231):877–81. PubMed Central PMCID: PMC2751663. doi: [10.1038/nature07744](https://doi.org/10.1038/nature07744) PMID: [19212409](https://pubmed.ncbi.nlm.nih.gov/19212409/)
22. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. *Genome Research*. 2001; 11(6):1005–17. doi: [10.1101/gr.187101](https://doi.org/10.1101/gr.187101) PMID: [11381028](https://pubmed.ncbi.nlm.nih.gov/11381028/)
23. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci U S A*. 2010; 107(20):9270–4. PubMed Central PMCID: PMC2889102. doi: [10.1073/pnas.0914697107](https://doi.org/10.1073/pnas.0914697107) PMID: [20439718](https://pubmed.ncbi.nlm.nih.gov/20439718/)
24. Singh PP, Affeldt S, Malaguti G, Isambert H. Human Dominant Disease Genes Are Enriched in Paralogous Originating from Whole Genome Duplication. *PLoS Comput Biol*. 2014; 10(7):e1003754. doi: [10.1371/journal.pcbi.1003754](https://doi.org/10.1371/journal.pcbi.1003754) PMID: [25080083](https://pubmed.ncbi.nlm.nih.gov/25080083/)
25. Makino T, McLysaght A, Kawata M. Genome-wide deserts for copy number variation in vertebrates. *Nat Commun*. 2013; 4:2283. doi: [10.1038/ncomms3283](https://doi.org/10.1038/ncomms3283) PMID: [23917329](https://pubmed.ncbi.nlm.nih.gov/23917329/)
26. Singh PP, Arora J, Isambert H. Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes. *PLoS Comput Biol*. 2015; 11(7):e1004394. doi: [10.1371/journal.pcbi.1004394](https://doi.org/10.1371/journal.pcbi.1004394) PMID: [26181593](https://pubmed.ncbi.nlm.nih.gov/26181593/)
27. Efron B. Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge; New York: Cambridge University Press; 2010. xii, 263 p.
28. 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(7538):197–206. doi: [10.1038/nature14177](https://doi.org/10.1038/nature14177) PMID: [25673413](https://pubmed.ncbi.nlm.nih.gov/25673413/)
29. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010; 467(7317):832–8. Epub 2010/10/01. PubMed Central PMCID: PMC2955183. doi: [10.1038/nature09410](https://doi.org/10.1038/nature09410) PMID: [20881960](https://pubmed.ncbi.nlm.nih.gov/20881960/)
30. 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(7307):707–13. Epub 2010/08/06. PubMed Central PMCID: PMC2939276. doi: [10.1038/nature09270](https://doi.org/10.1038/nature09270) PMID: [20686565](https://pubmed.ncbi.nlm.nih.gov/20686565/)
31. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*. 2010; 42(12):1118–25. Epub 2010/11/26. PubMed Central PMCID: PMC299551. doi: [10.1038/ng.717](https://doi.org/10.1038/ng.717) PMID: [21102463](https://pubmed.ncbi.nlm.nih.gov/21102463/)
32. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013; 45(12):1452–8. Epub 2013/10/29. PubMed Central PMCID: PMC3896259. doi: [10.1038/ng.2802](https://doi.org/10.1038/ng.2802) PMID: [24162737](https://pubmed.ncbi.nlm.nih.gov/24162737/)
33. Psychiatric GCBWDG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011; 43(10):977–83. Epub 2011/09/20. PubMed Central PMCID: PMC3637176. doi: [10.1038/ng.943](https://doi.org/10.1038/ng.943) PMID: [21926972](https://pubmed.ncbi.nlm.nih.gov/21926972/)
34. Kulski JK, Gaudieri S, Bellgard M, Balmer L, Giles K, Inoko H, et al. The Evolution of MHC Diversity by Segmental Duplication and Transposition of Retroelements. *Journal of Molecular Evolution*. 1997; 45(6):599–609. PMID: [9419237](https://pubmed.ncbi.nlm.nih.gov/9419237/)
35. 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(4):e1003455. PubMed Central PMCID: PMC3636100. doi: [10.1371/journal.pgen.1003455](https://doi.org/10.1371/journal.pgen.1003455) PMID: [23637625](https://pubmed.ncbi.nlm.nih.gov/23637625/)
36. 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 Genetics*. 2013; 9(4).
37. Andreassen OA, Thompson WK, Dale AM. Boosting the power of schizophrenia genetics by leveraging new statistical tools. *Schizophr Bull*. 2014; 40(1):13–7. PubMed Central PMCID: PMC3885310. doi: [10.1093/schbul/sbt168](https://doi.org/10.1093/schbul/sbt168) PMID: [24319118](https://pubmed.ncbi.nlm.nih.gov/24319118/)

38. 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. *Am J Hum Genet.* 2013; 92(2):197–209. Epub 2013/02/05. PubMed Central PMCID: PMC3567279. doi: [10.1016/j.ajhg.2013.01.001](https://doi.org/10.1016/j.ajhg.2013.01.001) PMID: [23375658](https://pubmed.ncbi.nlm.nih.gov/23375658/)
39. Consortium TEP. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004; 306(5696):636–40. doi: [10.1126/science.1105136](https://doi.org/10.1126/science.1105136) PMID: [15499007](https://pubmed.ncbi.nlm.nih.gov/15499007/)
40. Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS method. In: Kågström B, Ruhe A, editors. *Matrix Pencils: Proceedings of a Conference Held at Pite Havsbad, Sweden, March 22–24, 1982.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1983. p. 286–93.
41. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems.* 2001; 58(2):109–30.
42. Martens H, Martens M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference.* 2000; 11(1–2):5–16.
43. Zablocki RW, Schork AJ, Levine RA, Andreassen OA, Dale AM, Thompson WK. Covariate-Modulated Local False Discovery Rate for Genome-wide Association Studies. *Bioinformatics.* 2014.
44. Traherne JA. Human MHC architecture and evolution: implications for disease association studies. *International Journal of Immunogenetics.* 2008; 35(3):179–92. doi: [10.1111/j.1744-313X.2008.00765.x](https://doi.org/10.1111/j.1744-313X.2008.00765.x) PMID: [18397301](https://pubmed.ncbi.nlm.nih.gov/18397301/)
45. Crow TJ. The 'big bang' theory of the origin of psychosis and the faculty of language. *Schizophr Res.* 2008; 102(1–3):31–52. doi: [10.1016/j.schres.2008.03.010](https://doi.org/10.1016/j.schres.2008.03.010) PMID: [18502103](https://pubmed.ncbi.nlm.nih.gov/18502103/)
46. 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(5):1148–60. PubMed Central PMCID: PMC4408416. doi: [10.1093/molbev/msv031](https://doi.org/10.1093/molbev/msv031) PMID: [25681384](https://pubmed.ncbi.nlm.nih.gov/25681384/)