UC San Diego

UC San Diego Previously Published Works

Title

Genetic Markers of Human Evolution Are Enriched in Schizophrenia

Permalink

https://escholarship.org/uc/item/5615t7wn

Journal

Biological Psychiatry, 80(4)

ISSN

0006-3223

Authors

Srinivasan, Saurabh
Bettella, Francesco
Mattingsdal, Morten
et al.

Publication Date

2016-08-01

DOI

10.1016/j.biopsych.2015.10.009

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Archival Report

Biological Psychiatry

Genetic Markers of Human Evolution Are Enriched in Schizophrenia

Saurabh Srinivasan, Francesco Bettella, Morten Mattingsdal, Yunpeng Wang, Aree Witoelar, Andrew J. Schork, Wesley K. Thompson, Verena Zuber, The Schizophrenia Working Group of the Psychiatric Genomics Consortium, The International Headache Genetics Consortium, Bendik S. Winsvold, John-Anker Zwart, David A. Collier, Rahul S. Desikan, Ingrid Melle, Thomas Werge, Anders M. Dale, Srdjan Djurovic, and Ole A. Andreassen

ABSTRACT

BACKGROUND: Why schizophrenia has accompanied humans throughout our history despite its negative effect on fitness remains an evolutionary enigma. It is proposed that schizophrenia is a by-product of the complex evolution of the human brain and a compromise for humans' language, creative thinking, and cognitive abilities.

METHODS: We analyzed recent large genome-wide association studies of schizophrenia and a range of other human phenotypes (anthropometric measures, cardiovascular disease risk factors, immune-mediated diseases) using a statistical framework that draws on polygenic architecture and ancillary information on genetic variants. We used information from the evolutionary proxy measure called the Neanderthal selective sweep (NSS) score.

RESULTS: Gene loci associated with schizophrenia are significantly ($p = 7.30 \times 10^{-9}$) more prevalent in genomic regions that are likely to have undergone recent positive selection in humans (i.e., with a low NSS score). Variants in brain-related genes with a low NSS score confer significantly higher susceptibility than variants in other brain-related genes. The enrichment is strongest for schizophrenia, but we cannot rule out enrichment for other phenotypes. The false discovery rate conditional on the evolutionary proxy points to 27 candidate schizophrenia susceptibility loci, 12 of which are associated with schizophrenia and other psychiatric disorders or linked to brain development.

CONCLUSIONS: Our results suggest that there is a polygenic overlap between schizophrenia and NSS score, a marker of human evolution, which is in line with the hypothesis that the persistence of schizophrenia is related to the evolutionary process of becoming human.

Keywords: Evolution, GWAS, Human, Neanderthal, Polygenic, Schizophrenia

http://dx.doi.org/10.1016/j.biopsych.2015.10.009

Schizophrenia affects approximately 1% of the world's population and has accompanied humans through much of our recorded history (1–6). This seemingly human-specific disorder is characterized by hallucinations and delusions (often involving language), thought disorders, and higher order cognitive dysfunctions. The mechanisms of schizophrenia are not well understood, but its heritability is high, between 60% and 80% (7), and the fecundity of affected people is reduced (8). Nevertheless, the prevalence of the disease seems to remain stable across generations, giving rise to the yet unresolved "evolutionary enigma" of schizophrenia (3,4,9,10). Large variations in incidence across populations argue for environmental causes. However, by using standard, precisely drawn diagnostic criteria, the variation in incidence can be reduced (11). Classic explanations include a single, partially dominant gene with low penetrance giving slight physiologic advantages (12); balanced selection, where the gene variants conferring risk of the disease provide an advantage in particular environments; and hitchhiking, where disease variants are passed along with advantageous neighboring gene variants. Newer studies have focused on the polygenic nature of schizophrenia and have attributed the prevalence of the disease to the sporadic nature of complex disorders (13).

Archaeological and paleontologic evidence points to the appearance of various hominid forms such as *Homo habilis*, *Homo erectus*, *Homo neanderthalensis* (Neanderthals), and modern *Homo sapiens* (humans) over 2.5 million years from the Lower Paleolithic Age to the Neolithic Age. It is debated whether the emergence of the "modern human" was a morphologic or a behavioral process, a one-time event or a continuous process of adaptation and assimilation of different forms. Even as morphologic changes stopped, behavioral changes continued, rapidly leading to the ultimate success of humans (14).

Over the Pleistocene period, we see the appearance of specialized tools, the introduction of decorative arts, burial

 © 2016 Society of Biological Psychiatry.
 ISSN: 0006-3223

practices (15), and possibly the development of language (16). Research suggests that language acquisition played an important role in shaping the brain, helping humans to think abstractly and be more creative, but it also made humans vulnerable to psychiatric disorders such as schizophrenia (17). Changes that contributed to the ability to think more creatively and to improve executive function (18) could have also harbored susceptibility to this pathology (19). However, although archaeological evidence provides clues about other aspects of human evolution, it cannot offer insights into the origin of psychiatric disorders.

More recent developments in human genetics have provided unprecedented opportunities to investigate evolutionary aspects of schizophrenia. Genome-wide association studies (GWASs) have identified >100 schizophrenia risk loci and highlighted the polygenic architecture of the disease (20). The genome sequence of Neanderthals (21,22), close relatives of early modern humans, can help pinpoint the genomic regions affected by positive selection since the two species diverged. The genomic differences between the two *Homo* species may help explain specific human features and the relationship between human evolution and schizophrenia.

Several lines of evidence indicate that schizophrenia is a polygenic disorder (23,24) with a large number of risk loci, each with a small effect (20). We have recently developed statistical tools, building on an empirical Bayesian framework (25), that are specifically designed for polygenic architectures. These tools have been successfully applied to investigate several complex human phenotypes (26–32) but have not yet been used to study the evolutionary features thereof. We hypothesized that schizophrenia is the result of human polygenic adaptation (24) and investigated whether regions of the human genome, which may have undergone recent positive selection, are enriched of association with schizophrenia.

METHODS AND MATERIALS

Samples

We obtained summary statistics for ~1.0–2.5 million single nucleotide polymorphisms (SNPs) from GWASs of schizophrenia (conducted by the Psychiatric Genomics Consortium) and other phenotypes, including anthropometric measures (body mass index, height, waist-to-hip ratio), cardiovascular disease risk factors (systolic blood pressure, total cholesterol, triglycerides), immune-mediated diseases (celiac disease, Crohn's disease, rheumatoid arthritis, ulcerative colitis), and other psychiatric and central nervous system disorders (attention-deficit/hyperactivity disorder, Alzheimer's disease, bipolar disorder, and multiple sclerosis) (Supplemental Table S1). These studies included ~1.3 million phenotypic observations, although overlap between samples makes the number of unique subjects lower.

Neanderthal Selective Sweep Score

The Neanderthal selective sweep (NSS) score is obtained through alignment of human, Neanderthal, and primate consensus sequences (21,33) and is downloadable from the UCSC Genome Browser website (34) (http://genome.ucsc.edu; ntSssZScorePMVar track [S-scores]), developed and maintained by the University of California, Santa Cruz. This track consists of two entries per SNP (*z*-score + SD) and (*z*-score − SD). The NSS score provides a likelihood index of positive selection in humans sometime after the divergence of humans and Neanderthals (21,33) by measuring the relative abundance of ancestral/nonancestral (i.e., aligned/nonaligned with primate consensus) alleles in these two lineages. A negative NSS score indicates scarcity of nonancestral alleles in Neanderthals compared with modern humans and therefore possible positive selection in the latter. The (*z*-score + SD) entries in the genome track represent an upper bound on the statistic and are therefore conservative measures of positive selection likelihood. These were extracted for all SNPs in the GWASs of interest (Supplemental Table S1) and follow the distribution illustrated in Supplemental Figure S1. The (*z*-score + SD) entries, termed NSS scores, were used as ancillary information or covariates in the enrichment analyses. Using the NSS scores, the authors of the two articles on the Neanderthal genome identified regions of the human genome that are significantly likely to have undergone recent positive selection. The same analyses performed directly using the NSS scores were also performed using linkage disequilibrium (LD) weighted scores (see Analytical Approach) measuring affiliation to these regions.

Brain Genes

To control the enrichment analyses for affiliation to brain genes, we identified genes with a known function in the brain using information from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/gene). The query "human brain" in *Homo sapiens* revealed 2494 genes (March 2015). For comparison, we also used the list of brain genes from Kang *et al*. (35), which includes 1415 genes selected based on expression in various neural cells. The LD weighted procedure (see Analytical Approach) applied to the above-mentioned NSS regions was applied to these genes, yielding brain genes LD weighted affiliation scores.

Analytical Approach

We employed a genetic enrichment method developed to dissect the genetic architecture of complex traits (26,28,29,32). Specifically, we investigated the enrichment of associations concurrent with the NSS score selection index in a covariate-modulated statistical approach (36). We investigated whether SNPs with a low NSS score and therefore in regions possibly subjected to positive selection in humans are more likely associated with schizophrenia or other phenotypes. All statistical analyses were carried out with a covariate-modulated enrichment analysis package developed on R (www.r-project.org) and MATLAB (www.mathworks.se/products/matlab/) programming platforms.

Quantile-Quantile and Fold Enrichment Plots. Quantile-quantile (Q-Q) plots are designed to compare two distributions; we compared the nominal *p* value distribution with the empirical distribution. In the presence of null relationships only, the nominal *p* values form a straight line on a Q-Q plot when plotted against the empirical distribution. We plotted

$-\log_{10}$ nominal p values against $-\log_{10}$ empirical p values for the two SNP strata determined by the NSS score (conditional Q-Q plots) as well as for all SNPs. Leftward deflections of the observed distribution from the null line reflect increased tail probabilities in the distribution of test statistics (z-scores) and consequently an overabundance of low p values compared with that expected under the null hypothesis.

To graphically assess genetic enrichment, we used conditional fold enrichment plots (37). A direct measure of the enrichment is given by the degree of deflection from the expected null. The fold enrichment is derived as follows: first, the empirical cumulative distribution of $-\log_{10}(p)$ values for SNP association is computed for a given phenotype for all SNPs and for the two dichotomous SNPs strata determined by the NSS score. The fold enrichment of each stratum is then calculated as the ratio cumulative distribution function $(CDF)_{stratum}/CDF_{all}$ between the $-\log_{10}(p)$ CDF for that stratum and the CDF for all SNPs. The nominal $-\log_{10}(p)$ values are plotted on the x axis, and the fold enrichment is plotted on 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 $p > 5 \times 10^{-8}$).

Binomial Proportion Test. On randomly subdividing a set of SNPs into two disjoint subsets, one would expect these to present similar p value distributions. In particular, the proportion of SNPs with a p value below a certain threshold should be the same in the two subsets. The binomial proportion test (BPT) measures deviations from this null hypothesis below a threshold of interest (38). We compared the proportions of SNPs in the top $-\log_{10}(p)$ percentile within the two NSS strata. The BPT assumes independence of the data. Because of LD between SNPs, this independence requirement does not hold. We therefore subdivided the whole SNP set into blocks defined by 1-Mb windows and an LD r^2 threshold of .2 and randomly selected 10 sets of SNP representatives from all blocks. Then 10 sets of BPTs were carried out on the approximately independent randomly chosen SNPs, and the final p value was calculated from the median of the BPT statistics.

LD Weighted SNP Annotation Score. The use of GWAS SNPs in DNA regions of interest may underestimate the extent to which those regions are represented in the analysis. We used an LD weighted scoring algorithm developed previously (26) to identify SNPs that tag specific DNA regions even if they are not situated within them. For each GWAS SNP, a pairwise correlation coefficient approximation to LD (r^2) was calculated for all 1000 Genomes Project (1KGP) SNPs within 1 million base pairs (1 Mb). All r^2 values $< .2$ were set to 0, and each SNP was assigned an r^2 value of 1.0 with itself. The 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_i, its LD weighted region annotation score was computed as LD $score_i = \Sigma_j \delta_j r_{ij}^2$, where r_{ij}^2 is the LD r^2 between SNP_i and SNP_j, and δ_j takes values of 1 or 0 depending on whether or not the 1KGP SNP_j is within the region of interest.

Intergenic SNPs. Intergenic SNPs are defined as having LD weighted annotation scores for each of the genomic categories analyzed by Schork *et al.* (26) equal to 0 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 noncoding RNA, within a transcription factor binding site, or within a microRNA binding site. The SNPs singled out in this way are expected to form a collection of nongenic SNPs not belonging to any (annotated) functional elements within the genome (including through LD) and therefore represent a collection of SNPs more likely to be null.

Intergenic Correction. Intergenic SNPs were used to estimate the inflation of GWAS summary statistics secondary to cryptic relatedness. We used intergenic SNPs because their relative depletion of associations (26) suggests that they provide a set of SNPs whose statistics are less inflated by polygenic associations. The inflation factor, λ_{GC}, was estimated as the median squared z-score of independent (LD $r^2 < .2$) sets of intergenic SNPs across 100 LD-pruning iterations divided by the expected median of a χ^2 distribution with 1 df.

Squared z-Score Regression. The hypothesis is that there is some proportionality between a continuous covariate of interest and the incidence of SNP association with a phenotype. A viable proxy for the latter is the extent of the association z-scores. We regressed the squared z-scores against the NSS scores. Other covariates were included in the regression as well to account for possible confounding factors. These were exonic, intronic, 5′UTR, 3′UTR annotation scores (26,39); brain gene affiliation scores; genotypic variance; and total LD. As done for the BPT, regression analyses were performed on the 10 sets of SNP representatives, and the regression coefficient p values were calculated from the median of the 10 regression coefficient estimates.

Replication. The procedure used to compute the conditional rate of replication (Supplement) follows the one of Schork *et al*. (26). The 52 substudies were subdivided into two groups of 26 in 50 different ways, the first group, D_k, $k = 1 \ldots 50$, serving as discovery group, and the second group, R_k, $k = 1 \ldots 50$, serving as replication group. Cumulative replication rates were calculated over each of 1000 equally spaced bins spanning the range of $-\log_{10}$ (p values) observed in the discovery group and for each of the 50 subdivisions. Every cumulative replication rate was calculated as the fraction of SNPs with a discovery $-\log_{10}$ (p value) greater than the lower bound of the bin that had a replication p value $< .05$. Average cumulative replication rates were subsequently computed across the 50 subdivisions.

RESULTS

We first assessed the influence exerted on schizophrenia association propensity by the Neanderthal "character" of the DNA region of the SNP, as measured by the NSS score selection index (Supplemental Figure S1) (21). Using data from the recently published schizophrenia GWAS (20), we conditioned schizophrenia association p values on the NSS

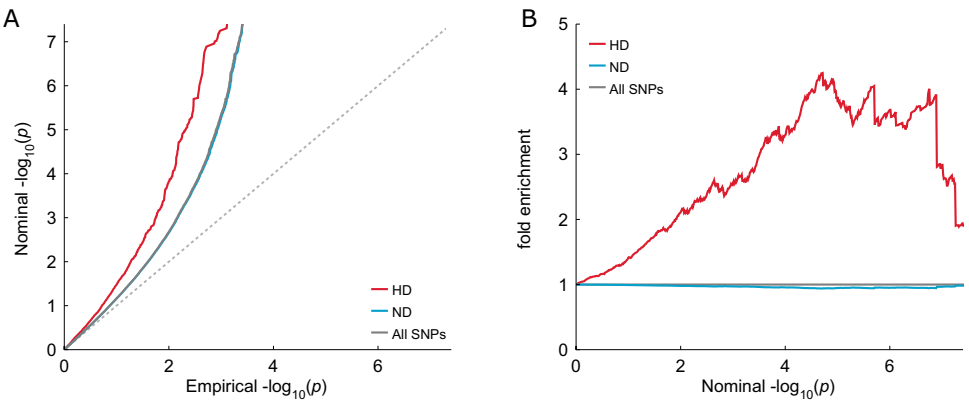


Figure 1. **(A)** Quantile-quantile and **(B)** fold enrichment plots of genome-wide association study summary statistics p values for schizophrenia, stratified based on Neanderthal selective sweep (NSS) score. The human divergent (HD) stratum comprises single nucleotide polymorphisms (SNPs) with negative NSS scores. The regions around these SNPs present fewer derived alleles in Neanderthal than expected given the frequency of derived alleles in humans and may have undergone recent positive selection in the latter. The nondivergent (ND) stratum comprises all SNPs with positive NSS scores. The HD SNPs show a marked leftward **(A)** and upward **(B)** deflection from the lines corresponding to all SNPs. This signifies a comparatively higher proportion of low p values among HD SNPs.

score. The conditional Q-Q (Figure 1A) and fold enrichment (Figure 1B) plots show that SNPs with negative NSS scores are enriched for associations with schizophrenia compared with SNPs with positive NSS scores. These results are nominally confirmed by the BPT ($p = 2.40 \times 10^{-2}$), and more robustly by the squared z-score regression against the NSS score ($\beta = -.067$, $p = 7.30 \times 10^{-9}$) (Table 1).

To control for the known effect of immune-related genes, all analyses were repeated after exclusion of SNPs in the major histocompatibility complex regions. These do not appear to affect the fold enrichment to any measurable extent (Supplemental Table S2 and Supplemental Figure S2). It appears that the SNPs in human DNA regions that diverge from their Neanderthal counterparts have a higher propensity to be associated with schizophrenia. Similar analyses were repeated using affiliation to NSS regions that were deemed significantly likely (top 5%) to have undergone positive selection on alignment with the Neanderthal genome. In this case, we investigated the original (21) and the more recently sequenced Neanderthal genome (22) and confirmed the initial results (Supplemental Table S3 and Supplemental Figure S3).

We carried out the same analyses on other phenotypes to assess the specificity of the evolutionary enrichment. As shown in Q-Q plots and fold enrichment plots (Figure 2), other phenotypes show mostly modest or scarce enrichment as a function of NSS compared with schizophrenia. The only other significant excesses of low p values were detected by the BPTs and the regression analyses for height and to some extent for body mass index (Table 1). Height in particular has effect size comparable to that of schizophrenia and possibly larger still, but its SE is larger. Targeted analyses of other psychiatric (attention-deficit/hyperactivity disorder, bipolar disorder, major depressive disorder) and neurologic (Alzheimer's disease, migraine, multiple sclerosis) disorders revealed no measurable enrichment effect (Supplemental Figures S4 and S5). Schizophrenia has by far the largest NSS effect size among the psychiatric and neurologic GWASs, all of which have similar SEs (Supplemental Figure S8).To test the extent to which the effect on schizophrenia depends on the power of the GWAS from 2014 (20), we performed the same analyses on the smaller schizophrenia GWAS from 2013 (40), which is comparable in size to several of the other GWASs. The enrichment was diminished (Supplemental Figure S6) but remained nominally significant according to the regression analysis ($\beta = -.038$, $p = 7.93 \times 10^{-3}$). We also tested the censored (Supplemental Methods and Materials) schizophrenia GWAS summary statistics and still found a significant (regression coefficient $p = 2.87 \times 10^{-6}$) residual enrichment.

The effect of brain genes affiliation on enrichment was investigated further by testing whether brain genes with negative NSS scores are more enriched of associations with schizophrenia than any brain genes. The enrichment plots (Figure 3) for brain genes with negative NSS scores show a wider deflection from baseline, and the BPT shows a significant difference in the proportion of association p values in the lowest percentile ($p = 5.5 \times 10^{-3}$).

We used the conditional false discovery rate (FDR) analysis (Supplemental Methods and Materials) to identify possible genomic loci associated with schizophrenia subject to the condition of having a negative NSS score. The 27 genomic loci identified (conditional FDR $< .01$) are listed in Supplemental Table S4 together with the annotated genes. A closer inspection of Supplemental Table S4 reveals no preferential direction of effect (Supplemental Figure S7) (i.e., positive and negative z-scores were equally represented). This lack of directionality is confirmed on regressing the SNP z-scores against their NSS scores (regression data not shown; i.e., no significant association between the two could be detected). None of the loci are identified by the analyses involving NSS region affiliation scores. This is probably due to the dichotomous origin of this measure, which is less well suited to the FDR lookup table smoothing procedure.

To assess the reliability of the genomic loci identified via conditional FDR, we investigated the association replication rates in independent schizophrenia substudies, defined as the proportion of SNPs declared significant in training samples with p values $< .05$ in the replication sample and with z-scores with the same sign in discovery and replication samples. We found that SNPs with NSS scores < 0 replicate at a higher rate than other SNPs (Figure 4). This finding confirms that the observed enrichment is due to associations and not to population stratification or other potential sources of spurious

Table 1. Neanderthal Selective Sweep Score

GWAS	β (Min, Max)	SE	*p* Value	CI	BPT(*p*) 1% (Min, Max)
AD	−.012 (−.029, .012)	.016	5.00E-01	−.041, .022	2.5E-01 (8.0E-04, 5.9E-01)
ADHD	.003 (−.013, .017)	.013	8.40E-01	−.022, .028	4.5E-01 (3.6E-02, 6.2E-01)
BD	−.004 (−.020, .013)	.010	7.10E-01	−.023, .015	4.8E-01 (1.4E-02, 9.1E-01)
MDD	−.018 (−.034, −.006)	.012	2.00E-01	−.042, .007	5.6E-01 (1.1E-01, 9.4E-01)
Migraine	−.006 (−.029, −.002)	.013	6.90E-01	−.032, .020	7.4E-01 (2.3E-01, 9.2E-01)
MS	−.003 (−.034, .022)	.020	8.80E-01	−.042, .035	5.1E-01 (8.7E-03, 8.5E-01)
SCZ 1	−.038 (−.052, −.026)	.013	7.90E-03	−.063, −.013	1.7E-01 (1.8E-02, 6.1E-01)
SCZ 2	−.067 (−.076, −.056)	.010	7.30E-09	−.088, −.047	2.4E-02 (5.7E-06, 3.7E-01)
BMI	−.050 (−.061, −.037)	.016	4.50E-03	−.079, −.023	4.2E-01 (6.2E-02, 9.2E-01)
Height	−.074 (−.098, −.058)	.015	8.90E-06	−.104, −.045	1.1E-01 (3.9E-04, 6.9E-01)
WHR	−.026 (−.039, −.019)	.011	2.50E-02	−.047, −.006	2.3E-01 (7.1E-03, 5.1E-01)
SBP	−.015 (−.024, −.003)	.010	1.80E-01	−.036, .005	3.7E-01 (9.3E-02, 7.3E-01)
TC	−.001 (−.019, .023)	.019	9.80E-01	−.037, .039	5.3E-01 (3.1E-01, 8.6E-01)
TG	−.017 (−.024, −.003)	.015	3.30E-01	−.048, .014	4.3E-01 (8.2E-03, 8.3E-01)
CD	−.025 (−.049, .003)	.019	2.50E-01	−.062, .014	5.3E-01 (2.6E-01, 8.6E-01)
CeD	−.000 (−.024, .018)	.018	9.90E-01	−.037, .035	3.7E-01 (1.2E-01, 8.4E-01)
RA	−.004 (−.020, .013)	.011	7.70E-01	−.025, .017	5.7E-01 (2.2E-02, 8.8E-01)
UC	−.017 (−.027, .015)	.015	3.10E-01	−.047, .014	5.3E-01 (1.7E-01, 9.2E-01)

BPT, binomial proportion test; CI, confidence interval; GWAS, genome-wide association study; Max, maximum; Min, minimum.

Squared *z*-scores vs. Neanderthal selective sweep score regression for various phenotypes controlling for other enrichment factors (genetic annotation scores, genotypic variance, linkage disequilibrium) and top 1% BPT *p* values. Schizophrenia is the only phenotype with a significant negative correlation between squared *z*-scores and Neanderthal selective sweep scores while controlling for other covariates. Also, in SCZ2, the top 1% single nucleotide polymorphisms include a nominally significant excess of single nucleotide polymorphisms with Neanderthal selective sweep score <0 (human divergent) compared with any single nucleotide polymorphisms (BPT).

Phenotypes: psychiatric and other neurologic diseases (Alzheimer's disease [AD], attention-deficit/hyperactivity disorder [ADHD], bipolar disorder [BD], major depressive disorder [MDD], migraine, multiple sclerosis [MS], first and second edition of the schizophrenia GWAS by the Psychiatric Genomic Consortium [SCZ1 and SCZ2]), anthropometric measures (body mass index [BMI], height, waist-to-hip ratio [WHR]), cardiovascular risk factors (systolic blood pressure [SBP], total cholesterol [TC], triglycerides [TG]), immune-mediated diseases (Crohn's disease [CD], celiac disease [CeD], rheumatoid arthritis [RA], ulcerative colitis [UC]).

effects. Replication rates were extrapolated for the 27 NSS candidate loci and are reported in Supplemental Table S4 as well.

DISCUSSION

Applying a polygenic statistical approach, we leveraged recent large GWAS data and showed that schizophrenia associations have a higher propensity to be found in genomic regions that diverge from their Neanderthal counterparts (negative NSS score). Such polygenic overlap between schizophrenia and a marker of human evolution is in accordance with the hypothesis of Crow (19), suggesting that many schizophrenia susceptibility factors might have arisen as a "side effect" of human achievements such as language and creative thinking (17). The current findings support the view that this evolutionary process also made humans vulnerable to schizophrenia.

Previous studies of evolutionary factors of schizophrenia focused on small sets of genes (41,42). The analysis of Bigdeli *et al.* (43) was more systematic but applied human accelerated regions as evolutionary proxy. Xu *et al.* (44), using special human accelerated regions, showed that genes next to human accelerated regions in primates were under greater selection pressure compared with other genes and are more likely to be associated with schizophrenia susceptibility loci. Green *et al.* (21), who reported the first Neanderthal draft sequence, introduced the selective sweep score and investigated its relationship with the disease association but only for the most significant genes singled out by their analysis. In the present study, we used the information from the original publication of Green *et al.* (21) and the more recent report by Prüfer *et al.* (22) on the complete Neanderthal genome sequence to identify evolutionary enrichment patterns with a polygenic approach. Another asset for our study was the availability of a large schizophrenia GWAS of >80,000 participants (20), which makes it feasible to investigate evolutionary factors in schizophrenia with adequate power.

The results presented here are in line with the idea of polygenic adaptation, which is believed to play a role in the development of many complex human diseases, as it likely happened in adaptation of humans to pathogens and in the variation of morphologic traits such as height (45–47). Classic selective sweeps, originating from strong selective pressure, are relatively rare in modern humans (48), and natural selection is not the only factor shaping human variation. Instead, polygenic selection involving subtle shifts of allele frequencies at many loci simultaneously has been suggested to be common for complex traits in humans (48). Selection acting simultaneously on many standing variants could be an efficient mechanism for phenotypic adaptation (49,50). Given these premises, it is desirable to use analytical tools designed to capture polygenicity. The methods applied in our analyses were useful in studying polygenic factors in schizophrenia

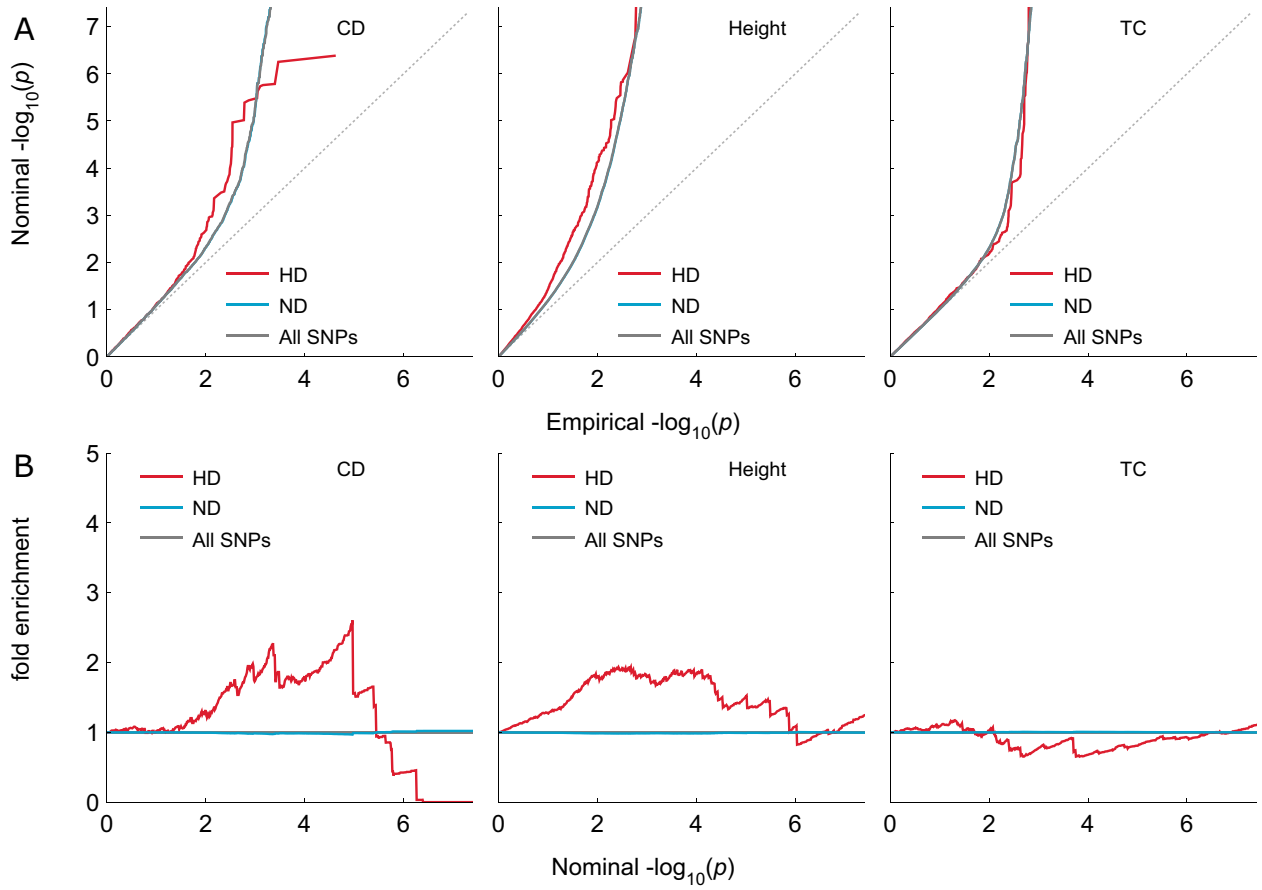
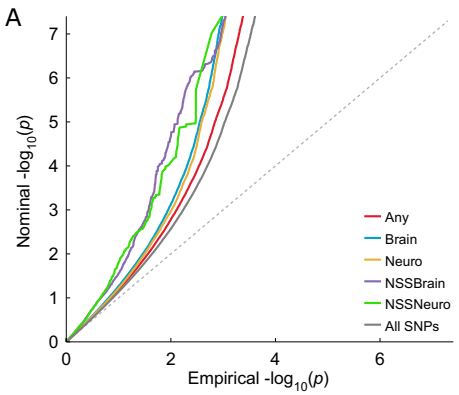


Figure 2. Quantile-quantile (Q-Q) and fold enrichment plots of three nonschizophrenia phenotypes (Crohn's disease [CD], height, and total cholesterol [TC]) stratified according to Neanderthal selective sweep scores. **(A)** The Q-Q plots show genome-wide association study summary statistics *p* values of single nucleotide polymorphisms (SNPs) tagging human divergent (HD) regions, comprising SNPs with negative Neanderthal selective sweep scores, nondivergent (ND) regions, comprising SNPs with positive Neanderthal selective sweep scores, and all SNPs. There is no indication of enrichment as seen in schizophrenia in Figure 1. **(B)** The fold enrichment counterparts of the Q-Q plots illustrate the lack of enrichment. However, the regression analysis shows significant enrichment for height (Table 1).



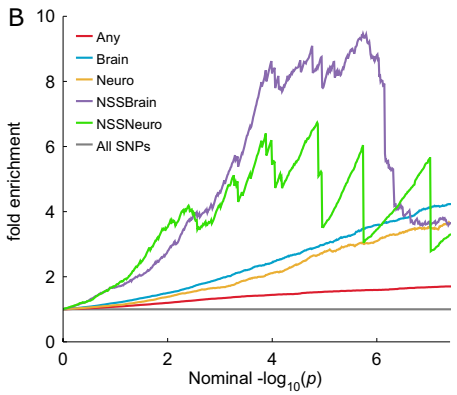


Figure 3. Quantile-quantile and fold enrichment plots showing schizophrenia association enrichment of brain genes with negative Neanderthal selective sweep (NSS) score. **(A)** Quantile-quantile and **(B)** fold enrichment plots are shown for single nucleotide polymorphisms (SNPs) annotated to generic genes (Any); SNPs annotated to genes associated with the brain, as established by a National Center for Biotechnology Information site search (Brain); SNPs annotated to genes associated with the brain, defined by Kang *et al*. (35) (Neuro); SNPs with negative NSS score and annotated to genes associated with the brain, as established by a National Center for Biotechnology Information site search (NSSBrain); SNPs with negative NSS score and annotated to genes defined by Kang *et al*. (35) (NSSNeuro); and all SNPs (All SNPs). The NSS Brain category is enriched (deflected left) compared with the other categories (i.e., presents a higher incidence of associations [lower *p* values] with schizophrenia). This is confirmed by the binomial proportion test comparing Brain and NSS Brain groups ($p = 5.5 \times 10^{-3}$).

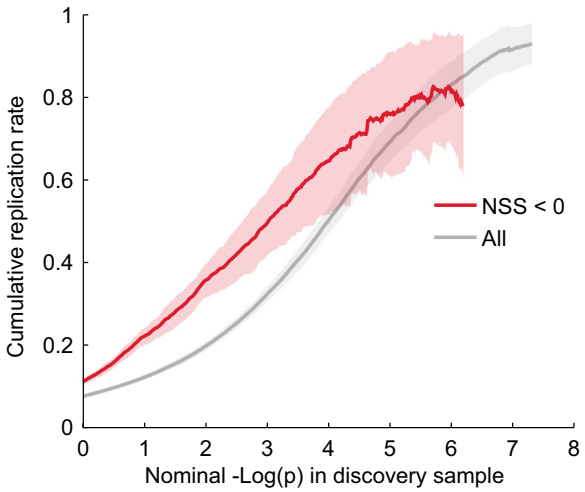


Figure 4. Replication plot for schizophrenia with and without conditioning on Neanderthal selective sweep (NSS) score < 0. Single nucleotide polymorphisms (SNPs) with negative NSS score tend to replicate better than baseline SNPs across the 52 schizophrenia meta-analysis substudies. For example, at a $-\log(p)$ value level of 4, the cumulative replication rate improves from 60% to about 80% when restricting the choice to SNPs with a negative NSS score. A negative NSS score seems to be a viable aid to identify nonspurious schizophrenia associations.

previously (28,29,32,36). Our results indicate that many schizophrenia susceptibility factors in modern humans may have emerged after their divergence from Neanderthals.

Several of the genes found that were likely to have undergone positive selection in modern humans (21) are involved in cognitive functions. The enrichment of SNP associations observed for schizophrenia may be due to an overlap between swept genomic regions and brain and other central nervous system genes and the regulatory regions thereof. This question is addressed by the regression analysis in which protein coding annotations are accounted for (Table 1). Even the inclusion of brain genes annotation scores in the regression did not reduce the enrichment for schizophrenia. However, among the brain genes themselves, the ones with a negative NSS score were more enriched of associations with schizophrenia compared with other brain genes, let alone just any genes (Figure 3).

The loci identified by the conditional FDR analysis harbor many genes that could plausibly play a role in the etiology of schizophrenia. Genes such as *DPYD*, *ZNF804A*, *NRXN1*, *NRG3*, and *VRK2*, which were previously known to be associated with schizophrenia, confirm its potentially evolutionary nature (51–53). Other interesting patterns emerge from genes such as *AGBL4*, *CEP170*, *IFT81*, and *SDCCAG8*, related to ciliogenesis and ciliary disorders (54–56), and *DPP10* and *FOXP1*, related to autism (57,58). The functional implications of the current associations based on tag SNPs need to be explored in experimental studies. It will be of interest at a later stage to investigate whether the current polygenic evolutionary signal in schizophrenia is associated with human specific brain structure variance. However, GWASs for relevant brain structures are not yet adequately powered ($N = \sim$15,000–21,000 participants) (59). Furthermore, the interplay of the polygenic effects and de novo mutations, such as schizophrenia risk copy number variations, should be examined, even if the latter appear to explain a very small proportion of schizophrenia cases (60).

The enrichment found here seems to be related to schizophrenia and some anthropomorphic human traits. However, we cannot rule out that there also may be enrichment in other disorders or diseases. The sample sizes available to some of the central nervous system GWAS might have limited the power to detect any enrichment. At any rate, the analysis of the smaller schizophrenia GWAS from 2013 (40) also revealed a nominally significant enrichment effect, further supporting the notion of a specific association between schizophrenia and positive selection.

In conclusion, the present findings of a prevalence of schizophrenia risk loci overlapping with some genetic signatures of human evolution support the argument that the emergence and the persistence of schizophrenia are connected to the human sapientia. This may help to explain the "evolutionary enigma" of schizophrenia.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the Research Council of Norway Grant Nos. 213837, 223273, and 225989 and South-East Norway Health Authority Grant No. 2013-123. The funding agencies had no role in the conception, design of the study, or collection of data. The iSelect chips were funded by the French National Foundation on Alzheimer's Disease and Related Disorders. European Alzheimer's Disease Initiative (EADI) was supported by the Laboratory of Excellence Program Investment for the Future Development of Innovative Strategies for a Transdisciplinary Approach to Alzheimer's Disease grant, INSERM, Institut Pasteur de Lille, Université de Lille 2, and the Lille University Hospital. The Genetic and Environmental Risk in Alzheimer's Disease consortium (GERAD) was supported by Medical Research Council Grant No. 503480, Alzheimer's Research UK Grant No. 503176, Wellcome Trust Grant No. 082604/2/07/Z, and German Federal Ministry of Education and Research: Competence Network Dementia Grant Nos. 01GI0102, 01GI0711, and 01GI0420. Cohorts for Heart and Aging Research in Genomic Epidemiology consortium (CHARGE) was partly supported by National Institutes of Health/National Institute on Aging Grant No. R01 AG033193 and National Institute on Aging Grant No. AG081220 and Age, Gene/Environment Susceptibility Study (AGES) contract N01-AG-12100, National Heart, Lung, and Blood Institute Grant No. R01 HL105756, Icelandic Heart Association, and Erasmus Medical Center and Erasmus University. The Alzheimer's Disease Genetics Consortium was supported by National Institutes of Health/National Institute on Aging Grant Nos. U01 AG032984, U24 AG021886, and U01 AG016976, and Alzheimer's Association Grant No. ADGC-10-196728.

We thank Dr. Tobias Kaufmann and Dr. Tuomo Mäki-Marttunen for technical advice and helpful support. We thank Cisca Wijmenga and David van Heel for access to celiac disease data. We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within the IGAP contributed to the design and implementation of the IGAP and/or provided data but did not participate in analysis or writing of this article. The IGAP was made possible by the generous participation of the control subjects, the patients, and their families.

FB, SD, and OAA designed the study. PGC, IHGC, BSW, J-AZ, SD, IM, TW, DAC, OAA, and MM provided data. MM, YW, AW, AJS, RSD, WKT, VZ, and AMD provided analytical tools or support. FB, SS, MM, and AW performed the analyses. SS, FB, and OAA wrote the first draft of the paper. All authors commented on and approved the final manuscript.

The members of the Schizophrenia Working Group of the Psychiatric Genomics Consortium and The International Headache Genetics Consortium are listed in the Supplement.

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

ARTICLE INFORMATION

From the Norwegian Centre for Mental Disorders Research (NORMENT) (SS, FB, MM, YW, AW, VZ, IM, SD, OAA), KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine (BSW, J-AZ), University of Oslo; Division of Mental Health and Addiction (SS, FB, MM, YW, AW, VZ, IM, SD, OAA) and Department of Medical Genetics (SD), Oslo University Hospital, Oslo, Norway; Multimodal Imaging Laboratory (YW, AJS, RSD, AMD), Cognitive Sciences Graduate Program (AJS), Center for Human Development (AJS), Department of Psychiatry (WKT, AMD, OAA), and Department of Neuroscience (AMD), University of California, San Diego, La Jolla, California; Centre for Molecular Medicine Norway (VZ), Nordic European Molecular Biology Laboratory Partnership, University of Oslo and Oslo University Hospital; Department of Neurology and Communication and Research Unit at the Division of Neuroscience (BSW, J-AZ), Oslo University Hospital, Ullevål, Oslo, Norway; Eli Lilly and Company (DAC, AMD), Windlesham, Surrey, United Kingdom; Neuroradiology Section (RSD), Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California; Institute of Biological Psychiatry (TW), Mental Health Center St. Hans, Mental Health Services Copenhagen, Roskilde; Department of Clinical Medicine (TW), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen; and The Lundbeck Foundation Initiative for Integrative Psychiatric Research (TW), Copenhagen, Capital Region of Denmark, Denmark.

SS and FB contributed equally to this work.

Address correspondence to Ole A. Andreassen, M.D., Ph.D., NORMENT, KG Jebsen Centre for Psychosis Research, Oslo University Hospital, Ullevål, Kirkeveien 166, PO Box 4956 Nydalen, 0424 Oslo, Norway; E-mail: o.a.andreassen@medisin.uio.no.

Received Mar 18, 2015; revised Sep 30, 2015; accepted Oct 5, 2015.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2015.10.009.

REFERENCES

1. Allen JS, Sarich VM (1988): Schizophrenia in an evolutionary perspective. Perspect Biol Med 32:132–153.
2. Crow TJ (1995): A Darwinian approach to the origins of psychosis [published erratum appears in Br J Psychiatry 1995 Sep;167:414]. Br Psychiatry 167:12–25.
3. Burns JK (2006): Psychosis: A costly by-product of social brain evolution in Homo sapiens. Prog Neuropsychopharmacol Biol Psychiatry 30:797–814.
4. Polimeni J, Reiss JP (2003): Evolutionary perspectives on schizophrenia. Can J Psychiatry 48:34–39.
5. Lewis DA, Hashimoto T, Volk DW (2005): Cortical inhibitory neurons and schizophrenia. Nat Rev Neurosci 6:312–324.
6. Brune M (2004): Schizophrenia—an evolutionary enigma? Neurosci Biobehav Rev 28:41–53.
7. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, *et al*. (2009): Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. Lancet 373:234–239.
8. Power RA, Kyaga S, Uher R, MacCabe JH, Langstrom N, Landen M, *et al*. (2013): Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry 70:22–30.
9. Uher R (2009): The role of genetic variation in the causation of mental illness: An evolution-informed framework. Mol Psychiatry 14: 1072–1082.
10. Pearlson GD, Folley BS (2008): Schizophrenia, psychiatric genetics, and Darwinian psychiatry: An evolutionary framework. Schizophr Bull 34:722–733.
11. Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper J, *et al*. (1992): Schizophrenia: Manifestations, incidence and course in different cultures. A World Health Organization Ten-Country Study. Psychol Med Monogr Suppl 20:1–97.
12. Huxley J, Mayr E, Osmond H, Hoffer A (1964): Schizophrenia as a genetic morphism. Nature 206:220–221.
13. Yang J, Visscher PM, Wray NR (2010): Sporadic cases are the norm for complex disease. Eur J Hum Genet 18:1039–1043.
14. Stringer C (2002): The morphological and behavioral origins of modern humans. In: Crow TJ, editor. The Speciation of Modern Homo Sapiens. Oxford: Oxford University Press.
15. Hublin JJ, Spoor F, Braun M, Zonneveld F, Condemi S (1996): A late Neanderthal associated with Upper Paleolithic artefacts. Nature 381: 224–226.
16. d'Errico F, Henshilwood C, Lawson G, Vanhaeren M, Tillier AM, Soressi M, *et al*. (2003): Archaeological evidence for the emergence of language, symbolism, and music—an alternative multidisciplinary perspective. J World Prehist 17:1–70.
17. Crow TJ (2000): Schizophrenia as the price that Homo sapiens pays for language: A resolution of the central paradox in the origin of the species. Brain Res Rev 31:118–129.
18. Wynn T, Coolidge FL (2011): The implications of the working memory model for the evolution of modern cognition. Int J Evol Biol 2011:741357.
19. Crow TJ (1997): Is schizophrenia the price that Homo sapiens pays for language? Schizophr Res 28:127–141.
20. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
21. 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.
22. Prufer 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.
23. Wagner GP, Zhang J (2011): The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. Nat Rev Genet 12:204–213.
24. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, *et al*. (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752.
25. Efron B (2010): Large-scale inference: Empirical Bayes methods for estimation, testing, and prediction. Cambridge: Cambridge University Press.
26. 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.
27. Liu JZ, Hov JR, Folseraas T, Ellinghaus E, Rushbrook SM, Doncheva NT, *et al*. (2013): Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet 45:670–675.
28. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, *et al*. (2013): Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. PLoS Genet 9:e1003455.
29. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, *et al*. (2013): Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet 92:197–209.
30. Andreassen OA, Zuber V, Thompson WK, Schork AJ, Bettella F, PRACTICAL Consortium, *et al*. (2014): Shared common variants in prostate cancer and blood lipids. Int J Epidemiol 43:1205–1214.
31. Andreassen OA, McEvoy LK, Thompson WK, Wang Y, Reppe S, Schork AJ, *et al*. (2014): Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. Hypertension 63:819–826.
32. Andreassen OA, Harbo HF, Wang Y, Thompson WK, Schork AJ, Mattingsdal M, *et al*. (2015): Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: Differential involvement of immune-related gene loci. Mol Psychiatry 20:207–214.
33. Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, *et al*. (2010): Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328:723–725.

34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, *et al*. (2002): The Human Genome Browser at UCSC. Genome Res 12: 996–1006.
35. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, *et al*. (2011): Spatio-temporal transcriptome of the human brain. Nature 478:483–489.
36. Andreassen OA, Thompson WK, Dale AM (2014): Boosting the power of schizophrenia genetics by leveraging new statistical tools. Schizophr Bull 40:13–17.
37. ENCODE Project Consortium (2004): The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306:636–640.
38. Ryan TP (2007): Modern Engineering Statistics. Hoboken, NJ: Wiley-Interscience.
39. Zablocki RW, Schork AJ, Levine RA, Andreassen OA, Dale AM, Thompson WK (2014): Covariate-modulated local false discovery rate for genome-wide association studies. Bioinformatics 30:2098–2104.
40. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, *et al*. (2013): Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 45:1150–1159.
41. Kamm GB, Pisciottano F, Kliger R, Franchini LF (2013): The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. Mol Biol Evol 30: 1088–1102.
42. Harrison PJ, Owen MJ (2003): Genes for schizophrenia? Recent findings and their pathophysiological implications. Lancet 361:417–419.
43. Bigdeli TB, Fanous AH, Riley BP, Reimers M, Schizophrenia Psychiatric Genome-Wide Association Study Consortium, Chen X, *et al*. (2013): On schizophrenia as a "disease of humanity". Schizophr Res 143:223–224.
44. 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.
45. Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, *et al*. (2013): Evidence for polygenic adaptation to pathogens in the human genome. Mol Biol Evol 30:1544–1558.
46. Perry GH, Foll M, Grenier JC, Patin E, Nedelec Y, Pacis A, *et al*. (2014): Adaptive, convergent origins of the pygmy phenotype in African rainforest hunter-gatherers. Proc Natl Acad Sci U S A 111:E3596–3603.
47. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, *et al*. (2012): Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat Genet 44:1015–1019.
48. Fu W, Akey JM (2013): Selection and adaptation in the human genome. Annu Rev Genomics Hum Genet 14:467–489.
49. Burger R, Gimelfarb A (1999): Genetic variation maintained in multilocus models of additive quantitative traits under stabilizing selection. Genetics 152:807–820.
50. Pritchard JK, Di Rienzo A (2010): Adaptation—not by sweeps alone. Nat Rev Genet 11:665–667.
51. Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR, *et al*. (2005): DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. Science 310:1187–1191.
52. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, *et al*. (2012): De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet 44:1365–1369.
53. Cacabelos R, Martínez-Bouza R (2011): Genomics and pharmacogenomics of schizophrenia. CNS Neurosci Ther 17:541–565.
54. Guarguaglini G, Duncan PI, Stierhof YD, Holmström T, Duensing S, Nigg EA (2005): The forkhead-associated domain protein Cep170 interacts with polo-like kinase 1 and serves as a marker for mature centrioles. Mol Biol Cell 16:1095–1107.
55. Lucker BF, Behal RH, Qin H, Siron LC, Taggart WD, Rosenbaum JL, *et al*. (2005): Characterization of the intraflagellar transport complex B core: Direct interaction of the IFT81 and IFT74/72 subunits. J Biol Chem 280:27688–27696.
56. Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, *et al*. (2010): Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. Nat Genet 42:840–850.
57. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, *et al*. (2013): Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet 92:221–237.
58. Bacon C, Schneider M, Le Magueresse C, Froehlich H, Sticht C, Gluch C, *et al*. (2015): Brain-specific Foxp1 deletion impairs neuronal development and causes autistic-like behaviour. Mol Psychiatry 20: 632–639.
59. Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, *et al*. (2012): Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 44: 552–561.
60. Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, *et al*. (2008): Large recurrent microdeletions associated with schizophrenia. Nature 455:232–236.