

## Polygenic selection underlies evolution of human brain structure and behavioral traits

Evan R. Beiter<sup>1†</sup>, Ekaterina A. Khramtsova<sup>2,3†</sup>, Celia Van Der Merwe<sup>4</sup>, Emile R. Chimusa<sup>5</sup>, Corinne Simonti<sup>6</sup>, Jason Stein<sup>7</sup>, Paul Thompson<sup>8</sup>, Simon E. Fisher<sup>9,10</sup>, Dan J. Stein<sup>4</sup>, John A. Capra<sup>6,11</sup>, James A. Knowles<sup>12</sup>, Barbara E. Stranger<sup>2,3,14\*§</sup>, Lea K. Davis<sup>6,15,16\*§</sup>

- 1) Department of Biology, Washington University, St. Louis, MO
- 2) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL
- 3) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL
- 4) Department of Psychiatry and MRC Unit on Risk & Resilience in Mental Disorders, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
- 5) Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa
- 6) Vanderbilt Genetics Institute; Vanderbilt University Medical Center, Nashville, TN
- 7) Department of Genetics & UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC
- 8) Imaging Genetics Center, Stevens Institute for Neuroimaging & Informatics, Keck USC School of Medicine, Los Angeles, CA
- 9) Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
- 10) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
- 11) Department of Biological Sciences, Vanderbilt University, Nashville, TN

- 12) Department of Cell Biology, State University of New York Downstate Medical Center, Brooklyn, New York
- 13) Division of Medical Genetics, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
- 14) Center for Data Intensive Science, University of Chicago, Chicago, IL
- 15) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
- 16) Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN

† These authors contributed equally

\* Corresponding author

§ These authors contributed equally

## **Abstract**

Seemingly paradoxical characteristics of psychiatric disorders, including moderate to high prevalence, reduced fecundity, and high heritability have motivated explanations for the persistence of common risk alleles for severe psychiatric phenotypes throughout human evolution. Proposed mechanisms include balancing selection, drift, and weak polygenic adaptation acting either directly, or indirectly through selection on correlated traits. While many mechanisms have been proposed, few have been empirically tested. Leveraging publicly available data of unprecedented sample size, we studied twenty-five traits (i.e., ten neuropsychiatric disorders, three personality traits, total intracranial volume,

seven subcortical brain structure volume traits, and four complex traits without neuropsychiatric associations) for evidence of several different signatures of selection over a range of evolutionary time scales. Consistent with the largely polygenic architecture of neuropsychiatric traits, we found no enrichment of trait-associated single-nucleotide polymorphisms (SNPs) in regions of the genome that underwent classical selective sweeps (i.e., events which would have driven selected alleles to near fixation). However, we discovered that SNPs associated with some, but not all, behaviors and brain structure volumes are enriched in genomic regions under selection since divergence from Neanderthals ~600,000 years ago, and show further evidence for signatures of ancient and recent polygenic adaptation. Individual subcortical brain structure volumes demonstrate genome-wide evidence in support of a mosaic theory of brain evolution while total intracranial volume and height appear to share evolutionary constraints consistent with concerted evolution. We further characterized the biological processes potentially targeted by selection, through expression Quantitative Trait Locus (eQTL) and Gene Ontology (GO) enrichment analyses and found evidence for the role of regulatory functions among selected SNPs in immune and brain tissues. Taken together, our results suggest that alleles associated with neuropsychiatric, behavioral, and brain volume phenotypes have experienced both ancient and recent polygenic adaptation in human evolution, acting through neurodevelopmental and immune-mediated pathways.

## Introduction

Common (minor allele frequency [MAF] > 5%) single-nucleotide polymorphisms (SNPs) account for a substantial proportion of the inter-individual variation in psychiatric disorders and other brain-related phenotypes within human populations<sup>1-14</sup>. Psychiatric disorders are often diagnosed before child-bearing years, and exhibit moderate to high prevalence worldwide, reduced fecundity<sup>15</sup>, and high heritability<sup>16</sup>. This raises a fundamental question of why risk alleles have persisted through the evolutionary history of our species<sup>17-19</sup>. One proposed explanatory mechanism is balancing selection, which can increase frequencies of risk alleles through an advantage conferred to heterozygotes, as observed in sickle cell anemia and malarial resistance<sup>20</sup>. Consistent with this hypothesis, siblings of individuals diagnosed with some neuropsychiatric disorders have significantly increased fecundity relative to the general population<sup>15</sup>. An alternative explanation is that positive selection on genes influencing largely beneficial traits may have indirectly influenced psychiatric disorders through pleiotropic mechanisms<sup>21-23</sup>. Another possibility that has been discussed in the literature is ancestral neutrality i.e., that reduced fecundity is only a modern phenomenon<sup>24</sup>. For example, the “shaman theory” proposes that individuals with schizophrenia or other severe mental illness may have been considered endowed with mystical powers and thus not subject to modern stigma.<sup>24</sup> A more likely scenario, perhaps, is that women with developmental delay or psychiatric illness may have been (and continue to be) more vulnerable to rape<sup>25-27</sup> than women without psychiatric illness. Still others suggest that recent selection on genes involved in cognitive

development may have over-selected the genes pre-disposing humans to mental disorders<sup>28, 29</sup>. Finally, according to a theory known as polygenic mutation-selection balance, proposed in 2006 by Keller and Miller, psychiatric illness reflects a small mutational load dispersed across thousands of genes<sup>24</sup>. Clearly, while many potential mechanisms have been hypothesized for risk variants maintaining a high allele frequency across evolutionary time, few have been empirically tested. Few genome-wide studies of adaptive selection have been reported for neuropsychiatric traits to date, thus motivating this study<sup>24, 28, 30</sup>.

Paleobiological evidence indicates that the size of the human skull has expanded massively over the last 200,000 years, likely mirroring increases in brain size. It is thought that these structural changes in human brain size precipitated our uniquely human cognitive capabilities. The observed evolutionary increase in human skull size, however, is not sufficient to explain how individual brain structures have changed over evolutionary time. Our recent insights into the genetic contribution to total intracranial volume<sup>13, 31, 32</sup> and to subcortical brain structure volumes now provides an opportunity to look backwards in time and infer the history of brain structure during the evolution of humankind. Two prominent evolutionary theories, the “concerted” and “mosaic” theories of human brain evolution, may offer insight into the evolution of behavioral traits in our species. The concerted theory states that human brains evolved in concert with changes in body size and share developmental pathways that allometrically constrain the evolution of the brain<sup>33</sup>. The mosaic theory, asserts that different brain structures are able to vary in size and evolve independently of one another

<sup>34-36</sup>. Importantly, these theories are not mutually exclusive and there is evidence for both concerted and mosaic evolution of the brain across species <sup>37, 38</sup>. Both theories carry implications for the evolutionary relationship between behavior and human brain development. For example, if evolution of the human brain is primarily a function of allometric scaling driven by genetic correlations with body size, behavior may have a more limited role in brain evolution. However, if behavioral traits are a target of selection, genetic correlation between those traits and individual brain structures could provide a mechanism to allow individual brain structures to adapt independently to a variety of evolutionary pressures. Finally, analysis of the signatures of selection across genetic variants associated with brain structure volumes may provide genetic and neurobiological insights into uniquely human capabilities.

Given recent advances in characterizing the genetic architecture of brain and behavioral phenotypes <sup>30, 32, 39-52</sup> and analytic advances in detecting polygenic adaptation <sup>53, 54</sup>, here we sought to systematically evaluate evidence for selection across a total of 25 complex traits including ten neuropsychiatric disorders, three personality traits, total intracranial volume, seven subcortical brain structure volume traits, and for comparison, four complex traits with no known neuropsychiatric associations (Supplementary Table 1). We tested summary statistics from publicly-available genome-wide association studies (GWAS) for multiple signatures of selection across different timescales of human evolutionary history, beginning with our divergence from the lineage that led to Neanderthals, estimated at ~600 kya. First, we examined trait-associated SNPs for

enrichment in regions of the human genome that are depleted of Neanderthal alleles and hence thought to indicate positively selected regions since divergence from Neanderthals (i.e., Neanderthal selective sweep score, NSS)<sup>55</sup>. Second, we tested for evidence of “soft selective sweeps” occurring up to 150 kya in which an adaptive mutation increases in frequency, but does not reach fixation, as indicated by global population differentiation (i.e., high global  $F_{st}$ , a metric of population differentiation due to genetic structure)<sup>56</sup>. Third, we tested for classical “hard sweeps” occurring ~25-30 kya, in which a new adaptive mutation on a single haplotype quickly rises to fixation (or near fixation), leaving behind an identifiable signature in the form of reduced genetic diversity at the haplotype locus and longer than expected haplotype lengths (i.e., integrated haplotype scores, iHS)<sup>57-59</sup>. Scans for signatures of hard sweeps have detected genes associated with traits such as lactase persistence<sup>60</sup>, malaria resistance<sup>61</sup> and Glucose-6-phosphate dehydrogenase enzyme activity<sup>62</sup>. Fourth, we employed an approach developed by Berg and Coop (2014) to detect weak positive selection that acts on standing variation (i.e., polygenic adaptation involving alleles at many different loci) and leaves behind a signature of increased covariance in trait-associated allele frequencies across 52 global populations compared to a model of genetic drift<sup>54, 63</sup>. This approach is powered to detect polygenic adaptation approximately 30 kya. Fifth, we applied the Singleton Density Score (SDS) test, to investigate the role of very recent polygenic adaptation, occurring up to 2 kya. The SDS test allows detection of variants experiencing very recent positive selection because they have rapidly increased in frequency - giving the surrounding haplotypes little time to accumulate new mutations (i.e.,

singletons). Thus, for phenotypes evolving under recent positive polygenic adaptation, we expect to observe a reduced number of singleton events on the selected haplotype in reference to sequencing panels <sup>54</sup>.

Typical effect sizes of common risk alleles for complex traits are small, so we hypothesized that there would be limited evidence of classical hard sweeps (typically targeting SNPs with large selection coefficients) on variants associated with these traits. Rather, we reasoned that if positive selection has played any role in maintaining risk allele frequencies for neuropsychiatric and brain structure volume traits, it would produce a distributed polygenic signal throughout the genome across many weakly selected variants. Furthermore, signatures of selection that differ across brain structure volumes would provide evidence in favor of mosaic brain evolution, while strong genetic correlations and consistent patterns of polygenic adaptation would provide evidence for concerted evolution.

Finally, recent studies have shown that several neuropsychiatric traits are genetically correlated with metabolic and immune-mediated phenotypes providing potential mechanisms through which pleiotropic evolutionary adaptation may influence trait-associated allele frequencies <sup>64, 65</sup>. For example, immune genes show evidence of pleiotropy in early neuronal development among humans and animal models <sup>66</sup> and the immune system has been a target of positive selection as humans have colonized land masses across the world, experiencing various environmentally-mediated immune challenges <sup>63, 67</sup>. Therefore, to characterize the potential biological targets of selection we

analyzed the biological functions of trait-associated SNPs for those phenotypes showing evidence of polygenic adaptation through studies of expression Quantitative Trait Loci (eQTL) detected in relevant tissues, including brain regions (e.g., frontal lobe, cerebellum, etc.) and immune tissues (e.g., monocytes, CD4+ cells, and whole blood) and gene set enrichment. In summary, we systematically assessed multiple signatures of selection across a comprehensive set of brain and behavior phenotypes and further investigated potential biological mechanisms underlying significant evidence of adaptation.

## Methods

### *Summary Statistics from Genome-Wide Association Studies*

GWAS summary statistics for attention deficit hyperactivity disorder (ADHD)<sup>1</sup>, Alzheimer's disease (ALZ)<sup>2</sup>, anorexia nervosa (AN)<sup>3</sup>, general anxiety disorder (GAD)<sup>4</sup>, autism spectrum disorder (ASD)<sup>68</sup>, bipolar disorder (BIP)<sup>5</sup>, major depressive disorder (MDD)<sup>7</sup>, obsessive compulsive disorder (OCD)<sup>69</sup>, Tourette Syndrome (TS)<sup>8, 70</sup>, schizophrenia (SCZ)<sup>52</sup>, extraversion (EXT)<sup>6, 71</sup>, neuroticism (NEO)<sup>71, 72</sup>, subjective well-being (SWB)<sup>14</sup>, type 2 diabetes (T2D)<sup>10</sup>, inflammatory bowel disease (IBD)<sup>9</sup>, height (HEI)<sup>11</sup>, body mass index (BMI)<sup>12</sup>, total intracranial volume (ICV)<sup>32</sup>, total intracranial volume controlled for height (ICV|height)<sup>32</sup> and seven separate subcortical brain regions (the nucleus accumbens<sup>13</sup>, amygdala<sup>13</sup>, caudate nucleus<sup>13</sup>, pallidum<sup>13</sup>, putamen<sup>13</sup>, thalamus<sup>13</sup> and hippocampus<sup>31</sup>) were obtained from consortium websites or directly from study authors (Supplementary Table 1). To protect research participants from re-

identification in summary GWAS data, many consortia replace the true allele frequencies with allele frequencies drawn from the 1000 Genomes Project or HapMap, depending on author preference. We therefore used a common reference and determined the allele frequencies across phenotypes by calculating each within the 1000 Genomes European population. For each phenotype, the analysis was either restricted to European populations or included a very small number of non-European cases and controls that were meta-analyzed with a majority of European samples. Integrated haplotype scores (iHS) <sup>59</sup> and global  $F_{st}$  scores <sup>56</sup> were downloaded from the 1000 Genomes Selection Browser 1.0 <sup>73</sup>. SNPs modestly associated with each trait ( $p < 5.0 \times 10^{-3}$ ) were chosen for subsequent analysis. To test the robustness of the results against varying significance thresholds, we also tested a subset of SNPs exceeding a more stringent p-value threshold ( $p < 5.0 \times 10^{-4}$ ). Trait-associated SNPs were clumped using an  $R^2$  threshold of 0.25, within a 500Kb window. Again, to determine the robustness of results against varying LD-pruning thresholds, we also generated a more stringently clumped set of SNPs for each phenotype ( $r^2 < 0.10$ , 1000Kb window). We retained the most significantly associated SNP within each LD block determined by these parameters. These sets of clumped, nominally trait-associated SNPs were then utilized in enrichment and polygenic analyses.

### *Enrichment of variants in regions of the genome under selection since Neanderthal divergence*

Each set of trait-associated SNPs was annotated with Neanderthal selective sweep

scores (NSS-scores) <sup>55</sup>. Extreme negative NSS-scores ( $< -4.32$ , 5% tail of genome-wide distribution) indicate regions of the human genome that are depleted of derived Neanderthal alleles and hypothesized to have experienced selective sweeps in modern human ancestors occurring since the time of human divergence from Neanderthals <sup>55</sup>. To test for enrichment of low NSS-scores among these top SNP associations, 500 sets of null matched SNPs were generated using SNPsnap <sup>74</sup>. For each phenotype, the 500 SNP sets were sampled without replacement from the European catalogue of 1000 Genomes SNPs. Randomly chosen SNPs were matched to the trait-associated SNPs by minor allele frequency ( $\pm 3\%$ ), gene density ( $\pm 50\%$ ), and distance to nearest gene ( $\pm 50$  kb) for each phenotype. An empirical enrichment p-value was calculated representing the proportion of random, matched SNP sets in which the number of SNPs within the Neanderthal-derived depleted regions matched or exceeded the actual number of observed trait-associated SNPs within the Neanderthal-derived depleted regions for each phenotype.

#### *Enrichment of signatures of selective sweeps among trait associated SNPs*

To test for enrichment of extreme  $iHS$  or  $F_{st}$  scores among these top SNP associations, 500 sets of randomly ascertained SNPs were generated using SNPsnap <sup>74</sup>. The same general enrichment approach used for NSS-scores was employed with one modification. SNPs of Neanderthal ancestry have demonstrated high  $iHS$  due to the extended LD caused by recent introgression <sup>75</sup>. Therefore, to avoid confounding signatures of positive selection and introgression, SNPs located in regions of Neanderthal

introgression were excluded from both trait-associated SNPs and the matching SNP sets<sup>76</sup>.<sup>77</sup> Because integrated haplotype scores (i.e., our tests of “hard sweeps”) are based on detection of haplotype lengths that are longer than expected, we also did not match SNPs on LD parameters, to avoid overmatching the randomly chosen SNP sets.

Each trait-associated SNP set was then tested for enrichment of high iHS scores at two thresholds ( $|iHS| > 2.0$  and  $|iHS| > 2.6$ ) and extreme global population differentiation at two thresholds ( $F_{st} > 0.30$  and  $F_{st} > 0.56$ ), representing the most extreme 5% and 1%, respectively, of both scores across the human genome. An empirical enrichment p-value was calculated as the proportion of random matched SNP sets in which the percentage of extreme scores (i.e., iHS or  $F_{st}$ , respectively) match or exceed the actual observed percentage of extreme scores in the set of trait-associated SNPs. (We summarize the enrichment analysis workflow in Supplementary Figure 1).

### *Polygenic Adaptation Analysis*

The polygenic adaptation detection methodology of Berg and Coop (2014) was used to test whether estimated genetic values (i.e., linear weighted sum of risk or trait-increasing alleles) exhibit more covariance among populations than expected due to genetic drift, tested empirically with a statistic termed the “Qx score”. For all twenty-one brain-related phenotypes, SNPs exceeding the nominal association threshold ( $p < 5.0 \times 10^{-3}$ ) were included in the analysis (Table 2). For non-neuropsychiatric traits (i.e., height, body mass index, type 2 diabetes, and inflammatory bowel disease) in which publicly-available

GWAS results report analysis of hundreds of thousands of individuals, we tested more stringent association thresholds ( $p < 5.0 \times 10^{-4}$ ,  $p < 5.0 \times 10^{-6}$ , or  $p < 5.0 \times 10^{-8}$ ) due to the computational burden of including nominally associated SNPs from these well-powered studies. For each phenotype, in each of 52 world populations, the genetic value was calculated as the linear sum of the number of risk (or trait-increasing) alleles in the population weighted by the effect size from the discovery GWAS (Supplementary Table 1).

In brief,  $Q_x$  scores quantify the variance and covariance of allele frequencies across 52 human populations to detect the over-dispersion of risk alleles across populations in comparison to an empirical model of neutral drift taking into account shared ancestry. To create the model of neutral drift, a null distribution was created for each set of trait-associated SNPs, composed of 1000 sets of SNPs matched to the trait-associated SNPs by MAF (2% bins) and B-value, a metric of neutral genetic drift (100 bins)<sup>78</sup>. Any SNP that was fixed in all 52 populations (or in the French European population) or was not part of the Human Genome Diversity Panel dataset was excluded from analysis (Supplementary Figure 1). The significance of the trait-associated SNP  $Q_x$  statistic was determined empirically as described above.

To ensure that the model of genetic drift (conditioned based on ancestral allele frequencies) is unbiased, we compared the genetic values for each of the 52 populations to their  $F_3$  statistic described by Patterson et al (2012)<sup>79</sup> in the form of:

$$F_3(\text{French}; Y, \text{San}) = \left(\frac{1}{L}\right) * \sum_i^L [(p_{i\text{French}} - p_{i\text{San}}) * (p_{iY} - p_{i\text{San}})] \quad (\text{eq. 1})$$

in which  $p_i$  is the allele frequency at locus  $i$  and  $L = 20,000$  randomly ascertained SNPs (Equation 1). No phenotypes demonstrated any correlation between genetic value and the  $F_3$  statistic indicating that our results cannot be explained by a biased modeling of genetic drift (Supplementary Table 2; Supplementary Figure 3).

### *Singleton Density Score Analysis*

The singleton density score approach is based on the assumption that alleles under positive selection will rise in frequency faster than their haplotypes can accumulate new mutations. This results in an observable decrease in the density of singleton events on the selected haplotype measured by sequencing of large reference populations<sup>54</sup>. Scores for each SNP can be calculated in which haplotypes are annotated as ancestral and derived (i.e., singleton density scores or SDS) or calculated with respect to risk (or trait-increasing) and protective (or trait-decreasing) alleles (i.e., trait-singleton density scores or tSDS). To test whether polygenic positive selection is acting on risk or protective (trait-increasing or trait-decreasing) alleles we tested the global Spearman correlation between tSDS and strength of association from the GWAS. As in Field et al., 2016, each complete set of GWAS summary statistics was divided into 100 blocks of contiguous SNPs and a block jackknife resampling approach (i.e., leave-one-block-out in 100 iterations) was used to calculate the standard error while accounting for LD. A positive tSDS implies very recent selection (~2,000 years) on the risk (trait-increasing) allele while negative tSDS suggests

recent selection on the protective (trait-decreasing) allele.

### *Assessment of significance*

For each of the five signatures selection, empirical assessments of the significance of the test statistic observation were conducted as described in respective methods sections above. In each case, we accounted for the number of phenotypes tested and considered the result significant if it met or exceeded the Bonferroni correction for all 25 phenotypes tested ( $p < 0.002$ ).

### *Enrichment of expression quantitative trait loci in brain and immune tissues*

To investigate the biological functions of the trait-associated SNPs for the phenotypes with evidence of polygenic selection, we annotated SNPs with expression quantitative trait loci (eQTLs) information. We used previously published significant cis-eQTL results derived from 10 brain regions and whole blood<sup>80</sup> (<https://www.gtexportal.org/home/>), as well as CD4+ T cells (adaptive immunity) and CD14+ monocytes (innate immunity)<sup>81</sup>.

To assess eQTL enrichment, specifically, to test for an enrichment for a gene regulatory role for trait-associated SNPs, we quantified the enrichment of the number of eQTL target genes (eGenes) associated with trait-associated SNPs compared to random, matched SNPs. One thousand randomly ascertained SNP sets were generated for each phenotype using SNPsnap<sup>74</sup>. Matching sets were sampled without replacement from the

European catalogue of 1000 Genomes SNPs, and matched for physical distance within 500 kb, MAF ( $\pm 3\%$ ), gene density ( $\pm 50\%$ ), distance to nearest gene ( $\pm 100$  kb), and LD buddies ( $\pm 50\%$ ) at  $r^2=0.8$ . Trait-associated SNPs and matching SNP sets were annotated as cis-eQTLs, and the genes they are associated with, as eGenes, in various tissues. The number of eGenes in each matched set yielded an empirical distribution. The enrichment P-value was calculated as the proportion of randomized sets in which the eGene count matches or exceeds the actual observed count in the list of the trait-associated SNPs. If several SNPs implicated the same eGene in a tissue, the eGene was counted once. The enrichment was considered significant if it met Bonferroni multiple testing correction threshold  $p < 0.004$  (i.e.,  $0.05/13$  tissues investigated). Due to a high degree of eQTL sharing across tissues<sup>82</sup>, when enrichment was assessed in combined brain tissues and combined immune tissues, only the set of unique eGenes was counted. To exclude the possibility of eQTL enrichment overestimation due to the gene-rich MHC region, we performed eQTL enrichment analysis both including and excluding SNPs in the HLA region.

To evaluate underlying biological pathways that may be regulated by the eQTLs identified in the brain and immune tissues, we applied Gene Ontology (GO) enrichment analysis as implemented in Gene Set Enrichment Analysis (GSEA)<sup>83, 84</sup> using the eGene lists as the input genes, and the whole genome as the background.

#### *Proportion of ancestral and derived alleles among risk conferring SNPs*

The SNP ancestral allele annotations were derived from the Ensembl 59

comparative species alignment or were obtained from dbSNP. For each phenotype, we determined the proportion of risk alleles ( $p < 5.0 \times 10^{-3}$ ) that were ancestral or derived. Previous work has shown that derived alleles are more often minor alleles ( $< 50\%$  allele frequency) and more often associated with risk than ancestral alleles<sup>85</sup>. For each phenotype, we investigated the proportion of derived risk alleles across minor allele frequency bins. To this end, the risk alleles were categorized into 5 MAF bins, ( $>0.05-0.1$ ,  $>0.1-0.2$ ,  $>0.2-0.3$ ,  $>0.3-0.4$ ,  $>0.4-0.5$ ), and we independently computed the fractions of derived alleles in each bin.

## Results

*Variants associated with several brain phenotypes are enriched in regions of the genome under selection since Neanderthal divergence*

SNPs nominally associated with schizophrenia ( $p_{SCZ} = 0.002$ ) and neuroticism ( $p_{NEU} < 0.002$ ), were significantly enriched in regions hypothesized to have experienced positive selection early in modern human evolution (i.e., low NSS-score SNPs) compared to randomly ascertained matched SNPs (Methods and Table 1). SNPs nominally associated with total intracranial volume ( $p_{ICV} = 0.004$ ) and Type 2 Diabetes ( $p_{T2D} = 0.004$ ) trended toward significant enrichment, however the remaining phenotypes demonstrated no enrichment, and none of the phenotypes demonstrated depletion.

*No evidence for strong selective sweeps among SNPs associated with brain and behavior*

*phenotypes.*

None of the neuropsychiatric trait-associated SNPs exhibited enrichment of extreme  $iHS$  scores characteristic of hard sweeps of positive selection (Table 1). No phenotype demonstrated consistent, robust evidence for population differentiation. While SNPs associated with bipolar disorder and subjective well-being, respectively, showed significant evidence ( $p_{BPD} < 0.002$ ,  $p_{SWB} < 0.002$ ) of enrichment for global population differentiation at the less extreme  $F_{st}$  threshold ( $F_{st} > 0.30$ , representing the top 5% of differentiated SNPs in the genome), they demonstrated no enrichment at the more stringent  $F_{st}$  threshold ( $F_{st} > 0.56$ , representing the top 1% of differentiated SNPs in the genome). All analyses, including enrichment of extreme  $F_{st}$  scores, can be influenced by residual population structure that may remain even after extensive QC procedures are performed. To test whether our results were influenced by residual population stratification, we calculated LD-score regression intercepts and lambda values (Supplementary Table 1), which reflect the contribution of confounding bias due to cryptic relatedness and residual population stratification (i.e., scores close to 1 indicate minimal residual confounding). No significant evidence of residual population stratification or cryptic relatedness was observed.

#### *Schizophrenia associated loci exhibit evidence of ancient polygenic adaptation*

All traits were evaluated for evidence of ancient polygenic adaptation (approximately 30 kya) by testing each set of nominally associated SNPs for excess genetic

covariance, after accounting for background selection, compared to a null model of genetic drift. Schizophrenia ( $Q_X = 208.36$ ,  $p_{SCZ} < 0.001$ ) was the only neuropsychiatric disorder that provided statistically significant evidence of ancient polygenic adaptation (Figure 1, Table 2, Supplementary Figure 2). Among the personality traits, extraversion ( $Q_X = 88.04$ ,  $p_{EXT} < 0.001$ ) and subjective well-being ( $Q_X = 113.90$ ,  $p_{SWB} < 0.001$ ), but not neuroticism ( $Q_X = 77.52$ ,  $p_{NEU} = 0.156$ ) demonstrated significant evidence of ancient polygenic adaptation (Figure 1). Moreover, among the brain structure volume traits, hippocampus ( $Q_X = 108.98$ ,  $p_{HIP} < 0.001$ ) and putamen ( $Q_X = 115.76$ ,  $p_{PUT} < 0.001$ ) demonstrated significant evidence of ancient selection (Figure 1, Table 2, Supplementary Figure 2). For comparison, we also examined common complex phenotypes with no known neuropsychiatric component. Inflammatory bowel disease ( $Q_X = 101.13$ ,  $p_{IBD} < 0.001$ ), type 2 diabetes ( $Q_X = 117.77$ ,  $p_{T2D} < 0.001$ ), and height ( $Q_X = 303.29$ ,  $p_{HEI} < 0.001$ ) demonstrated significant evidence of polygenic adaptation (Table 2, Figure 1), consistent with previous reports<sup>53, 86-88</sup>.

When a more stringent GWAS association threshold was applied (GWAS  $p < 5.0 \times 10^{-4}$ ), the number of SNPs included in each selection analysis dropped, resulting in attenuation of the polygenic adaptation signal for extraversion ( $Q_X = 49.71$ ,  $p_{EXT} = 0.545$ ), subjective well-being ( $Q_X = 74.87$ ,  $p_{SWB} = 0.029$ ), hippocampus volume ( $Q_X = 79.65$ ,  $p_{HIP} = 0.01$ ), putamen volume ( $Q_X = 72.69$ ,  $p_{PUT} = 0.029$ ), inflammatory bowel disease ( $Q_X = 77.25$ ,  $p_{IBD} = 0.017$ ), and type 2 diabetes ( $Q_X = 50.61$ ,  $p_{T2D} = 0.525$ ). Results for schizophrenia and height, both of which suffered less SNP attrition at the more stringent

GWAS p-value, were robust ( $Q_x = 101.30$ ,  $p_{SCZ} < 0.001$ ;  $Q_x = 209.27$ ,  $p_{HEI} < 0.001$ ) to the more stringent association threshold, likely due to the power of the original GWAS (Table 2).

Because the null model of genetic drift is computed with respect to the derived and ancestral alleles (not necessarily the trait-associated alleles), the direction of selection with respect to risk can be difficult to determine. Most phenotypes tested exhibited a higher proportion of derived risk or trait-increasing alleles (for quantitative traits); however, there were six phenotypes with a slightly increased proportion of ancestral risk (or trait-increasing) alleles – namely major depressive disorder, attention deficit hyperactivity disorder, body mass index, type 2 diabetes, height and bipolar disorder (Supplementary Figure 7).

#### *Brain structure volumes exhibit evidence for very recent mosaic polygenic adaptation*

Several brain structure volumes exhibited evidence of very recent polygenic adaptation (approximately 2kya). Total intracranial volume exhibited significant positively shifted tSDS ( $Z = 6.23$ ,  $p_{ICV} = 4.79 \times 10^{-10}$ ) and also when intracranial volume was controlled for height ( $Z = 3.35$ ,  $p_{ICV|HEIGHT} = 7.95 \times 10^{-4}$ ), while amygdala ( $Z = -3.50$ ,  $p_{AMY} = 4.69 \times 10^{-4}$ ), pallidum ( $Z = -3.40$ ,  $p_{PAL} = 6.80 \times 10^{-4}$ ), and putamen ( $Z = -4.93$ ,  $p_{PUT} = 8.15 \times 10^{-7}$ ) exhibited significant negatively shifted tSDS (Supplementary Figure 5) indicating positive selection for *increased* total intracranial volume, but *decreased* volume of subcortical brain structures.

*Schizophrenia-protective alleles exhibit evidence for very recent polygenic adaptation*

Schizophrenia demonstrated significant negatively shifted tSDS ( $Z = -3.07$ ;  $p = 0.002$ ) indicating very recent positive selection on schizophrenia-protective alleles (Figure 1, Table 3, Supplementary Figure 4). Height exhibited significant positively shifted tSDS ( $Z = 17.14$ ,  $p_{HEI} = 6.90 \times 10^{-66}$ ) indicating very recent positive selection on height-increasing alleles; while T2D exhibited significant negative shifted tSDS ( $Z = -6.68$ ,  $p_{T2D} = 2.34 \times 10^{-11}$ ) indicating positive selection on T2D-protective alleles, consistent with prior reports<sup>54</sup>.

*Brain and immune eQTLs are enriched among schizophrenia and brain structure volume-associated variants*

For any phenotype exhibiting evidence of polygenic selection (significant  $Q_x$  or tSDS after multiple testing correction), we tested whether trait-associated alleles are enriched for regulatory function in either brain or immune tissues. Given the fact that the polygenic selection scores are unable to robustly distinguish individual SNPs or alleles that have experienced selection, we examined in bulk the trait-associated SNPs that contributed to signatures of selection for these phenotypes. We expand on previous studies characterizing eQTL enrichment among neuropsychiatric trait-associated variants<sup>8, 89</sup>, by characterizing the functions of the implicated eGenes. We chose to focus on brain and immune eQTLs because both tissues have been previously implicated in neuropsychiatric

and brain-related phenotypes and both systems have demonstrated adaptation throughout human evolution<sup>49,90-92</sup>. We performed cis-eQTL enrichment analysis using multiple brain tissues (putamen, nucleus accumbens, hypothalamus, hippocampus, frontal cortex, cortex, cerebellum, cerebellar hemisphere, caudate, and anterior cingulate cortex) and immune tissues (whole blood, CD4+ T cells, and CD14+ monocytes)<sup>80,81</sup>. Specifically, we tested for an enrichment for a gene regulatory role for trait-associated SNPs, as quantified by an enrichment of the number of eQTL target genes (eGenes) associated with trait-associated SNPs compared to random, matched SNPs. (Methods). This differs from previous eQTL enrichment analyses for neuropsychiatric phenotypes<sup>8,89</sup>, in which the enrichment was based on whether a SNP was or was not an eQTL (ie., not accounting for the number of genes it may be associated with).

Among the brain and behavioral phenotypes with Bonferroni-corrected significant evidence for ancient or recent polygenic adaptation, only schizophrenia was significantly enriched for brain eQTLs ( $p_{\text{brain\_all}} = 0.002$ ,  $p_{\text{cerebellum}} < 0.001$ ). eQTLs from immune cell types were strongly enriched among SNPs associated with subjective well-being ( $p_{\text{immune\_all}} < 0.001$ ,  $p_{\text{monocytes}} < 0.001$ ,  $p_{\text{CD4+cells}} < 0.001$ ), intracranial volume ( $p_{\text{immune\_all}} < 0.001$ ,  $p_{\text{monocytes}} = 0.001$ ,  $p_{\text{CD4+cells}} = 0.005$ ,  $p_{\text{whole blood}} < 0.001$ ), and putamen volume ( $p_{\text{immune\_all}} = 0.003$ ). Hippocampus, pallidum and amygdala volumes and extraversion showed no significant enrichment for brain or immune eQTLs (Figure 2, Supplementary Table 3). Including SNPs located in the HLA region did not significantly affect the results (Supplementary Figure 5, Supplementary Table 3). Additionally, we replicated previous

findings demonstrating enrichment of brain eQTLs among BMI-associated SNPs ( $p_{\text{brain\_all}} < 0.001$ ), and enrichment of immune tissue eQTLs among SNPs associated with height ( $p_{\text{immune\_all}} < 0.001$ ), T2D ( $p_{\text{immune\_all}} = 0.003$ ), and IBD ( $p_{\text{immune\_all}} = 0.004$ ).

To characterize the sets of genes implicated by brain or immune eQTLs (i.e., eGenes, Supplementary Table 4), we performed GO term enrichment analysis on the eGenes derived from each phenotype that showed a significant tissue-specific enrichment (Supplementary Table 5). Here we report the top two significant molecular processes and biological processes for each phenotype-tissue pair. For schizophrenia, the brain eGenes showed the strongest enrichment for phosphate binding ( $q = 1.24 \times 10^{-5}$ ) and protein phosphatase binding ( $q = 1.36 \times 10^{-4}$ ) molecular functions, as well as negative regulation of molecular function ( $q = 1.36 \times 10^{-4}$ ) and establishment of localization in cell ( $q = 2.19 \times 10^{-4}$ ) biological processes. Intracranial volume immune eGenes showed the strongest enrichment for phosphoric ester hydrolase activity ( $q = 5.54 \times 10^{-6}$ ) and hydrolase activity acting on ester bonds ( $q = 5.54 \times 10^{-6}$ ) molecular functions, as well as dephosphorylation ( $q = 4.68 \times 10^{-6}$ ) and phosphate containing compound metabolic processes ( $q = 3.20 \times 10^{-5}$ ) biological processes. Subjective well-being immune eGenes showed the strongest enrichment for adenylyl nucleotide binding ( $q = 4.78 \times 10^{-6}$ ) and ribonucleotide binding ( $q = 1.41 \times 10^{-5}$ ) molecular functions, as well as small molecule metabolic process ( $q = 1.62 \times 10^{-9}$ ) and organonitrogen compound metabolic process ( $q = 1.22 \times 10^{-7}$ ).

#### *Proportion of ancestral and derived alleles among risk conferring SNPs*

For common risk alleles (MAF > 5%), the fraction that were also derived alleles was close to 50% for each phenotype tested (Supplementary Figure 6). For some phenotypes, such as eating disorders, Tourette Syndrome, Alzheimer's disease, obsessive-compulsive disorder, and Type 2 diabetes, > 60% of risk alleles with MAF <5% were also derived alleles (Supplementary Figure 6).

## Discussion

The largely polygenic architecture of neuropsychiatric and other brain related phenotypes has been well established by studies demonstrating that the individual effect size of each common SNP is small, but cumulatively these SNPs account for a significant proportion of the estimated trait heritability<sup>51, 93-96</sup>. For most phenotypes, a small proportion of cases appear to result from *de novo* mutations of large effect which disrupt the coding or regulatory sequences of critical brain genes<sup>90, 97-99</sup>. The high penetrance and very low frequency of these variants in the general population suggest that they are likely targets of negative selection. In contrast, little is known about the history of selection on common genetic variation that confers risk for neuropsychiatric phenotypes. Here we present a systematic study of genetic evidence for strong and weak positive selection across a catalogue of twenty-five phenotypes including neuropsychiatric traits, personality traits, and brain structure volumes.

Consistent with the polygenic architecture of common variation underlying

complex traits, we find little evidence of widespread global population differentiation and no evidence of strong selective sweeps among trait-associated SNPs (Table 1). While prior studies have reported significant genetic correlation between polygenic risk across ancestries<sup>100</sup>, we cannot rule out the possibility that SNPs associated with these phenotypes in non-European ancestries may demonstrate greater population differentiation. Nevertheless, in contrast to autoimmune phenotypes<sup>87</sup>, there is currently no evidence of strong positive selection acting on common risk alleles for neuropsychiatric disorders, personality traits, or brain structure volumes.

We observed a significant enrichment of trait-associated SNPs for schizophrenia and neuroticism within regions of the genome depleted of derived Neanderthal alleles, and nearly significant enrichment for total intracranial volume, consistent with the hypothesis that these genomic regions harbor variation fundamental to the development of the modern human brain<sup>30</sup>. Importantly, this enrichment analysis focuses on regional genomic annotations, as opposed to allelic annotations thus we can make no claim regarding risk or protective alleles for any trait. Furthermore, we found no evidence of enrichment in these regions in some of the most well-powered traits we studied – including height and BMI – indicating that the findings in schizophrenia, neuroticism, and total intracranial volume were not solely a function of power achieved in the discovery GWAS.

Ancient polygenic adaptation (~30 kya), which is manifest in small but coordinated shifts in trait-associated frequencies, was observed among neuropsychiatric phenotypes (schizophrenia), personality traits (extraversion and subjective well-being), volumes of

subcortical brain structures (putamen and hippocampus), autoimmune disorders (inflammatory bowel disease), height, and type 2 diabetes (Table 2). Loci associated with these traits demonstrated significant evidence of shifts in allele frequency consistent with a positive selection model. However, a limitation of this approach is that the null model of neutral drift is computed with respect to ancestral and derived allelic annotations, thus complicating the interpretation of whether ancient positive selection has acted on risk/trait-increasing alleles or protective/trait-decreasing alleles.

We employed a directional test of selection (tSDS) which overcomes this limitation, albeit covering a narrower evolutionary timescale. The tSDS approach is powered to detect selection that has been active during the past ~2,000 years - which makes it difficult to determine with certainty whether the observed direction of selection also reflects more ancient selection pressures<sup>54</sup>. Using this test, we reproduced significant findings of positive polygenic selection acting on height-increasing alleles, T2D-protective alleles, and schizophrenia-protective alleles. Moreover, our tSDS analysis revealed strong evidence of recent polygenic selection acting on volumes of the amygdala, pallidum and putamen, as well as on total intracranial volume, and total intracranial volume adjusted for height.

The results provide evidence for both the mosaic and concerted theories of brain evolution. Specifically, we find that recent positive polygenic selection has strongly favored total intracranial-volume *increasing* alleles, which closely mirrors the adaptation signal observed for height-*increasing* alleles. A significant genetic correlation between

height and total ICV has also been reported ( $R_g = 0.241$ ,  $p = 2.4 \times 10^{-10}$ )<sup>32</sup>. In the analysis of ICV adjusted for height, the strength of the signal for polygenic adaptation was greatly attenuated (Table 3) but remained significant even after accounting for substantial evolutionarily constraint related to overall body size. Integration of eQTLs from the GTEx project further characterizes this relationship revealing that approximately 30% of the brain-tissue eGenes targeted by ICV-associated SNPs are also targeted by height-associated SNPs, suggesting a shared set of genes influencing both phenotypes and experiencing positive polygenic selection.

Among subcortical brain volume traits, we observed signatures of ancient and recent adaptation in some (e.g., hippocampus, putamen, pallidum, amygdala), but not all (e.g., nucleus accumbens, caudate nucleus, thalamus) brain structures (Figure 1). The differential pattern of signatures of selection suggest that individual brain structures have been able to evolve independently of each other, consistent with a mosaic theory of brain evolution<sup>34-36, 101</sup>. Furthermore, our results indicate very recent positive selection acting on total intracranial volume-*increasing* alleles, but putamen, amygdala and pallidum volume-*decreasing* alleles (Supplementary Figure 5). While increases in some regions could be genetically correlated to decreases in other regions, GWAS findings from Hibar et al. (2015) also confirm that despite high heritability, the genetic correlations between the structures are low. Thus, in the context of previous findings, our results provide direct genetic evidence for a model of selection that has differentially influenced both volume-increasing and volume-decreasing alleles, perhaps providing opportunities for modern

human-specific neocortex enlargement over time. If correct, falsifiable hypotheses also emerge from these findings; for example, cortical surface area-increasing alleles should demonstrate evidence of positive selection during both recent and ancient evolutionary history. Additional work will be required to replicate these results and test hypotheses consistent with this model.

Finally, we investigated whether our observed signatures of selection could be driven by adaptations of the immune system or by biological innovations in brain development. Given that polygenic selection scores do not identify individual SNPs or alleles that have experienced selection, we are limited to examine possible function of the trait-associated SNPs in bulk that contributed to observed signatures of selection. The cis-genetic component of gene expression contributes up to 30-50% of overall heritability of complex traits, but there is considerable variation among traits<sup>102</sup>. For example, while cis-eQTLs contribute only 8% (s.e.=2%) to heritability of schizophrenia, they contribute up to 32% (s.e.=10%) of putamen volume<sup>102</sup>. We discovered a significant enrichment of combined immune tissue eQTLs including whole blood, monocytes, and CD4+ T-cells among SNPs influencing putamen and intracranial volumes, subjective well-being, height, and type 2 diabetes (Figure 2; Supplementary Figure 5, Supplementary Table 4). We also observed an enrichment of brain tissue eQTLs among SNPs influencing schizophrenia, height, and BMI. These findings are concordant with previously reported contribution of specific cell types to heritability of complex traits (Figure 2; Supplementary Figure 5, Supplementary Table 4)<sup>50</sup>. Gene set enrichment analysis indicated that the genes

implicated by eQTLs in the top GWAS associations for several traits under selection cluster in biologically plausible molecular functions and biological processes, such as “neurogenesis, neuron differentiation, and neuron development” (Supplementary Table 5).

In conclusion, we demonstrate that weak polygenic adaptation, not strong selective sweeps, has influenced the evolutionary history of common genetic variation underlying neuropsychiatric disorders, behavioral traits, and brain structure volumes. The lack of evidence for strong sweeps is consistent with evolutionary scenarios in which selection coefficients for individual trait associated variants were not strong, but also consistent with hypotheses relating to fitness trade-offs between effects on multiple traits for highly polygenic phenotypes. Schizophrenia-protective alleles may have experienced both ancient and very recent positive polygenic selection, consistent with hypotheses positing shifting selection pressures between ancient and modern times due to ancestral adaptation or neutrality and fluctuating environments<sup>103</sup>. Furthermore, we provide genetic evidence of concerted evolution between height and total ICV, but mosaic evolution of subcortical brain structure volumes. Finally, we suggest that adaptive innovations of the immune system may also influence brain evolution among humans.

### **Acknowledgements**

We thank Dr. Jeremy Berg, Yair Field, and Jonathan Pritchard for helpful discussions and again thank Dr. Field for sharing scripts used to calculate standard error and p-value for the tSDS analysis. Additionally, we would like to thank Mr. Donald Hucks for assistance

in manuscript preparation and Mr. Peter Straub for his assistance in data preparation. We wish to thank the individuals who participated in each study and the investigators who shared their results with the larger scientific community. Finally, we wish to acknowledge the Psychiatric Genomics Consortium, Social Science Genetic Association Consortium, Diabetes Genetics Replication and Meta-Analysis Consortium, Enhancing Neuro-Imaging Genetics through Meta-Analysis consortium, Genetic Investigation of Anthropometric Traits consortium, and the International Genomics of Alzheimer's Project.

**Funding:** This work was supported, in part, by the Conte Center for Computational Neuropsychiatric Genomics: E.R.B., E.A.K., L.K.D., and B.E.S. are supported, in part by NIH/NIMH grant 3P50MH094267-04S1. J.L.S. and P.T. are also supported, in part, by NIH grant U54 EB20403. S.E.F is supported by the Max Planck Society.

*Psychiatric Genomics Consortium TS/OCD Workgroup*

We wish to acknowledge the contribution of the TS/OCD PGC Working group for sharing summary results of GWAS prior to their publication, and for their helpful reviews of this manuscript.

*International Genomics of Alzheimer's Project*

We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the

design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients, and their families. The i-Select chips was funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD was supported by the Medical Research Council (Grant n° 503480), Alzheimer's Research UK (Grant n° 503176), the Wellcome Trust (Grant n° 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) grant n° 01GI0102, 01GI0711, 01GI0420. CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01-AG-12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association, and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants: U01 AG032984, U24 AG021886, U01 AG016976, and the Alzheimer's Association grant ADGC-10-196728.

### **IGAP Material and methods**

International Genomics of Alzheimer's Project (IGAP) is a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyse four previously-published GWAS datasets

consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative – EADI the Alzheimer Disease Genetics Consortium – ADGC The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium – CHARGE The Genetic and Environmental Risk in AD consortium – GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 & 2.

## **Main Paper Tables**

Table 1. Enrichment of trait-associated variants in genomic regions of extreme iHS, Fst, and S Score. The #SNPs denotes the number of trait-associated SNPs used in the analysis (Methods and Supplementary Figure 1). Enrichment p-values significant after multiple-testing correction ( $p < 0.002$ ) are denoted with an asterisk. 1: GWAS significance  $P < 5.0 \times 10^{-6}$ , 2: GWAS significance  $P < 5.0 \times 10^{-8}$ , 3: GWAS significance  $P < 5.0 \times 10^{-4}$ , 4: GWAS significance  $P < 5.0 \times 10^{-4}$ .

Phenotypes	NSS-Score Enrichment		Extreme Population Differentiation			Extreme Integrated Haplotype Score Enrichment		
	#SNPs	Enrichment p-value ( $S < -4.32$ )	#SNPs	Enrichment p-value ( $F_{st} > 0.30$ )	Enrichment p-value ( $F_{st} > 0.56$ )	#SNPs	Enrichment p-value ( $iHS > 2.0$ )	Enrichment p-value ( $iHS > 2.6$ )
<b>Neuropsychiatric and Behavioral Traits</b>								
Attention Deficit and Hyperactivity Disorder	2,319	0.544	1,036	0.048	0.624	514	0.552	0.760
Alzheimer's Disease	7,328	0.948	3,863	0.518	0.638	1,613	0.914	0.968
Anorexia Nervosa	8,200	0.934	4,247	0.868	0.168	1,512	0.944	0.714
Generalized Anxiety Disorder	5,564	0.896	2,504	0.934	0.678	1,317	0.974	0.368
Autism Spectrum Disorder	6,713	0.096	3,467	0.792	0.390	1,383	0.464	0.120
Bipolar Disorder	3,832	0.050	1,847	<0.002*	0.036	924	0.868	0.910
Extraversion	6,854	0.162	3,316	0.814	0.972	1,586	0.286	0.242
Major Depressive Disorder	2,412	0.044	1,162	0.146	0.632	595	0.966	0.890
Neuroticism	6,833	<0.002*	3,306	0.970	0.838	1,566	0.862	0.810
Obsessive Compulsive Disorder	7,475	0.230	3,930	1.000	0.994	1,624	0.904	0.960
Schizophrenia	16,200	0.002*	8,759	0.378	0.860	3,845	0.974	0.972
Subjective Well-Being	3,858	0.436	1,806	<0.002*	0.234	1,011	0.898	0.928
Tourette Syndrome	7,917	0.532	4,246	0.878	0.750	1,694	0.594	0.774
<b>Brain Structure Volumes</b>								
Accumbens	5,164	0.488	2,709	0.996	0.486	1,060	0.510	0.590
Amygdala	5,237	0.794	2,710	0.250	0.122	1,057	0.770	0.596
Caudate Nucleus	5,198	0.214	2,661	0.762	0.170	1,098	0.592	0.104
Hippocampus	6,621	0.594	3,462	0.944	0.966	1,366	0.998	0.906
Total Intracranial Volume	7,446	0.004	3,888	0.680	0.758	1,561	0.776	0.890
Pallidum	5,271	0.026	2,707	0.856	0.678	1,075	0.768	0.948
Putamen	5,064	0.766	2,662	0.640	0.504	1,097	0.942	0.778
Thalamus	5,204	0.288	2,649	0.902	0.220	1,064	0.508	0.776
<b>Complex Traits</b>								
BMI <sup>1</sup>	306	0.032	179	0.260	0.124	107	0.968	0.578
Height <sup>2</sup>	1,449	0.108	838	0.150	0.394	405	0.480	0.972
Inflammatory Bowel Disease <sup>3</sup>	2,170	0.400	1,250	0.686	0.612	482	0.150	0.452
Type 2 Diabetes <sup>4</sup>	669	0.004	344	0.002*	0.134	168	0.346	0.576

Table 2. Enrichment of trait-associated variants in genomic regions experiencing polygenic adaptation. The selection test statistic ( $Q_x$ ) was computed as described in Berg and Coop (2014). Enrichment p-values significant after multiple-testing correction are denoted with an asterisk.

Phenotypes	#SNPs	Qx	P(Qx)	#SNPs	Qx	P(Qx)
<b>Neuropsychiatric Traits</b>	$P < 5.0 \times 10^{-3}$			$P < 5.0 \times 10^{-4}$		
Alzheimer's Disease	1,777	83.80	0.014	259	46.81	0.688
Anorexia Nervosa	1,196	68.96	0.112	166	51.41	0.556
Generalized Anxiety Disorder	1,420	66.41	0.141	183	59.21	0.235
Autism Spectrum Disorder	1,149	58.02	0.334	138	42.69	0.833
Bipolar Disorder	2,634	57.26	0.310	449	54.30	0.356
Extraversion	1,485	88.04	0.001*	196	49.71	0.545
Major Depressive Disorder	1,806	63.31	0.146	224	70.11	0.065
Neuroticism	1,617	77.52	0.156	205	75.69	0.029
Obsessive Compulsive Disorder	1,198	70.33	0.082	172	39.18	0.905
Schizophrenia	3,307	208.36	<0.001*	1,029	101.30	<0.001*
Subjective Well-Being	2,717	113.90	<0.001*	434	74.87	0.029
Tourette Syndrome	1,441	73.51	0.082	209	42.79	0.828
<b>Brain Structure Volumes</b>	$P < 5.0 \times 10^{-3}$			$P < 5.0 \times 10^{-4}$		
Accumbens	1,152	61.61	0.203	134	63.86	0.162
Amygdala	1,164	55.05	0.378	137	53.34	0.470
Caudate Nucleus	1,210	48.33	0.626	153	56.18	0.348
Hippocampus	1,237	108.98	<0.001*	177	79.65	0.010
Intracranial Volume	1,249	54.37	0.386	175	50.79	0.526
Pallidum	1,183	52.76	0.559	176	48.78	0.625
Putamen	1,203	115.76	<0.001*	155	72.69	0.029
Thalamus	1,123	70.98	0.088	153	57.51	0.288
<b>Complex Traits</b>	$P < 5.0 \times 10^{-4}$			$P < 5.0 \times 10^{-6}$		
Inflammatory Bowel Disease	423	101.13	<0.001*	81	77.25	0.017
Type 2 Diabetes	478	117.77	<0.001*	42	50.61	0.525
	$P < 5.0 \times 10^{-6}$			$P < 5.0 \times 10^{-8}$		
BMI	246	78.20	0.010	97	82.61	0.003
Height	2,002	303.29	<0.001*	1,087	209.27	<0.001*

Table 3. Recent positive selection on trait-associated variant using Singleton Density Score (SDS) analysis. The mean tSDS for the top 1,000 associated SNPs are shown: a negative value indicates positive selection on the trait-increasing allele, a positive value indicates positive selection on the trait-decreasing allele. The p-value for a two-tailed T-test represents deviation from the expected mean of 0. Results significant after multiple-testing correction are denoted with an asterisk. 1 GWAS significance  $P < 5.0 \times 10^{-6}$ , 2 GWAS significance  $P < 5.0 \times 10^{-8}$ , 3 GWAS significance  $P < 5.0 \times 10^{-4}$ , 4 GWAS significance  $P < 5.0 \times 10^{-4}$ .

Phenotypes	Spearman rho	Jackknife SE	Z-Score	P-value
<b>Neuropsychiatric and Behavioral Phenotypes</b>				
Attention Deficit and Hyperactivity Disorder	-0.008279	0.003476	-2.382	0.017
Alzheimer's Disease	-0.001977	0.003538	-0.559	0.576
Anorexia Nervosa	0.00138	0.004159	+0.332	0.740
Generalized Anxiety Disorder	0.000357	0.003508	+0.102	0.919
Autism Spectrum Disorder	-6.564E-06	0.003456	-0.002	0.999
Bipolar Disorder	-0.009223	0.004048	-2.278	0.023
Extraversion	-0.003696	0.003478	-1.063	0.288
Major Depression	-0.0006429	0.003412	-1.884	0.060
Neuroticism	-0.0003826	0.003404	-0.112	0.911
Obsessive Compulsive Disorder	-0.008432	0.003392	-2.486	0.013
Schizophrenia	-0.01296	0.004226	-3.066	0.002*
Subjective Well-Being	-0.001072	0.003812	-0.281	0.779
Tourette Syndrome	-0.000783	0.003497	-0.224	0.823
<b>Brain Structure Volumes</b>				
Accumbens	-0.006426	0.003612	-1.779	0.075
Amygdala	-0.01237	0.003537	-3.498	4.69e-4*
Caudate Nucleus	0.002604	0.003658	+0.712	0.477
Hippocampus	-0.002703	0.003718	-0.7269	0.467
Intracranial Volume	0.02272	0.00365	+6.226	4.79e-10*
Intracranial Volume   height	0.01244	0.00371	3.354	7.95e-4*
Pallidum	-0.01209	0.003559	-3.398	6.8e-4*
Putamen	-0.01878	0.003808	-4.932	8.15e-07*
Thalamus	-0.005964	0.003567	-1.672	0.096
<b>Complex Traits</b>				
BMI <sup>1</sup>	-0.009211	0.004681	-1.968	0.049
Height <sup>2</sup>	0.07902	0.004609	+17.140	6.90e-66*
Inflammatory Bowel Disease <sup>3</sup>	0.007365	0.004039	+1.824	0.068
Type 2 Diabetes <sup>4</sup>	-0.02337	0.003497	-6.683	2.34e-11*

## Main Paper Figures

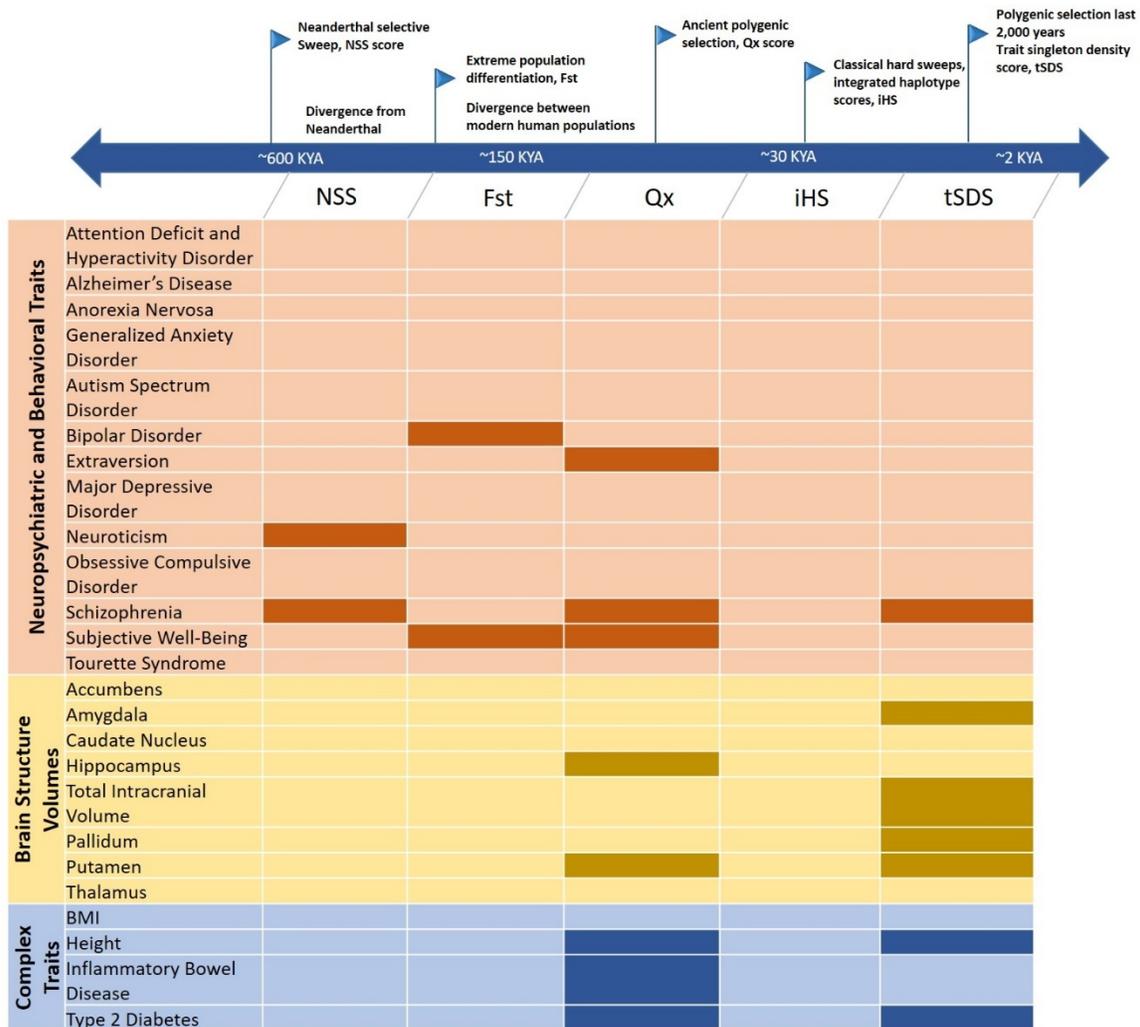


Figure 1. Summary of various ancient and recent signatures of polygenic adaptation analysis in neuropsychiatric and behavioral traits (orange), brain structure volumes (yellow), and complex traits (blue). The tested signatures are selection since Neanderthal divergence (Neanderthal selective sweep, NSS score), extreme population differentiation (fixation index, Fst), ancient polygenic selection (Qx score), classical hard sweeps

(integrated haplotype score, iHS), and polygenic selection in the past 2,000 years (trait singleton density score, tSDS). The boxes filled with a darker color for each of the tests represent significant signature of selection for a given phenotype after multiple testing correction for 25 phenotypes ( $p < 0.002$ ).

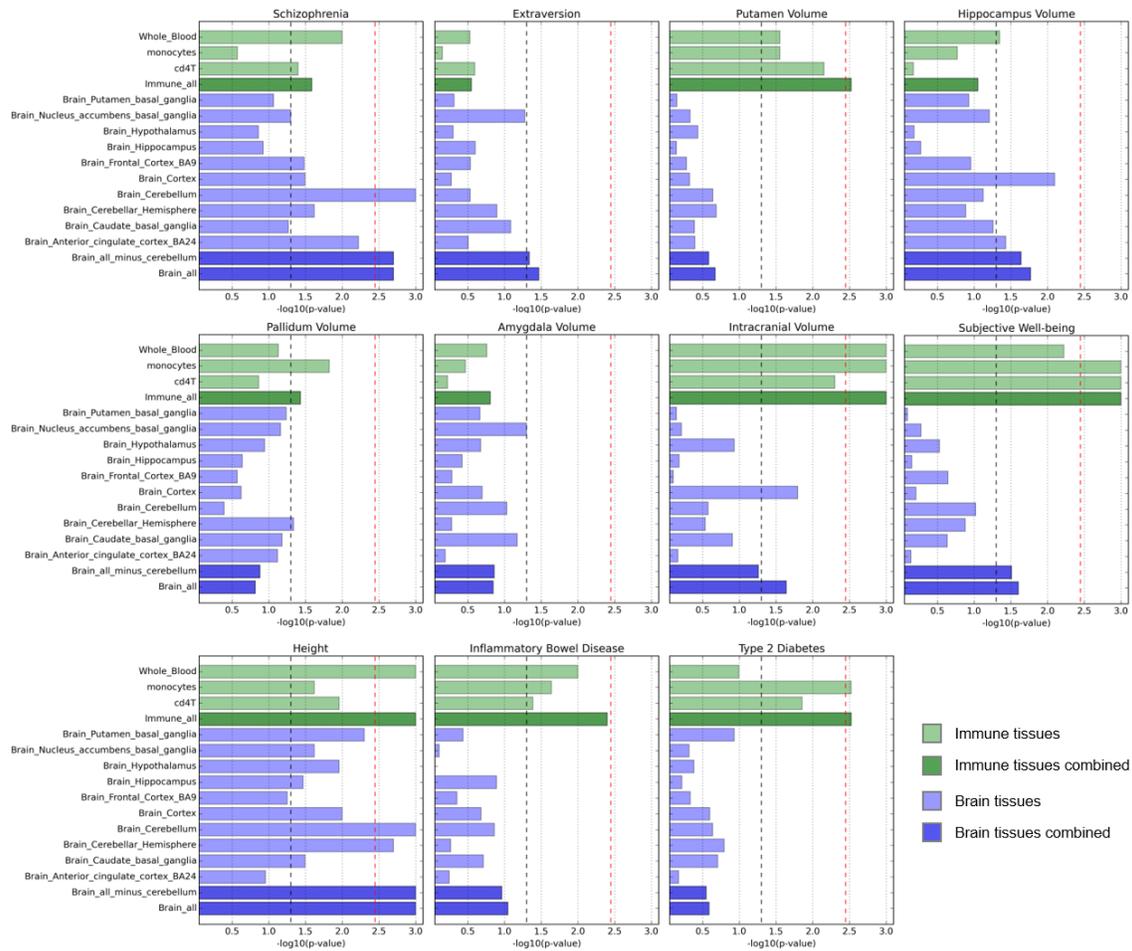


Figure 2. eQTL enrichment in the brain and immune tissues for eleven phenotypes (excluding SNPs in the HLA region) exhibiting evidence of polygenic selection (significant Qx or tSDS after multiple testing correction). Light green bars represent each

immune tissue or cell type: whole blood, monocytes, and CD4+ T cells, while the dark green represents enrichment in a combination of the three immune tissues. Light blue bars represent each brain tissue, while the dark blue represents enrichment in a combination of ten brain tissues, or all ten brain tissues minus cerebellum. The black dashed line represents a p-value of 0.05. The red dashed line represents the significant p-value threshold (0.004) after accounting for 13 tissues.

### **Supplementary Tables**

Supplementary Table 1. Description of GWAS studies used for 25 phenotypes. The total number of participant subjects, number of SNPs tested, lambda GC, LD score regression, and number of SNPs reaching significance at varying thresholds after PLINK clumping ( $R^2 = 0.25$ , window = 500kb) are shown.

Supplementary Table 2.  $F_3$  statistic and genetic value spearman correlations. Spearman's rho for the relationship between the  $F_3$  statistic and the genetic values (Methods) of the 52 Human Genome Diversity Panel populations included in the  $Q_x$  score analysis.

Supplementary Table 3. Tables for eQTL enrichment in the brain and immune tissues for eleven phenotypes (each phenotype in a separate tab) exhibiting evidence of polygenic selection (significant  $Q_x$  or tSDS after multiple testing correction). Each tab shows results of eQTL enrichment analysis with and without inclusion of HLA SNPs.

N\_eGenes is the number of eGenes implicated by variants in each tissue tested, median and SD are the median number of eGenes and the standard deviation, respectively, in the 1000 matching sets. Enrichment P-value was calculated as the proportion of randomized sets in which the eGene count matches or exceeds the actual observed count in the list of the trait-associated variants.

Supplementary Table 4. Tables showing eQTLs (rs id, chromosome, position, statistics associated with eQTL tests, associated eGene, and whether it is in the HLA region) among the trait-associated variants for eleven phenotypes exhibiting evidence of polygenic selection. Each phenotype and eQTL dataset (GTEx and ImmVar) are listed in a separate tab.

Supplementary Table 5. Tables listing biological processes and molecular functions enriched among the eGenes implicated by trait-associated variants in tissues and phenotypes showing significant eQTL enrichment (i.e. schizophrenia\_brain, intracranial volume\_immune, and subjective well-being\_immune). Gene set enrichment analysis tool (<http://software.broadinstitute.org/gsea/index.jsp>) was used to perform enrichment analysis of gene sets.

## **Supplementary Figures**

Supplementary Figure 1. Workflow for enrichment analyses.  $iHS$  and  $F_{st}$  enrichment analyses excluded Neanderthal SNPs while the NSS analysis did not. SNPs fixed in all 52 Human Genome Diversity Panel populations, and SNPs fixed in the French Reference Population were excluded for the ancient polygenic selection analysis. AF= Allele Frequencies, 1KF = 1000 Genomes Project, HLA = Human leukocyte antigen.

Supplementary Figure 2. Ancient polygenic adaptation enrichment distributions. The five brain-related phenotypes showing significant evidence of ancient polygenic adaptation are shown. The blue line represent the  $Q_x$  test statistic compared to a background of 1000 permuted genetic drift models represented as histograms.

Supplementary Figure 3.  $F_3$  statistic correlation with genetic values from 52 Human Genome Diversity Panel populations. Populations are grouped by broad regional groups: orange = Africa, green = Oceania, pink = East Asia, purple = Americas, red = East Asia, yellow = Middle East, blue = Europe. None of these correlations are significant (see Supplementary Table 2).

Supplementary Figure 4. Recent polygenic adaptation across the genome. The five brain-related phenotypes with significant evidence of recent polygenic adaptation are shown. Each gold dot represents the mean  $tSDS$  of 1,000 SNPs. SNP bins are ranked from

GWAS p-value  $\sim 1$  (left) to GWAS p-value  $\sim 0$  (right). Positive tSDS corresponds to selection acting on the risk/trait-increasing allele while a negative tSDS corresponds to selection acting on the protective/trait-decreasing allele. The dashed-line represents the null expectation of no selection.

Supplementary Figure 5. eQTL enrichment in the brain and immune tissues for eleven phenotypes (including SNPs in the HLA region) exhibiting evidence of polygenic selection (significant Qx or tSDS after multiple testing correction). Light green bars represent each immune tissue or cell type: whole blood, monocytes, and cd4+ T cells, while the dark green represents enrichment in a combination of the three immune tissues. Light blue bars represent each brain tissue, while the dark blue represents enrichment in a combination of ten brain tissues, or all ten brain tissues minus cerebellum. The black dashed line represents a p-value of 0.05. The red dashed line represents the significant p-value threshold (0.004) after accounting for 13 tissues.

Supplementary Figure 6. Ancestral and derived trait-associated alleles and their frequencies. (A) Proportion of trait-associated derived alleles by minor allele frequency bins for ten neuropsychiatric disorders and three personality traits; (B) Proportion of trait associated derived alleles by minor allele frequency bins for total intracranial volume, seven subcortical brain structure volume traits, and four complex traits without neuropsychiatric associations.

## References

1. Neale, B.M., *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-897 (2010).
2. Lambert, J.C., *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458 (2013).
3. Boraska, V., *et al.* A genome-wide association study of anorexia nervosa. *Mol Psychiatry* **19**, 1085-1094 (2014).
4. Otowa, T., *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry* **21**, 1391-1399 (2016).
5. Sklar, P., Psychiatric GWAS Consortium Bipolar Disorder Working Group, Ripke, S., Scott, L.J. & Andreassen, O.A. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**, 977-983 (2011).
6. van den Berg, S.M., *et al.* Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet* **46**, 170-182 (2016).
7. Ripke, S., *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511 (2013).
8. Scharf, J.M., *et al.* Genome-wide association study of Tourette's syndrome. *Mol Psychiatry* **18**, 721-728 (2013).
9. Liu, J.Z., *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
10. Morris, A.P., *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-990 (2012).
11. Wood, A.R., *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186 (2014).
12. Locke, A.E., *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
13. Hibar, D.P., *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229 (2015).
14. Okbay, A., *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* **48**, 624-633 (2016).

15. Power, R.A., *et al.* Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* **70**, 22-30 (2013).
16. Polderman, T.J., *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* **47**, 702-709 (2015).
17. Bigdeli, T.B., *et al.* On schizophrenia as a "disease of humanity". *Schizophr Res* **143**, 223-224 (2013).
18. Huxley, J., Mayr, E., Osmond, H. & Hoffer, A. Schizophrenia as a Genetic Morphism. *Nature* **204**, 220-221 (1964).
19. Jablensky, A., *et al.* Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol Med Monogr Suppl* **20**, 1-97 (1992).
20. Allison, A.C. The discovery of resistance to malaria of sickle-cell heterozygotes. *Biochemistry and Molecular Biology Education* **30**, 279-287 (2002).
21. Gazal, S. Linkage disequilibrium dependent architecture of human complex traits reveals action of negative selection. *BioRxiv* (2016).
22. Luo, X.J., *et al.* Systematic Integration of Brain eQTL and GWAS Identifies ZNF323 as a Novel Schizophrenia Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function. *Schizophr Bull* **41**, 1294-1308 (2015).
23. Crespi, B., Summers, K. & Dorus, S. Adaptive evolution of genes underlying schizophrenia. *Proc Biol Sci* **274**, 2801-2810 (2007).
24. Keller, M.C. & Miller, G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci* **29**, 385-404; discussion 405-352 (2006).
25. Darves-Bornoz, J.M., Lemperiere, T., Degiovanni, A. & Gaillard, P. Sexual victimization in women with schizophrenia and bipolar disorder. *Soc Psychiatry Psychiatr Epidemiol* **30**, 78-84 (1995).
26. Coverdale, J.H. & Turbott, S.H. Sexual and physical abuse of chronically ill psychiatric outpatients compared with a matched sample of medical outpatients. *J Nerv Ment Dis* **188**, 440-445 (2000).
27. Khalifeh, H., *et al.* Domestic and sexual violence against patients with severe mental illness. *Psychol Med* **45**, 875-886 (2015).
28. Pearlson, G.D. & Folley, B.S. Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. *Schizophr Bull* **34**, 722-733 (2008).
29. Ogawa, L.M. & Vallender, E.J. Evolutionary conservation in genes underlying human psychiatric disorders. *Front Hum Neurosci* **8**, 283 (2014).
30. Srinivasan, S., *et al.* Genetic Markers of Human Evolution Are Enriched in Schizophrenia. *Biol Psychiatry* **80**, 284-292 (2016).
31. Hibar, D.P., *et al.* Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry* (2017).

32. Adams, H.H., *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* **19**, 1569-1582 (2016).
33. Montgomery, S.H., Mundy, N.I. & Barton, R.A. Brain evolution and development: adaptation, allometry and constraint. *Proc Biol Sci* **283** (2016).
34. Hager, R., Lu, L., Rosen, G.D. & Williams, R.W. Genetic architecture supports mosaic brain evolution and independent brain-body size regulation. *Nat Commun* **3**, 1079 (2012).
35. Gomez-Robles, A., Hopkins, W.D. & Sherwood, C.C. Modular structure facilitates mosaic evolution of the brain in chimpanzees and humans. *Nat Commun* **5**, 4469 (2014).
36. Noreikiene, K., *et al.* Quantitative genetic analysis of brain size variation in sticklebacks: support for the mosaic model of brain evolution. *Proceedings Biological Sciences* **282** (2015).
37. Herculano-Houzel, S., Manger, P.R. & Kaas, J.H. Brain scaling in mammalian evolution as a consequence of concerted and mosaic changes in numbers of neurons and average neuronal cell size. *Front Neuroanat* **8**, 77 (2014).
38. Gutierrez-Ibanez, C., *et al.* Mosaic and concerted evolution in the visual system of birds. *PLoS One* **9**, e90102 (2014).
39. Boedhoe, P.S., *et al.* Distinct Subcortical Volume Alterations in Pediatric and Adult OCD: A Worldwide Meta- and Mega-Analysis. *Am J Psychiatry* **174**, 60-69 (2017).
40. Franke, B., *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nat Neurosci* **19**, 420-431 (2016).
41. Thompson, P.M., *et al.* ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide. *Neuroimage* **145**, 389-408 (2017).
42. Li, M., *et al.* Allelic differences between Europeans and Chinese for CREB1 SNPs and their implications in gene expression regulation, hippocampal structure and function, and bipolar disorder susceptibility. *Mol Psychiatry* **19**, 452-461 (2014).
43. Perlman, G., *et al.* Symptoms of psychosis in schizophrenia, schizoaffective disorder, and bipolar disorder: A comparison of African Americans and Caucasians in the Genomic Psychiatry Cohort. *Am J Med Genet B Neuropsychiatr Genet* **171**, 546-555 (2016).
44. Estrada, E., *et al.* Nicotine dependence and psychosis in Bipolar disorder and Schizoaffective disorder, Bipolar type. *Am J Med Genet B Neuropsychiatr Genet* **171**, 521-524 (2016).
45. Tang, V.K., *et al.* Substance use associated with short sleep duration in patients with schizophrenia or schizoaffective disorder. *Am J Med Genet B Neuropsychiatr Genet* **171**, 525-533 (2016).
46. Cieslak, K., *et al.* Traumatic brain injury and bipolar psychosis in the Genomic Psychiatry Cohort. *Am J Med Genet B Neuropsychiatr Genet* **171**, 506-512 (2016).
47. Hartz, S.M., *et al.* Comorbidity of severe psychotic disorders with measures of substance use. *JAMA Psychiatry* **71**, 248-254 (2014).

48. Pato, M.T., *et al.* The genomic psychiatry cohort: partners in discovery. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 306-312 (2013).
49. Sekar, A., *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177-183 (2016).
50. Finucane, H.K., *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235 (2015).
51. Bulik-Sullivan, B., *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
52. Ripke, S., *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
53. Berg, J.J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet* **10**, e1004412 (2014).
54. Field, Y., *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**, 760-764 (2016).
55. Green, R.E., *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
56. Weir, B.S. & Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370 (1984).
57. Sabeti, P.C., *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-837 (2002).
58. Vitti, J.J., Grossman, S.R. & Sabeti, P.C. Detecting natural selection in genomic data. *Annu Rev Genet* **47**, 97-120 (2013).
59. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
60. Bersaglieri, T., *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* **74**, 1111-1120 (2004).
61. Hamblin, M.T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* **66**, 1669-1679 (2000).
62. Tishkoff, S.A., *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455-462 (2001).
63. Pritchard, J.K., Pickrell, J.K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* **20**, R208-215 (2010).
64. Tylee, D.S., *et al.* Genetic correlations among brain-behavioral and immune-related phenotypes based on genome-wide association data. *bioRxiv* (2016).
65. Zheng, P., *et al.* Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Mol Psychiatry* **21**, 786-796 (2016).
66. Filiano, A.J., Gadani, S.P. & Kipnis, J. Interactions of innate and adaptive immunity in brain development and function. *Brain Res* **1617**, 18-27 (2015).
67. Quintana-Murci, L. & Clark, A.G. Population genetic tools for dissecting innate

- immunity in humans. *Nat Rev Immunol* **13**, 280-293 (2013).
68. Autism Spectrum Disorder Working Group of the Psychiatry Genomics Consortium. Dataset: PGC-ASD summary statistics from a meta-analysis of 5,305 ASD-diagnosed cases and 5,305 pseudocontrols of European descent (based on similarity to CEPH reference genotypes) (March 2015). (available at: <http://www.med.unc.edu/pgc/results-and-downloads>).
69. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OC GAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* (in press).
70. Yu, D., *et al.* Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *Am J Psychiatry* **172**, 82-93 (2015).
71. van den Berg, S.M., *et al.* Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet* **44**, 295-313 (2014).
72. de Moor, M.H., *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **72**, 642-650 (2015).
73. Pybus, M., *et al.* 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res* **42**, D903-909 (2014).
74. Pers, T.H., Timshel, P. & Hirschhorn, J.N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418-420 (2015).
75. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sanchez, E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* **16**, 359-371 (2015).
76. Vernot, B. & Akey, J.M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017-1021 (2014).
77. Simonti, C.N., *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737-741 (2016).
78. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* **5**, e1000471 (2009).
79. Patterson, N., *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
80. Aguet, F., *et al.* Local genetic effects on gene expression across 44 human tissues. *bioRxiv* (2016).
81. Raj, T., *et al.* CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. *Hum Mol Genet* **23**, 2729-2736 (2014).
82. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
83. Mootha, V.K., *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).

84. Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
85. Gorlova, O.Y., *et al.* Derived SNP alleles are used more frequently than ancestral alleles as risk-associated variants in common human diseases. *J Bioinform Comput Biol* **10**, 1241008 (2012).
86. Luca, F., Perry, G.H. & Di Rienzo, A. Evolutionary adaptations to dietary changes. *Annu Rev Nutr* **30**, 291-314 (2010).
87. Raj, T., *et al.* Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet* **92**, 517-529 (2013).
88. Robinson, M.R., *et al.* Population genetic differentiation of height and body mass index across Europe. *Nat Genet* **47**, 1357-1362 (2015).
89. Stewart, S.E., *et al.* Genome-wide association study of obsessive-compulsive disorder. *Mol Psychiatry* **18**, 788-798 (2013).
90. Purcell, S.M., *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190 (2014).
91. Heneka, M.T., Golenbock, D.T. & Latz, E. Innate immunity in Alzheimer's disease. *Nat Immunol* **16**, 229-236 (2015).
92. Marsh, S.E., *et al.* The adaptive immune system restrains Alzheimer's disease pathogenesis by modulating microglial function. *Proc Natl Acad Sci U S A* **113**, E1316-1325 (2016).
93. Davis, L.K., *et al.* Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet* **9**, e1003864 (2013).
94. Gandal, M.J., Leppa, V., Won, H., Parikshak, N.N. & Geschwind, D.H. The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci* **19**, 1397-1407 (2016).
95. Purcell, S.M., *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748-752 (2009).
96. Wray, N.R., *et al.* Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* **17**, 36-48 (2012).
97. Kirov, G. CNVs in neuropsychiatric disorders. *Hum Mol Genet* **24**, R45-49 (2015).
98. Li, M., Chang, H. & Xiao, X. BDNF Val66Met polymorphism and bipolar disorder in European populations: A risk association in case-control, family-based and GWAS studies. *Neurosci Biobehav Rev* **68**, 218-233 (2016).
99. Szatkiewicz, J.P., *et al.* Copy number variation in schizophrenia in Sweden. *Mol Psychiatry* **19**, 762-773 (2014).
100. de Candia, T.R., *et al.* Additive genetic variation in schizophrenia risk is shared by populations of African and European descent. *Am J Hum Genet* **93**, 463-470 (2013).
101. Barton, R.A. & Harvey, P.H. Mosaic evolution of brain structure in mammals. *Nature* **405**, 1055-1058 (2000).

102. O'Connor, L.J., *et al.* Estimating the proportion of disease heritability mediated by gene expression levels. *BioRxiv* (2017).
103. van Dongen, J. & Boomsma, D.I. The evolutionary paradox and the missing heritability of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 122-136 (2013).