The contribution of Neanderthal introgression and natural selection to neurodegenerative diseases

Zhongbo Chen[1,2,3]*, Regina H. Reynolds[2,3]*, Antonio F. Pardiñas[4], Sarah A. Gagliano Taliun[5,6], Wouter van Rheenen[7], Kuang Lin[8], Aleksey Shatunov[9], Emil K. Gustavsson[2,3], Isabella Fogh[9], Ashley R. Jones[9], Wim Robberecht[10,11,12], Philippe Corcia[13], Adriano Chiò[14,15], Pamela J. Shaw[16], Karen E. Morrison[17], Jan H. Veldink[7], Leonard H. van den Berg[7], Christopher E. Shaw[9], John F. Powell[9], Vincenzo Silani[18,19], John A. Hardy[1,20,21,22,23], Henry Houlden[24], Michael J. Owen[4], Martin R. Turner[25], Mina Ryten[2,3], Ammar Al-Chalabi[9]

1. Department of Neurodegenerative Disease, Queen Square Institute of Neurology, University College London (UCL), London, UK.
2. Department of Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, UCL, London, UK.
3. NIHR Great Ormond Street Hospital Biomedical Research Centre, UCL, London, UK.
4. MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK.
5. Department of Medicine & Department of Neurosciences, Université de Montréal, Montréal, Québec, Canada.
6. Montréal Heart Institute, Montréal, Québec, Canada.
7. Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands.
8. Nuffield Department of Population Health, Oxford University, Oxford, UK.
9. Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.
10. Department of Neurology, University Hospital Leuven, Leuven, Belgium.
11. Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease, Leuven, Belgium.
12. Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium.
13. ALS Center, Department of Neurology, CHRU Bretonneau, Tours, France.
14. Rita Levi Montalcini Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy.
15. Azienda Ospedaliera Universitaria Città della Salute e della Scienza, Torino, Italy.
16. Academic Neurology Unit, Department of Neuroscience, Faculty of Medicine, Dentistry and Health, University of Sheffield, Sheffield, UK.
17. School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK.

18. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy.
19. Department of Pathophysiology and Transplantation, Dino Ferrari Center, Università degli Studi di Milano, 20122 Milano, Italy.
20. Reta Lila Weston Institute, Queen Square Institute of Neurology, UCL, London, UK.
21. UK Dementia Research Institute, Queen Square Institute of Neurology, UCL, London, UK.
22. NIHR University College London Hospitals Biomedical Research Centre, London, UK.
23. Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China.
24. Department of Neuromuscular Disease, Queen Square Institute of Neurology, UCL, London, UK.
25. Nuffield Department of Clinical Neurosciences, Oxford University, Oxford, UK.

*ZC and RHR contributed equally to the work.

Correspondence to Professor Ammar Al-Chalabi:
Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, United Kingdom SE5 9RX
Email: ammar.al-chalabi@kcl.ac.uk
Telephone: + 44(0)207 8485244 Fax: + 44(0)207 7080017

Keywords

Neurodegenerative diseases, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, genetics, Neanderthal, natural selection, evolution

Abbreviations

single nucleotide polymorphisms (SNPs); amyotrophic lateral sclerosis (ALS); genome-wide association studies (GWAS); linkage disequilibrium score regression (LDSC); linkage disequilibrium (LD); False discovery rate (FDR)

Word count: 3,188

Reference count: 57

Abstract

Humans are thought to be more susceptible to neurodegeneration than equivalently-aged primates. It is not known whether this vulnerability is specific to anatomically-modern humans or shared with other hominids. The contribution of introgressed Neanderthal DNA to neurodegenerative disorders remains uncertain. It is also unclear how common variants associated with neurodegenerative disease risk are maintained by natural selection in the population despite their deleterious effects. In this study, we aimed to quantify the genome-wide contribution of Neanderthal introgression and positive selection to the heritability of complex neurodegenerative disorders to address these questions.

We used stratified-linkage disequilibrium score regression to investigate the relationship between five SNP-based signatures of natural selection, reflecting different timepoints of evolution, and genome-wide associated variants of the three most prevalent neurodegenerative disorders: Alzheimer's disease, amyotrophic lateral sclerosis and Parkinson's disease.

We found no evidence for enrichment of positively-selected SNPs in the heritability of Alzheimer's disease, amyotrophic lateral sclerosis and Parkinson's disease, , suggesting that common deleterious disease variants are unlikely to be maintained by positive selection. There was no enrichment of Neanderthal introgression in the SNP-heritability of these disorders, suggesting that Neanderthal admixture is unlikely to have contributed to disease risk.

These findings provide insight into the origins of neurodegenerative disorders within the evolution of *Homo sapiens* and addresses a long-standing debate, showing that Neanderthal admixture is unlikely to have contributed to common genetic risk of neurodegeneration in anatomically-modern humans.

Introduction

Encephalisation and the evolution of complex human-specific traits are thought to have increased the susceptibility of *Homo sapiens* to disorders of the brain compared to their aged non-human primate counterparts. [1-3] This is seen in Alzheimer's and Parkinson's disease, which seldom occur naturally on a pathological or phenotypic level in non-human species.[1,2] Likewise, unique motor dysfunction in amyotrophic lateral sclerosis (ALS) supports the selective vulnerability of the highly-developed corticomotoneuronal system in humans.[4-6] Human-lineage-specific genomic sequences have been shown to be enriched for brain-specific elements and risk loci for neurodegenerative disorders.[7] Thus, while positive natural selection may have driven a proportion of human-specific adaptive evolution,[8,9] it may be possible that the same selected variants also influence the risk of neurodegenerative disease.

It is not known whether this neuro-vulnerability arose after divergence from other species or whether it represents a more recent phenomenon, characteristic of modern-day humans over other hominids. As anatomically-modern humans migrated out of Africa 50 to 100 thousand years ago, they interbred with archaic hominins including Neanderthals.[10] As a result, Neanderthal DNA accounts for approximately 1–4% of the modern Eurasian genome.[11-13] While most Neanderthal DNA experienced purifying selective pressures,[14] positive selection of these archaic alleles may have contributed to modern human adaptation to the non-African environment[15] through modulating dermatological[14], immunological[16,17] and metabolic function.[18,19] These introgressed Neanderthal alleles have also been implicated in contributing to the risk of some conditions, including actinic keratosis, depression and obesity.[20]

It remains unclear how much Neanderthal admixture has affected our risk of neurodegenerative disorders. While Neanderthal single nucleotide polymorphisms (SNPs) may be associated with "neurological" phenotypes in electronic health records of European patients, these nervous system traits were not representative of neurodegenerative diseases.[20] More recently, two studies aimed to address the complex nature of medically-relevant traits and the genome-wide influence of Neanderthal admixture using UK Biobank data.[21,22] The former study found that Neanderthal introgression did not significantly contribute to neurological and psychiatric traits and the latter found that introgressed variants were depleted for heritability of high-level cognitive traits.[22] However, neither looked specifically at neurodegenerative diseases of interest.

Thus, to quantify the contribution of Neanderthal admixture to the heritability of neurodegenerative diseases and to examine whether natural selection maintains common genetic risk of these disorders, we tested the relationship between alleles associated with Alzheimer's disease[23], ALS[24] and Parkinson's disease[25] from recent genome-wide association studies (GWAS) with SNP-based signatures of natural selection.[26] Using stratified-linkage disequilibrium score regression (LDSC), we found that there was no significant enrichment of Neanderthal introgression or positive selection in the SNP-heritability of Alzheimer's disease, ALS or Parkinson's disease. Thus, positive selection is unlikely to have played a significant role in the maintenance of common deleterious variants in the genetic architecture of these neurodegenerative diseases.

Methods

Heritability is defined as the fraction of a trait that is explained by inherited genetic variants in a given environment, and is important for understanding the biology of disease.[27] More specific to stratified-LDSC, narrow-sense heritability is defined as the proportion of phenotypic variance that can be attributed to variation in the additive effects of genes.[28] Stratified-LDSC analysis estimates the SNP-based heritability (h^2_{SNP}) of complex traits stratified across different annotations using GWAS data.[29,30] Thus, we were able to use stratified-LDSC to assess the enrichment and depletion of common-h^2_{SNP} of complex neurodegenerative diseases for metrics of Neanderthal introgression and positive selection.[26]

For the metric of Neanderthal introgression, we used the average posterior probability of each human haplotype being the result of Neanderthal admixture estimated by comparing human and Neanderthal genomes (LA).[14,26] We used four metrics of positive selection: integrated haplotype score (iHS), composite of multiple signals (CMS), cross-population extended haplotype homozygosity (XP-EHH), and composite likelihood ratio (CLR). These metrics were chosen to reflect the different timeframes of the selective processes used and described in previous analyses.[26] A detailed summary of the calculation of these metrics is shown in **Supplementary Table 1** and code is available in https://github.com/RHReynolds/als-neanderthal-analysis. LA, CLR and CMS were directly retrieved from published references. For both iHS and XP-EHH, we used the metrics calculated previously[26] using the European superpopulation of the 1,000 Genomes Project Panel 3 dataset, with the African superpopulation used as the second population for XP-EHH.[26,31] Further, we transformed the absolute iHS and $-\log_{10}$ XP-EHH metrics to ensure that all metrics were on a common scale, in which larger values indicate stronger effect of selection or increased probability of

introgression. A summary of the workflow is found in **Figure 1.** iHS measures the amount of extended haplotype homozygosity at a given SNP in the ancestral allele relative to the derived allele and estimates positive selective sweep.[8] CMS identifies the regions under positive selection by combining long-range haplotypes, differentiated alleles and high frequency derived alleles.[32] Both iHS and CMS detect more recent selective sweeps in the last 30,000 years.[8,32] XP-EHH compares two populations to detect an allele that has reached fixation in one population but remains polymorphic in another, identifying alleles that have undergone different selective pressures since population divergence.[33] Lastly, CLR detects incomplete selective sweeps, quantifying the relative influence of recombination and selection and corrects for background selection.[34] It can thus detect older signals from ~60,000 to 240,000 years ago.[26] We did not use any other metrics of purifying selection such as McVicker B-statistic[35] as a metric of background selection is already incorporated into the baseline LDSC model (v.2.2.).

.

We used stratified-LDSC v.1.0.1 (https://github.com/bulik/ldsc/wiki) to test whether these natural selection metrics contribute significantly to heritability of neurodegenerative disease. All natural selection metrics were annotated to the ~9,997,000 SNPs present in the baseline LDSC model (v.2.2), which only includes SNPs with a minor allele frequency of more than 5%. Binary annotations were generated from the natural selection metrics, with thresholds at the top 1%, 2%, 3%, 4% and 5% of the genome-wide values of each metric in the full set of baseline SNPs.[26] This centile approach was used in previous studies due to difficulties defining the thresholds for selection.[26] Annotations were then added individually to the baseline LDSC model of 97 annotations (v.2.2, GRCh37),[36] comprising genome-wide annotations reflecting genetic and LD architecture. HapMap Project Phase 3 (HapMap3)[37] SNPs and 1000 Genomes Project Phase 3[38] European population SNPs were used for the regression and LD reference

panels, respectively. The major histocompatibility complex region was excluded from all analyses owing to the region's complicated and long-range LD patterns. This analysis generated a regression coefficient (τ_c), from which we calculated a two-tailed p-value, testing whether the regression coefficient of the annotation category contributes (either through enrichment or depletion) to the trait heritability conditional upon other annotations in the baseline-LD model. False discovery rate (FDR) multiple testing correction was applied across each GWAS, accounting for the number of annotations run (totalling 25 annotations per GWAS). A stringent FDR-corrected p-value (FDR P) < 0.05 was taken to be significant. We assessed the annotation for h^2_{SNP} contribution in complex brain-related disorders of apparently sporadic Alzheimer's disease[23], Parkinson's disease (excluding 23&Me participants)[25] and ALS (European cases only)[24] using recent corresponding GWAS summary statistics (further description in **Supplementary Table 2**). We chose these three neurodegenerative diseases given that they are the most prevalent of such disorders, have varying mean ages of onset which might affect selection pressures, and have correspondingly well-powered GWAS. All analyses were carried out in R version 4.0.5 (https://www.R-project.org/) and code is available from: https://github.com/RHReynolds/als-neanderthal-analysis.

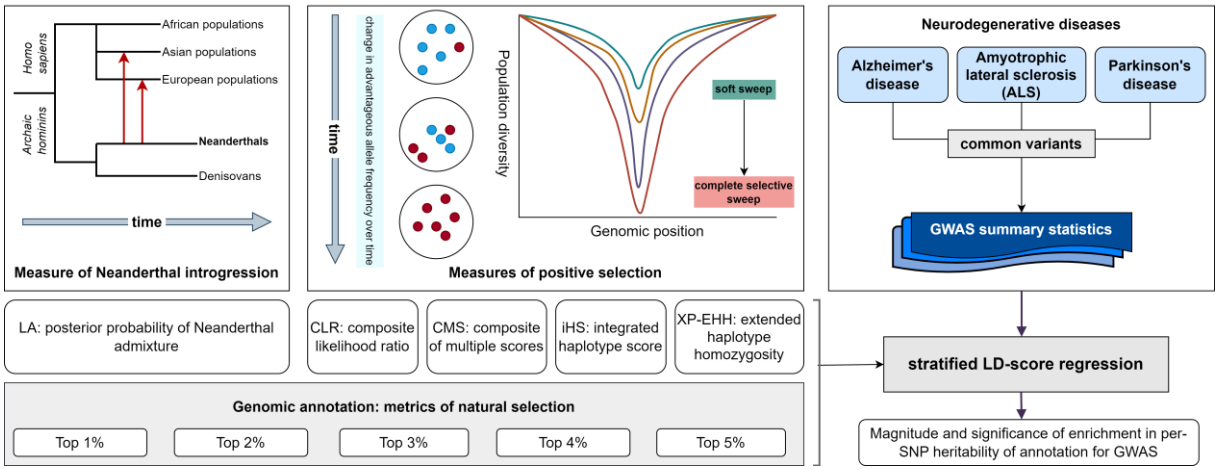

Figure 1. Study workflow. Schematic figure showing key concepts of the study workflow. The red arrows signify admixture from Neanderthals to Eurasian populations; the posterior probability of Neanderthal admixture is measured by LA and partitioned into the top 1–5% metrics and inputted into the stratified-linkage disequilibrium (LD) score regression. Not all gene flow from archaic hominins to *Homo* sapiens is represented in the figure and phylogenetic branches are not to scale. For simplified measures of positive selection, the middle panel shows that with time, an advantageous allele rises in frequency in the population. At the same time, the population diversity decreases at that particular genomic position, and also in other SNPs in linkage disequilibrium that hitchhike with the beneficial allele. This resultant dip in diversity can be shallower in incomplete sweeps and deeper in more complete selective sweeps. Four metrics can measure this positive selection (CLR; CMS; iHS and XP-EHH). These four measures of positive selection were partitioned into the top 1–5% for use in the stratified-LD score regression. We used stratified-LD score regression analysis to test whether these natural selection metrics contribute significantly to heritability of neurodegenerative disease. We assessed the annotation for SNP-heritability contribution in the three most prevalent complex neurodegenerative disorders of Alzheimer's disease, Parkinson's disease and amyotrophic lateral scelosis using recent corresponding GWAS summary statistics. Illustrative figures for natural selection are adapted from a review by Vitti *et al.*[39]

Results

To quantify the contribution of Neanderthal introgression and positive natural selection to the heritability of three neurodegenerative disorders, we used stratified-LDSC to assess enrichment or depletion of the h^2_{SNP} of Alzheimer's disease, ALS and Parkinson's disease for each SNP-based signature of natural selection. For all analyses, we reported an FDR-corrected two-tailed coefficient p-value that tested whether the regression coefficient (τ_c) of an annotation category contributes to trait heritability, conditional upon other annotations that account for the underlying genetic architecture (**Table 1**, **Figure 2**, full results in **Supplementary Table 3**).

After correction for multiple testing, we found no evidence for enrichment in Alzheimer's disease, ALS or Parkinson's disease h^2_{SNP} in alleles most subjected to Neanderthal introgression, as annotated by LA across the top five percentiles (**Figure 2**). This was the case when using both the τ_c taking into other annotations in the baseline model (**Figure 2a**) and enrichment estimate (**Figure 2b**).

A significant depletion of Parkinson's disease h^2_{SNP} was also observed in SNPs under positive selection, as defined by the top 1 to 4% XP-EHH metric (τ_c:-1.120×10^{-8}, -7.780×10^{-9}, -6.770×10^{-9}, -5.890×10^{-9}; FDR P = 0.035, 0.043, 0.043, 0.043 respectively) (**Table 1, Figure 2a**). Importantly, the baseline-LD model controls for metrics of background selection, suggesting that SNPs under positive selection, but under weak or no background selection, are more likely depleted for association with Parkinson's disease compared to other annotations in the baseline model (**Supplementary Table 4**). However, it should be noted that the τ_c values are relatively small. Furthermore, given the enrichment estimates for these metrics were

negative with larger standard errors towards the top centiles reflecting their lower SNP coverage (0.009% at the top 1% metric annotated by XP-EHH) (**Figure 2b**, **Supplementary Table 3**), the interpretation of a negative τ_c here should be interpreted with caution. We then interrogated the contribution of the baseline-LD annotation for background selection for h^2_{SNP} in models including each of our natural selection metrics. For each of the five natural selection metrics across all five centiles used, we found that the background selection metric used in the baseline-LD model was not enriched (τ_c $p < 0.05$) for Parkinson's disease h^2_{SNP} over and above the enrichment of any of the natural selection metrics and the other 96 baseline-LD annotations (**Supplementary Table 4, Figure 2a**). Therefore, these findings suggest that there is unlikely to be an enrichment of positively-selected SNPs in the genetic architecture of Parkinson's disease.

A significant depletion of Alzheimer's disease h^2_{SNP} was found using the top 2, 3 and 4% CLR statistic (τ_c -2.12×10^{-9}, -2.20×10^{-9}, -2.24×10^{-9}; FDR P = 0.023, 0.023, 0.023 respectively). However, there was no nominal depletion across the top 1% of this positive selection metric and given that the corresponding enrichment values are negative, this finding is not conclusive in suggesting that SNPs under positive selection are depleted for association with Alzheimer's disease (**Figure 2**). Given that there was no evidence for other metrics of natural selection to be enriched for Alzheimer's disease h^2_{SNP}, it is also unlikely that positive selection contributed to the evolution of common variants contributing towards the risk of this disease.

After correction for multiple testing, no significant enrichment or depletion of ALS h^2_{SNP} were observed in metrics of positive selection, suggesting that positive selection did not play a role in the common genetic risk of ALS.

		Top 1% scores		Top 2% scores		Top 3% scores		Top 4% scores		Top 5% scores	
GWAS	**Metric**	**τc**	**FDR P**	**τc**	**FDR P**	**τc**	**FDR P**	**τc**	**FDR P**	**τc**	**FDR P**
Alzheimer's disease (AD2019)	**CLR**	-1.48E-09	0.406	-2.12E-09	**0.023**	-2.20E-09	**0.023**	-2.24E-09	**0.023**	-1.91E-09	0.054
	CMS	-3.71E-10	0.911	-1.24E-09	0.703	-1.29E-09	0.639	-1.28E-09	0.639	-1.38E-09	0.635
	iHS	-7.29E-09	0.520	-4.17E-09	0.520	-2.16E-09	0.635	-8.97E-10	0.753	-3.72E-10	0.908
	LA	-1.50E-09	0.635	-1.50E-09	0.544	2.01E-10	0.908	1.06E-09	0.639	9.16E-10	0.639
	XP-EHH	-3.58E-09	0.563	-3.97E-09	0.406	-3.26E-09	0.406	-2.90E-09	0.406	-2.21E-09	0.518
ALS (ALS2021. EUR)	**CLR**	-1.11E-08	0.160	-1.01E-08	0.160	-1.01E-08	0.160	-3.76E-09	0.852	-3.75E-09	0.852
	CMS	-5.99E-09	0.852	-4.66E-09	0.852	-3.38E-09	0.852	-1.12E-09	0.953	3.00E-10	0.976
	iHS	6.21E-08	0.160	2.80E-08	0.219	1.51E-08	0.473	1.03E-08	0.667	6.50E-09	0.852
	LA	2.82E-09	0.852	-1.95E-09	0.852	2.68E-09	0.852	2.63E-09	0.852	1.65E-09	0.852
	XP-EHH	3.97E-08	0.215	2.89E-08	0.208	1.88E-08	0.295	1.89E-08	0.219	1.88E-08	0.215
Parkinson's disease (PD2019.meta 5.ex23&Me)	**CLR**	-3.54E-10	0.955	2.20E-10	0.955	-7.29E-11	0.955	-5.65E-10	0.691	-7.13E-10	0.563
	CMS	-1.95E-09	0.763	-2.19E-10	0.955	-1.21E-10	0.955	-1.22E-10	0.955	-1.28E-10	0.955
	iHS	-6.15E-09	0.558	-4.77E-09	0.426	-4.64E-09	0.285	-4.61E-09	0.183	-4.31E-09	0.163
	LA	-2.98E-09	0.111	-1.78E-09	0.263	-1.10E-09	0.426	-1.23E-09	0.340	-1.06E-09	0.422
	XP-EHH	-1.12E-08	**0.035**	-7.78E-09	**0.043**	-6.77E-09	**0.043**	-5.89E-09	**0.043**	-4.86E-09	0.062

Table 1. Heritability analysis of natural selection metrics. Stratified-LDSC results for SNPs defined by top genome-wide percentiles (1–5%) of all SNPs annotated for each natural selection metric. The regression coefficient (τ_c) represents the contribution of the annotation to trait SNP-heritability (h^2_{SNP}), controlling for all other annotations within the baseline model. The FDR-corrected p-value (FDR P) is the coefficient p-value following correction for multiple testing. Coefficient FDR P<0.05 and corresponding τ_c values are highlighted in bold. Neanderthal introgression metric (LA) indicates posterior probability of Neanderthal admixture. Positive selection metrics are composite likelihood ratio statistic (CLR), composite of multiple scores (CMS), integrated haplotype score (iHS), cross-population extended haplotype homozygosity (XP-EHH). Genome-wide association studies (GWAS) include AD2019 (Alzheimer's disease, Jansen *et al.* 2019)[23], ALS2021.EUR (European cases only from amyotrophic lateral sclerosis GWAS, van Rheenen *et al.* 2021)[24] and PD2019.met5.ex23&Me (Nalls et al. 2019, excluding 23&Me participants).[25] Full results are shown in **Supplementary Table 3**.

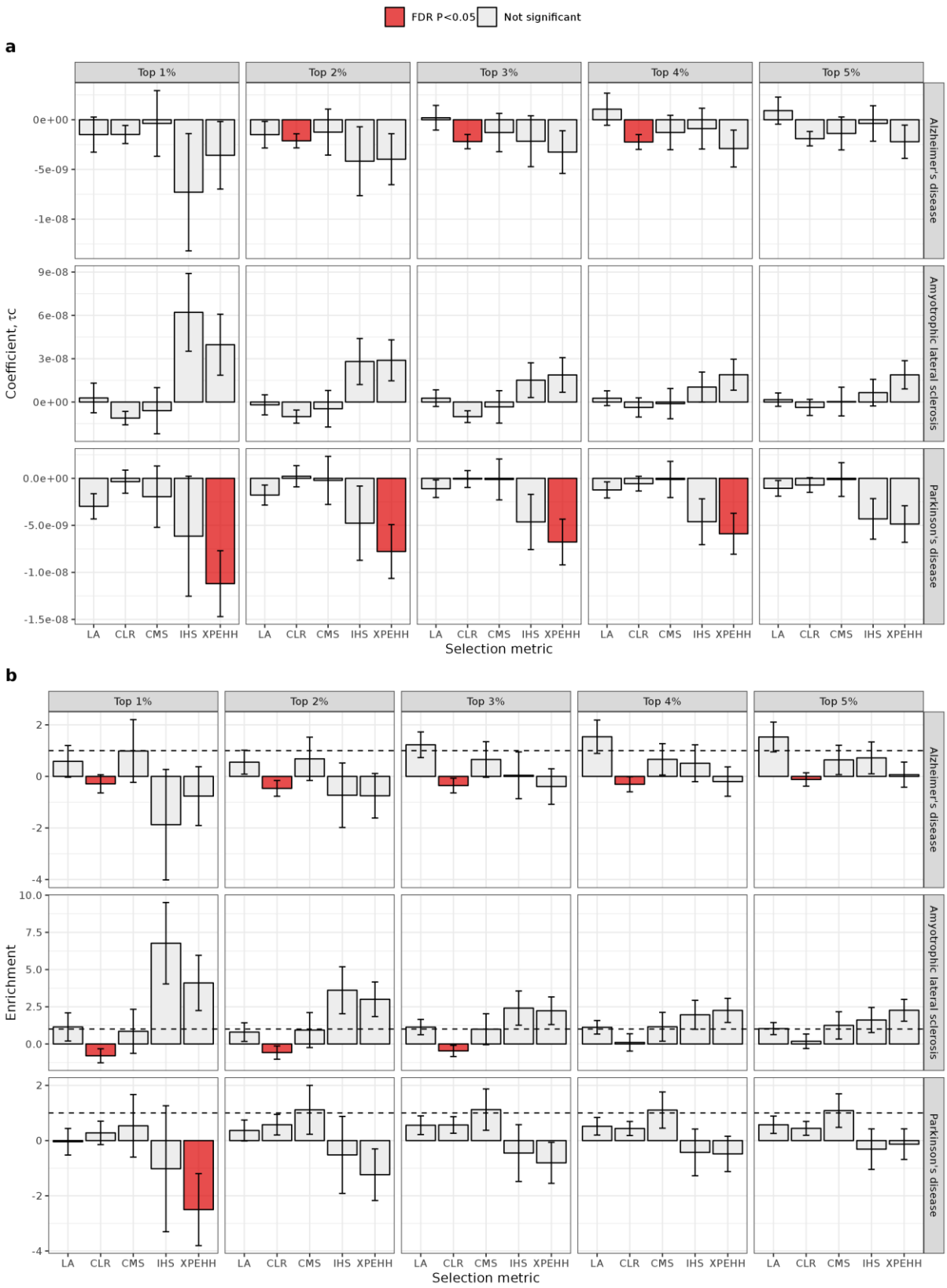

Figure 2. Stratified-linkage disequilibrium score regression analysis across three neurodegenerative disorders comparing SNP-based signatures of Neanderthal introgression and natural selection, stratified across top genome-wide percentiles (1–5%) of annotated SNPs. (a)The regression coefficient (τ_c) represents the contribution of the

annotation to trait SNP-heritability (h^2_{SNP}), controlling for all other annotations within the baseline model. The error bars represent 95% confidence intervals. Positive values represent enrichment of trait h^2_{SNP} within that annotation. Negative values represent depletion of trait h^2_{SNP} within that annotation compared to other annotations within the baseline model. Bars filled with red shading represent significant FDR-corrected two-sided p-values of the regression coefficient (FDR P < 0.05). **(b) Enrichment estimates** from LD score regression analysis. Enrichment values between 0 and 1 indicate a depletion of h^2_{SNP} in an annotation category (i.e. showing a lower contribution than expected for a given number of SNPs). Enrichment values larger than 1 indicate an enrichment of h^2_{SNP} in an annotation category. The error bars represent 95% confidence intervals. Bars filled with red shading represent significant FDR-corrected two-sided p-values of the regression coefficient (FDR P<0.05). Neanderthal introgression metric (LA) indicates posterior probability of Neanderthal admixture. Positive selection metrics are composite likelihood ratio statistic (CLR), composite of multiple scores (CMS), integrated haplotype score (iHS), cross-population extended haplotype homozygosity (XP-EHH). Genome-wide association studies (GWAS) include AD2019 (Alzheimer's disease, Jansen *et al.* 2019)[23], ALS2021.EUR (European cases only from amyotrophic lateral sclerosis GWAS, van Rheenen *et al.* 2021)[24] and PD2019.met5.ex23&Me (Nalls *et al.* 2019, excluding 23&Me participants).[25]

Discussion

Humans are particularly vulnerable to neurodegeneration.[1,2] For example, in ALS, the most frequent neurodegenerative disease of mid-life, motor neurone degeneration results in disruption of multiple motor functions that are key to survival, and therefore of importance for evolutionary adaptation.[40] Any genetic variation predisposing to motor neurone degeneration might therefore be expected to be under major negative selection pressures. Thus, several possible evolutionary explanations may exist for common alleles contributing to neurodegenerative disease risk to persist in the population despite their deleterious effects.

First, any genetic variation predisposing to disease could have a corresponding benefit and therefore be positively selected. This is a mechanism by which the persistence of Neanderthal-derived sequences in the modern Eurasian human genome has been explained.[14] For example, a Neanderthal haplotype associated with protection against severe forms of SARS-CoV-2 infection (and other RNA viruses) is also linked to Alzheimer's disease risk.[41,42 43]Alternatively, deleterious variants in LD with an advantageous allele may have hitchhiked during positive selection, rising in frequency in the population.[44] This would be consistent with the proposed Northern founders of the p.91D>A *SOD1* variant in ALS[45] and the pathogenic *C9orf72* hexanucleotide repeat expansion associated with ALS and frontotemporal dementia.[46,47] Secondly, because neurodegenerative disorders are diseases of ageing, any genetic susceptibility might act after child-rearing years, and therefore outside the age window in which negative selection pressure could have an impact on allele frequency.[48] In this scenario, the negative effect of the genetic variant is mitigated by the timing of neurodegeneration. Thirdly, neurodegenerative diseases might result from multiple rare variants, each unique in the affected person, and therefore be too rare for selection to have a significant impact.[49] With the increased availability of high-depth next generation sequencing, many rare variants have now been found

to be associated with ALS[24,50,51], Parkinson's disease[25,52] and Alzheimer's disease[53,54], but common variants also contribute to risk. To address these hypotheses of how natural selection may have led to the persistence of common genetic risk variants of neurodegenerative disorders, we used stratified-LDSC to test the relationship between neurodegeneration-related SNPs and SNP-based signatures of natural selection.

We found no significant enrichment of Alzheimer's disease, ALS or Parkinson's disease h^2_{SNP} associated with Neanderthal admixture. This observation is consistent with recent studies using UK Biobank data that showed introgressed variants were depleted for contribution to the heritability of most complex traits or at least not enriched.[21,22] These studies did not specifically focus on neurodegenerative diseases. Our findings suggest that Neanderthal admixture is unlikely to have maintained the common genetic variant architecture of these neurodegenerative diseases in modern humans.

Positive selection is seen as an evolutionary mechanism for adaptation to a new environment.[39] In this situation, the beneficial allele sweeps to high frequencies and towards fixation (defined as 100% frequency) together with other variants on the same haplotype, reducing the population genetic diversity in a selective sweep.[39] SNPs under positive selection were found to be depleted for association with Parkinson's disease h^2_{SNP} compared to other annotations in the baseline model (τ_c) but the simultaneous presence of a negative enrichment estimate suggests that there is at least no evidence of enrichment for positive selection in the heritability of these diseases. This suggests that positive selection does not maintain common risk alleles in Parkinson's disease. Heritability needs to be considered using this approach to estimate the contribution of natural selection to disease risk. The negative enrichment estimates may also, in part, stem from the relatively low heritability of Parkinson's disease compared to other neurodegenerative and

psychiatric diseases (0.22 according to estimates from genome-wide meta-analysis)[25], making it more difficult to definitely interpret the depletion in τ_c as a true depletion in Parkinson's disease h^2_{SNP}. Therefore, given the relatively low heritability of Parkinson's disease, a strong environmental component is implicated in the contribution to disease. This would mean that natural selection would have a limited role in Parkinson's disease risk. The lack of enrichment of alleles that have been positively selected through Neanderthal admixture or other selective sweeps implicate a strong gene-environment interaction. Indeed, gene-environment interaction frameworks have previously been proposed for the persistence of deleterious common variants in the population gene pool.[55]

In Alzheimer's disease, this depletion of positively selected SNPs was even less clear, indicating that positive selection is unlikely to be associated with persistence of these common disease-associated variants.

For ALS, we saw no evidence for the enrichment of positive selection in contributing to its trait heritability. This relationship with positive selection is intriguing, providing support for an increased common allele frequency despite the selective removal of alleles that are deleterious, perhaps in combination with mitigation from the timing of disease onset. This is consistent with a previous study that uses UK Biobank data that found positive selection, as tagged by CMS, is not enriched for heritability of neurological traits.[56] In addition, these findings showing the lack of enrichment from positive selection imply that the genetic architecture of Parkinson's and Alzheimer's diseases, and ALS, may stem from multiple, rare, pleiotropic variants rather than polygenic variants with small effects.

This study has several limitations. Firstly, our analyses only studied individuals of European ancestry while Neanderthal introgression also occurred in non-European populations, where the extent of introgression may have differed and thus affects the sensitivity of the analysis. Secondly, the h^2_{SNP} estimates using stratified-LDSC analysis are limited by the quality of LD information underpinning the heritability calculations[30] and the sample size of the GWAS, although we attempted to use only well-powered studies. LDSC also does not take into account the major histocompatibility complex region while Neanderthal alleles have been shown to play a role in modern immune function, thus missing those variants with pleiotropic effects in the nervous system.[16] We are also excluding rarer Neanderthal introgressed and positively-selected variants by using a MAF cut-off and thus are not able to conclude on the heritability contribution of these variants with lower population frequency. Thirdly, each metric of natural selection and Neanderthal admixture has its own strengths and shortcomings and our analysis is limited by these measures.[39] However, we attempted to overcome this issue by using a range of SNP-based signatures.[26] We also note that the negative enrichment estimates are difficult to interpret in this context, possibly due to the SNP coverage of the annotations, which we have tried to compensate for by using higher percentiles of each selection metric. Lastly, this analysis did not take into account structural variants, short tandem repeats or other repetitive genomic elements that may have been acquired through positive selection. A recent study showed a number of repeat elements have risen *ab initio* in *Homo sapiens* which may have implications for positive selection in disease given that these are the most mutable regions of the genome.[57]

In this analysis, we quantified the contribution of positive selection and Neanderthal introgression to the heritability of Alzheimer's disease, ALS and Parkinson's disease using stratified-LDSC. We found no significant enrichment of Neanderthal introgression in the SNP-

heritability of these neurodegenerative diseases. We also found that genomic regions with positive selection showed no evidence for the contribution in Alzheimer's disease, ALS and Parkinson's disease SNP-heritability. This suggests that positive selection does not maintain common risk variants for these disorders in the population even after controlling for background selection. These findings provide further insight into the evolution of the genetic architecture of these disorders.

Contributorship statement

ZC, MRT, MR and AAC conceived and designed the study. ZC, RHR and SAGT critically analysed and interpreted the data. RHR set up and ran the stratified-LDSC analysis pipeline. AFP provided metrics for natural selection. WvR provided summary statistics from the ALS GWAS used. ZC wrote the first draft of the manuscript. JAH, HH, MJO, MRT, MR and AAC supervised the study. KL, AS, EKG, IF and ARJ contributed to the analyses. WR, PC, AC, PJS, KEM, JHV, LHvdV, CES, JFP, and VS contributed to GWAS data used in this study. All authors approved the manuscript and contributed to generation of the submitted draft.

Acknowledgements

Z.C. is funded by a Leonard Wolfson Foundation clinical research fellowship in neurodegeneration. I.F. receives salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London and Maudsley NHS Foundation Trust and King's College London. I.F. is also funded by the Motor Neurone Disease Association (Grant number 905-793, 6058). W.R. is supported through the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the Laevers Fund for ALS Research, the ALS Liga België, the fund 'Een Hart voor ALS' and the fund 'Opening the Future'. MRT is supported by the Motor Neurone Disease Association. AAC is an NIHR Senior Investigator (NIHR202421). This is an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organisations under the aegis of JPND - www.jpnd.eu (United Kingdom, Medical Research Council (MR/L501529/1; MR/R024804/1) and Economic and Social Research Council (ES/L008238/1)) and through the Motor

Neurone Disease Association, My Name'5 Doddie Foundation, and Alan Davidson Foundation. This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The work leading up to this publication was funded by the European Community's Health Seventh Framework Programme (FP7/2007–2013; grant agreement number 259867 and 340429). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests

AAC reports consultancies or advisory boards for Amylyx, Apellis, Biogen, Brainstorm, Cytokinetics, GenieUs, GSK, Lilly, Mitsubishi Tanabe Pharma, Novartis, OrionPharma, Quralis, and Wave Pharmaceuticals.**Patient consent for publication:** Not applicable.

Ethics approval: This study does not involve human participants.

Data availability statement: All data relevant to the study are included in the article or uploaded as supplementary information. Code is available from: https://github.com/RHReynolds/als-neanderthal-analysis.

References

1. Walker LC, Jucker M. The Exceptional Vulnerability of Humans to Alzheimer's Disease. *Trends in molecular medicine*. Jun 2017;23(6):534-545. doi:10.1016/j.molmed.2017.04.001
2. Diederich NJ, James Surmeier D, Uchihara T, Grillner S, Goetz CG. Parkinson's disease: Is it a consequence of human brain evolution? *Movement disorders : official journal of the Movement Disorder Society*. Feb 13 2019;doi:10.1002/mds.27628
3. Cookson MR. Evolution of neurodegeneration. *Current biology : CB*. Sep 11 2012;22(17):R753-61. doi:10.1016/j.cub.2012.07.008
4. Eisen A, Kim S, Pant B. Amyotrophic lateral sclerosis (ALS): a phylogenetic disease of the corticomotoneuron? *Muscle Nerve*. Feb 1992;15(2):219-24. doi:10.1002/mus.880150215
5. Chen Z, Lin K, Macklis JD, Al-Chalabi A. Proposed association between the hexanucleotide repeat of C9orf72 and opposability index of the thumb. *Amyotrophic lateral sclerosis & frontotemporal degeneration*. May 2017;18(3-4):175-181. doi:10.1080/21678421.2016.1257024
6. Henderson RD, Garton FC, Kiernan MC, Turner MR, Eisen A. Human cerebral evolution and the clinical syndrome of amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*. 2019;90(5):570-575. doi:10.1136/jnnp-2017-317245
7. Chen Z, Zhang D, Reynolds RH, et al. Human-lineage-specific genomic elements are associated with neurodegenerative disease and APOE transcript usage. *Nature communications*. 2021/04/06 2021;12(1):2076. doi:10.1038/s41467-021-22262-5
8. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. Mar 2006;4(3):e72. doi:10.1371/journal.pbio.0040072

9. Mangan RJ, Alsina FC, Mosti F, et al. Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. *Cell*. 2022/11/23/ 2022;185(24):4587-4603.e23. doi:https://doi.org/10.1016/j.cell.2022.10.016
10. Stewart JR, Stringer CB. Human evolution out of Africa: the role of refugia and climate change. *Science (New York, NY)*. Mar 16 2012;335(6074):1317-21. doi:10.1126/science.1215627
11. Green RE, Krause J, Briggs AW, et al. A draft sequence of the Neandertal genome. *Science (New York, NY)*. May 7 2010;328(5979):710-22. doi:10.1126/science.1188021
12. Prüfer K, de Filippo C, Grote S, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (New York, NY)*. Nov 3 2017;358(6363):655-658. doi:10.1126/science.aao1887
13. Skov L, Coll Macià M, Sveinbjörnsson G, et al. The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature*. Jun 2020;582(7810):78-83. doi:10.1038/s41586-020-2225-9
14. Sankararaman S, Mallick S, Dannemann M, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*. Mar 20 2014;507(7492):354-7. doi:10.1038/nature12961
15. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*. 2015/06/01 2015;16(6):359-371. doi:10.1038/nrg3936
16. Abi-Rached L, Jobin MJ, Kulkarni S, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science (New York, NY)*. Oct 7 2011;334(6052):89-94. doi:10.1126/science.1209202
17. Yan SM, Sherman RM, Taylor DJ, et al. Local adaptation and archaic introgression shape global diversity at human structural variant loci. *Elife*. Sep 16 2021;10doi:10.7554/eLife.67615
18. Khrameeva EE, Bozek K, He L, et al. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nature communications*. 2014;5(1):1-8.
19. Silvert M, Quintana-Murci L, Rotival M. Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation. *American journal of human genetics*. Jun 6 2019;104(6):1241-1250. doi:10.1016/j.ajhg.2019.04.016
20. Simonti CN, Vernot B, Bastarache L, et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science (New York, NY)*. Feb 12 2016;351(6274):737-41. doi:10.1126/science.aad2149
21. Dannemann M, Milaneschi Y, Yermakovich D, et al. Neandertal introgression partitions the genetic landscape of neuropsychiatric disorders and associated behavioral phenotypes. *Translational Psychiatry*. 2022/10/05 2022;12(1):433. doi:10.1038/s41398-022-02196-2
22. McArthur E, Rinker DC, Capra JA. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nature communications*. 2021/07/22 2021;12(1):4481. doi:10.1038/s41467-021-24582-y
23. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics*. 2019/03/01 2019;51(3):404-413. doi:10.1038/s41588-018-0311-9
24. van Rheenen W, van der Spek RAA, Bakker MK, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nature genetics*. Dec 2021;53(12):1636-1648. doi:10.1038/s41588-021-00973-1
25. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*. Dec 2019;18(12):1091-1102. doi:10.1016/s1474-4422(19)30320-5
26. Pardinas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics*. Mar 2018;50(3):381-389. doi:10.1038/s41588-018-0059-2
27. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*. 2008/04/01 2008;9(4):255-266. doi:10.1038/nrg2322
28. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. *Nature Reviews Genetics*. 2013/02/01 2013;14(2):139-149. doi:10.1038/nrg3377

29. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*. Mar 2015;47(3):291-5. doi:10.1038/ng.3211
30. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*. Nov 2015;47(11):1228-35. doi:10.1038/ng.3404
31. Consortium GP, Auton A, Brooks L, Durbin R, Garrison E, Kang H. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
32. Grossman SR, Shylakhter I, Karlsson EK, et al. A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science (New York, NY)*. 2010;327(5967):883-886. doi:doi:10.1126/science.1183863
33. Sabeti PC, Schaffner SF, Fry B, et al. Positive natural selection in the human lineage. *Science (New York, NY)*. 2006;312(5780):1614-1620.
34. Vy HM, Kim Y. A Composite-Likelihood Method for Detecting Incomplete Selective Sweep from Population Genomic Data. *Genetics*. Jun 2015;200(2):633-49. doi:10.1534/genetics.115.175380
35. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS genetics*. May 2009;5(5):e1000471. doi:10.1371/journal.pgen.1000471
36. Hujoel MLA, Gazal S, Hormozdiari F, van de Geijn B, Price AL. Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species. *American journal of human genetics*. Apr 4 2019;104(4):611-624. doi:10.1016/j.ajhg.2019.02.008
37. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. Sep 2 2010;467(7311):52-8. doi:10.1038/nature09298
38. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. Nov 1 2012;491(7422):56-65. doi:10.1038/nature11632
39. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annu Rev Genet*. 2013;47:97-120. doi:10.1146/annurev-genet-111212-133526
40. Eisen A, Turner MR, Lemon R. Tools and talk: an evolutionary perspective on the functional deficits associated with amyotrophic lateral sclerosis. *Muscle & nerve*. Apr 2014;49(4):469-77. doi:10.1002/mus.24132
41. Sabeti PC, Schaffner SF, Fry B, et al. Positive Natural Selection in the Human Lineage. *Science (New York, NY)*. 2006;312(5780):1614-1620. doi:doi:10.1126/science.1124309
42. Magusali N, Graham AC, Piers TM, et al. A genetic link between risk for Alzheimer's disease and severe COVID-19 outcomes via the OAS1 gene. *Brain : a journal of neurology*. 2021;144(12):3727-3741. doi:10.1093/brain/awab337
43. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. 2020/11/01 2020;587(7835):610-612. doi:10.1038/s41586-020-2818-3
44. Chun S, Fay JC. Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS genetics*. Aug 2011;7(8):e1002240. doi:10.1371/journal.pgen.1002240
45. Saeed M, Yang Y, Deng HX, et al. Age and founder effect of SOD1 A4V mutation causing ALS. *Neurology*. May 12 2009;72(19):1634-9. doi:10.1212/01.wnl.0000343509.76828.2a
46. Smith BN, Newhouse S, Shatunov A, et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *European journal of human genetics : EJHG*. Jan 2013;21(1):102-8. doi:10.1038/ejhg.2012.98
47. Pliner HA, Mann DM, Traynor BJ. Searching for Grendel: origin and global spread of the C9ORF72 repeat expansion. *Acta neuropathologica*. Mar 2014;127(3):391-6. doi:10.1007/s00401-014-1250-x
48. Gluckman PD, Low FM, Buklijas T, Hanson MA, Beedle AS. How evolutionary principles improve the understanding of human health and disease. *Evol Appl*. Mar 2011;4(2):249-63. doi:10.1111/j.1752-4571.2010.00164.x
49. Dilliott AA, Abdelhady A, Sunderland KM, et al. Contribution of rare variant associations to neurodegenerative disease presentation. *npj Genomic Medicine*. 2021/09/28 2021;6(1):80. doi:10.1038/s41525-021-00243-3

50. Smith BN, Ticozzi N, Fallini C, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron*. Oct 22 2014;84(2):324-31. doi:10.1016/j.neuron.2014.09.027
51. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science (New York, NY)*. Mar 27 2015;347(6229):1436-41. doi:10.1126/science.aaa3650
52. Robak LA, Jansen IE, van Rooij J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain : a journal of neurology*. Dec 1 2017;140(12):3191-3203. doi:10.1093/brain/awx285
53. Prokopenko D, Morgan SL, Mullin K, et al. Whole-genome sequencing reveals new Alzheimer's disease–associated rare variants in loci related to synaptic function and neuronal development. *Alzheimer's & Dementia*. 2021;17(9):1509-1527. doi:https://doi.org/10.1002/alz.12319
54. Sims R, van der Lee SJ, Naj AC, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics*. 2017/09/01 2017;49(9):1373-1384. doi:10.1038/ng.3916
55. Uher R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Molecular psychiatry*. 2009/12/01 2009;14(12):1072-1082. doi:10.1038/mp.2009.85
56. Song W, Shi Y, Wang W, et al. A selection pressure landscape for 870 human polygenic traits. *Nature Human Behaviour*. 2021/12/01 2021;5(12):1731-1743. doi:10.1038/s41562-021-01231-4
57. Sulovari A, Li R, Audano PA, et al. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proceedings of the National Academy of Sciences*. 2019;116(46):23243-23253. doi:10.1073/pnas.1912175116