No phenotypic consequences of archaic hominin alleles in present-day humans

Barbara Molz[1†], Mikel Lana Alberro[1,2†], Else Eising[1], Dick Schijven[1,4], Gökberk Alagöz[1], Clyde Francks[1,3,4], Simon E. Fisher[1,4*]

Affiliations:
[1]Language & Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands;
[2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany;
[3]Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, Netherlands;
[4]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands;

† These authors contributed equally
* Corresponding author: Simon E. Fisher simon.fisher@mpi.nl

ABSTRACT

Recent advances in paleo-genetics allowed the identification of protein-coding changes apparently fixed on the lineage leading to *Homo sapiens*, by comparing genomes of present-day humans and archaic hominins. Although such genomic differences are thought to make key contributions to distinctly modern human traits, experimental validation of their potential impact was so far restricted to functional assays and model organisms. With the availability of large-scale genetically informative population databases, it now becomes possible to identify present-day carriers of rare archaic alleles of interest and to directly assess putative phenotypic consequences in living humans. We queried exome sequencing data of around half a million people in the UK Biobank in search of carriers of archaic alleles at 37 genomic positions with supposedly fixed human-specific changes. This search yielded 103 carriers of the archaic allele for 17 positions, with diverging allele counts across ancestries. We contrasted carriers of an exemplary archaic allele in *SSH2* with a curated set of non-carriers, observing no deviation from the norm in a range of health, psychological, and cognitive traits. We also identified 62 carriers of the archaic allele of a missense change in the *TKTL1* gene, previously reported to have large effects on cortical neurogenesis based on functional analyses in brain organoids and animal models. However, human carriers of the archaic *TKTL1* allele did not show differences in anatomical brain measures and qualification level, compared to non-carriers. These results highlight the importance of investigating diverse ancestral populations for a more accurate representation of shared human variation and challenge the notion of permanently fixed genetic changes that set *Homo sapiens* apart from Neandertals and Denisovans. Lastly, we propose that future investigations should assess effects of multiple archaic alleles in aggregate, since any single genetic change is unlikely to itself explain the emergence of complex human traits.

Introduction

Understanding the origins of modern humans and how our ancestors developed sophisticated cultural, social and behavioural skills has been a central issue for many fields of science[1–3]. While latest research is gradually reaching a consensus that cognitive capacities of Neandertals were greater than previously appreciated, the question remains why *Homo sapiens* outlived its archaic cousins and was able to migrate all across the globe[2,4–6]. Advances in high-throughput DNA sequencing and the availability of three high quality Neandertal genomes[7–9] enabled comparative genomic approaches, opening up new ways to reconstruct aspects of the evolutionary history of *Homo sapiens*. In particular, such approaches yielded catalogues of missense variants (changes that substitute one amino acid for another in an encoded protein) that occurred after *Homo sapiens* split from its common ancestor with Neandertals ~600,000 years ago, and that reached (near) fixation on our lineage. These human-specific fixed derived alleles have been hailed as promising entry points for explaining human origins, given their enrichment in genes that are relevant for human-specific traits and involved in cortical development and neurogenesis[1,2,7].

Since a missense variant can potentially arise and spread through a population without any consequence to properties or functions of the encoded protein, experimental validation is crucial to determine the functional significance of derived alleles. In a prominent example, Pinson et al.[10] investigated the impacts of a lysine-to-arginine substitution in human *TKTL1* (chrX:154,315,258; G->A) by comparing the archaic and derived alleles using genome-edited cerebral organoid and *in vivo* models, as well as in primary brain tissue. The authors observed substantial differences between samples carrying the Neandertal and *Homo sapiens* versions of *TKTL1* in basal radial glia abundance and neurogenesis, and suggested that the modern human-derived allele might have played a key role in evolutionary expansion of the brain's frontal lobe. However, despite the strengths of cerebral organoids for modelling events in early embryogenesis[11], cellular diversity and transcriptomic programmes of these models do not fully recapture human brain development, and lack insights from diversity across genetic ancestries[12]. Similarly, expression of "humanised" genes in primary brain tissue of non-human species may lead to non-specific artefacts[12–14], due to inter-species differences in genetic background. Thus, the actual consequences of any such modern human-derived genetic changes may be more complex than those which can be observed in cellular/animal models[15].

An alternative approach for evaluating broader biological impact depends on the identification of present-day carriers of archaic alleles at genomic positions that differ between modern humans and Neandertals[3]. Indeed, databases like gnomAD highlight the existence of individuals carrying these archaic single nucleotide variants (aSNVs)[12], albeit in low numbers. The recent availability of large-scale biobanks with exome sequencing and trait data enables not only the detection of aSNVs in living humans, but also the investigation of their putative phenotypic consequences.

In this study, we used the UK Biobank (UKB), a large-scale population resource with both exome and dense phenotype data available from around half a million individuals[16]. This offers a unique opportunity to i) determine the frequencies of present-day aSNV carriers, and ii) assess how phenotypic profiles of carriers of the archaic allele compare to individuals that are homozygous for the derived present-day human allele. We focused our efforts on a catalogue of putative fixed genomic positions established from a prior survey of potential human-specific changes[2] and searched for

carriers of ancestral alleles among UKB participants. To gain insight into the phenotypic profile of an exemplary aSNV in *SSH2*, we contrasted identified carriers with a curated set of non-carriers, homozygous for the derived allele, assessing a range of phenotypic traits. Given the especially dramatic effects of the *TKTL1* aSNV on neurogenesis reported by Pinson et al. in their cellular and animal models[10], we also included this high-frequency human-specific change in our investigations. Specifically, we identified carriers of the archaic *TKTL1* allele and used the available neuroimaging data[17] to study putative effects of the aSNV on brain morphology and cognitive traits.

Results

103 carriers of archaic SNVs in 17 positions identified in UK Biobank across different ancestries

Based on the Kuhlwilm & Boeckx[2] catalogue of single nucleotide changes that distinguish modern humans and archaic hominins, we curated a list of 42 fixed missense changes with an allele frequency of one (AF = 1) at the time of publication, indicating complete fixation within the investigated modern human populations (see Methods, Supplementary Table 1). After quality control (see Methods) we then queried the whole-exome sequencing data of approximately 455,000 individuals[18–20] of the UKB to identify possible carriers of the archaic allele at 37 positions. We investigated four ancestry superclusters: European, African, East & South Asian (Supplementary Figure 1A). In total, we identified 103 unique individuals carrying 118 aSNVs in 13 protein-coding genes. All were heterozygous carriers, except for a female carrying a homozygous aSNV in *GRM6* (chr5:178994530) (Table 1), a gene encoding the ON bipolar metabotropic glutamate receptor, which overall also represents the genomic position with the largest carrier count.

We observed diverging carrier counts for aSNVs across ancestry superclusters. Even though the UKB is comprised of predominantly European ancestry individuals[16] with only around 0.5% individuals of African and 1% of East Asian descent, nearly equal numbers of aSNV carriers were identified in the European, East Asian and African ancestry superclusters, highlighting allele frequency differences for these rare variants, and a bias towards European data being used previously to identify aSNVs.
We identified five individuals carrying a combination of three aSNVs in *SPAG5* (chr17:28,592,759; chr17:28,598,560; chr17:28,598,560), and two pairs of carriers who carry a combination of two aSNVs on *ADAM18* and *KNL1,* respectively (*ADAM18*: chr8:39,680,099; chr8:39,706,833; *KNL1*: chr15:40,620,662; chr15:40,623,442). In each case the aSNVs found in the same carriers were in tight linkage disequilibrium, thus were likely inherited together.

We also queried the relatedness status (up to 3rd degree) of identified carriers and found only one related pair carrying an aSNV in *SSH2*. Thus, it is unlikely that the allele counts of identified aSNVs in our study are inflated due to relatedness.

Gene	*KIF26B*		*NOTO*		*GRM6*			*ADAM18*		*ADAM18*		*DCHS1*		*KNL1*		*KNL1*	
Chromosome	1		2		5			8		8		11		15		15	
Position (hg 38)	245419603		73210883		178994530			39680099		39706833		6633538		40620662		40623442	
Reference Allele	A		T		G			C		G		C		G		A	
Archaic allele	G		A		T			T		A		T		A		G	
	Het	Hom (Ref)	Het	Hom (Ref)	Hom (Alt)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)
SAS		8585		8585			8585		8585		8570		8585		8583		8585
EAS		1887	3	1884			1886	2	1885	2	1881		1887	1	1885	1	1887
EUR		423885		423883		3	423838	1	423882		423457		423887	3	423652		423885
AFR	2	5091		5092		12	5077		5092		5075		5092		5087		5092
Uncategorized	3	13343	3	13343	1	28	13316		13346		13319	1	13345	2	13339	1	13346
Total N	5	452791	6	452787	1	43	452702	3	452790	2	452302	1	452796	6	452546	2	452795

Gene	*ZNF106*		*SPAG5*		*SPAG5*		*SPAG5*		*SSH2*		*RFNG*		*GREB1L*		*LMNB2*		*C3*	
Chromosome	15		17		17		17		17		17		18		19		19	
Position (hg 38)	42450114		28592016		28592759		28598560		29632016		82049104		21505418		2434035		6685100	
Reference Allele	C		G		C		A		T		G		A		A		G	
Archaic allele	T		C		T		G		C		A		G		T		A	
	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)	Het	Hom (Ref)
SAS		8585		8585		8584		8585		8585		8584		8585		8585		8584
EAS	4	1883		1887		1887		1887		1887		1887	1	1886	2	1885		1887
EUR		423887		423887	1	423883		423885	21	423866	2	423832		423877		423886	2	423861
AFR		5092	5	5092	5	5092	5	5092		5092		5092		5092		5092		5092
Uncategorized	1	13345		13346		13346		13346		13345		13343		13346		13346		13346
Total N	5	452792	5	452797	6	452792	5	452795	21	452775	2	452738	1	452786	2	452794	2	452770

Table 1: Overview of identified aSNV carriers in UK Biobank. Genotype count of carriers of each aSNV and respective individuals homozygous for the derived allele are noted per ancestry supercluster. Genomic positions are based on hg38; Het = Heterozygous, Hom = Homozygous, Ref = reference allele; SAS = South Asian; EAS = East Asian; EUR = European; AFR = African.

Phenotypes of aSNV carriers in *SSH2* do not deviate from matched non-carriers.

Next, we showed how availability of biobank trait data can be used to query whether aSNVs have major phenotypic consequences in living humans. As an example, we chose an aSNV in *SSH2* (chr17:29632016), since the encoded protein is a protein phosphatase with enzymatic properties regulating actin filament dynamics and possible functions in neurite outgrowth[2,21–23], and because this variant was found in a relatively large number of unrelated carriers within a strict ancestry cluster (N=19, see Methods). We chose the following traits for phenotypic assessment: body composition measures (body mass index, whole body fat mass), height, overall health rating, smoking status and highest qualification level as an indication of educational attainment. These phenotypes were selected a priori based on previously identified GWAS trait associations of *SSH2*[24] that further overlapped with traits linked to Neandertal admixture[1,25–29].

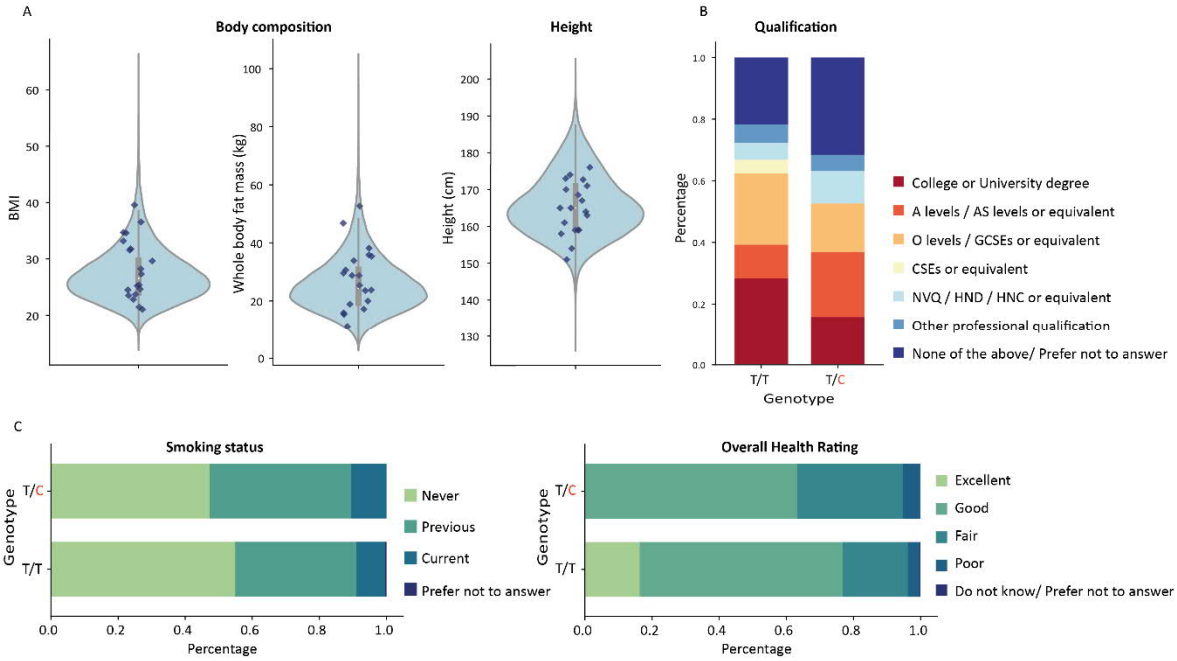

Figure 1: Phenotypic effects of aSNV carriers in *SSH2*. (A) Values of continuous traits are shown for each aSNV carrier as dark blue diamonds. Violin plots show the phenotypic distribution of matched set of non-carriers, with boxplots indicating the 25th and 75th percentiles, and whiskers representing 1.5 times the inter quartile range (IQR); (B & C) Stacked bar plots showing the percentage of highest qualification level, as well as health related measures for each genotype: T/T for matched non-carriers and T/C for aSNV carriers (N_{aSNV} = 19; $N_{Non\text{-}carrier}$ = 39,501). A level = Advanced level, AS level = Advanced Subsidiary level, O level = Ordinary level, GCSE = General Certificate of Secondary Education, CSE = Certificate of Secondary Education, NVQ = National Vocational Qualification, HND = Higher National Diploma, HNC = Higher National Certificate.

For all continuous traits, carriers of the aSNVs on *SSH2* were within the standard trait distribution based on a matched set of individuals homozygous for the derived alleles (non-carriers; see Methods), and did not show a trend towards extreme values (Figure 1A). A similar pattern was observed for categorical traits, where carriers show no strong deviating pattern from the matched non-carrier cohort (Figure 1B-C). Given its putative roles in neurite outgrowth, prior associations of common variants with a broad array of brain imaging metrics[24], and the general involvement of protein phosphatases in psychiatric and neurological disorders[30], we tested the possible consequences of

carrying aSNVs in *SSH2* for a range of neuropsychological traits. We did not observe diverging patterns for aSNV carriers compared to the non-carrier group (Figure S2).

No consequences of the archaic allele in *TKTL1* for frontal pole morphology and overall cognition

We went on to study the archaic allele (A) of the rs111811311 polymorphism (A/G) of the *TKTL1* gene, located on the X chromosome. This missense change (yielding a lysine-to-arginine change at residue 317 of the long isoform) gained considerable prominence in recent literature when it was proposed by Pinson et al. as a major driver of human/Neandertal brain differences in evolution based on functional experiments[10]. Note that the variant was not among our curated list of aSNVs above, since it did not fit the criteria of full fixation in Kuhlwilm & Boeckx[2] (AF = 1), while a critique of Pinson et al. has highlighted the existence of rs111811311 archaic allele carriers in gnomAD[12], but without any phenotypic follow-up. Querying the UKB resource for the *TKTL1* archaic allele, we identified 45 heterozygous and one homozygous female carrier, as well as 16 hemizygous male carriers across multiple ancestry groups (Table 2, Supplementary Figure 1C). Among these 62 carriers, we identified four pairs with a kinship coefficient below 0.042, indicating again that relatedness (up to the 3rd degree) is unlikely to explain the larger number of carriers. One individual was identified with an archaic allele in both *TKTL1* and *KIF26B*.

TKTL1		Population	Homozygotes (Alt)	Hemizygotes (Alt)	Heterozygotes	Hemizygotes (Ref)	Homozygotes (Ref)
Chromosome	X	SAS				4612	3960
Position (hg38)	154315258	EAS				600	1283
Reference Allele	G	EUR		4	12	193302	229987
Archaic Allele	*A*	AFR		1	10	2251	2822
rsID	rs111811311	Uncategorized	1	11	23	5774	7513
		Total N	1	16	45	206539	241605

Table 2: Overview of identified aSNV carriers for *TKTL1* in UK Biobank. Genotype count of carriers of the archaic allele and individuals homozygous or hemizygous for the derived allele are noted per ancestry supercluster. Genomic positions based on hg38; REF = reference allele; ALT = Alternative/Ancestral allele;

Given that the cellular/animal work of Pinson et al.[10] linked the human-derived allele of *TKTL1* to substantial increases in neuron production in the prefrontal cortex, we contrasted imaging derived structural brain metrics of the frontal lobe in unrelated aSNV carriers (N = 5) and matched non-carriers, homozygous for the derived allele (N = 2145) to investigate the effects of carrying an archaic allele on frontal lobe surface area and cortical thickness in living human adults (Figure 2).

We found that the range of phenotypic variation of aSNV carriers lies in general within the 25th and 75th percentiles of the non-carriers for all cortical measures. This is in stark contrast with the pronounced effects shown in the various functional assays performed by Pinson et al.[10] which would predict substantial reductions in prefrontal cortex brain metrics of carriers of the archaic allele. As a sensitivity analysis, we repeated this approach in an ancestry matched cohort of only European carriers (N = 3) and a matched non-carrier cohort (n = 30), and obtained an even clearer overlap in phenotypic distributions (Figure S3).

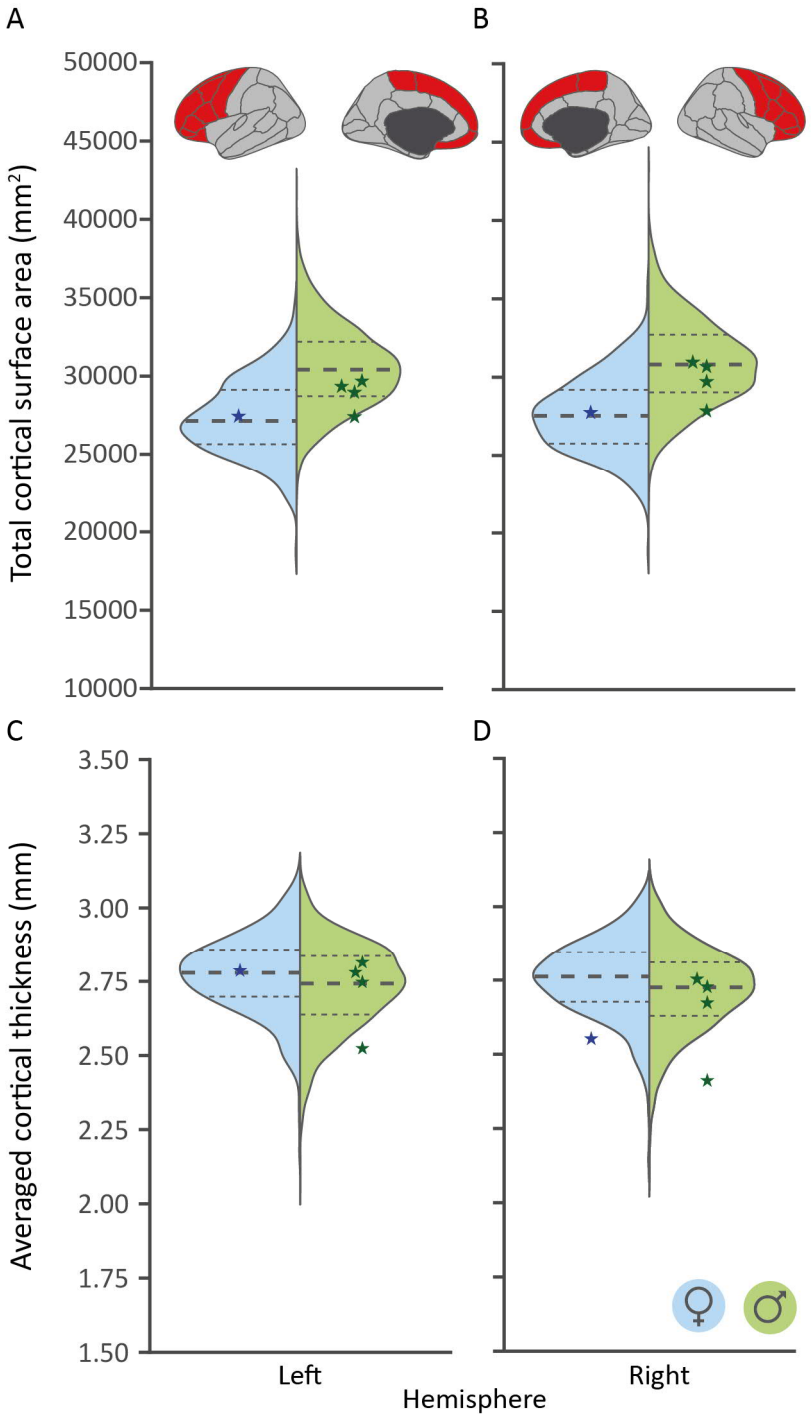

Figure 2: Carriers of the archaic allele of the *TKTL1* aSNV show no diverging cortical measures compared to matched set of non-carriers. Archaic allele carrier values for each sex and metric are depicted as diamonds. These are overlaid over split violin plots indicating the phenotypic variability for both the female (left, blue) and male (right, green) matched non-carrier sample for total cortical surface area for both left and right hemisphere (A and B, respectively) and averaged cortical thickness (C and D, respectively); G: Reference/derived allele, A: Archaic allele. Tick dotted line indicates the median, while the thin dotted lines highlight both the 25th and 75th percentiles.

Increased neuronal proliferation and expansion of the neocortex along the lineage leading to modern humans is argued by some to have been a driver of increased cognitive capacities of our species[31,32]. Indeed, based on the Pinson et al[10] findings, it has been proposed that the *TKTL1* protein-coding aSNV contributed to differences in cognition between *Homo sapiens* and extinct archaic humans[33]. Thus, we

also assessed educational qualification levels of carriers of the archaic *TKTL1* allele (N = 30) compared to matched non-carriers (N = 600) (Figure 2). Due to the difference in zygosity, this was done separately for males and females.

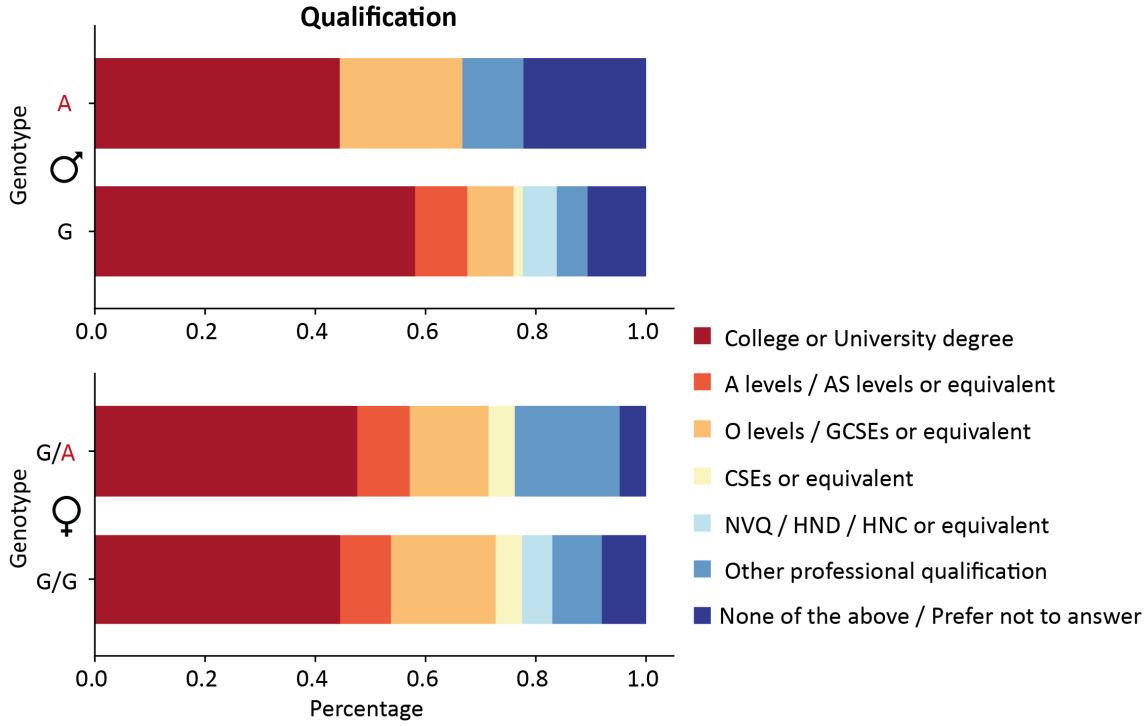

Figure 2: Qualification levels of carriers of archaic alleles of the *TKTL1* aSNV are similar to matched set of non-carriers. Stacked bar plots showing the percentage of highest qualification level by genotype. (Male sample: N_{aSNV} = 9; $N_{Non\text{-}carriersl}$ = 180; Female sample: N_{aSNV} = 21; $N_{Non\text{-}carriers}$ = 420); A level = Advanced level, AS level = Advanced Subsidiary level, O level = Ordinary level, GCSE = General Certificate of Secondary Education, CSE = Certificate of Secondary Education, NVQ = National Vocational Qualification, HND = Higher National Diploma, HNC = Higher National Certificate;

While the percentage of males with the highest qualification level was slightly lower for those with the archaic allele of *TKTL1*, it is striking that in both sexes a substantial proportion of carriers of this allele have a college or University degree. In particular, > 44% of males with only an archaic allele on this polymorphic site of the X chromosome have a college/University degree, contradicting the idea that the human-derived change in *TKTL1* was a key player in the evolution of enhanced human cognitive abilities[33].

DISCUSSION

Taken together, this study identified 165 unique carriers of archaic SNVs for 18 out of a total of 38 interrogated genomic positions in around 450,000 individuals with exome data in UKB. Regarding phenotypic consequences of an examplar aSNV in *SSH2*, one for which relatively large numbers of carriers were available, all interrogated traits fell within the typical range of variation, with no obvious divergence from the norm. A similar pattern was observable for *TKTL1* for frontal lobe structural measures as well as overall qualification level, despite this variant being reported to have large effects on neocortical development in cellular/animal models.

Ever since the first high-coverage genome sequence of a Neandertal resulted in a catalogue of fixed missense aSNVs[7], the overall number has continually decreased, as more high quality Neandertal

genomes and ever-increasing population databases of present-day humans have become available. For such protein-coding changes, the present study reduces the number of potential fully fixed genomic positions described in Kuhlwilm and Boeckx[2] that we investigated here from 37 to only 20, while the true number is likely even smaller. This raises questions over whether (some of) the aSNV carriers are explained by rare back-mutations, whether these sites were never fixed to begin with, or whether the ancestral allele was reintroduced post-fixation during admixture events[3]. While for some genomic positions only a handful of carriers were found in the UKB, some positions present with higher carrier counts that make back-mutations an unlikely explanation[1]. Further, higher ancestral allele counts were often evident in non-European ancestry groups. High genetic diversity within African populations[34] might partially explain this pattern, but considering the skewedness of UKB towards White European ancestry[16] it remains intriguing. While it is known that some isolated populations have higher levels of archaic ancestry, either because they persisted since a common ancestor, as seen in the Khoe-San[35] or due to relatively recent admixture with Neandertals/Denisovans (e.g., Oceanian populations)[36,37], there is no detailed catalogue of fixed human-specific changes across a range of ancestries that could be used as a reference point, given that most results of genomic studies are solely based on populations with European ancestry[3].

The presence of aSNV carriers in population databases, however, does not rule out the possibility that these DNA changes contributed to the formation of anatomically modern humans. While experimental validation in model systems is crucial to understand the impact of variation at these genomic positions, current approaches are laborious, with a range of known pitfalls[12,38]. The availability of phenotypic data in UKB makes it possible for the first time to query possible phenotypic consequences in present-day living humans that carry the variants of interest. As contrasting a range of traits of interest indicated no systematic differences between matched individuals homozygous for the derived allele and aSNV carriers, this could be seen as evidence that there are no major phenotypic consequences of carrying an ancestral rather than a derived version at the queried position. However, while the aSNV on *SSH2* was carefully chosen because of its potential effects on the enzymatic properties of the encoded protein, increasing the likelihood of observable phenotypic consequences[3], and based on the relatively large number of identified carriers, the possibility remains that we did not have a sufficient sample size to detect trait differences[3]. Of note, all individuals were heterozygous carriers and still had one copy of the derived allele, therefore not reflecting the homozygous state observed in the Neandertal genome. Moreover, lacking power for a large phenome-wide screen, we chose phenotypes to target a priori based on broader literature, and might have thus missed a trait that is truly impacted by the variant. This raises a larger question of importance for the field: which phenotype(s) would best represent 'the human condition' in investigations of this kind? Latest archaeological evidence increasingly suggests cognitive and behavioural similarities with our extinct archaic cousins, meaning that differences, especially for complex traits, may well be subtle[2,4–6]. A large phenome-wide scan sensitive enough to detect small deviations from the norm might highlight the most important phenotypes, as well as the actual contribution of these genomic positions, but this will only be feasible when even larger sample sizes are available than at present.

While only reported as a human-specific high-frequency variant by Kuhlwilm and Boeckx[2], several reasons led us to include the aSNV in *TKTL1* in the current investigation. Firstly, the phenotypic consequences of ancestral versus derived alleles of this aSNV are well described in Pinson et al., based on their experiments in animal/cellular models[10], allowing for a more targeted phenotypic selection.

Secondly, the effect sizes reported in Pinson et al.[10] were substantial, indicating that even with only a small number of identified carriers there should be good prospects of detecting such phenotypic consequences. Thirdly, the position of the aSNV on the X chromosome should lead to even more pronounced effects in males, who are hemizygous for either a derived or ancestral allele. Still, we saw no differences for carriers and matched sample of non-carriers in either neuroanatomical properties of the frontal lobe or in qualification level, even in male carriers. This result likely suggests we cannot extrapolate effect sizes identified in functional assays and model organisms to the true consequences of these changes for human kind[15].

Beyond general challenges related to rare variant analysis and the choice of target phenotypes, as discussed above, limitations of the current study include those related to the nature of the UKB cohort (restricted age range, lack of diversity in ancestral background, existence of participation bias etc.[16,39]), and the need for more and better-quality archaic hominin genomes to understand the genetic variation patterns in their populations.

Against this background, our work points towards promising avenues for future investigations: i) The inclusion of more large-scale diverse population databases[40–43] together with the information from the third high quality Neandertal genome[9] (and additional archaic genomes that might be sequenced) would likely yield a more representative catalogue of human-specific changes to help reconstruct how natural selection, archaic gene flow, and our demographic history together shaped our genome [1,3,34,44]. ii) Given the ever-decreasing number of such sites, we recommend abandoning the notion of fully-fixed variants and expanding the scope to also take high-frequency non-fixed changes into account. Kuhlwilm and Boeckx[2] already made a start in this direction and expanded their catalogue accordingly, but with the availability of larger and more diverse databases, this list will need updating. As more population databases are also including whole genome sequencing, the search can be expanded further to include high-frequency changes in regulatory regions[45]. iii) Our results indicate that looking at each of these genomic positions individually might not be so informative and that future work focusing on their aggregated effects could be valuable[2,3]. One way to achieve this would be by grouping high-frequency changes according to their potential functions [see Kuhlwilm and Boeckx[2] for an initial categorization]. Further, a list of high-frequency variants could also be used for burden testing, which would additionally allow formal statistical analyses of possible effects[46,47].

Overall, by leveraging the availability of archaic variation in modern biobanks, our study has challenged the notion of fixed genomic changes on the human lineage, highlighted that individual interrogation of the key sites is unlikely to yield major insights into emergence of complex human traits, and emphasises again the importance of including diverse ancestral backgrounds in studies on the origins of our species.

MATERIALS AND METHODS

Dataset

All data used were obtained from the UK Biobank (UKB) under the research application 16066 with Clyde Francks as the principal investigator. Detailed descriptions of the data used as well as sample, genotype and variant specific quality control (QC) are given below. The UK Biobank has received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382) and all of their procedures were performed in accordance with the World Medical Association. Informed consent was obtained for all participants by UKB with details about data collection and ethical procedures described elsewhere[16,48].

Whole exome sequencing data

Whole-exome sequencing was performed and data were processed by the UKB according to protocols described elsewhere[18–20]. Briefly, the samples were multiplexed and then sequenced using 75-base-pair paired-end reads with two 10-base-pair index reads on the Illumina NovaSeq 6000 platform using either S2 (first exome release) or S4 flow cells. Sample-specific FASTQ files, representing all the reads generated for that sample, were mapped to the GRCh38 genome reference using the BWA-meme algorithm 60. Subsequently, the binary alignment files (BAM) for each sample contained the mapped reads' genomic coordinates, quality information, and the degree to which a particular read differed from the reference at its mapped location. Duplicated reads were removed with the Picard[49] MarkDuplicates tool. GVCF files were then produced using the weCall variant caller. Upon completion of variant calling, individual sample BAM files were converted to fully lossless CRAM files using samtools[50].

For this project, we made use of the Broad 455k exome gnomAD VCF files (UKB data field 24068): population VCF files that have been returned to UK Biobank as part of the 'alternative exome processing' (UKB Category 172). Here, original UKB CRAM files were re-processed according to the GATK Best Practices, aligning reads using BWA-MEM 0.7.15.r1140 and processing reads using Picard and GATK with protocols described in detail in Karczewski et al.[51]

Variant selection

We based our selection of fixed, amino acid changing SNVs on a list of high-frequency human-specific missense changes described in Kuhlwilm & Boeckx[2], where only those SNVs with an allele frequency of one were selected, indicating total fixation at the time of their publication. These 42 genomic positions were lifted to GRCh38 using Liftover (https://liftover.broadinstitute.org/). We further confirmed that each SNV was indeed located in a translated exon using gnomAD v3.1.2. (https://gnomad.broadinstitute.org/) and the UCSC Genome Browser (https://genome.ucsc.edu/), which lead to exclusion of 3 variants located in *C1orf159* (chr1:1091245), *DNHD1* (chr11: 6534188) and *DNMT3L (*21: 44251169). One genomic position in *TBC1D3* (chr17:38202786) was excluded due to ambiguous results after liftover to GRCh38.

Due to its reported profound effects on brain development in cellular/animal models[10], a SNV in *TKTL1* on chromosome X (X: 154315258), previously reported as fixed by Prüfer et al.[7], was also included in our analysis despite being reported as a high-frequency variant (AF: 0.999694) in Kuhlwilm & Boeckx[2]. This resulted in a total of 39 SNVs in 33 genes that were put forward for further analysis (see Supplementary table 1).

Sample specific quality control

On all available individuals included in the gnomAD VCFs (N = 454,672), we first applied sample level quality control measures. This entailed excluding individuals with a mismatch of their self-reported (UKB data field 31) and genetically inferred sex (UKB data field 22001), as well as individuals with putative aneuploidies (UKB data field 22019), or individuals who were determined as outliers based

on heterozygosity (PC corrected heterozygosity >0.1903) or genotype missingness rate (missing rate >0.05) (UKB data field 22027), leading to a final sample of 452,797 individuals.

Variant and genotype QC

For further analysis we moved to the UK Biobank Research Analysis Platform (UKB RAP, https://ukbiobank.dnanexus.com) and queried the curated list of genomic position detailed above using bcftools (version 1.17)[52] to identify possible carriers of the archaic allele.
To assure that identified carriers did not represent sequencing errors, we only included archaic SNVs who were called with PASS in the VCF. This variant filter is based on a combination of a random forest classifier and hard filters, detailed in Karczewski et al.[51]. Only one SNV (chr9:6606647) did not pass these filters and was discarded for further analysis. We further used Hail (https://github.com/hail-is/hail) as implemented in JupyterLab on the UKB RAP for quality control of individual genotype data, where genotypes at the specific positions were filtered based on genotype quality (QUAL> 20), depth (DP> 10), and allele balance for heterozygous genotypes (AB > 25% < 75%), leading to different sample counts per queried position.

Variant distribution per ancestry cluster

We inferred ancestry for all individuals surviving the quality control outlined above. We first used the self-reported ethnicity[16] (UKB data field 21000-0.0) as provided by UKB and grouped each individual into 4 major ancestry clusters [European (EUR), African (AFR), South Asian (SAS), East Asian (EAS)]. Individuals who reported 'Mixed', 'Other', 'Do not know' and 'Do not want to answer' were grouped as 'Uncategorized'. We then used the first four provided genetic ancestry principal components (UKB data field Data-Field 22009-0.1 - 22009-0.4) and assigned each individual to one of the respective major ancestry clusters using hard cut-offs. Individuals labelled as 'Uncategorized', which only showed a highly dispersed cluster, were reassigned to one of four superclusters if their PC's fell within the respective boundaries. Otherwise, the individuals kept their initial category. This resulted in 5092 individuals within the AFR ancestry cluster, 1887 individuals within the EAS cluster, 423,887 individuals within the EUR ancestry cluster and 8585 individuals within the SAS cluster, while 13,346 individuals could not be assigned to one of the superclusters and were thus labelled as 'uncategorized'.

Relatedness

To infer if relatedness could explain an accumulation of archaic SNVs at certain positions, we identified if carriers had a kinship coefficient > 0.0442 (UKB data field 22021), but initially did not exclude any individuals based on relatedness. For phenotypic analysis, this information was used, and one individual from each pair of relatives was excluded, where we prioritised the exclusion of non-carriers, as well as individuals related to a larger number of other individuals.

Phenotypic analyses

SSH2

For phenotypic analysis, we chose one exemplary genomic position for a qualitative comparison of carriers to non-carriers. Our initial query of the exome data highlighted a group of 21 individuals with an aSNV on chr17:29632016 within *SSH2*, where all 21 were grouped within the EUR ancestry cluster, and 20 were also assigned to the more stringent 'White British' ancestry (UKB data field 22006)[16]. We further excluded one carrier due to relatedness, leading to a final carrier count of 19 unrelated, White British individuals (14 female; mean age ± SD: 59.42 ± 7.94 years).

We derived an age- (UKB data field 21003-0.0), sex- (UKB data field 31-0.0) and ancestry-matched (white British, UKB data field 22006) unrelated sample of individuals homozygous for the derived allele, where each carrier was paired with 2079 unique non-carriers. This led to a matched non-carrier cohort of 39,501 individuals (29,106 females, mean age ± SD: 60.69 ± 7.73 years).

Trait selection

Based on literature detailing the phenotypic legacy of previous admixture events, and previous genetic associations with *SSH2* variants as detailed in the GWAS catalogue, we selected a range of traits to investigate potential phenotypic effects of carrying an aSNV: BMI (UKB data field 21001-0.0), Whole-body-fat-mass (UKB data field 23100-0.0), Height (UKB data filed 50-0.0), Overall Health rating (2178-0.0), and Qualification (UKB data field 6138-0.0). For each individual we included only the highest qualification level reported in the analysis. As the GWAS catalogue highlighted several associations of *SSH2* with different brain phenotypes and given prior proposed links of archaic admixture to psychiatric disorders, we also included a broad range of neuropsychiatric metrics: "Seen doctor (GP) for nerves, anxiety, tension or depression" (UKB data field 2190-0.0), "Seen Psychiatrist for nerves, anxiety, tension or depression" (UKB data field 2100-0.0), Moodiness (UKB data field 1920-0.0), Miserableness (UKB data field 1930-0.0), loneliness (UKB data field 2020-0.0) and risk-taking (UKB data field 2040-0.0).

TKTL1

Given the profound effects of the derived allele of *TKTL1* on brain development in experiments described by Pinson at al.[10], we also included targeted investigations of genotype/phenotype relationships for this aSNV. As that prior study had highlighted increase neuronal counts specifically in the frontal lobe[10] we focused our phenotypic analysis first on relevant neuroanatomical data, making use of imaging-derived phenotypes generated by an imaging-processing pipeline developed and run on behalf of the UK Biobank[17,53]. Imaging-derived structural measures (UK Biobank category 192) were available for 5 unrelated carriers (1 female; mean age ± SD: 71.2 ± SD 6.14 years). To estimate total frontal lobe surface area, we summed for each brain hemisphere the following imaging-derived phenotypes Superior Frontal (UKB data field 26748-2.0 & 26849-2.0), Rostral Middle Frontal (UKB data field 26747-2.0 & 26848-2.0), Caudal Middle Frontal (UKB data field 26724-2.0 & 26825-2.0), Pars Opercularis (UKB data field 26738-2.0 & 26839-2.0), Pars Triangularis (UKB data field 26740-2.0 & 26841-2.0), Pars Orbitalis (UKB data field 26739-2.0 & 26840), Lateral Orbifrontal (UKB data field 26732-2.0 26833-2.0) , Medial Orbifrontal (UKB data field 26734-2.0 & 26835-2.0), Precentral (UKB data field 26744-2.0 & 26845-2.0), Paracentral (UKB data field 26737-2.0 & 26838-2.0) and frontal pole (UKB data field 26752-2.0 & 26853-2.0). For the same cortical parcellations we used the imaging-derived cortical thickness (UKB data field 26782-2.0 & 26883-2.0, UKB data field 26781-2.0 & 26882-2.0, UKB data field 26758-2.0 & 26859-2.0, UKB data field 26772-2.0 & 26873-2.0, UKB data field 76774-2.0 & 26875-2.0, UKB data field 26773-2.0 & 26874-2.0, UKB data field 26766-2.0 & 26867-2.0, UKB data field 26768-2.0 & 26869-2.0, UKB data field 26778-2.0 & 26879-2.0, UKB data field 26771-2.0 & 26872-2.0, UKB data field 26786-2.0 & 26887-2.0) to estimate the averaged cortical thickness for each hemisphere in the frontal pole. As Pinson et al.[10] and others[33] have suggested that the change from archaic to human *TKTL1* could have played an important role for the evolution of complex behaviour, we also looked at qualification level (UK Biobank field 6138-0.0). For each individual we included only the highest qualification level reported in the analysis.

As *TKTL1* carriers were found in more than one ancestry cluster, the matched samples of non-carriers were set up as following.

- For brain imaging phenotypes, we first used all 5 carriers with imaging data and identified an equal number of individuals homozygous for the derived allele (N = 429) per carrier which were only matched by age (UK Biobank field 21003-2.0) and sex (UKB data field 31-0.0). This resulted in a non-carrier sample across ancestries of 2145 individuals (429 female; mean age ± SD: 71.2 ± SD 5.49 years).
- For a sensitivity analysis of the above, we only used the 3 European carriers (1 female, mean age ± SD: 73.67 ± SD 3.21 years), where 10 non-carrier individuals were matched to each carrier by the first two genetic principal components (PC1 & PC2 ± 2.5, respectively; UKB data field 22009-0.1 and 22009-0.2), age (± 2.5 years; UK Biobank field 21003-2.0)

and sex (UKB data field 31-0.0). This resulted in 30 unique non-carriers (10 females; mean age ± SD: 73.2± SD 3.21 years).

- For our analysis of qualification level, we only selected unrelated carriers, where we could identify 20 unique, matched (PC1 & PC2 ± 2.5, respectively; age ± 2.5 years, sex) non-carriers, which led to a final sample comprised of 30 carriers (21 female; 11 AFR, 10 EUR, 9 UNC; mean age ± SD: 54.4 ± SD 7.67 years) and 600 matched non-carriers (420 females; mean age ± SD: 54.49 ± SD 7.73 years);

For all included qualitative traits used in this study we combined possible answer options 'Do not know', 'Do not want to answer' and/or 'None of the above' to one item.

Data availability:

Whole exome sequencing data, imaging-derived phenotypes and other phenotypic data used in this study are available from UK Biobank (https://www.ukbiobank.ac.uk).

Code availability:

All scripts used for the analyses are available on the project GitLab repository (https://gitlab.gwdg.de/barmol/fixedVariant).

References

1. Pääbo, S. The Human Condition—A Molecular Approach. *Cell* 157, 216–226 (2014).
2. Kuhlwilm, M. & Boeckx, C. A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci. Rep.* 9, 8463 (2019).
3. Zeberg, H., Jakobsson, M. & Pääbo, S. The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell* 187, 1047–1058 (2024).
4. Scerri, E. M. L. & Will, M. The revolution that still isn't: The origins of behavioral complexity in Homo sapiens. *J. Hum. Evol.* 179, 103358 (2023).
5. Wynn, T. & Coolidge, F. L. The expert Neandertal mind. *J. Hum. Evol.* 46, 467–487 (2004).
6. Nowell, A. Rethinking Neandertals. *Annu. Rev. Anthropol.* 52, annurev-anthro-052621-024752 (2023).
7. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49 (2014).
8. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* 358, 655–658 (2017).
9. Mafessoni, F. *et al.* A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci.* 117, 15132–15136 (2020).
10. Pinson, A. *et al.* Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* 377, eabl6422 (2022).
11. Chiaradia, I. & Lancaster, M. A. Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo. *Nat. Neurosci.* 23, 1496–1508 (2020).
12. Herai, R. H., Semendeferi, K. & Muotri, A. R. Comment on "Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals". *Science* 379, eadf0602 (2023).
13. Bolognesi, B. & Lehner, B. Reaching the limit. *eLife* 7, e39804 (2018).

14. Enard, W. *et al.* A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell* 137, 961–971 (2009).
15. Pinson, A., Maricic, T., Zeberg, H., Pääbo, S. & Huttner, W. B. Response to Comment on "Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals". *Science* 379, eadf2212 (2023).
16. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209 (2018).
17. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* 166, 400–424 (2018).
18. Szustakowski, J. D. *et al.* Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat. Genet.* 53, 942–948 (2021).
19. Van Hout, C. V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 586, 749–756 (2020).
20. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* 599, 628–634 (2021).
21. Cuberos, H. *et al.* Roles of LIM kinases in central nervous system function and dysfunction. *FEBS Lett.* 589, 3795–3806 (2015).
22. Niwa, R., Nagata-Ohashi, K., Takeichi, M., Mizuno, K. & Uemura, T. Control of Actin Reorganization by Slingshot, a Family of Phosphatases that Dephosphorylate ADF/Cofilin. *Cell* 108, 233–246 (2002).
23. Ohta, Y. *et al.* Differential activities, subcellular distribution and tissue expression patterns of three members of Slingshot family phosphatases that dephosphorylate cofilin. *Genes Cells* 8, 811–824 (2003).
24. Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 51, D977–D985 (2023).
25. De Sousa, A. A. *et al.* From fossils to mind. *Commun. Biol.* 6, 636 (2023).

26. McArthur, E., Rinker, D. C. & Capra, J. A. Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat. Commun.* 12, 4481 (2021).
27. Dannemann, M. *et al.* Neandertal introgression partitions the genetic landscape of neuropsychiatric disorders and associated behavioral phenotypes. *Transl. Psychiatry* 12, 433 (2022).
28. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507, 354–357 (2014).
29. Simonti, C. N. *et al.* The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 351, 737–741 (2016).
30. An, N. *et al.* Dual-specificity phosphatases in mental and neurological disorders. *Prog. Neurobiol.* 198, 101906 (2021).
31. Rakic, P. Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* 10, 724–735 (2009).
32. Kovach, C. K. *et al.* Anterior Prefrontal Cortex Contributes to Action Selection through Tracking of Recent Reward Trends. *J. Neurosci.* 32, 8434–8442 (2012).
33. Malgrange, B. & Nguyen, L. Scaling brain neurogenesis across evolution. *Science* 377, 1155–1156 (2022).
34. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, eaay5012 (2020).
35. Schlebusch, C. M. *et al.* Khoe-San Genomes Reveal Unique Variation and Confirm the Deepest Population Divergence in Homo sapiens. *Mol. Biol. Evol.* 37, 2944–2954 (2020).
36. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338, 222–226 (2012).
37. Peyrégne, S., Slon, V. & Kelso, J. More than a decade of genetic research on the Denisovans. *Nat. Rev. Genet.* 25, 83–103 (2024).

38. Maricic, T. *et al.* Comment on “Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment”. *Science* 374, eabi6060 (2021).
39. Schoeler, T. *et al.* Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat. Hum. Behav.* 7, 1216–1227 (2023).
40. The All of Us Research Program Genomics Investigators *et al.* Genomic data in the All of Us Research Program. *Nature* 627, 340–346 (2024).
41. Kurki, M. I. *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. Preprint at https://doi.org/10.1101/2022.03.03.22271360 (2022).
42. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* 27, S2–S8 (2017).
43. Walters, R. G. *et al.* Genotyping and population structure of the China Kadoorie Biobank. Preprint at https://doi.org/10.1101/2022.05.02.22274487 (2022).
44. Gokcumen, O. Archaic hominin introgression into modern human genomes. *Am. J. Phys. Anthropol.* 171, 60–73 (2020).
45. Moriano, J. & Boeckx, C. Modern human changes in regulatory regions implicated in cortical development. *BMC Genomics* 21, 304 (2020).
46. Povysil, G. *et al.* Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat. Rev. Genet.* 20, 747–759 (2019).
47. Cirulli, E. T. *et al.* Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat. Commun.* 11, 542 (2020).
48. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* 12, e1001779 (2015).
49. Broad Institute. Picard toolkit. Broad Institute (2019).
50. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).

51. Karczewski, K. J. *et al.* Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* 2, 100168 (2022).

52. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008 (2021).

53. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536 (2016).

Acknowledgments:
This research was conducted using the UK Biobank resource under application no. 16066 with CF as the principal applicant. Our study made use of brain imaging derived phenotypes and pre-processed imaging data generated by an image processing pipeline developed and run on behalf of UK Biobank. SEF is a member of the Center for Academic Research and Training in Anthropogeny. BM, GA, EE, DS, CF and SEF were supported by the Max Planck Society. EE is also supported by the Dutch Research Council (NWO; VI.Veni.202.072).

Author contributions:

Conceptualization: BM, MLA, EE, GA, SEF
Resources: CF, SEF
Methodology: BM, EE, GA, DS, SEF
Data analysis: BM, MLA
Writing – original draft: BM
Writing – review & editing: BM, MLA, EE, GA, DS, CF, SEF

Competing interests: Authors declare that they have no competing interests.

Supplementary Information is available for this paper

Correspondence and requests for materials should be addressed to Simon Fisher.