Title: Evaluating the effects of archaic protein-altering variants in living human adults

Authors:

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

Affiliations:

[1]Language & Genetics Department, Max Planck Institute for Psycholinguistics; Nijmegen, 6525XD, Netherlands.

[2]Department of Cognitive Neuroscience, Radboud University Medical Center; Nijmegen, 6525 EN, Netherlands.

[3]Donders Institute for Brain, Cognition and Behaviour; Nijmegen, 6525 EN, Netherlands.

† These authors contributed equally to this work

‡ Present address: Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology; Leipzig, 04103, Germany.

§ Present address: Donders Institute for Brain, Cognition and Behaviour; Nijmegen, 6525 EN, Netherlands.

*Corresponding author. Email: simon.fisher@mpi.nl

Abstract: Advances in paleo-genetics allowed the identification of protein-coding changes arising on the lineage leading to *Homo sapiens*, by comparing genomes of present-day and archaic hominins. Experimental validation of the potential impact of such changes has so far been restricted to functional assays and model organisms. Large-scale biobanking now makes it possible to identify present-day carriers of archaic alleles and to directly assess phenotypic consequences in living adults. We queried exomes of half a million people in the UK Biobank at 37 genomic positions with supposedly fixed human-specific protein-coding changes. This yielded 103 carriers at 17 positions, with variable allele counts across ancestries. Contrasting carriers and non-carriers of an exemplary archaic allele in *SSH2*, we observed no deviation from the norm in a range of health, psychological, and cognitive traits. We also identified 62 archaic-allele carriers for a *TKTL1* missense change, previously shown to have large effects on cortical neurogenesis in brain organoids and animal models. Carriers did not show differences in relevant anatomical brain measures, and a substantial proportion had college/university degrees. This work offers an empirical demonstration of how large-scale biobank investigations of living adults can transform our understanding of human evolution. The findings challenge the notion of fixed human-specific genomic changes, highlight that individual interrogation of relevant sites is unlikely to yield major insights into the emergence of complex human traits, and emphasise the importance of including diverse ancestries when investigating origins of our species.

Main Text:

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 multiple 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*).

A complementary approach for evaluating broader biological impact that has only recently become feasible, depends on the identification of present-day living 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. With availability of large-scale biobanks with exome sequencing and trait data it is now possible not only to detect aSNVs in living humans, but also to investigate putative phenotypic consequences in a way that could not be done before.

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. We use our findings to make recommendations about how to optimize future biobank-based investigations of human evolution.

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, table S1). 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 (fig. S1A). In total, we identified 103 unique individuals carrying 118 aSNVs in 13 protein-coding genes (Table 1). All were heterozygous carriers, except for a female carrying a homozygous aSNV in *GRM6* (chr5:178994530), 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 assessed in carriers of the *SSH2* aSNV 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. We chose an aSNV in *SSH2* (chr17:29632016) to exemplify this, 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 the 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*).

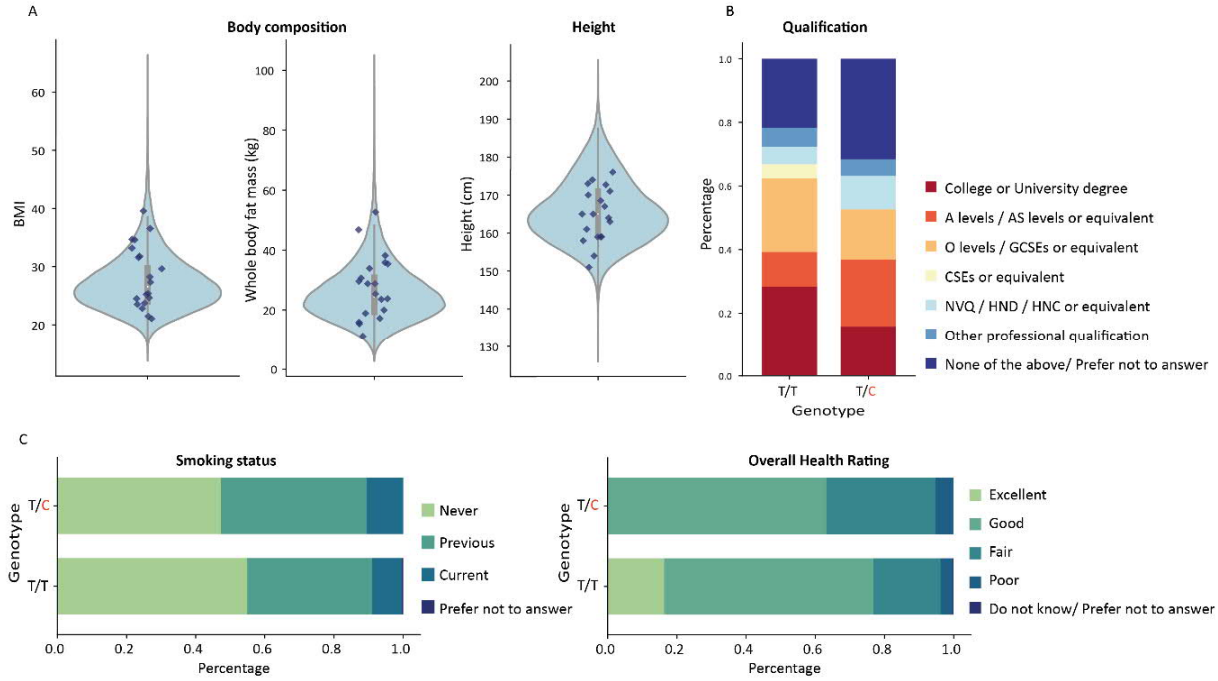

Figure 1: Investigating phenotypic effects in carriers of the *SSH2* aSNV. (A) Values of continuous traits are shown for each aSNV carrier as dark blue diamonds. Violin plots show the phenotypic distribution of matched set of non-carriers, with boxplots indicating the 25th and 75th percentiles, and whiskers 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 *SSH2* aSNV 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 (Fig. 1A). A similar pattern was observed for categorical traits, where carriers show no strong deviating pattern from the matched non-carrier cohort (Fig. 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 (fig. S2).

The archaic allele in *TKTL1* shows little consequence for frontal pole morphology and overall cognition in adult humans

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 an array of 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. (*10*) 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, fig S1C). 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 (Fig. 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 (fig. S3).

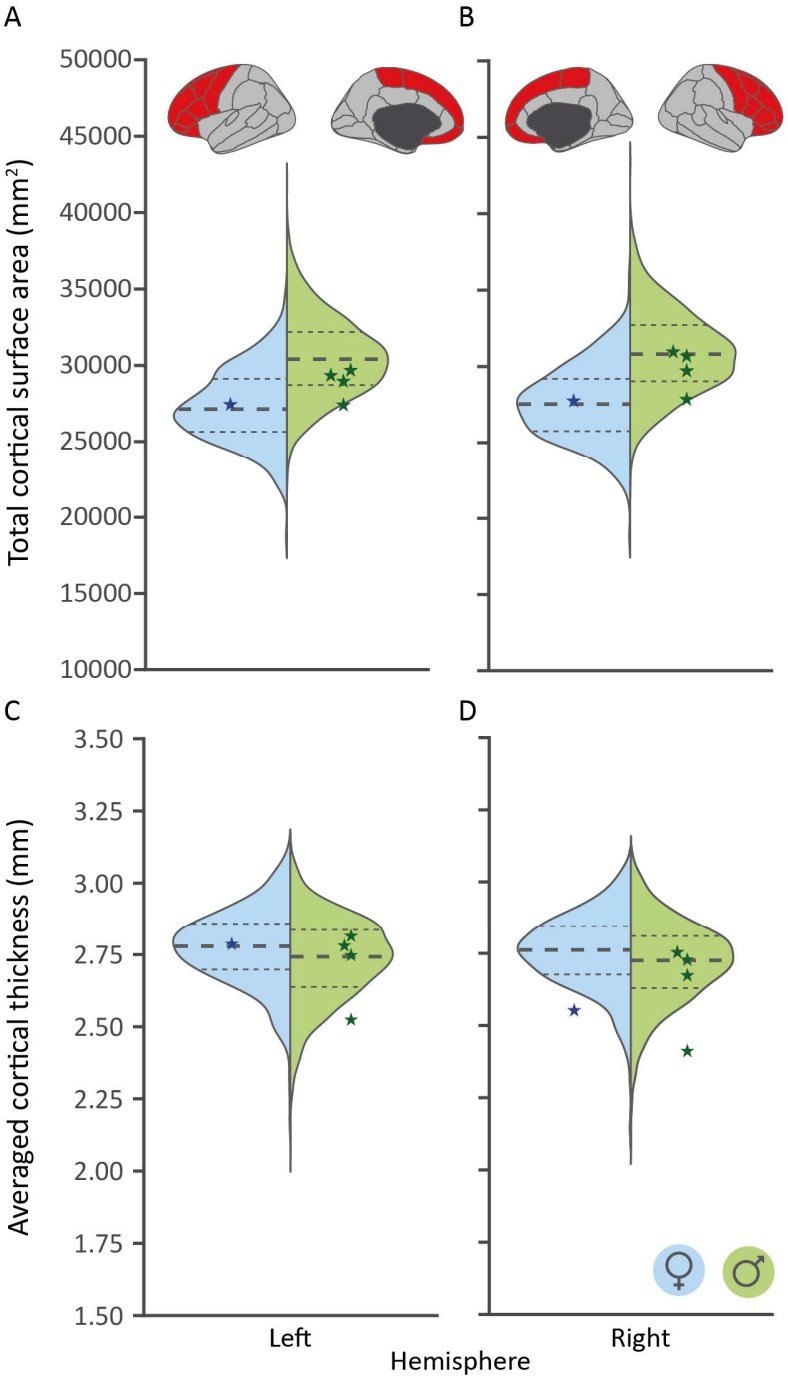

Figure 2: Carriers of the archaic allele of the *TKTL1* aSNV show no diverging cortical measures compared to a matched set of non-carriers. Archaic allele carrier values for each sex and metric are depicted as diamonds. These are overlaid over split violin plots indicating the phenotypic variability 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, some other researchers and

commentators have 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) (Fig. 2). Due to the difference in zygosity, this was done separately for males and females.

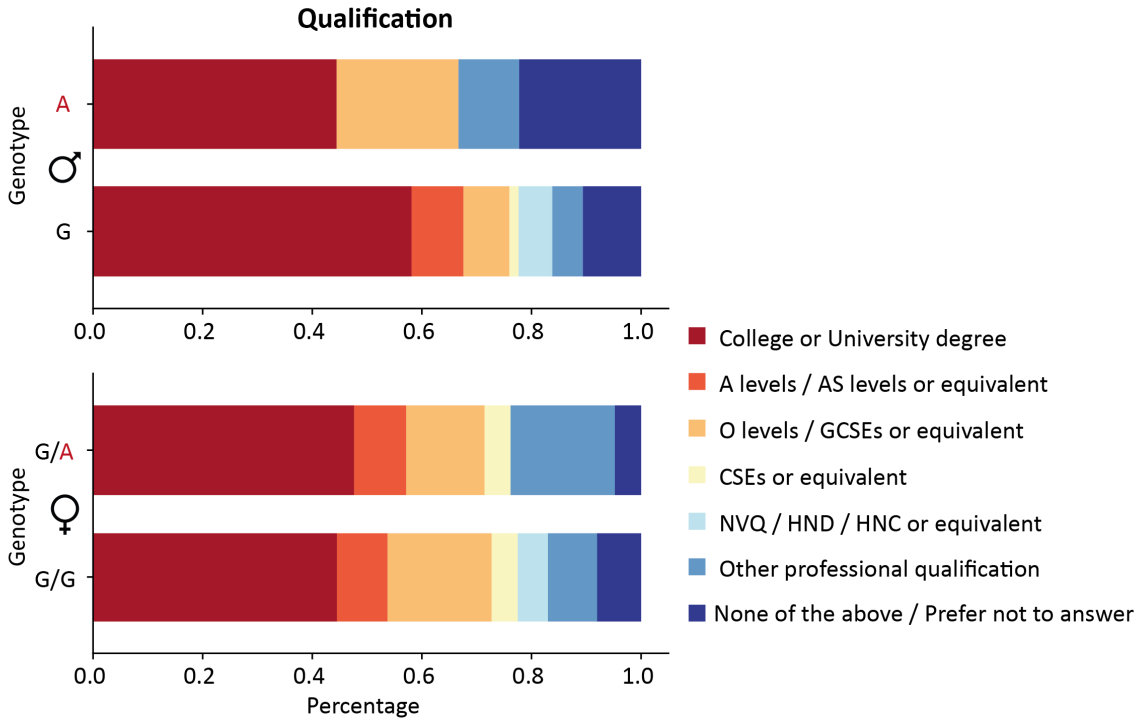

Figure 2: **Qualification levels of carriers of archaic alleles of the *TKTL1* aSNV are similar to matched set of non-carriers**. Stacked bar plots showing the percentage of highest qualification level by genotype. (Male sample: N_{aSNV} = 9; $N_{Non\text{-}carriersl}$ = 180; Female sample: N_{aSNV} = 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. This pattern of findings casts doubt on the idea that the human-derived change in *TKTL1* was a key player in the evolution of enhanced human cognitive abilities (*33*).

DISCUSSION

This study brings a novel source of empirical data to questions regarding evolutionary impacts of protein-coding variants that distinguish between modern humans and our extinct archaic cousins, adding to the rich prior literature in this area, recently reviewed by Zeberg and colleagues (*3*). Our work 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 exemplar 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 previously showing 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 skewness 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 adult 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, 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*). The lower carrier numbers of other aSNVs (at least within ancestry clusters) further limit the scope of currently feasible phenotypic investigations. A large phenome-wide scan sensitive enough to detect small deviations from the norm might highlight the most important phenotypes, as well as clarifying contributions

of these genomic positions, but this will only be feasible when even larger sample sizes are available than at present.

While reported as only 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. Complementing findings from these models, the researchers also reported that disrupting *TKTL1* expression in fetal human brain neocortical tissue significantly reduced basal radial glial progenitors (*10*). Secondly, the effect sizes of the aSNV allele 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 neuroanatomical properties of the frontal lobe even in male carriers, and a substantial proportion of these had a college/university degree, arguing against a major impact on cognitive functions. While the absence of consequences for adults might possibly be explained by compensatory mechanisms with mitigating effects on the developing frontal lobes, our results show that effect sizes identified in functional assays and model organisms cannot be directly extrapolated to the consequences of carrying these changes for adult human phenotypes (*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 (*16*, *39*)), and the need for more and better-quality archaic hominin genomes to understand the genetic variation patterns in their populations.

The findings here resonate well with the recent perspective of the field set out by Zeberg and colleagues (*3*). With our concrete demonstration of biobank analyses, we provide new impetus 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) will 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, it seems warranted to abandon the notion of fully-fixed variants, broadening 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 provided evidence against 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 the emergence of complex human traits, and emphasises again the importance of including diverse ancestral backgrounds in studies on the origins of our species.

References

1. S. Pääbo, The Human Condition—A Molecular Approach. *Cell* **157**, 216–226 (2014).
2. M. Kuhlwilm, C. Boeckx, A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci Rep* **9**, 8463 (2019).
3. H. Zeberg, M. Jakobsson, S. Pääbo, The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell* **187**, 1047–1058 (2024).
4. E. M. L. Scerri, M. Will, The revolution that still isn't: The origins of behavioral complexity in Homo sapiens. *Journal of Human Evolution* **179**, 103358 (2023).
5. T. Wynn, F. L. Coolidge, The expert Neandertal mind. *Journal of Human Evolution* **46**, 467–487 (2004).
6. A. Nowell, Rethinking Neandertals. *Annu. Rev. Anthropol.* **52**, annurev-anthro-052621-024752 (2023).
7. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
8. K. Prüfer, C. De Filippo, S. Grote, F. Mafessoni, P. Korlević, M. Hajdinjak, B. Vernot, L. Skov, P. Hsieh, S. Peyrégne, D. Reher, C. Hopfe, S. Nagel, T. Maricic, Q. Fu, C. Theunert, R. Rogers, P. Skoglund, M. Chintalapati, M. Dannemann, B. J. Nelson, F. M. Key, P. Rudan, Ž. Kućan, I. Gušić, L. V. Golovanova, V. B. Doronichev, N. Patterson, D. Reich, E. E. Eichler, M. Slatkin, M. H. Schierup, A. M. Andrés, J. Kelso, M. Meyer, S. Pääbo, A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
9. F. Mafessoni, S. Grote, C. De Filippo, V. Slon, K. A. Kolobova, B. Viola, S. V. Markin, M. Chintalapati, S. Peyrégne, L. Skov, P. Skoglund, A. I. Krivoshapkin, A. P. Derevianko, M. Meyer, J. Kelso, B. Peter, K. Prüfer, S. Pääbo, A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15132–15136 (2020).
10. A. Pinson, L. Xing, T. Namba, N. Kalebic, J. Peters, C. E. Oegema, S. Traikov, K. Reppe, S. Riesenberg, T. Maricic, R. Derihaci, P. Wimberger, S. Pääbo, W. B. Huttner, Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* **377**, eabl6422 (2022).
11. I. Chiaradia, M. A. Lancaster, Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo. *Nat Neurosci* **23**, 1496–1508 (2020).
12. R. H. Herai, K. Semendeferi, A. R. Muotri, Comment on "Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals." *Science* **379**, eadf0602 (2023).
13. B. Bolognesi, B. Lehner, Reaching the limit. *eLife* **7**, e39804 (2018).
14. W. Enard, S. Gehre, K. Hammerschmidt, S. M. Hölter, T. Blass, M. Somel, M. K. Brückner, C. Schreiweis, C. Winter, R. Sohr, L. Becker, V. Wiebe, B. Nickel, T. Giger, U. Müller, M. Groszer, T. Adler, A. Aguilar, I. Bolle, J. Calzada-Wack, C. Dalke, N. Ehrhardt, J. Favor, H. Fuchs, V. Gailus-Durner, W. Hans, G. Hölzlwimmer, A. Javaheri, S. Kalaydjiev, M. Kallnik, E. Kling, S. Kunder, I. Moßbrugger, B. Naton, I. Racz, B. Rathkolb, J. Rozman, A. Schrewe, D. H. Busch, J. Graw, B. Ivandic, M. Klingenspor, T. Klopstock, M. Ollert, L. Quintanilla-Martinez, H. Schulz, E. Wolf, W. Wurst, A. Zimmer, S. E. Fisher, R. Morgenstern, T. Arendt, M. Hrabé De Angelis, J. Fischer, J. Schwarz, S. Pääbo, A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. *Cell* **137**, 961–971 (2009).
15. A. Pinson, T. Maricic, H. Zeberg, S. Pääbo, W. B. Huttner, Response to Comment on "Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals." *Science* **379**, eadf2212 (2023).
16. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
17. F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. R. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee, D. Vidaurre, M. Webster, P. McCarthy, C. Rorden, A. Daducci, D. C. Alexander, H. Zhang, I. Dragonu, P. M. Matthews, K. L. Miller, S. M. Smith, Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
18. J. D. Szustakowski, S. Balasubramanian, E. Kvikstad, S. Khalid, P. G. Bronson, A. Sasson, E. Wong, D. Liu, J. Wade Davis, C. Haefliger, A. Katrina Loomis, R. Mikkilineni, H. J. Noh, S. Wadhawan, X. Bai, A. Hawes,

O. Krasheninina, R. Ulloa, A. E. Lopez, E. N. Smith, J. F. Waring, C. D. Whelan, E. A. Tsai, J. D. Overton, W. J. Salerno, H. Jacob, S. Szalma, H. Runz, G. Hinkle, P. Nioi, S. Petrovski, M. R. Miller, A. Baras, L. J. Mitnaul, J. G. Reid, UKB-ESC Research Team, O. Moiseyenko, C. Rios, S. Saha, G. Abecasis, N. Banerjee, C. Beechert, B. Boutkov, M. Cantor, G. Coppola, A. Economides, G. Eom, C. Forsythe, E. D. Fuller, Z. Gu, L. Habegger, M. B. Jones, R. Lanche, M. Lattari, M. LeBlanc, D. Li, L. A. Lotta, K. Manoochehri, A. J. Mansfield, E. K. Maxwell, J. Mighty, M. Nafde, S. O'Keeffe, M. Orelus, M. S. Padilla, R. Panea, T. Polanco, M. Pradhan, A. Rasool, T. D. Schleicher, D. Sharma, A. Shuldiner, J. C. Staples, C. V. Van Hout, L. Widom, S. E. Wolf, S. John, C.-Y. Chen, D. Sexton, V. Kupelian, E. Marshall, T. Swan, S. Eaton, J. Z. Liu, S. Loomis, M. Jensen, S. Duraisamy, J. Tetrault, D. Merberg, S. Badola, M. Reppell, J. Grundstad, X. Zheng, A. M. Deaton, M. M. Parker, L. D. Ward, A. O. Flynn-Carroll, C. Austin, R. March, M. N. Pangalos, A. Platt, M. Snowden, A. Matakidou, S. Wasilewski, Q. Wang, S. Deevi, K. Carss, K. Smith, M. Sogaard, X. Hu, X. Chen, Z. Ye, Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank. *Nat Genet* **53**, 942–948 (2021).

19. C. V. Van Hout, I. Tachmazidou, J. D. Backman, J. D. Hoffman, D. Liu, A. K. Pandey, C. Gonzaga-Jauregui, S. Khalid, B. Ye, N. Banerjee, A. H. Li, C. O'Dushlaine, A. Marcketta, J. Staples, C. Schurmann, A. Hawes, E. Maxwell, L. Barnard, A. Lopez, J. Penn, L. Habegger, A. L. Blumenfeld, X. Bai, S. O'Keeffe, A. Yadav, K. Praveen, M. Jones, W. J. Salerno, W. K. Chung, I. Surakka, C. J. Willer, K. Hveem, J. B. Leader, D. J. Carey, D. H. Ledbetter, Geisinger-Regeneron DiscovEHR Collaboration, L. Cardon, G. D. Yancopoulos, A. Economides, G. Coppola, A. R. Shuldiner, S. Balasubramanian, M. Cantor, Regeneron Genetics Center, M. R. Nelson, J. Whittaker, J. G. Reid, J. Marchini, J. D. Overton, R. A. Scott, G. R. Abecasis, L. Yerges-Armstrong, A. Baras, Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

20. J. D. Backman, A. H. Li, A. Marcketta, D. Sun, J. Mbatchou, M. D. Kessler, C. Benner, D. Liu, A. E. Locke, S. Balasubramanian, A. Yadav, N. Banerjee, C. E. Gillies, A. Damask, S. Liu, X. Bai, A. Hawes, E. Maxwell, L. Gurski, K. Watanabe, J. A. Kosmicki, V. Rajagopal, J. Mighty, Regeneron Genetics Center, DiscovEHR, M. Jones, L. Mitnaul, E. Stahl, G. Coppola, E. Jorgenson, L. Habegger, W. J. Salerno, A. R. Shuldiner, L. A. Lotta, J. D. Overton, M. N. Cantor, J. G. Reid, G. Yancopoulos, H. M. Kang, J. Marchini, A. Baras, G. R. Abecasis, M. A. R. Ferreira, Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).

21. H. Cuberos, B. Vallée, P. Vourc'h, J. Tastet, C. R. Andres, H. Bénédetti, Roles of LIM kinases in central nervous system function and dysfunction. *FEBS Letters* **589**, 3795–3806 (2015).

22. R. Niwa, K. Nagata-Ohashi, M. Takeichi, K. Mizuno, T. Uemura, Control of Actin Reorganization by Slingshot, a Family of Phosphatases that Dephosphorylate ADF/Cofilin. *Cell* **108**, 233–246 (2002).

23. Y. Ohta, K. Kousaka, K. Nagata-Ohashi, K. Ohashi, A. Muramoto, Y. Shima, R. Niwa, T. Uemura, K. Mizuno, Differential activities, subcellular distribution and tissue expression patterns of three members of Slingshot family phosphatases that dephosphorylate cofilin. *Genes to Cells* **8**, 811–824 (2003).

24. E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Güneş, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. A. L. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorff, F. Cunningham, S. A. Lambert, M. Inouye, H. Parkinson, L. W. Harris, The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* **51**, D977–D985 (2023).

25. A. A. De Sousa, A. Beaudet, T. Calvey, A. Bardo, J. Benoit, C. J. Charvet, C. Dehay, A. Gómez-Robles, P. Gunz, K. Heuer, M. P. Van Den Heuvel, S. Hurst, P. Lauters, D. Reed, M. Salagnon, C. C. Sherwood, F. Ströckens, M. Tawane, O. S. Todorov, R. Toro, Y. Wei, From fossils to mind. *Commun Biol* **6**, 636 (2023).

26. E. McArthur, D. C. Rinker, J. A. Capra, Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat Commun* **12**, 4481 (2021).

27. M. Dannemann, Y. Milaneschi, D. Yermakovich, V. Stiglbauer, H. M. Kariis, K. Krebs, M. A. Friese, C. Otte, Estonian Biobank Research Team, T. Esko, A. Metspalu, L. Milani, R. Mägi, M. Nelis, K. Lehto, B. W. J. H. Penninx, J. Kelso, S. M. Gold, Neandertal introgression partitions the genetic landscape of neuropsychiatric disorders and associated behavioral phenotypes. *Transl Psychiatry* **12**, 433 (2022).

28. S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, D. Reich, The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).

29. C. N. Simonti, B. Vernot, L. Bastarache, E. Bottinger, D. S. Carrell, R. L. Chisholm, D. R. Crosslin, S. J. Hebbring, G. P. Jarvik, I. J. Kullo, R. Li, J. Pathak, M. D. Ritchie, D. M. Roden, S. S. Verma, G. Tromp, J. D. Prato, W. S. Bush, J. M. Akey, J. C. Denny, J. A. Capra, The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).

30. N. An, K. Bassil, G. I. Al Jowf, H. W. M. Steinbusch, M. Rothermel, L. De Nijs, B. P. F. Rutten, Dual-specificity phosphatases in mental and neurological disorders. *Progress in Neurobiology* **198**, 101906 (2021).
31. P. Rakic, Evolution of the neocortex: a perspective from developmental biology. *Nat Rev Neurosci* **10**, 724–735 (2009).
32. C. K. Kovach, N. D. Daw, D. Rudrauf, D. Tranel, J. P. O'Doherty, R. Adolphs, Anterior Prefrontal Cortex Contributes to Action Selection through Tracking of Recent Reward Trends. *Journal of Neuroscience* **32**, 8434–8442 (2012).
33. B. Malgrange, L. Nguyen, Scaling brain neurogenesis across evolution. *Science* **377**, 1155–1156 (2022).
34. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanché, J.-F. Deleuze, H. Cann, S. Mallick, D. Reich, M. S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
35. C. M. Schlebusch, P. Sjödin, G. Breton, T. Günther, T. Naidoo, N. Hollfelder, A. E. Sjöstrand, J. Xu, L. M. Gattepaille, M. Vicente, D. G. Scofield, H. Malmström, M. De Jongh, M. Lombard, H. Soodyall, M. Jakobsson, Khoe-San Genomes Reveal Unique Variation and Confirm the Deepest Population Divergence in Homo sapiens. *Molecular Biology and Evolution* **37**, 2944–2954 (2020).
36. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. De Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222–226 (2012).
37. S. Peyrégne, V. Slon, J. Kelso, More than a decade of genetic research on the Denisovans. *Nat Rev Genet* **25**, 83–103 (2024).
38. T. Maricic, N. Helmbrecht, S. Riesenberg, D. Macak, P. Kanis, M. Lackner, A. D. Pugach-Matveeva, S. Pääbo, Comment on "Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment." *Science* **374**, eabi6060 (2021).
39. T. Schoeler, D. Speed, E. Porcu, N. Pirastu, J.-B. Pingault, Z. Kutalik, 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, Manuscript Writing Group, A. G. Bick, G. A. Metcalf, K. R. Mayo, L. Lichtenstein, S. Rura, R. J. Carroll, A. Musick, J. E. Linder, I. K. Jordan, S. D. Nagar, S. Sharma, R. Meller, All of Us Research Program Genomics Principal Investigators, M. Basford, E. Boerwinkle, M. S. Cicek, K. F. Doheny, E. E. Eichler, S. Gabriel, R. A. Gibbs, D. Glazer, P. A. Harris, G. P. Jarvik, A. Philippakis, H. L. Rehm, D. M. Roden, S. N. Thibodeau, S. Topper, Biobank, Mayo, A. L. Blegen, S. J. Wirkus, V. A. Wagner, J. G. Meyer, M. S. Cicek, Genome Center: Baylor-Hopkins Clinical Genome Center, D. M. Muzny, E. Venner, M. Z. Mawhinney, S. M. L. Griffith, E. Hsu, H. Ling, M. K. Adams, K. Walker, J. Hu, H. Doddapaneni, C. L. Kovar, M. Murugan, S. Dugan, Z. Khan, E. Boerwinkle, Genome Center: Broad, Color, and Mass General Brigham Laboratory for Molecular Medicine, N. J. Lennon, C. Austin-Tse, E. Banks, M. Gatzen, N. Gupta, E. Henricks, K. Larsson, S. McDonough, S. M. Harrison, C. Kachulis, M. S. Lebo, C. L. Neben, M. Steeves, A. Y. Zhou, Genome Center: University of Washington, J. D. Smith, C. D. Frazar, C. P. Davis, K. E. Patterson, M. M. Wheeler, S. McGee, C. M. Lockwood, B. H. Shirts, C. C. Pritchard, M. L. Murray, V. Vasta, D. Leistritz, M. A. Richardson, J. G. Buchan, A. Radhakrishnan, N. Krumm, B. W. Ehmen, Data and Research Center, S. Schwartz, M. M. T. Aster, K. Cibulskis, A. Haessly, R. Asch, A. Cremer, K. Degatano, A. Shergill, L. D. Gauthier, S. K. Lee, A. Hatcher, G. B. Grant, G. R. Brandt, M. Covarrubias, E. Banks, A. Able, A. E. Green, R. J. Carroll, J. Zhang, H. R. Condon, Y. Wang, M. K. Dillon, C. H. Albach, W. Baalawi, All of Us Research Demonstration Project Teams, S. H. Choi, X. Wang, E. A. Rosenthal, NIH All of Us Research Program Staff, A. H. Ramirez, S. Lim, S. Nambiar, B. Ozenberger, A. L. Wise, C. Lunt, G. S. Ginsburg, J. C. Denny, Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
41. M. I. Kurki, J. Karjalainen, P. Palta, T. P. Sipilä, K. Kristiansson, K. Donner, M. P. Reeve, H. Laivuori, M. Aavikko, M. A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. Della Briotta Parolo, A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H. O. Heyne, K. Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S. Koskelainen, T. Paajanen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A. G. Elnahas, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro, S. Biswas, A. Bizaki-Vallaskangas, O. Carpen, C.-Y. Chen, O. A. Dada, Z. Ding, M. G. Ehm, K. Eklund, M. Färkkilä, H. Finucane, A. Ganna, A. Ghazal, R. R. Graham, E. Green, A. Hakanen, M. Hautalahti, Å. Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J. Hunkapiller, H. Jacob, J.-N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K. Kaarniranta, M. Kähönen, R. M. Kajanne, L. Kallio, R.

Kälviäinen, J. Kaprio, N. Kerimov, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V.-M. Kosma, T. Kuopio, V. Kurra, T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longerich, R. Mägi, J. Mäkelä, A. Mäkitie, A. Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S. V. Mozaffari, M. Ek. Niemi, M. Niemi, T. Niiranen, C. J. O'Donnell, M. Obeidat, G. Okafo, H. M. Ollila, A. Palomäki, T. Palotie, J. Partanen, D. S. Paul, M. Pelkonen, R. K. Pendergrass, S. Petrovski, A. Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal, N. A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V. Salomaa, J. Schleutker, R. Serpi, H. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H. Soininen, J. H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T. Tuomi, J. A. Turunen, J. C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang, M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A. S. Havulinna, M. Perola, S. Ripatti, A. Jalanko, T. Laitinen, T. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M. J. Daly, A. Palotie, FinnGen: Unique genetic insights from combining isolated population and national health register data. [Preprint] (2022). https://doi.org/10.1101/2022.03.03.22271360.

42. A. Nagai, M. Hirata, Y. Kamatani, K. Muto, K. Matsuda, Y. Kiyohara, T. Ninomiya, A. Tamakoshi, Z. Yamagata, T. Mushiroda, Y. Murakami, K. Yuji, Y. Furukawa, H. Zembutsu, T. Tanaka, Y. Ohnishi, Y. Nakamura, M. Kubo, M. Shiono, K. Misumi, R. Kaieda, H. Harada, S. Minami, M. Emi, N. Emoto, H. Daida, K. Miyauchi, A. Murakami, S. Asai, M. Moriyama, Y. Takahashi, T. Fujioka, W. Obara, S. Mori, H. Ito, S. Nagayama, Y. Miki, A. Masumoto, A. Yamada, Y. Nishizawa, K. Kodama, H. Kutsumi, Y. Sugimoto, Y. Koretsune, H. Kusuoka, H. Yanai, Overview of the BioBank Japan Project: Study design and profile. *Journal of Epidemiology* **27**, S2–S8 (2017).
43. R. G. Walters, I. Y. Millwood, K. Lin, D. S. Valle, P. McDonnell, A. Hacker, D. Avery, N. Cai, W. W. Kretzschmar, M. A. Ansari, P. A. Lyons, R. Collins, P. Donnelly, M. Hill, R. Peto, H. Shen, X. Jin, C. Nie, X. Xu, Y. Guo, C. Yu, J. Lv, R. J. Clarke, L. Li, Z. Chen, the China Kadoorie Biobank Collaborative Group, Genotyping and population structure of the China Kadoorie Biobank. [Preprint] (2022). https://doi.org/10.1101/2022.05.02.22274487.
44. O. Gokcumen, Archaic hominin introgression into modern human genomes. *American J Phys Anthropol* **171**, 60–73 (2020).
45. J. Moriano, C. Boeckx, Modern human changes in regulatory regions implicated in cortical development. *BMC Genomics* **21**, 304 (2020).
46. G. Povysil, S. Petrovski, J. Hostyk, V. Aggarwal, A. S. Allen, D. B. Goldstein, Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet* **20**, 747–759 (2019).
47. E. T. Cirulli, S. White, R. W. Read, G. Elhanan, W. J. Metcalf, F. Tanudjaja, D. M. Fath, E. Sandoval, M. Isaksson, K. A. Schlauch, J. J. Grzymski, J. T. Lu, N. L. Washington, Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nat Commun* **11**, 542 (2020).
48. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, 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); https://github.com/broadinstitute/picard.
50. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
51. K. J. Karczewski, M. Solomonson, K. R. Chao, J. K. Goodrich, G. Tiao, W. Lu, B. M. Riley-Gillis, E. A. Tsai, H. I. Kim, X. Zheng, F. Rahimov, S. Esmaeeli, A. J. Grundstad, M. Reppell, J. Waring, H. Jacob, D. Sexton, P. G. Bronson, X. Chen, X. Hu, J. I. Goldstein, D. King, C. Vittal, T. Poterba, D. S. Palmer, C. Churchhouse, D. P. Howrigan, W. Zhou, N. A. Watts, K. Nguyen, H. Nguyen, C. Mason, C. Farnham, C. Tolonen, L. D. Gauthier, N. Gupta, D. G. MacArthur, H. L. Rehm, C. Seed, A. A. Philippakis, M. J. Daly, J. W. Davis, H. Runz, M. R. Miller, B. M. Neale, Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* **2**, 100168 (2022).
52. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
53. K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. R. Andersson, L. Griffanti, G. Douaud, T. W. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, P. M. Matthews, S. M. Smith, 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.

Funding

Max Planck Society core funding (BM, MLA, GA, EE, DS, CF, SEF)

Dutch Research Council NWO; VI.Veni.202.072 (EE)

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.

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

Supplementary Materials

Materials and Methods

Figs. S1 to S3

Tables S1

References (*48–53*)