The trait specific timing of accelerated genomic change in the human lineage

Eucharist Kun[1], Mashaal Sohail[2+], Vagheesh M. Narasimhan[1,3+]

[1] Department of Integrative Biology, The University of Texas at Austin
[2] Centro de Ciencias Genómicas (CCG), Universidad Nacional Autónoma de México (UNAM)
[3] Department of Statistics and Data Science, The University of Texas at Austin
[+] Co-corresponding Authors

Summary

Humans exhibit distinct characteristics compared to our primate and ancient hominin ancestors including bipedal locomotion and enhanced neurocognitive ability, but the timing of accelerated changes in these traits is uncertain. To investigate if specific trait-associated variation show enrichment during particular periods of human evolution, we combine genome wide association study (GWAS) data from 70 traits, spanning multiple categories including AI-based image-derived morphological phenotypes of the brain, heart, and skeletal tissues with data from 12 different evolutionary regions obtained from comparative functional genomics, multi-species alignments from long read sequencing, and ancient DNA reflecting 4 different major evolutionary divergence points. These regions cover epigenetic differences in the brain between humans and rhesus macaques, various human accelerated regions (HARs) including regions from the Zoonomia Project, ancient selective sweeps, and Neanderthal introgressed alleles. Using two complementary approaches to examine enrichment between GWAS loci and genomic regions, we show that more phenotypes are enriched in earlier periods of divergence of humans with macaques and chimps, and less so during the divergence with Neanderthals. These traits span respiratory, dermatological, reproductive, metabolic, and psychiatric domains along with skeletal and brain imaging traits, consistent with striking morphological changes between humans and other primates. Among brain imaging traits, we observe an enrichment of SNPs associated with the longitudinal fasciculus in human-gained epigenetic elements since macaques, the visual cortex in HARs, and the thalamus proper in Neanderthal introgressed alleles, implying associated functions such as language processing, decision making, relay of sensory signals, and motor control are enriched at different evolutionary depths.

Introduction

Humans have distinct features compared to other primates despite our high genetic similarities, encompassing modified skeletal physiology, unique brain size and organization, as well as specific diseases and behaviors[1–5]. Evidence in both the fossil record and comparative anatomy of hominin species reveals shifts in bone structure and size, influencing both our walking patterns and cognitive abilities. In particular, changes in arm length relative to leg length and pelvic morphology facilitates bipedal locomotion, and the increased cranial size in humans results in a brain surface area three times larger than that of our closest living ancestor, the chimpanzee[4,6,7]. Compared to our most recent known hominin ancestors, Neanderthals and Denisovans, there are also variations in limb proportions despite both species also being bipedal, as well as differences in hindbrain and forebrain volume[8].

While these changes can be observed by examining the fossil evidence and comparisons with the other great apes, it is unclear when along the past few million years that accelerated genomic change for those phenotypes occurred. Two types of datasets from different domains of genomic science may help to address this question. First, over the past decade, there has been major progress in mapping human genotypes to phenotypes. Recent studies have also combined large-scale medical imaging data and genomics to uncover the genetic architecture of skeletal proportions, brain shape and size, as well as heart structure, while biobank data along with large disease-specific cohorts have found novel gene-trait associations across a variety of diseases[9–12].

Second, on the other side of genomic research, there have been remarkable strides in acquiring and sequencing DNA from both existing and extinct species. New techniques for sequencing ancient DNA have led to the identification of Neanderthal-introgressed alleles, which has unveiled the varying presence of extant hominin DNA in the genomes of modern humans stemming from admixture events around 50,000 to 60,000 years ago[13]. Moreover, advances in sequencing technology, especially long-read sequencing, have broadened the field of comparative genomics, enabling us to generate highly accurate sequences of various organisms[14–17]. By comparing genomic sequences across multiple species, researchers have discovered regions that have acquired more mutations specifically in humans compared with chimpanzees and other mammals[18]. Apart from examining changes in genomic sequences, enhancer elements unique to the human lineage were identified by comparing post translational histone modification profiles across cortical brain tissue in humans, macaques, and mice. These elements, known as human-gained enhancers and promoters (HGEPs), offer a potential avenue for insights into the evolutionary dynamics of the brain over the last approximately 25 million years[19,20]. Through combining genome-wide association studies (GWAS) with evolutionary annotations that have identified human-gained epigenetic marks, accelerated evolution, selective sweeps, and archaic introgression, we can tackle the questions of when human traits likely underwent accelerated evolution.

Previous attempts to answer these questions have used stratified linkage disequilibrium score (S-LDSC) regression to examine the distribution of heritability for phenotypes across the genome. Hujoel et al showed an enrichment of heritability of disease traits in ancient enhancers and promoters as well as in promoters of constrained genes for numerous traits[21]. Other studies have also found that variants introgressed from Neanderthals and shared across multiple Neanderthal populations are enriched in heritability for traits such as skin pigmentation and hair[22–24].

Lastly, another study discovered heritability enrichment in fetal brain human-gained enhancers as well as Neanderthal introgressed variants (NIVs) for brain MRI phenotypes related to cortical surface area and white matter[25]. These approaches have reliably united GWAS and genomic data to further our understanding of the overall evolution of human traits. We expand on this field of work by using genetic variants associated with 70 independent complex traits, 12 genomic annotations marking sequences that have evolved at different periods of human evolutionary history, and 2 methods of enrichment analysis: S-LDSC as well as a new approach, a gene enrichment pipeline called HARE, which was developed to provide a flexible pipeline for estimating gene overlap enrichment of GWAS traits with genomic regions of interest and provide a complimentary approach to S-LDSC which has known biases in accurately estimating partitioned heritability on smaller size annotations which do not cover enough common SNPs[26,27].

Results

Stratified linkage disequilibrium score regression and gene enrichment analysis of 70 GWAS traits across 12 annotations

We utilized two methods to analyze the enrichment of GWAS trait loci in genomic regions. The first method, stratified LD score regression (S-LDSC), partitions heritability from GWAS summary statistics based on genomic regions, known as annotations, and determines if a particular region explains more heritability for a given trait than would be expected from the proportion of SNPs present in that annotation[28]. For this method, we analyzed our test annotations in a model simultaneously incorporating several other regulatory elements, measures of selective constraint, and linkage statistics (baselineLDv2.2 with 97 annotations) to estimate heritability enrichment ($h^2(C)$) while minimizing bias due to model misspecification (**Methods**: Stratified LD score regression (S-LDSC) framework)[21,28–30]. The second method, HARE, maps genome-wide significant SNPs to nearby genes and scans for elevated levels of overlap between these genes and genomic annotations of interest. We then estimate overlap enrichment (*I*) by calculating the percent difference between the amount of overlap between a test annotation and a set of genes related to a GWAS trait versus the amount of overlap between the annotation and a random set of genes matched for gene length and number of genes for each trait (**Methods**: Gene

enrichment analysis for evolutionary annotations)[27]. Percentage difference values greater than 0% are considered enriched in an annotation.

In order to examine a range of complex traits and diseases that cover human specific phenotypes as well as those common to other great apes, we chose a set of 42 relatively unrelated GWAS datasets previous analyzed in Hujoel et al. across distinct trait domains including psychiatric, immunological, dermatological, skeletal, reproductive, metabolic, cardiovascular, gastrointestinal, or endocrine as defined in the GWAS Atlas (heritability (hg) = 0.92% - 0.67%, genetic correlations (rg) < 0.9, average sample size (N)=320,000)[21,31]. Additionally, we analyzed GWAS results from 3 image-derived phenotyping studies carried out on medical images. These include 16 brain MRI phenotypes spanning brain size, structure, and MRI activity, 6 right heart MRI phenotypes related to atrium size and function, and 6 skeletal dual-energy X-ray absorptiometry (DXA) phenotypes related to skeletal proportions (hg = 8% - 51%, rg < 0.9, N = 31,221 – 41135) (**Fig. 1**, **Table S1** and **Table S2**)[11,32,33].

We chose to examine the role of comparative functional genomic data, multispecies alignments from long read sequencing, and ancient DNA in shaping polygenic traits and disease evolution across a timescale of 25 million years through these genomic annotations marking various evolutionary periods: (A) Human-gained enhancers and promoters (HGEPs) in the brain since human divergence with rhesus macaque at different post-conception weeks (p.c.w) and adulthood[19,20], (B) The fastest evolving unique regions of the human genome when compared to various mammal and primate genomes (human accelerated regions (HARs), lineage-specific accelerated regions (LinARs), high-confidence HARs (zooHARs), and human ancestor quickly evolved regions (HAQERs))[16,17,34,35], (C) Ancient selective sweeps in humans relative to Neanderthals and Denisovans[36], and (D) Neanderthal introgressed variants[37]. These genomic annotations constitute different numbers of regions, genomic sizes, and proportion of total common variation (**Fig. 2**, **Table S3**). Because of previously mentioned S-LDSC limitations for accurately estimating partitioned heritability on smaller size annotations as well as strength to control for epigenetic annotations in different cell types, we performed S-LDSC analysis on only the HGEPs and HARE analysis on the remaining evolutionary annotations[26].

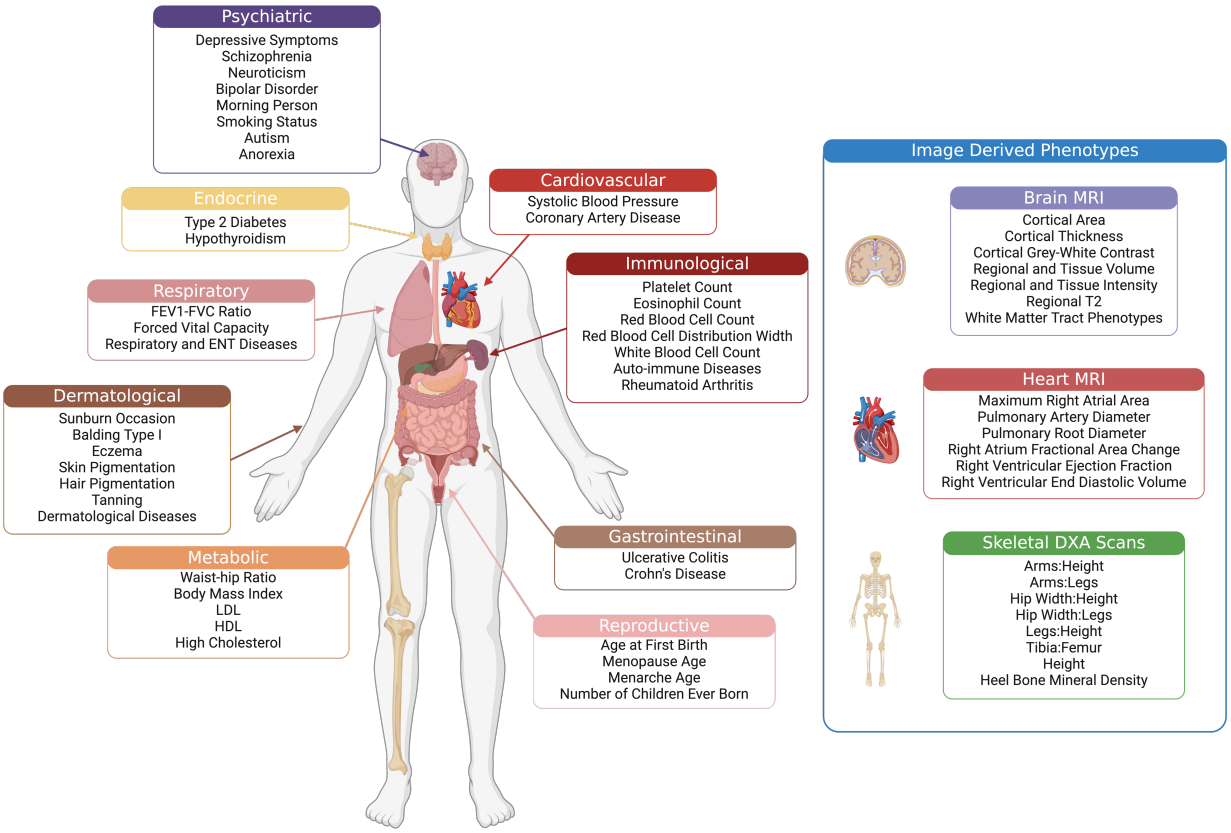


Fig. 1. Phenotypes examined in our gene and heritability enrichment analyses. Phenotypes were grouped according to study and/or GWAS Atlas category (psychiatric, endocrine, respiratory, dermatological, metabolic, cardiovascular, immunological, gastrointestinal, and reproductive). Morphological phenotypes of the human body obtained specifically from GWAS carried out on UK Biobank medical imaging cohorts (brain MRI, heart MRI, and full body DXA) are shown to the right and grouped in a blue box.

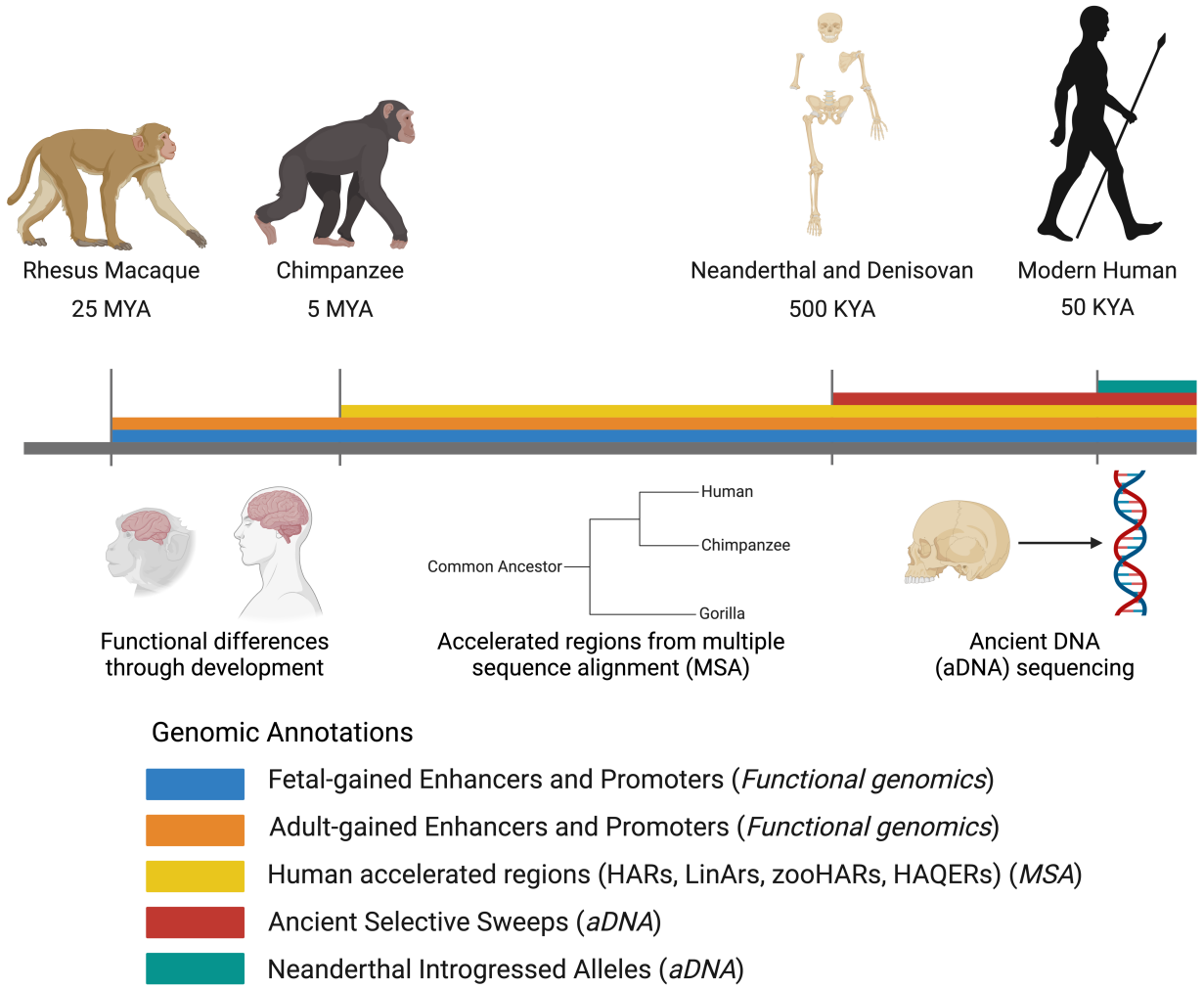


Fig. 2. Evolutionary time periods. Major time points of primate evolution relevant in this study are highlighted. Genomic annotations corresponding to evolutionary time periods are shown in color on the timeline. These annotations include fetal HGEPs (blue), adult HGEPs (orange), human accelerated regions (yellow), ancient selective sweeps (Extended Lineage Sorting ELS) (red), and putatively introgressed variants from Neanderthals (teal). Methods used to obtain each genomic region of interest are also shown below their corresponding annotations. Blue and orange intervals mark epigenetic gains in the cerebral cortex while the other color intervals mark genetic gains discovered through comparative genomics from primate and ancient DNA sequences.

Fetal human-gained enhancers and promoters in early development are enriched in traits across skeletal, respiratory, and dermatological domains

First, considering the period from our divergence from our most recent common ancestor with rhesus macaque (~25 million years ago (MYA)), we analyzed HGEPs in the brain at 7, 8, and 12 p.c.w and adulthood for a total of 6 annotations. We performed S-LDSC on 70 traits across the 6 annotations while performing additional control by adding all epigenetic fetal and

adult brain regulatory elements from the Epigenome Roadmap Project 25 state model to our model (**Methods**: Stratified LD score regression (S-LDSC) framework). We found that fetal human-gained regulatory elements at 7 p.c.w were enriched in heritability for traits across several domains at FDR-adjusted p-value < 0.05 (heel T score, balding type I, FEV1-FVC ratio, height, forced vital capacity, systolic blood pressure, skeletal proportions, and weighted-mean isotropic or free water volume fraction (ISOVF) in tract left superior longitudinal fasciculus) ($h^2(C)$ = 4.26 – 11.30, p = 4.00×10^{-2} – 6.94×10^{-5}) (**Fig. 3**). Comparatively, HGEPs at other time points either fetal or adult showed little to no enrichment for specific traits, with only forced vital capacity showing significant enrichment at 8 p.c.w ($h^2(C)$ = 6.52, p = 3.84×10^{-2}) (**Table S4**).

To check whether our heritability enrichment was affected by tissue specificity, we performed a secondary S-LDSC analysis on the 7 p.c.w time point annotation. We obtained genomic regions pertaining to regulatory elements from various tissues in the Epigenome Roadmap 25 state model and created an annotation set of enhancers and promoters that were common to all major tissues in the Roadmap Epigenome Consortium dataset[38]. We intersected the human gained elements found specifically in the brain with this pan tissue annotation set to generate two subsets of annotations, regulatory elements that overlapped with genomic regions shared by Epigenome Roadmap tissues and elements that had no overlap. We then performed S-LDSC on each set and correlated the heritability enrichment values of each subset with the original annotation set. In doing so, we found that our heritability enrichment results for both subsets were over 90% correlated with the original annotation, and we confirmed that HGEPs found only in brain tissue at 7 p.c.w contribute significant heritability enrichment across various domains including dermatological, skeletal, and respiratory traits ($h^2(C)$ 4.53 – 12.55, p = 9.56×10^{-3} – 1.23×10^{-4}) (**Fig. S1**) (**Table S5**) (**Methods**: Tissue specificity analysis). This suggests that the majority of the evolutionary enrichment signals we observe are driven by a core set of enhancers and promoters that are tissue agnostic that are differentially regulated in humans vs rhesus macaques, and is also suggestive of pleiotropic effects across the brain-body axis through epigenetic regulation.

Various sets of human accelerated regions are enriched for psychiatric, respiratory, and metabolic traits as well as the area of the visual cortex

We then carried out gene enrichment analysis on various types of human accelerated regions of the genome, HAQERs, HARs, LinARs, and zooHARs, which all pertain to the period spanning the last known common ancestor between humans and chimpanzees (~5 MYA). These comparative genomic annotations were generated through different genomic species alignments and methodologies and share various amounts of overlap as 18% of LinARs overlap with HARs and 45% of zooHARs overlap with HARs, but each annotation also exhibits distinct features (**Methods**: Differences among various types of human accelerated regions). Our results show that enrichment analysis for HARs and LinARs were largely similar (Pearson correlation = 0.71),

with any trait that was significantly enriched in LinARs also being significantly enriched in HARs (neuroticism, smoking status, forced vital capacity, BMI, and area of the primary visual cortex in the right hemisphere). However, zooHARs deviated slightly from HARs (Pearson correlation = 0.42) and LinARs (Pearson correlation = 0.29) and had unique significance (FDR-adjusted p-value < 0.05) in several traits (eosinophil count, height, and FEV1-FVC ratio) (I = 68%, p = 3.47×10^{-2}; I = 51%, p = 9.52×10^{-5}; I = 76%, p = 1.30×10^{-2}) (**Fig. 3**). BMI and forced vital capacity were significantly enriched across these three types of accelerated regions (I = 37% – 100%, p = 2.36×10^{-3} – 2.18×10^{-4}; I = 34% – 81%, p = 3.27×10^{-4} – 4.84×10^{-84}) and neuroticism, smoking status, and area of the primary visual cortex in the right hemisphere were enriched across HARs and LinARs (I = 85% – 117%, p = 3.72×10^{-4} – 8.78×10^{-23}; I = 68% – 126%, p = 4.63×10^{-3} – 1.96×10^{-27}; I = 130% – 141%, p = 2.67×10^{-2} – 1.14×10^{-3}) (**Fig. 3**) (**Table S6**). As HAQERs are found in the non-conserved genome and do not typically overlap any genes, there was no enrichment for any trait which falls in line with our expectations from this type of analysis (**Methods**: Differences among various types of human accelerated regions)[35].

Ancient selective sweeps and Neanderthal introgressed regions are enriched in immunological, dermatological, respiratory, and reproductive traits

Lastly, we examined human evolution over the past several hundred thousand years by carrying out gene enrichment analysis on ancient selective sweeps and Neanderthal introgressed regions. We observed significant enrichment (FDR-adjusted p-value < 0.05) for several traits and trait domains in both types of annotations. In agreement with hypotheses regarding differences between Neanderthals and humans, immunological, dermatological, and respiratory traits were enriched in ancient selective sweeps (eosinophil count, red blood cell distribution width, balding, FEV1-FVC ratio, and forced vital capacity) (I = 47% – 80%, p = 2.03×10^{-2} – 8.83×10^{-4}) while reproductive traits were enriched in introgressed alleles (menarche and menopause age) (I = 36%, p = 2.22×10^{-4}; I = 54%, p = 2.02×10^{-3}) (**Fig. 3**)[23,39–41]. Of particular interest, our analysis found that genes associated with the thalamus, which is responsible for relaying sensory information in the brain, were also enriched in NIVs (I = 117%, p = 1.93×10^{-2}) as well as certain body proportions known to differ between modern humans and Neanderthals (waist-hip ratio and leg-height ratio) (I = 28%, p = 2.65×10^{-4}; I = 74%, p = 2.55×10^{-2}) (**Fig. 3**) (**Table S6**)[8].

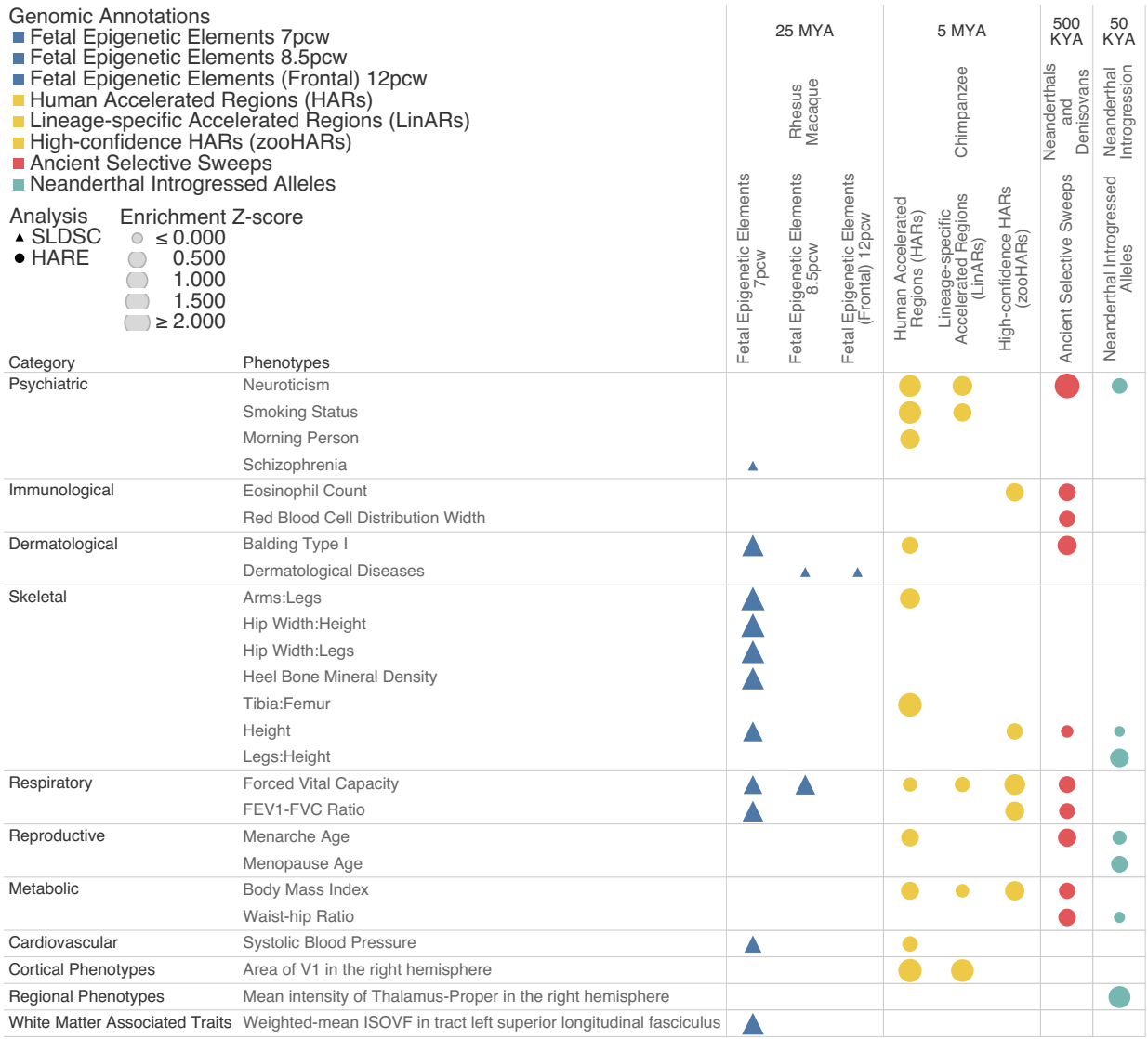


Fig. 3. Heritability and gene enrichment of human phenotypes across evolutionary time periods. S-LDSC and HARE analysis of 70 traits for 12 genomic annotations marking various time periods in human evolutionary history. Color corresponds to each major evolutionary period analyzed in accordance with the schema in **Fig. S2** and each shape marks the type of analysis performed - triangular for S-LDSC and circular for HARE. Enrichment z-scores were calculated separately for each type of analysis and the resulting enrichment value is denoted by the overall size of each shape. All results shown are significant at FDR-adjusted p-value < 0.05.

Meta-analysis of S-LDSC and gene overlap enrichment show significant enrichment across several domains

To determine if overall trait domains were enriched in genomic annotations pertaining to specific evolutionary time periods, we carried out a meta-analysis across our GWAS categories (**Methods**: S-LDSC and HARE meta-analysis). Overall, the meta-analysis of the S-LDSC results shows that the highest number of significantly enriched categories (FDR-adjusted p-value <

0.05) are in the 7 p.c.w time point (immunological, skeletal, respiratory, gastrointestinal, cortical brain MRI, regional brain MRI, and heart MRI), and that respiratory and skeletal traits are enriched across all fetal timepoints (**Fig. 4**). Additionally, psychiatric and cardiovascular traits are the sole categories to show an increase in enrichment across the developmental periods and only show significant enrichment at the 12 p.c.w time period ($h^2(C)$ = 2.56, p = 3.40 × 10^{-3}, $h^2(C)$ = 3.20, p = 1.09 × 10^{-12}) in occipital and frontal regions respectively (**Fig. 4**) (**Table S7**). Lastly, white matter associated traits which form the network of the brain by connecting different brain structures and play a crucial role in cognition, particularly cognitive speed, were the only category to be significantly enriched in heritability in adult human gained enhancers and was not enriched in any other developmental period[42,43]. We performed the same meta-analysis using the HARE results and found only a few categories of traits to be significantly enriched (FDR-adjusted p-value < 0.05) across the genomic annotations. These include psychiatric traits in HARs (I = 100%, p = 4.32 × 10^{-7}) and dermatological and respiratory traits in zooHARs, regions accelerated in our genomic evolution since divergence with chimps (I = 84%, p = 2.13 × 10^{-9}; enrichment = 0.92, p = 1.53 × 10^{-8}), but only dermatological traits were significantly enriched in regions under selection since our divergence with Neanderthals (I = 35%, p = 6.82 × 10^{-4}) (**Fig. 4**) (**Table S8**).

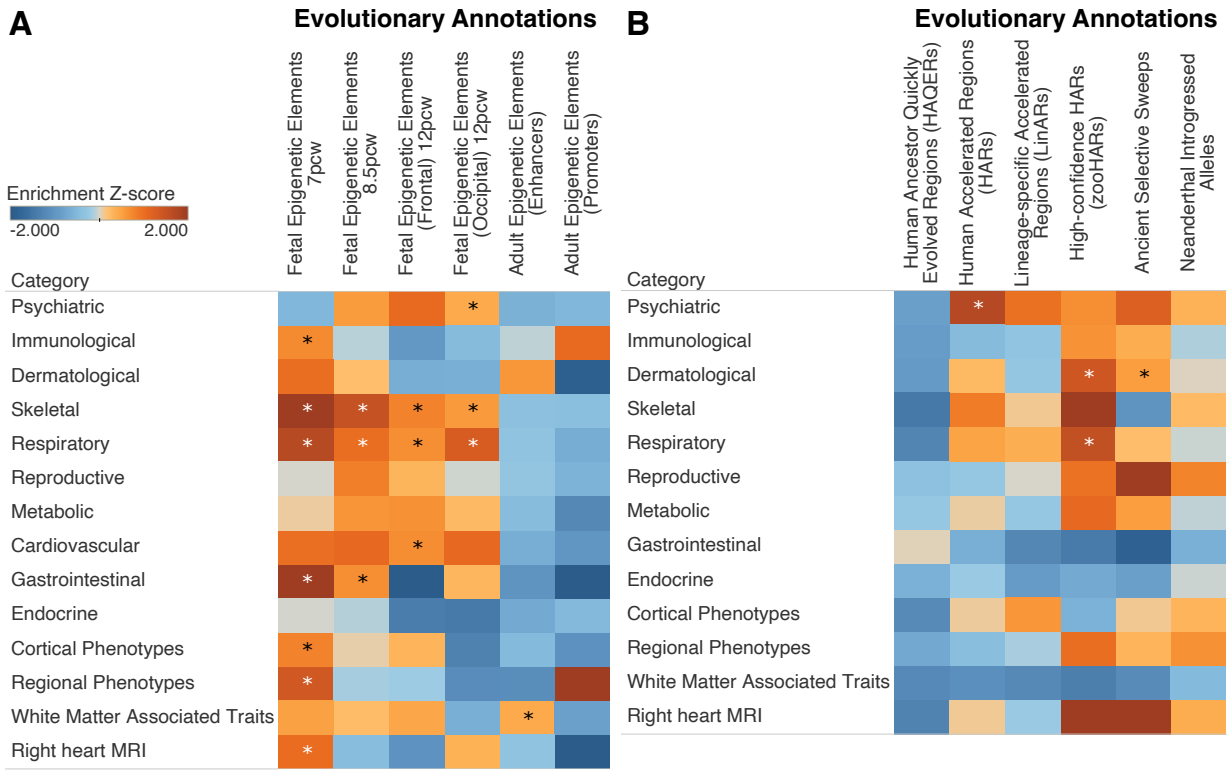


Fig. 4. Meta-analysis of heritability and gene enrichment. **A**) Meta-analysis of S-LDSC results across HGEPs based on trait categories. **B**) Meta-analysis of HARE results across evolutionary annotations based on trait categories. Asterisks denote significance at FDR-adjusted p-value < 0.05 across each type of analysis.

Genomic annotation robustness and tissue overlap sensitivity analysis

As many of the genomic annotations we used in our analysis were derived from computational approaches that may not be able to accurately determine these regions at base pair level resolution, we ran additional sensitivity analyses to test the robustness of both the various methods used to generate the annotations we tested as well as our enrichment methods. We accomplished this by modifying our genomic regions of interest in various ways. First, we examined the impact of randomly sampling just 90% of each of the annotations over 3 replicates and running S-LDSC and HARE on these subsets (**Fig. S2**). Additionally, we took a separate approach where we modified the lengths of each individual region within each annotation, first making each region 10% smaller followed by making each region 10% larger and ran S-LDSC and HARE on each modified subset. We then correlated the corresponding enrichment estimates and p-values to the original results and found that all three types of modified annotation sets had an average Pearson correlation of 95% with the original annotations (**Methods**: Annotation robustness analysis) (**Table S9** to **Table S11**).

Additionally, to determine whether the significant heritability enrichment we observed in fetal gained enhancers and promoter at 7 p.c.w and not the other fetal timepoints was largely due to overlap with epigenetic elements present in other tissues, we examined base pair overlap among all fetal timepoint annotations with regulatory elements from various tissues in the Epigenome Roadmap 25 state model (**Table S12** – **Table S14**)[38]. Our results show that base pair overlap between the brain and other tissues actually increases on average from 7 p.c.w to 12 p.c.w, indicating that the signals we observe in the 7 p.c.w annotation are not solely driven by greater overlap with regulatory regions as compared to the other fetal timepoint annotations (**Fig. S4** and **Fig. S5**) (**Methods**: Tissue specificity analysis).

Discussion

Overall, our study examines when various phenotypes underwent accelerated evolution across developmental time points and evolutionary events ranging from human divergence from macaques to ancient hominin introgression. We first examined evolutionary changes in the brain since our divergence from rhesus macaque over 25 MYA using S-LDSC and discovered that 12 of our 70 traits showed heritability enrichment in this time period. Across the various developmental time points, the 7 p.c.w period showed heritability enrichment in far more traits and trait domains than other stages of development as determined by both the individual and the meta-analysis. Our results demonstrate that epigenetic elements that gained novel function in the fetal brain since human divergence with rhesus macaque are likely involved in the development of several skeletal, dermatological, and respiratory traits at 7 p.c.w. suggesting pleiotropic effects across the brain-body axis from our early evolution and life history. Next, we analyzed 4 types of fast evolving regions of the human genome since our divergence from chimpanzees over 5 MYA using HARE and found that 14 traits showed significant gene enrichment for these genomic regions, and a meta-analysis of our results showed that psychiatric traits overall were also

307 significantly enriched in HARs. Though different methods and genomic alignments spanning
308 various species were used to generate these human accelerated regions of the genome, our results
309 largely show similar trait enrichment across each annotation (**Methods**: Differences among
310 various types of human accelerated regions). Lastly, we analyzed the last 50 - 100 thousand
311 years of human evolution through leveraging information from Neanderthal and Denisovan DNA
312 and discovered that 10 traits showed gene enrichment in ancient selective sweeps and 7 traits
313 showed gene enrichment in Neanderthal introgressed alleles. In line with observed changes in the
314 fossil record and previous genomic evidence linking selective sweeps and introgressed hominin
315 DNA in humans to various adaptive phenotypes, these traits spanned immunological,
316 dermatological, skeletal, respiratory, reproductive, and metabolic categories. Overall, our results
317 show that more traits were enriched between the periods of divergence related to humans and
318 macaques as well as humans and chimpanzees rather than the period of divergence related to
319 Neanderthals and Denisovans.
320
321 Of particular interest, traits that were enriched in multiple annotations showed continuity
322 over consecutive time periods. For example, we observed that BMI was enriched in both HARs
323 and ancient selective sweeps while height and waist to hip ratio were enriched in ancient
324 selective sweeps and NIVs, but no traits were enriched in only HARs and NIVs or only HGEPs
325 and NIVs. Balding and forced vital capacity were enriched across HGEPs at 7 p.c.w, HARs, and
326 ancient selective sweeps while neuroticism and menarche age were significantly enriched across
327 HARs, ancient selective sweeps, and NIVs. Balding, forced vital capacity, and menarche age are
328 related to observed or hypothesized phenotypic differences in lung function, hair, and
329 reproduction across primates, Neanderthals, and humans while neuroticism could be associated
330 with general changes in behavior and emotional complexity which also differ in humans from
331 our evolutionary ancestors, and our study provides firsthand genomic evidence of accelerated
332 evolution for these traits across multiple periods of evolution[23,44–48].
333
334 Across the deep-learning-based image derived phenotypes of the skeleton, brain, and
335 heart, we show evidence that all three major body systems underwent accelerated change in our
336 human lineage. In particular, our meta-analysis results show that all three modalities were
337 enriched in HGEPs since our divergence with macaques, and individual skeletal traits such as
338 arm to leg ratio and leg to height ratio as well as brain traits such as area of the visual cortex and
339 intensity of the thalamus were also enriched in accelerated regions as well as introgressed alleles
340 (**Fig. 3**), hinting at continued evolution of these traits across timepoints. These observations
341 mirror common observed morphological differences between humans and other great apes as our
342 skeleton structures, brain size and organization, and heart morphology have all undergone
343 significant changes[49–52].
344 Looking closer at the three brain imaging traits that were enriched at different
345 evolutionary timepoints, we found that each was related to a brain MRI category as well as a
346 substructure of the brain. In our earliest evolutionary timeframe regarding human divergence

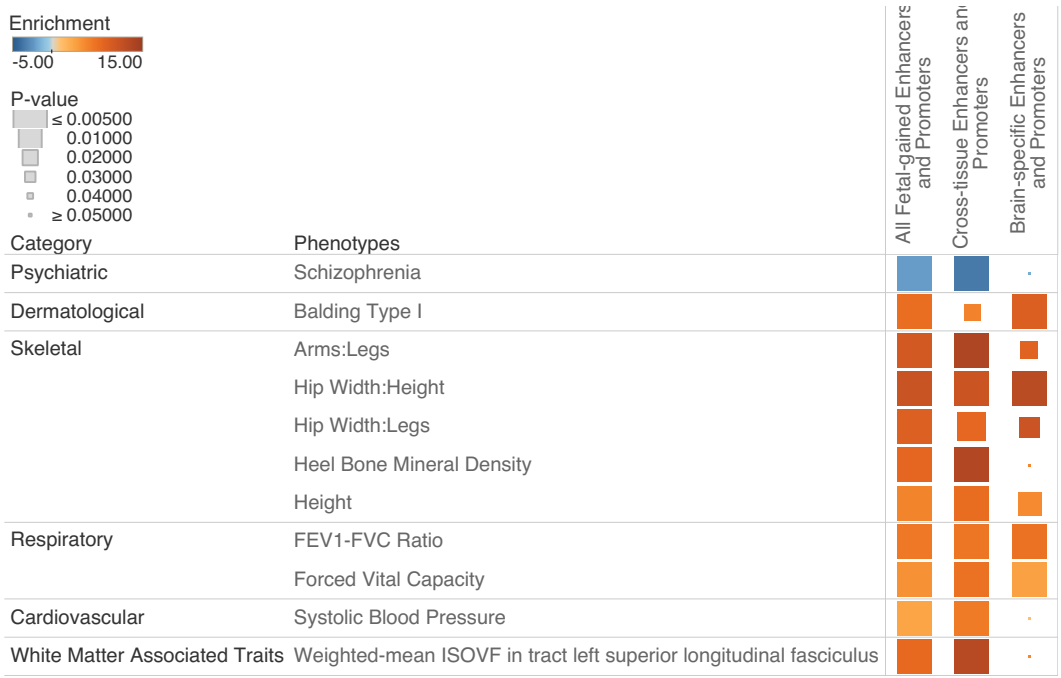
347 with macaques, we found that white matter substructure measurements in the left superior
348 longitudinal fasciculus, which is part of the complex network connecting different brain
349 structures to each other and is associated with language processing, decision making, and motor
350 function, were enriched in fetal-gained epigenetic elements at 7 p.c.w[53]. Next, we found that the
351 area of the visual cortex in the right hemisphere, which processes visual information from the
352 retinas, was enriched in human accelerated regions since our divergence with chimpanzees[54].
353 Lastly, the mean intensity of the thalamus proper in the right hemisphere, which is primarily
354 associated with relaying information from vision, taste, touch, and hearing sensory pathways to
355 the cerebral cortex as well as relaying motor signals from the cerebellum to the motor cortex,
356 was enriched in Neanderthal introgressed alleles[55]. The temporal approach of this study paints a
357 possible roadmap of the differential timepoints various brain structures underwent accelerated
358 evolution.
359
360 Our work as well as other forays into genomic enrichment analysis to take advantage of
361 continuous advances in generating large amounts of GWAS and genomic annotation data. As
362 such, we face similar limitations outlined by Hujoel et al and Alagoz et al[21,25]. Firstly, our data is
363 limited only to GWAS carried out on populations with European ancestry due to a lack of large-
364 scale cohorts of more diverse populations. Additionally, although we use some of the current
365 largest sample size GWAS available, some of the traits we analyze are likely insufficiently
366 powered for truly accurate heritability estimation. Our work would greatly benefit from increases
367 in GWAS data through larger and more diverse sample sizes as well as study of novel traits.
368
369 However, the key limitation is one of evolutionary annotation. While ancient DNA has
370 been a revolutionary technology in our ability to obtain genetic data from our archaic ancestors,
371 there is a lack of genomic data between human divergence from chimpanzees and the emergence
372 of Neanderthals. Some of the most overt changes to the human physiological form such as
373 bipedal walking and increased cranial size occurred during this particular period and is evident
374 through the hominin fossil record[56]. An important avenue moving forward with this work would
375 involve advances in aDNA sequencing techniques to elucidate genomic changes that occurred
376 from ~5 MYA to 500 thousand years ago (KYA) across various extinct hominin species. Even
377 using the latest methods, high quality ancient human genomes are only available for samples up
378 to 40,000 years old and no ancient DNA has ever been extracted from any sample dating older
379 than 430 KYA[57,58]. Lastly, our analysis of HGEPs since divergence with rhesus macaque was
380 limited to epigenetic elements found in the brain tissue. As many of our analyzed phenotypes
381 may not be affected by brain specific HGEPs, our work can only give a partial glimpse at
382 accelerated evolutionary trends during this time period. Further work to discover more human-
383 gained enhancers and promoters across a variety of tissues would provide an opportunity for a
384 more unbiased approach of examining when human traits experienced accelerated evolution.
385

386 Overall, in this study we probe a variety of complex traits and diseases for accelerated
387 evolution across four major points of human evolutionary divergence related to macaques,
388 chimpanzees, and Neanderthals/Denisovans ranging from 25 MYA to 50 KYA. We found
389 significant enrichment across various time points for traits spanning from psychiatric-related to
390 skeletal morphology. Combined, we paint a temporal landscape of when phenotypes underwent
391 accelerated evolution in the human lineage.

392 Acknowledgments

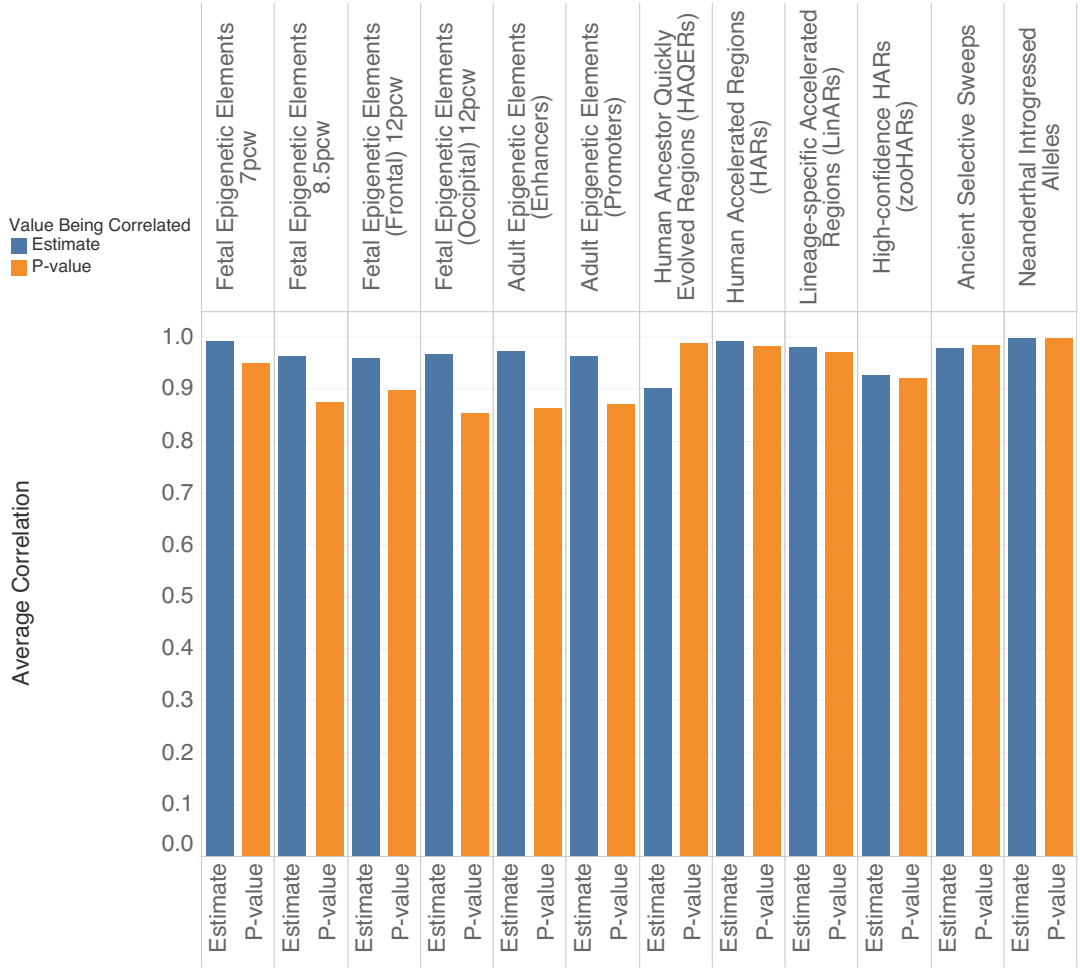
393 Figs 1 and 2 were created with Biorender.
394 Funding: V.M.N. and E.K. were supported by a grant from the Allen Discovery Center program,
395 a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation. M.S.
396 was supported by UNAM-PAPIIT grant **IA209024**. GPU and compute resources were supported
397 by a Director's Discretionary Award from the Texas Advanced Computing Cluster.
398 Author contributions: M.S. conceived of the initial project. M.S. and E.K. performed analysis.
399 E.K., M.S. and V.M.N. wrote the paper. M.S. and V.M.N. jointly supervised the work.
400 Competing interests: The authors declare no competing interests.
401 Data and materials availability: The HARE enrichment pipeline is available at
402 https://github.com/ossmith/HARE.

403 Supplementary Figures



404
405 **Fig. S1. Comparison of S-LDSC heritability enrichment** across fetal-gained enhancers and
406 promoters at 7 p.c.w in the brain. Heritability enrichment carried out on the original set shown on
407 the left column followed by cross-tissue enhancers and promoters then brain-specific enhancers

408 and promoters. All traits shown are significant in the leftmost column at FDR-adjusted p-value <
409 0.05. Larger squares represent lower p-values.



410
411 **Fig. S2. Comparison of annotation robustness.** Average estimate and p-value correlations
412 between enrichment analyses carried out on the full annotation set with three different subsets of
413 90% of the annotation. All fetal and adult epigenetic element sets were analyzed with S-LDSC
414 while the remaining genomic regions were analyzed with HARE.

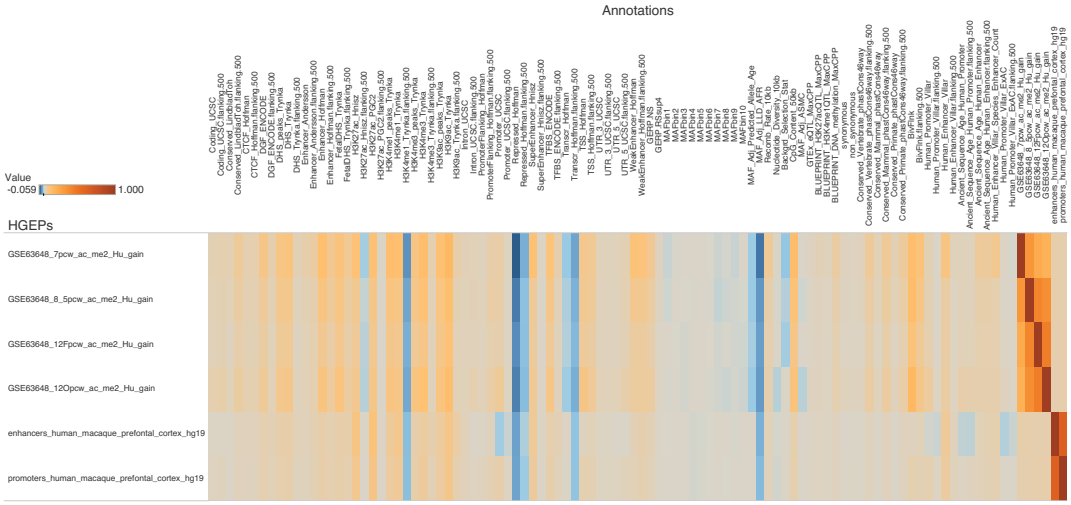


Fig. S3. Correlations between baselineLDv2.2 annotations and HGEPs.

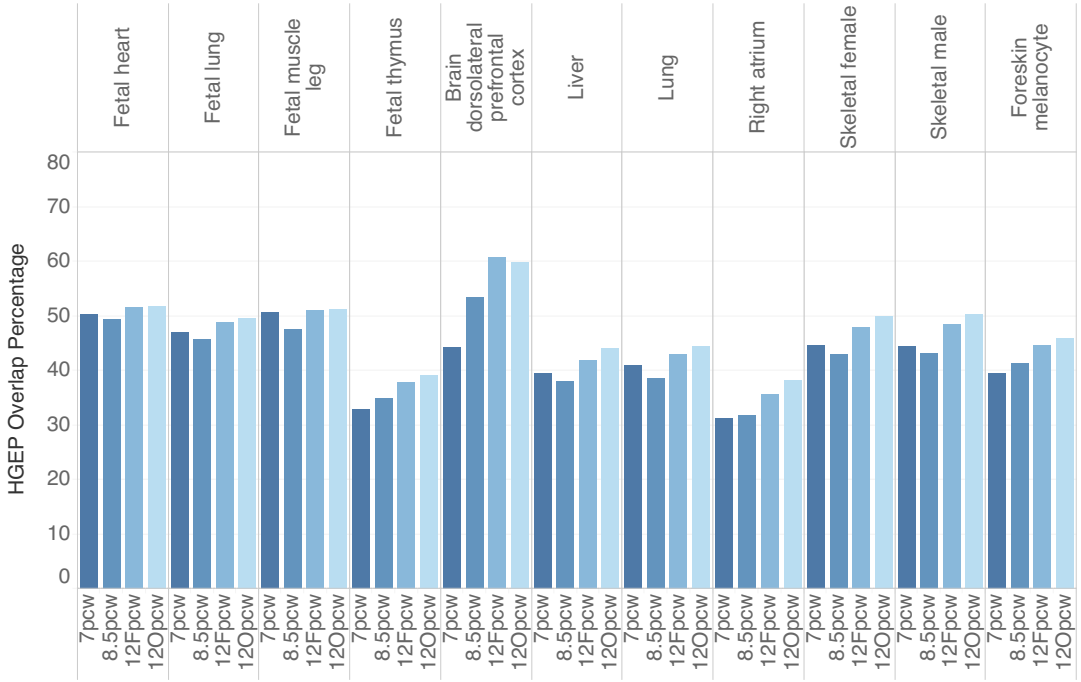


Fig. S4. Comparison of base pair overlap. Shared overlap of base pairs between enhancers and promoters of various tissues with HGEP in the brain at various fetal time points. A total of 11 tissues obtained from various origins from the Epigenome Roadmap 25 state model were analyzed. The percent overlap is calculated as the number of bases that overlap between each enhancer/promoter annotation with the Epigenome Roadmap tissue divided by the total length of the enhancer/promoter annotation - i.e. (base pair overlap of 7 p.c.w and E066 (liver) / length of 7 p.c.w)

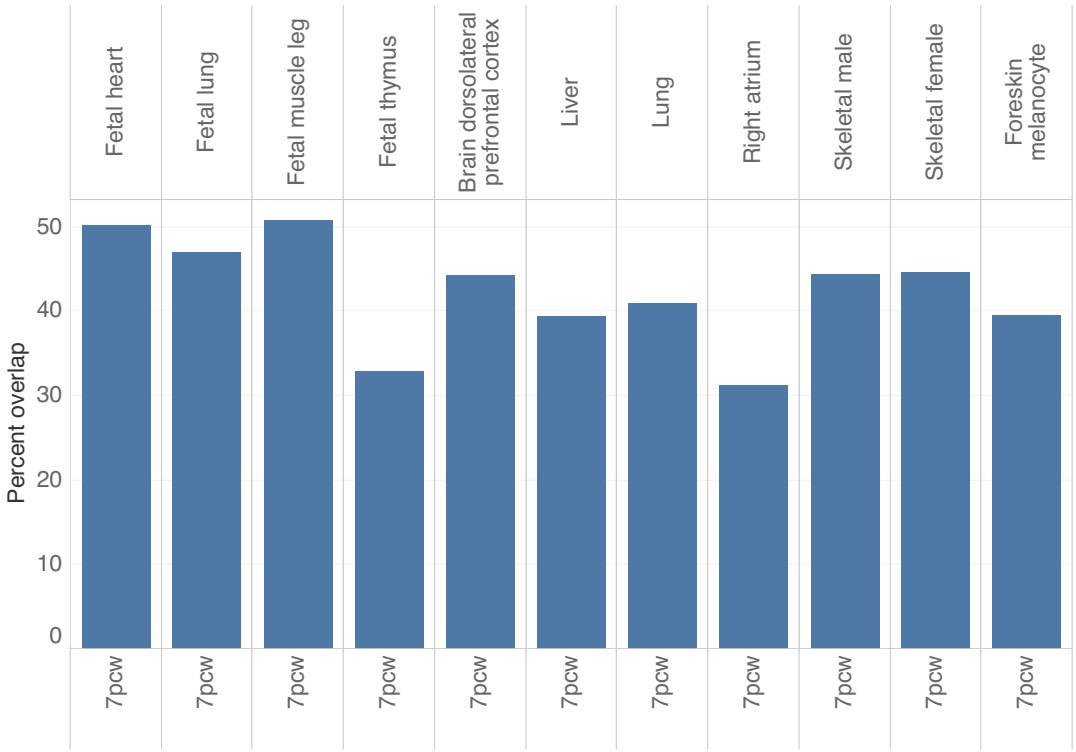


Fig. S5. Shared overlap of base pairs between enhancers and promoters of various tissues with HGEP in the brain at 7 p.c.w only.

Supplementary Tables

Table S1 - GWAS phenotypes

Table S2 - Genetic correlations

Table S3 - Evolutionary Annotations

Table S4 - S-LDSC Results

Table S5 - S-LDSC Tissue Specificity

Table S6 - HARE Results

Table S7 - S-LDSC Meta-analysis

Table S8 - HARE Meta-analysis

Table S9 - HARE Robustness 1

Table S10 - S-LDSC Robustness

Table S11 - HARE Robustness 2

Table S12 - Epigenome Roadmap Annotations

Table S13 - Epigenome Roadmap State 25 Model

Table S14 - Annotation Overlap

Table S15 - HARE HGEP Results

STAR Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
LDSC	Bulik-Sullivan, et al.[59]	https://github.com/bulik/ldsc
S-LDSC	Finucane, et al.[28]	https://github.com/bulik/ldsc
Ensembl Assembly Converter	Martin, et al.[60]	https://grch37.ensembl.org/Homo_sapiens/Tools/AssemblyConverter
HARE	GitHub	https://github.com/ossmith/HARE
BEDTools	Quinlan, et al.[61]	https://github.com/arq5x/bedtools2

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Eucharist Kun (eucharistkun@utexas.edu).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

- This paper does not report original code, but all code used are listed in the key resources table and are publicly available as of the date of publication.
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

Method details

GWAS datasets

We analyzed a set of 41 independent GWAS summary statistics that were previously curated and analyzed[21,62]. These were curated by screening 34 GWAS summary statistics that are publicly available and 55 UK Biobank traits for which summary statistics were computed using BOLT-LMM (up to N= 459,324 European-ancestry samples). BOLT-LMM software is available at https://data.broadinstitute.org/alkesgroup/BOLT-LMM and the BOLT- LMM summary statistics for UK Biobank traits are available through the Alkes group. The set was first filtered to 47 traits

465 with z-scores of total SNP heritability of at least 6[62]. Since the set of 47 included 6 traits that
466 were duplicated in two different datasets (genetic correlation >= 0.9), the final set of GWAS was
467 reduced to 41 independent GWAS datasets (N=10,263 – 459,324; average N=320,000). We also
468 analyzed GWAS obtained from 3 imaging studies carried out on UKB participants. We obtained
469 16 brain MRI traits (average N = 31,435), 10 right heart MRI traits (N = 41,135), and 6 skeletal
470 DXA traits (N = 31,221)[11,32,33]. The brain MRI traits were chosen by taking the trait with the
471 highest heritability from each MRI trait category as denoted by Smith et al[33]. Traits within each
472 imaging study that had a genetic correlation higher than 0.9 were again filtered out. In total, we
473 analyzed a set of 70 traits (**Table S1** and **Table S2**).

474 ***Evolutionary genomic annotations***

475 Genomic annotations were obtained from the following sources: (A) Epigenetic elements that
476 gained novel function in the brain since our divergence with rhesus macaque at different
477 developmental stages or post-conception weeks (p.c.w) and adulthood (Fetal human-gained (HG)
478 enhancers and promoters at 7 p.c.w, 8.5 p.c.w, and 12 p.c.w and adult human-gained (HG)
479 enhancers and promoters)[19,20], (B) The fastest evolving regions of the human genome when
480 compared to various sets of mammals and primates (human accelerated regions (HARs), lineage-
481 specific accelerated regions (LinARs), high-confidence HARs (zooHARs), and human ancestor
482 quickly evolved regions (HAQERs))[16,17,34,35], (C) Ancient selective sweeps from the extended
483 lineage sorting method capturing human-specific sweeps relative to Neanderthal/Denisovan[36],
484 and (D) Putatively introgressed variants from Neanderthals (**Table S3**)[37].

485 ***Annotation conversion to GRCH37***

486 Annotations that were not in GRCH37 format were converted from GRCH38 to GRCH37 using
487 the Ensembl Assembly Converter tool available at
488 https://grch37.ensembl.org/Homo_sapiens/Tools/AssemblyConverter and stored in BED format.
489 Resulting regions were merged using BEDTools 'merge'[61].

490 ***Stratified LD score regression (S-LDSC) framework***

491 Stratified LD score regression (S-LDSC) was developed in Finucane et al 2015[28]. The
492 framework takes as input summary statistics from a GWAS, and LD-scores for SNPs in a set of
493 test and control genomic categories. S-LDSC estimates whether a genomic region is enriched or
494 depleted in heritability for a set of traits, capturing the contribution of variants in that genomic
495 region towards trait variation, and whether this contribution is more or less than expected given
496 the relative proportion of variants in that region. The per-SNP heritability of a SNP i and
497 heritability enrichment of a genomic annotation or category are estimated from:
498

499 $$Var(\beta_i) = \sum_k \tau_k \, 1\,\{i \in C_k\}$$

Where total heritability is represented as:

$$\sum_i Var(\beta_i)$$

And heritability in category C_k:

$$\sum_{i \in C_k} Var(\beta_i)$$

Then heritability enrichment ($h^2(C)$) in category C_k is calculated as:

$$\frac{\sum_{i \in C_k} Var(\beta_i) / \sum_i Var(\beta_i)}{|C_k|/M}$$

Lastly, jackknife standard error estimates generated by S-LDSC are used to compute confidence intervals for $h^2(C)$. Further details are outlined in Finucane et al 2015.

To account for multiple significance testing, the false discovery rate (FDR) was controlled at 0.05 using the approach of Benjamini-Hochberg[63].

S-LDSC genomic panels

The framework of S-LDSC uses three SNP panels. Here, we describe each and the set of SNPs used for each in our analysis, following previous analyses using S-LDSC[21,28,30].

Reference SNPs (9,997,231 SNPs): These are SNPs used by S-LDSC to compute LD scores. We used the 1000 Genomes Phase 3 SNPs[64] from the European super population with a minor allele count > 5.

Regression SNPs (1,187,349 SNPs): These are SNPs used by S-LDSC to estimate τ_k from marginal association statistics. We used the HapMap Project Phase 3 SNPs (MHC region excluded)[65].

Heritability SNPs (5,961,159 SNPs): These are SNPs used by S-LDSC to compute h^2, $h^2(C)$, $|c|$ and sd_c. We used common variants (MAF > 0.05) among the set of reference SNPs (see above).

S-LDSC joint model and control annotations

We used the S-LDSC framework to estimate the heritability enrichment in a given genomic annotation while jointly modeling the same parameters across a set of genomic annotations (Fig 3) (**Table S4**). The idea of the joint model is to account for confounding effects of other genomic annotations while estimating the heritability enrichment of a given annotation. The baseline model was presented in Finucane et al[28] and includes a set of 53 annotations, those that partition the gene body as well as several epigenetic annotations. The baselineLDv1.1[30] further includes a

set of annotations related to selection such as allele frequency, allelic age, LD, nucleotide diversity, GERP scores and background selection for a total of 75 annotations. The BaselineLDv2.2[21,28–30] is the most comprehensive and conservative set of annotations to include in a joint model with a test annotation and includes 97 genomic annotations (**Fig. S3**). This is the recommended baseline model and further includes annotations relating to conservation and age of epigenetic marks and whether a promoter is a promoter of an ExAC gene[21]. All heritability enrichment analyses in this study jointly model test annotations with the baselineLDv2.2 model. In our analysis of fetal human-gained regulatory elements, we further included in our model fetal brain regulatory elements from the Epigenome Roadmap 25 State Model[38], taking the union of male (E081) and female (E082) marks. In our analysis of adult human-gained regulatory elements, we further included in our model adult brain prefrontal cortex regulatory elements (E073) from the Epigenome Roadmap 25 State Model[38]. For the 25 state model, a chromatin state model based on the imputed data for 12 marks, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H4K20me1, H3K79me2, H3K36me3, H3K9me3, H3K27me3, H2A.Z, and DNase, across all 127 reference epigenomes with 25-states was learned[38]. We included the following states that correspond to all regulatory elements (**Table S12** and **Table S13**).

Tissue specificity analysis

As a proportion of the human-gained elements in the brain also overlap with regulatory elements in other tissues (**Table S14**), we performed additional analysis to determine the main driver of heritability enrichment across the different time points. Firstly, we examined whether the enrichment signals in the fetal gained enhancers and promoters at 7 p.c.w were driven by a greater number of base pair overlap with other tissues such as the heart, lung, or bone when compared to the other fetal timepoints. We obtained the genomic regions pertaining to regulatory elements from various tissues in the Epigenome Roadmap 25 state model to examine regions common to all tissues as well as each set of fetal HGEP (**Table S12** – **Table S14**)[38]. The number of base pair overlap between the brain and other tissues increases from 7 p.c.w to 12 p.c.w, ruling out increased overlap in regulatory regions at earlier time points rather than later time points as the main driving factor behind the increased heritability enrichment we observe at 7 p.c.w (**Fig. S4** and **Fig. S5**). To further elucidate the driving factor behind enrichment in the 7 p.c.w HGEPs, we split the annotation into two sets, regulatory elements that overlapped with genomic regions shared by all Epigenome Roadmap tissues and elements that had no overlap and performed S-LDSC on each set. We then correlated the heritability enrichment values of each subset with the original annotation. We found that the subset of brain HGEPs that overlapped regulatory elements in other tissues had an enrichment correlation of 0.91 with the original annotation. Similarly, the subset of brain HGEPs that shared no overlap with regulatory elements in other tissues had an enrichment correlation of 0.90 with the original results. Our results also show that heritability enrichment for some traits was only significant in the subset of brain HGEPs that overlapped regulatory elements in other tissues (p-value < 0.05) (**Fig. S1**). However, we see that the fetal brain-specific HGEPs at 7 p.c.w do contribute significant heritability

579 enrichment across various trait domains (dermatological, skeletal, and respiratory) suggesting
580 pleiotropic effects across the brain-body axis from our early evolution and life history.

581 ***Gene enrichment analysis for evolutionary annotations***

582 We scanned for elevated levels of intersections between genes containing genome-wide
583 significant SNPs and our genomic annotations through a modified version of the method outlined
584 in Xu et al[66]. For each phenotype, we first created annotations of protein coding regions that lie
585 on our genome-wide significant SNPs using Ensembl's GRCh37 Variant Effect Predictor version
586 105. We selected the closest protein coding feature within 5,000 base pairs up- or downstream of
587 the SNP. Using biotype categorizations identified by VEP, these protein coding features were:
588 ("protein_coding", "IG_C_gene", "IG_D_gene", "IG_J_gene", "IG_LV_gene", "IG_M_gene",
589 "IG_V_gene", "IG_Z_gene", "nonsense_mediated_decay", "nontranslating_CDS",
590 "non_stop_decay", "polymorphic_pseudogene", "TR_C_gene", "TR_D_gene", "TR_J_gene").
591 We refer to the list of features for all independent genome-wide significant loci significantly
592 associated with the trait as the *element set* for the phenotype being analyzed. Phenotypes with
593 fewer than 50 elements in their set were removed from the analysis due to insufficient power.
594 We then used BioMart command line queries to generate the genomic locations (chromosome,
595 start, stop) of each feature within the human genome. In order to scan for selection, we used
596 BEDTools 'intersect' to compute the number of intersections found in the gene set with genomic
597 annotations sourced from literature.
598 To generate a background distribution of intersections per bp, we computed the annotation-
599 element intersections per bp of 1,000 length-matched element sets. Because the distribution of
600 these feature lengths is non-normal, we binned the element sets into deciles based on gene length
601 and computed the average length l within each bin of size n. For each bin in the simulation, we
602 sampled n random elements of length l to create our complete element set which was then used
603 to compute the intersections per base pair of the simulated set. Due to the large differences in
604 element set sizes and lengths across phenotypes, a background distribution was generated
605 independently for each phenotype analyzed. On this background we fit a Weibull distribution for
606 computation of p-values of the observed intersections in comparison to the background. Overlap
607 enrichment was then calculated as a percent difference between GWAS gene overlap with an
608 annotation compared to random gene overlap with an annotation. This was calculated by
609 subtracting the number of intersections between genes related to a GWAS trait and the
610 annotation of interest from the average number of intersections between random genes and the
611 annotation of interest and then dividing by the average number of intersections between random
612 genes and the annotation of interest[27].
613

614 $$\frac{(\text{GWAS gene overlap} - \text{average random gene overlap})}{\text{average random gene overlap}} \times 100\%$$

615 ### *S-LDSC and HARE meta-analysis*

616 To meta-analyze results while accounting for genetic correlations among the various traits, we
617 performed a random effects meta-analysis by each trait category using the rma.mv function from
618 the metafor package. This computes a summary estimate and summary standard error from a
619 collection of effect estimates and standard errors for each trait and also considers a genetic
620 variance-covariance matrix among the traits. These were used to compute confidence intervals
621 and z-scores to test for significance of the summary estimate for each annotation. To account for
622 multiple significance testing, the false discovery rate (FDR) was controlled at 0.05 using the
623 approach of Benjamini-Hochberg.

624 ### *Annotation robustness analysis*

625 In order to assess the sensitivity and specificity of the various approaches used in finding these
626 various evolutionary genomic annotations, we examined the impact of randomly sampling just
627 90% of each of the regions over 3 replicates and running S-LDSC and HARE on these subsets.
628 We then correlated the corresponding enrichment estimates and p-values to the original results
629 and found that these subsets had an average Pearson correlation of 95% with the original
630 annotations (**Fig. S2**). Another approach we took was to shorten or lengthen the ends of every
631 genomic region in each annotation by 5% and re-run HARE on the altered set of regions. The
632 correlation between these sets and the original sets was greater than 99% (**Table S9** – **Table**
633 **S11**). All correlations were performed with base R functions.

634 ### *Annotation comparison*

635 Annotations were tested for base pair overlap using BEDTools[61]. Annotations were paired and
636 the overlapping regions between the annotations was calculated using BEDTools 'intersect'. The
637 proportion of intersecting base pairs between two annotations with the length of each original
638 annotation was then calculated (**Table S14**).

639 ### *Differences among various types of human accelerated regions*

640 As the comparative genomic annotations were generated through slight differences in either the
641 number or type of species used in genome alignment, we wanted to probe the variation across the
642 4 annotations. The set of HARs we examined were taken from a study that combined 4 datasets
643 that used a variety of filters and limited multi-species alignment (generally chimpanzee, mouse,
644 and rat genomes)[34] while the LinARs and zooHARs were generated from aligning new long read
645 sequences of many more species. In particular, LinARs were generated from aligning 49 high-
646 quality reference genomes from primate species while the zooHARs were generated from an
647 alignment of 241 mammalian genomes[16,17]. These 3 comparative genomic analyses focused on
648 conserved regions of the genome, which is only 5% of the total human genome. Our last
649 genomic annotation, HAQERs, were generated by performing comparative genomic analysis on

650 only great ape species, which allowed for the discovery of fast evolving regions of the human
651 genome present in the non-conserved region of the genome[35]. Comparison of the location and
652 size of these various genomic annotations shows that HARs, LinARs, and zooHARs are closely
653 related and share some amount of overlap while HAQERs are vastly different (**Table S14**). We
654 found that 18% of LinARs overlapped HARs while 45% of zooHARs overlapped HARs.
655 However, in terms of gene enrichment, LinARs and HARs had an enrichment correlation of 0.71
656 while zooHARs and HARs only had an enrichment correlation of 0.42.

657 ### *S-LDSC and HARE comparison*

658 We carried out HARE analysis on the fetal and adult human gained enhancers and promoters for
659 all traits to see if the results were in agreement with S-LDSC results. Overall, the correlation
660 between heritability enrichment and gene enrichment was low (Pearson correlation = 0.23),
661 however traits that were significantly enriched in S-LDSC were also significant in HARE (FDR-
662 adjusted p-value < 0.05) (**Table S15**).

663 ## References

664 1. Finch, C. E. Evolution of the human lifespan and diseases of aging: Roles of infection,
665 inflammation, and nutrition. *Proc Natl Acad Sci U S A* **107**, 1718–1724 (2010).
666 2. Varki, N. M., Strobert, E., Dick, E. J., Benirschke, K. & Varki, A. Biomedical differences
667 between human and nonhuman hominids: Potential roles for uniquely human aspects of
668 sialic acid biology. *Annual Review of Pathology: Mechanisms of Disease* **6**, 365–393
669 (2011).
670 3. Ackermann, R. R. Craniofacial Variation and Developmental Divergence in Primate and
671 Human Evolution. *Tinkering: The Microevolution of Development* 262–279 (2007)
672 doi:10.1002/9780470319390.CH17.
673 4. Van Essen, D. C., Donahue, C. J. & Glasser, M. F. Development and Evolution of
674 Cerebral and Cerebellar Cortex. *Brain Behav Evol* **91**, 158–169 (2018).
675 5. Matsuzawa, T. *Primate Origins of Human Cognition and Behavior*. (Springer-Verlag
676 Publishing, 2001).
677 6. Young, M. *et al.* The developmental impacts of natural selection on human pelvic
678 morphology. *Sci Adv* **8**, 4884 (2022).
679 7. Young, N. M., Wagner, G. P. & Hallgrímsson, B. Development and the evolvability of
680 human limbs. *Proc Natl Acad Sci U S A* **107**, 3400–3405 (2010).
681 8. Helmuth, H. Body height, body mass and surface area of the Neandertals. *Z Morphol*
682 *Anthropol* **82**, 1–12 (1998).
683 9. Pirruccello, J. P. *et al.* Deep learning enables genetic analysis of the human thoracic aorta.
684 *Nat Genet* **54**, 40–51 (2021).
685 10. Pan-UKB team. https://pan.ukbb.broadinstitute.org. *2020*.

686 11. Kun, E. *et al.* The genetic architecture and evolution of the human skeletal form. *Science*
687 *(1979)* **381**, (2023).
688 12. Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK
689 Biobank. *Nature 2018 562:7726* **562**, 210–216 (2018).
690 13. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328**, 710 (2010).
691 14. Nurk, S. *et al.* The complete sequence of a human genome. *Science (1979)* **376**, 44–53
692 (2022).
693 15. Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing
694 and its applications. *Nature Reviews Genetics 2020 21:10* **21**, 597–614 (2020).
695 16. Bi, X. *et al.* Lineage-specific accelerated sequences underlying primate evolution. *Sci Adv*
696 **9**, (2023).
697 17. Whalen, S. *et al.* Three-dimensional genome rewiring in loci with human accelerated
698 regions. *Science (1979)* **380**, (2023).
699 18. Pollard, K. S. *et al.* Forces Shaping the Fastest Evolving Regions in the Human Genome.
700 *PLoS Genet* **2**, 1599–1611 (2006).
701 19. Reilly, S. K. *et al.* Evolutionary changes in promoter and enhancer activity during human
702 corticogenesis. *Science (1979)* **347**, 1155–1159 (2015).
703 20. Vermunt, M. W. *et al.* Epigenomic annotation of gene regulatory alterations during
704 evolution of the primate brain. *Nat Neurosci* **19**, 494–503 (2016).
705 21. Hujoel, M. L. A., Gazal, S., Hormozdiari, F., van de Geijn, B. & Price, A. L. Disease
706 Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient
707 Sequence Age and Conserved Function across Species. *Am J Hum Genet* **104**, 611–624
708 (2019).
709 22. McArthur, E., Rinker, D. C. & Capra, J. A. Quantifying the contribution of Neanderthal
710 introgression to the heritability of complex traits. *Nature Communications 2021 12:1* **12**,
711 1–14 (2021).
712 23. Simonti, C. N. *et al.* The phenotypic legacy of admixture between modern humans and
713 Neanderthals. *Science* **351**, 737 (2016).
714 24. Sankararaman, S. *et al.* The landscape of Neandertal ancestry in present-day humans.
715 *Nature* **507**, 354 (2014).
716 25. Alagöz, G. *et al.* Using neuroimaging genomics to investigate the evolution of human
717 brain structure. *Proc Natl Acad Sci U S A* **119**, e2200638119 (2022).
718 26. Tashman, K. C., Cui, R., O'Connor, L. J., Neale, B. M. & Finucane, H. K. Significance
719 testing for small annotations in stratified LD-Score regression. *medRxiv* (2021)
720 doi:10.1101/2021.03.13.21249938.
721 27. Smith, O. HARE. *GitHub* (2023) doi:10.5281/zenodo.7793834.
722 28. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
723 wide association summary statistics. *Nat Genet* **47**, 1228–1235 (2015).
724 29. Marnetto, D. *et al.* Evolutionary Rewiring of Human Regulatory Networks by Waves of
725 Genome Expansion. *Am J Hum Genet* **102**, 207–218 (2018).

726 30. Gazal, S. *et al.* Linkage disequilibrium–dependent architecture of human complex traits
727 shows action of negative selection. *Nat Genet* **49**, 1421–1427 (2017).
728 31. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex
729 traits. *Nature Genetics 2019 51:9* **51**, 1339–1348 (2019).
730 32. Pirruccello, J. P. *et al.* Genetic analysis of right heart structure and function in 40,000
731 people. *Nature Genetics 2022 54:6* **54**, 792–803 (2022).
732 33. Smith, S. M. *et al.* An expanded set of genome-wide association studies of brain imaging
733 phenotypes in UK Biobank. *Nature Neuroscience 2021 24:5* **24**, 737–745 (2021).
734 34. Capra, J. A., Erwin, G. D., Mckinsey, G., Rubenstein, J. L. R. & Pollard, K. S. Many
735 human accelerated regions are developmental enhancers. *Philosophical Transactions of*
736 *the Royal Society B: Biological Sciences* **368**, (2013).
737 35. Mangan, R. J. *et al.* Adaptive sequence divergence forged new neurodevelopmental
738 enhancers in humans. *Cell* **185**, 4587-4603.e23 (2022).
739 36. Peyregne, S., Boyle, M. J., Dannemann, M. & Prufer, K. Detecting ancient positive
740 selection in humans using extended lineage sorting. *Genome Res* **27**, 1563–1572 (2017).
741 37. Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S. & Akey, J. M. Analysis of Human
742 Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* **173**, 53-61.e9
743 (2018).
744 38. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human
745 epigenomes. *Nature* **518**, 317–330 (2015).
746 39. Harvati, K. Neanderthals. *Evolution: Education and Outreach* **3**, 367–376 (2010).
747 40. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of
748 Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* **26**, 1241–
749 1247 (2016).
750 41. Zeberg, H., Kelso, J. & Paabo, S. The Neandertal Progesterone Receptor. *Mol Biol Evol*
751 **37**, 2655–2660 (2020).
752 42. Filley, C. M. & Fields, R. D. Glial Cells and Neuronal Signaling: White matter and
753 cognition: making the connection. *J Neurophysiol* **116**, 2093 (2016).
754 43. Bullock, D. N. *et al.* A taxonomy of the brain's white matter: twenty-one major tracts for
755 the 21st century. *Cerebral Cortex (New York, NY)* **32**, 4524 (2022).
756 44. Hill, W. D. *et al.* Genetic contributions to two special factors of neuroticism are associated
757 with affluence, higher intelligence, better health, and longer life. *Mol Psychiatry* **25**, 3034
758 (2020).
759 45. Schwartz, G. G. & Rosenblum, L. A. Allometry of primate hair density and the evolution
760 of human hairlessness. *Am J Phys Anthropol* **55**, 9–12 (1981).
761 46. Rowell, T. E. Variation in age at puberty in monkeys. *Folia Primatol (Basel)* **27**, 284–296
762 (1977).
763 47. Papadimitriou, A. The Evolution of the Age at Menarche from Prehistorical to Modern
764 Times. *J Pediatr Adolesc Gynecol* **29**, 527–530 (2016).

765 48. Gea, J. The evolution of the human species: A long journey for the respiratory system.
766 *Arch Bronconeumol* **44**, 263–270 (2008).
767 49. Ríos, L. *et al.* The aorta in humans and African great apes, and cardiac output and
768 metabolic levels in human evolution. *Scientific Reports 2023 13:1* **13**, 1–13 (2023).
769 50. Sayers, K., Raghanti, M. A. & Lovejoy, C. O. Human evolution and the chimpanzee
770 referential doctrine. *Annu Rev Anthropol* **41**, 119–138 (2012).
771 51. Sherwood, C. C., Subiaul, F. & Zawidzki, T. W. A natural history of the human mind:
772 tracing evolutionary changes in brain and cognition. *J Anat* **212**, 426 (2008).
773 52. Bert, A. A. *et al.* Transesophageal echocardiography in healthy young adult male baboons
774 (Papio hamadryas anubis): Normal cardiac anatomy and function in subhuman primates
775 compared to humans. *Prog Pediatr Cardiol* **35**, 109–120 (2013).
776 53. Janelle, F., Iorio-Morin, C., D'amour, S. & Fortin, D. Superior Longitudinal Fasciculus: A
777 Review of the Anatomical Descriptions With Functional Correlates. *Front Neurol* **13**,
778 (2022).
779 54. Huff, T., Mahabadi, N. & Tadi, P. Neuroanatomy, Visual Cortex. *StatPearls* (2023).
780 55. Herrero, M. T., Barcia, C. & Navarro, J. M. Functional anatomy of thalamus and basal
781 ganglia. *Childs Nerv Syst* **18**, 386–404 (2002).
782 56. Wood, B. & Lonergan, N. The hominin fossil record: taxa, grades and clades. *J Anat* **212**,
783 354 (2008).
784 57. Meyer, M. *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los
785 Huesos hominins. *Nature 2015 531:7595* **531**, 504–507 (2016).
786 58. Prüfer, K. *et al.* A genome sequence from a modern human skull over 45,000 years old
787 from Zlatý kůň in Czechia. *Nat Ecol Evol* **5**, 820 (2021).
788 59. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from polygenicity
789 in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
790 60. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Res* **51**, D933–D941 (2023).
791 61. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
792 features. *Bioinformatics* **26**, 841–842 (2010).
793 62. Hormozdiari, F. *et al.* Functional disease architectures reveal unique biological role of
794 transposable elements. *Nature Communications 2019 10:1* **10**, 1–8 (2019).
795 63. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and
796 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*
797 *(Methodological)* **57**, 289–300 (1995).
798 64. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
799 (2015).
800 65. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human
801 populations. *Nature* **467**, 52–58 (2010).
802 66. Xu, K., Schadt, E. E., Pollard, K. S., Roussos, P. & Dudley, J. T. Genomic and Network
803 Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions.
804 *Mol Biol Evol* **32**, 1148 (2015).