Human-specific lncRNAs contributed critically to human evolution by distinctly regulating gene expression

Jie Lin[1,2], Yujian Wen[1], Ji Tang[1], Xuecong Zhang, Huanlin Zhang, Hao Zhu[1,3,4*]

1 Bioinformatics Section, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China

2 Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics, Institute of Aging Research, Guangdong Medical University, Dongguan, 523808, China

3 Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence, Southern Medical University, Guangzhou, 510515, China

4 Guangdong Provincial Key Lab of Single Cell Technology and Application, Southern Medical University, Guangzhou, 510515, China

* Corresponding author. Email: zhuhao@smu.edu.cn (H.Z.)

ABSTRACT

What genes and regulatory sequences make humans and apes that share substantial genes but show distinct phenotypes has puzzled researchers for decades. Genomic studies have examined species-specific genes and regulatory sequences (e.g., transcription factor binding sites, TFBS); birth, loss, and changes in these genes and sequences can greatly drive speciation and evolution. However, those involved in epigenetic regulation - species-specific lncRNA genes and regulatory sequences - remain poorly explored. We identified human-specific (HS) lncRNAs from GENCODE-annotated human lncRNAs, predicted their DNA binding domains (DBDs) and binding sites (DBSs) genome-wide, analyzed DBSs in modern humans (CEU, CHB, and YRI), archaic humans (Altai Neanderthals, Denisovans, and Vindija Neanderthals), and their counterparts in chimpanzees, and analyzed how DBSs influence gene expression in modern and archaic humans. Our results suggest that HS lncRNAs and their DBSs have substantially rewired gene expression human-specifically, the rewiring has evolved continuously from archaic to modern humans, and rewired gene expression has promoted brain development, made humans adapt to new environments and lifestyles, and caused differences in modern humans.

Keywords: Long noncoding RNA; Human-specific lncRNA; Human evolution; Archaic human

genomes

1. INTRODUCTION

Genomic changes that drive evolution include gene birth (Kaessmann, 2010), gene loss (Albalat and Canestro, 2016), and changes in regulatory sequences (Prud'homme et al., 2007). The limited genomic differences but the substantial phenotypic and behavioral divergence between humans and other hominids make what genomic changes have critically driven human evolution an enduring puzzle (Anton et al., 2014; Pollen et al., 2023). Waves of studies have identified genes promoting human brain development (Evans et al., 2005; Fiddes et al., 2018; Mekel-Bobrov et al., 2005; Pinson et al., 2022; Suzuki et al., 2018), but conclusions from gene birth-centric studies are controversial (Currat et al., 2006; Timpson et al., 2007; Yu et al., 2007). Moreover, genes important for brain development (e.g., *NOTCH2NL* and *ASPM*) may not regulate other human traits that coincide with or even precede brain enlargement. Studies revealed regulatory sequence changes important for human evolution, including new sequences (Liu et al., 2021), lost sequences (McLean et al., 2011), human accelerated regions (HARs) (Dong et al., 2016; Mangan et al., 2022; Prabhakar et al., 2006; Whalen and Pollard, 2022), and the turnover of transcription factor binding sites (TFBS) (Krieger et al., 2022; Otto et al., 2009; Zhang et al., 2023). However, how epigenetic regulatory sequences contribute to human evolution remains barely explored.

Experiments have generated substantial findings on epigenetic regulation. (a) About one-third of human lncRNA are primate-specific (Derrien et al., 2012), and the chimpanzee and human genomes are almost 99% identical (Chimpanzee Sequencing and Analysis Consortium, 2005). (b) Species-specific lncRNAs have distinct expression profiles in organs of different species (Sarropoulos et al., 2019). (c) Many lncRNAs can bind to DNA sequences by forming RNA:DNA triplexes (Abu Almakarem et al., 2012), and recruit histone and DNA modification enzymes to binding sites (Lee, 2009; Roadmap Epigenomics et al., 2015). (d) Approximately 40% of differentially expressed human genes result from interspecies epigenetic differences (Hernando-Herraez et al., 2015). Thus, besides human-specific (HS) protein-coding genes and related regulatory sequences (e.g., TFBS), HS lncRNAs and their DNA binding sites may also critically promote human evolution and determine human traits by human-specifically rewiring gene expression.

Since RNA sequence search by structure and sequence alignment can reliably identify orthologous lncRNA genes in genomes (Nawrocki, 2014; Nawrocki and Eddy, 2013), HS lncRNAs can be identified upon human lncRNAs' orthologues in other mammals. Many lncRNAs contain DNA binding domains (DBDs) and can use DBDs to bind to DNA binding sites (DBSs) in the genome. Since RNA:DNA triplexes follow specific base-pairing rules (Abu Almakarem et al., 2012), DBDs and DBSs can be computationally predicted (Lin et al., 2019; Wen et al., 2022). Detected gene expression in organs and tissues, especially in human organs and tissues reported by the GTEx project (GTEx Consortium, 2017), provides data for examining whether the expression of lncRNAs and their potential targets (genes with DBSs) are correlated. Further, genomes of modern humans, archaic humans, and multiple apes have been sequenced (1000 Genomes Project Consortium et al., 2012; Chimpanzee Sequencing and Analysis Consortium, 2005; Meyer et al., 2012; Prufer et al., 2017; Prüfer et al., 2013), making the above analyses performed across species. Thus, it is time to examine whether HS lncRNAs and their DBSs have substantially and human-specifically rewired gene expression and determined human traits.

This study identified HS lncRNAs upon the orthologues of the GENCODE-annotated human lncRNA genes in 16 mammalian genomes, predicted HS lncRNAs' DBDs and DBSs in the genomes of modern humans (CEU, CHB, and YRI), archaic humans (Altai Neanderthals, Denisovans, Vindija Neanderthals), and these DBSs' counterparts in the genome of chimpanzees (Figure 1A). We experimentally validated DBS prediction using CRISPR/Cas9 experiment and multiple cell lines. Cross-population and cross-species analyses allowed us to exploit reasonable methods, controls, gene numbers, and data sources. To analyze DBSs accurately, which have different lengths, unclear biochemical properties, and significant variations across populations and species, we computed the binding affinity for each DBS upon its length and identity (percent of paired nucleotides) and classified DBSs into strong and weak ones. We also classified DBSs into old and young ones upon when most of their sequence changes occurred. These strategies ensure robust genomic analyses, which include evolutionary/population genetics analyses, and transcriptomic analyses, which cover multiple datasets, including the GTEx data and data of multiple experiments. Our analyses reveal that HS lncRNAs and their DBSs have distinctly and continuously rewired gene expression during

human evolution, and the rewired gene expression determines or greatly influences diverse human traits from a rapidly evolved brain to a modified metabolism system.

2. RESULTS

2.1. HS lncRNAs regulate diverse genes and transcripts

How many human lncRNA genes are human-specific is an important question still without a clear answer. After searching for orthologues of the 13562 GENCODE-annotated human lncRNA genes (v18) in 16 mammalian genomes and in the genomes of archaic humans using the *Infernal* program that performs RNA sequence and structure alignment (Figure 1A) (Lin et al., 2019; Nawrocki, 2014; Nawrocki et al., 2009), we identified 66 HS lncRNA genes (Supplementary Table 1). Then, we predicted DBSs of HS lncRNAs in the 5000 bp promoter regions of the 179128 Ensembl-annotated transcripts (release 79) using *the LongTarget* program (Supplementary Note 1) (Lin et al., 2019). The binding affinity of a pair of DBD and DBS depends on their length and identity (i.e., the percent of paired nucleotides). Analyzing DBSs at known imprinted genes predicted using different *LongTarget* parameters and comparing DBSs with CHART-seq-reported binding sequences ("peaks") reveal the reasonable affinity threshold of 60 (He et al., 2015) (Supplementary Fig. 2). Since weak DBSs could be generated recently and play subtle functions, we used affinity>=60 and 36<affinity<60 as the thresholds for identifying strong and weak DBSs. 105141 strong DBSs (mean length>147 bp) were identified in 96789 transcripts of 23396 genes; 152836 weak DBSs were identified in 127898 transcripts of 33185 genes (Figure 1B; Supplementary Table 2, 3). These genes and transcripts are HS lncRNAs' potential targets. Several HS lncRNAs (e.g., RP11-423H2.3) have abundant targets; many targets have DBSs for multiple HS lncRNAs; many HS lncRNAs have multiple DBSs in a single target; and about 1.5% of target genes (0.6% of target transcripts) are human-specific (Figure 1C). Further, HS lncRNAs themselves show complex targeting relationships (Figure 1D; Supplementary Note 2). These results indicate that HS lncRNAs have distinctly rewired the expression of substantial conserved genes instead of regulating evolutionarily new genes.

We validated DBS prediction using multiple methods and datasets (Supplementary Note 1; Supplementary Fig 1-5). (a) HS lncRNA DBSs overlap well with experimentally identified DNA methylation and histone modification signals in multiple cell lines (https://genome.UCSC.edu). (b)

Many DBSs co-localize with annotated cis-regulatory elements (cCREs) in genes' promoter regions (https://genome.UCSC.edu). (c) Predicted DBSs overlap well with experimentally detected DBSs of NEAT1, MALAT1, and MEG3 (Mondal et al., 2015; West et al., 2014). (d) We used CRISPR/Cas9 to knock out predicted DBDs (100-200 bp) of seven lncRNAs (three HS lncRNAs and four wrongly transcribed lncRNAs) in multiple cell lines, performed RNA-seq before and after DBD knockout, analyzed differential gene expression, and found that the |fold change| of target genes was significantly larger than the |fold change| of non-target genes (one-sided Mann-Whitney test, p = 3.1e-72, 1.49e-114, 1.12e-206, 2.58e-09, 6.49e-41, 0.034, and 5.23e-07, respectively). This result suggests that the changed expression of target genes is caused by DBD knockout. (e) The *LongTarget* program uses the Smith-Waterman local alignment to identify DBD/DBS binding; thus, taking a DBS of 147 bp as an example, it is extremely unlikely to be generated by chance (p < 8.2e-19 to 1.5e-48).

HARs are conserved genomic loci yet show accelerated evolution in the human lineage. Researchers assume that HARs critically influence human evolution, but results obtained from HAR studies are controversial (Levchenko et al., 2018). The Zoonomia Project recently identified 312 HARs important for 3D genome rewiring and neurodevelopment (Keough et al., 2023). Since only 8 HARs overlap 26 DBSs of 14 HS lncRNAs, we infer that DBS and HAR are functionally different sequences.

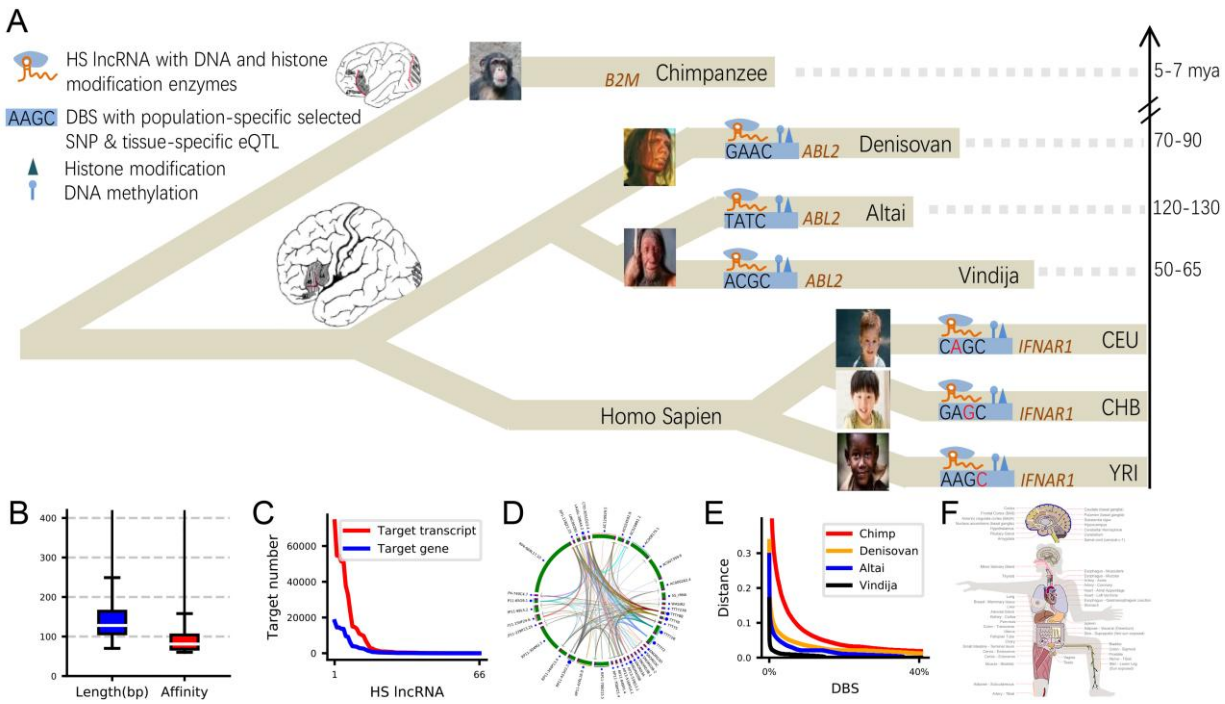

Figure 1. Study Overview. (A) The relationships between the chimpanzees, the three archaic humans (Altai Neanderthals, Denisovans, and Vindija Neanderthals), and the three modern human populations. Dashed lines indicate the phylogenetic distances from modern humans to archaic humans and chimpanzees. Genes and DBSs have the following implications – the DBS in *B2M* lacks a counterpart in chimpanzees; the DBS in *ABL2* has great differences between archaic and modern humans; the DBS in *IRNAR1* is highly polymorphic in modern humans. Red letters in DBSs indicate tissue-specific eQTLs or population-specific favored mutations. (B) DBSs' mean length and binding affinity. (C) The target genes and transcripts of 66 HS lncRNAs. (D) The targeting relationships between HS lncRNAs. (E) The sequence distances of DBSs (the top 40%) from modern humans to chimpanzees and archaic humans. (F) The impacts of interactions between HS lncRNAs and DBSs on gene expression in GTEx tissues.

2.2. HS lncRNAs' target genes characterize human evolution and traits

Since strong and weak DBSs may occur and function in different periods in human evolution, it is interesting and necessary to examine whether their target genes are enriched in distinct functions. About 5% of genes have significant sequence differences in humans and chimpanzees, but more show expression differences due to regulatory sequences. We sorted target genes by their DBS affinity and, to be prudential, chose the top 2000 genes (DBS length>252 bp and binding affinity>151) and bottom 2000 genes (DBS length<60 bp but binding affinity>36) to conduct over-representation analysis using the *g:Profiler* program and Gene Ontology (GO) database to the two gene sets (Supplementary Table 4). The top and bottom 2000 genes (with the whole genome as the reference) are co-enriched in GO terms important for human evolution, including "neuron projection development" (GO:0031175), "learning" (GO:0007612), "regulation of anatomical structure size" (GO: GO:0090066), "head development" (GO:0060322), and "regulation of locomotion" (GO:0040012) (Supplementary Note 3; Supplementary Table 4). The genes containing strongest DBSs (514 DBSs have length >500 bp) include *IFNAR1* and *NFATC1* (important for immune function), *KIF21B* and *NTSR1* (critical for neural development), *SLC2A11* and *SLC2A1* (involved in glucose usage), *BAIAP3* (a brain-specific angiogenesis inhibitor) and *TAS1R3* (a receptor for the sweet taste response), and some primate-specific noncoding RNA genes (e.g., *CTD-3224I3.3* which is highly expressed in the cerebellum, lung, and testis) (Table 1; Supplementary

Table 4, 5). On the other hand, genes with weak DBSs are enriched in such GO terms as "response to alcohol" (GO:0097305), "response to temperature stimulus" (GO:0009266), "DNA methylation or demethylation" (GO:0044728), and "female pregnancy" (GO:0007565) (Supplementary Note 3). These findings justify the classification of DBSs into strong and weak ones and indicate that strong and weak DBSs have different functions.

Table 1 Genes with DBSs that have largest binding affinity and mostly changed sequence distances (from modern humans to archaic humans and chimpanzees)

Target gene	Annotation	Binding affinity	Mostly changed
IFNAR1	i.e., Interferon Alpha And Beta Receptor Subunit 1.	794	C,D
NFATC1	A TF that induces gene transcription during immune responses.	736	C
NFATC1	A TF that induces gene transcription during immune responses.	491	C,A,D
ANKLE2	Diseases associated with ANKLE2 include microcephaly.	527	C,D
SEMA4D	Regulating phosphatidylinositol 3-kinase signaling, neuron projection development, and phosphate metabolic process.	495	C,A,D
KIF21B	Essential for neuronal morphology, synapse function, and learning and memory.	471	C
ALDH3B2	An aldehyde dehydrogenase for alcohol metabolism.	444	C
NTSR1	A brain and gastrointestinal peptide that mediates functions of neurotensin (e.g., hypotension, hyperglycemia, hypothermia, and antinociception).	402	C,A,D
MC5R	A receptor for melanocyte-stimulating hormone and adrenocorticotropic hormone.	397	C

THEG	Specifically expressed in the germ cells and involved in spermatogenesis.	395	C,D
HERC6	In pathways including class I MHC mediated antigen processing and presentation and innate immune system.	369	C,A,D
SLC2A11	Facilitating glucose transporter.	356	C
NGEF	Playing a role in axon guidance regulating ephrin-induced growth cone collapse and dendritic spine morphogenesis.	354	C
SHC2	Involved in the signal transduction pathways of neurotrophin-activated Trk receptors in cortical neurons.	345	C,D
BAIAP3	Regulating behavior and food intake by controlling calcium-stimulated exocytosis of neurotransmitters and serotonin and hormones like Insulin.	321	C
SLURP1	A marker of late differentiation of the skin.	319	C
MLPH	Involved in melanosome transport.	307	C
TAS1R3	Responding to the umami taste stimulus and recognizing diverse natural and synthetic sweeteners.	304	C,D
SLC2A1	A major glucose transporter in the mammalian blood-brain barrier.	356	C
CTD-3224I3.3	A lncRNA highly expressed in the cerebellum, lung, and testis.	312	C,A,D

Letters 'C', 'A', 'D', and 'V' indicate that the DBS in the gene has mostly changed sequence distances from modern humans to chimpanzees, Altai Neanderthals, Denisovans, and Vindija Neanderthals, respectively. SNORA59B and TTTY8/TTTY8B have quite different DBSs (whose binding affinity is 736 and 491) in *NFATC1*, so they are displayed separately.

2.3. The rewiring of gene expression has greatly evolved

The human genome is almost 99%, 99%, and 98% identical to the genomes of chimpanzees, bonobos, and gorillas (Chimpanzee Sequencing and Analysis Consortium, 2005; Prufer et al., 2012), but what sequence variants critically differentiate these species remains unknown. We assessed whether HS lncRNAs' DBSs are conserved in these species or human-specific. Notably, 97.81% of the 105141 strong DBSs have counterparts in chimpanzees, suggesting that these DBSs are similar to HARs in evolution and have undergone human-specific evolution.

HS lncRNAs may be generated before, with, or after their DBSs. To analyze the evolution of HS lncRNAs and their DBSs, we identified their counterparts in not only chimpanzees but also Altai Neanderthals, Denisovans, and Vindija Neanderthals (Meyer et al., 2012; Prufer et al., 2017; Prüfer et al., 2013). Counterparts of HS lncRNAs and strong DBSs were identified in these archaic humans, but the distances (distance per base) of HS lncRNAs are smaller than the distances of DBSs between archaic and modern humans, suggesting that many DBSs are older than HS lncRNAs and HS lncRNAs were generated after most of their DBSs. This suggests that HS lncRNAs make conversed genes in apes expressed specifically in humans.

We computed DBS distances in two ways to examine DBS evolution. The first is from the reconstructed human ancestor (downloaded from the EBI website) to chimpanzees, archaic humans, and modern humans; the second is from chimpanzees and archaic humans to modern humans. The first set of distances has the anomaly that many distances from the human ancestor to modern humans are smaller than to archaic humans. For example, DBSs in the top 20% genes in the human ancestor have distances >0.015 to modern humans, but more DBSs have distances >0.015 to Altai Neanderthals, Denisovans, and even chimpanzees (Supplementary Note 3). This anomaly is caused by the human ancestor being built using six primates without archaic humans. The second set of distances agrees well with the phylogenetic distances between chimpanzees, archaic humans, and modern humans. Large distances are in a small portion of DBSs (Figure 1E), thus targets containing these DBSs may greatly characterize human evolution. Based on a recent study that identified 5,984 genes differentially expressed between human-only and chimpanzee-only iPSC lines (Song et al., 2021), we estimated that the top 20% (4248) genes in chimpanzees may well characterize the human-chimpanzee differences, and we found that these genes have DBS

distances >0.034. Taking 0.034 as the threshold, we identified 1256, 2514, and 134 genes in Altai Neanderthals, Denisovans, and Vindija Neanderthals with DBS distance >0.034. These numbers agree with the phylogenetic relationships between these archaic and modern humans (Figure 1A; Supplementary Table 6).

To reveal whether target genes with mostly changed DBSs are enriched in specific functions, we applied over-representation analysis to target genes in chimpanzees, Altai Neanderthals, Denisovans, and Vindija Neanderthals using the *g:Profiler* program and Gene Ontology (GO) database. To make statistical significance comparable, we used just the top 1256 genes as the inputs. The 1256 genes in chimpanzees and Altai Neanderthals are enriched in many common GO terms. However, “behavior” (GO:0007610) and "cognition" (GO:0050890) have much higher significance in Altai Neanderthals than in chimpanzees (Figure 2), and some GO terms such as "learning" (GO:0007612), “learning or memory” (GO:0007611), and “glucose homeostasis” (GO:0042593) is enriched only in Altai Neanderthals. In addition, agreeing with the closer relationship between Vindija Neanderthals and modern humans, the 134 genes in Vindija Neanderthals are enriched only in two GO terms with low significance (“smoothened signaling pathway” (GO:0007224) and “regulation of smoothened signaling pathway” (GO:0008589)) (Supplementary Table 7). These results reveal when and how HS lncRNA-mediated epigenetic regulation influences human evolution.

GO terms	Adjusted_p	GO terms	Adjusted_p
regulation of vesicle-mediated transport	2.78E-04	**behavior**	1.32E-06
maintenance of location	7.46E-04	trans-synaptic signaling	1.43E-06
mitochondrial RNA metabolic process	9.22E-04	**learning**	2.16E-05
positive regulation of GTPase activity	1.12E-03	neuron projection development	9.91E-05
protein localization to plasma membrane	2.20E-03	**cognition**	1.49E-04
modification-dependent macromolecule catabolic process	2.35E-03	regulation of lipid localization	2.40E-04
regulation of GTPase activity	2.77E-03	**learning or memory**	4.74E-04
ubiquitin-dependent protein catabolic process	2.89E-03	**response to hypoxia**	5.76E-04
miRNA metabolic process	3.29E-03	**associative learning**	7.42E-04
receptor-mediated endocytosis	3.50E-03	regulation of neuron projection development	1.63E-03
neuron projection development	3.78E-03	response to virus	2.25E-03
behavior	5.23E-03	brain development	3.10E-03
brain development	7.47E-03	cell morphogenesis involved in neuron differentiation	3.93E-03
regulation of anatomical structure size	7.79E-03	response to nutrient levels	5.76E-03
negative regulation of neuron projection development	8.50E-03	neuron projection morphogenesis	6.68E-03
head development	8.61E-03	**glucose homeostasis**	6.86E-03
regulation of neuron projection development	9.23E-03	head development	7.90E-03
cell morphogenesis involved in neuron differentiation	1.07E-02	visual behavior	1.04E-02
axon development	2.96E-02	**glucose metabolic process**	1.41E-02
retina development in camera-type eye	3.12E-02	axon development	1.66E-02
response to starvation	3.12E-02	response to light stimulus	1.69E-02
regulation of lipid localization	3.78E-02	regeneration	1.71E-02
lipid storage	3.78E-02	**sensory perception of mechanical stimulus**	2.38E-02
adult locomotory behavior	4.81E-02	**sensory perception of sound**	2.98E-02
cognition	4.99E-02	response to alcohol	4.31E-02

Figure 2. 1256 target genes whose DBSs have the largest distances from modern humans to chimpanzees and Altai Neanderthals are enriched in different *Biological Processes* GO terms. Upon significance threshold = 0.05 (Benjamini-Hochberg FDR), the two gene sets in chimpanzees and Altai Neanderthals are enriched in 199 and 409 GO terms (50<terms size<1000), respectively. Shared GO terms are color-marked. Left: Top GO terms enriched in genes in chimpanzees. Right: Top GO terms enriched in genes in Altai Neanderthals.

2.4. Epigenetic regulation by HS lncRNAs shows differences in modern humans

To check whether the rewiring has been population-specifically revised in modern humans, we analyzed single nucleotide polymorphisms (SNPs, with minimal allele frequency (MAF)≥0.1) in strong DBSs using the 1000 Human Genome data (1000 Genomes Project Consortium et al., 2012). We computed the average number of SNPs for all DBSs, identified highly polymorphic DBSs (SNP numbers in CEU, CHB, YRI are 5 times larger than the average), and identified DBSs with mostly changed sequence distances from modern humans to chimpanzees, Altai Neanderthals, and Denisovans (sequence distance>0.034 to chimpanzees, Altai Neanderthals, and Denisovans which are indicated by C, A, D in Table 1,2). The intersection of the two classes of DBSs (highly polymorphic DBSs and mostly changed DBSs), wherein SNPs may greatly change binding affinity, are found in genes important for human evolution (Supplementary Table 8), including *IFNAR1*, *NFATC1*, *TAS1R3*, *INS*, *ST3GAL4*, and *FN3KRP* (Table 2). Genes involved in sugar metabolism are especially notable. A key feature of modern human diets is high sugar intake which causes non-enzymatic oxidation of proteins (i.e., glycation) and deactivates protein functions. Among proteins encoded by genes that have these two classes of DBSs, TAS1R3 recognizes diverse natural and synthetic sweeteners, Insulin decreases blood glucose concentration, ST3GAL4 regulates protein glycosylation, and FN3KRP deglycates proteins. These DBSs indicate continuous evolution of gene regulation human-specifically and population-specifically by HS lncRNAs.

We next used multiple statistical tests, including XP-CLR (Chen et al., 2010), iSAFE (Akbari et al., 2018), Tajima's D (Tajima, 1989), Fay-Wu's H (Fay and Wu, 2000), the fixation index (Fst) (Weir and Cockerham, 1984), and linkage disequilibrium (LD) (Slatkin, 2008), with widely-used parameters and thresholds (Supplementary Note 4, 5), to detect selection signals in HS lncRNAs and strong

DBSs in CEU, CHB, and YRI which are the representative populations of Europeans, Asians, and Africans. These tests detected selection signals in several HS lncRNA genes in CEU and CHB but not in YRI and selection signals in more DBSs in CEU and CHB than in YRI. Selection signals in YRI may be underestimated due to fewer samples and smaller sample sizes (than CEU and CHB), yet the overall results agree well with features of human evolution. Signals in considerable DBSs were detected by multiple tests, indicating the reliability of the analysis (Supplementary Note 4, 5; Supplementary Table 9, 10, 11). DBSs in the two pigmentation-related genes *MC1R* and *MFSD12* and the two odor reception-related genes *OR6C1* and *TAS1R3* are notable examples. The rich SNPs, population-specific selection signals, and specific functions of related genes together indicate the ongoing influence of rewired gene expression on human evolution.

Table 2 Genes with most polymorphic DBSs and DBSs with mostly changed sequence distances from modern humans to archaic humans and chimpanzees

Target gene	Annotation	SNP number	Mostly changed
IFNAR1	i.e., Interferon Alpha And Beta Receptor Subunit 1.	31	C,D
DECR2	The related pathways include metabolism and regulation of lipid metabolism.	17	C,A,D
DOK7	Essential for neuromuscular synaptogenesis.	17	C,D
TAS1R3	Responding to the umami taste stimulus and recognizing diverse natural and synthetic sweeteners.	17	C,D
NFATC1	A TF that induces gene transcription during immune responses.	16	C
ELSPBP1	Binding to spermatozoa and playing a role in sperm capacitation.	15	C,D
ST3GAL4	Involved in protein glycosylation.	15	C,D

CAMK2B	Calcium/calmodulin-dependent protein kinase important for dendritic spine and synapse formation and maintaining synaptic plasticity.	13	C,D
HLA-DQB1-AS1	Highly expressed in EBV-transformed lymphocytes, lung, spleen.	13	C,A,D,V
ANKLE2	Diseases associated with ANKLE2 include microcephaly.	12	C,D
KRTAP1-3	The KAP proteins form a matrix of keratin intermediate filaments that contribute to the structure of hair fibers.	12	C,A,D
INS, INS-IGF2	Insulin decreases blood glucose concentration.	11	C,A,D
SHC2	Involved in the signal transduction pathways of neurotrophin-activated Trk receptors in cortical neurons	11	C,D
FN3KRP	Deglycating proteins to restore their function, important for modern humans adaptive to high glucose intake and functions in all tissues.	10	C,D
TFB1M	The encoded protein is part of the basal mitochondrial transcription complex and is necessary for mitochondrial gene expression.	10	C,A,D

The most SNP-riched DBSs are also DBSs with mostly changed sequence distances from modern humans to archaic humans and chimpanzees. C, A, D, and V indicate that the DBS in the gene has mostly changed sequence distances from modern humans to chimpanzees (C), Altai Neanderthals (A), Denisovans (D), and Vindija Neanderthals (V), respectively.

2.5. Young weak DBSs may have greatly promoted recent human evolution

To further examine whether the rewiring has been revised recently, we classified strong and weak DBSs with significant sequence changes into "old" ones, "young" ones, and others. We defined old DBSs as having "Human-Chimp distance>0.034 AND Human-Altai Neanderthals distance==0" and

young DBSs as having "Human-Altai Neanderthals distance>0.034 OR Human-Denisovan distance>0.034", which demark early and late sequence changes in human evolution. We identified favored and hitchhiking mutations in 17 human populations (Tang et al., 2022; Tang et al., 2023), classified DBSs into six classes (strong old, strong young, strong other, weak old, weak young, weak others), and examined the numbers of favored and hitchhiking mutations in each class (Supplementary Table 16). Favored and hitchhiking mutations are most enriched in young weak DBSs, suggesting that these DBSs may have greatly promoted recent human evolution.

2.6. SNPs in DBSs have cis-effects on gene expression

To verify the above genomic and population genetics analysis, we examined SNPs in DBSs and their impact on gene expression using transcriptomic data. Based on the Genotype-Tissue Expression (GTEx) data (GTEx Consortium, 2017), we identified expression quantitative trait loci (eQTLs) in DBSs using the widely adopted criteria of MAF≥0.1 and cis-effect size (ES)≥0.5. 1727 SNPs, with MAF ≥0.1 in at least one population and absolute values of ES≥0.5 in at least one tissue, were identified in DBSs in autosomal genes (Supplementary Table 12,13). These eQTLs include 372 "conserved" ones (i.e., also in DBSs in the three archaic humans) and 1020 "novel" ones (i.e., only in DBSs in modern humans). Many conserved eQTLs are in brain tissues and have high derived allele frequencies (DAFs) in all three modern populations. A notable eQTL that has high DAFs in all three populations and has positive ES≥0.5 in 44 out of 48 GTEx tissues is rs2246577 in the DBS in *FN3KRP,* whose encoded protein deglycates proteins to restore their function (Table 2). In contrast, many novel eQTLs are in a particular tissue and have population-specific DAFs (Supplementary Note 6).

Next, we performed two analyses to examine how eQTLs in DBSs influence gene expression via lncRNA-DBS interactions. First, we computed the expression correlation between HS lncRNAs and target transcripts with DBSs having eQTLs in specific tissues. Most (94%) HS lncRNA-target transcript pairs showed significant expression correlation in tissues in which the eQTLs were identified (|Spearman's rho| >0.3 and FDR <0.05). Second, we examined whether eQTLs are more enriched in DBSs than in promoter regions by computing and comparing the density of eQTLs in each DBS and each Ensembl-annotated promoter. Using promoters as the control for the comparison is reasonable as they contain and also are longer than DBSs. The results indicate that

eQTLs are more enriched in DBSs than in promoters (one-sided Mann-Whitney test, p = 0.0) and that these eQTLs influence gene expression via lncRNA-DBS interactions (Supplementary Note 6). Thus, tissue-specific eQTLs and population-specific SNPs in DBSs support the genomic and population genetics analysis.

2.7. HS lncRNAs promote brain evolution from archaic to modern humans

To further confirm that SNPs in DBSs have cis-effects on gene expression, we examined the impact of HS lncRNA-DBS on gene expression in the GTEx tissues (GTEx Consortium, 2017). 40 autosomal HS lncRNAs are expressed in at least one tissue (median TPM>0.1) and form 198876 HS lncRNA-DBS pairs with target transcripts in all tissues. Of these pairs, 45% show a significant expression correlation in specific tissues (Spearman's |rho|>0.3 and FDR<0.05). To check how likely such correlation could be obtained by chance, we randomly sampled 10000 pairs of lncRNAs and protein-coding transcripts genome-wide and found that only 2.3% of pairs show significant expression correlation (Spearman's |rho|>0.3 and FDR<0.05). Moreover, a higher percentage (56%) of HS lncRNA-target transcript pairs with significant correlation is detected in at least one brain region, indicating more extensive gene expression regulation by HS lncRNAs in the brain (Figure 3A).

To obtain detailed supporting evidence, we analyzed nine experimental datasets (Supplementary Note 7). **First**, we analyzed two datasets of epigenetic studies, one examining H3K27ac and H3K4me2 profiling of human, macaque, and mouse corticogenesis, and the other examining gene expression and H3K27ac modification in eight brain regions in humans and four other primates (Reilly et al., 2015; Xu et al., 2018). Compared with the genome-wide background, 84% and 73% of genes in the two datasets have DBSs for HS lncRNAs, indicating significantly higher enrichment in HS lncRNA targets (p = 1.21e-21 and 1.2e-56, two-sided Fisher's exact test). When taking gene expression in chimpanzees as the control, 1851 genes show human-specific transcriptomic differences in one or multiple brain regions, but only 240 genes show chimpanzee-specific transcriptomic differences. **Second**, we analyzed two datasets of PsychENCODE studies, one examining spatiotemporally differentially expressed genes and spatiotemporally differentially methylated sites in 16 brain regions, and the other examining the spatiotemporal transcriptomic divergence across human and macaque brain development (Li et al., 2018; Zhu et al., 2018). In the

first dataset, 65 HS lncRNAs are expressed in all 16 brain regions, 109 transcripts have spatiotemporally differentially methylated sites in their promoters, and 56 of the 109 transcripts have DBSs for HS lncRNAs. In the second dataset, 8951 genes show differential expression between the human and macaque brains, and 72% of differentially expressed genes in the human brain have DBSs of HS lncRNAs. Thus, both datasets show significant enrichment for HS lncRNA regulation. **Third**, three studies identified genes critically regulating cortical expansion (Florio et al., 2018; Johnson et al., 2018; Suzuki et al., 2018). Of the 40 reported protein-coding genes, 29 have DBSs of HS lncRNAs in promoter regions (Supplementary Table 14). Thus, these genes are enriched with DBSs of HS lncRNAs compared with the genome-wide background ($p<0.01$, two-sided Fisher's exact test). **Finally**, we analyzed two datasets of brain organoids studies. By establishing and comparing cerebral organoids between humans, chimpanzees, and macaques, Pollen et al. identified 261 human-specific gene expression changes (Pollen et al., 2019). By fusing human and chimpanzee iPS cells and differentiating the hybrid iPS cells into hybrid cortical spheroids, Agoglia et al. generated a panel of tetraploid human-chimpanzee hybrid iPS cells and identified thousands of genes with divergent expression between humans and chimpanzees (Agoglia et al., 2021). We found that the 261 and 1102 genes in the two datasets are enriched with DBSs of HS lncRNAs compared with the genome-wide background ($p=1.2e-16$ and $3.4e-74$).

Finally, we next examined the evolution of the impact of HS lncRNAs and their DBSs on gene expression using the GTEx data. For each DBS in a pair of HS lncRNA-target transcript that show correlated expression in a GTEx tissue, we computed the sequence distances of this DBS from modern humans to archaic humans. Then, we compared the distribution of DBS sequence distances in each tissue with the distribution of DBS sequence distances in all tissues (the control) (one-sided two-sample Kolmogorov-Smirnov test). DBSs in HS lncRNA-target transcript pairs correlated in brain regions have significantly changed sequence distances since the Altai Neanderthals and Denisovans (Figure 3B), suggesting that gene expression regulation by HS lncRNAs in the brain has experienced significant evolution and that HS lncRNAs contributed greatly to recent brain evolution (Ponce de Leon et al., 2021). Results from other GTEx tissues, including Kidney-Cortex, Ovary, and Skin-Sun Expressed (Figure 3B), also suggest continuous influences of rewired gene expression on human evolution.

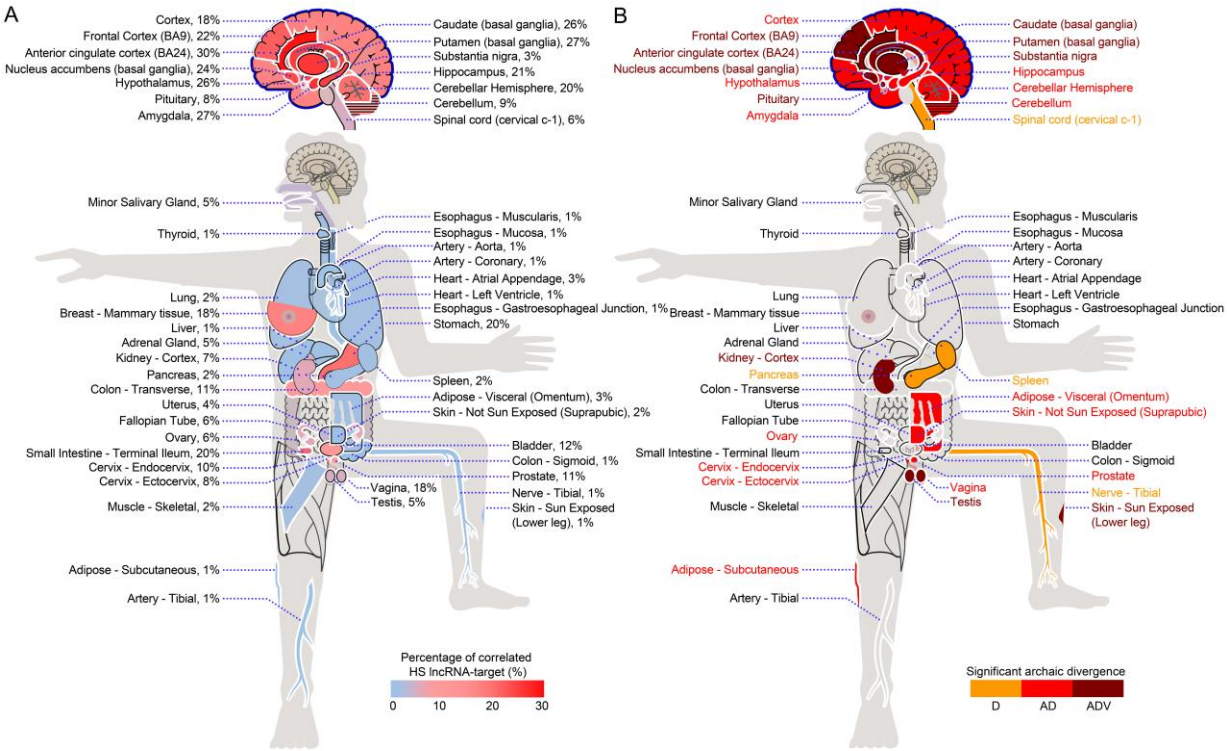

Figure 3. The impact of HS lncRNA-DBS interaction on gene expression in GTEx tissues and organs. (A) The percentage distribution of correlated HS lncRNA-target transcript pairs across GTEx tissues and organs. Higher percentages of correlated pairs are in the brain than in other tissues and organs. (B) The distribution of significantly changed DBSs in HS lncRNA-target transcript pairs across GTEx tissues and organs between archaic and modern humans. Orange, red, and dark red indicate significant DBS changes from Denisovans (D), Altai Neanderthals and Denisovans (AD), and all of the three archaic humans (ADV). DBSs in HS lncRNA-target transcript pairs correlated in seven brain regions (in dark red) have changed significantly and consistently since the Altai Neanderthals, Denisovans, and Vindija Neanderthals (one-sided two-sample Kolmogorov-Smirnov test, significant changes determined by FDR <0.001).

2.8. HS lncRNAs mediate human-specific correlated gene expression in the brain

Species-specificity of expression correlation highlights species-specific transcriptional regulation. To further explore the evidence in this aspect, we used the GTEx data from the human frontal cortex (BA9) and anterior cingulate cortex (BA24) (n=101 and 83, respectively) (GTEx Consortium, 2017), the transcriptomic data for the same brain regions in macaques (n=22 and 25, respectively) (Zhu et al., 2018), and the *eGRAM* program we developed recently, to identify co-expressed gene modules.

HS lncRNAs' target genes form distinct modules in the human brain, in which genes have rich correlation relationships (Pearson's r>0.8) and are enriched in neurodevelopment-related KEGG pathways (hypergeometric distribution test, FDR<0.05). In contrast, their orthologues in the same macaque brain regions do not have these features (Figure 4; Supplementary Figure 24; Supplementary Table 17). The distinct differences in gene modules in these brain regions further support that gene expression has been substantially rewired by HS lncRNAs in the human brain.

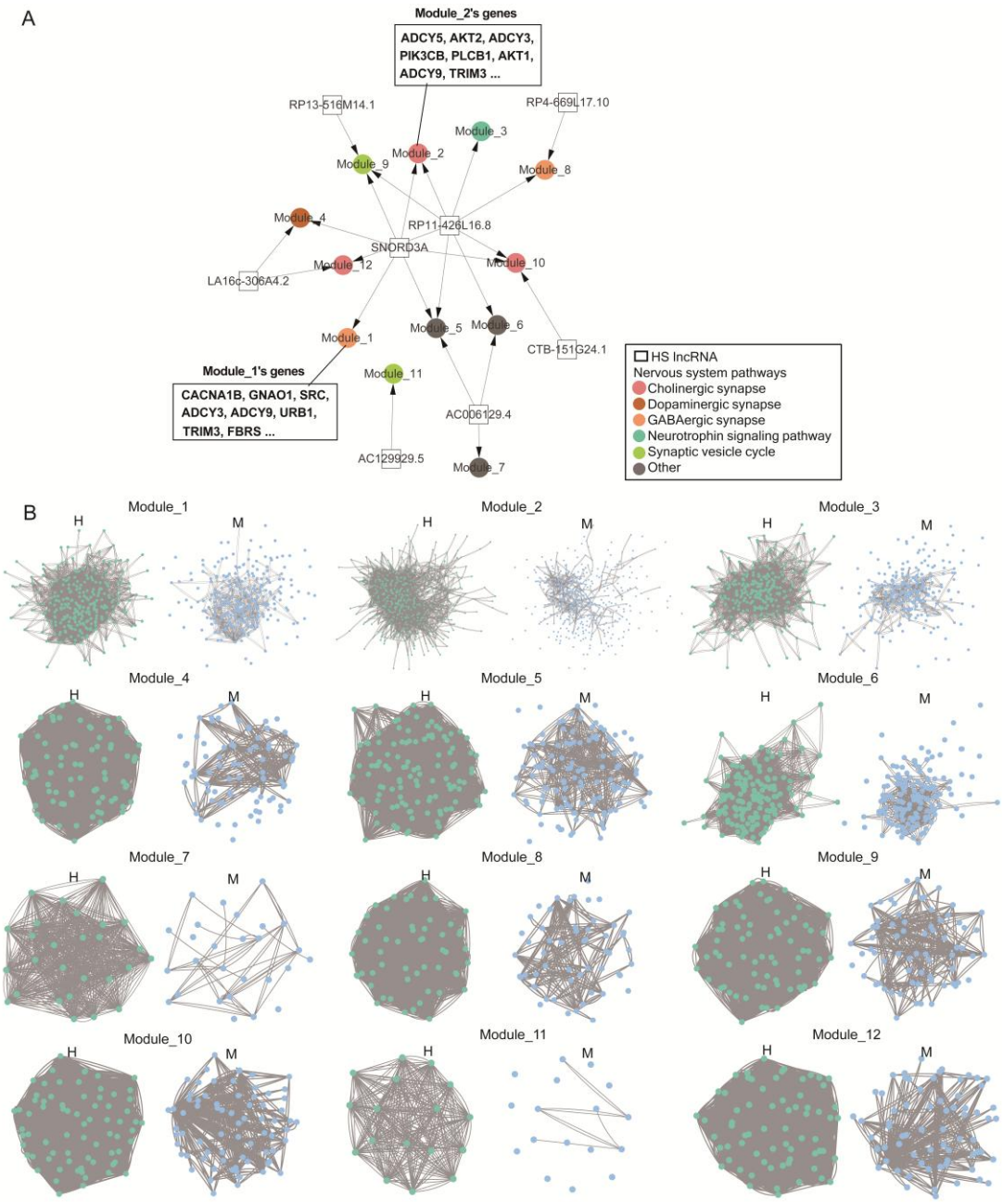

Figure 4. Human-specifically rewired gene expression by HS lncRNAs in the frontal cortex (BA9). (A)

Genes expressed in the frontal cortex are enriched for HS lncRNAs' target genes and neurodevelopment-related pathways. Squares indicate HS lncRNAs; dots indicate gene modules (Module_1 and Module_2 are illustrated); colors indicate enriched KEGG pathways. (B) Comparison of modules of the same genes in humans and macaques (indicated by H and M). In each pair of modules, green dots, and blue dots denote human genes and their orthologues in the corresponding positions, and lines indicate correlation. Correlated expression is more prominent in humans than in macaques.

3. DISCUSSION

Modern humans differ from archaic humans and great apes trivially in genomes but greatly in phenotypes, making what genomic differences critically determine modern humans a profound question. This question has been addressed by many researchers along multiple approaches. Previous attention has been paid to new protein-coding genes, HARs, and TF-TFBS interactions, and findings include that HS protein-coding genes promote rapid neocortex expansion (Fiddes et al., 2018; Florio et al., 2018; Pinson et al., 2022), that HARs are significantly enriched in domains important for human-specific 3D genome organization (Keough et al., 2023), and that gene expression is rewired by TFs and TFBSs (Johnson, 2017). These findings make us postulate that HS lncRNAs have greatly rewired gene expression. To verify this postulation, this study identified HS lncRNAs, their DBSs, and these DBSs' counterparts in ancient humans and multiple apes. Our analyses of these lncRNAs, DBSs, and target genes, including their evolution and interaction, indicate that HS lncRNAs have greatly promoted human evolution by distinctly rewiring gene expression. This conclusion is supported by our analyses of multiple experimental datasets. The rewiring has greatly evolved from humans' ancestors to archaic humans and modern humans and still influences modern human populations.

As DBS prediction is the basis of the study, we performed CRISPR/Cas9 experiments to validate predicted DBSs. When running *Infernal*, *LongTarget, MAFFT7*, *Phylip*, and *eGRAM*, default parameters that suit most situations were used. When analyzing gene sets, the whole genome or the chimpanzee genome were used as controls. When identifying HS lncRNAs' target genes that probably critically characterize human evolution, the top 20% of genes having the largest DBS

distances from humans to chimpanzees were used because this percentage of genes may well characterize the human-chimpanzee differences. Accordingly, the DBS distance of 0.034 was used as the threshold to identify genes that may well characterize the differences between modern and archaic humans. The numbers of critical genes agree with the phylogenetic relationships between modern humans, archaic humans, and chimpanzees. Importantly, we differentiated strong/weak DBSs and old/young DBSs. We found that different classes of DBSs have different evolutionary and population genetics features and genes with these DBSs are enriched in distinct functions. Other values of parameters and thresholds may generate slightly different results, but the core of the results and the main conclusion probably still hold. Finally, not all detected signals reliably indicate positive selection.

Several findings are noteworthy. First, Neanderthal-inherited sequences have measurable impacts on gene expression in modern humans, and the impacts are least detectable in the brain (McCoy et al., 2017). In comparison, we find that HS lncRNAs, which may have evolved quite recently, have more significantly influenced gene expression in the brain than in other tissues. Second, the regulation of many genes by HS lncRNAs has undergone great evolution, notable examples including those making humans adapt to high sugar intake. Third, intensively regulated genes are enriched in different functions at different periods of human evolution. Specifically, "old" SNPs in DBSs are in genes regulating neural development, and "new" SNPs in DBSs are in genes regulating such processes as glucose metabolism. Fourth, recently emerged regulations (i.e., target genes with young weak DBSs) have significantly promoted recent human evolution. These findings reveal diverse, continuous, and subtle contributions of HS lncRNA to human evolution.

The above findings also raise new questions. First, the protein-coding regions in the mouse and human genomes are about 85% identical; however, many mouse lncRNAs are rodent-specific (Yue et al., 2014). Thus, two questions are how mouse-specific lncRNAs specifically rewire gene expression in mice and how human- and mouse-specific rewiring influences the cross-species transcriptional differences (Breschi et al., 2017; Hodge et al., 2019; Zhu et al., 2018). Second, whether mice and other mammals (especially model animals) have species-specific lncRNAs such as RP11-423H2.3 that substantially rewires gene expression species-specifically. Third, whether

and how clade- and species-specific lncRNAs would make evolution preordained. These questions are theoretically interesting and important for translational medicine.

4. MATERIAL AND METHODS

4.1. Data sources

The sequences of human (GRCh38/hg38, GRCh37/hg19) and chimpanzee genomes (panTro5) were obtained from the UCSC Genome Browser (http://genome.UCSC.edu). Three high-quality archaic human genomes were obtained from the Max Planck Institute for Evolutionary Anthropology (https://www.eva.mpg.de/genetics/index.html), which include an Altai Neanderthal which lived approximately 122 thousand years ago (kya) (Prüfer et al., 2013), a Denisovan (an Asian relative of Neanderthals) which lived approximately 72 kya (Meyer et al., 2012), and a Vindija Neanderthal which lived approximately 52 kya (Prufer et al., 2017). The ancestral sequences for the human genome (GRCh37) (which were generated upon the EPO alignments of six primate species human, chimpanzee, gorilla, orangutan, macaque, and marmoset) were downloaded from the EBI website (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/). The SNP data of modern humans were obtained from the 1000 Genomes Project (phase I) (1000 Genomes Project Consortium et al., 2012). The three modern human populations are CEU (Utah residents with Northern and Western European ancestry), CHB (Han Chinese in Beijing, China), and YRI (Yoruba in Ibadan, Nigeria), which contain 100, 99, and 97 individuals, respectively. The genetic map files of modern humans were downloaded from the 1000 Genomes Project website (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20110106_recombination_hotspots/). The 1000 Genomes phased genotype data in VCF format were downloaded from the International Genome Sample Resource website (http://www.internationalgenome.org/). Annotated human genes and transcripts were obtained from the Ensembl website (http://www.Ensembl.org). FASTA-format nucleotide sequences were converted from VCFs using the *vcf-consensus* program in the *VCFtools* package when necessary (Danecek et al., 2011). Data in the GeneCards Human Gene Database (www.genecards.org) were used to annotate genes and transcripts.

RNA-sequencing (RNA-seq) data (transcripts per million [TPM]) and eQTL data of human tissues

were obtained from the Genotype-Tissue Expression Project (GTEx, v8) website (https://gtexportal.org/) (GTEx Consortium, 2017). Histone modification and DNA methylation signals in multiple cell lines and ENCODE Candidate Cis-Regulatory Elements (cCRE) were obtained from the UCSC Genome Browser (http://genome.UCSC.edu). The genome-wide DNA binding sites of lncRNA NEAT1, MALAT1, and MEG3 were obtained from experimental studies (Mondal et al., 2015; West et al., 2014). The predicted and experimentally detected DBSs in target genes of MALAT1, NEAT1, and MEG3 are given in supplementary tables (Supplementary Table 15).

4.2. Identifying HS lncRNA genes

We used the *Infernal* program (Nawrocki et al., 2009), which searches orthologous RNA sequences upon sequence and structure alignment, to identify orthologous exons in 16 mammalian genomes for each exon in each of the 13562 GENCODE (v18)-annotated human lncRNA genes (Lin et al., 2019). The 16 mammals were chimpanzee (CSAC 2.1.4/panTro4), macaque (BGI CR_1.0/rheMac3), marmoset (WUGSC 3.2/calJac3), tarsier (Broad/tarSyr1), mouse lemur (Broad/micMur1), tree shrew (Broad/tupBel1), mouse (GRCm38/mm10), rat (Baylor3.4/rn4, RGSC6.0/rn6), guinea pig (Broad/cavPor3), rabbit (Broad/oryCun2), dog (Broad CanFam3.1/canFam3), cow (Baylor Btau_4.6.1/bosTau7), elephant(Broad/loxAfr3), hedgehog (EriEur2.0/eriEur2), opossum (Broad/monDom5), and platypus (WUGSC 5.0.1/ornAna1) (http://genome.UCSC.edu). If the number of orthologous exons of a human lncRNA gene in a genome exceeded half the exon number of the human lncRNA gene, these orthologous exons were assumed to form an orthologous lncRNA gene. If a human lncRNA gene had no orthologous gene in all of the 16 mammals, it was assumed to be human-specific.

4.3. Identifying DBSs of HS lncRNAs

LncRNAs bind to DNA sequences by forming RNA:DNA triplexes. Each triplex comprises triplex-forming oligonucleotides (TFO) in the lncRNA and a triplex-targeting site (TTS) in the DNA sequence. We used the *LongTarget* program to predict HS lncRNAs' DNA binding domains (DBD) and binding sites (DBS) with the default parameters (*Ruleset* = all, *T.T. penalty* = -1000, *CC penalty* = 0, *Offset* = 15, *Identity* ≥ 60, *Nt* ≥ 50) (Lin et al., 2019; Wen et al., 2022). *LongTarget* program predicts DBDs and DBSs simultaneously, with a DBD comprising a set of densely overlapping TFOs and a DBS

comprising a set of densely overlapping TTSs. For each HS lncRNA, we predicted its DBSs in the 5000 bp promoter regions (-3500 bp upstream and +1500 bp downstream the transcription start site) of the 179128 Ensembl-annotated transcripts (release 79). For each DBS, a binding affinity was calculated by multiplexing the DBS length and the averaged *Identity* score of all TTSs (the *Identity* score is the percentage of paired nucleotides). Strong and weak DBSs were classified upon binding affinity ≥ 60 and < 60. A transcript whose promoter region contains a strong DBS of an HS lncRNA was assumed to be a target transcript of the HS lncRNA, and the gene containing this transcript was assumed to be a target gene of the HS lncRNA. As the 1000 Genomes Project (phase I) data and the archaic human genomes are based on GRCh37/hg19, the coordinates of DBSs were converted from GRCh38/hg38 to GRCh37/hg19 using the *liftover* program from the UCSC Genome Browser (Kuhn et al., 2013).

4.4. Experimentally validating DBS prediction

A 157 bp sequence (chr17:80252565-80252721, hg19) containing the DBD of RP13-516M14.1, a 202 bp sequence (chr1:113392603-113392804, hg19) containing the DBD of RP11-426L16.8, and a 198 bp sequence (chr17:19460524-19460721, hg19) containing the DBD of SNORA59B, were knocked out in the HeLa cell line, RKO cell line, and SK-MES-1 cell line, respectively. Two sequences (chr1:156643524-156643684, chr10:52445649-52445740, hg38) containing the DBD of two wrongly transcribed noncoding sequences were knocked out in the HCT-116 and A549 cell lines, respectively. The seven knockouts were performed by UBIGENE, Guangzhou, China (http://www.ubigene.com) using CRISPR-U™, a revised version of CRISPR/Cas9 technology. Before and after the seven DBD knockouts, RNA-sequencing (RNA-seq) was performed by Novogene, Beijing, China (https://cn.novogene.com) and HaploX, Shenzhen, China (https://www.haplox.cn/). The reads were aligned to the human GRCh38 genome using the *Hiasat2* program (Kim et al., 2019), and the resulting Sam files were converted to bam files using *Samtools* (Li et al., 2009). The *Stringtie* program was used to quantify gene expression levels (Pertea et al., 2015). Fold change of gene expression was computed using the *edgeR* package (Robinson et al., 2010), and significant up- and down-regulation of target genes after DBD knockout was determined upon |log2(fold change)| > 1 with FDR < 0.1.

Genome-wide DBSs of NEAT1, MALAT1, and MEG3 were experimentally detected (Mondal et al., 2015; West et al., 2014). We also used these data to validate DBS prediction by predicting DBSs of the three lncRNAs and checking the overlap between predicted and experimentally detected DBSs (Supplementary Table 15).

4.5. Mapping DBSs of HS lncRNAs in the chimpanzee and archaic human genomes

We used the *liftover* program from the UCSC Genome Browser (http://genome.UCSC.edu) to map loci of DBSs in the human genome hg38 to the chimpanzee genome Pan_tro 3.0 (panTro5). The mapping results were examined by checking the human-chimpanzee pairwise alignment (hg38 vs panTro5) in the UCSC Genome Browser (http://genome.UCSC.edu). 97.81% (100403/102651) of DBSs were mapped, and the 2248 unmapped DBSs in 429 genes were considered human-specific. We used *vcf-consensus* in the *VCFtool*s package to extract the DBSs of HS lncRNAs from the VCF files of Altai Neanderthals, Denisovans, and Vindija Neanderthals. The variant with the highest quality score was selected whenever multiple variant calls were observed at a given locus. The obtained DBS sequences in the chimpanzees and three archaic humans are called counterparts of DBSs in these genomes.

4.6. Estimating sequence distances of DBSs between different genomes

We first aligned DBS sequences in the genomes of humans, chimpanzees, Altai Neanderthals, Denisovans, and Vindija Neanderthals using the *MAFFT*7 program for measuring sequence distances from modern humans to chimpanzees and archaic humans (Katoh and Standley, 2013). We then computed sequence distances using the *dnadist* program with the Kimura 2-parameter model in the *PHYLIP* (3.6) package (http://evolution.genetics.washington.edu/phylip.html) and using the *Tamura-Nei model* in the *MEGA7* package (Kumar et al., 2016). The two methods generated equivalent results. The largest distance between DBSs in humans and their counterparts in chimpanzees is 5.383. Since 2248 DBSs in 429 human genes do not have counterparts in chimpanzees, we assumed that the sequence distances of these DBSs between humans and chimpanzees were 10.0. Since the top 20% of DBSs in chimpanzees (including these 2248 DBSs) have distances ≥ 0.034, we used 0.034 as the threshold to identify the "mostly changed" DBSs in the chimpanzees and three archaic humans.

We determined human ancestral sequences of DBSs upon the human ancestor sequences from the EBI website, which were generated upon the EPO alignments of six primate species. We used the above-mentioned methods to compute DBSs' sequence distances from the human ancestor to chimpanzees, archaic humans, and modern humans. We found that when the human-chimpanzee ancestral sequence has the ancestral sequence (which means the inferring of ancestral allele is of high confidence), DBS distances from the human ancestor to modern humans are larger than to archaic humans, but this situation only accounts for about 63.8%. For many DBSs, the distances from the human ancestor to modern humans are smaller than to archaic humans (especially Neanderthals and Denisovans) and even to chimpanzees. This defect may be caused by the absence of archaic humans in building the human ancestral sequence.

4.7. Detecting positive selection signals in HS lncRNA genes and DBSs

We used multiple tests to detect positive selection signals in HS lncRNA genes. First, we used the XP-CLR test (parameters = -w1 0.001 300 100 -p0 0.95, window size = 0.1 cM, grid size = 100 bp) to perform six pairwise genome-wide scans (i.e., CEU-CHB, CEU-YRI, CHB-CEU, CHB-YRI, YRI-CEU, and YRI-CHB) (Chen et al., 2010). The upper 1% of scores across the entire genome in each pairwise scan were 34.6 in the CEU-YRI scan, 16.8 in the CEU-CHB scan, 45.0 in the CHB-YRI scan, 26.9 in the CHB-CEU scan, 14.1 in the YRI-CEU scan, and 14.1 in the YRI-CHB scan. These scores were used as the thresholds of positive selection signals in these populations. Second, we used the *iSAFE* program to scan each genomic region that contains a HS lncRNA gene and its 500-kb upstream and downstream sequences (Akbari et al., 2018). Strongly selected loci were detected only in CEU and CHB. Third, we used the *VCFtools* program to calculate Tajima's D values for each HS lncRNA gene in CEU, CHB, and YRI (Danecek et al., 2011). The calculation was performed using a 1500-bp non-overlapping sliding window because the lengths of these genes exceed 1500 bp. To generate the background reference for judging the significant increase or decrease of Tajima's D for HS lncRNA genes in a population, we calculated Tajima's D values across the whole genome using a sliding window of 1500 bp. As values of Tajima's D were compared with the background reference, a significant D<0 and D>0 indicate positive (or directional) selection and balancing selection, respectively, instead of population demography dynamics (Tajima, 1989). Fourth, we used the

integrated Fst to detect positive selection signals in HS lncRNA genes. The Fst of each HS lncRNA gene was computed for three comparisons, i.e., CEU-YRI, CHB-YRI, and CHB-CEU. Extreme Fst values of SNPs were detected in HS lncRNA genes in the comparisons of CEU-YRI and CHB-YRI. Since allele frequencies for different loci vary in a genome and genetic drift may have different effects at different loci, we used a sliding window of 1500 for comparing Fst values of HS lncRNA genes with the genome-wide background. Extreme Fst values indicate positive selection. Finally, we applied linkage disequilibrium (LD) analysis to each HS lncRNA gene. We computed the pairwise LD (r^2) in CEU, CHB, and YRI for common SNPs (with minimal allele frequency (MAF) ≥ 0.05 in at least one population) in HS lncRNA genes and DBSs using the *PLINK* program (Purcell et al., 2007). Significantly increased LD was detected in SNPs in HS lncRNA genes in CEU and CHB. The LD patterns were portrayed using the *Haploview* program (Barrett et al., 2005).

Next, we used the above tests to detect positive selection signals in DBSs. First, the 100 bp grid size of the XP-CLR test also allowed the detection of selection signals in DBSs. Second, we performed the Tajima's D and Fay-Wu's H tests (Fay and Wu, 2000; Tajima, 1989). We calculated Tajima's D values for each DBS in CEU, CHB, and YRI using the *VCFtools* program (Danecek et al., 2011), with a sliding window of 147 bp (the mean length of strong DBSs). To generate the background reference for judging the significant increase or decrease of Tajima's D in a population, we calculated Tajima's D values across the whole genome using the same sliding window. As values of Tajima's D were compared with the background reference, a significant D<0 and D>0 indicate positive (or directional) selection and balancing selection, respectively. Fay-Wu's H values were calculated similarly using the *VariScan* program (Vilella et al., 2005). Calculating Fay-Wu's H demands the ancestral sequences as the outgroup. We extracted the ancestral sequences of DBSs from the human ancestral sequence, which was generated upon the EPO alignments of six primate species (human, chimpanzee, gorilla, orangutan, macaque, and marmoset) (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/). Third, we computed the Fst to measure the frequency differences of alleles in DBSs between populations. For the CEU-YRI, CHB-YRI, and CHB-CEU pairwise comparisons, we used the revised *VCFtools* program to compute the weighted Fst for all SNPs in each DBS (Weir and Cockerham, 1984). Fourth, we integrated the weighted Fst values in the three populations into an "integrated Fst"

which indicated whether the DBS locus was under selection in a certain population (Nielsen and Slatkin, 2013). We used sliding windows of 147 bp for comparing Fst values of DBSs with the genome-wide background. We empirically denoted the values of integrated Fst in the upper 10% of scores across the entire genome as statistically significant. To determine positive selection more reliably, we used Tajima's D and the integrated Fst to jointly determine if a DBS was under positive selection in a population. The thresholds that determined the upper 10% of Tajima's D values across the entire genome in CEU, CHB, and YRI were -0.97, -0.96, and -0.97, respectively, and the threshold that determined the upper 10% of integrated Fst values across the entire genome in the three populations was 0.22. For example, a DBS was assumed to be under positive selection in CEU if (a) the DBS had a Tajima's D<-0.97 in CEU and Tajima's D>0.0 in the two other populations and (b) the DBS had an integrated Fst>0.22. Analogously, a DBS was assumed to be under positive selection in both CEU and CHB if the DBS had a Tajima's D <-0.97 in CEU, <-0.96 in CHB, and >0.0 in YRI, and had an integrated Fst>0.22.

4.8. Functional enrichment analysis of genes

We used the *g:Profiler* program (with the parameter settings: Organism=Homo sapiens, Ordered query = No, Significance threshold = Benjamini-Hochberg FDR, User threshold = 0.05, 50 < terms size < 1000) and the Gene Ontology (GO) database to perform over-representation analysis (Raudvere et al., 2019). This analysis is used to determine what pre-defined gene sets (GO terms) are more present (over-represented) in a list of "interesting" genes than what would be expected by chance. The lists of genes included genes with strongest DBSs in humans, genes with weakest DBSs in humans, and genes with mostly changed DBSs in chimpanzees and three archaic humans. To choose genes with strong DBSs, we ranked genes in descending order based on binding affinity and chose the top 2000 genes. To choose genes with weak DBSs, we first ranked genes in ascending order based on binding affinity, then ranked genes in descending order based on the number of regulatory HS lncRNAs, and finally chose the top 2000 genes with binding affinity ≤40 and regulatory HS lncRNAs ≥5. We used the distance threshold of 0.034 to choose genes with mostly changed DBSs. Genes with mostly changed DBSs in chimpanzees include the 429 genes in which 2248 DBSs lack counterparts. 4248, 1256, 2514, and 134 genes in chimpanzees, Altai Neanderthals, Denisovans, and Vindija Neanderthals have DBSs with distance ≥0.034. To make the statistical

significance comparable across chimpanzees, Altai Neanderthals, and Denisovans, we used the same number of genes (the top 1256 genes whose DBSs have the largest DBS distances) as the input to the *g:Profiler* program.

4.9. Analysing SNP frequencies in human populations

The frequency of common SNPs (MAF ≥0.05) in DBSs in the three modern human populations was computed using the *VCFtools* package (Danecek et al., 2011). The ancestral/derived states of SNPs were inferred upon the human ancestor sequences and were used to determine derived allele frequency.

4.10. Analysing the cis-effect of SNPs in DBSs on target gene expression

SNPs with MAF >0.1 in DBSs in any of the three modern human populations and absolute values of cis-effect size >0.5 (FDR <0.05) in any of the GTEx tissues were examined for an influence on the expression of the target genes (GTEx Consortium, 2017). SNPs that are eQTLs in the GTEx tissues and have biased derived allele frequencies (DAF) in the three modern human populations were examined to estimate whether the eQTL is population-specific.

4.11. Examining the tissue-specific impact of HS lncRNA-regulated gene expression

First, we examined the expression of HS lncRNA genes across the GTEx tissues. HS lncRNA genes with a median TPM value >0.1 in a tissue were considered robustly expressed in that tissue. Upon this criterion, 40 HS lncRNA genes were expressed in at least one tissue and were used to examine the impact of HS lncRNA regulation on gene and transcript expression (other HS lncRNAs may function in the cytoplasm). Since an HS lncRNA gene may have multiple transcripts, we adopted the transcript, which contains the predicted DNA binding domain (DBD) and has the largest TPM value as the representative transcript of the HS lncRNA. We calculated the pairwise Spearman's correlation coefficient between the expression of an HS lncRNA (the representative transcript) and the expression of each of its target transcripts using the *scipy.stats.spearmanr* program in the *scipy* package. The expression of an HS lncRNA gene and a target transcript was considered to be significantly correlated if the |Spearman's rho| > 0.3, with Benjamini-Hochberg FDR <0.05. We examined the percentage distribution of correlated HS lncRNA-target transcript pairs across GTEx

tissues and organs (Figure 3A).

We examined all GTEx tissues to determine in which tissues HS lncRNA-regulated gene expression may have undergone significant changes from archaic to modern humans. If a pair of HS lncRNA and target transcript are robustly expressed in a tissue and their expression shows a significant correlation (|Spearman's rho| >0.3, with Benjamini-Hochberg FDR <0.05) in the tissue, we computed the sequence distance of the HS lncRNA's DBS in the transcript from the three archaic humans to modern humans. We compared the sequence distances of all DBSs in each tissue with the sequence distances of all DBSs in all GTEx tissues (i.e. the distribution of all DBS distances was taken as the background). A one-sided two-sample Kolmogorov-Smirnov test was used to examine whether the sequence distances of all DBSs in a specific tissue deviate from the background distribution (which reflects the "neutral evolution" of gene expression). For each tissue, if the test result was Benjamini-Hochberg FDR < 0.001, the tissue was considered to have significantly changed gene expression regulated by the HS lncRNA. We used different colors to mark tissues with significantly changed gene expression regulation since Altai Neanderthals, Denisovans, and Vindija Neanderthals, and used "D", "A.D.", and "ADV" to indicate changes since Denisovans, since Altai Neanderthals and Denisovans, and since Altai Neanderthals, Denisovans, and Vindija Neanderthals, respectively (Figure 3B).

4.12. Examining enrichment of favored and hitchhiking mutations in DBSs

Using the deep learning network *DeepFavored* we developed that integrates multiple statistical tests for identifying favored mutations (Tang et al., 2022), we identified 13339 favored mutations and 244098 hitchhiking mutations in 17 human populations (Tang et al., 2023). In this study, we classified DBSs in two ways - into strong ones (affinity>60) and weak ones (36<affinity<60), and into old ones (Human-Chimp distance>0.034 AND Human-Altai Neanderthals distance==0), young ones (Human-Altai Neanderthals distance>0.034 OR Human-Denisovan distance>0.034), and others. We then examined the numbers of favored and hitchhiking mutations in each class of DBSs. The weak young DBSs have the largest proportion of favored and hitchhiking mutations.

4.13. Identifying and analyzing transcriptional regulatory modules

Most clustering algorithms classify genes into disjoint modules based on expression correlation without considering regulatory relationships (Saelens et al., 2018). The *GRAM* program identifies gene regulatory modules based on correlation and TF-TFBS binding (Bar-Joseph et al., 2003). LncRNAs transcriptionally regulate genes based on lncRNA-DBS binding and correlate gene expression. We developed the *eGRAM* program to identify gene modules based on correlated expression, TF-TFBS interaction, and lncRNA-DBS interaction. In this study, we used *eGRAM* to identify gene modules in the same regions in the human and macaque brain and enriched KEGG pathways using reported RNS-seq datasets (n=101 and 83 for frontal cortex and n=22 and 25 for anterior cingulate cortex) (GTEx Consortium, 2017; Zhu et al., 2018). The default parameters of DBS affinity=60, Pearson correlation=0.5, module size=50, and FDR=0.01 (hypergeometric distribution test) were used. The key steps of program are as follows. (a1) Identify each lncRNA's correlated lncRNAs, which may form a set of co-regulators. (a2, optional) Identify each TF's correlated TFs, which may form a set of co-regulators. (b1) Compute the correlation between each lncRNA and all genes. (b2, optional) Compute the correlation between each TF and all genes. (c1) Identify each lncRNA's target genes. (c2, optional) Identify each TF's target genes. (d1) Identify each lncRNA set's target module upon correlation and targeting relationships. (d2, optional) Identify each TF set's target module upon correlation and targeting relationships. (e) Check whether TFs' modules contain lncRNAs' targets and whether lncRNAs' modules contain TFs' targets, which reveal genes co-regulated by TFs and lncRNAs and genes independently regulated by TFs and lncRNAs. (f) Performs pathway enrichment analysis for all modules.

DECLARATION OF COMPETING INTEREST

The authors declare no competing interest.

AUTHOR CONTRIBUTIONS

H.Z. designed the study and drafted the manuscript. J.L. performed most data analyses. Y.W. and J.T. performed XP-CLR and iSAFE analyses. All authors have read the manuscript and consent to its publication.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (31771456 to H.Z.) and the China Postdoctoral Science Foundation (2020M682788 to J.L.).

DATA AND CODE AVAILABILITY

GENCODE human lncRNAs are available at the website (https://www.gencodegenes.org/human/). Human lncRNAs' orthologues in 16 mammals and the *LongTarget* program are available at the *LongTarget*/*LongMan* website (http://www.gaemons.net/). Other programs and tests are open-source and freely available. The RNA-seq data of the three HS lncRNAs before and after the CRISPR/Cas9 experiments have been deposited in the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo) under the accession number GSE213231. The RNA-seq data of the four wrongly transcribed noncoding sequences before and after the CRISPR/Cas9 experiments have been deposited in the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo) (the accession number is to be obtained). All other data are in Supplementary Tables. The data of favored and hitchhiking mutations are available from the PopTradeOff database or from the authors upon request. The *eGRAM* code is available on the GitHub website (https://github.com/LinjieCodes/eGRAMv2R1).

SUPPLEMENTARY DATA

Two supplementary files are available with the manuscript online. One PDF file contains 8 Supplementary Notes, in which there are 24 Supplementary Figures, and one Excel file contains 17 Supplementary Tables.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Abu Almakarem, A.S., Petrov, A.I., Stombaugh, J., Zirbel, C.L., and Leontis, N.B. (2012). Comprehensive survey and geometric classification of base triples in RNA structures. Nucleic Acids Res *40*, 1407-1423.

Agoglia, R.M., Sun, D., Birey, F., Yoon, S.J., Miura, Y., Sabatini, K., Pasca, S.P., and Fraser, H.B. (2021). Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. Nature *592*, 421-427.

Akbari, A., Vitti, J.J., Iranmehr, A., Bakhtiari, M., Sabeti, P.C., Mirarab, S., and Bafna, V. (2018). Identifying the favored mutation in a positive selective sweep. Nat Methods *15*, 279-282.

Albalat, R., and Canestro, C. (2016). Evolution by gene loss. Nat Rev Genet *17*, 379-391.

Anton, S.C., Potts, R., and Aiello, L.C. (2014). Human evolution. Evolution of early Homo: an integrated biological perspective. Science *345*, 1236828.

Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A.*, et al.* (2003). Computational discovery of gene modules and regulatory networks. Nat Biotechnol *21*, 1337-1342.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.

Breschi, A., Gingeras, T.R., and Guigo, R. (2017). Comparative transcriptomics in human and mouse. Nat Rev Genet *18*, 425-440.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res *20*, 393-402.

Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature *437*, 69-87.

Currat, M., Excoffier, L., Maddison, W., Otto, S.P., Ray, N., Whitlock, M.C., and Yeaman, S. (2006). Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". Science *313*, 172; author reply 172.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T.*, et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G.*, et al.* (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res *22*, 1775-1789.

Dong, X., Wang, X., Zhang, F., and Tian, W. (2016). Genome-Wide Identification of Regulatory Sequences Undergoing Accelerated Evolution in the Human Genome. Mol Biol Evol *33*, 2565-2575.

Evans, P.D., Gilbert, S.L., Mekel-Bobrov, N., Vallender, E.J., Anderson, J.R., Vaez-Azizi, L.M., Tishkoff, S.A., Hudson, R.R., and Lahn, B.T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science *309*, 1717-1720.

Fay, J.C., and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. Genetics *155*, 1405-1413.

Fiddes, I.T., Lodewijk, G.A., Mooring, M., Bosworth, C.M., Ewing, A.D., Mantalas, G.L., Novak, A.M., van den Bout, A., Bishara, A., Rosenkrantz, J.L.*, et al.* (2018). Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. Cell *173*, 1356-1369 e1322.

Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., Wimberger, P., Huttner, W.B., and Hiller, M. (2018). Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. Elife *7*.

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204-213.

He, S., Zhang, H., Liu, H., and Zhu, H. (2015). LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via Hoogsteen base-pairing analysis. Bioinformatics *31*, 178-186.

Hernando-Herraez, I., Garcia-Perez, R., Sharp, A.J., and Marques-Bonet, T. (2015). DNA Methylation: Insights into Human Evolution. PLoS Genet *11*, e1005661.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O.*, et al.* (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61-68.

Johnson, A.D. (2017). The rewiring of transcription circuits in evolution. Curr Opin Genet Dev *47*, 121-127.

Johnson, M.B., Sun, X., Kodani, A., Borges-Monroy, R., Girskis, K.M., Ryu, S.C., Wang, P.P., Patel, K., Gonzalez, D.M., Woo, Y.M.*, et al.* (2018). Aspm knockout ferret reveals an evolutionary mechanism governing cerebral cortical size.

Nature *556*, 370-375.
Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res *20*, 1313-1326.
Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772-780.
Keough, K.C., Whalen, S., Inoue, F., Przytycki, P.F., Fair, T., Deng, C., Steyert, M., Ryu, H., Lindblad-Toh, K., Karlsson, E.*, et al.* (2023). Three-dimensional genome rewiring in loci with human accelerated regions. Science *380*, eabm1696.
Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915.
Krieger, G., Lupo, O., Wittkopp, P., and Barkai, N. (2022). Evolution of transcription factor binding through sequence variations and turnover of binding sites. Genome Res *32*, 1099-1111.
Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief Bioinform *14*, 144-161.
Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution *33*, 1870-1874.
Lee, J.T. (2009). Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev *23*, 1831-1842.
Levchenko, A., Kanapin, A., Samsonova, A., and Gainetdinov, R.R. (2018). Human Accelerated Regions and Other Human-Specific Sequence Variations in the Context of Evolution and Their Relevance for Brain Development. Genome Biol Evol *10*, 166-188.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y.*, et al.* (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*.
Lin, J., Wen, Y., He, S., Yang, X., Zhang, H., and Zhu, H. (2019). Pipelines for cross-species and genome-wide prediction of long noncoding RNA binding. Nat Protoc *14*, 795-818.
Liu, J., Mosti, F., and Silver, D.L. (2021). Human brain evolution: Emerging roles for regulatory DNA and RNA. Curr Opin Neurobiol *71*, 170-177.
Mangan, R.J., Alsina, F.C., Mosti, F., Sotelo-Fonseca, J.E., Snellings, D.A., Au, E.H., Carvalho, J., Sathyan, L., Johnson, G.D., Reddy, T.E.*, et al.* (2022). Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. Cell *185*, 4587-4603 e4523.
McCoy, R.C., Wakefield, J., and Akey, J.M. (2017). Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. Cell *168*, 916-+.
McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T.*, et al.* (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature *471*, 216-219.
Mekel-Bobrov, N., Gilbert, S.L., Evans, P.D., Vallender, E.J., Anderson, J.R., Hudson, R.R., Tishkoff, S.A., and Lahn, B.T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science *309*, 1720-1722.
Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C.*, et al.* (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science *338*, 222-226.
Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E.*, et al.* (2015). MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA

triplex structures. Nat Commun *6*, 7743.
Nawrocki, E.P. (2014). Annotating functional RNAs in genomes using Infernal. Methods Mol Biol *1097*, 163-197.
Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.
Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335-1337.
Nielsen, R., and Slatkin, M. (2013). An Introduction to Population Genetics. Sinauer Associates Inc Publishers.
Otto, W., Stadler, P.F., Lopez-Giraldez, F., Townsend, J.P., Lynch, V.J., and Wagner, G.P. (2009). Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. Genome Biol Evol *1*, 85-98.
Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol *33*, 290-295.
Pinson, A., Xing, L., Namba, T., Kalebic, N., Peters, J., Oegema, C.E., Traikov, S., Reppe, K., Riesenberg, S., Maricic, T.*, et al.* (2022). Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. Science *377*, eabl6422.
Pollen, A.A., Bhaduri, A., Andrews, M.G., Nowakowski, T.J., Meyerson, O.S., Mostajo-Radji, M.A., Di Lullo, E., Alvarado, B., Bedolli, M., Dougherty, M.L.*, et al.* (2019). Establishing Cerebral Organoids as Models of Human-Specific Brain Evolution. Cell *176*, 743-756 e717.
Pollen, A.A., Kilik, U., Lowe, C.B., and Camp, J.G. (2023). Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution. Nat Rev Genet, 1-25.
Ponce de Leon, M.S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., Alatorre Warren, J.L., Lordkipanidze, D., Kurniawan, I., Murti, D.B., Suriyanto, R.A.*, et al.* (2021). The primitive brain of early Homo. Science *372*, 165-171.
Prabhakar, S., Noonan, J.P., Paabo, S., and Rubin, E.M. (2006). Accelerated evolution of conserved noncoding sequences in humans. Science *314*, 786.
Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A *104 Suppl 1*, 8605-8612.
Prufer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlevic, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P.S., Peyregne, S.*, et al.* (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655-658.
Prufer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J.R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R.*, et al.* (2012). The bonobo genome compared with the chimpanzee and human genomes. Nature *486*, 527-531.
Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C.*, et al.* (2013). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43-49.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet *81*, 559-575.
Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res *47*, W191-W198.
Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155-1159.
Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J.*, et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
Saelens, W., Cannoodt, R., and Saeys, Y. (2018). A comprehensive evaluation of module detection methods for gene expression data. Nat Commun *9*, 1090.
Sarropoulos, I., Marin, R., Cardoso-Moreira, M., and Kaessmann, H. (2019). Developmental dynamics of lncRNAs across mammalian organs and species. Nature *571*, 510-514.
Slatkin, M. (2008). Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet *9*, 477-485.
Song, J.H.T., Grant, R.L., Behrens, V.C., Kucka, M., Roberts Kingman, G.A., Soltys, V., Chan, Y.F., and Kingsley, D.M. (2021). Genetic studies of human-chimpanzee divergence using stem cell fusions. Proc Natl Acad Sci U S A *118*.
Suzuki, I.K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F.*, et al.* (2018). Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. Cell *173*, 1370-1384 e1316.
Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585-595.
Tang, J., Huang, M., He, S., Zeng, J., and Zhu, H. (2022). Uncovering the extensive trade-off between adaptive evolution and disease susceptibility. Cell Rep *40*, 111351.
Tang, J., Zhang, H., Zhang, H., and Zhu, H. (2023). PopTradeOff: A database for exploring population-specificity of adaptive evolution, disease susceptibility, and drug responsiveness. Comput Struct Biotechnol J *21*, 3443-3451.
Timpson, N., Heron, J., Smith, G.D., and Enard, W. (2007). Comment on papers by Evans et al. and Mekel-Bobrov et al. on Evidence for Positive Selection of MCPH1 and ASPM. Science *317*, 1036; author reply 1036.
Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics *21*, 2791-2793.
Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution *38*, 1358-1370.
Wen, Y., Wu, Y., Xu, B., Lin, J., and Zhu, H. (2022). Fasim-LongTarget enables fast and accurate genome-wide lncRNA/DNA binding prediction. Comput Struct Biotechnol J *20*, 3347-3350.
West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell *55*, 791-802.
Whalen, S., and Pollard, K.S. (2022). Enhancer Function and Evolutionary Roles of Human Accelerated Regions. Annu Rev Genet *56*, 423-439.
Xu, C., Li, Q., Efimova, O., He, L., Tatsumoto, S., Stepanova, V., Oishi, T., Udono, T., Yamaguchi, K., Shigenobu, S.*, et al.* (2018). Human-specific features of spatial gene expression and regulation in eight brain regions. Genome Res *28*, 1097-1110.
Yu, F., Hill, R.S., Schaffner, S.F., Sabeti, P.C., Wang, E.T., Mignault, A.A., Ferland, R.J., Moyzis, R.K., Walsh, C.A., and Reich, D. (2007). Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens". Science *316*, 370.
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D.*, et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.
Zhang, X., Fang, B., and Huang, Y.F. (2023). Transcription factor binding sites are frequently under accelerated evolution in primates. Nat Commun *14*, 783.
Zhu, Y., Sousa, A.M.M., Gao, T., Skarica, M., Li, M., Santpere, G., Esteller-Cucala, P., Juan, D., Ferrandez-Peral, L., Gulden, F.O.*, et al.* (2018). Spatiotemporal transcriptomic divergence across human and macaque brain development. Science *362*.