

Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation

Martin Silvert, Lluis Quintana-Murci, Maxime Rotival

▶ **To cite this version:**

Martin Silvert, Lluis Quintana-Murci, Maxime Rotival. Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation. American Journal of Human Genetics, 2019, 104 (6), pp.1241-1250. 10.1016/j.ajhg.2019.04.016 . pasteur-02549232

HAL Id: pasteur-02549232

https://pasteur.hal.science/pasteur-02549232

Submitted on 13 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Manuscript format:

Report

Title:

Impact and evolutionary determinants of Neanderthal introgression on transcriptional and post-transcriptional regulation

Authors:

Martin Silvert[1,2], Lluis Quintana-Murci[1,3] and Maxime Rotival[1,3]

Affiliations:

[1]Human Evolutionary Genetics Unit, Institut Pasteur, CNRS UMR 2000, 75015 Paris, France;

[2]Sorbonne Universités, École Doctorale Complexité du Vivant, 75005 Paris, France.

[3]These authors contributed equally to this work

*Correspondence: quintana@pasteur.fr (L.Q.-M.), maxime.rotival@pasteur.fr (M.R.)

Abstract

Archaic admixture is increasingly recognized as an important source of diversity in modern humans, with Neanderthal haplotypes covering 1-3% of the genome of present-day Eurasians. Recent work has shown that archaic introgression has contributed to human phenotypic diversity, mostly through the regulation of gene expression. Yet, the mechanisms through which archaic variants alter gene expression, and the forces driving the introgression landscape at regulatory regions remain elusive. Here, we explored the impact of archaic introgression on transcriptional and post-transcriptional regulation, focusing on promoters and enhancers across 127 different tissues as well as microRNA-mediated regulation. Although miRNAs themselves harbor few archaic variants, we found that some of these variants may have a strong impact on miRNA-mediated gene regulation. Enhancers were by far the regulatory elements most affected by archaic introgression, with up to one third of the tissues tested presenting significant enrichments. Specifically, we found strong enrichments of archaic variants in adipose-related tissues and primary T cells, even after accounting for various genomic and evolutionary confounders such as recombination rate and background selection. Interestingly, we identified signatures of adaptive introgression at enhancers of some key regulators of adipogenesis, raising the interesting hypothesis of a possible adaptation of early Eurasians to colder climates. Collectively, this study sheds new light onto the mechanisms through which archaic admixture have impacted gene regulation in Eurasians and, more generally, increases our understanding of the contribution of Neanderthals to the regulation of acquired immunity and adipose homeostasis in modern humans.

Main text

The sequencing of the genomes of extinct human forms, such as Neanderthals or Denisovans, has enabled the mapping of archaic variants in the genomes of modern humans.[1-7] This archaic introgression has functional consequences today, as introgressed variants have been reported to alter a variety of phenotypes, ranging from skin pigmentation to sleeping patterns and mood disorders.[8,9] Furthermore, several studies have shown that Neanderthal haplotypes are enriched in regulatory variants, with respect to non-archaic haplotypes,[10,11] suggesting that archaic introgression may impact complex, organismal phenotypes through changes in gene expression. Indeed, up to a quarter of Neanderthal-introgressed haplotypes have been estimated to present *cis*-regulatory effects across tissues, with a bias towards down-regulation of Neanderthal alleles in brain and testes.[12] Furthermore, genes involved in innate immunity and interactions with RNA viruses have been reported to be enriched in Neanderthal ancestry,[13,14] with archaic variants affecting, in particular, transcriptional responses to viral challenges.[11,15] A depletion of Neanderthal introgression has recently been documented in conserved coding regions and, surprisingly, promoters,[16] suggesting that archaic introgression could affect gene expression through promoter-independent mechanisms. One such example is found in post-transcriptional regulation by miRNAs, which has been reported to contribute to phenotypic differences between archaic and modern humans.[17,18] Thus, the relative contribution of transcriptional and post-transcriptional mechanisms to the effects of archaic variants on gene expression remain to be determined.

Our understanding of the selective forces that shaped the landscape of archaic introgression is also rapidly growing. In most cases, archaic variants were selected against, with regions of higher selective constraint being depleted in archaic ancestry, and in particular those that are X-linked or contain testis-expressed and meiotic-related genes.[1,2,19] Some studies have also suggested that Neanderthals had a reduced effective population size,[6,7] owing to a prolonged bottleneck or a deeply structured population.[6,20,21] Natural selection in Neanderthals would thus have been less efficient to purge deleterious

mutations,[22,23] a large proportion of which were removed from the genome of modern humans following their admixture with Neanderthals.[16] However, archaic variants have also contributed, in some cases, to human adaptation,[15,24-28] shortly after their introduction into modern humans or after an initial period of genetic drift.[29,30] Given the rapid evolution of regulatory regions and their potential adaptive nature,[31,32] the evolutionary dynamics of Neanderthal introgression at regulatory elements needs to be explored in further detail.

In this study, we aimed to increase knowledge on the impact that archaic introgression has had on transcriptional and post-transcriptional mechanisms, focusing on promoter-, enhancer- and microRNA-mediated regulation.[33,34] To this end, we first characterized the set of variants of putative Neanderthal origin — archaic SNPs (aSNPs) — as those for which one allele is both present in the Neanderthal Altai genome[6] and absent in the Yoruba African population of the 1000 Genomes project[35] (**Supplemental Methods**). We further required aSNPs to be located in genomic regions where Neanderthal introgression has already been detected in Europe or Asia.[1] We then investigated deviations in the presence or absence of aSNPs among specific classes of functional elements, by measuring the density of aSNPs, with respect to that of non-aSNPs, in the European (CEU) and Asian (CHB) populations of the 1000 Genomes project.[35] We then compared the relative density of aSNPs at specific functional regions to that of the rest of the genome. Genomic regions were considered as enriched/depleted in aSNPs if the resulting odds ratio was significantly different from 1.

Overall, we observed a strong depletion of aSNPs in coding regions (OR = 0.71, *p*-value$<10^{-4}$) and similar levels of introgression at regulatory regions compared to those of non-functional elements. We then divided genetic variants according to the frequency of their minor allele, which corresponds to the Neanderthal allele in 99.8% of all aSNPs. Minor allele frequency (MAF) was computed in Eurasian populations combined, and variants were split into three bins (rare – MAF < 1%, low-frequency – 1% ≤ MAF < 5%, and common – MAF ≥ 5%). In doing so, we found that the depletion in coding regions was driven by rare and low-frequency variants (OR < 0.79, *p*-value < 2×10^{-5}; **Figure 1A**), whereas regulatory regions were weakly enriched in low-frequency and common aSNPs (OR > 1.04, *p*-value < 0.05). We

then used the ancestral/derived states of each aSNP to distinguish between derived alleles that originated in the Neanderthal lineage (i.e. derived-aSNPs, 91% of all aSNPs) and ancestral alleles that were re-introduced by Neanderthals following the fixation of the derived allele in the human lineage (i.e., ancestral-aSNPs, 9% of all aSNPs, **Figure S1**). When comparing derived-aSNPs to non-archaic variants of similar derived allele frequency (DAF), we observed a depletion of both coding and regulatory regions in rare archaic variants (DAF < 1%, OR < 0.91, $p < 2 \times 10^{-4}$), and no significant enrichment of common/low-frequency aSNPs was detected in regulatory regions. Interestingly, when focusing on variants presenting a DAF ≥ 95% (i.e., ancestral-aSNPs at ≤ 5% frequency), we observed an enrichment of archaic variants in regulatory regions (OR > 1.25, $p < 2 \times 10^{-5}$), highlighting the contribution of ancestral alleles re-introgressed by Neanderthal to gene regulation in humans.

To understand how introgression has impacted genetic diversity across various types of regulatory elements, we then investigated the relative density of aSNPs across promoters, enhancers, and miRNA binding sites. While this metric differed markedly across frequency bins, it did not differ across categories of regulatory elements, despite their important differences in strength of negative or background selection (**Figures 1B-D**). However, when measuring the rate at which Neanderthal alleles were introgressed in Europe or Asia, we found that they were less likely to reach high frequency (MAF>5%) in coding or regulatory regions, with respect to non-functional regions ($p< 10^{-10}$, **Figures 1E** and **S2**). This effect was less marked among enhancers, which, together with their larger size (**Figure 1F**), suggests that Neanderthal variants are quantitatively more likely to affect gene regulation via modification of enhancer activity.

Given the low fraction of the genome that is covered by miRNAs and miRNA binding sites (miRNABS) (**Figure 1F**), they are expected to be, quantitatively, the least affected by archaic introgression. Indeed, we only found 6 aSNPs that overlap the sequence of mature miRNAs, two of which alter the seed region (**Figure 2A**): rs74904371 in miR-2682-3p ($MAF_{CHB} = 0$, $MAF_{CEU} = 3\%$) and rs12220909 in miR-4293 ($MAF_{CHB} = 17\%$, $MAF_{CEU} = 0$). The presence of

aSNPs in four of these miRNAs, particularly those located in seed regions, affected the set of genes they bind (**Figure 2B** and **Table S1**). We also detected 2,909 aSNPs in miRNABS, 29% of which were common (**Table S2**). We found a direct linear relationship between the number of genes bound by a miRNA and the number of aSNPs in its binding sites ($r = 0.56$, p-value < 10^{-10}; **Figure 2C**), suggesting that introgression affected miRNABS independently of their cognate miRNAs. Highlighting a pertinent example, the *ONECUT2* locus (MIM: 604894) presents the highest number of aSNPs altering conserved miRNABS (**Figure 2D**), and has been previously reported as a likely target of adaptive introgression.[24] This gene, which encodes a member of the onecut family of transcription factors, contains 13 aSNPs that alter miRNABS, 6 of which are highly conserved (GerpRS > 2). Interestingly, these aSNPs fall within the 0.4% most differentiated aSNPs between Europeans and Asians at the genome-wide level (F_{ST} > 0.38). We also detected aSNPs, mostly population specific, that alter conserved miRNABS at several, key immune genes, including *CXCR5* (MIM: 601613; MAF_{CHB} = 16%, MAF_{CEU} = 1%), *TLR6* (MIM: 605403; MAF_{CHB} = 8%, MAF_{CEU} = 0), *IL7R* (MIM: 146661; MAF_{CHB} = 8%, MAF_{CEU} = 0) or *IL21* (MIM: 605384; MAF_{CHB} = 0, MAF_{CEU} = 8%).

Next, we focused on how archaic introgression has affected promoters and enhancers. Given the tissue-specific impact of archaic introgression on gene regulation,[10,12] we searched for enrichments in Neanderthal ancestry across regulatory elements in 127 different tissues.[33] The impact of archaic introgression in promoters was similar to that of the remainder of the genome, in all tissues and frequency bins (**Table S3)**. Conversely, we found that enhancers are enriched in common aSNPs in 42 tissues (FDR < 5%, **Figure 3A** and **Table S4**), with similar patterns being detected in CEU and CHB populations ($r = 0.62$, **Figure S3**). Among the 42 tissues presenting significant enrichments, adipose-derived mesenchymal stem cells (AdMSC) and mesenchymal stem cell-derived adipocytes were the most enriched (OR > 1.13, p-value < 3×10^{-5}), followed by fetal heart (OR = 1.15, p-value = 8×10^{-5}), small intestine (OR = 1.21, p-value = 2×10^{-4}) and different T cell tissues (OR > 1.14, p-value < 1.5×10^{-2}). When restricting our analyses to derived-aSNPs (using SNPs with DAF<50% as background set), we replicated the enrichments at enhancers for 27 tissues

(FDR < 5%, **Table S3** and **S4**), indicating that the impact of archaic introgression for these tissues is driven by Neanderthal-derived variants.

Focusing on circulating immune cell types (**Figure 3B**), we found enrichments among enhancers of various types of primary T cells, the most significant being $CD4^+/CD25^-$ memory T cells (OR = 1.21, *p*-value = 2.2×10^{-4}), while enhancers of B cells, monocytes and natural killer cells exhibited a density of common aSNPs similar to genome-wide expectations. We also observed that shared enhancers across different T cell subtypes (i.e., active in more than half of T cells subtypes, "core T cells enhancers") display an enrichment in aSNPs (OR = 1.22, *p*-value = 5×10^{-4}, **Figure 3C**), with respect to more specialized enhancers that are only active in a small fraction of T cell subtypes.

We sought to assess whether the enrichment in aSNPs detected in enhancers resulted from an excessive divergence of these elements in the Neanderthal lineage, or a higher rate of archaic introgression at enhancers. We quantified the number of fixed human/Neanderthal differences at enhancers, across the 127 tissues, focusing on sites where both the Altai and Vindija Neanderthal genomes[6,7] differ from the ancestral sequence. We uncovered large tissue variability, with enhancers active in induced pluripotent stem cells presenting the highest divergence (290 differences/Mb) and those active in pancreas showing the lowest (220 differences/Mb). However, given that the number of fixed differences strongly correlates with genetic diversity (i.e., density of common variants, r = 0.71, *p*-value $<10^{-20}$), we measured the ratio of the number of fixed human/Neanderthal differences to that of common, segregating SNPs in the region. Using this metric, we found that enhancers of T cells displayed the strongest divergence (7% increase compared to the mean across tissues, Wilcoxon *p*-value < 2×10^{-8}), whereas stem cells showed the lowest (4% decrease, Wilcoxon *p*-value < 7×10^{-6}) (**Figure 4A**). Focusing on the rate of introgression, defined as the proportion of Neanderthal-descent alleles that are present in the human genome at a MAF > 5%, we found that enhancers of T cells showed the highest percentage (5% increase, Wilcoxon *p*-value < 2×10^{-5}), while brain the lowest (7% decrease, Wilcoxon *p*-value< 4×10^{-5}) (**Figure 4A**).

We then explored the factors that may drive, at the genome-wide level, the detected variation in Neanderthal divergence and archaic introgression. We correlated, using 100kb-windows, divergence and introgression with metrics that capture local variation in neutral (mutation, recombination) and selected (negative and background selection) diversity. Specifically, we measured the percentage of GC to account for their higher mutability, genetic size as measure of recombination rate, density of conserved sites (GerpRS > 2) as a measure of negative selection, and background selection derived as (1-B), where B is the mean B-statistic in the window.[36] We found that background selection correlates with a lower rate of archaic introgression (r=-0.049, *p*-value $<10^{-15}$, **Figure 4B**), consistently with previous findings,[1,2] but also with increased local divergence (r = 0.22, *p*-value < 10^{-20}, **Figure 4C**) and reduced density of both common variants and fixed differences (r = -0.46 and -0.05 respectively, *p*-value < 5×10^{-20}, **Figure S4)**. We also found that negative selection and recombination rate correlate with both divergence and introgression, even after adjusting for background selection (**Figure S5**).

To understand further how these factors could account for the variation in divergence and introgression detected at enhancers (**Figure S6**), we focused on 3 model tissues: T cells (enhancers with high divergence and introgression), AdMSC (enhancers with low divergence and high introgression), and prefrontal cortex (enhancers with high divergence and low introgression) (**Figure 4D**). When correcting for the various neutral and selective factors, introgression at T cell enhancers did not exceed that of other tissues (*p*-value > 0.11), but the high divergence and relative density of aSNPs remained significant (*p*-value < 8×10^{-3}). For AdMSC, introgression remained higher than expected (*p*-value = 4×10^{-3}), leading to an excess of aSNPs despite their depletion in divergence (*p*-value = 3.8×10^{-2}). For enhancers active at prefrontal cortex, all variables were within expected bounds. Collectively, these analyses indicate that variation of several neutral and selective factors are not sufficient to explain the excess of Neanderthal introgression detected at enhancers. Some enhancers may have undergone past adaptation in the Neanderthal lineage or adaptive introgression in modern humans, as illustrated by T-cells and AdMSC, respectively.

Finally, we explored the impact of archaic introgression at enhancers on gene expression. To identify genes whose expression is altered by Neanderthal introgression at enhancers, we focused on tissues where promoter capture-HiC data were available[37,38] and assigned each enhancer located in a promoter-interacting region to the corresponding gene(s). Archaic variants at enhancers predicted to interact with a gene were strongly enriched in eQTLs (OR=2.6, p-value<10^{-3}, **Supplemental Note 1** and **Figure S7**), supporting further the regulatory potential of aSNPs. Genes interacting with T cell enhancers that harbor common aSNPs (N = 1629, **Table S5**) were not enriched in any specific biological function. However, 285 of these genes are highly expressed in T cells (FPKM > 100), and include known regulators of the immune response (e.g., *CXCR4* [MIM: 162643], *IL7R* [MIM: 146661], *IL10RA* [MIM: 146933], *NFKBIA* [MIM: 164008] and *PTPRC* [MIM: 151460]). We found 14 loci presenting signatures of adaptive introgression; i.e., genes that interact with enhancers harboring very high frequency aSNPs (99th percentile of MAF: $MAF_{CEU} > 0.29$ or $MAF_{CHB} > 0.35$; **Figures 5A,B)**. Among these, we found *ANKRD27*, associated with eosinophilic esophagitis (MIM: 610247),[39] and *MED15* (MIM: 607372), involved in several cancers.[40-42] With respect to adipose-related tissues, we identified 690 genes — 43 of which being highly expressed (FPKM > 100) in the adipose tissue — interacting with AdMSC enhancers that contain common aSNPs (**Table S6**). These genes were enriched in functions related to the regulation of cell motility (GO:2000145, p-value < 2.0×10^{-8}) and insulin-like growth factor binding protein complex (GO:0016942, p-value < 2.7×10^{-5}) (**Table S7**). We detected 16 aSNPs at AdMSC enhancers that present strong signatures of adaptive introgression (**Figures 5C,D**).

This study reconstructs the history of how Neanderthal introgression has affected various types of regulatory elements as well as the mechanistic bases through which archaic variants have altered gene regulation. Previous studies have shown that archaic variants are more likely to correlate with gene expression than non-archaic variants segregating at the same frequency.[10,11] Our approach differs from these studies in that it excludes indirect effects from non-archaic variants segregating on introgressed haplotypes, and focuses on the direct

regulatory potential of archaic alleles. In doing so, we find little evidence for an enrichment of common archaic variants in regulatory regions taken as a whole, which may seem at odds with previous studies. Yet, one should note that the functional impact of the archaic material we measure can be decomposed in two separate components: (i) the frequency at which Neanderthal haplotypes are introgressed into the human lineage, which corresponds to the rate of introgression measured by the f4-ratio statistics,[16] and (ii) the degree of human-Neanderthal divergence at regulatory elements, which determines the probability that introgressed haplotypes carry a functional variant. Indeed, when focusing on the rate of introgression, we find that Neanderthal alleles were introgressed at a lower rate in regulatory regions, consistent with recent findings for promoter regions.[16] This lower introgression rate is nevertheless compensated by a higher human-Neanderthal divergence at regulatory regions (**Figure S8**), which is consistent with an increased probability of Neanderthal haplotypes to associate with gene expression.[10,11]

We also explored how Neanderthal introgression has impacted miRNA-mediated regulation, and showed that although miRNAs harbor few archaic variants *per se*, some of them may impact strongly miRNA-mediated gene regulation and disease risk. For example, the archaic allele at miR-4293 (rs12220909) is responsible for the loss of 95% of its targets, and has been associated with diminished cancer susceptibility.[43,44] Archaic introgression has also affected miRNA binding sites, as illustrated by *ONECUT2* (MIM: 604894), where an archaic haplotype that is present at high frequency in Asia (MAF_{CHB} = 0.49) alters multiple conserved miRNA binding sites. *ONECUT2* is involved in liver, pancreas and nervous system development,[45] and has recently been proposed as a regulator of tumor growth in ovarian cancer (MIM: 167000).[46]

Finally, our study reveals that archaic introgression has impacted enhancers in a tissue-specific manner, reflecting either high human-Neanderthal differentiation, as observed in T-cell enhancers, or increased archaic introgression, as detected in AdMSC. Interestingly, the AdMSC enhancers impacted by archaic introgression interact preferentially with genes involved in the regulation of adipocyte differentiation and adipogenesis. These include

receptors such as *PDGFRB* (MIM: 173410) and *TGFBR2* (MIM: 190182), the insulin growth factor *IGF1* (MIM: 147440) and its binding partners *IGFBP2* (MIM: 146731) and *IGFBP3* (MIM: 146732), or the *CXCR4* chemokine (MIM: 162643).[47-53] Furthermore, two of the enhancers harboring archaic variants at the highest frequencies interact with key regulators of adipocyte differentiation, such as *KLF3* (MIM: 609392) and *PRRX1* (MIM: 167420),[54,55] suggesting that introgression at AdMSC may have been adaptive in humans. In support of this notion, Dannemann and colleagues have found that more than half of aSNPs associated to gene expression in subcutaneous adipose tissue had increased in frequency over the last 10,000 years, whereas the majority of aSNPs had decreased in frequency over the same period of time.[10] Given the proposed adaptation of Neanderthals to cold environments,[56] it is tempting to speculate that archaic alleles at enhancers of adipose-derived mesenchymal stem cells provided a selective advantage to early modern humans during their migration out-of-Africa. This hypothesis becomes particularly interesting in the light of previously reported cases of adaptive introgression at *LEPR* (MIM: 601007) and *WARS2*/*TBX15* (MIM: 604733/604127), both involved in the regulation of adipose tissue differentiation and body-fat distribution.[25,57] Further studies aiming to functionally characterize the regulatory effects of Neanderthal variants on adipocyte differentiation and fat distribution are now required, as these archaic variants may have contributed to the adaptation of early Eurasians to colder climates.

Supplemental Data

Supplemental Data include eight figures, seven tables, supplemental Methods, and one Note, which can be found with this article online at http://dx.doi.org/XXX.

Acknowledgments

This work was supported by the *Institut Pasteur*, the *Centre National de la Recherche Scientifique* (CNRS), and the *Agence Nationale de la Recherche* (ANR) grants: “IEIHSEER” ANR-14-CE14-0008-02 and “TBPATHGEN” ANR-14-CE14-0007-02. The laboratory of LQM has received funding from the French Government’s Investissement d’Avenir program, Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (grant no. ANR-10- LABX-62-IBEID). MS was funded by the Ecole Doctorale “Complexité du vivant”, Sorbonne Université (contract n°2532/2016).

Declaration of Interests

The authors declare no competing interests.

Web resources

OMIM, http://www.omim.org

1000genomes, http://www.internationalgenome.org/

Epigenomic Roadmaps, https://egg2.wustl.edu/roadmap/web_portal/

miRanda Software, http://www.microrna.org/microrna/

Neanderthal Genomes, http://cdna.eva.mpg.de/neandertal/

References

1. Sankararaman, S., Mallick, S., Dannemann, M., Prufer, K., Kelso, J., Paabo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354-357.
2. Sankararaman, S., Mallick, S., Patterson, N., and Reich, D. (2016). The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol *26*, 1241-1247.
3. Vernot, B., and Akey, J.M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. Science *343*, 1017-1021.
4. Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., et al. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science *352*, 235-239.
5. Browning, S.R., Browning, B.L., Zhou, Y., Tucci, S., and Akey, J.M. (2018). Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell *173*, 53-61.e59.
6. Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43-49.
7. Prufer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlevic, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyregne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655-658.
8. Dannemann, M., and Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. Am J Hum Genet *101*, 578-589.
9. Simonti, C.N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D.S., Chisholm, R.L., Crosslin, D.R., Hebbring, S.J., Jarvik, G.P., Kullo, I.J., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. Science *351*, 737-741.

10. Dannemann, M., Prufer, K., and Kelso, J. (2017). Functional implications of Neandertal introgression in modern humans. Genome Biol *18*, 61.
11. Quach, H., Rotival, M., Pothlichet, J., Loh, Y.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643-656.
12. McCoy, R.C., Wakefield, J., and Akey, J.M. (2017). Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. Cell *168*, 916-927.
13. Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.L., Patin, E., and Quintana-Murci, L. (2016). Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am J Hum Genet *98*, 5-21.
14. Enard, D., and Petrov, D.A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. Cell *175*, 360-371.
15. Sams, A.J., Dumaine, A., Nedelec, Y., Yotova, V., Alfieri, C., Tanner, J.E., Messer, P.W., and Barreiro, L.B. (2016). Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol *17*, 246.
16. Petr, M., Paabo, S., Kelso, J., and Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. Proc Natl Acad Sci U S A *116*, 1639-1644.
17. Lopez-Valenzuela, M., Ramirez, O., Rosas, A., Garcia-Vargas, S., de la Rasilla, M., Lalueza-Fox, C., and Espinosa-Parrilla, Y. (2012). An ancestral miR-1304 allele present in Neanderthals regulates genes involved in enamel formation and could explain dental differences with modern humans. Mol Biol Evol *29*, 1797-1806.
18. Gunbin, K.V., Afonnikov, D.A., Kolchanov, N.A., Derevianko, A.P., and Rogaev, E.I. (2015). The evolution of Homo sapiens denisova and Homo sapiens neanderthalensis miRNA targeting genes in the prenatal and postnatal brain. BMC Genomics *16 Suppl 13*, S4.
19. Jegou, B., Sankararaman, S., Rolland, A.D., Reich, D., and Chalmel, F. (2017). Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry. Mol Biol Evol *34*, 1974-1980.

20. Kuhlwilm, M., Gronau, I., Hubisz, M.J., de Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A., Lalueza-Fox, C., de la Rasilla, M., et al. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. Nature *530*, 429-433.
21. Rogers, A.R., Bohlender, R.J., and Huff, C.D. (2017). Early history of Neanderthals and Denisovans. Proc Natl Acad Sci U S A *114*, 9859-9863.
22. Harris, K., and Nielsen, R. (2016). The Genetic Cost of Neanderthal Introgression. Genetics *203*, 881-891.
23. Juric, I., Aeschbacher, S., and Coop, G. (2016). The Strength of Selection against Neanderthal Introgression. PLoS Genet *12*, e1006340.
24. Racimo, F., Marnetto, D., and Huerta-Sanchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. Mol Biol Evol *34*, 296-317.
25. Racimo, F., Gokhman, D., Fumagalli, M., Ko, A., Hansen, T., Moltke, I., Albrechtsen, A., Carmel, L., Huerta-Sanchez, E., and Nielsen, R. (2017). Archaic Adaptive Introgression in TBX15/WARS2. Mol Biol Evol *34*, 509-524.
26. Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., and Akey, J.M. (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. Curr Biol *26*, 3375-3382.
27. Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sanchez, E. (2015). Evidence for archaic adaptive introgression in humans. Nat Rev Genet *16*, 359-371.
28. Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature *512*, 194-197.
29. Jagoda, E., Lawson, D.J., Wall, J.D., Lambert, D., Muller, C., Westaway, M., Leavesley, M., Capellini, T.D., Mirazon Lahr, M., Gerbault, P., et al. (2017). Disentangling Immediate Adaptive Introgression from Selection on Standing Introgressed Variation in Humans. Mol Biol Evol *35*, 623-630.
30. Dannemann, M., and Racimo, F. (2018). Something old, something borrowed: admixture and adaptation in human evolution. Curr Opin Genet Dev *53*, 1-8.

31. Kudaravalli, S., Veyrieras, J.B., Stranger, B.E., Dermitzakis, E.T., and Pritchard, J.K. (2009). Gene expression levels are a target of recent natural selection in the human genome. Mol Biol Evol *26*, 649-658.
32. Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554-566.
33. Roadmap Epigenomics Consortium, 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.
34. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biol *5*, R1.
35. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68-74.
36. McVicker, G., Gordon, D., Davis, C., and Green, P. (2009). Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet *5*, e1000471.
37. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Varnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369-1384.
38. Pan, D.Z., Garske, K.M., Alvarez, M., Bhagat, Y.V., Boocock, J., Nikkola, E., Miao, Z., Raulerson, C.K., Cantor, R.M., Civelek, M., et al. (2018). Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. Nat Commun *9*, 1512.
39. Sleiman, P.M., Wang, M.L., Cianferoni, A., Aceves, S., Gonsalves, N., Nadeau, K., Bredenoord, A.J., Furuta, G.T., Spergel, J.M., and Hakonarson, H. (2014). GWAS identifies four novel eosinophilic esophagitis loci. Nat Commun *5*, 5593.

40. Syring, I., Weiten, R., Muller, T., Schmidt, D., Steiner, S., Kristiansen, G., Muller, S.C., and Ellinger, J. (2018). The knockdown of the Mediator complex subunit MED15 restrains urothelial bladder cancer cells' malignancy. Oncol Lett *16*, 3013-3021.
41. Weiten, R., Muller, T., Schmidt, D., Steiner, S., Kristiansen, G., Muller, S.C., Ellinger, J., and Syring, I. (2018). The Mediator complex subunit MED15, a promoter of tumour progression and metastatic spread in renal cell carcinoma. Cancer Biomark *21*, 839-847.
42. Shaikhibrahim, Z., Offermann, A., Halbach, R., Vogel, W., Braun, M., Kristiansen, G., Bootz, F., Wenzel, J., Mikut, R., Lengerke, C., et al. (2015). Clinical and molecular implications of MED15 in head and neck squamous cell carcinoma. Am J Pathol *185*, 1114-1122.
43. Fan, L., Chen, L., Ni, X., Guo, S., Zhou, Y., Wang, C., Zheng, Y., Shen, F., Kolluri, V.K., Muktiali, M., et al. (2017). Genetic variant of miR-4293 rs12220909 is associated with susceptibility to non-small cell lung cancer in a Chinese Han population. PLoS One *12*, e0175666.
44. Zhang, P., Wang, J., Lu, T., Wang, X., Zheng, Y., Guo, S., Yang, Y., Wang, M., Kolluri, V.K., Qiu, L., et al. (2015). miR-449b rs10061133 and miR-4293 rs12220909 polymorphisms are associated with decreased esophageal squamous cell carcinoma in a Chinese population. Tumour Biol *36*, 8789-8795.
45. Kropp, P.A., and Gannon, M. (2016). Onecut transcription factors in development and disease. Trends Dev Biol *9*, 43-57.
46. Lu, T., Wu, B., Yu, Y., Zhu, W., Zhang, S., Zhang, Y., Guo, J., and Deng, N. (2018). Blockade of ONECUT2 expression in ovarian cancer inhibited tumor cell proliferation, migration, invasion and angiogenesis. Cancer Sci *109*, 2221-2234.
47. Gao, Z., Daquinag, A.C., Su, F., Snyder, B., and Kolonin, M.G. (2018). PDGFRalpha/PDGFRbeta signaling balance modulates progenitor cell differentiation into white and beige adipocytes. Development *145*, dev155861.
48. Onogi, Y., Wada, T., Kamiya, C., Inata, K., Matsuzawa, T., Inaba, Y., Kimura, K., Inoue, H., Yamamoto, S., Ishii, Y., et al. (2017). PDGFRbeta Regulates Adipose Tissue

Expansion and Glucose Metabolism via Vascular Remodeling in Diet-Induced Obesity. Diabetes *66*, 1008-1021.

49. Kim, Y.J., Hwang, S.J., Bae, Y.C., and Jung, J.S. (2009). MiR-21 regulates adipogenic differentiation through the modulation of TGF-beta signaling in mesenchymal stem cells derived from human adipose tissue. Stem Cells *27*, 3093-3102.
50. Chang, H.R., Kim, H.J., Xu, X., and Ferrante, A.W., Jr. (2016). Macrophage and adipocyte IGF1 maintain adipose tissue homeostasis during metabolic stresses. Obesity (Silver Spring) *24*, 172-183.
51. Yau, S.W., Russo, V.C., Clarke, I.J., Dunshea, F.R., Werther, G.A., and Sabin, M.A. (2015). IGFBP-2 inhibits adipogenesis and lipogenesis in human visceral, but not subcutaneous, adipocytes. Int J Obes (Lond) *39*, 770-781.
52. Chan, S.S., Schedlich, L.J., Twigg, S.M., and Baxter, R.C. (2009). Inhibition of adipocyte differentiation by insulin-like growth factor-binding protein-3. Am J Physiol Endocrinol Metab *296*, E654-663.
53. Yao, L., Heuser-Baker, J., Herlea-Pana, O., Zhang, N., Szweda, L.I., Griffin, T.M., and Barlic-Dicen, J. (2014). Deficiency in adipocyte chemokine receptor CXCR4 exacerbates obesity and compromises thermoregulatory responses of brown adipose tissue in a mouse model of diet-induced obesity. FASEB J *28*, 4534-4550.
54. Sue, N., Jack, B.H., Eaton, S.A., Pearson, R.C., Funnell, A.P., Turner, J., Czolij, R., Denyer, G., Bao, S., Molero-Navajas, J.C., et al. (2008). Targeted disruption of the basic Kruppel-like factor gene (Klf3) reveals a role in adipogenesis. Mol Cell Biol *28*, 3967-3978.
55. Du, B., Cawthorn, W.P., Su, A., Doucette, C.R., Yao, Y., Hemati, N., Kampert, S., McCoin, C., Broome, D.T., Rosen, C.J., et al. (2013). The transcription factor paired-related homeobox 1 (Prrx1) inhibits adipogenesis by activating transforming growth factor-beta (TGFbeta) signaling. J Biol Chem *288*, 3036-3047.
56. Steegmann, A.T., Jr., Cerny, F.J., and Holliday, T.W. (2002). Neandertal cold adaptation: physiological and energetic factors. Am J Hum Biol *14*, 566-583.

57. Sazzini, M., Schiavo, G., De Fanti, S., Martelli, P.L., Casadio, R., and Luiselli, D. (2014). Searching for signatures of cold adaptations in modern and archaic humans: hints from the brown adipose tissue genes. Heredity (Edinb) *113*, 259-267.
58. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martin, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398-1414 e1324.
59. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet *45*, 580-585.

Figures legends

Figure 1. Enrichment of Neanderthal variants in regulatory regions. (A) Odds ratio depicting the excess or depletion of Neanderthal variants in coding regions and regulatory elements (promoters, enhancers and miRNA binding sites), compared to the remainder of the genome. Enrichments are shown for three bins of minor allele frequencies (MAF), together with 95% bootstrap confidence intervals: **p*-value < 0.05, ***p*-value <0.01, ****p*-value < 0.001. (B) Relative density of aSNPs in promoters, enhancers and miRNA binding sites in different bins of MAF, with 95% bootstrap confidence intervals. (C-D) Comparison of density of conserved sites (GerpRS > 2) and mean B-statistic of promoters, enhancers, and miRNA binding sites. (E) Percentage of alleles fixed in Neanderthal and absent from the African Yoruba from Nigeria (YRI) that are introgressed at a MAF > 5% in Eurasians. For each type of region, boxplots show the variability of the estimates based on 1,000 bootstrap resamples of 100kb genomic windows. The dashed vertical line indicates the genome-wide average. (F) Total length of promoters, enhancers, and miRNA binding sites.

Figure 2. Effects of archaic introgression on miRNA-mediated regulation. (A) Representation of the archaic (red) and modern (green) human alleles for the 6 miRNAs presenting a Neanderthal-introgressed variant in their mature sequence. The seed region of the miRNAs is shaded in grey. (B) Total number of genes bound by the archaic/modern human allele of each of the 6 miRNAs harboring a Neanderthal variant in their mature sequence. (C) Relationship between the number of targets of each miRNA and the number of common aSNPs in the corresponding miRNA binding sites. (D) Introgression of aSNPs altering the miRNABS at the *ONECUT2* locus (MIM: 604894). Gene structure is shown in upper panel, with miRNA binding sites that are altered by archaic introgression highlighted in green. The middle panel represents the density of conserved sites (GerpRS > 2) in 1000bp-windows, and the bottom panel the repartition and frequency of archaic alleles at the locus

(blue for CEU, red for CHB). aSNPs that overlap miRNABS are represented with a darker shade, and aSNPs that disrupt a conserved site are marked with stars.

Figure 3. Effects of archaic introgression at enhancers. (A) Volcano plot illustrating the enrichment of common aSNPs in the enhancers of 127 different tissues from the Epigenomic Roadmap Consortium. Tissues with an FDR < 5% (triangles) are significantly enriched. (B) Enrichments of common aSNPs in the enhancers of different immune tissues. Vertical bars indicate 95% confidence intervals computed by bootstrap. (C) Enrichment of common aSNPs in the enhancers that are active in more than half of the investigated T cell types (dark red, referred to as "core T cells"), and in enhancers that are active in each T cell type and are not part of core T cell enhancers (light red, referred to as "cell type specific enhancers"). (B,C) Note that $CD4^+$ T cells are separated based on CD25 to distinguish T_{reg} ($CD25^+$), T_{EM} ($CD25^{low}$) and T_{helper} ($CD25^-$).

Figure 4. Factors shaping human-Neanderthal divergence and archaic introgression at enhancers. (A) Comparison of the relative density of fixed human-Neanderthal differences and rate of introgression in the enhancers of the 127 tissues studied. The size of the circles is proportional to the relative density of common aSNPs in the enhancers of the corresponding tissue; a black circle is added when the relative density of common aSNPs is significantly higher in these enhancers (FDR < 5%) than in the rest of the genome. The densities of each tissue category along the two axes are also presented. (B-C) Genome-wide correlations, using 100kb-windows, between the rate of Neanderthal introgression (B) or the relative density of fixed human-Neanderthal differences (C), and neutral and selective forces. *p-value < 10^{-2}, **p-value < 10^{-10}, ***p-value < 10^{-20}. (D) Observed values of rate of introgression and relative density of fixed differences and common aSNPs at the enhancers of core T cells, AdMSC, and prefrontal cortex, with respect to expectations based on 100 kb windows matched for length of enhancers alone, or for length of enhancers, percentage of GC, recombination rate, density of conserved sites and mean B-statistic of their enhancers

(see **Supplemental Methods**). n.s. not significant, *p-value < 0.05, **p-value < 0.01, ***p-value < 10^{-3}.

Figure 5. Manhattan plots of genes interacting with enhancers that contain archaic variants. (A-B) Genome-wide distribution of MAF in CEU or CHB at aSNPs that overlap enhancers active in T Cells (core T cell enhancers). For each window of 1Mb along the genome, only the aSNP with the highest MAF is shown. Point sizes reflect FPKM of the most expressed genes (max FPKM across T lymphocytes from Blueprint database[58]) among genes interacting with the enhancer in T cells.[37] (C-D) Similar plots for enhancers active in AdMSC. Point sizes reflect the FPKM of the most expressed gene (max FPKM in GTEx tissues *Adipose – subcutaneous* and *Adipose – Visceral (Omentum)*[59]) among genes interacting with the enhancer in adipose tissue.[38]

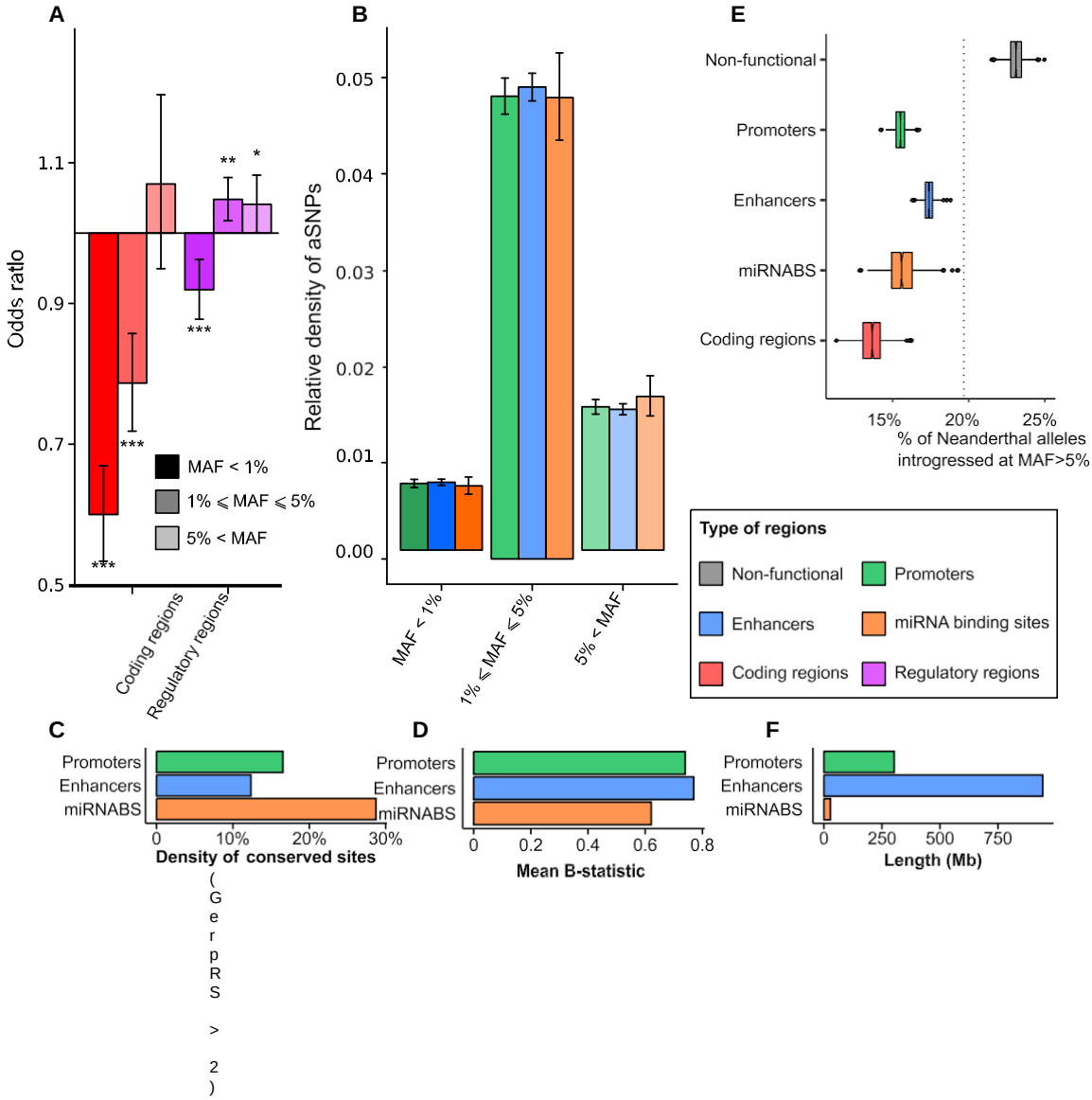


Figure 1. Enrichment of Neanderthal Variants in Regulatory Regions

(A) Odds ratio depicting the excess or depletion of Neanderthal variants in coding regions and regulatory elements (promoters, en- hancers, and miRNA binding sites) compared to the remainder of the genome. Enrichments are shown for three bins of minor-allele frequencies (MAFs), together with 95% bootstrap confidence intervals: *p value < 0.05, **p value < 0.01, ***p value < 0.001.

(B) Relative density of aSNPs in promoters, enhancers, and miRNA binding sites in different MAF bins, with 95% bootstrap confidence intervals.

(C and D) Comparison of density of conserved sites (GerpRS > 2) and mean B statistic of promoters, enhancers, and miRNA binding sites.

(E) Percentage of alleles that are fixed in Neanderthal, absent from the African Yoruba from Nigeria (YRI), and introgressed at a MAF $> 5\%$ in Eurasians. For each type of region, box plots show the variability of the estimates based on 1,000 bootstrap resamples of 100 kb genomic windows. The dashed vertical line indicates the genome-wide average.

(F) Total length of promoters, enhancers, and miRNA binding sites.

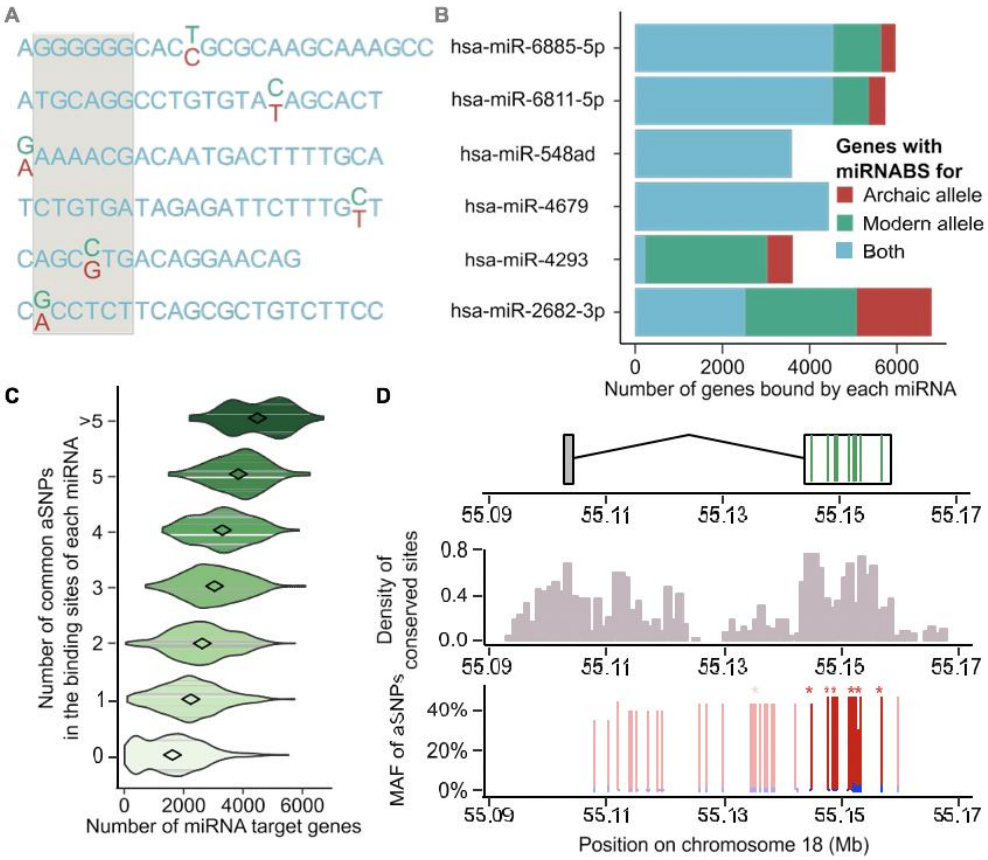


Figure 2. Effects of Archaic Introgression on miRNA-Mediated Regulation

(A) Representation of the archaic (red) and modern (green) human alleles for the six miRNAs presenting a Neanderthal-introgressed variant in their mature sequence. The seed region of the miRNAs is shaded in gray.

(B) Total number of genes bound by the archaic and/or modern human allele of each of the six miRNAs harboring a Neanderthal variant in their mature sequence.

(C) Relationship between the number of targets of each miRNA and the number of common aSNPs in the corresponding miRNA binding sites.

(D) Introgression of aSNPs altering the miRNABS at the ONECUT2 locus (MIM: 604894). Gene structure is shown in the upper panel, and miRNA binding sites that are altered by archaic introgression are highlighted in green. The middle panel represents the density of conserved sites (GerpRS > 2) in 1,000 bp windows, and the bottom panel represents the repartition and frequency of archaic alleles at the locus (blue for CEU, red for CHB). aSNPs that overlap miRNABS are represented with a darker shade, and aSNPs that disrupt a conserved site are marked with stars.

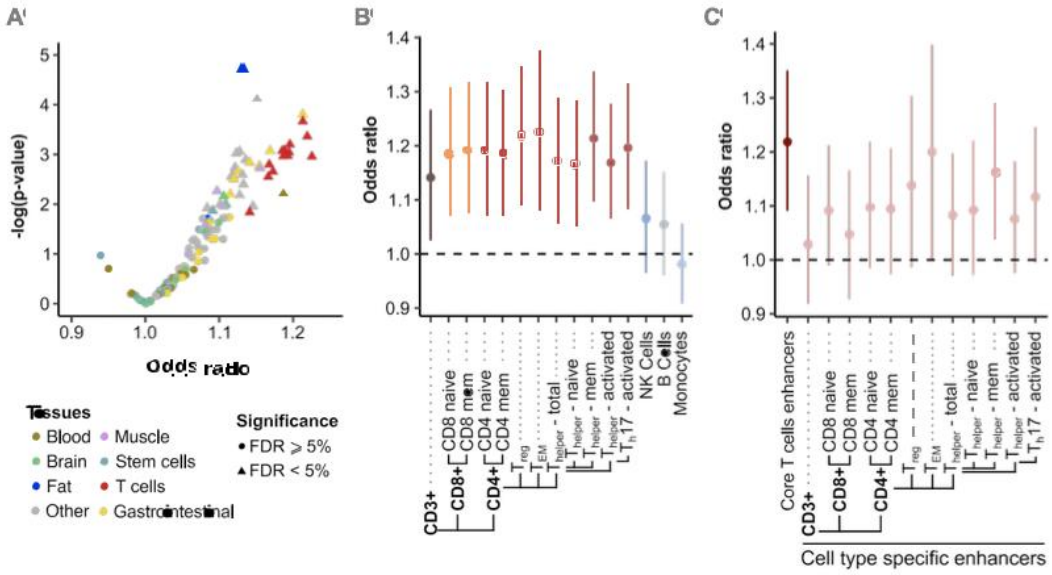


Figure 3. Effects of Archaic Introgression at Enhancers

(A) Volcano plot illustrating the enrichment of common aSNPs in the enhancers of 127 different tissues from the Epigenomic Roadmap Consortium. Tissues with FDR < 5% (triangles) are significantly enriched.

(B) Enrichments of common aSNPs in the enhancers of different immune tissues. Vertical bars indicate 95% confidence intervals computed by bootstrap analysis.

(C) Enrichment of common aSNPs in the enhancers that are active in more than half of the investigated T cell subtypes (dark red, referred to as "core T cells") and in enhancers that are active in each T cell subtype and are not part of core T cell enhancers (light red, referred to as "cell-type-specific enhancers"). Vertical bars indicate 95% confidence intervals computed by bootstrap.

(B and C) Note that $CD4^{+}$ T cells are separated on the basis of CD25 so that T_{reg} ($CD25^{+}$), T_{EM} ($CD25^{low}$), and T_{helper} ($CD25^{-}$) are distin- guished from one another.

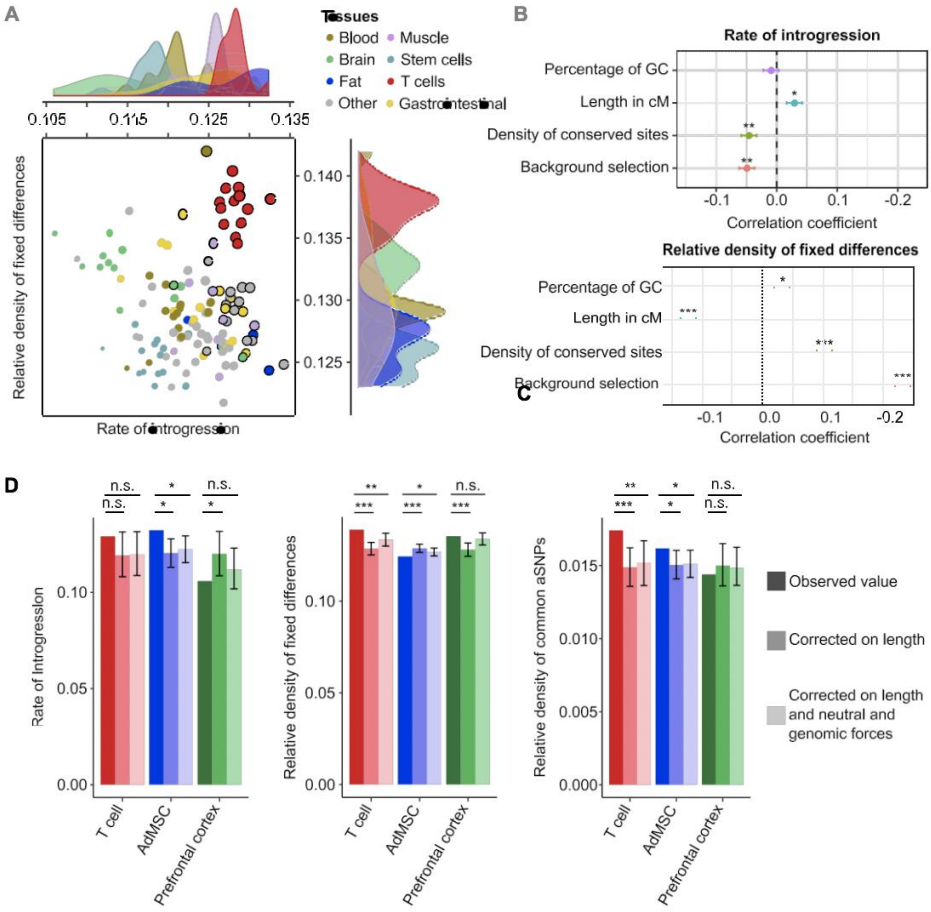


Figure 4. Factors Shaping Human-Neanderthal Divergence and Archaic Introgression at Enhancers

(A) Comparison of the relative density of fixed human-Neanderthal differences and rate of introgression in the enhancers of the 127 tissues studied. The size of the circles is proportional to the relative density of common aSNPs in the enhancers of the corresponding tissue; a black circle is added when the relative density of common aSNPs is significantly higher in these enhancers (FDR $<$ 5%) than in the rest of the genome. The density of each tissue category along the two axes is also presented.

(B and C) Genome-wide correlations, using 100 kb windows, between either the rate of Neanderthal introgression (B) or the relative den- sity of fixed human-Neanderthal differences (C) and neutral and selective forces. *p value $< 10^{-2}$, **p value $< 10^{-10}$, and *** p value $< 10^{-20}$. For each correlation, horizontal lines indicate 95% confidence interval.

(D) Observed values of rate of introgression and relative density of fixed differences and common aSNPs at the enhancers of core T cells, AdMSCs, and prefrontal cortex, with respect to expectations based on 100 kb windows matched for length of enhancers alone or for length of enhancers, percentage of GC, recombination rate, density of conserved sites, and mean B statistic of their enhancers (see Supplemental Data). n.s. ¼ not significant; *p value < 0.05, **p value < 0.01, and ***p value $< 10^{-3}$. Errors bars indicate 95% confidence intervals of the expected values obtained by resampling.

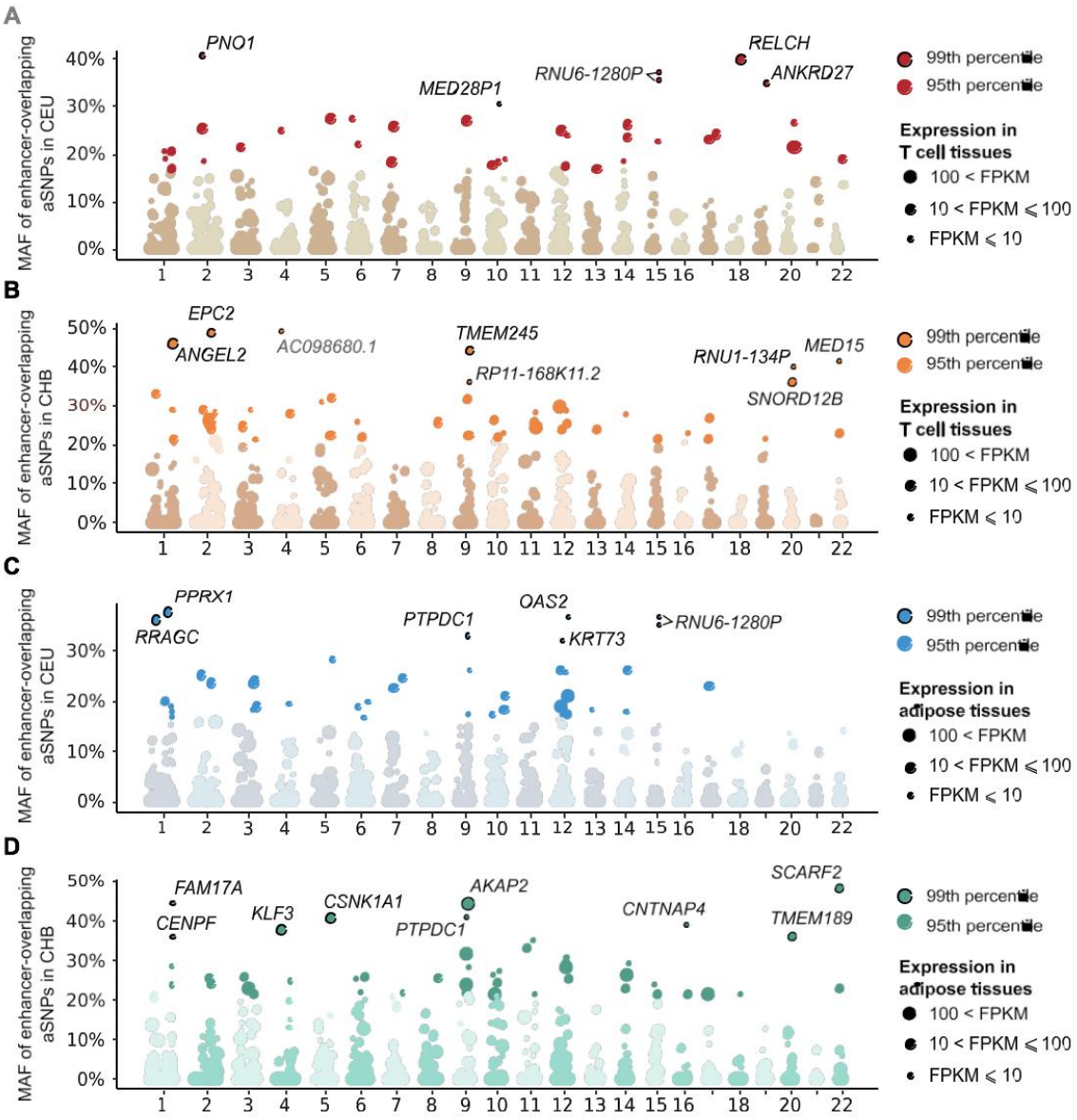


Figure 5. Manhattan Plots of Genes Interacting with Enhancers That Contain Archaic Variants

(A and B) Genome-wide distribution of MAFs in CEU or CHB at aSNPs that overlap enhancers active in T Cells (core T cell enhancers). For each window of 1 Mb along the genome, only the aSNP with the highest MAF is shown. Point sizes reflect FPKM of the most expressed genes (max FPKM across T lymphocytes from Blueprint database[58]) among genes interacting with the enhancer in T cells.[37]

(C and D) Similar plots for enhancers active in AdMSCs. Point sizes reflect the FPKM of the most expressed gene (max FPKM in GTEx tissues Adipose—Subcutaneous and Adipose—Visceral [Omentum][59]) among genes interacting with the enhancer in adipose tissue.[38]