bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reconstructing the 3D genome organization of Neanderthals reveals that chromatin folding shaped phenotypic and sequence divergence

Evonne McArthur[1,2], David C. Rinker[3], Erin N. Gilbertson[2,4], Geoff Fudenberg[5], Maureen Pittman[4,6], Kathleen Keough[6,7], Katherine S. Pollard[2,4,6,7,8], and John A. Capra[1,2,4,8,9,*]

[1]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN
[2]Bakar Computational Health Sciences Institute, University of California, San Francisco, CA
[3]Department of Chemistry, Vanderbilt University, Nashville, TN
[4]Biomedical Informatics Graduate Program, University of California San Francisco, San Francisco, CA
[5]Department of Computational and Quantitative Biology, University of Southern California, Los Angeles, CA
[6]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA
[7]Chan Zuckerberg Biohub, San Francisco, CA, USA
[8]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA
[9]Department of Biological Sciences, Vanderbilt University, Nashville, TN

[*] *Correspondence to* tony@capralab.org

Abstract

Changes in gene regulation were a major driver of the divergence of archaic hominins (AHs)—Neanderthals and Denisovans—and modern humans (MHs). The three-dimensional (3D) folding of the genome is critical for regulating gene expression; however, its role in recent human evolution has not been explored because the degradation of ancient samples does not permit experimental determination of AH 3D genome folding. To fill this gap, we apply novel deep learning methods for inferring 3D genome organization from DNA sequence to Neanderthal, Denisovan, and diverse MH genomes. Using the resulting 3D contact maps across the genome, we identify 167 distinct regions with diverged 3D genome organization between AHs and MHs. We show that these 3D-diverged loci are enriched for genes related to the function and morphology of the eye, supra-orbital ridges, hair, lungs, immune response, and cognition. Despite these specific diverged loci, the 3D genome of AHs and MHs is more similar than expected based on sequence divergence, suggesting that the pressure to maintain 3D genome organization constrained hominin sequence evolution. We also find that 3D genome organization constrained the landscape of AH ancestry in MHs today: regions more tolerant of 3D variation are enriched for introgression in modern Eurasians. Finally, we identify loci where modern Eurasians have inherited novel 3D genome folding from AH ancestors, which provides a putative molecular mechanism for phenotypes associated with these introgressed haplotypes. In summary, our application of deep learning to predict archaic 3D genome organization illustrates the potential of inferring molecular phenotypes from ancient DNA to reveal previously unobservable biological differences.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Highlights

- The 3D genome organization of archaic hominins can be inferred from sequence to facilitate comparisons to modern humans.
- Loci with 3D genome folding divergence between humans and Neanderthals highlight functional differences in the eye, supra-orbital ridges, hair, lungs, immune response, and cognition.
- 3D genome organization constrained recent human evolution.
- Tolerance to variation in 3D genome organization shaped the landscape of Neanderthal ancestry in modern humans.
- Neanderthal introgression contributed novel 3D genome folding patterns to Eurasians.

2 Introduction

The sequencing of archaic hominin (AH) and modern human (MH) genomes has transformed our understanding of human history, evolution, and biology [1–5]. However, even with these whole-genome sequences available, our understanding of how and why AHs differed from MHs is limited [6]. A major challenge in understanding the phenotypic and sequence differences between AHs and MHs is bridging the gap between genetic variation and function. The evolution of hominins is largely driven by changes in the regulation of conserved proteins [7–13], but the mechanisms through which archaic variants influence gene expression, and ultimately phenotype, are incompletely understood [6, 13, 14].

Many studies that investigate the gene regulatory differences between MHs and AHs leverage Neanderthal ancestry remaining in modern Eurasians. Because MHs interbred with many AH groups over the past 50,000 years, more than one-third of the Neanderthal genome remains in introgressed sequences in MH genomes [15, 16]. These investigations have found widespread expression differences between Neanderthal and MH alleles [11, 12], many of which are hypothesized to contribute to trait variation in diverse MHs [17–21]. Phenotypes associated with Neanderthal ancestry range from immune system response [18, 19, 22–29], hair and skin coloration [18, 19, 30–32], metabolism [33–36], cardiopulmonary function [19, 37], skeletal morphology [19, 38], and behavior [18, 19]. However, since most regions of MH genomes have little or no evidence of introgression [11, 12, 30, 31, 39–41], considering only introgressed variation provides a very limited view into hominin biology and cannot address why certain regions of MH genomes tolerated Neanderthal DNA better than others.

Colbran et al. [13] addressed this challenge by inferring AH gene regulation genome-wide through predictive models trained on gene expression data in MHs [42]. They estimated that over 1900 genes had different patterns of regulation between AHs and MHs. However, the specific molecular mechanisms through which archaic variants alter gene expression remain unclear. Gokhman et al. [43] and Batyrev et al. [44] aimed to elucidate these mechanisms by computationally reconstructing maps of AH DNA methylation. They found 2,000 differentially methylated regions that associate with genes predominantly related to facial and limb anatomy. Together, these illustrate the potential to mechanistically link archaic genotypes with regulatory functions via prediction of molecular phenotypes.

Yet, previous work has been unable to address a fundamental aspect of gene regulation and genome function—the physical three-dimensional (3D) organization of the genome. Regulation of gene expression is facilitated by the 3D looping and folding of chromatin in the cell nucleus, which is central to enhancer-promoter (E-P) communication and insulation [45–52]. The 3D genome also plays a role in determining cell-type identity, cellular differentiation, replication timing, and risk for multiple diseases [53–59]. Advances in chromosome-conformation-capture technologies (3C, 4C, 5C, Hi-C, MicroC) [60–64] allow quantification of genome folding at increasing resolution from chromosomal territories, megabase-scale topologically associating domains (TADs), to smaller-scale loops [62] and "architectural stripes," which can reflect enhancer activity and gene activation [65–67]. Disrupting 3D genome folding can cause

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inappropriate E-P interactions and alter gene expression in ways that lead to disease [49, 50, 68–72]. Accordingly, there is preliminary evidence suggesting the 3D genome constrains variation at different scales of evolution [73–77] and that reorganization of chromatin may contribute to gene regulatory evolution and inter-species gene expression divergence [78].

Thus, to fully understand the consequences of genetic variation between AHs and MHs, we must consider the 3D genome folding. However, the role of 3D genome organization in the divergence between AHs and MHs has never been explored because chromatin contacts cannot be assayed in ancient DNA. 3D genome folding is facilitated by a complex interplay of CTCF binding with cohesin and other architectural factors [50, 62, 79, 80]. Recent deep learning methods have been developed that learn the sequence "grammar" underlying 3d genome folding patterns [81–84]. We hypothesized that these deep learning methods would allow us to infer genome-wide 3D chromatin contact maps of Neanderthals and Denisovans. Because the molecular mechanisms that determine genome organization, like CTCF binding and co-localization with cohesin, are largely evolutionarily conserved [85, 86], models trained using human data perform well even when applied to DNA sequences from distantly related species, such as mouse [82]. Thus, unlike genome-wide methods for predicting organism-level phenotype (e.g., polygenic risk scores), these models can be applied across diverse hominins.

To elucidate the contribution of 3D genome folding to recent hominin evolution, we apply novel deep learning methods for inferring 3D genome organization from DNA sequence patterns to Neanderthal, Denisovan, and diverse MH genomes. Using the resulting genome-wide 3D genome folding maps, we identify 167 loci that are divergent in 3D organization between AHs and MHs. We show that these 3D-diverged loci are enriched for physical links to genes related to the function and morphology of the eye, supra-orbital ridge, hair, lung function, immune response, and cognition. We also find that 3D genome organization constrained recent human evolution and patterns of introgression. Finally, we evaluate the legacy of introgression on the 3D organization of humans and identify examples where introgression imparted divergent 3D genome folding to Eurasians. In summary, our application of deep learning to predict archaic 3D genome folding provides a window into previously unobservable molecular mechanisms linking genetic differences to phenotypic consequences in hominin evolution.

3 Results

3.1 Reconstructing the 3D genome organization of archaic hominins

To evaluate the role of 3D genome organization changes in recent human evolution, we apply deep learning to infer 3D genome organization from DNA sequences of archaic hominins (AHs) and modern humans (MHs) (Fig. 1). We consider the genomes of four AHs—one Denisovan and three Neanderthals, each named for where they were discovered (*Altai* mountains, *Vindija* and *Chagyrskaya* caves) [1–4]. We compare these to 20 diverse MHs from the 1000 Genomes Project (Table S1) [87].

For each individual, we predict chromatin contact maps across the genome. Each contact map gives a 2D representation of the predicted 3D chromatin physical contacts, which will refer to as "3D genome organization". We predict these maps using approximately 1 Mb (1,048,576 bp) tiled sliding windows overlapping by half with Akita, a convolutional neural network (CNN) trained on high-quality experimental chromatin contact maps (Hi-C and Micro-C) [82]. Each resulting contact map represents pairwise physical 3D contact frequencies at approximately 2 kb (2,048 bp) resolution for a single individual. Previous work demonstrated that Akita accurately infers 3D contact organization at this resolution [82]. We only consider windows with full (100%) sequence coverage in the MH reference, and we conservatively mask missing archaic sequence with the human reference sequence (Figs. S1,S2,S3 and Methods).

We compare contact maps from two genomes using a "3D divergence" score, namely, one minus the Spearman's rank correlation coefficient ($1 - \rho$) for all pixels in the maps. Genomic windows with more different 3D genome maps have higher 3D divergence and, conversely, a window with lower 3D divergence will reflect more 3D similarity (Fig. 1). Other divergence metrics (e.g., based on Pearson's correlation coefficient and mean squared difference) are strongly correlated (Fig. S4). Akita is trained

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

simultaneously on Hi-C and Micro-C across five cell types in a multi-task framework. In the main text we focus on predictions from the highest resolution cell type, human foreskin fibroblast (HFF). Results are similar when considering other cell types (e.g. embryonic stem cells) (Fig. S5), likely because of limited cell-type-specific differences in both available experimental data and model predictions [82].

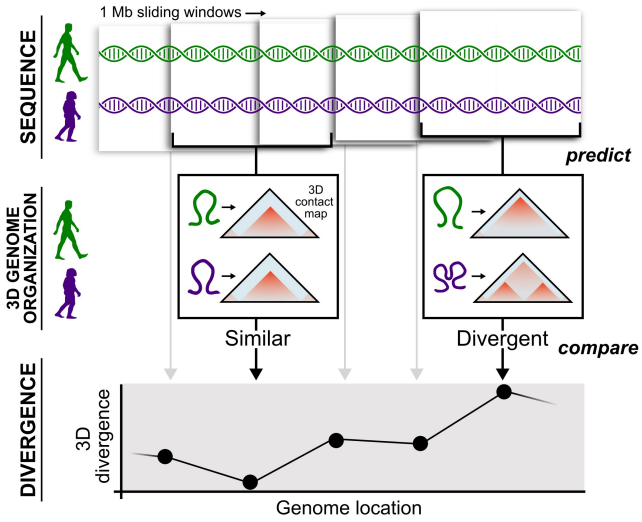

Figure 1: Reconstructing the 3D genome organization of archaic hominins. We infer 3D genome organization from sequence across the genomes of modern humans (MHs, green) and archaic hominins (AHs, purple). Using approximately 1 Mb (1,048,576 bp) sliding windows (overlapping by half), we input the genome sequences into Akita, a convolutional neural network, to predict 3D genome contact maps [82]. The resulting contact maps are compared between MHs and AHs to identify regions that have similar 3D genome organization (left, low divergence) and regions that have different 3D organization (right, high divergence).

3.2 Archaic hominin and modern human genomes exhibit a range of 3D divergence

Reconstructing the genome-wide 3D genome organization of AHs and MHs revealed genomic windows with a range of 3D divergence (Fig. 2A). Most of the genome has very similar 3D genome organization between AHs and MHs (circle example in Fig. 2A-B). However, we also found regions of AH-MH 3D genome divergence. Some of these differences are changes in predicted chromatin contact intensity but similar overall organization (diamond example in Fig. 2A-B). Others reveal reorganization with evidence of new sub-organization (neo-TADs or -loops) or lost structures (fused TADs or loops) (indicated with an "x" example in Fig. 2A-B). At the 95th percentile of observed divergence, differences in the contact maps are substantial. However, because the 3D divergence measure considers the entire window, strong focal changes may not rank as highly as structural differences that influence a large segment of the window (diamond vs. "x" examples in Fig. 2B).

To illustrate genome-wide patterns of divergence in 3D organization, we plotted the average divergence of each of the AHs to five modern African individuals from different subpopulations (Fig. 2C). We show the landscape of 3D divergence across the entire genome for all four AHs in Fig. S6. Some AH-MH divergences are shared across all four archaics, while others are specific to a single lineage like the Denisovan individual (Fig. 2C). We only considered sub-Saharan Africans in these comparisons, because they have low levels of AH introgression. We consider how introgressed variation in Eurasians influences 3D divergence in a subsequent section.

3.3 3D genome organization diverges between AH and MH at 167 genomic loci

To consistently identify regions with divergent 3D genome organization between AH and MH, we compared the 3D contact maps at each locus for each AH to 20 MH (African) individuals. We applied

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

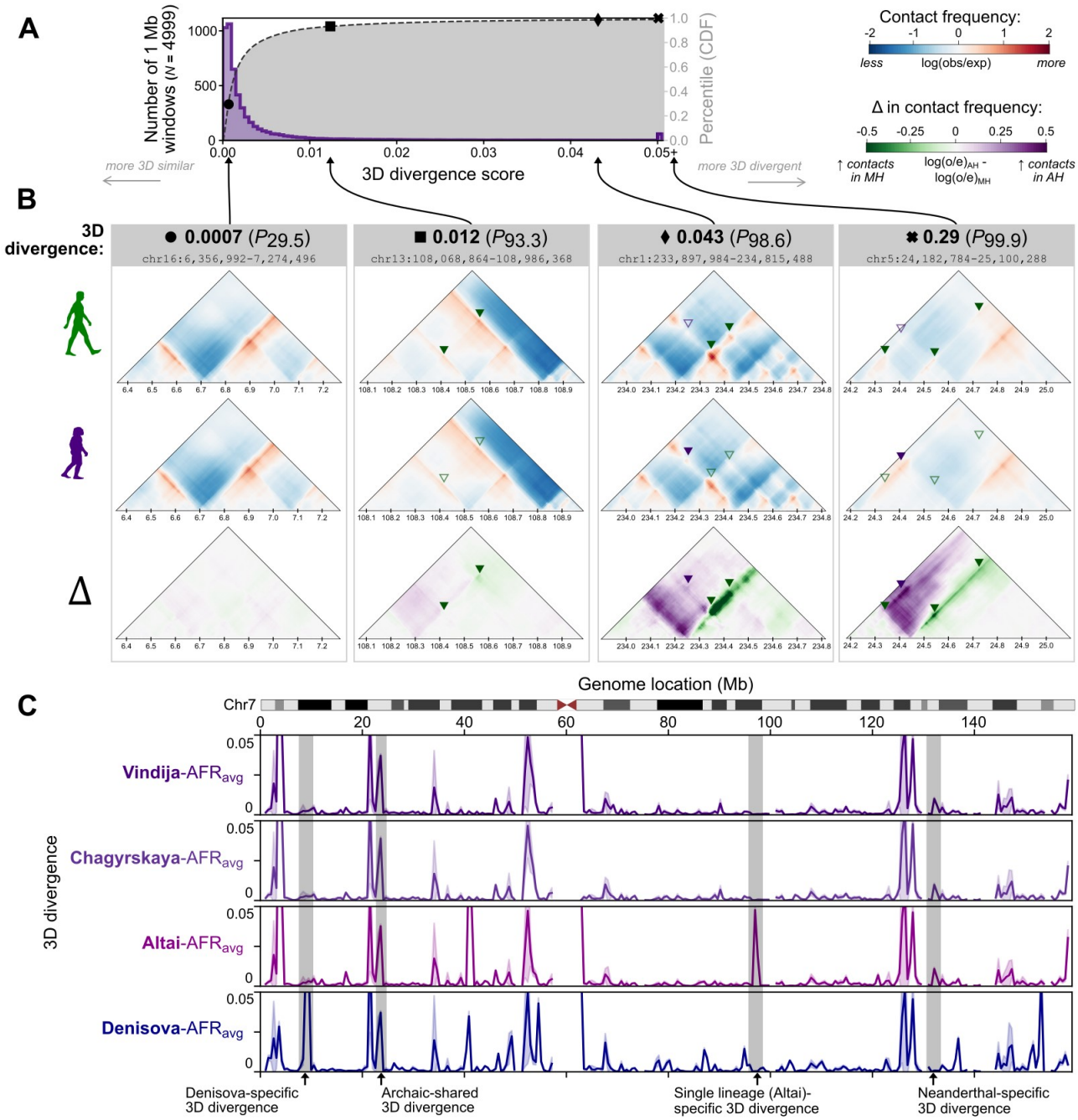

Figure 2: 3D genome divergence between archaic hominins (AHs) and modern humans (MHs) varies across the genome. **(A)** Distribution of 3D genome divergence between AHs and modern humans MHs for 1 Mb windows across the genome. Most windows have similar 3D genome organization between MHs and AHs (low 3D divergence). The cumulative density function (CDF) of this distribution is overlaid in gray with percentiles on the right vertical axis. **(B)** We highlight four examples (shapes) along the 3D divergence distribution illustrating low 3D divergence (left) to high divergence (right). Each example compares a representative African MH (top, HG03105) to a Neanderthal (bottom, Vindija) in terms of both raw score and relative percentile of 3D divergence. Examples with scores near the 95th percentile have visible contact map differences, but the type of differences vary from re-organization (neo-TADs or TAD-fusions) to altered contact intensity (stronger vs. weaker TAD/loop). Green and purple triangles indicate regions with increased contact frequency in MH versus AH, respectively. **(C)** Average 3D divergence along chromosome 7 between AHs and five representative African MHs. The error band indicates the 95% confidence interval (CI). Comparing the 3D genomes of Neanderthals (purple) or Denisova (blue) with MHs reveals windows of both similarity and divergence (peaks). Featured examples (gray overlays) highlight regions of 3D divergence that are shared (e.g., shared across all archaics) or lineage-specific (e.g., specific to the Denisovan individual).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a conservative procedure that required all 20 AH-MH comparisons to be more 3D divergent than all MH-MH comparisons (Fig. 3A). In other words, the differences between the 3D genome organization of an AH to all MHs must be more extreme than the differences between each MHs to all other MHs. Furthermore, we required the average AH-MH 3D divergence to be in the 80th percentile of the most diverged. This identified regions with consistent 3D differences between AHs and MHs (Fig. 3A, left) while excluding regions with a large 3D diversity in modern humans (Fig. 3A, right) (Methods).

We find 167 total AH-MH consistently 3D diverged loci: 67, 70, 71, and 73 for Altai, Vindija, Chagyrskaya, and Denisova compared to MHs, respectively (Fig. 3B). 3D diverged loci are found throughout the genome on every chromosome (Fig. 3B). As suggested by Fig. 2C, some 3D divergences are shared by all four AHs ($N = 7$), and many are shared by all three Neanderthals ($N = 43$) (Fig. 3B). We summarize the AH-MH 3D divergent windows in Tables S2,S3 and report a larger set of windows based on less conservative criteria in Table S4.

To illustrate the properties of a AH-MH 3D divergent window, we highlight a divergent locus on chromosome 2 that is nearby several immune genes (Fig. 3C). MHs have an approximately 140 kb loop linking the promoter of *ICOS* at 204.80 Mb to a CTCF motif at 204.94 Mb. This CTCF motif is overlapped by many ChIP-seq peaks for transcription factors (TFs) involved in determining chromatin folding (CTCF, RAD21, SMC3, and ZNF143). The contact maps for both Vindija and Altai Neanderthal show a more prominent "architectural stripe"—an asymmetric loop-like contact often reflecting enhancer activity [65–67]—starting near the promoter of *ICOS*. However, in contrast to MHs, the loop does not end at the same CTCF site and instead has greater contact frequency with a CTCF site at 205.2 Mb. Thus, the resulting loop in Neanderthals is predicted to be over 400 kb—three times as large as the MH loop.

To determine which AH-MH nucleotide differences cause the largest change in the contact maps, we used *in silico* mutagenesis (Methods). Using an African MH (HG03105) background, we inserted every allele unique to the AH genome one-by-one and measured the resulting 3D genome divergence. This identifies the archaic variant resulting in the largest 3D organization changes between the AH and MH genomes, a G to C change at chr2:204,937,347 (Methods). This change disrupts a high information-content site in the CTCF binding site described above. All MHs carry an ancestral C allele, but Vindija and Altai have a derived G allele. In summary, we predict that the Neanderthal-derived allele weakens CTCF binding leading to reduced insulation between *ICOS*, a T-cell costimulator, with downstream contacts.

3.4 Regions with 3D divergence highlight AH-MH phenotypic differences

To explore the functional effects of AH-MH 3D genome divergence, we tested for phenotypic annotation enrichment. We considered the 43 loci with shared divergence between MHs and all three Neanderthals (Fig. 3B). Although the loci were identified at approximately 1 Mb resolution, most 3D modifications disrupt a smaller sub-window. Thus, as described in the example above (Fig. 3C), we used *in silico* mutagenesis to identify the AH-MH sequence change(s) that produced the largest disruption in the contact maps. We will refer to these as "3D-modifying variants" (Methods). We then intersected the predicted 3D-modifying variants with experimentally defined TADs to determine the genes to which they are physically linked. Ultimately, we found 88 physical links to protein-coding genes (85 unique genes) for the 45 3D-modifying variants in the 43 Neanderthal-MH 3D divergent loci (Tables S2,S5).

We tested if these genes are enriched for phenotypic annotations using both gene-phenotype links from rare disease (OMIM Human Phenotype Ontology [HPO] terms) and common disease databases (GWAS Catalog 2019) [88–92]. 3D genome organization perturbation has been linked to both types of disease: large-scale disruption leading to severe disease and subtle changes in regulatory insulation contributing to complex traits disease [69–72, 74]. We find links to 271 and 208 candidate traits from the rare and common disease ontologies, respectively. For each trait, we test if the observed overlap with 3D divergent loci is more than expected by chance using an empirically-generated null distribution (Methods). In summary, this sequential process links 3D divergent windows to variants to TADs to genes

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

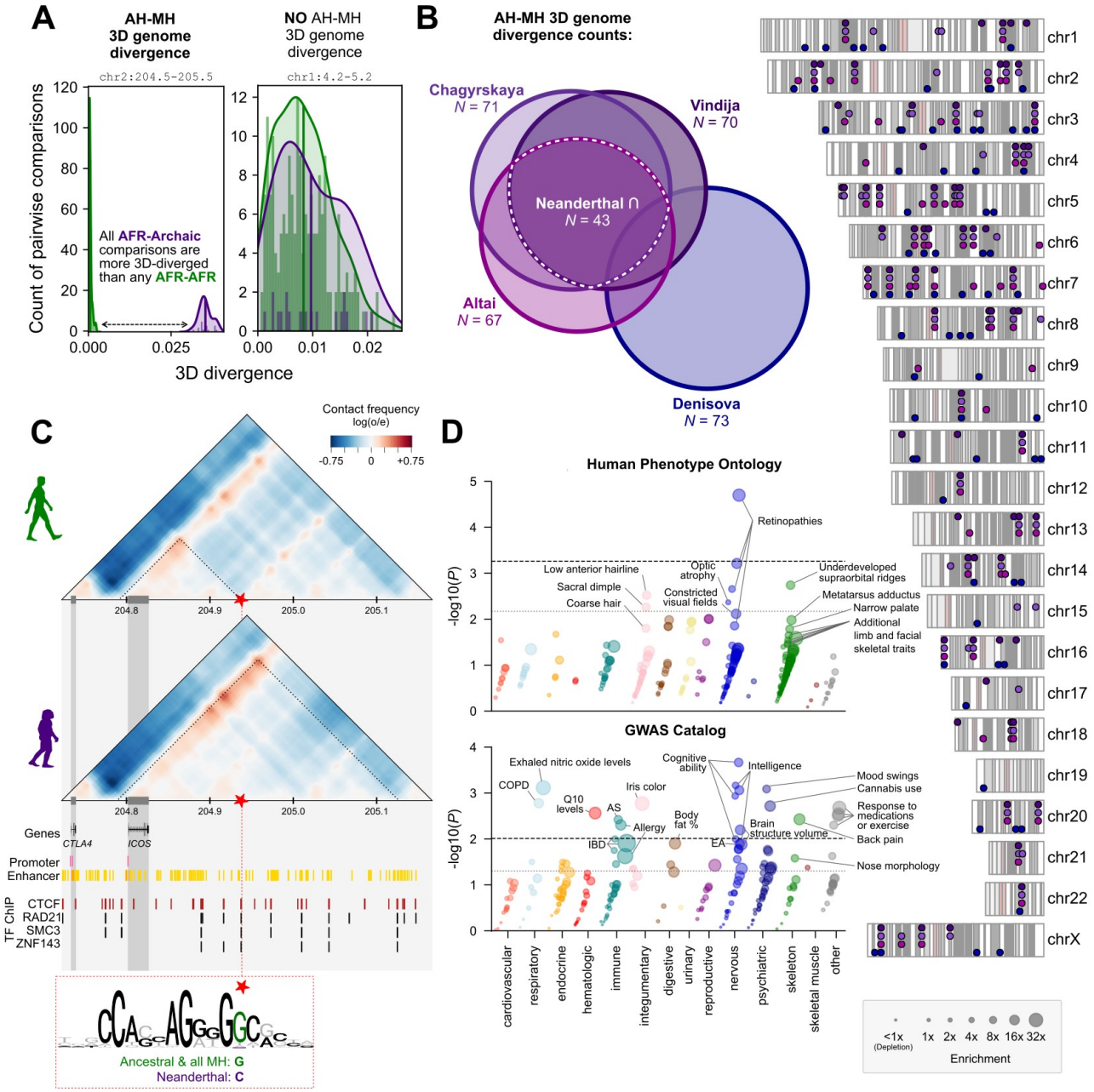

Figure 3: Regions with 3D divergence between MHs and AHs highlight loci linked to phenotypic differences. **(A)** We identified genomic windows with 3D divergence between AH and MH by comparing distributions of pairwise divergence in 3D contact maps. We used a conservative procedure that required all 20 comparisons of each AH to 20 MH (African) individuals (purple, $n = 20$) to be more 3D-diverged than all MH-MH comparisons (green, $n = \binom{20}{2} = 190$) and the mean of the AH-MH divergences (purple) to be in the 95th percentile of most diverged. The left plot shows an example that meets these criteria (chr2:204,472,320-205,520,896). The right shows an example where there is diversity in 3D genome organization, but not an AH-MH divergence (chr1:4,194,304-5,242,880). **(B)** We identified 167 AH-MH 3D divergent windows across the genome. Many are shared (Euler-diagram), but some are unique to a single lineage, with the most unique divergence in the Denisovan. **(C)** Contact maps for the example Neanderthal-MH 3D divergent window shown in **A** (zoomed to chr2:204,722,176-205,166,592). All MHs have a smaller domain insulated by a CTCF site (red star). In Neanderthals (Vindija and Altai), the CTCF motif is disrupted with a C instead of a G (red dashed box, chr2:204,937,347). We predict that this leads to ectopic connections with the promoter of *ICOS* (T-cell costimulator). **(D)** Phenotype enrichment for the 43 Neanderthal 3D diverged loci identified in **B** (white dashed line). We computed functional annotation enrichment for genes physically linked to 3D-modifying variants at these 3D divergent loci using HPO (top, $n = 271$) and GWAS catalog (bottom, $n = 208$) annotations (Methods). Within each phenotypic domain, traits are organized along the vertical axis by significance and along the horizontal axis by enrichment (also indicated by size). Genes nearby AH-MH 3D divergence are enriched for functions related to the retina and visual field, skeletal morphology (notably, supra-orbital ridge), hair, lung function, immune and medication response, and cognitive traits. Significance lines represent the P-value thresholds that controls the FDR with $q = 0.05$ (dotted) and $q = 0.1$ (dashed). (COPD: chronic obstructive pulmonary disease, AS: ankylosing spondylitis, IBD: inflammatory bowel disease, EA: educational attainment)

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and, ultimately, phenotypes (Fig. S7).

With the HPO annotations, we found enrichment for effects of these genes related to the eye (retinopathies, optic atrophy, constricted visual field [most significant association: 27× enriched, $P = 2 \times 10^{-5}$]), skeletal system (notably, supraorbital ridge morphology [12×, $P = 0.002$]), and hair (e.g. low anterior hairline [12×, $P = 0.003$]) (Fig. 3D, top). In the GWAS Catalog annotations, we find enrichment related to intelligence and cognition (13×, $P = 0.0002$), lung function (NO levels, COPD [35×, $P = 0.0008$]), response to certain medications (30×, $P = 0.002$), immunologic response (ankylosing spondylitis, allergy, inflammatory bowel disease [12×, $P = 0.004$]), and brain region volumes (putamen, subcortex [17×, $P = 0.006$]) (Fig. 3D, bottom). Trait enrichments for 3D-modifying variants found in Denisova are highlighted in Fig. S8. Because Denisova and Neanderthal share many alleles, some similar traits are enriched (retinopathy, intelligence, lung function, etc.); however, overall, we find fewer enriched traits.

In summary, genomic loci with 3D divergence between Neanderthals and MHs are enriched for physical proximity to genes associated with a diversity of traits related to the skeleton, eye, hair, lung, immune response, brain region volume, and cognitive ability. These findings align with and expand what we know from both the fossil-record and previous work based on variants in MHs [11, 14–20]. Importantly, our approach permitted the interrogation of variants unobserved in MHs (76% of predicted 3D-modifying variants), and it provides a putative molecular mechanism for the phenotypic differences.

3.5 Relationship between sequence divergence and 3D divergence

Given that we observe 3D differences between AH and MH genomes, we quantified the relationship between 3D and sequence divergence on both genome-wide and more local scales. First, we computed the genome-wide 3D genome divergence for all pairs of AH and MH individuals. We find the mean 3D genome divergence largely follows sequence divergence (Figs. 4A,S9). Neanderthals are the most similar in 3D genome organization to other Neanderthals, then to the Denisova, and then to MHs (mean 3D divergences: 9.8×10^{-4}, 3.4×10^{-3}, and 4.3×10^{-3}, respectively). Genome-wide 3D divergence also tracks with sequence divergence within the Neanderthal: Vindija and Chagyrskaya are more similar than they are to the outgroup Altai (Vindija-Chagyrskaya mean 3D divergence of 8.4×10^{-4} vs. Vindija-Altai of 1.0×10^{-3}) [3].

Next, we evaluated if sequence divergence and 3D divergence are correlated on the local scale. We find a very weak positive relationship between 3D and sequence divergence at the 1 Mb window level (Fig. 4B, $r^2 = 0.01$, $P = 2.3 \times 10^{-13}$). As suggested by the weak correlation, many windows with low sequence divergence have high 3D divergence, and many windows with high sequence divergence have low 3D divergence.

Given the weak relationship between sequence and 3D divergence, we sought to identify some properties of sequence differences that result in large 3D divergence. Based on the importance of CTCF-binding in maintaining 3D genome organization [50, 62, 79, 80], we quantified the effects of AH-MH nucleotide differences overlapping CTCF binding motifs. Disruption of CTCF binding sites is important, but not all disruptions are likely to influence 3D divergence. Leveraging additional functional genomics data on CTCF binding and TAD boundaries, we find that the quantity, quality, and context (e.g., strength of a motif and proximity to a TAD boundary) influence whether AH-MH sequence divergence will result in a 3D organization divergence (Fig. S10). For example, if a window has at least one AH-MH nucleotide difference overlapping a strong CTCF-bound motif near a TAD boundary (within 15 kb), the AH-MH 3D divergence is 1.64-times greater ($P = 0.00077$, $N = 260/4999$ windows, Fig. 4B). Thus, we are observing complex sequence patterns underlying 3D genome folding that could not be determined by simply considering sequence divergence or intersecting AH variants with all CTCF sites. This is concordant with previous results which suggest that 3D genome folding is governed by a complex CTCF binding grammar [50, 80, 82, 83].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

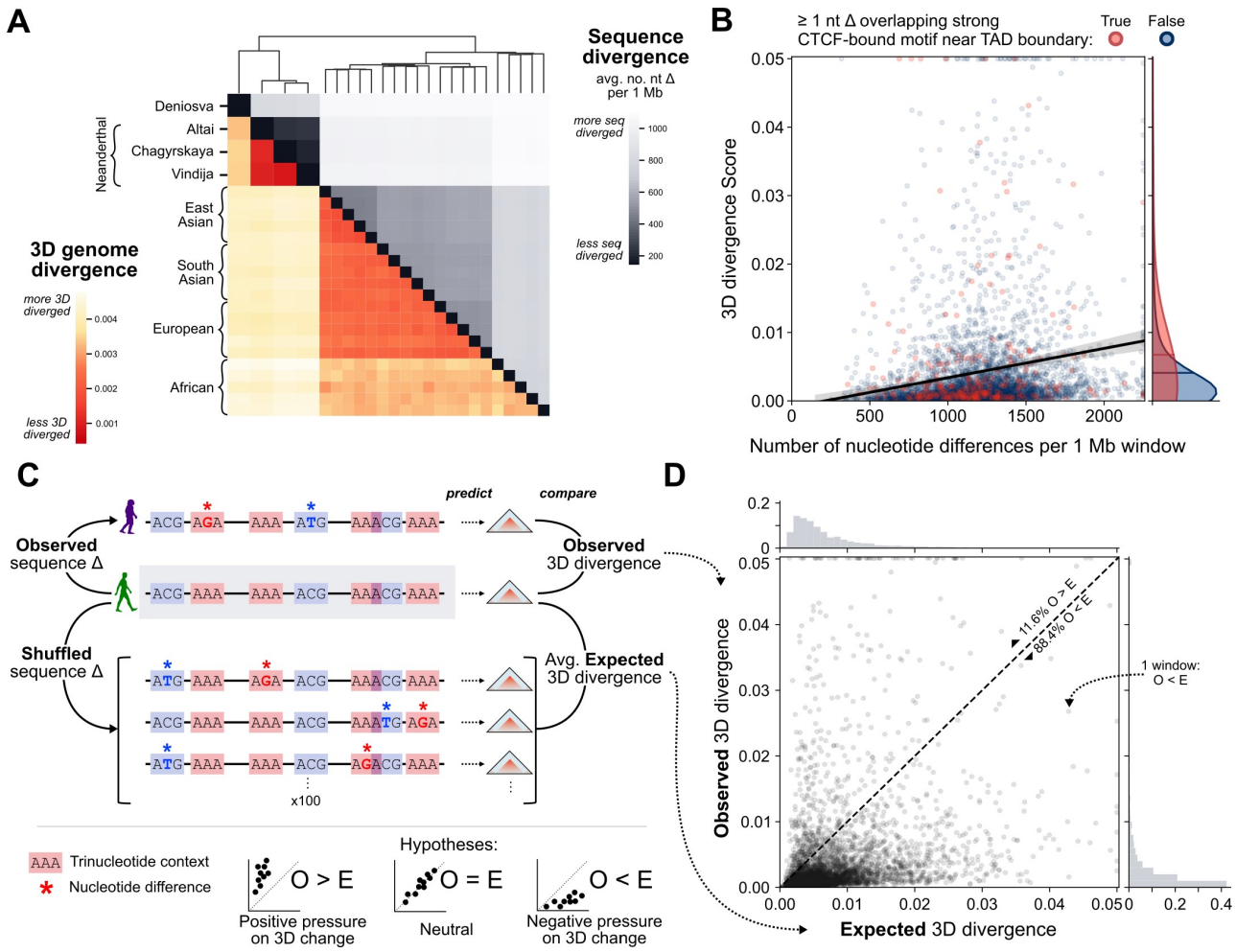

Figure 4: 3D genome organization constrained human sequence divergence. **(A)** 3D genome divergence (lower triangle) follows patterns of sequence divergence (upper triangle). AHs have more similar 3D genome organization to each other than to 15 MHs from different 1000G super-populations. Clustering is based on sequence divergence; see Fig. S9 for clustering by 3D genome divergence and data for each sub-population. **(B)** Sequence divergence is only very modestly correlated with 3D genome divergence ($r^2 = 0.011$, $P = 2.3 \times 10^{-13}$, $N = 4999$). Each point represents a 1 Mb window from a genome-wide comparison between the 3D genome organization of a Neanderthal (Vindija) and African MH (HG03105) individual and the black line with band represents a linear regression with 95% CI. Windows with large 3D divergence are enriched for MH-AH nucleotide (nt) differences overlapping a strong CTCF-bound motif within 15 kb of a TAD boundary (red) (two-tailed Mann–Whitney U $P = 0.00077$). **(C)** To evaluate whether 3D genome organization constrained sequence divergence, we estimate the null distribution of expected 3D divergence based on sequence differences between the Neanderthal (Vindija) and African MH (HG03105) genomes. We shuffle observed nucleotide differences (stars) while preserving tri-nucleotide context (colored rectangles) and predict 3D genome organization for 100 shuffled sequences for each window. Under a model of no sequence constraint due to 3D organization, observed 3D divergence would equal the expected 3D divergence ($O = E$). Alternatively, observing more 3D divergence than expected would suggest positive selection on sequence changes that cause 3D divergence ($O > E$). Finally, observing less 3D divergence than expected would suggest negative pressure on sequence changes that cause 3D divergence ($O < E$). **(D)** Observed 3D divergence is significantly less than the mean expected 3D divergence based on sequence ($O < E$: 88.4% of $N = 4,999$ windows below the diagonal, binomial-test $P < 5 \times 10^{-324}$). The mean expected 3D divergence is on average 1.78-times higher than the observed 3D divergence (t-test $P = 1.8 \times 10^{-48}$). 3D divergence scores greater than 0.05 and nucleotide differences greater than 2250 are clipped to the baseline for visualization purpose

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.6 Maintenance of 3D genome organization constrained sequence divergence in recent hominin evolution

Next, we evaluated if the pressure to maintain 3D genome organization constrained recent human sequence evolution. We estimated whether the amount of 3D divergence between AHs and MHs is more or less than expected given the observed sequence divergence. To compute the expected 3D divergence distribution for each 1 Mb window, we shuffled observed nucleotide differences between an African MH (HG03105) and AH (Vindija Neanderthal) 100 times and applied Akita to predict the resulting 3D genome divergence (Fig. 4C). We controlled for the non-uniform probability of mutation across sites using a model that preserved the tri-nucleotide context of all variants in each window with each shuffle. For each 1 Mb window, we compared the observed 3D divergence with the expected 3D divergence from the 100 shuffled sequences with the same nucleotide divergence.

If the 3D genome does not influence sequence divergence, the observed 3D divergence would be similar to the expected 3D divergence (Fig. 4C, bottom-middle). Alternatively, if the observed 3D divergence is greater than expected based on sequence divergence (Fig. 4C, bottom-left), this suggests positive selection on variation contributing to 3D differences. Finally, if the observed 3D divergence is less than expected based on sequence divergence (Fig. 4C, bottom-right), this suggests negative pressure on variation contributing to 3D differences.

We find that observed 3D divergence is significantly less than expected based on sequence divergence (Fig. 4D). 88.4% of 1 Mb windows have less 3D divergence that expected based on their observed sequence differences (binomial-test $P < 5 \times 10^{-324}$). Genome-wide, the mean expected 3D divergence is 78% higher than the observed 3D divergence (t-test $P = 1.8 \times 10^{-48}$). This suggests that, in recent hominin evolution, pressure to maintain 3D genome organization constrained sequence divergence. This aligns with previous studies that demonstrated depletion of variation at 3D genome-defining elements (e.g., TAD boundaries, CTCF sites) [73–77], but it specifically implicates 3D genome folding.

3.7 3D genome organization constrained introgression in MHs

Eurasian individuals have on average 2% AH ancestry due to introgression; however, AH ancestry is not evenly distributed throughout the genome [2, 15, 31]. Our previous analyses demonstrate that AH and MH exhibit a range of 3D genome organization divergence across the genome (Fig. 2C) and that pressure to maintain 3D genome organization constrained sequence divergence (Fig. 4D). Thus, we hypothesized that for a given genomic window, its tolerance to 3D genome organization variation in MHs would influence the probability that introgressed AH DNA is maintained in MH.

To test this, we first quantified the levels of 3D genome diversity for 20 modern Africans in 1 Mb sliding windows across the genome. We then computed the average African-African 3D genome divergence and term this "3D genome variability". Genomic windows with low 3D genome variability have similar 3D genome organization among all Africans, suggesting these loci are less tolerant of 3D folding changes. In contrast, regions with high 3D genome variability suggest a diversity of 3D genome organization present. Finally, we computed the amount of introgressed sequence in Eurasian populations for each window (Methods, [93]).

Genomic windows with high levels of introgression across Eurasians are enriched for windows with higher 3D genome variability (Fig. 5A, Mann-Whitney U $P = 0.0007$). On average, windows with evidence of introgression have 72% higher 3D genome variability than windows without introgression. Moreover, the magnitude of 3D genome variability is predictive of the average amount (proportion of bp) of introgressed sequence remaining in a 1 Mb window ($P = 5.7 \times 10^{-9}$, Fig 5B, vertical axis). Even when conditioning on sequence variability, 3D genome variability provides additional information about the amount of AH ancestry in a window (Fig 5B, conditional $P = 5.7 \times 10^{-4}$). In other words, even if two windows have the same level of sequence variability in MHs, windows that are more 3D variable are more likely to retain introgressed sequence. We also find that 3D genome variability is more strongly predictive of introgression shared among all three super-populations than an introgressed sequence unique

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to a single super-population (Supplemental Text, Tables S7,S8). Using earlier introgressed Neanderthal haplotype predictions from Vernot et al. [15] and other thresholds yield similar results (Figs. S11,S12). Because we compute variability in Africans with very low levels of AH ancestry, the increased 3D genome variability in MHs is not a result of introgression.

These results suggest that 3D genome organization shaped the landscape of AH introgression in modern Eurasian genomes. Previous findings demonstrated Neanderthal ancestry is depleted in regions of the genome with strong background selection, evolutionary conservation, and annotated molecular function (e.g. genes and regulatory elements) [11, 30, 31, 40, 41]. Our results expand this to implicate the 3D genome as a contributor to the landscape of AH ancestry in MHs today.

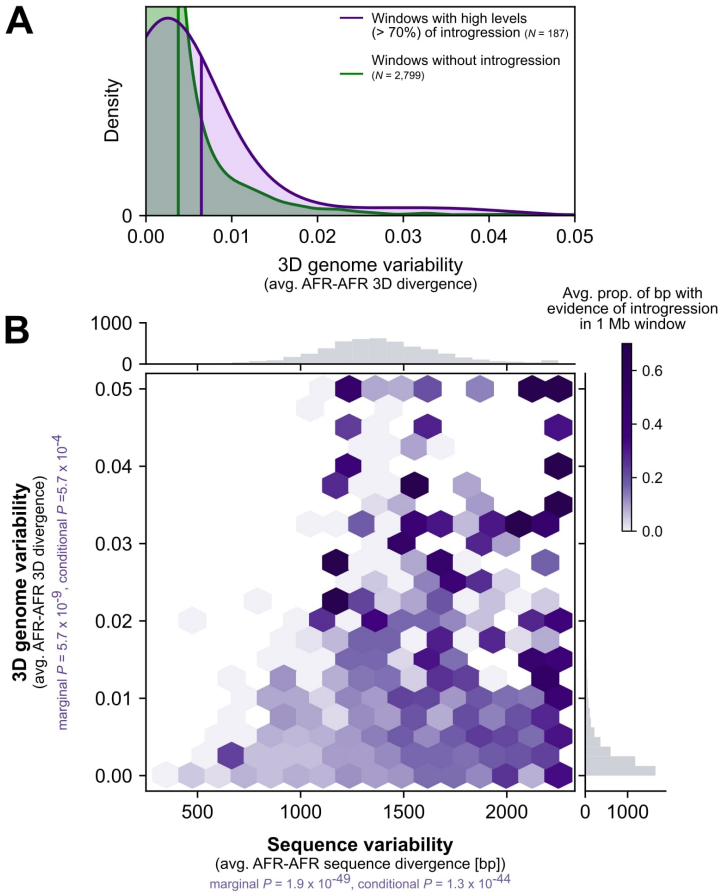

Figure 5: 3D variable windows in MH have more evidence of AH introgression. **(A)** Windows with high levels of introgression across present-day non-African populations (purple, $N = 187$) are more 3D-variable in modern Africans (horizontal axis) than windows without evidence of introgression (green, $N = 2,799$; two-tailed Mann–Whitney U $P = 0.0007$). Vertical lines represent the distribution means. Introgression is called based on Sprime [93]. To focus on regions consistently tolerant of AH ancestry, we considered introgression shared across 1000 Genomes super-populations and covering at least 70% of bases in a 1 Mb window (Methods). Results from other introgression sets and thresholds are similar (Figs. S11–S12 and Tables S7–S8). **(B)** The relationship between sequence variability (horizontal axis) and 3D genome variability (vertical axis) with amount of AH ancestry in a window. Darker purple indicates a higher proportion of introgression in a 1 Mb genomic window. Sequence variability ($P = 1.9 \times 10^{-49}$) and 3D genome variability ($P = 5.7 \times 10^{-9}$) both independently predict amount of introgression. Additionally, even when controlling for sequence variability in a window, 3D genome variability is informative about the amount of introgression ($P = 5.7 \times 10^{-4}$).

3.8 Introgression shaped the 3D genome organization of present-day Eurasians

Given the differences between AH and MH 3D genome organization at many loci, we hypothesized that introgressed AH sequences could have introduced novel 3D contact patterns to Eurasian MHs. To test this, we integrated Eurasians into our previous comparisons of AHs and African MHs.

For example, we found an AH-MH 3D divergent window on chromosome 7 with a striking pattern of 3D genome diversity across modern Eurasians (Fig. 6A). As required to be an AH-MH divergent locus,

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the 3D genome divergence between all Africans and AH (Vindija Neanderthal) was consistently high. And, out of 15 Eurasians, 11 had similar divergent organization compared to the Neanderthal 3D contact map. However, four Eurasians had very low 3D divergence from the Neanderthal.

When examining the contact maps of this window, all Africans have a large approximately 450 kb loop domain starting near the promoter of *IGFBP3*, a gene encoding insulin-like growth factor binding protein 3 (Fig. 6B). In contrast, Neanderthals (Vindija, Chagyrskaya, and Altai) have two smaller sub-domains insulated by a CTCF site. Using *in silico* mutagenesis, we identify that the variant with the largest effect on 3D organization is a G to A change at chr7:46,169,621 (rs12536129). The derived A allele, which strengthens the CTCF motif, appeared along the Neanderthal lineage. The four Eurasians (two Europeans (EUR), two South Asians (SAS)) with 3D genome organization very similar to Neanderthals all have an introgressed haplotype carrying the Neanderthal-derived A allele overlapping this CTCF site [94]. None of the other 11 Eurasians have introgression at this site (although some have introgression in the larger 1 Mb window). Across human populations, this introgressed allele remains at high-frequency today, especially in Peru (28% AMR, 2% EAS, 16% EUR, 11% SAS, 0% AFR, Fig. S13A).

In addition to influencing the strength of a CTCF site, this introgressed allele is also an eQTL in GTEx for the physically linked gene *IGFBP3*, Insulin-like growth factor-binding protein 3 (Fig. S13B, $P = 0.00014$ in artery tissue) [42]. In MHs, this variant is associated with traits including standing height ($P = 9.9 \times 10^{-7}$), fat distribution (trunk fat ratio, impedance measures, $P = 1.3 \times 10^{-5}$), and diastolic blood pressure ($P = 2.1 \times 10^{-5}$) (Fig. S13C).

Of the 191 3D-modifying variants identified in 167 AH-MH 3D diverged windows, 45 are observed in MHs (Table S2). Of note, 18 are common (> 5% MAF) and 6 are at high frequency (> 10%) in at least one MH 1000 Genomes Project (1KGP) super-population which motivates the hypothesis that some introgressed 3D changes were adaptive. We find very modest non-significant enrichment for these loci in previously proposed adaptive haplotypes [94] (2.3-fold enrichment, $P = 0.24$). We annotate all 3D-modifying variants with their nearby genes, allele frequency, and eQTL associations in Table S5.

Given these examples of Neanderthal introgression contributing novel 3D folding to present-day Eurasians, we searched for similar patterns genome-wide. We considered 4,749 autosomal 1 Mb windows for 15 Eurasians (total $n = 71,235$) to quantify the relationship between the amount of introgression and 3D similarity to Neanderthals. We find that the amount of introgression (bp per window) is significantly correlated with 3D divergence to the Vindija Neanderthal ($P = 0.00011$, Fig. 6C). Results from comparisons to the other Neanderthals are consistent (Fig. S14). On average, in a 1 Mb window, if an individual has 80% Neanderthal ancestry, their 3D genome is 2.4 times more similar to the Neanderthal 3D genome than if they have no (0%) Neanderthal ancestry.

In summary, we find that Eurasians with more Neanderthal ancestry in a window have more Neanderthal-like 3D genome folding patterns. Furthermore, at an example locus, we demonstrate how the influence of Neanderthal introgression on 3D genome organization highlights a putative molecular mechanism for the effect of Neanderthal ancestry on human traits.

4 Discussion

The role of 3D genome organization in human biology is increasingly recognized [62, 73–77]; however, current techniques for measuring 3D folding cannot be applied to the study of ancient DNA. Furthermore, despite methodological improvements in assays of the 3D genome, high-resolution experiments across many diverse individuals, species, and cell types remain prohibitive. To address these gaps, we provide a framework for inferring 3D genome organization at population-scale that facilitates evaluation of previously untestable hypotheses.

First, we apply this framework to resurrect archaic 3D genome organization. We find that 3D genome organization constrained sequence divergence and patterns of introgression in hominin evolution. We catalog genomic regions where AH and MH 3D genome organization diverged and illustrate how this novel mechanism links sequence differences to phenotypic differences. Importantly, our approach permitted

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

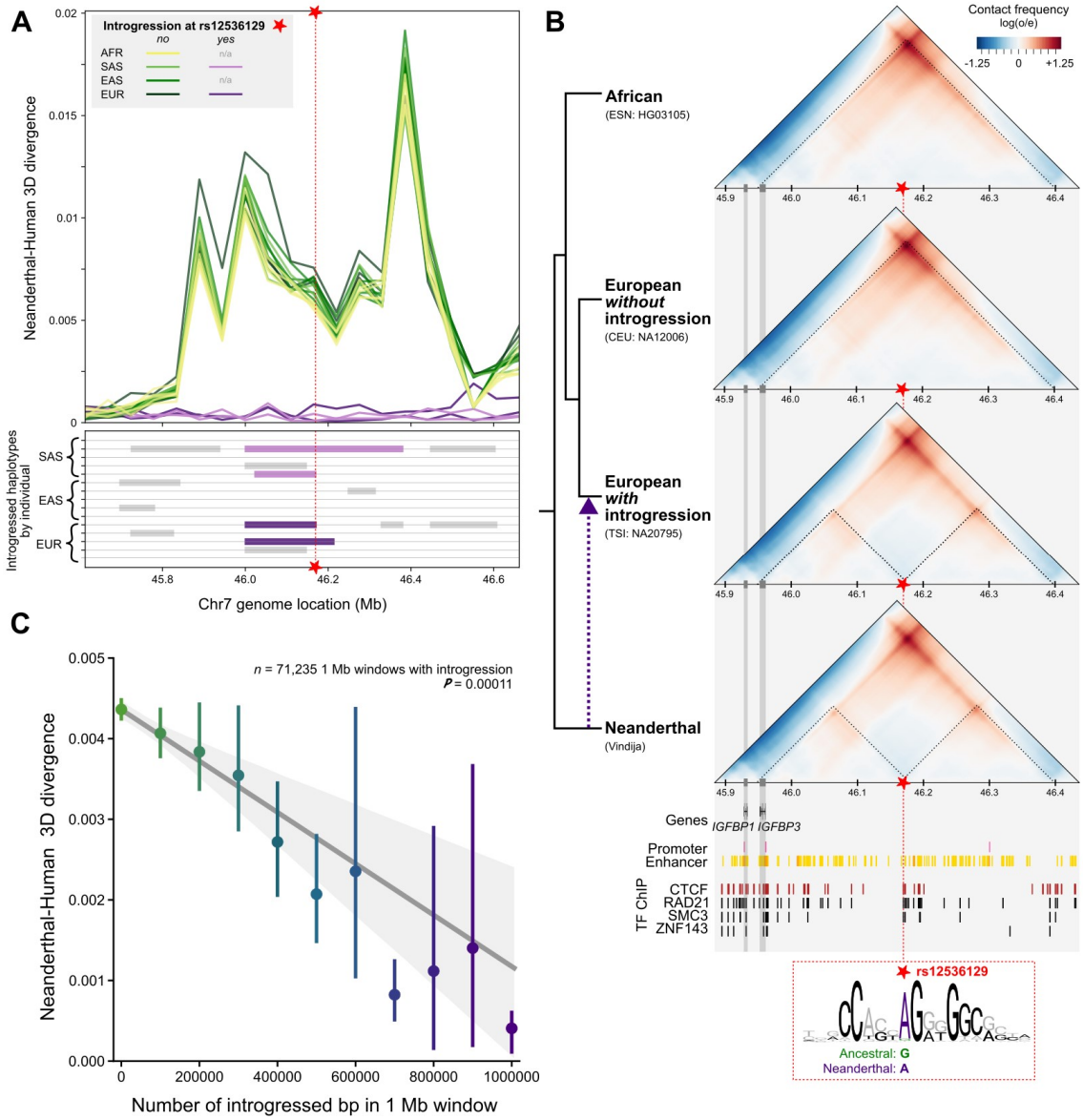

Figure 6: Introgression introduced novel 3D genome organization patterns to modern Eurasians. **(A)** Comparison of the 3D contact maps between Neanderthal (Vindija) and 20 MHs for a window on chromosome 7 reveals that most MHs (yellow, green) have different 3D organization compared to Neanderthals. In contrast, four MHs with introgression (purple boxes) overlapping chr7:46,169,621 (red star) have similar 3D organization compared to Neanderthals across this part of the genome (purple). (AFR: African, SAS: Southeast Asian, EAS: East Asian, EUR: European) This example 3D-divergent locus **(B)** was introgressed into MH and remains at high frequency (28% AMR, 2% EAS, 16% EUR, 11% SAS, 0% AFR, Fig. S13). At this locus (zoomed to chr7:45,883,392-46,436,352), Neanderthals and individuals with introgression have two domains insulated by a CTCF site (red box). In MHs without introgression, this motif is disrupted with a G instead of an A (star, chr7:46,169,621, rs12536129) leading to a larger fused domain and differential contacts with the promoter of *IGFBP3*. **(C)** The amount of introgression in a 1 Mb window (number of bp, horizontal axis) is significantly correlated with the similarity of an individual's 3D genome organization to a Neanderthal's (Vindija) genome organization (vertical axis) ($P = 0.00011$, $n = 71,235$ 1 Mb windows across 15 Eurasians). The error bars signify 95% bootstrapped CIs and the error band signifies the 95% bootstrapped CI for the linear regression estimate.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the evaluation of variants unobserved in MHs, and it provides a putative molecular mechanism for AH-MH phenotypic differences including those that may have been selected against after hybridization (e.g. cognitive and brain morphology traits) [11, 19, 30, 31, 39–41]. Finally, we identify regions in which introgression introduced AH 3D genome folding that are novel to MHs in Eurasians with Neanderthal ancestry. Together, these results illustrate the power of imputing unobservable molecular phenotypes to resolve evolutionary questions about functional divergence.

Second, we anticipate that our framework for comparing and interpreting hundreds of genome-wide 3D genome contact maps will be helpful for testing hypotheses beyond archaic DNA. In the interpretation of genetic variants of unknown significance, it will be key to consider the effect of inter-individual and inter-species variation on 3D genome architecture, especially given recent evidence that even common DNA sequence variants can influence 3D organization and human phenotypic variation [72]. Our work establishes the groundwork to answer many diverse questions. For example, we illustrate how *in silico* mutagenesis can highlight the role of a variant in 3D genome organization and how to integrate this with other functional annotations. This allows us to examine the 3D effects of variants never before observed in MHs, which is essential to non-coding variant interpretation from the lens of both evolution and disease. Our new measure of "3D genome variability" provides genome-wide quantification of how different regions tolerate variation in 3D genome folding. We also demonstrate a simulation approach for testing how 3D genome folding constrains sequence evolution across the genome. Finally, we develop a method to robustly identify 3D divergent windows between populations. With the recent growth of 3D genome *in silico* predictors [81–84], we anticipate that our work can provide a foundation for both hypothesis generation and prioritization of experimental resources.

Although our approach provides many novel benefits, it also has limitations that we hope future work will address. First, our comparisons likely underestimate 3D diversity. We only investigate windows of the genome with complete sequence coverage. Because of ancient sample degradation, we do not have full coverage of AH genomes. We use a conservative approach to effectively mask regions of the genome lacking coverage in AHs (Fig. S1 and Methods). Furthermore, we only consider the effects single nucleotide variants. We do not consider structural variation (SV) due to the challenges of calling SV accurately in ancient samples. We anticipate new methods in ancient DNA sequencing will allow us to model the 3D genome organization of AHs more completely. Second, our 3D genome organization comparisons are based on a correlation-based metric. We demonstrate concordance with comparisons using other more biologically informed methods (Fig. S4); however, more sophisticated methods to quantify the type and resolution of change (e.g. neo-TAD vs TAD-fusion event, scale of TAD vs. loop) would benefit the 3D genome community [81]. Third, although Akita is trained simultaneously across five cell types, 3D genome organization is largely conserved across cell types and predictors only identify limited cell-type-specific differences. Therefore, we focused on the highest resolution predictions in a single context (HFF). As more high-resolution Hi-C and Micro-C becomes available across diverse cell types, our framework can be applied to identify cell-type-specific AH-MH differences.

Several practical caveats must be considered when interpreting some of our results. For example, to conduct *in silico* mutagenesis we manipulate every single nucleotide separately against the same background rather than considering the prohibitively large number of possible combinatorial variant sets. Additionally, while our null model of genome divergence accounts for context-dependent mutation probabilities, we suggest that future study of the influence of 3D folding on genome evolution would benefit from the use of forward-time genomic simulations. The annotations that link 3D-modifying variants to genes and functions are also based on studies in MHs (HPO and GWAS). It is possible, though unlikely, that a gene disrupted in MHs would not lead to the same traits in AHs. Finally, given the scope of our study and the nature of archaic DNA, direct experimental validation is not possible with current technology. To date, Gorkin et al. [72] provides the largest set of Hi-C across 19 MH individuals in the same cell type (LCL GM12878). However, the resolution is too low to call chromatin loops (40 kb vs. 2 kb in our analyses), and 13 of the 19 individuals are African and have almost no Neanderthal ancestry. Thus, we use complementary experimental data, like CTCF ChIP-seq and experimentally-derived TAD maps, to provide independent support for the influence of variants on 3D genome organization and to

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

link variants with genes in true physical proximity. Moreover, even if high-resolution Hi-C were available across many Eurasians, an experimental approach would still not capture all AH variation, highlighting the necessity of our computational approach.

In conclusion, our framework for inferring archaic 3D genome organization provides a window into previously unobservable molecular mechanisms which shaped the sequence and phenotypic evolution of hominins.

5 Methods

5.1 Modern human and archaic genomes

Obtaining genomes

All genomic analysis was conducted using the GRCh37 (hg19) genome assembly and coordinates (www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/). Genomic variation within modern humans (MH) came from 1000 Genomes Project (1KGP), Phase 3 from Auton et al. [87]. All MH genomes were selected randomly from each subpopulation with a filter for females only to facilitate comparisons of the X chromosome. The 1KGP individuals used are listed in Table S1. Archaic genomes are from Prüfer et al. [1] (Altai), Prüfer et al. [2] (Vindija), Mafessoni et al. [3] (Chagyrskaya), and Meyer et al. [4] (Denisova).

Building individual genomes

We constructed full-length genomes for each MH or AH based upon the genotyping information in their respective vcf file. Given the difficulty of distinguishing heterozygous genotypes in the ancient DNA samples, we treated all individuals as if they were homozygous (pseudo-haploid). We built each individual genome using GATK's FastaAlternateReferenceMaker tool [95]. If an individual had an alternate allele (homozygous or heterozygous), we inserted it into the reference genome to create a pseudo-haploid, or "flattened" genome for each individual. This procedure is illustrated in step 1 of Fig. S1.

Accounting for missing data in the archaic genomes

Ancient DNA is both fragmented and degraded. These characteristics present challenges to both sequencing and alignment, resulting in gaps in coverage, particularly in genomic regions of low complexity. To account for this missing data, we "masked" all genomic regions lacking archaic genotyping information by reverting nucleotide states to the hg19 reference. For analyses that compared 3D genome organization between MHs and AHs, we masked both MH and AH genomes. This procedure is illustrated in steps 2-4 of Fig. S1. Archaic genome coverage is shown in Fig. S2. For analyses that only considered MHs (e.g. quantifying 3D genome variability across the genome in MHs), this masking procedure was not applied.

5.2 3D genome organization predictions with Akita

After the genomes were prepared, we input them into Akita for predictions using a 1 Mb sliding window (1,048,576 bp) overlapping by half (e.g. 524,288-1,572,864, 1,048,576-2,097,152, 1,572,864-2,621,440). Although Akita is trained simultaneously on Hi-C and Micro-C across five cell types in a multi-task framework to achieve greater accuracy, we focus on predictions in the highest resolution maps, human foreskin fibroblast (HFF). We note that the results are similar when considering other cell types (e.g. embryonic stem cells), likely because of limited cell-type-specific differences (Fig. S5). Akita considers the full window to generate predictions, but the resulting predictions are generated for only the middle 917,504 bp. Each contact map is a prediction for a single individual, and each cell represents physical 3D contacts at approximately 2 kb (2,048 bp) resolution. The value in each cell is $\log_2(obs/exp)$-scaled to account for the distance-dependent nature of chromatin contacts. Darker red pixels indicate more physical contacts and darker blue pixels denote fewer physical contacts. For all analyses, we only

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

considered windows with full (100%) coverage in the hg19 reference genome for a total of 4749 autosomal and 250 chromosome X windows. Fudenberg et al. [82] provides further details on the CNN architecture and training data used.

5.3 3D genome comparisons

After predictions were made on all 1 Mb windows for all individuals, we compared the resulting predictions using a variety of measures. All measures are scaled to indicate divergence: higher indicates more difference while lower indicates more similarity. In the maintext we transform the Spearman's rank correlation coefficient $(1-\rho)$ to describe 3D divergence. We consider measures based on the Pearson correlation coefficient $(1-r)$ and mean squared difference $(\frac{1}{n}\sum_{i=1}^{n}(x_i-y_i)^2)$ in Fig. S4. Percentiles of 3D divergence shown in Fig. 2A-B are calculated with reference to a universe of 4 AHs $\times$ 5 African MHs $\times$ 4999 genomic windows for a total of 99,980 comparisons. Figs. 4A,S9 averages the 3D divergence $(1-\rho)$ across all 4999 1 Mb windows (lower triangle) to compare to the average number of bp differences (after the masking procedure described above) in the same pair of individuals (upper triangle). Clustering is done with the "complete" (Farthest Point) method.

5.4 Sequence comparisons

Some analyses compare 3D genome divergence with sequence divergence. To calculate the sequence divergence between two individuals, we counted the proportion of bases at which the two individuals differ in the 1 Mb window. For comparisons of divergence when including AHs, we applied the same masking procedure as used to facilitate 3D genome comparisons (i.e. windows with missingness in AHs are filled with hg19 reference).

5.5 CTCF motif overlap

We consider how nucleotide differences in a window (between Neanderthal [Vindija] and an African MH [HG03105]) impacts 3D genome divergence in Figs. 4B,S10. We stratified variants by if they overlap a bound CTCF motif and their distance to TAD boundaries. CTCF motifs are from Vierstra et al. [96]. CTCF-bound open chromatin candidate cis-regulatory elements (cCREs) in the HFF cell type are from Abascal et al. [97]. TAD boundaries in the HFF cell type are from processed MicroC data from Akgol Oksuz et al. [98]. These annotations were all lifted over to hg19 [99]. A window was considered to have a CTCF-overlapping variant if an AH-MH nucleotide difference intersected a CTCF-bound HFF cCRE and a CTCF motif. Results were further stratified by varying levels of motif strength ("match_score" in the top 10th,25th, 50th, or any percentile), distance to TAD boundary (within 15 kb, 30 kb, or anywhere), and whether the CTCF motif overlap occurs in the middle 50% of the 1 Mb window or not.

5.6 Empirical distribution of expected 3D genome divergence

To compute the expected 3D divergence in a window given the observed sequence divergence, we generate genomes with shuffled nucleotide differences. We match these shuffled differences to the same number and tri-nucleotide context of the observed sequence differences between the Neanderthal (Vindija) and an African MH (HG03105) genome (Fig. 4C). Variants are not shuffled into masked regions of the genome. For each 1 Mb window of the genome ($N = 4999$) we generate 100 shuffled sequences. We calculate an empirical distribution of expected 3D divergence from comparing the contact maps of the shuffled sequences with the MH sequence. Finally, we compare the average expected 3D divergence from this distribution to the observed AH-MH 3D divergence.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5.7 AH-MH 3D divergent loci

Identifying loci

To identify loci with AH-MH 3D genome organization divergence, we compared the 3D contact map at each 1 Mb loci between each AH and 20 African MHs. To call a region as divergent, we required all 20 AH-MH comparisons to be more 3D divergent than all MH-MH comparisons (Fig. 3A). This identifies regions with consistent 3D differences between AHs and MHs while excluding regions with a large 3D diversity in modern humans. We also required the minimum AH-MH 3D divergence to be in the 80th percentile or greater of most 3D diverged (Fig, 2A, 3D divergence > 0.0042). Because 20 MHs do not capture the full MH genome diversity, it is possible that these criteria would still capture 3D patterns segregating in modern Africans that are not truly AH-MH diverged. Thus, we removed any windows where the 3D-modifying variant determined by *in silico* mutagenesis (below) was observed in 1KGP MHs if it was not introgressed (LD of $r^2 = 1$ with introgressed variants called by Browning et al. [93] or Vernot et al. [15]). For the counts of AH-MH divergent windows (Fig. 3B), we considered overlapping 1 Mb windows as a single observation. We summarize and report the AH-MH 3D divergent windows in Tables S2,S3 and a larger set of windows based on less conservative criteria in Table S4.

In silico mutagenesis

To identify the variant(s) contributing to the most prominent 3D differences in each identified AH-MH divergent window, *3D-modifying variants*, we use *in silico* mutagenesis. For example, for an Altai Neanderthal divergent window, we identify every bp difference that is unique to the Altai genome when compared to 20 African MH genomes. In the background of the MH (HG03105) genome, we insert each different Altai allele one-at-a-time. We then compare the resulting contact map between the original MH genome and the MH genome with each Altai allele. We then identify both the allele resulting in the largest 3D divergence and any other variants that contribute to a 3D divergence $>= 0.0042$ and term these "3D-modifying variants" (Table S2,S5).

Phenotype ontology enrichment

To test if AH-MH 3D-modifying variants are enriched near genes related to particular phenotypes we follow a procedure visually described in Fig. S7. 3D-modifying variants (above) are linked to genes in their TAD because this provides evidence of physical proximity. TADs are defined as regions between TAD boundaries as defined with MicroC data in HFF from Akgol Oksuz et al. [98] (lifted over to hg19). Genes are defined as the longest transcript from protein-coding genes (NM prefix) from NCBI RefSeq downloaded from the UCSC Table Browser [100]. Genes are linked to phenotypes from the Human Phenotype Ontology (HPO) and GWAS Catalog 2019 downloaded from Enrichr [90–92]. Annotations are further grouped into phenotypic systems via system-level annotations from Gene ORGANizer [101] and manual curation. HPO largely considers rare disease annotations and has 1779 terms with 3,096 genes annotated [88]. The GWAS Catalog largely considers common disease annotations and has 737 terms with 19,378 genes annotated [89]. Through this procedure, we counted the number of ontology terms linked to the set of 3D-modifying variants. We considered 3 different sets, those shared (intersect) by all Neanderthals (Fig. 3), those in any Neanderthal (union), and those in Denisova (Fig. S8, Table S2).

We test enrichment for ontology terms linked to at least one 3D-modifying variant. While the annotations are downloaded from Enrichr, we did enrichment analyses with a more appropriate null. For each set, we shuffle the observed 3D-modifying variants into the background genome. We defined the background genome as any place where a 3D-modifying variant could have been identified (i.e. regions with full coverage in modern humans used for Akita predictions). We then use the same procedure (Fig. S7 to link the shuffled variants to genes and then ontology terms. We repeat this shuffle 500,000 times to create an empirical distribution for how many times we would observe each annotation under the null. We used these distributions to calculate an enrichment and P-value for each ontology term. The FDR-corrected significance level was determined empirically using these null observations (a subset

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of $n = 50,000$). We select the highest p-value threshold that led to a $V/R < Q$ where V is the mean number of expected false discoveries and R is the observed discoveries (which includes both true and false positives).

5.8 Relationship between 3D genome organization and introgression

3D genome variability

To consider how 3D organization may have constrained where we observe introgression in the genome, we calculated 3D genome variability across the genome in MHs. Because we are not comparing these predictions with AH 3D genome organization, we did not mask the genomes before 3D genome predictions (above). In the same 1 Mb sliding windows across the genome, we predicted the contact maps for 20 modern Africans (because they have no or very little introgression). For each window, we calculate the 3D genome divergence between all 190 $\binom{20}{2}$ pairs of contact maps. We then computed the "3D genome variability" by taking the mean of these 190 divergences for each 1 Mb window across the genome. High 3D genome variability indicates a high average pairwise 3D divergence (i.e. diversity of 3D organization), while low 3D genome variability indicates low pairwise 3D divergence (i.e. similar 3D organization across all individuals).

Genomic windows with evidence of introgression

To define genomic regions with Neanderthal ancestry we used "segments" identified by Browning et al. [93] using Sprime, a heuristic scoring strategy that compares high-LD regions in a target admixed population (i.e., Europeans) with an unadmixed outgroup (i.e., Africans) to identify putatively introgressed regions. We considered a set of Sprime-identified segments shared (intersection) among East Asians (EAS), EUR, and SAS. We repeat the analysis using a more stringent subset of Sprime segments that (1) have at least 30 putatively introgressed variants that could be compared to the Altai Neanderthal genome and (2) had a match rate of at least 30% to the Altai Neanderthal allele (Neanderthal filter). We also considered the introgressed Neanderthal haplotypes previously identified by Vernot et al. [15] identified using the S* statistic. Finally, we consider introgressed segments unique to a single population (EAS, EUR, or SAS). Because these introgression calls only consider autosomes, we do not use the X chromosome for these analyses. Results from these sets of Neanderthal ancestry are in Figs. 5,S11,S12 and Tables S7,S8.

In the main text (Fig. 5), we compare the 3D genome variability between 1 Mb windows with no introgression (0%) versus windows where at least 70% of the bp have evidence of introgression. Other thresholds are shown in Fig. S11.

Predicting the amount of introgression

To test if 3D genome variability can be uniquely informative to predict tolerance of introgression, we conducted a simple linear regression. We predict the amount of introgression in a 1 Mb window while conditioning on the amount of sequence variability in a window. $Y = B_0 + B_1 X_{\text{3D variability}} + B_2 X_{\text{Sequence Variability}}$, where Y is the proportion of the 1 Mb window with evidence of introgression defined using the previously described sets of Neanderthal ancestry. For comparison, we also conducted some regressions where Y was modeled from only 3D variability or sequence variability alone. Results from these models are in Figs. 5B,S12, Tables S7,S8.

5.9 Individual-level introgression calls

We used introgression calls in 1KGP individuals from Chen et al. [94], which applied IBDmix with the Altai Neanderthal genome to identify introgressed segments in MHs. We identified windows with AH-MH divergence with evidence of introgression by intersecting with the introgression calls.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We also test the relationship between the amount of introgression an individual has and their 3D divergence from AHs. For each window, we compare the amount of introgression (% of bp) for an individual in a 1 Mb window with that individual's 3D divergence from Neanderthals. We do this analysis for 15 Eurasians across 4,749 1 Mb autosomal windows (total $n = 71,235$). In Fig. 6C we compare Eurasians to the Vindija Neanderthal 3D genome and in Fig. S14 we compare to Altai and Chagyrskaya. We also repeat the analysis removing windows with no evidence (0% bp) of introgression.

5.10 eQTL and PheWAS analysis

eQTL analysis and plots were generated using the Genotype-Tissue Expression (GTEx) Project (V8 release) Portal (lifted over to hg19) [42]. PheWAS results are from the GWAS Atlas and consider 4756 traits [102]. Allele frequencies come from 1KGP Phase 3 [87].

5.11 Examples

The examples visualized in Figs. 3,6 are annotated using the UCSC genome browser [99]. They were each manually zoomed to highlight the regions of interest. We use ENCODE open chromatin candidate cis-regulatory elements (cCREs) [97] to highlight promoters (promoter-like signature, pink) and enhancers (proximal [orange] and distal [yellow] enhancer-like signature) combined from all cell types downloaded from the UCSC table browser (lifted over to hg19) [100]. We use Transcription Factor (TF) ChIP-seq Clusters (130 cell types) from ENCODE 3 [103, 104] downloaded from UCSC table browser [100]. We show the motif sequence logo with reference to the positive strand of hg19.

5.12 Data analysis and figure generation

The datasets we generated are available in the GitHub repository "neanderthal-3d-genome" available here `https://github.com/emcarthur/neanderthal-3D-genome/` which will be formally cited and versioned upon publication.

All genomic coordinates and analysis refer to Homo sapiens (human) genome assembly GRCh37 (hg19), unless otherwise specified. All P values are two-tailed, unless otherwise specified. All measures of central tendencies are means, unless otherwise specified. Data and statistical analyses were conducted using Python 3.6.10 (Anaconda distribution), Jupyter Notebook, BedTools v2.26, and PLINK 1.9 [105, 106]. Figure generation was significantly aided by Matplotlib, Seaborn, and Inkscape [107–109].

5.13 Data availability

The publicly available data used for analysis are available in the following repositories. MH genome vcfs are from 1000 Genomes Project (1KGP) (`ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000_genomes_project/release/20190312_biallelic_SNV_and_INDEL/`[87]. Archaic genotypes are from the following repositories: Altai Neanderthal [1] (`ftp.eva.mpg.de/neandertal/Vindija/VCF/Altai/`), Denisova (`ftp.eva.mpg.de/neandertal/Vindija/VCF/Denisova/`) [4], Vindija Neanderthal [2] (`ftp.eva.mpg.de/neandertal/Vindija/VCF/Vindija33.19/`), and Chagyrskaya Neanderthal [3] (`ftp.eva.mpg.de/neandertal/Chagyrskaya/VCF/`). Introgressed variants and segments are from Sprime Version 1 (`https://data.mendeley.com/datasets/y7hyt83vxr`)[93]. An alternative set of introgressed variants and segments are from S* (`https://drive.google.com/drive/folders/0B9Pc7_zItMCVWUp6bWtXc2xJVkk?resourcekey=0-Cj8G4QYndXQLVIGPoWKUjQ`)[15]]. Individual level 1KGP introgression calls are from the Akey Lab (`https://drive.google.com/drive/folders/1mDQaDFS-j22Eim5_y7LAsTTNt5GWsoow`)[94].

CTCF motifs are from genome-wide motif scans v1.0 (`https://resources.altius.org/~jvierstra/projects/motif-clustering/releases/v1.0/`, all models in the CTCF archetype motif cluster, lifted-over to hg19)[96], CTCF-bound open chromatin candidate cis-regulatory elements (cCREs) in the HFF cell type (`https://screen.encodeproject.org/` > Downloads > by cell type > HFF-Myc male

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

newborn originated from foreskin fibroblast, lifted-over to hg19)[97], TAD boundaries in the HFF cell type are from processed MicroC data available at the 4D nucleome data portal (https://data.4dnucleome.org/experiment-set-replicates/4DNES9X112GZ/, lifted-over to hg19)[98]. RefSeq genes, TF ChIP-seq Clusters, enhancer and promoter cCREs are downloaded from the UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables)[100]. Gene ontology annotations are downloaded from Enrichr (https://maayanlab.cloud/Enrichr/#libraries)[90–92]. System-level groupings of disease ontology terms were aided by Gene ORGANizer annotations(http://geneorganizer.huji.ac.il/downloads/)[101]. eQTL data is from the GTEx Portal (https://www.gtexportal.org/, lifted-over to hg19)[42]. PheWAS results are from the GWAS Atlas (https://atlas.ctglab.nl/)[102].

5.14 Code availability

Akita is in the "basenji" GitHub repository available here https://github.com/calico/basenji/tree/master/manuscripts/akita [82]. The "neanderthal-3d-genome" GitHub repository (above) contains a Jupyter notebook with custom code used for data analysis and all figure generation.

5.15 Acknowledgements

The authors would like to thank Colby Tubbs, Mary Lauren Benton, Douglas Ruderfer, Colin Brand, and other members of the Capra and Pollard labs for helpful discussions and manuscript comments. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education (ACCRE) at Vanderbilt University, Nashville, TN.

5.16 Funding sources

This work was supported by the National Institutes of Health (NIH) General Medical Sciences award R35GM127087 to JAC, NIH National Human Genome Research Institute award F30HG011200 to EM, and T32GM007347. GF is supported by R35 GM143116-01. The funding bodies had no role in the design of the study and collection, analysis, or interpretation of data, or in writing the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

5.17 Authors' contributions

EM, DCR, ENG, GF, MP, KK, KSP, JAC conceived and designed the work presented here. EM and DCR conducted all the analyses. EM, DCR, ENG, GF, KSP, JAC interpreted the results, drafted the work, and substantively revised the manuscript. EM, DCR, ENG, GF, MP, KK, KSP, JAC have approved the submitted version and have agreed to be accountable for their contributions.

5.18 Competing interests

The authors declare no competing interests.

References

[1] K. Prüfer, F. Racimo, N. Patterson, et al. "The complete genome sequence of a Neanderthal from the Altai Mountains". In: *Nature* 505.7481 (2014), pp. 43–49. ISSN: 00280836. DOI: 10.1038/nature12886.

[2] K. Prüfer, C. De Filippo, S. Grote, et al. "A high-coverage Neandertal genome from Vindija Cave in Croatia". In: *Science* 358.6363 (2017), pp. 655–658. ISSN: 10959203. DOI: 10.1126/science.aao1887.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[3] F. Mafessoni, S. Grote, C. D. Filippo, et al. "A high-coverage neandertal genome from chagyrskaya cave". In: *Proceedings of the National Academy of Sciences of the United States of America* 117.26 (June 2020), pp. 15132–15136. ISSN: 10916490. DOI: 10.1073/pnas.2004944117.

[4] M. Meyer, M. Kircher, M. T. Gansauge, et al. "A high-coverage genome sequence from an archaic Denisovan individual". In: *Science* 338.6104 (Oct. 2012), pp. 222–226. ISSN: 10959203. DOI: 10.1126/science.1224344.

[5] M. Kuhlwilm and C. Boeckx. "A catalog of single nucleotide changes distinguishing modern humans from archaic hominins". In: *Scientific Reports* 9.1 (June 2019), pp. 1–14. ISSN: 20452322. DOI: 10.1038/s41598-019-44877-x.

[6] A. B. Wolf and J. M. Akey. "Outstanding questions in the study of archaic hominin admixture". In: *PLoS Genetics* 14.5 (May 2018), e1007349. ISSN: 15537404. DOI: 10.1371/journal.pgen.1007349.

[7] M. C. King and A. C. Wilson. "Evolution at two levels in humans and chimpanzees". In: *Science* 188.4184 (1975), pp. 107–116. ISSN: 00368075. DOI: 10.1126/science.1090005.

[8] G. A. Wray. "The evolutionary significance of cis-regulatory mutations". In: *Nature Reviews Genetics* 8.3 (Mar. 2007), pp. 206–216. ISSN: 14710056. DOI: 10.1038/nrg2063.

[9] S. B. Carroll. *Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution*. 2008. DOI: 10.1016/j.cell.2008.06.030.

[10] P. J. Wittkopp and G. Kalay. *Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence*. 2012. DOI: 10.1038/nrg3095.

[11] R. C. McCoy, J. Wakefield, and J. M. Akey. "Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression". In: *Cell* 168.5 (2017), 916–927.e12. ISSN: 10974172. DOI: 10.1016/j.cell.2017.01.038.

[12] M. Dannemann, K. Prüfer, and J. Kelso. "Functional implications of Neandertal introgression in modern humans". In: *Genome Biology* 18.1 (Apr. 2017), pp. 1–11. ISSN: 1474760X. DOI: 10.1186/s13059-017-1181-7.

[13] L. L. Colbran, E. R. Gamazon, D. Zhou, P. Evans, N. J. Cox, and J. A. Capra. "Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences". In: *Nature Ecology and Evolution* 3.11 (2019), pp. 1598–1606. ISSN: 2397334X. DOI: 10.1038/s41559-019-0996-x.

[14] M. Silvert, L. Quintana-Murci, and M. Rotival. "Impact and Evolutionary Determinants of Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation". In: *American Journal of Human Genetics* 104.6 (June 2019), pp. 1241–1250. ISSN: 15376605. DOI: 10.1016/j.ajhg.2019.04.016.

[15] B. Vernot, S. Tucci, J. Kelso, et al. "Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals". In: *Science* 352.6282 (2016), pp. 235–239. ISSN: 10959203. DOI: 10.1126/science.aad9416.

[16] S. Sankararaman, S. Mallick, N. Patterson, and D. Reich. "The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans". In: *Current Biology* 26.9 (May 2016), pp. 1241–1247. ISSN: 09609822. DOI: 10.1016/j.cub.2016.03.037.

[17] M. Dannemann and J. Kelso. "The Contribution of Neanderthals to Phenotypic Variation in Modern Humans". In: *American Journal of Human Genetics* 101.4 (2017), pp. 578–589. ISSN: 15376605. DOI: 10.1016/j.ajhg.2017.09.010.

[18] C. N. Simonti, B. Vernot, L. Bastarache, et al. "The phenotypic legacy of admixture between modern humans and Neandertals". In: *Science* 351.6274 (Feb. 2016), pp. 737–741. ISSN: 10959203. DOI: 10.1126/science.aad2149.

[19] E. McArthur, D. C. Rinker, and J. A. Capra. "Quantifying the contribution of Neanderthal introgression to the heritability of complex traits". In: *Nature Communications* 12.1 (July 2021), p. 2020.06.08.140087. ISSN: 20411723. DOI: 10.1038/s41467-021-24582-y.

[20] M. Dannemann. "The Population-Specific Impact of Neandertal Introgression on Human Disease". In: *Genome biology and evolution* 13.1 (Jan. 2021). ISSN: 17596653. DOI: 10.1093/gbe/evaa250.

[21] D. Koller, F. R. Wendt, G. A. Pathak, A. D. Lillo, and F. De. "The impact of evolutionary processes in shaping the genetics of complex traits in East Asia and Europe : a specific contribution from Denisovan and Neanderthal introgression". In: *bioRxiv* 1.203 (Aug. 2021), p. 2021.08.12.456138. DOI: 10.1101/2021.08.12.456138.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[22] L. Abi-Rached, M. J. Jobin, S. Kulkarni, et al. "The shaping of modern human immune systems by multiregional admixture with archaic humans". In: *Science* 334.6052 (2011), pp. 89–94. ISSN: 10959203. DOI: 10.1126/science.1209202.

[23] F. L. Mendez, J. C. Watkins, and M. F. Hammer. "A haplotype at STAT2 introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea". In: *American Journal of Human Genetics* 91.2 (2012), pp. 265–274. ISSN: 00029297. DOI: 10.1016/j.ajhg.2012.06.015.

[24] A. J. Sams, A. Dumaine, Y. Nédélec, V. Yotova, C. Alfieri, J. E. Tanner, P. W. Messer, and L. B. Barreiro. *Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans*. 2016. DOI: 10.1186/s13059-016-1098-6.

[25] M. Dannemann, A. M. Andrés, and J. Kelso. "Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors". In: *American Journal of Human Genetics* 98.1 (2016), pp. 22–33. ISSN: 15376605. DOI: 10.1016/j.ajhg.2015.11.015.

[26] M. Deschamps, G. Laval, M. Fagny, Y. Itan, L. Abel, J. L. Casanova, E. Patin, and L. Quintana-Murci. "Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes". In: *American Journal of Human Genetics* 98.1 (2016), pp. 5–21. ISSN: 15376605. DOI: 10.1016/j.ajhg.2015.11.014.

[27] H. Quach, M. Rotival, J. Pothlichet, et al. "Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations". In: *Cell* 167.3 (2016), 643–656.e17. ISSN: 10974172. DOI: 10.1016/j.cell.2016.09.024.

[28] Y. Nédélec, J. Sanz, G. Baharian, et al. "Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens". In: *Cell* 167.3 (2016), 657–669.e21. ISSN: 10974172. DOI: 10.1016/j.cell.2016.09.025.

[29] D. Enard and D. A. Petrov. "Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans". In: *Cell* 175.2 (2018), 360–371.e13. ISSN: 10974172. DOI: 10.1016/j.cell.2018.08.034.

[30] B. Vernot and J. M. Akey. "Resurrecting surviving Neandertal lineages from modern human genomes". In: *Science* 343.6174 (2014), pp. 1017–1021. ISSN: 10959203. DOI: 10.1126/science.1245938.

[31] S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich. "The genomic landscape of Neanderthal ancestry in present-day humans". In: *Nature* 507.7492 (2014), pp. 354–357. ISSN: 14764687. DOI: 10.1038/nature12961.

[32] Q. Ding, Y. Hu, S. Xu, J. Wang, and L. Jin. "Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in east asians". In: *Molecular Biology and Evolution* 31.3 (2014), pp. 683–695. ISSN: 15371719. DOI: 10.1093/molbev/mst260.

[33] F. Racimo, D. Marnetto, and E. Huerta-Sánchez. "Signatures of archaic adaptive introgression in present-day human populations". In: *Molecular Biology and Evolution* 34.2 (2017), pp. 296–317. ISSN: 15371719. DOI: 10.1093/molbev/msw216.

[34] F. Racimo, D. Gokhman, M. Fumagalli, A. Ko, T. Hansen, I. Moltke, A. Albrechtsen, L. Carmel, E. Huerta-Sanchez, and R. Nielsen. "Archaic adaptive introgression in TBX15/WARS2". In: *Molecular Biology and Evolution* 34.3 (2017), pp. 509–524. ISSN: 15371719. DOI: 10.1093/molbev/msw283.

[35] E. E. Khrameeva, K. Bozek, L. He, et al. "Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans". In: *Nature Communications* 5 (2014). ISSN: 20411723. DOI: 10.1038/ncomms4584.

[36] A. Gouy, L. Excoffier, and R. Nielsen. "Polygenic Patterns of Adaptive Introgression in Modern Humans Are Mainly Shaped by Response to Pathogens". In: *Molecular Biology and Evolution* 37.5 (May 2020), pp. 1420–1433. ISSN: 15371719. DOI: 10.1093/molbev/msz306.

[37] E. Huerta-Sánchez, X. Jin, Asan, et al. "Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA". In: *Nature* 512.7513 (2014), pp. 194–197. ISSN: 14764687. DOI: 10.1038/nature13408.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[38] P. Gunz, A. K. Tilot, K. Wittfeld, et al. "Neandertal Introgression Sheds Light on Modern Human Endocranial Globularity". In: *Current Biology* 29.1 (Jan. 2019), 120–127.e5. ISSN: 09609822. DOI: 10.1016/j.cub.2018.10.065.

[39] M. Hajdinjak, F. Mafessoni, L. Skov, et al. "Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry". In: *Nature* 592.7853 (Apr. 2021), pp. 253–257. ISSN: 14764687. DOI: 10.1038/s41586-021-03335-3.

[40] M. Petr, S. Pääbo, J. Kelso, and B. Vernot. "Limits of long-term selection against Neandertal introgression". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.5 (Jan. 2019), pp. 1639–1644. ISSN: 10916490. DOI: 10.1073/pnas.1814338116.

[41] N. Telis, R. Aguilar, and K. Harris. "Selection against archaic hominin genetic variation in regulatory regions". In: *Nature Ecology and Evolution* 4.11 (Aug. 2020), pp. 1558–1566. ISSN: 2397334X. DOI: 10.1038/s41559-020-01284-0.

[42] J. Lonsdale, J. Thomas, M. Salvatore, et al. "The Genotype-Tissue Expression (GTEx) project". In: *Nature Genetics* 45.6 (June 2013), pp. 580–585. ISSN: 10614036. DOI: 10.1038/ng.2653.

[43] D. Gokhman, E. Lavi, K. Prüfer, M. F. Fraga, J. A. Riancho, J. Kelso, S. Pääbo, E. Meshorer, and L. Carmel. "Reconstructing the DNA methylation maps of the neandertal and the Denisovan". In: *Science* 344.6183 (May 2014), pp. 523–527. ISSN: 10959203. DOI: 10.1126/science.1250368.

[44] D. Batyrev, E. Lapid, L. Carmel, and E. Meshorer. "Predicted Archaic 3D Genome Organization Reveals Genes Related to Head and Spinal Cord Separating Modern from Archaic Humans". In: *Cells* 9.1 (Dec. 2019). ISSN: 20734409. DOI: 10.3390/cells9010048.

[45] T. Cremer and C. Cremer. *Chromosome territories, nuclear architecture and gene regulation in mammalian cells*. 2001. DOI: 10.1038/35066075.

[46] G. Cavalli and T. Misteli. *Functional implications of genome topology*. 2013. DOI: 10.1038/nsmb.2474.

[47] G. Duggal, H. Wang, and C. Kingsford. "Higher-order chromatin domains link eQTLs with the expression of far-away genes". In: *Nucleic Acids Research* 42.1 (2014), pp. 87–96. ISSN: 03051048. DOI: 10.1093/nar/gkt857.

[48] F. L. Le Dily, D. Baù, A. Pohl, et al. "Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation". In: *Genes and Development* 28.19 (2014), pp. 2151–2162. ISSN: 15495477. DOI: 10.1101/gad.241422.114.

[49] J. A. Beagan and J. E. Phillips-Cremins. "On the existence and functionality of topologically associating domains". In: *Nature Genetics* 52.1 (Jan. 2020), pp. 8–16. ISSN: 15461718. DOI: 10.1038/s41588-019-0561-1.

[50] T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, and R. Tjian. "Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1". In: *bioRxiv* (July 2021), p. 2021.07.14.452365. DOI: 10.1101/2021.07.14.452365.

[51] B. Baur, J. Schreiber, J. Shin, S. Zhang, Y. Zhang, M. Manjunath, J. S. Song, W. S. Noble, and S. Roy. "Leveraging epigenomes and three-dimensional genome organization for interpreting regulatory variation". In: *bioRxiv* (Aug. 2021), p. 2021.08.29.458098. DOI: 10.1101/2021.08.29.458098.

[52] P. J. Batut, X. Y. Bing, Z. Sisco, J. Raimundo, M. Levo, and M. S. Levine. "Genome organization controls transcriptional dynamics during development." In: *Science (New York, N.Y.)* 375.6580 (Feb. 2022), pp. 566–570. ISSN: 1095-9203. DOI: 10.1126/science.abi7178.

[53] G. Fudenberg, G. Getz, M. Meyerson, and L. A. Mirny. "High order chromatin architecture shapes the landscape of chromosomal alterations in cancer". In: *Nature Biotechnology* 29.12 (2011), pp. 1109–1113. ISSN: 10870156. DOI: 10.1038/nbt.2049.

[54] D. Hnisz, A. S. Weintrau, D. S. Day, et al. "Activation of proto-oncogenes by disruption of chromosome neighborhoods". In: *Science* 351.6280 (2016), pp. 1454–1458. ISSN: 10959203. DOI: 10.1126/science.aad9024.

[55] K. J. Meaburn, P. R. Gudla, S. Khan, S. J. Lockett, and T. Misteli. "Disease-specific gene repositioning in breast cancer". In: *Journal of Cell Biology* 187.6 (2009), pp. 801–812. ISSN: 00219525. DOI: 10.1083/jcb.200909127.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[56] T. Misteli. *Higher-order genome organization in human disease.* 2010. DOI: 10.1101/cshperspect.a000794.

[57] B. Bonev, N. Mendelson Cohen, Q. Szabo, et al. "Multiscale 3D Genome Rewiring during Mouse Neural Development". In: *Cell* 171.3 (Oct. 2017), 557–572.e24. ISSN: 10974172. DOI: 10.1016/j.cell.2017.09.043.

[58] S. E. de Bruijn, A. Fiorentino, D. Ottaviani, et al. "Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa". In: *American Journal of Human Genetics* 107.5 (Oct. 2020), pp. 802–814. ISSN: 15376605. DOI: 10.1016/j.ajhg.2020.09.002.

[59] B. D. Pope, T. Ryba, V. Dileep, et al. "Topologically associating domains are stable units of replication-timing regulation". In: *Nature* 515.7527 (Nov. 2014), pp. 402–405. ISSN: 14764687. DOI: 10.1038/nature13986.

[60] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. "Capturing chromosome conformation". In: *Science* 295.5558 (2002), pp. 1306–1311. ISSN: 00368075. DOI: 10.1126/science.1067799.

[61] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome". In: *Science* 326.5950 (Oct. 2009), pp. 289–293. ISSN: 00368075. DOI: 10.1126/science.1181369.

[62] J. R. Dixon, D. U. Gorkin, and B. Ren. "Chromatin Domains: The Unit of Chromosome Organization". In: *Molecular Cell* 62.5 (June 2016), pp. 668–680. ISSN: 10974164. DOI: 10.1016/j.molcel.2016.05.018.

[63] N. Krietenstein, S. Abraham, S. V. Venev, et al. "Ultrastructural Details of Mammalian Chromosome Architecture". In: *Molecular Cell* 78.3 (May 2020), 554–565.e7. ISSN: 10974164. DOI: 10.1016/j.molcel.2020.03.003.

[64] T. H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq. "Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding". In: *Molecular Cell* 78.3 (May 2020), 539–553.e8. ISSN: 10974164. DOI: 10.1016/j.molcel.2020.03.002.

[65] L. Vian, A. Pękowska, S. S. Rao, et al. "The Energetics and Physiological Impact of Cohesin Extrusion". In: *Cell* 173.5 (May 2018), 1165–1178.e20. ISSN: 10974172. DOI: 10.1016/j.cell.2018.03.072.

[66] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny. "Formation of Chromosomal Domains by Loop Extrusion". In: *Cell Reports* 15.9 (May 2016), pp. 2038–2049. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.04.085.

[67] K. Kraft, A. Magg, V. Heinrich, et al. "Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations". In: *Nature Cell Biology* 21.3 (Feb. 2019), pp. 305–310. ISSN: 14764679. DOI: 10.1038/s41556-019-0273-x.

[68] S. Oh, J. Shao, J. Mitra, et al. "Enhancer release and retargeting activates disease-susceptibility genes". In: *Nature* 595.7869 (May 2021), pp. 735–740. ISSN: 14764687. DOI: 10.1038/s41586-021-03577-1.

[69] D. G. Lupiáñez, K. Kraft, V. Heinrich, et al. "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions". In: *Cell* 161.5 (May 2015), pp. 1012–1025. ISSN: 10974172. DOI: 10.1016/j.cell.2015.04.004.

[70] W. W. Greenwald, H. Li, P. Benaglio, et al. "Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression". In: *Nature Communications* 10.1 (Dec. 2019), p. 1054. ISSN: 20411723. DOI: 10.1038/s41467-019-08940-5.

[71] M. Spielmann, D. G. Lupiáñez, and S. Mundlos. "Structural variation in the 3D genome". In: *Nature Reviews Genetics* 19.7 (July 2018), pp. 453–467. ISSN: 14710064. DOI: 10.1038/s41576-018-0007-0.

[72] D. U. Gorkin, Y. Qiu, M. Hu, et al. "Common DNA sequence variation influences 3-dimensional conformation of the human genome". In: *Genome Biology* 20.1 (Nov. 2019), pp. 1–25. ISSN: 1474760X. DOI: 10.1186/s13059-019-1855-4.

[73] G. Fudenberg and K. S. Pollard. "Chromatin features constrain structural variation across evolutionary timescales". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.6 (Feb. 2019), pp. 2175–2180. ISSN: 10916490. DOI: 10.1073/pnas.1808631116.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[74] E. McArthur and J. A. Capra. "Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability". In: *American Journal of Human Genetics* 108.2 (Feb. 2021), pp. 269–283. ISSN: 15376605. DOI: 10.1016/j.ajhg.2021.01.001.

[75] Y. Liao, X. Zhang, M. Chakraborty, and J. J. Emerson. "Topologically associating domains and their role in the evolution of genome structure and function in Drosophila". In: *Genome Research* 31.3 (Mar. 2021), pp. 397–410. ISSN: 15495469. DOI: 10.1101/GR.266130.120.

[76] J. Krefting, M. A. Andrade-Navarro, and J. Ibn-Salem. "Evolutionary stability of topologically associating domains is associated with conserved gene regulation". In: *BMC Biology* 16.1 (Dec. 2018), p. 87. ISSN: 17417007. DOI: 10.1186/s12915-018-0556-x.

[77] Y. Yang, Y. Zhang, B. Ren, J. R. Dixon, and J. Ma. "Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF". In: *Cell Systems* 8.6 (June 2019), 494–505.e14. ISSN: 24054720. DOI: 10.1016/j.cels.2019.05.011.

[78] I. E. Eres, K. Luo, C. J. Hsiao, L. E. Blake, and Y. Gilad. "Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates". In: *PLoS Genetics* 15.7 (July 2019). Ed. by H. S. Malik, e1008278. ISSN: 15537404. DOI: 10.1371/journal.pgen.1008278.

[79] M. J. Rowley and V. G. Corces. "Organizational principles of 3D genome architecture". In: *Nature Reviews Genetics* 19.12 (Dec. 2018), pp. 789–800. ISSN: 14710064. DOI: 10.1038/s41576-018-0060-8.

[80] L.-H. Chang, S. Ghosh, A. Papale, et al. "A complex CTCF binding code defines TAD boundary structure and function". In: *bioRxiv* (Apr. 2021), p. 2021.04.15.440007. DOI: 10.1101/2021.04.15.440007.

[81] P. Belokopytova and V. Fishman. "Predicting Genome Architecture: Challenges and Solutions". In: *Frontiers in Genetics* 11 (Jan. 2021), p. 1776. ISSN: 16648021. DOI: 10.3389/fgene.2020.617202.

[82] G. Fudenberg, D. R. Kelley, and K. S. Pollard. "Predicting 3D genome folding from DNA sequence with Akita". In: *Nature Methods* 17.11 (Oct. 2020), pp. 1111–1117. ISSN: 15487105. DOI: 10.1038/s41592-020-0958-x.

[83] R. Schwessinger, M. Gosden, D. Downes, R. C. Brown, A. M. Oudelaar, J. Telenius, Y. W. Teh, G. Lunter, and J. R. Hughes. "DeepC: predicting 3D genome folding using megabase-scale transfer learning". In: *Nature Methods* 17.11 (Oct. 2020), pp. 1118–1124. ISSN: 15487105. DOI: 10.1038/s41592-020-0960-3.

[84] J. Zhou. "Sequence-based modeling of genome 3D architecture from kilobase to chromosome-scale". In: *bioRxiv* (May 2021), p. 2021.05.19.444847. DOI: 10.1101/2021.05.19.444847.

[85] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. "Topological domains in mammalian genomes identified by analysis of chromatin interactions". In: *Nature* 485.7398 (Apr. 2012), pp. 376–380. ISSN: 00280836. DOI: 10.1038/nature11082.

[86] M. Vietri Rudan, C. Barrington, S. Henderson, C. Ernst, D. T. Odom, A. Tanay, and S. Hadjur. "Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture". In: *Cell Reports* 10.8 (Mar. 2015), pp. 1297–1309. ISSN: 22111247. DOI: 10.1016/j.celrep.2015.02.004.

[87] A. Auton, G. R. Abecasis, D. M. Altshuler, et al. *A global reference for human genetic variation.* 2015. DOI: 10.1038/nature15393.

[88] S. Köhler, M. Gargano, N. Matentzoglu, et al. "The human phenotype ontology in 2021". In: *Nucleic Acids Research* 49.D1 (Jan. 2021), pp. D1207–D1217. ISSN: 13624962. DOI: 10.1093/nar/gkaa1043.

[89] A. Buniello, J. A. Macarthur, M. Cerezo, et al. "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019". In: *Nucleic Acids Research* 47.D1 (Jan. 2019), pp. D1005–D1012. ISSN: 13624962. DOI: 10.1093/nar/gky1120.

[90] E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan. "Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool". In: *BMC Bioinformatics* 14.1 (Apr. 2013), pp. 1–14. ISSN: 14712105. DOI: 10.1186/1471-2105-14-128.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[91] M. V. Kuleshov, M. R. Jones, A. D. Rouillard, et al. "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update". In: *Nucleic acids research* 44.W1 (July 2016), W90–W97. ISSN: 13624962. DOI: 10.1093/nar/gkw377.

[92] Z. Xie, A. Bailey, M. V. Kuleshov, et al. "Gene Set Knowledge Discovery with Enrichr". In: *Current Protocols* 1.3 (Mar. 2021), e90. ISSN: 26911299. DOI: 10.1002/cpz1.90.

[93] S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, and J. M. Akey. "Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture". In: *Cell* 173.1 (2018), 53–61.e9. ISSN: 10974172. DOI: 10.1016/j.cell.2018.02.031.

[94] L. Chen, A. B. Wolf, W. Fu, L. Li, and J. M. Akey. "Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals". In: *Cell* 180.4 (Feb. 2020), 677–687.e16. ISSN: 10974172. DOI: 10.1016/j.cell.2020.01.012.

[95] G. A. Van der Auwera and B. O'Connor. *Genomics in the cloud : using Docker, GATK, and WDL in Terra*. 2020. ISBN: 1-4919-7518-0.

[96] J. Vierstra, J. Lazar, R. Sandstrom, et al. "Global reference mapping of human transcription factor footprints". In: *Nature* 583.7818 (July 2020), pp. 729–736. ISSN: 14764687. DOI: 10.1038/s41586-020-2528-x.

[97] F. Abascal, R. Acosta, N. J. Addleman, et al. "Expanded encyclopaedias of DNA elements in the human and mouse genomes". In: *Nature* 583.7818 (July 2020), pp. 699–710. ISSN: 14764687. DOI: 10.1038/s41586-020-2493-4.

[98] B. Akgol Oksuz, L. Yang, S. Abraham, et al. "Systematic evaluation of chromosome conformation capture assays". In: *Nature Methods* 18.9 (Sept. 2021), pp. 1046–1055. ISSN: 15487105. DOI: 10.1038/s41592-021-01248-7.

[99] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, Haussler, and David. "The Human Genome Browser at UCSC". In: *Genome Research* 12.6 (June 2002), pp. 996–1006. ISSN: 1088-9051. DOI: 10.1101/gr.229102.

[100] D. Karolchik, A. S. Hinricks, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, and W. J. Kent. "The UCSC table browser data retrieval tool". In: *Nucleic Acids Research* 32.DATABASE ISS. (Jan. 2004). ISSN: 03051048. DOI: 10.1093/nar/gkh103.

[101] D. Gokhman, G. Kelman, A. Amartely, G. Gershon, S. Tsur, and L. Carmel. "Gene ORGANizer: Linking genes to the organs they affect". In: *Nucleic Acids Research* 45.W1 (July 2017), W138–W145. ISSN: 13624962. DOI: 10.1093/nar/gkx302.

[102] K. Watanabe, S. Stringer, O. Frei, M. Umićević Mirkov, C. de Leeuw, T. J. Polderman, S. van der Sluis, O. A. Andreassen, B. M. Neale, and D. Posthuma. "A global overview of pleiotropy and genetic architecture in complex traits". In: *Nature Genetics* 51.9 (2019), pp. 1339–1348. ISSN: 15461718. DOI: 10.1038/s41588-019-0481-0.

[103] J. Wang, J. Zhuang, S. Iyer, et al. "Factorbook.org: A Wiki-based database for transcription factor-binding data generated by the ENCODE consortium". In: *Nucleic Acids Research* 41.D1 (Jan. 2013), pp. D171–D176. ISSN: 03051048. DOI: 10.1093/nar/gks1221.

[104] J. Wang, J. Zhuang, S. Iyer, et al. "Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors". In: *Genome Research* 22.9 (Sept. 2012), pp. 1798–1812. ISSN: 10889051. DOI: 10.1101/gr.139105.112.

[105] A. R. Quinlan and I. M. Hall. "BEDTools: A flexible suite of utilities for comparing genomic features". In: *Bioinformatics* 26.6 (2010), pp. 841–842. ISSN: 13674803. DOI: 10.1093/bioinformatics/btq033.

[106] S. Purcell, B. Neale, K. Todd-Brown, et al. "PLINK: A tool set for whole-genome association and population-based linkage analyses". In: *American Journal of Human Genetics* 81.3 (Sept. 2007), pp. 559–575. ISSN: 00029297. DOI: 10.1086/519795.

[107] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science and Engineering* 9.3 (2007), pp. 90–95. ISSN: 15219615. DOI: 10.1109/MCSE.2007.55.

[108] M. Waskom, O. Botvinnik, D. O'Kane, et al. "mwaskom/seaborn: v0.9.0 (July 2018)". In: (July 2018). DOI: 10.5281/ZENODO.1313201.

[109] InkscapeProject. *Inkscape*. 2018.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6 Supplementary Information

6.1 Supplementary Text

When evaluating the relationship between 3D genome variability and introgression (Results section 3.7:"3D genome organization constrained introgression in MHs"), we considered a variety of subsets of genomic windows to fully explore these results. We show that the maintext results (Fig. 5) replicate when using earlier introgressed Neanderthal haplotype predictions from Vernot et al. [15] and other thresholds (Figs. S11,S12). We also find that 3D genome variability is more strongly predictive of introgression shared among all three super-populations than an introgressed sequence unique to a single super-population (Table S7). We hypothesize this is because the maintenance of a haplotype across diverse populations indicates stronger tolerance of the AH 3D organization pattern in diverse human genomic contexts. Additionally, 3D variability is relatively more informative about the amount of introgression when only considering windows of the genome with any introgressed sequence present (Table S8). Thus, we hypothesize that in 1 Mb windows with strong purifying selection against a large-effect introgressed variant (e.g., a deleterious protein-coding variant), 3D genome variability is less relevant. Ultimately, the pressures shaping the landscape of introgression across the genome were multi-factorial, but we demonstrate that 3D genome organization likely played a role.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6.2 Supplementary Figures

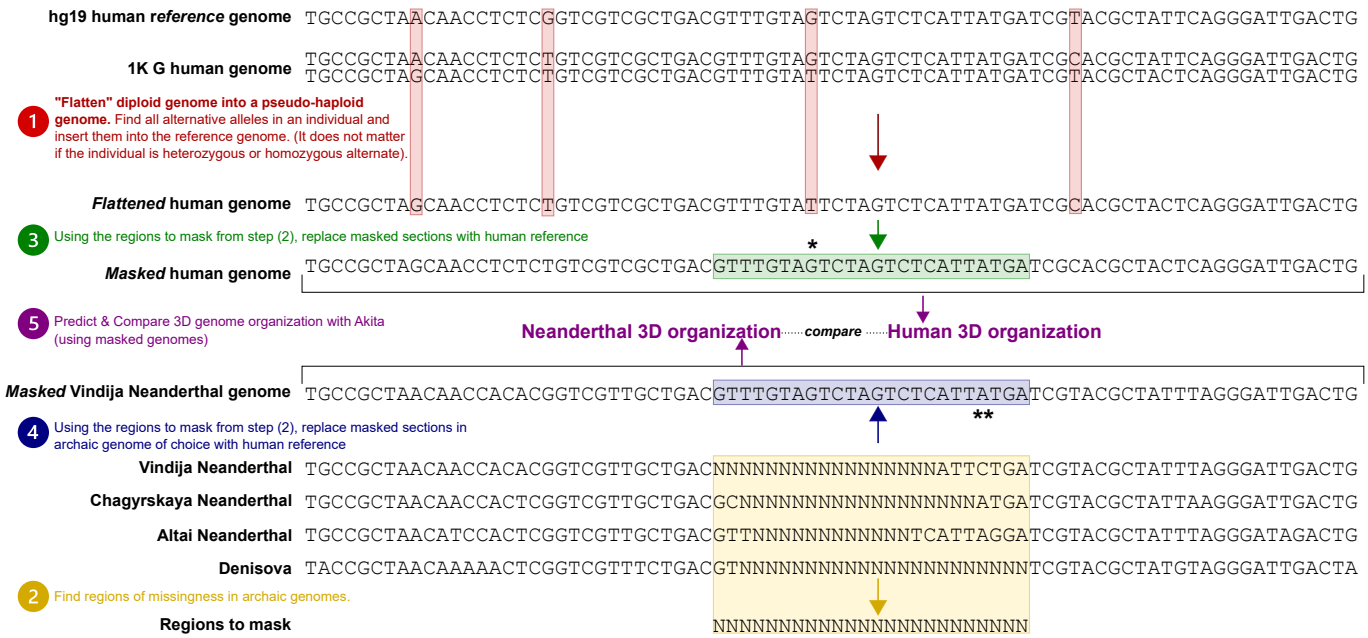

Figure S1: Handling missingness in the archaic hominin genomes. We constructed full-length genomes for each MH or AH based upon their genotyping information. Here, we illustrate a schematic of the procedure used to account for the challenges of archaic DNA. **(1)** Given the difficulty of distinguishing heterozygous genotypes in the ancient DNA samples, we treated all individuals as if they were homozygous (pseudo-haploid). If an individual had an alternate allele (homozygous or heterozygous), we inserted it into the reference genome to create a pseudo-haploid, or "flattened" genome for each individual (hightlighted in red boxes). **(2)** Because of gaps in coverage resulting from the challenges of ancient DNA, particularly in genomic regions of low complexity, we "masked" all genomic regions lacking archaic genotyping information by reverting nucleotide states to the hg19 reference (yellow box). For analyses that compared 3D genome organization between MHs and AHs, and MHs we do this masking procedure for both **[3]** MHs (green box) and **[4]** AHs (blue box) to facilitate appropriate comparisons. **[5]** We run Akita on each processed genome separately and then compare the resulting contact maps. By filling both genomes with the same sequence, there will be no differences between the AH-MH predictions or resulting comparisons. Although AHs and MHs certainly did not have the same genome sequences in these regions of missingness, we preferred this as a conservative approach to minimize identifying regions of interest if there were missing data. For example, we illustrate that at the nucleotide *****, although we observe an MH alternative allele (`T`), it gets masked and replaced with the hg19 reference (`G`) because that locus is not comparable to AH genomes. Many of the regions of missingness are shared by all or most of the AHs because those regions are just inherently difficult to sequence (Fig. S2). However, at the nucleotide ******, we illustrate another example where an allele observed in the Vindija genome (`C`) is masked with hg19 reference (`A`) so that it facilitates comparisons between the AHs (some of which have missingness at that locus).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

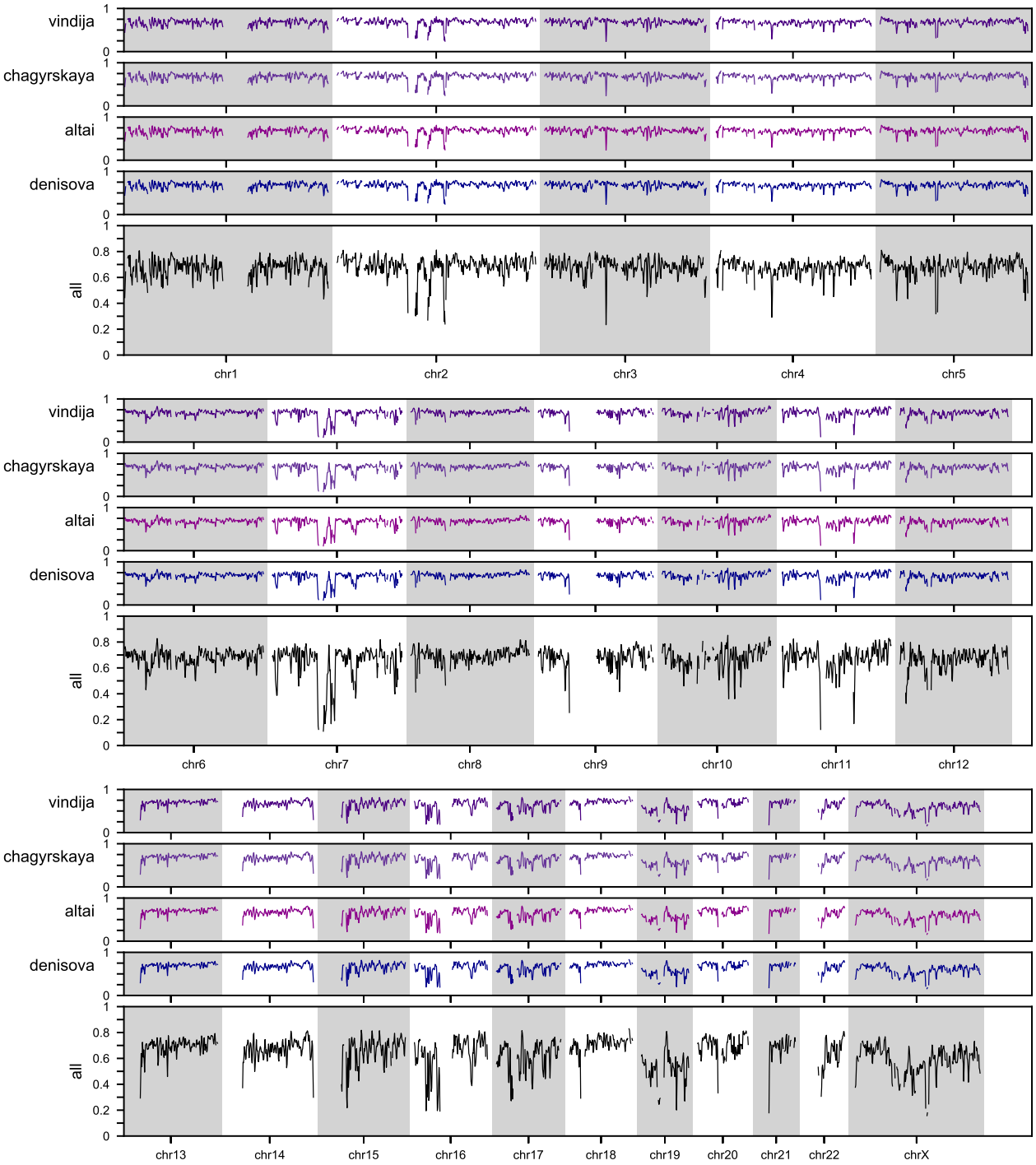

Figure S2: Archaic hominin sequence coverage across the genome. Ancient DNA fragmentation and degradation present challenges to both sequencing and alignment resulting in gaps in coverage, particularly in genomic regions of low complexity. Here, we show coverage across the genome for the 4 AHs. The horizontal axis represents genomic loci at the same sliding approximately 1 Mb window resolution ($N = 4,999$) used to do all analyses (Methods). The vertical axis unit is the proportion of bp with coverage (for the 1 Mb window). Bins without full coverage in modern humans (often near centromeres or telomeres) are excluded from all analyses and this figure. The bottom trace (black, labeled "all") represents the union of the missing segments for all 4 AHs. These regions are masked (Methods, Fig. S1) to facilitate 3D genome and sequence variation comparisons.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

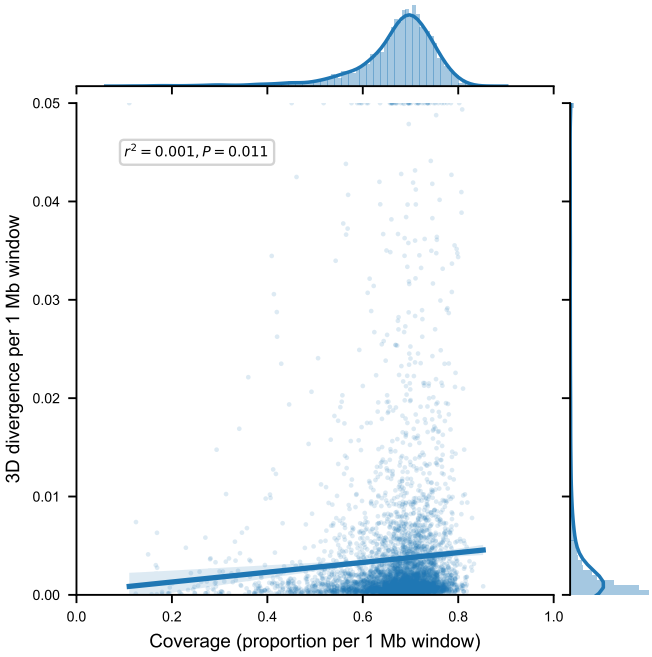

Figure S3: 3D divergence in 1 Mb genomic window is weakly correlated with coverage. Because we mask archaic missingness (Methods, Fig. S1,S2), regions with less coverage have more masking and the resulting processed sequences may have less AH-MH sequence variation. For 1 Mb windows across the genome ($N = 4999$), we compare AH (Vindija Neanderthal) and African MH (HG03105) 3D divergence (vertical axis) with the amount of coverage in that window (horizontal axis). The amount masked is equal to $1 -$ coverage. 3D divergence is positively correlated with coverage ($r^2 = 0.001$, $P = 0.01$). This is likely because there is more opportunity to find variation that results in contact map changes when less of the region is masked; however, this correlation is very weak suggesting that more coverage of the archaic genomes may not uncover many additional examples of divergent organization.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

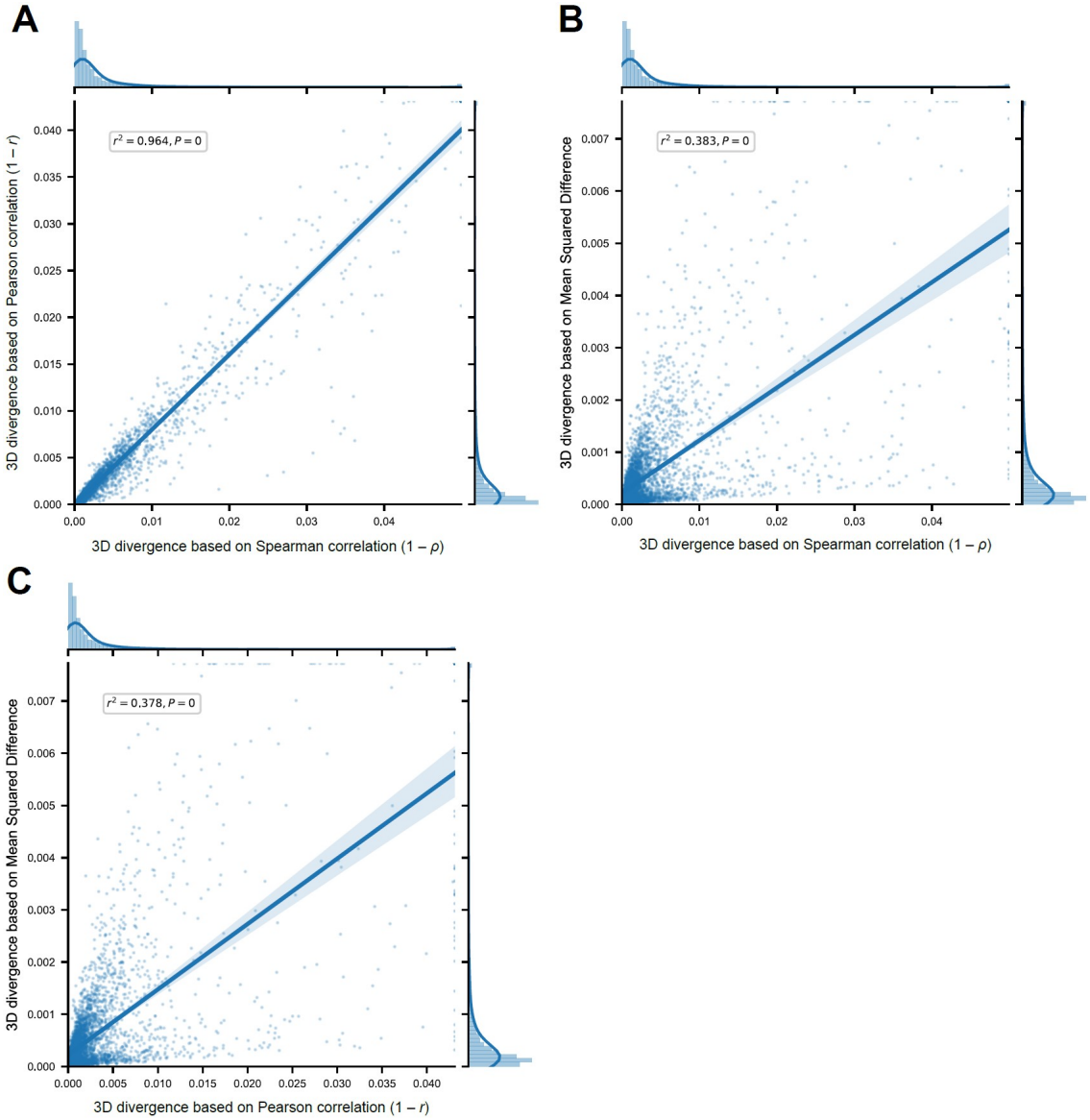

Figure S4: Alternative measures of contact map comparison correlate with the 3D divergence derived from the Spearman's rank correlation coefficient. In the main text, we compare chromatin contact maps using a 3D divergence score based on Spearman's rank correlation coefficient $(1 - \rho)$. Here, for the same windows across the genome ($N = 4999$), we compare AH (Vindija Neanderthal) and African MH (HG03105) predictions using this Spearman-derived 3D divergence to others based on **(A)** Pearson's correlation coefficient $(1 - r)$ ($r^2 = 0.964$) and **(B)** mean squared difference ($\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$) ($r^2 = 0.383$). We also compare **(C)** these alternative measures (mean squared difference vs. Pearson's correlation) to each other ($r^2 = 0.378$). The correlations between all measures are highly significant (all $P < 5 \times 10^{-324}$).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

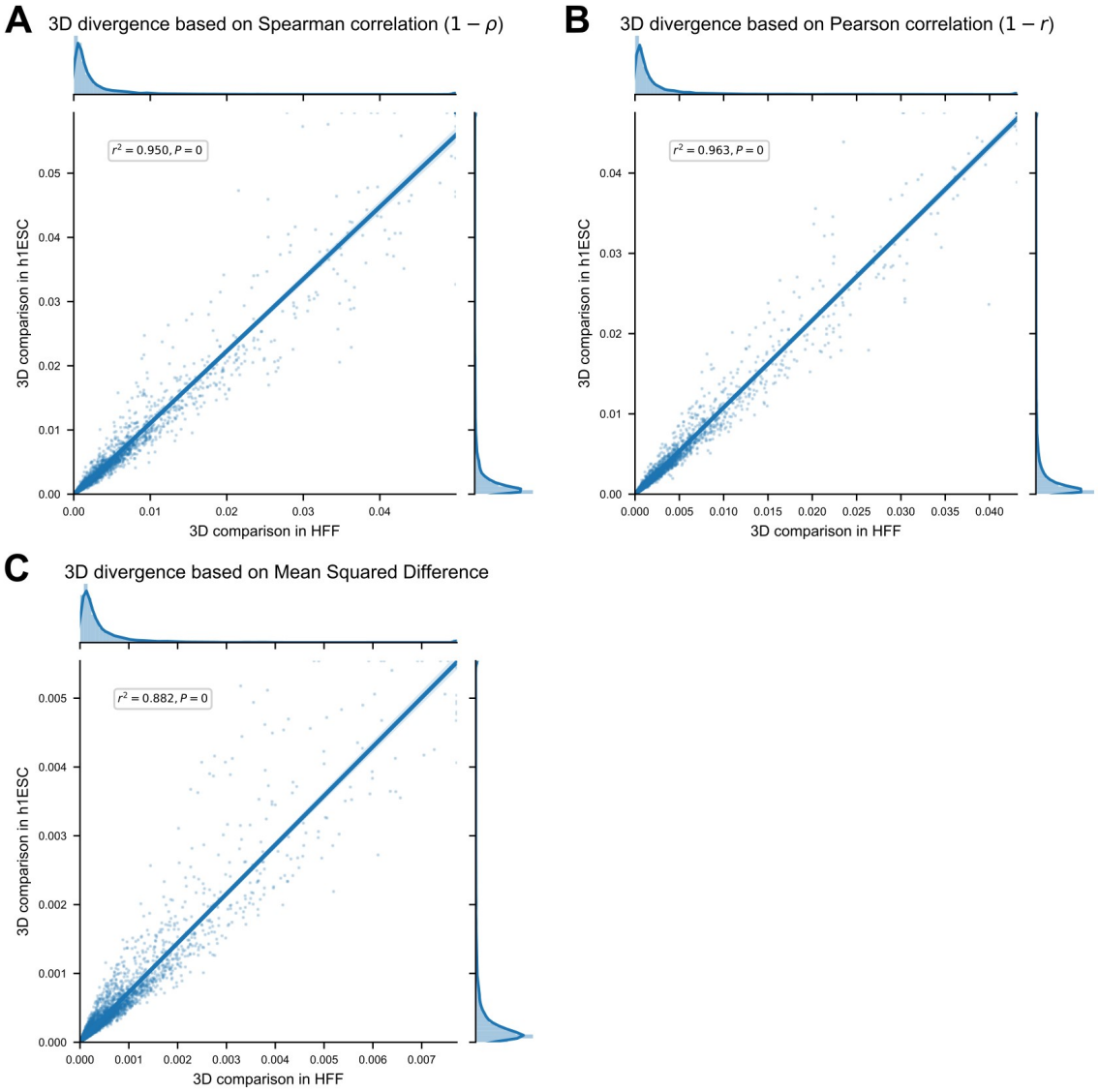

Figure S5: 3D genome organization comparisons with chromatin contact maps from embryonic stem cell (ESC) are similar to those from human foreskin fibroblast (HFF). For the same windows across the genome ($N = 4999$), we compare AH (Vindija Neanderthal) and African MH (HG03105) predictions in embryonic stem cell (ESC) (vertical axis) versus human foreskin fibroblast (HFF) (horizontal axis) cell types. The comparisons across cell types are highly correlated regardless of the measure used to quantify their divergence. We consider comparison measures defined using the **(A)** Spearman correlation ($r^2 = 0.95$), **(B)** Pearson correlation ($r^2 = 0.96$), and **(C)** mean squared difference ($r^2 = 0.88$) (all $P < 5 \times 10^{-324}$).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

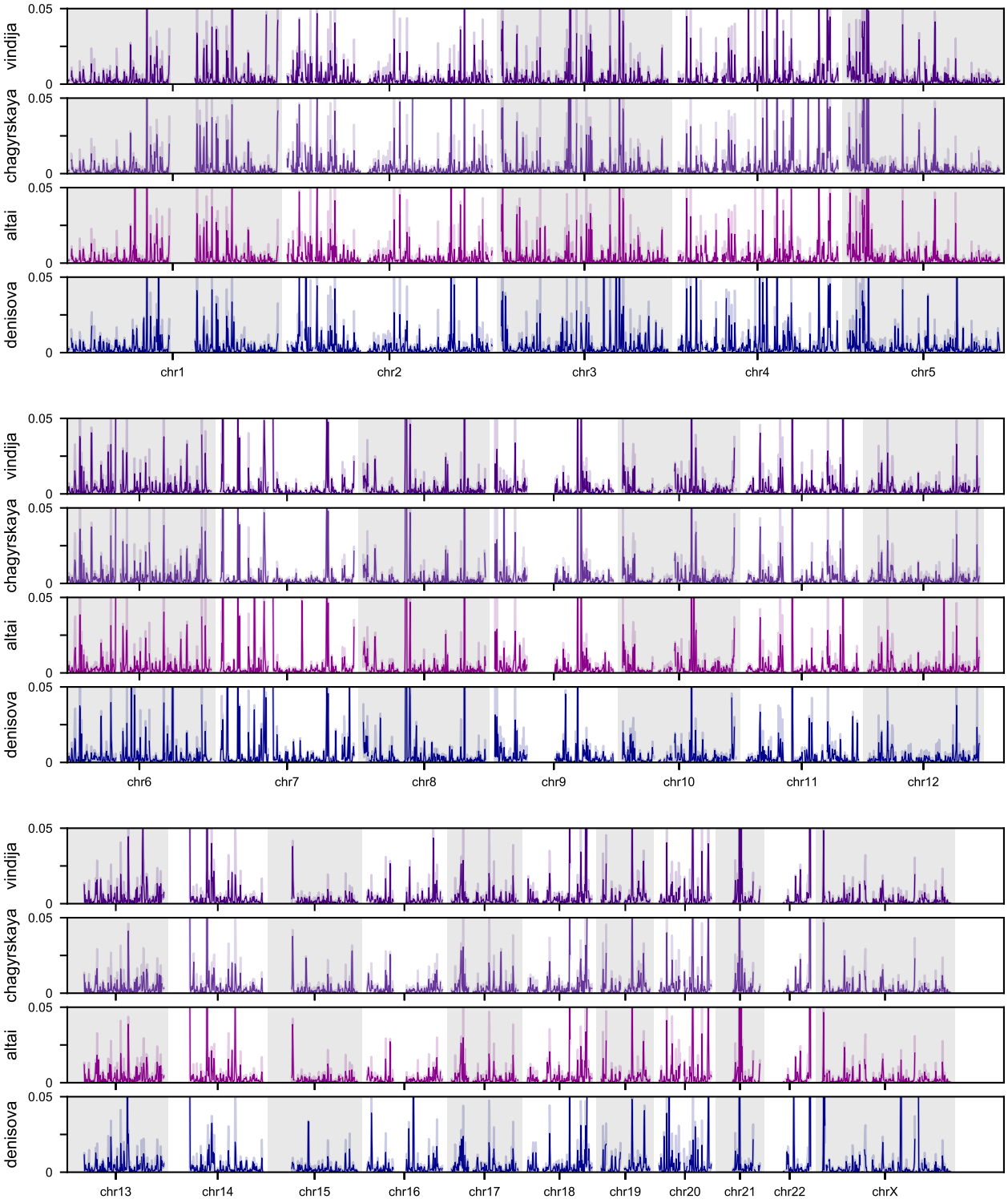

Figure S6: AH-MH 3D divergence across the whole genome. Across the genome, we plotted the average divergence of each of the AHs to five modern African individuals from different subpopulations. The horizontal axis represents genomic loci at the same sliding 1 Mb window resolution ($N = 4,999$) used to do all analyses (Methods). This expands Fig. 2C from chr7 to the whole genome. The error band indicates the 95% CI. Comparing the 3D genomes of Neanderthals (purple) or Denisova (blue) with MHs reveals windows of both similarity and divergence (peaks).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

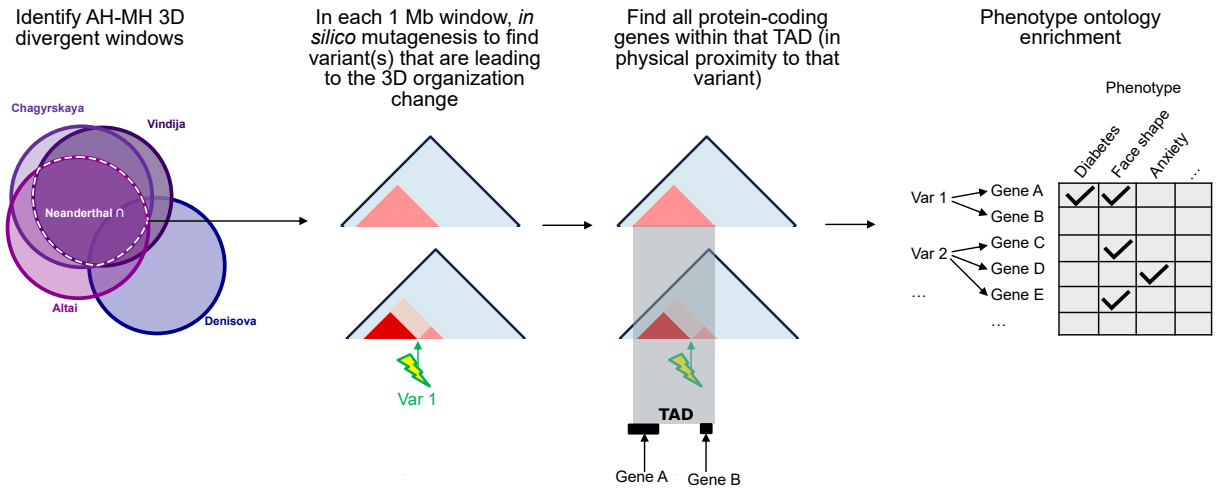

Figure S7: Method for linking 3D divergent windows to test phenotype ontology term enrichment. To test if differences in AH-MH 3D organization are enriched near genes related to particular phenotypes we follow a procedure that sequentially links 3D divergent windows to variants to TADs to genes and, ultimately, to phenotypes. We identify AH-MH 3D divergent windows in Fig. 3A–B. We consider three different sets of AH-MH divergent windows, those shared (intersect) by all Neanderthals, those in any Neanderthal (union), and those in Denisova. Results from the set shared by all Neanderthals ($N = 43$ windows) are shown in the main text (Fig. 3D). In each 1 Mb 3D divergent window, we identify the variant(s) contributing to the most prominent 3D differences using *in silico* mutagenesis (lightning bolt) (Methods). 3D-modifying variants are then linked to protein-coding genes (black bars) in their TAD (gray rectangle) because this provides evidence of physical proximity. Genes are linked to phenotypes from the Human Phenotype Ontology (HPO) and genome-wide association studies (GWAS) Catalog 2019. Through this procedure, we counted the number of ontology terms linked to the set of 3D-modifying variants. We test enrichment for ontology terms linked to at least one 3D-modifying variant using a shuffling approach to create an empirical distribution for how many times we would observe each annotation under the null. We used these distributions to calculate an enrichment and P-value for each ontology term. The specific data sets used in this procedure are detailed in the Methods. Counts of the number of windows, 3D-modifying variants, genes, and phenotypes for each set are in Table S2. Results for enrichment are in Figs. 3D,S8.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

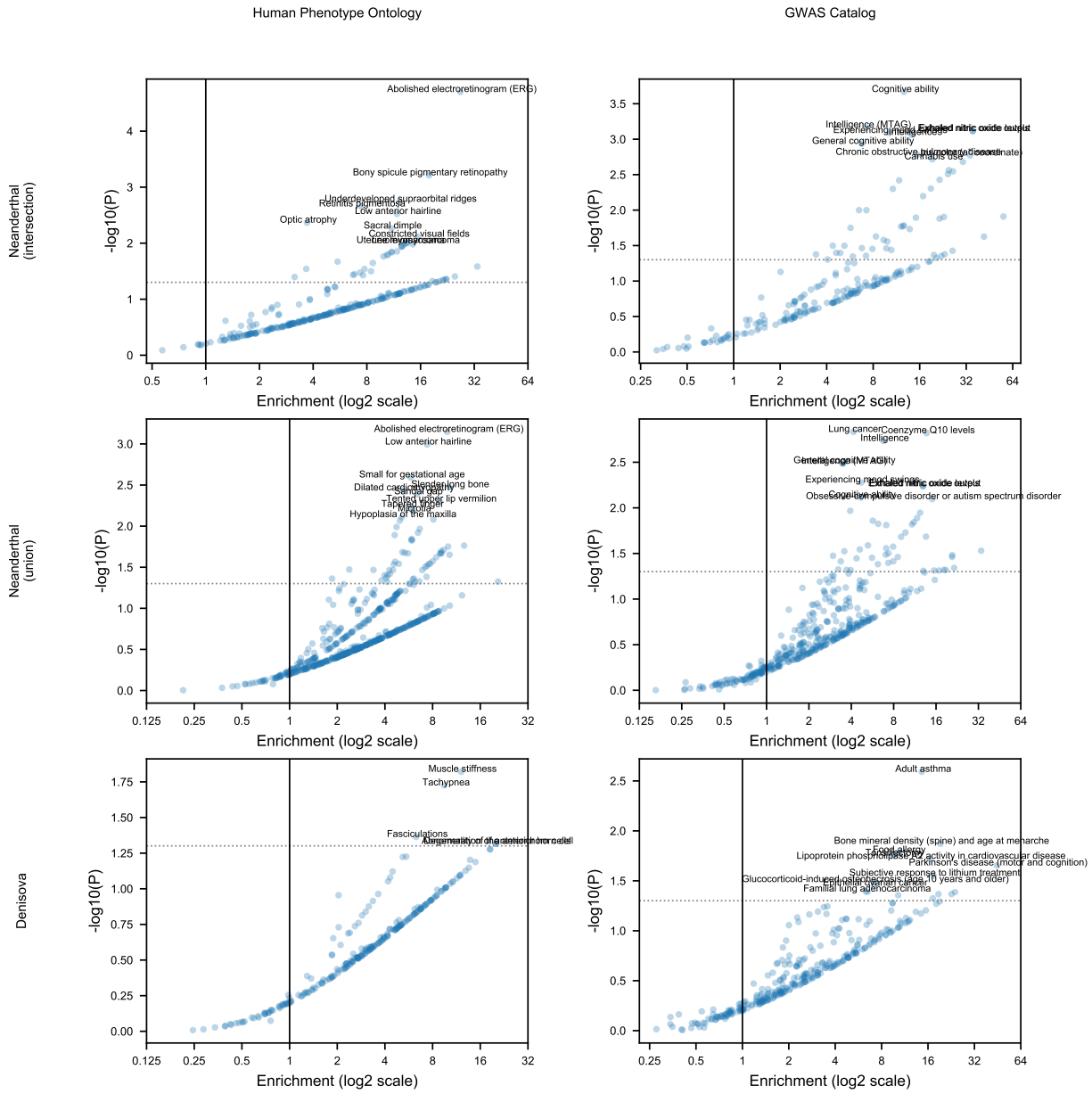

Figure S8: Phenotype ontology enrichment across other sets of AH-MH 3D divergent windows implicate similar phenotypes. When testing if differences in AH-MH 3D organization are enriched near genes related to particular phenotypes, we used three different sets of AH-MH 3D divergent windows (rows) and two different sets of gene-phenotype links (columns). The top set is from 43 3D-divergent windows shared by Neanderthals (intersect) (also shown in the main text, Fig. 3D). The middle is from 110 divergent windows in any Neanderthal (union). The bottom is from 73 divergent windows in Denisova. Each volcano plot has enrichment on the horizontal axis and significance on the vertical axis which were calculated with reference to a shuffled null distribution ($n = 500,000$, Methods). Each point represents one ontology term. Only terms linked to the 3D divergent windows in each set were tested for enrichment or depletion. The most significant 10 terms are labeled if $P < 0.05$ (dotted line). Similar to the Neanderthal (intersection) set, phenotypes related to the retina, hair, immune response, skeleton, cognition, and lung capacity are highlighted. Additional phenotypes at nominal significance include traits related to the heart, muscle, cancer, and bone density. Details about the process to link the 3D divergent windows to genes and phenotypes are in the Methods and Fig. S7. Details about the number of windows, variants, and phenotypes considered for each set are in Table S2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

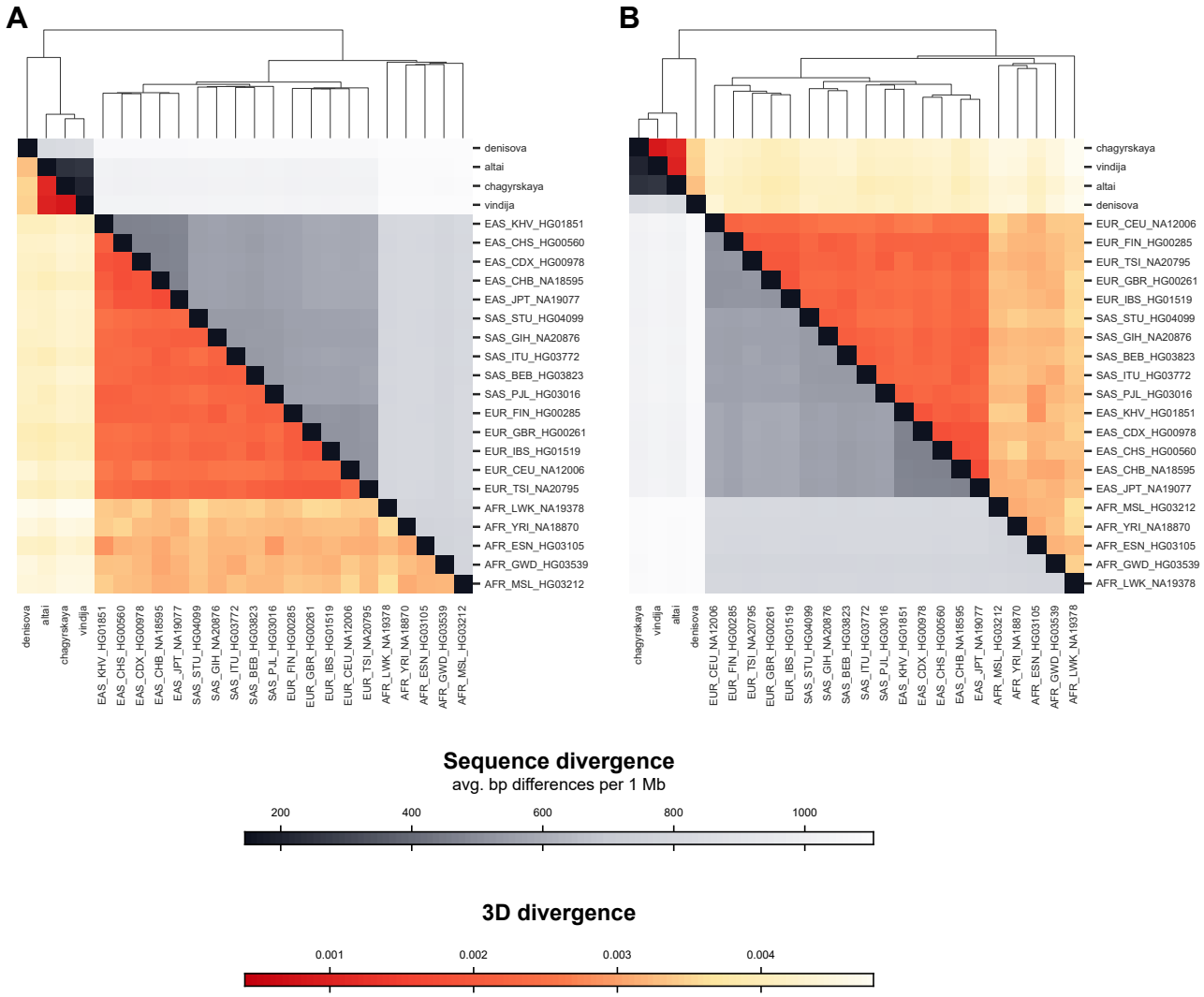

Figure S9: Full pairwise heatmaps clustered by both sequence 3D divergence and sequence divergence. We calculated the mean genome-wide 3D divergence for all pairs of AH and MH individuals (oranges) to compare with the genome-wide mean sequence divergence (grays). Fig. 4A displays these heatmaps when clustered by sequence divergence. Fig. 4A is reproduced in **(A)** with the full labels of all 1KGP individuals and their sub- and super-population information. **(B)** We also show the heatmap clustered by 3D genome divergence. Overall, global patterns of 3D genome divergence follow global patterns of sequence divergence. Lists of 1KGP individuals used and their abbreviation codes are defined in Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

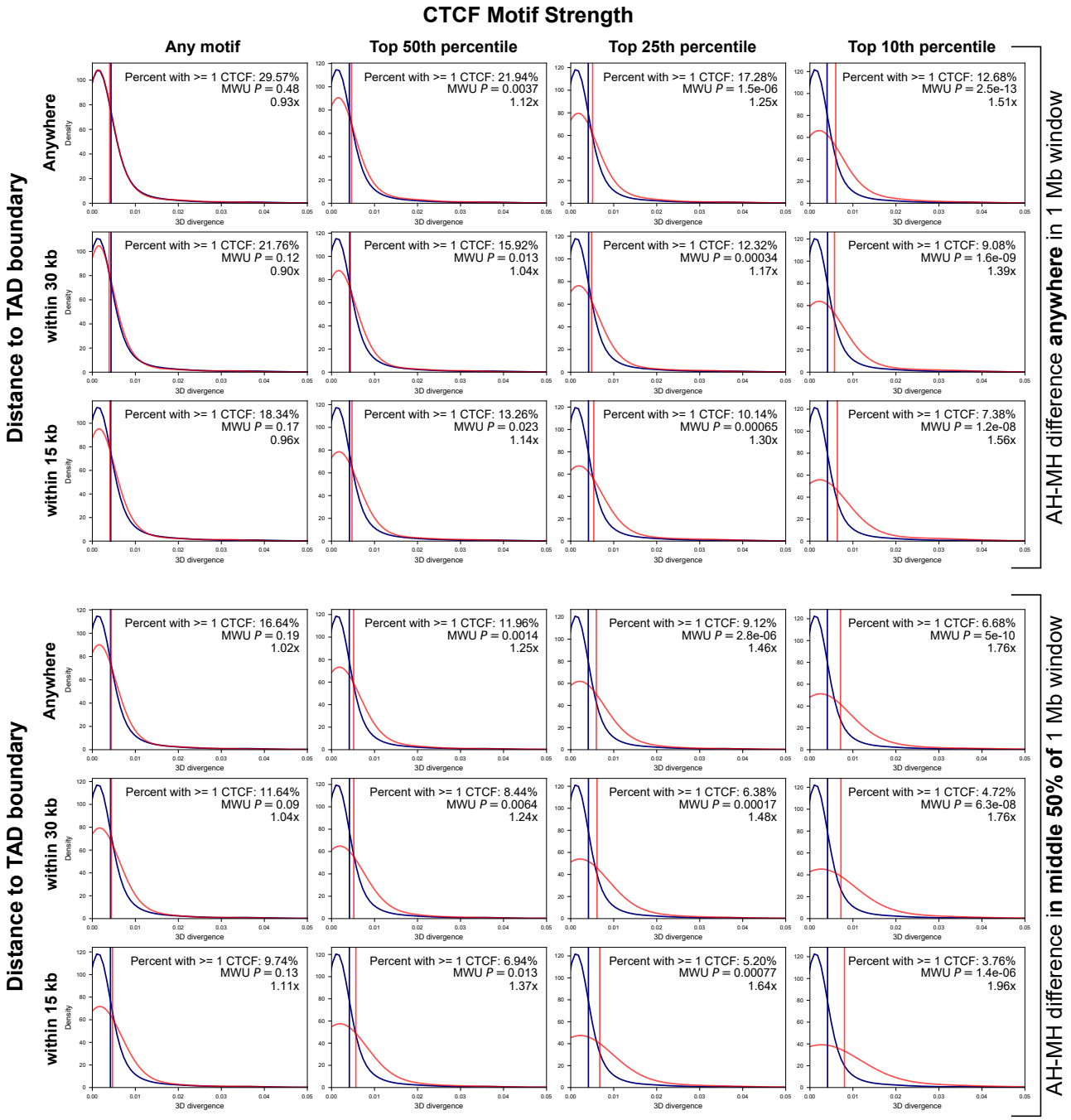

Figure S10: 3D genome divergence depends on both the strength and context of the CTCF motif disrupted. Based on the importance of CTCF-binding in maintaining 3D genome organization, we quantified the effects of AH-MH nucleotide differences overlapping CTCF binding motifs on 3D divergence. Given the complexity in the "grammar" of encoding 3D genome organization, we hypothesized that not all CTCF disruptions are equally likely to influence 3D divergence. Fig. 4B demonstrates this. But, here we replicate this with other thresholds and filters. We considered if each 1 Mb window ($N = 4,999$) had a sequence difference between a Neanderthal (Vindija) and a MH (HG03105) genome that overlapped a CTCF site. We plotted the distribution of 3D divergence in a window by whether there was a "CTCF overlapping variant" (red) or not (blue). We further filtered windows by multiple annotations describing the context and strength of the CTCF site overlapped. First, we stratified windows by if the "CTCF overlapping variant" occurs within the middle half of the 1 Mb window (right vertical axis). Second, we stratified windows by the proximity of the "CTCF overlapping variant" to a TAD boundary (anywhere, within 30 kb, or within 15 kb) (left vertical axis). Finally, we stratified windows by the strength of the overlapped CTCF motif in percentiles (any, top 50%, 25%, or 10%) (horizontal axis). All three features describing context and strength are informative about the likelihood of 3D divergence. For example, when filtering for the strongest CTCF motifs overlapped by a variant, 3D divergence increases 1.96-fold compared to 1.11-fold if strength is ignored (bottom left vs. bottom right). When considering by proximity to TAD boundaries, 3D divergence always increases when a "CTCF overlapping variant" is closer to a TAD boundary (4th row vs. 6th row). This illustrates that our approach has learned the complex sequence patterns underlying 3D genome folding that could not be determined by simply intersecting AH variants with all CTCF sites.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

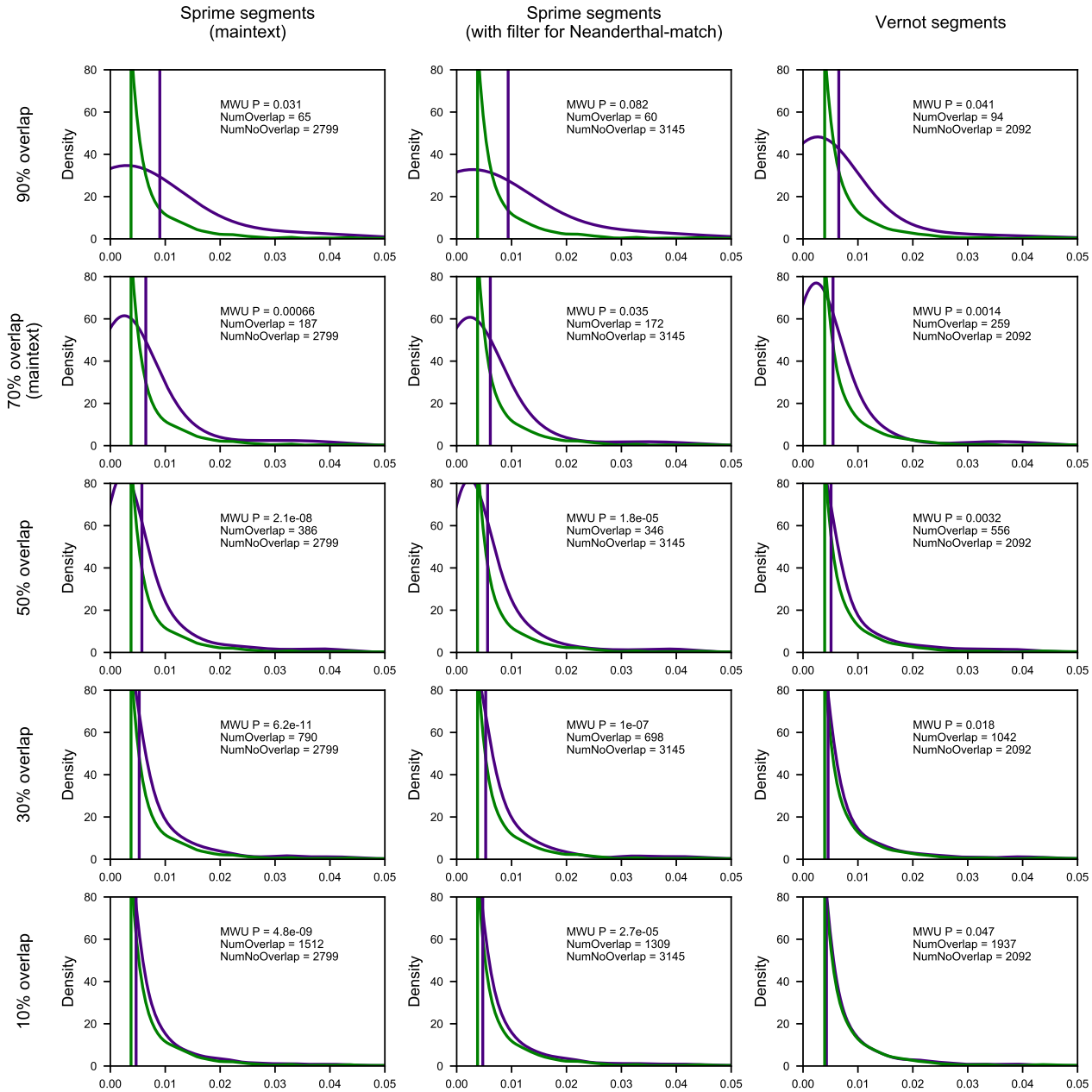

Figure S11: Windows with evidence of AH introgression are more 3D variable in MHs even when using different definitions of introgression. Genomic windows with high levels of introgression across present-day non-African populations (purple distribution) are more 3D-variable in modern Africans (horizontal axis) than windows without evidence of introgression (green distribution). In the main text, we considered introgression defined by segments from Browning et al. [93] (first column) covering at least 70% of bases in a 1 Mb window (second row). This identifies 187 autosomal 1 Mb windows with introgression and 2,799 without (same figure as Fig. 5A). Here, we show that this trend is consistent even when using different sets of introgressed haplotypes (columns) and thresholds for overlap (rows). Sprime segments are from Browning et al. [93]. Sprime segments with Neanderthal-matching filter are a subset of the Browning et al. [93] introgressed segments that have 30 putatively introgressed variants that could be compared to the Altai Neanderthal genome and had a match rate of at least 30% to the Altai Neanderthal allele. S* Vernot segments are from Vernot et al. [15]. Vertical lines represent the distribution means. P-values are from a two-tailed Mann–Whitney U test.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

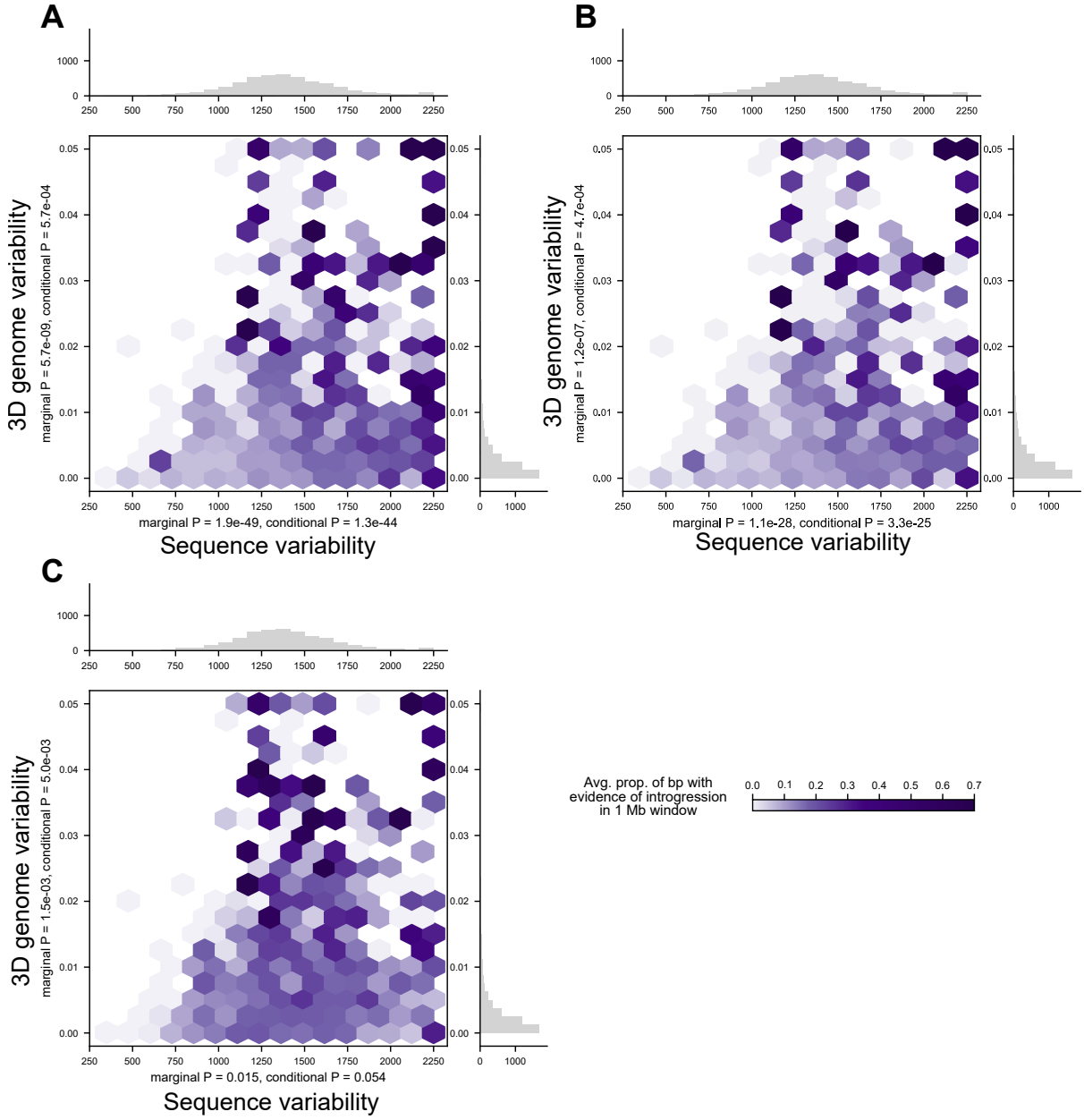

Figure S12: 3D variable windows in MH have more evidence of AH introgression even when using different definitions of introgression. For three different sets of introgressed haplotypes (**A**-**C**), we plot the relationship between sequence variability (horizontal axis) and 3D genome variability (vertical axis) with amount of AH ancestry in a window (purples). Darker purple indicates a higher proportion of introgression in a 1 Mb genomic window. 3D genome variability is defined as the average modern-African pairwise 3D genome diversity. Sequence variability is defined as the average pairwise nucleotide differences per modern-African in a 1 Mb window. *P*-values correspond to the significance of sequence variability or 3D genome variability to predict amount of introgression in a 1 Mb window. 3D genome variability is predictive of the amount of introgression both independently and when conditioned on sequence variability for all three sets of introgression. For, **A**,**B**, and **C**, respectively, introgressed haplotypes are from Sprime segments, Sprime segments with a Neanderthal-sequence match filter, and S* segments. **A** is shown in the maintext in Fig. 5B. Sprime segments are from Browning et al. [93]. Sprime segments with Neanderthal-matching filter are a subset of the Browning et al. [93] introgressed segments that have 30 putatively introgressed variants that could be compared to the Altai Neanderthal genome and had a match rate of at least 30% to the Altai Neanderthal allele. Vernot segments are from Vernot et al. [15].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

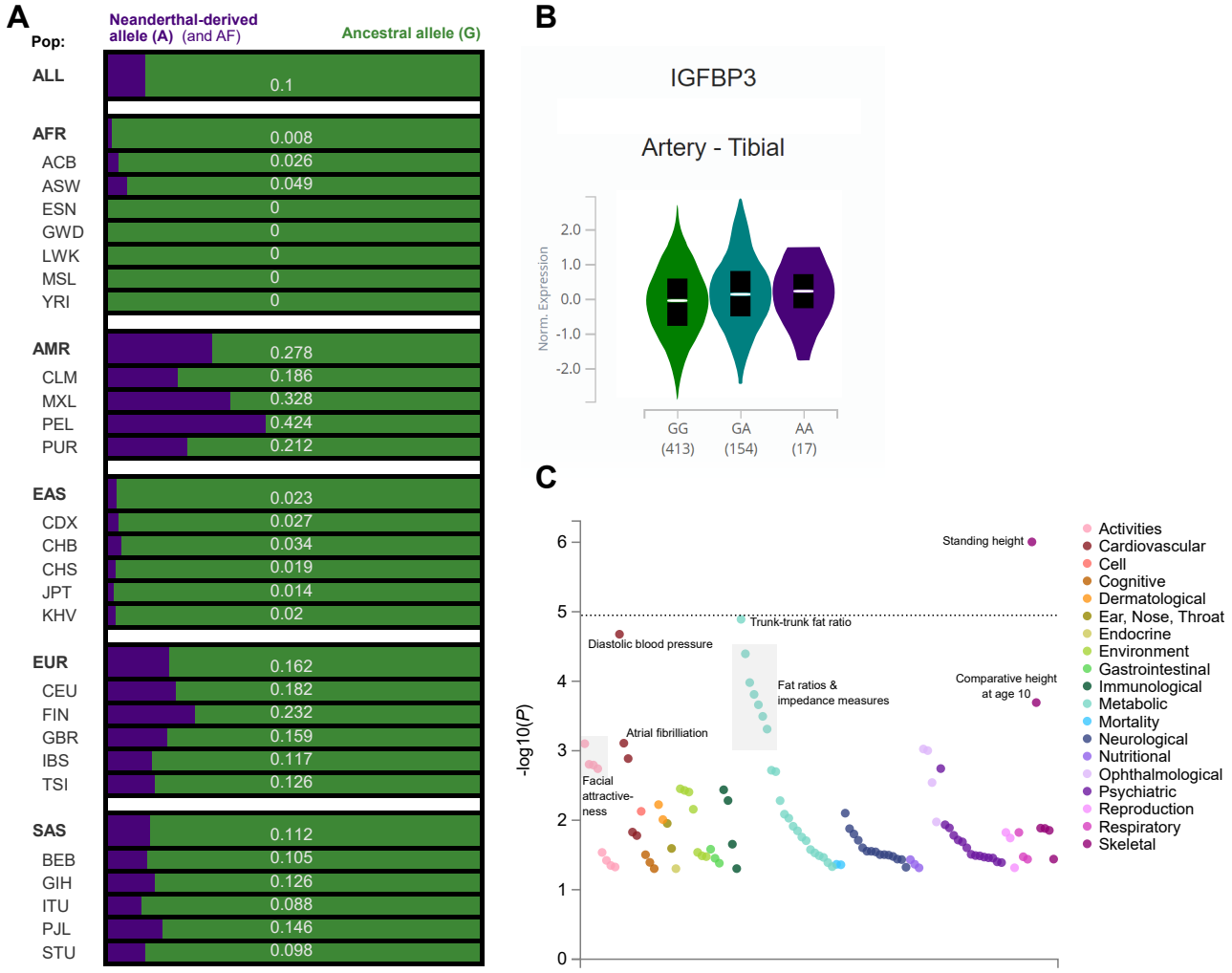

Figure S13: rs12536129 is a high-frequency introgressed allele with regulatory and phenotypic associations. In Fig. 6A–B, we describe an AH-MH 3D divergent window that was introgressed into some modern Eurasians. *In silico* mutagenesis of this window revealed a G to A change at chr7:46,169,621 (rs12536129) associated with the largest change in 3D genome organization. **(A)** Across human populations, this introgressed allele remains at high-frequency today, especially in Peru (28% AMR, 2% EAS, 16% EUR, 11% SAS, 0% non-admixed sub-Saharan AFR). Purple bars represent the frequency of the introgressed Neanderthal-derived allele. **(B)** This introgressed allele is also an eQTL in GTEx for the physically linked gene *IGFBP3*, Insulin-like growth factor-binding protein 3 ($P = 0.00014$ in artery tissue) [42]. **(C)** In MHs, this variant is associated with traits including standing height ($P = 9.9 \times 10^{-7}$), fat distribution (trunk fat ratio, impedance measures, $P = 1.3 \times 10^{-5}$), and diastolic blood pressure ($P = 2.1 \times 10^{-5}$). This figure was generated with the GWASAtlas from Watanabe et al. [102] and is sorted by domain and P-value. The dotted line represents a highly conservative Bonferroni corrected P-value (1.05×10^{-5}) for testing 4756 traits (including many correlated traits and GWASs in which the SNP was not tested).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

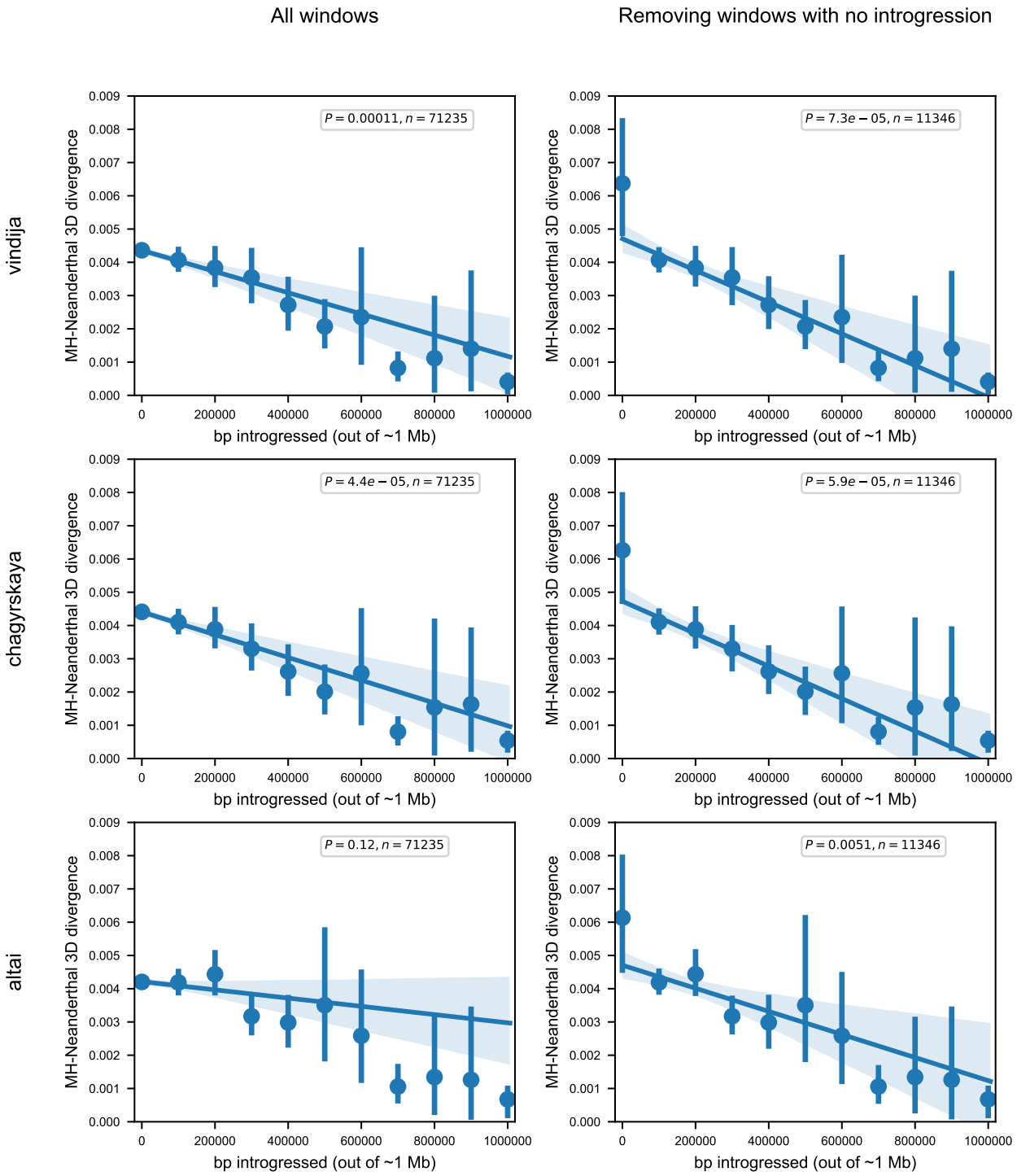

Figure S14: Amount of introgression is negatively correlated with 3D divergence to all Neanderthal individuals. The amount of introgression in a 1 Mb window (number of bp, horizontal axis) is significantly correlated with the similarity of an individual's 3D genome organization to a Neanderthal's genome organization (vertical axis). This is demonstrated across all three Neanderthal individuals: Vindija in the top panel (also shown in Fig. 6C), Chagyrskaya in the middle, and Altai at the bottom. We hypothesize the trend is weakest in Altai because it is less related to the introgressing Neanderthal population compared to the Vindija Neanderthal [2]. The left column considers all 4,749 autosomal 1 Mb windows for 15 Eurasians (total $n = 71,235$, 1KGP individuals in Table S1). In the right column, this trend also holds when you remove 1 Mb windows with no (0 bp) introgression in the 15 considered Eurasian individuals $n = 11,346$. The P-values are the significance of the correlation. The error bars signify 95% bootstrapped confidence intervals and the error band signifies the 95% bootstrapped confidence interval for the linear regression estimate.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6.3 Supplementary Tables

	Superpopulation	Subpopulation	ID	Subpopulation Description
Individuals in initial and Eurasian introgression analyses	EAS	CDX	HG00978	Chinese Dai in Xishuangbanna, China
	EAS	CHB	NA18595	Han Chinese in Beijing, China
	EAS	CHS	HG00560	Han Chinese South
	EAS	JPT	NA19077	Japanese in Tokyo, Japan
	EAS	KHV	HG01851	Kinh in Ho Chi Minh City, Vietnam
	EUR	CEU	NA12006	Utah residents (CEPH) with Northern and Western European ancestry
	EUR	FIN	HG00285	Finnish in Finland
	EUR	GBR	HG00261	British in England and Scotland
	EUR	IBS	HG01519	Iberian populations in Spain
	EUR	TSI	NA20795	Toscani in Italia
	SAS	BEB	HG03823	Bengali in Bangladesh
	SAS	GIH	NA20876	Gujarati Indian in Houston, TX
	SAS	ITU	HG03772	Indian Telugu in the UK
	SAS	PJL	HG03016	Punjabi in Lahore, Pakistan
	SAS	STU	HG04099	Sri Lankan Tamil in the UK
	AFR	GWD	HG03539	Gambian in Western Division, The Gambia
	AFR	LWK	NA19378	Luhya in Webuye, Kenya
	AFR	MSL	HG03212	Mende in Sierra Leone
	AFR	YRI	NA18870	Yoruba in Ibadan, Nigeria
	AFR	ESN	HG03105*	Esan in Nigeria
Africans in AH-MH divergence and 3D genome variability analyses	AFR	ESN	HG03105	Esan in Nigeria
	AFR	ESN	HG03499	Esan in Nigeria
	AFR	ESN	HG03511	Esan in Nigeria
	AFR	ESN	HG03514	Esan in Nigeria
	AFR	ESN	HG02922	Esan in Nigeria
	AFR	GWD	HG03539	Gambian in Western Division, The Gambia
	AFR	GWD	HG03025	Gambian in Western Division, The Gambia
	AFR	GWD	HG03028	Gambian in Western Division, The Gambia
	AFR	GWD	HG03040	Gambian in Western Division, The Gambia
	AFR	GWD	HG03046	Gambian in Western Division, The Gambia
	AFR	LWK	NA19378	Luhya in Webuye, Kenya
	AFR	LWK	NA19017	Luhya in Webuye, Kenya
	AFR	LWK	NA19434	Luhya in Webuye, Kenya
	AFR	LWK	NA19445	Luhya in Webuye, Kenya
	AFR	LWK	NA19019	Luhya in Webuye, Kenya
	AFR	MSL	HG03212	Mende in Sierra Leone
	AFR	MSL	HG03086	Mende in Sierra Leone
	AFR	MSL	HG03085	Mende in Sierra Leone
	AFR	MSL	HG03437	Mende in Sierra Leone
	AFR	MSL	HG03378	Mende in Sierra Leone

Table S1: 1000 Genomes Project (1KGP) individual genomes used for 3D genome predictions. The top set of individuals were used in the initial 3D genome survey (Figs. 2, 4A) and introgression analyses (Fig. 6). The bottom set of African individuals was used to more robustly call AH-MH 3D genome divergence windows (Fig. 3) and to calculate MH 3D genome variability (Fig. 5). For consistency, the genome of HG03105 was used for all examples.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

‘	Number of 1 Mb 3D divergent windows (includes partially overlapped windows)	Number of unique 3D divergent windows (merging overlapping windows)	Number of “3D-modifying variants” observed in the 3D divergent windows	Number of “3D-modifying variants” that have evidence of introgression	Proportion of “3D-modifying variants” that have evidence of introgression	Number of gene-”3D-modifying variants” links*	Number of unique genes linked to 3D-modifying variants”*	Number of unique HPO terms linked to “3D-modifying variants” (used in enrichment test)*	Number of unique GWAS terms linked to “3D-modifying variants” (used in enrichment test)*
Vindija	93	70	76	38	0.500				
Chagyrskaya	95	71	78	32	0.410				
Altai	82	67	73	33	0.452				
Denisova	105	73	83	9	0.108	130	129	248	318
All Neanderthals (intersection)	54	43	45	28	0.622	88	85	271	208
All Neanderthals (union)	144	110	121	43	0.355	224	206	535	435
All archaic hominins (intersection)	10	7	8	6	0.750				
All archaic hominins (union)	234	167	191	45	0.236				

Table S2: Counts of 3D divergent windows and 3D-modifying variants. The number of 3D divergent windows per different AH individuals (rows) are in the first two columns. The first column is the raw 1 Mb windows, while the second column counts overlapping windows as one merged window (these values are depicted in Fig. 3B). The number of 3D-modifying variants in each window is in column three. The number and fraction of these 3D-modifying variants that are introgressed are in columns four and five. To conduct the phenotype ontology enrichment analyses shown in Figs. 3D,S8, we linked the 3D-modifying variants to genes (column six and seven, see Methods). These genes were then linked to terms using Human Phenotype Ontology (HPO) (column eight) and the GWAS Catalog. These two sets of terms were then tested for enrichment. The phenotype ontology enrichment analysis parts (columns denoted with *) were only calculated on certain sets of 3D-diverged windows.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

See supplementary excel file for large tables.

Table S3: AH-MH 3D divergent windows. Coordinates (in hg19) for 167 AH-MH 3D divergent windows identified in Fig. 3A–B. Windows were identified at approximately 1 Mb resolution (Methods) and overlapping windows were merged. The "AH" column details for which AH(s) the 3D divergent window was identified (A: Altai, C: Chagyrskaya, D: Denisova, V: Vindija). If a window was identified in two different AHs but they were only partially overlapping, the specific coordinates are reported in the "AH" column. For example, at chr13:92798976-94371840, Altai and Chagyrskaya have a 3D divergent window identified at chr13:92798976-93847552, while Vindija has a slightly longer window at chr13:92798976-94371840. Table S5 reports the 3D-modifying variants identified in each window.

See supplementary excel file for large tables.

Table S4: AH-MH 3D divergent windows with less strict thresholds. In addition to the AH-MH divergent windows characterized in the maintext (Figs. 3A–B) and reported in Tables S2,S3, we report a set of AH-MH windows using less stringent criteria. Instead of requiring all 20 AH-MH comparisons to be more 3D divergent than all MH-MH comparisons, we required the average AH-MH comparison to be more 3D divergent than all MH-MH comparisons. We considered regions in the 75th percentile most diverged using either the mean squared error (MSE) or Spearman-based $(1-\rho)$ measures. Otherwise, the procedure to identify AH-MH 3D divergent windows is the same as in the Methods. This identifies 252 windows. Although the windows were identified at approximately 1 Mb resolution (Methods), overlapping windows were merged. The "AH" column details for which AH(s) the 3D divergent window was identified (A: Altai, C: Chagyrskaya, D: Denisova, V: Vindija). If a window was identified in two different AHs but they were only partially overlapping, the specific coordinates are reported in the "AH" column. For example, at chr14:69206016-70778880, Altai and Vindija have a 3D divergent window identified at chr14:69206016-70778880, while Chagyrskaya has a slightly shorter window at chr14:69206016-70254592. Table S6 reports the 3D-modifying variants identified in each window.

See supplementary excel file for large tables.

Table S5: 3D-modifying variants identified inside AH-MH 3D divergent windows. Each 3D-modifying variant that was identified in an AH-MH 3D divergent window (Fig. 3A–B, Table S3) is reported and described. Columns one through four detail the position (in hg19) and alleles. Column five details for which AH(s) the variant and window was found (A: Altai, C: Chagyrskaya, D: Denisova, V: Vindija). It also provides the 3D divergence score. The format is "AH : *in silico* mutagenesis 3D divergence score : AH-MH 3D-divergent window". For example, chr1:74305804 is a 3D-modifying variant identified in the 1 Mb window chr1:73924608-74973184 in Chagyrksyaka, Altai, and Vindija with a 3D divergence of 0.0279 in *in silico* mutagenesis (Methods). Many 3D-modifying variants are identified in overlapping windows. For example, chr1:159131001 is a 3D-modifying variant identified in both chr1:158859264-159907840 and chr1:158334976-159383552 with 3D divergences of 0.0048 and 0.0049 from *in silico* mutagenesis, respectively. Column six provides the coordinates of the TAD in which the 3D-modifying variant is located and column seven provides the protein coding genes within that TAD. Column eight provides overlap with GTEx eQTL in the format "gene:P-value:tissue". Column nine provides overlap with putatively adaptive high-frequency haplotypes from Chen et al. [94]. Column ten provides 1KGP phase 3 allele frequencies by super-population (note: "AFR" includes admixed individuals from the Caribbean and southwestern USA). If alelle frequencies are not present, this variant was not introgressed. Column ten provides a CTCF motif match score for 3D-modifying variants that overlapped CTCF sites defined by Vierstra et al. [96]. For details about all the resources used for these annotations, see the Methods.

See supplementary excel file for large tables.

Table S6: 3D-modifying variants identified in AH-MH divergent windows with less strict thresholds. Each 3D-modifying variant that was identified in an AH-MH 3D divergent windows with less strict thresholds is reported. See Table S4 for a list of these windows and the criteria used to identify them. Columns one through four detail the position (in hg19) and alleles. Column five details for which AH(s) the variant and window was found (A: Altai, C: Chagyrskaya, D: Denisova, V: Vindija). It also provides the 3D divergence score and the measure (Spearman-based [spe] or mean squared error [mse]) used to identify the variant. The format is "AH : *in silico* mutagenesis 3D divergence score based on $1-\rho$: *in silico* mutagenesis 3D divergence score based on MSE : AH-MH 3D-divergent window". For example, chr1:74305804 is a 3D-modifying variant identified in the window chr1:73924608-74973184 in Chagyrksyaka, Altai, and Vindija by the Spearman-based and MSE measures (with 3D divergence 0.0279 and 0.0015, respectively). This variant is also identified for the overlapping window chr1:73400320-74448896 in Chagyrksyaka, Altai, and Vindija with the Spearman-based measure (0.0063, but not identified with the MSE).

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		Sequence variability		3D genome variability	
		marginal P	conditional P	marginal P	conditional P
Browning introgressed haplotypes	introgression SHARED across populations	1.9E-49	1.3E-44	5.7E-09	0.00057
	introgression UNIQUE to one population	0.039	0.019	0.14	0.066
Browning introgressed haplotypes with Neanderthal filter	introgression SHARED across populations	1.1E-28	3.3E-25	1.2E-07	0.00047
	introgression UNIQUE to one population	0.067	0.014	0.00054	0.00013
Vernot introgressed haplotype	introgression SHARED across populations	0.015	0.054	0.0015	0.005
	introgression UNIQUE to one population	0.48	0.79	0.0094	0.012

Table S7: Both 3D genome and sequence variability are more important in predicting introgression shared across super-populations than introgression unique to a single super-population. When considering the relationships between 3D genome variability, sequence variability, and amount of introgression (Supplemental Text, Figs. 5, S12), we consider introgression that was shared across 1KGP super-populations (EAS, EUR, SAS) (white rows) compared to introgression unique to only one super-population (gray rows). We find that 3D genome variability (last two columns) is more strongly predictive of introgression shared among all three super-populations. The analysis was replicated on three sets of introgressed haplotypes. Browning introgressed haplotypes are Sprime segments Browning haplotypes with Neanderthal-matching filter are a subset of the Browning et al. [93] introgressed segments that have 30 putatively introgressed variants that could be compared to the Altai Neanderthal genome and had a match rate of at least 30% to the Altai Neanderthal allele. Vernot haplotypes are S* segments from Vernot et al. [15].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.07.479462; this version posted February 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		Sequence variability		3D genome variability	
		marginal P	conditional P	marginal P	conditional P
Browning introgressed haplotypes	ALL windows (N = 4749)	1.90E-49	1.30E-44	5.70E-09	0.00057
	ONLY windows with any evidence of introgression (N = 1950)	0.0004	0.0072	1.90E-06	3.00E-05
Browning introgressed haplotypes with Neanderthal filter	ALL windows (N = 4749)	1.10E-28	3.30E-25	1.20E-07	0.00047
	ONLY windows with any evidence of introgression (N = 1604)	0.042	0.19	0.0001	0.00038
Vernot introgressed haplotype	ALL windows (N = 4749)	1.50E-02	5.40E-02	0.0015	0.005
	ONLY windows with any evidence of introgression (N = 2657)	3.40E-05	8.40E-07	0.00068	1.60E-05

Table S8: Compared to sequence variability, 3D variability is a relatively more informative predictor of amount of introgression when considering windows of the genome with any introgression. When considering the relationships between 3D genome variability, sequence variability, and amount of introgression (Supplemental Text, Figs. 5, S12), we consider a subset of windows with any evidence of introgression (gray rows) compared to all windows (white rows). 3D variability is relatively more informative about the amount of introgression when only considering windows of the genome with any introgressed sequence present (last column). The analysis was replicated on three sets of introgressed haplotypes. Browning introgressed haplotypes are Sprime segments Browning haplotypes with Neanderthal-matching filter are a subset of the Browning et al. [93] introgressed segments that have 30 putatively introgressed variants that could be compared to the Altai Neanderthal genome and had a match rate of at least 30% to the Altai Neanderthal allele. Vernot haplotypes are S* segments from Vernot et al. [15].