Recent Regulatory Changes Shaped Human Facial and Vocal Anatomy

David Gokhman[1], Lily Agranat-Tamir[1,13], Genevieve Housman[2], Malka Nissim-Rafinia[1], Maria Nieves-Colon[3], Hongcang Gu[4], Manuel Ferrando[5], Pere Gelabert[5], Iddi Lipende[20], Ellen E. Quillen[3], Alexander Meissner[4], Anne C. Stone[2,6,7], Anne E. Pusey[8], Deus Mjungu[20], Leonid Kandel[9], Meir Liebergall[9], María E. Prada[10], Julio M. Vidal[11], Johannes Krause[12], Benjamin Yakir[13], Svante Pääbo[14], David Reich[4,15,16], Carles Lalueza-Fox[5], Tomas Marques-Bonet[5,17,18], Eran Meshorer[1,19,*], Liran Carmel[1,*]

[1] Department of Genetics, The Alexander Silberman Institute of Life Sciences, Faculty of Science, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, Israel.
[2] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA.
[3] Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas 85287, USA.
[4] Broad Institute, Cambridge MA 02138, USA.
[5] Institute of Evolutionary Biology (UPF-CSIC), 08003 Barcelona, Spain.
[6] Center for Bioarchaeological Research, Arizona State University, Tempe, AZ 85287, USA.
[7] Institute of Human Origins, Arizona State University, Tempe, AZ 85287, USA.
[8] Department of Evolutionary Anthropology , Duke University , Durham, NC 27708, USA.
[9] Orthopaedic Department, Hadassah – Hebrew University Medical Center, Jerusalem, Israel.
[10] I.E.S.O. 'Los Salados'. Junta de Castilla y León, Spain
[11] Junta de Castilla y León, Servicio de Cultura de León, Spain
[12] Max Planck Institute for the Science of Human History, 07745 Jena, Germany.
[13] Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel.
[14] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany.
[15] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.
[16] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA.
[17] Catalan Institution of Research and Advanced Studies (ICREA), 08010 Barcelona, Spain.
[18] Centro Nacional de Análisis Genómico (CRG-CNAG), 08028 Barcelona, Spain.
[19] The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, 91904, Israel
[20] Gombe Stream Research Center, Jane Goodall Institute, Kigoma, Tanzania

Identifying changes in gene regulation that shaped human-specific traits is critical to understanding human adaptation. Here, we use dozens of ancient and present-day DNA methylation maps to detect regulatory changes that emerged in modern humans. We show that genes affecting vocalization and facial features went through particularly extensive changes in methylation. Especially, we identify expansive changes in a network of genes regulating skeletal development (*SOX9*, *ACAN* and *COL2A1*), and in *NFIX*, which controls facial projection and voice box (larynx) development. We propose that these changes played a key role in shaping the human face, and in forming the human 1:1 vocal tract configuration that is considered optimal for speech. Our results provide insights into the molecular mechanisms that shaped the modern human face and voice, and suggest that they arose after the split from Neanderthals and Denisovans.

The advent of high-quality ancient genomes of archaic humans (the Neanderthal and Denisovan) and anatomically modern humans (AMHs) opened up the possibility to identify genomic changes that are unique to AMHs, and may underlie some of our defining traits[1]. Although the biological impact of mutations in protein-coding regions is easier to predict, these mutations are rare[2]. The overwhelming majority of the ~30,000 fixed substitutions between archaic and present-day humans occurred at noncoding regions[2]. While most of them are probably nearly neutral, a sizeable minority likely affects gene regulation, especially those mutations that reside in such regions as promoters and enhancers. In fact, regulatory changes are thought to account for much of the phenotypic variation that differentiates between closely-related groups[3]. Today, our ability to understand the regulatory function and phenotypic effect of a noncoding substitution is very limited. Thus the importance in looking at regulatory layers such as DNA methylation, which provide direct information on gene regulation.

To this end, we have previously developed a method to reconstruct DNA methylation maps along ancient genomes[4]. This method is based on a predominant degradation process in ancient DNA, where methylated cytosines are deaminated with time into thymines, while unmethylated cytosines are deaminated into uracils[4–6]. The subsequent removal of uracils during ancient DNA library preparation creates a signal that enables the distinction between pre-mortem methylated and unmethylated positions[4–6]. Using this method, we have reconstructed the methylomes of a Neanderthal and a Denisovan, and compared them to a present-day osteoblast methylation map[4]. However, the ability to identify differentially methylated regions (DMRs) between the human groups was confined by the incomplete reference map, the differences in sequencing technologies, the lack of an outgroup and the restricted set of skeletal samples (see Methods).

Therefore, we have sought to establish a comprehensive assembly of skeletal DNA methylation maps. To the previously reconstructed Denisovan and Altai Neanderthal maps, we added the methylome of the ~40,000 years old (yo) Vindija Neanderthal, and three methylomes of ancient AMHs – the ~45,000 yo Ust'-Ishim indvidual[7], the ~8,000 yo Loschbour individual[8], and the ~7,000 yo Stuttgart individual[8]. In addition, we sequenced to high-coverage the ancient genome of the ~7,000 yo La Braña 1 AMH individual, which was previously sequenced only to low-coverage[9]. To obtain a full modern methylation map, we produced whole-genome bisulfite sequencing (WGBS) methylomes from the bones of two present-day individuals (hereinafter

bone1 and bone2). To this we added 53 publically available bone methylation maps from present-day individuals, produced using reduced-representation bisulfite-sequencing (RRBS)[10] and 450K methylation arrays[11,12]. The ancient and modern samples come from a variety of bones and teeth, from both sexes, from individuals of various ages and of different ancestries (Extended Data Table 1). Therefore, DMRs that are identified between the human groups are unlikely to stem from variability that is driven by any of these factors. As an outgroup, we produced six chimpanzee methylomes (WGBS, RRBS, and four 850K methylation arrays, Extended Data Table 1). Together, these data establish a unique and comprehensive platform to study DNA methylation dynamics at the population level in recent human history.

In order to minimize artifacts that might arise from comparing maps produced through different technologies, we used the reconstructed Ust'-Ishim methylome as the AMH reference, to which we compared the Altai Neanderthal and the Denisovan (see Methods). Over 99% of the genome showed no significant variation between the three samples, but 18,080 loci showed methylation differences separating these individuals. Notably, these DMRs do not necessarily represent differences between the human groups. Rather, many of them could be attributed to variability within the AMH population, within archaic humans, or between extant and extinct humans. To account for this, we used our large sample collection to filter out regions where such variability is detected, leaving us with a set of 6,371 DMRs that discriminate between human groups. Finally, using the chimpanzee samples, we were able to polarize 3,869 of these DMRs into AMH-derived (1,667), archaic-derived (1,103), Neanderthal-derived (597), and Denisovan-derived (502). These DMRs were ranked according to their significance level (Extended Data Table 2, see Methods). Our focus in this work was on AMH-derived DMRs, as the large number of AMH samples used to filter out within-population variability makes these DMRs particularly robust. Throughout this work, we refer to three hierarchies of AMH-derived DMRs, differing by the number of samples used to determine them: i. DMRs that show little variability in methylation across full AMHs bone methylomes (1,667 DMRs, hereinafter full bone AMH-derived). ii. DMRs that show little variability across all AMH bone methylomes (including 450K and 850K methylation arrays, 1,100 DMRs, hereinafter bone AMH-derived). iii. DMRs that show little variability across all skeletal AMH methylomes, including teeth (881 DMRs, hereinafter skeletal AMH-derived, Extended Data Table 2, see Methods).

To gain insight into the function of these DNA methylation changes, we first analyzed the gene ontology (GO) of the AMH-derived differentially methylated genes (DMGs). As expected from a comparison between skeletal tissues, there are many enriched terms associated with the skeleton (e.g., *chondrocyte differentiation*, *proteoglycan biosynthetic process*, *cartilage development*, *embryonic skeletal system development* and *ossification*, FDR < 0.05). Also notable are terms associated with the skeletal muscle, cardiovascular and nervous system (Extended Data Table 3). To get a more precise picture of the possible functional consequences of these DMGs, we used Gene ORGANizer (geneorganizer.huji.ac.il, paper submitted). This is a phenotype-based tool that links genes to the organs where their phenotypes are observed, and allowed us to identify organs that are significantly over-represented. We tested our three hierarchies of AMH-derived DMGs in Gene ORGANizer and found that, regardless of the hierarchy level, genes that affect the voice are the most enriched (Fig. 1a, Extended Data Table 4). For example, when running the list of 881 skeletal AMH-derived DMRs, we identify 14 enriched body parts, with the strongest enrichment in the vocal cords (x2.18, FDR = 0.01), followed by the voice box (larynx, x1.73, FDR = 0.029) and then by body parts belonging primarily to the face, spine and pelvis (Fig. 1a, Extended Data Table 4). Interestingly, these parts are considered some of the most morphologically derived regions between Neanderthals and AMHs[13]. The voice-affecting DMGs (Table 1) were shown to shape voice production mainly through skeletal alterations of the larynx (where voice is produced) and vocal tract (the pharyngeal, oral and nasal cavities, where sound is filtered to specific frequencies). The phenotypes associated with these genes range from slight changes of the pitch and hoarseness of the voice, to a complete loss of speech ability (Table 1). When taking the top quartile of the most significant DMGs, the over-representation of voice-affecting genes becomes even more pronounced, with the vocal cords being enriched over 3-fold (FDR = 4.2×10^{-3}), and the larynx over 2-fold (FDR = 6.1×10^{-3}, Fig. 1b, Extended Data Table 4). The enrichment of the larynx and the vocal tract within AMH-derived DMGs is also apparent when examining patterns of gene expression. The tissues where these DMGs are most enriched are the pharynx and larynx (x1.78, 4.9×10^{-6}, x1.64, 5.3×10^{-7}, respectively, Extended Data Table 5). Importantly, we ruled out the possibility that the enrichment of the larynx within AMH-derived DMGs is attributed to potential biases driven by inherent characteristics of genes affecting the voice (e.g., potential differences in gene length or genomic distribution). We found no enrichment when simulating DNA

degradation and subsequent reconstruction of methylation maps (see Methods), nor do any of the other hominin branches show enrichment of voice-affecting genes (Extended Data Fig. 1, Extended Data Table 4). Finally, when comparing the chimpanzee samples to all human samples (archaic and AMH) to identify DMRs that separate chimpanzees from humans, we did not find any enrichment of genes that affect the voice, larynx or any parts of the vocal tract, supporting the notion that this trend emerged exclusively along the AMH lineage.

The ancient DNA methylation maps we use for DMR detection come from limb bones. At first glance, it may seem that the ability to make inferences about other body parts, such as the larynx, is limited. However, the laryngeal skeleton, and specifically the cricoid and arytenoid cartilages, which are central in vocalization, are very close developmentally to limbs, as both derive from the somatic layer of the lateral plate mesoderm. Furthermore, skeletal DMRs are largely devoid of regions that are variable across bones, and between bones and teeth, and are thus likely to be relevant to all skeletal parts. We conclude that voice-affecting genes are the most over-represented DMGs along the AMH lineage, regardless of inter-skeletal variability, coverage by methylation array probes, the extent to which a DMR is variable across human or chimpanzee populations, or the significance level of the DMRs (see Methods).

To date, it is still unclear what enabled humans to evolve such exceptional speech and language abilities, and why other primates did not develop similar capabilities. The capacity of humans to communicate vocally is attributed not only to neural changes, but also to structural alterations to the vocal tract[14,15]. The relative role of anatomy in our speech skills is still under debate[14,16], but it is nevertheless widely accepted that even with a human brain, other apes probably could not reach the human level of articulation[14,15]. Chimpanzees, for example, are restricted not only in their linguistic capacity (e.g., they can hardly learn grammar[15]), but also in their ability to produce the phonetic range that humans can. As a result, chimpanzees communicate through sign language and symbols much better than they do vocally, even after being raised in an entirely human environment[15]. Phonetic range is determined by the different conformations that the vocal tract can produce. These conformations are largely shaped by the position of the larynx, tongue, lips and mandible. Modern humans have a 1:1 proportion between the horizontal and vertical dimensions of the vocal tract, which is unique among primates[14,17] (Fig. 1c). The extent to which this configuration is a prerequisite for speech is still debated, but it was nonetheless shown to be

optimal for speech[14,17–19]. It enables the tongue to move both vertically and horizontally within the vocal tract, thus increasing the range of vocal tract shapes, and expanding the discriminable phonetic repertoire[14,15,19].

The 1:1 proportion was reached through a relative shortening of the human face, together with the descent of the larynx[20] (Fig. 1c). In order to test whether this anatomy is shared by archaic humans, researchers turned to analyze Neanderthal archaeological remains[21]. However, cartilage and soft tissues do not survive long after death and the only remnant from the laryngeal region is the hyoid bone[15]. Based on this single bone, on computer simulations or on tentative vocal tract reconstruction, it is difficult to characterize the full anatomy of the Neanderthal vocal apparatus. Thus, opinions remain split as to whether the Neanderthal, and even more so, the Denisovan, had similar vocal anatomy[15,21,22]. Thus, investigating archaic methylomes opens an opportunity to directly study the regulatory mechanisms that may underlie these changes.

To further examine DNA methylation changes in genes affecting vocal anatomy, we quantified the expanse of methylation change along the genome. To do so, we scanned the genome in windows of 100 kb and computed the fraction of CpGs which are differentially methylated in AMHs (hereinafter, AMH-derived CpGs). We found that this fraction is more than twice as high within genes affecting the voice compared to other genes (0.142 vs. 0.055, $P = 3.7\text{x}10^{-5}$, *t*-test). In fact, three out of the six body parts associated with genes with the highest fraction of AMH-derived CpGs are in the laryngeal region (vocal cords: 0.142, FDR = $2.9\text{x}10^{-4}$; epiglottis: 0.140, FDR = $4.4\text{x}10^{-3}$; larynx: 0.133, FDR = $2.6\text{x}10^{-4}$, Extended Data Table 6). Moreover, three of the top five DMGs with the highest fraction of AMH-derived CpGs affect the laryngeal skeleton (*ACAN*, *SOX9* and *COL2A1*, Fig. 2a,b). The fact that so many of the most derived genes affect the larynx is particularly surprising considering that only ~2% of the genome (502 genes) are known to affect it. Interestingly, the extra-cellular matrix genes *ACAN* and *COL2A1*, and their key regulator *SOX9*, form a network of genes that regulate skeletal growth, pre-ossification processes, and spatio-temporal patterning of skeletal development[23,24]. Although mainly involved in chondrogenesis, these genes were also shown to be active in osteogenic tissues of facial membranous bones. In late stages of development, *SOX9* activity was shown to persist in chondrocytes of the larynx and in osteoblasts of the lower face[25]. Hypermethylation of the *SOX9* promoter was shown to down-regulate its activity, and consequently, its targets[26]. *SOX9* is also

regulated by a series of upstream enhancers[27]. We show that its promoter and proximal (20kb upstream) enhancer[27] (covered by DMR #26), as well as its targets – *ACAN* (DMR #200) and *COL2A1* (the most significant AMH-derived DMR, DMR #1) – have all become hypermethylated in AMHs (Fig. 2c). Additionally, a more distant putative enhancer, located ~350kb upstream of *SOX9*, was shown to bear strong active histone modification marks in chimpanzee craniofacial progenitor cells, while in humans these marks are almost absent (~x10 stronger in chimpanzee)[28]. These epigenetic changes suggest that *SOX9* became down-regulated along the AMH lineage, followed by hypermethylation and possibly down-regulation of its targets – *ACAN* and *COL2A1* (Fig. 2c).

While this group of genes shapes many skeletal parts, the flattening of the face is the most common phenotype associated with their reduced activity. Heterozygous loss-of-function mutations in *SOX9*, which result in a reduction of ~50% in its activity, were shown to cause a retracted lower face, as well as affect the pitch of the voice[23,24]. *ACAN* was shown to affect facial prognathism and the hoarseness of the voice[29]. *COL2A1* is key for proper laryngeal skeletal development[30], and its decreased activity results in a retracted face[31]. Given that these genes have likely become down-regulated along the AMH lineage, and that they are key players in skeletal, and particularly facial development, we turned to investigate AMH-derived facial features. One of the main features separating archaic from modern humans is facial retraction. It was shown that the lower and midface of AMH is markedly retracted compared to apes, Australopithecines, and other *Homo* groups[20]. The developmental alterations that underlie the flattened ontogeny of the human face are still under investigation. Cranial base length and flexion were shown to play a role in the retracted face[20], but reduced growth rate, and heterochrony of spatio-temporal switches are thought to be involved as well[32]. Importantly, *SOX9* and *COL2A1* were implemented in the elongation and ossification of the basicranium[33,34], and *SOX9* is a key regulator of skeletal growth rate, and the developmental switch to ossification[23,24].

In order to further explore expression changes driven by changes in methylation, we focused on DMRs where methylation levels are strongly correlated with expression. Particularly noteworthy is *NFIX*, one of the most derived genes in AMH (Fig. 2b), which controls the balance between lower and upper projection of the face[35]. *NFIX* harbors two AMH-derived DMRs, where DNA methylation alone explains 73.9% and 81.7% of variation in its expression (FDR = 6.2×10^{-3} and

$7.5x10^{-4}$, Fig. 3a-c, see Methods). This relationship between *NFIX* methylation and expression was also shown previously[36], and suggests that it became down-regulated through hypermethylation along the AMH lineage. Interestingly, NFI proteins were shown to bind the upstream enhancers of *SOX9*[37], suggesting a possible mechanism to the simultaneous change in the voice- and face-affecting genes. In order to test whether these changes could explain morphological changes in AMHs, we examined the skeletal phenotypes that are driven by *NFIX*. Mutations in *NFIX* were shown to be behind the Marshall-Smith and Malan syndromes, which include various skeletal alterations such as hypoplasia of the midface, retracted lower jaw, and depressed nasal bridge[35], as well as limited speech capabilities[38]. In many cases, the syndromes are driven by nonsense-mediated decay, deletions or loss-of-function nonsense mutations, suggesting that the phenotypes associated with NFIX are dosage-dependent[35]. Given that reduced activity of *NFIX* drives these symptoms, a simplistic hypothesis would be that increased *NFIX* activity in the Neanderthal would result in changes in the opposite direction. Indeed, we found that in 18 out of 22 Marshall-Smith syndrome skeletal phenotypes, and in 8 out of 9 Malan syndrome skeletal phenotypes, the observed morphology is a mirror image of the Neanderthal. In other words, from the *NFIX*-related syndromes, through healthy AMHs, to the Neanderthal, the level of phenotype manifestation corresponds to the level of *NFIX* activity (Fig. 3c, Extended Data Table 7). Interestingly, many cases of laryngeal malformations in the Marshall-Smith syndrome have been reported[39]. Some of the patients exhibit positional changes of the larynx, changes in its width, and structural alterations to the arytenoid cartilage – the anchor point of the vocal cords, which controls their movement[39]. In fact, these laryngeal and facial anatomical changes are thought to underlie the limited speech capabilities observed in some patients[38].

Discussion

We have shown here that genes affecting vocal and facial anatomy have gone through particularly extensive regulatory changes in recent AMH evolution. These changes are evident both in the number of genes that changed and in the extent of changes within each gene. Interestingly, we found that it is not only at the DNA methylation level that such changes are detected. When using Gene ORGANizer to analyze genes associated with human accelerated regions (HARs, i.e., regions that are conserved across vertebrates but are significantly derived in humans[40]), we found that the epiglottis and vocal cords are the most over-represented body parts

(although the epiglottis enrichment is not significant; x2.28 and x1.91, $P = 0.052$ and $P = 0.033$, respectively). When examining primate accelerated regions, on the other hand, the representation of these body parts does not deviate from expectation (x1.03 and x1.04, $P = 0.445$ and $P = 0.447$, respectively).

We report here AMH-specific regulatory changes in genes affecting the larynx and face. Importantly, most of the studies that linked these genes to the organs they affect were based on sequence, rather than epigenetic, changes. While heterozygous loss-of-function mutations could be paralleled to partial silencing of a gene and thus provide more direct evidence of expression changes in the genes, further study is nevertheless required in order to fully characterize the phenotypic effect of the changes we report.

Our analysis account for the fact that methylation in bone tissues changes throughout development and across bone types. First, both adult and juvenile AMHs have similar methylation patterns in the DMGs we report, and both adult and non-adult archaic and chimpanzee samples are differentially methylated compared to AMH samples. Second, these differences hold throughout a wide range of bone types (Extended Data Table 1). This suggests that the observed DMRs probably reflect a true AMH-specific evolutionary shift, rather than variability related to age or bone type. This is also supported by the phenotypic observation that facial prognathism in general, and facial growth rates in particular, are derived and reduced in AMH[41].

SOX9, *ACAN* and *COL2A1* are active mainly in early stages of osteochondrogenesis, making the observation of differential methylation in mature bones puzzling at first glance. This could be explained by two factors: i. The DMRs might reflect earlier changes in the mesenchymal progenitors of these cells that are carried on to later stages of osteogenesis. ii. Although these genes are downregulated with the progress towards skeletal maturation, they were shown to persist into later skeletal developmental stages in the larynx, vertebras, limbs, and jaws, including in their osteoblasts[25,42,43]. Interestingly, these are also the organs that are most affected by mutations in these genes[23,24,29–31].

References

1. Krause, J. & Paabo, S. Genetic Time Travel. *Genetics* **203,** 9–12 (2016).

2. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–9 (2014).

3. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116 (1975).

4. Gokhman, D. *et al.* Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344,** 523–527 (2014).

5. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38,** (2010).

6. Pedersen, J. S. *et al.* Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res.* **24,** 454–466 (2014).

7. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–9 (2014).

8. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).

9. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).

10. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22,** 1680–1688 (2012).

11. Horvath, S. *et al.* The cerebellum ages slowly according to the epigenetic clock. *Aging (Albany. NY).* **7,** 294–306 (2015).

12. Lokk, K. *et al.* DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* **15,** r54 (2014).

13. Weaver, T. D. The meaning of Neandertal skeletal morphology. *Proc. Natl. Acad. Sci.* **106,** 16028–16033 (2009).

14. Lieberman, P. The Evolution of Human Speech: Its Anatomical and Neural Bases. *Curr. Anthropol.* **48,** 39–66 (2007).

15. Fitch, W. T. The evolution of speech: A comparative review. *Trends in Cognitive Sciences* **4,** 258–267 (2000).

16. Fitch, W. T., de Boer, B., Mathur, N. & Ghazanfar, A. A. Monkey vocal tracts are speech-ready. *Sci. Adv.* **2,** (2016).

17. Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M. & Palmer, J. B. Ontogeny of postnatal hyoid and larynx descent in humans. *Arch. Oral Biol.* **46,** 117–128 (2001).

18. Stevens, K. in *Human Communication A Unified View* 51–66 (1972).

19. De Boer, B. Modelling vocal anatomy's significant effect on speech. *J. Evol. Psychol.* **8,** 351–366 (2010).

20. Lieberman, D. E. *The Evolution of the Human Head*. (Harvard University Press, 2011).

21. Steele, J., Clegg, M. & Martelli, S. Comparative morphology of the hominin and african ape hyoid bone, a possible marker of the evolution of speech. *Hum Biol* **85,** 639–672 (2013).

22. Lieberman P. and McCarthy C. The Evolution of Speech and Language. *Handb. Paleoanthropology* (2014).

23. Meyer, J. *et al.* Mutational analysis of the SOX9 gene in campomelic dysplasia and autosomal sex reversal: Lack of genotype/phenotype correlations. *Hum. Mol. Genet.* **6,** 91–98 (1997).

24. Lee, Y. H. & Saint-Jeannet, J. P. Sox9 function in craniofacial development and disease. *Genesis* **49,** 200–208 (2011).

25. Ng, L. J. *et al.* Sox9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse. *Dev. Biol.* **183,** 108–121 (1997).

26. Kim, K. Il, Park, Y. S. & Im, G. Il. Changes in the epigenetic status of the SOX-9 promoter in human osteoarthritic cartilage. *J. Bone Miner. Res.* **28,** 1050–1060 (2013).

27. Yao, B. *et al.* The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. *Nucleic Acids Res.* **43,** 5394–5408 (2015).

28. Prescott, S. L. *et al.* Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp

Neural Crest. *Cell* **163,** 68–84 (2015).

29. Tompson, S. W. *et al.* A Recessive Skeletal Dysplasia, SEMD Aggrecan Type, Results from a Missense Mutation Affecting the C-Type Lectin Domain of Aggrecan. *Am. J. Hum. Genet.* **84,** 72–79 (2009).

30. Frenzel, K., Amann, G. & Lubec, B. Deficiency of laryngeal collagen type II in an infant with respiratory problems. *Arch. Dis. Child.* **78,** 557–9 (1998).

31. Hoornaert, K. P. *et al.* Stickler syndrome caused by COL2A1 mutations: genotype–phenotype correlation in a series of 100 patients. *Eur J Hum Genet.* **18,** 872–880 (2010).

32. Bastir, M., O'Higgins, P. & Rosas, A. Facial ontogeny in Neanderthals and modern humans. *Proc. R. Soc. B Biol. Sci.* **274,** 1125–1132 (2007).

33. Horton WA, Rimoin DL, Hollister DW, L. R. Further heterogeneity within lethal neonatal short-limbed dwarfism: the platyspondylic types. *J Pediatr.* **94,** 736–42 (1979).

34. Yan, Y.-L. *et al.* A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. *Development* **132,** 1069–1083 (2005).

35. Malan, V. *et al.* Distinct effects of allelic NFIX mutations on nonsense-mediated mRNA decay engender either a sotos-like or a Marshall-Smith Syndrome. *Am. J. Hum. Genet.* **87,** 189–198 (2010).

36. Carrió, E. *et al.* Deconstruction of DNA methylation patterns during myogenesis reveals specific epigenetic events in the establishment of the skeletal muscle lineage. *Stem Cells* **33,** 2025–2036 (2015).

37. Pjanic, M. *et al.* Nuclear Factor I genomic binding associates with chromatin boundaries. *BMC Genomics* **14,** 99 (2013).

38. Shaw, A. C. *et al.* Phenotype and natural history in Marshall-Smith syndrome. *Am. J. Med. Genet. Part A* **152,** 2714–2726 (2010).

39. A. Cullen á T. A. Clarke á T. P. O'Dwyer. The Marshall-Smith syndrome: a review of the laryngeal complications. *Eur J Pediatr* **156,** 463–464 (1997).

40. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).

41. Lacruz, R. S. *et al.* Ontogeny of the maxilla in Neanderthals and their ancestors. *Nat. Commun.* **6,** 8996 (2015).

42. Rojas-Peña, M. L. *et al.* Characterization of distinct classes of differential gene expression in osteoblast cultures from non-syndromic craniosynostosis bone. *J. genomics* **2,** 121–30 (2014).

43. Moriarity, B. S. *et al.* A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nat. Genet.* **47,** 615–24 (2015).

44. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* **2,** 1756–1762 (2007).

45. Simbolo, M. *et al.* DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS One* **8,** (2013).

46. Boyle, P. *et al.* Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol.* **13,** R92 (2012).

47. Barnett, R. & Larson, G. A phenol-chloroform protocol for extracting DNA from ancient samples. *Methods Mol. Biol.* **840,** 13–19 (2012).

48. Triche, T. J., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41,** e90 (2013).

49. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15,** 503 (2014).

50. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30,** 1363–1369 (2014).

51. Fortin, J.-P., Triche, T. & Hansen, K. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array. *bioRxiv* 65490 (2016). doi:10.1101/065490

52. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12,** 115–121 (2015).

53. Hernando-Herraez, I. *et al.* Dynamics of DNA Methylation in Recent Human and Great Ape Evolution. *PLOS Genet* **9,** e1003763 (2013).

54. Ong, M.-L. *et al.* Infinium Monkeys: Infinium 450K Array for the Cynomolgus macaque (Macaca fascicularis). *G3 Genes|Genomes|Genetics* **4,** 1227–1234 (2014).

55. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

56. McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9,** 22–24 (2016).

57. Gokhman, D., Meshorer, E. & Carmel, L. Epigenetics: It's Getting Old. Past Meets Future in Paleoepigenetics. *Trends Ecol. Evol.* **31,** (2016).

58. Hanghøj, K. *et al.* Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps with epiPALEOMIX. *Mol. Biol. Evol.* msw184 (2016). doi:10.1093/molbev/msw184

59. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–329 (2015).

60. K??hler, S. *et al.* The Human Phenotype Ontology project: Linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **42,** (2014).

Acknowledgements

We would like to thank Sagiv Shifman, Yoel Rak, Philip Lieberman, Rodrigo Lacruz, Achinoam Blau, and Daniel Lieberman for their useful advice, and Maayan Harel for illustrations. FUNDING

Tables and Figures

Table 1. DMRs in genes that affect the voice and the larynx.

DMG	Associated phenotype	Chr	DMR start	DMR end
ALPL	Abnormality of the voice	1	21901961	21907487
AHDC1	Laryngomalacia	1	27869253	27871400
AHDC1	Laryngomalacia	1	27917471	27921806
SATB2	Abnormality of the voice	2	200236735	200244763
SPEG	Dysphonia	2	220316303	220319764
COLQ	Weak cry	3	15508914	15512536
TGFBR2	Abnormality of the voice	3	30649533	30658854
TGFBR2	Abnormality of the voice	3	30674279	30680742
TGFBR2	Abnormality of the voice	3	30706167	30710950
POC1A	High pitched voice	3	52110680	52112683
PLXND1	Abnormality of the voice	3	129312022	129315078
SH3BP2	Abnormality of the voice	4	2796208	2800983
SDHA	Hoarse voice, loss of voice, vocal cord paralysis	5	251676	254993
GLI3	Laryngeal cleft	7	42212811	42214593
CHD7	Abnormality of the voice, Laryngomalacia	8	61679558	61684133
HNRNPA1	Bowing of the vocal cords, hoarse voice	12	54679251	54682731
TRPV4	Vocal cord paresis	12	110248589	110250088
MEIS2	Laryngomalacia	15	37217518	37219852
ACAN	Hoarse voice	15	89333945	89344957
CREBBP	Laryngomalacia	16	3828787	3834862
CREBBP	Laryngomalacia	16	3891316	3900883
XYLT1	High-pitched voice	16	17428938	17431410
WWOX	Abnormality of the voice	16	78707061	78709972
WWOX	Abnormality of the voice	16	79038137	79040340
SOX9	Laryngomalacia	17	70077734	70113643
SOX9	Laryngomalacia	17	70119247	70120418
GNAL	Laryngeal dystonia	18	11747116	11748993
NFIX	Laryngomalacia	19	13155588	13158871
NFIX	Laryngomalacia	19	13185658	13192650
POLD1	High-pitched voice	19	50883926	50885758
RIN2	High-pitched voice	20	19944783	19947262
TBX1	Abnormality of the voice, nasal speech	22	19748985	19750495

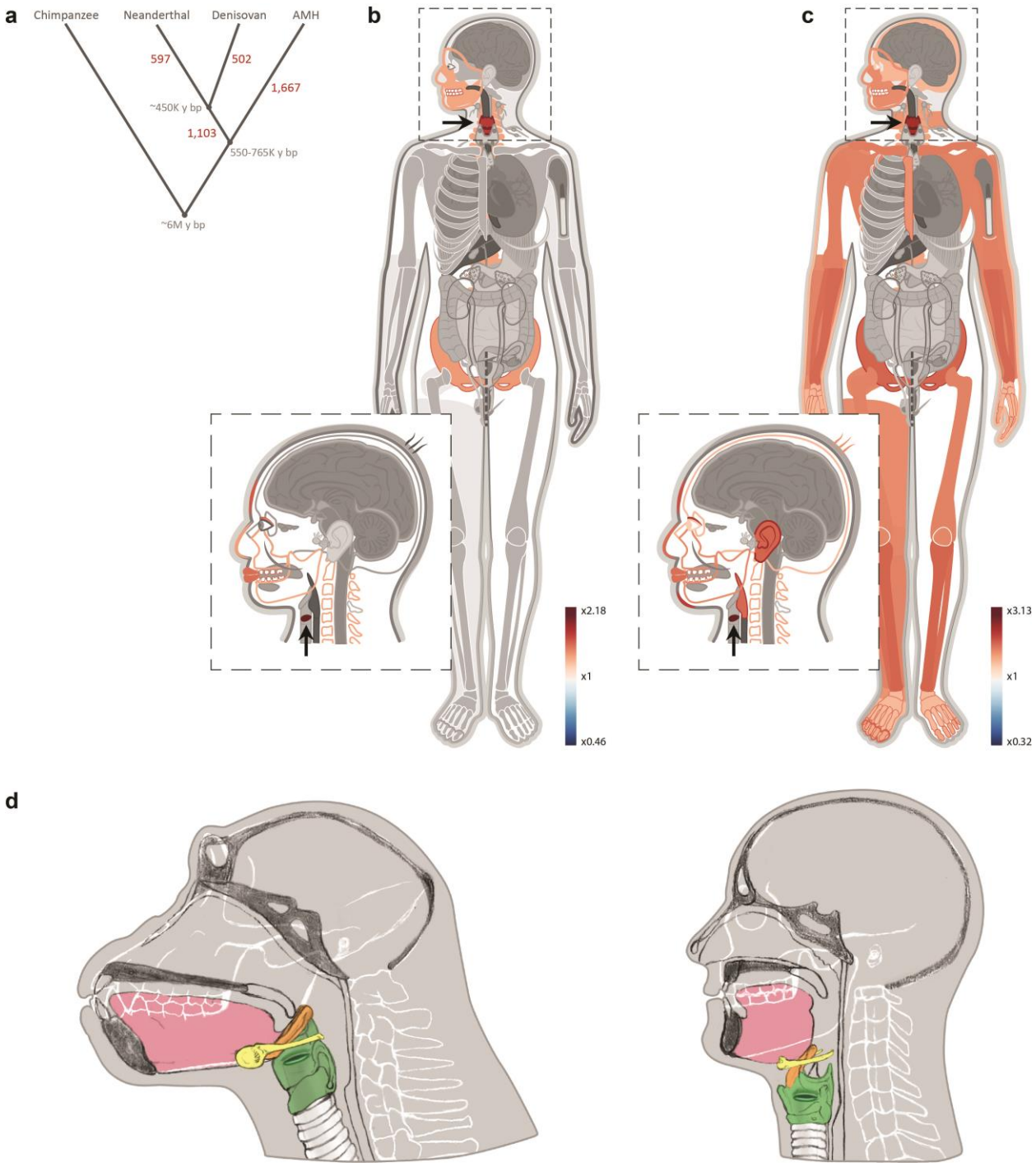


Fig. 1. Genes affecting the voice are the most over-represented within AMH-derived DMRs. a. The number of DMRs along each of the human branches. Split times are in years (y) before present (bp). **b.** A heat map representing the level of enrichment of each anatomical part within the skeletal AMH-derived DMRs. Only body parts that are significantly enriched (FDR < 0.05) are colored. Most enriched parts are within the head and neck region, with genes affecting

the vocal cords and voice box (larynx, marked with arrows) being the most over-represented. **c.** Enrichment levels of the most significant (first quartile) skeletal AMH-derived DMRs, showing an even more pronounced over-representation of genes affecting vocal anatomy. **d.** Vocal anatomy of the chimpanzee and AMH. The flattening of the AMH face, as well as the descent of the larynx, produced a 1:1 configuration of the cavity from the lips to the larynx (i.e., the vocal tract). In AMH, the horizontal and vertical proportions of the vocal tract are equal, whereas chimpanzees have a longer horizontal and a shorter vertical vocal tract. Colors mark central body parts: larynx and vocal cords (green), epiglottis (orange), hyoid bone (yellow), and tongue (pink).

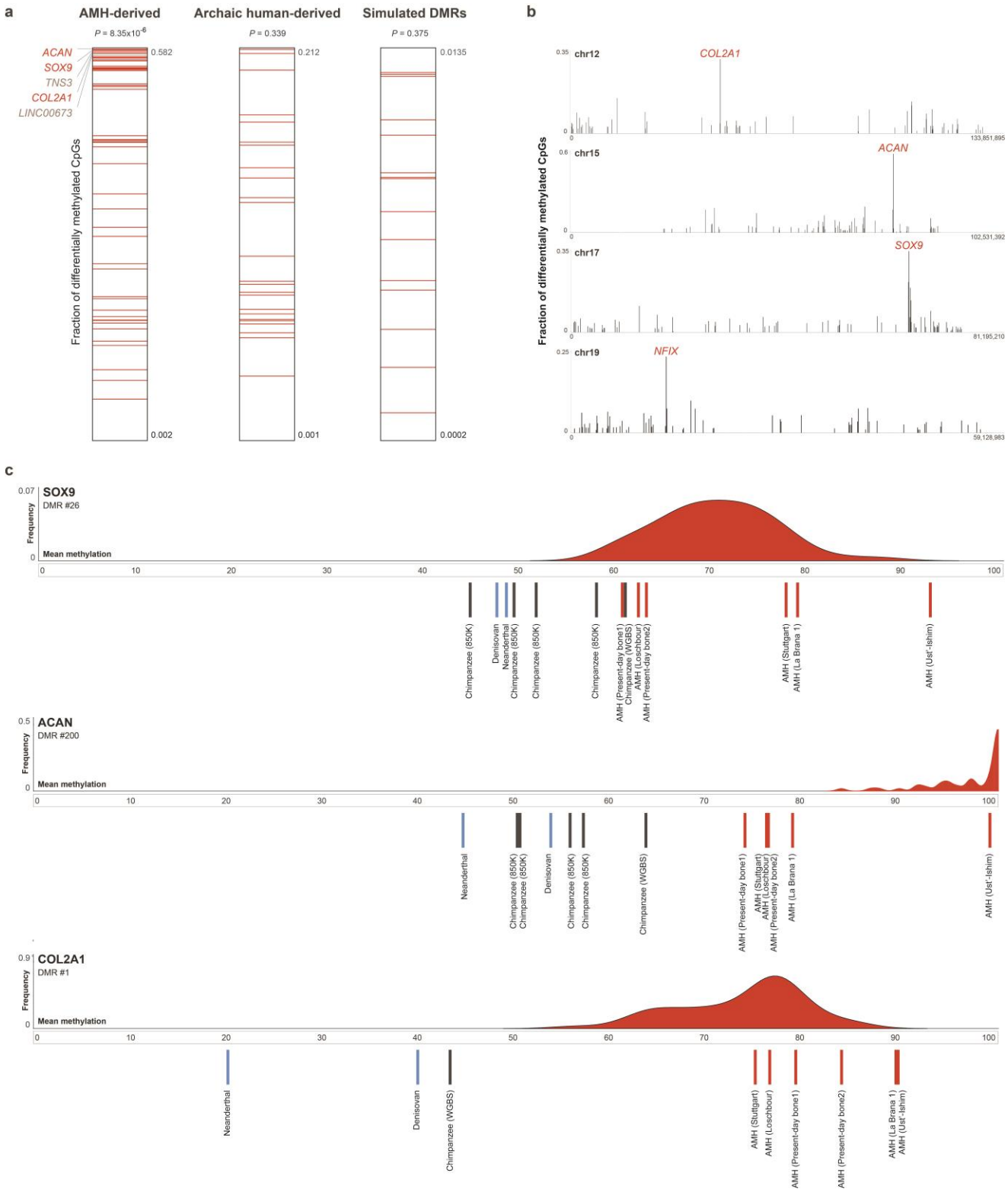


Fig. 2. The extent of differential methylation in AMHs is highest among genes affecting the larynx. a. The number of AMH-derived CpGs per 100 kb centered around the middle of each

DMR. Genes were ranked according to the fraction of derived CpG positions within them. Genes affecting the voice are marked with red lines. In AMH-derived DMGs, these genes went through more extensive changes compared to other genes, and tend to be ranked higher. Although these genes comprise ~2% of the genome, three of the top five AMH-derived windows overlap genes affecting the voice. In archaic-derived DMRs and in simulated DMRs, voice-affecting genes do not harbor more changes compared to the rest of the genome. **b.** The fraction of AMH-derived CpGs along chromosomes 12, 15, 17 and 19. The most extensive changes are found within the genes *COL2A1*, *SOX9*, *ACAN*, and *NFIX*. All of these genes control facial projection and the development of the larynx. **c.** Methylation levels in the skeletal AMH-derived DMRs in *SOX9*, *ACAN* and *COL2A1*. AMH samples are marked with red lines, archaic human samples are marked with blue lines and chimpanzee samples are marked with grey lines. The distribution of methylation across 52 AMH samples (450K methylation arrays) is presented as a red plot on top.

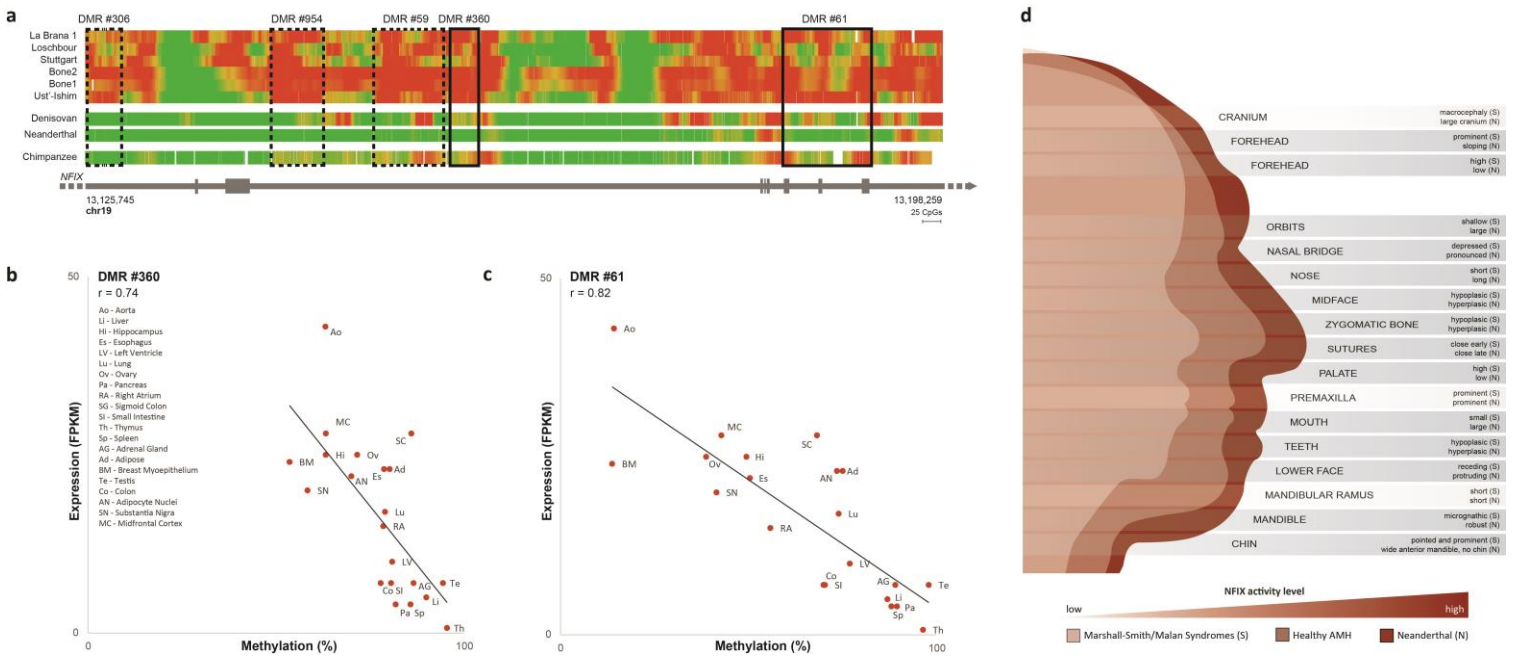
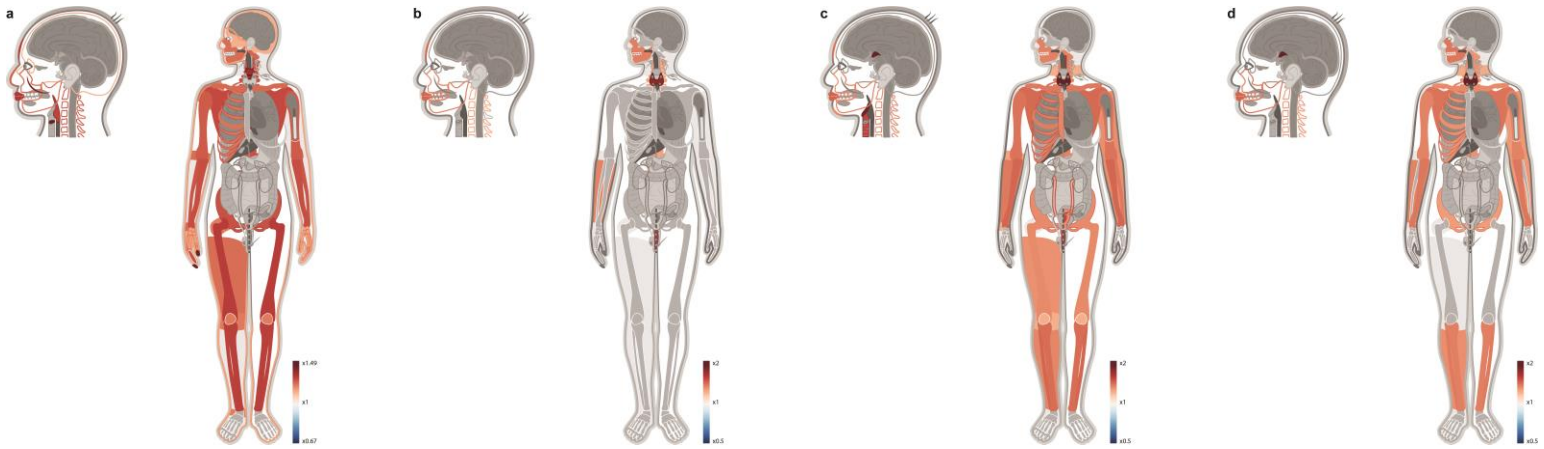
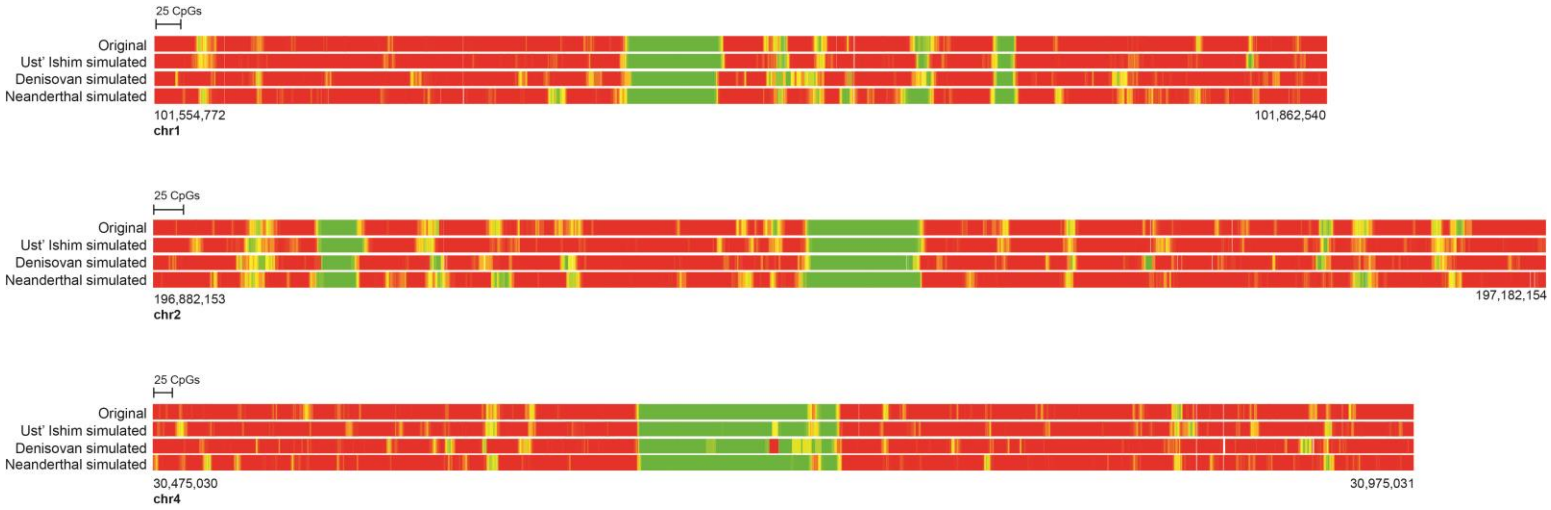


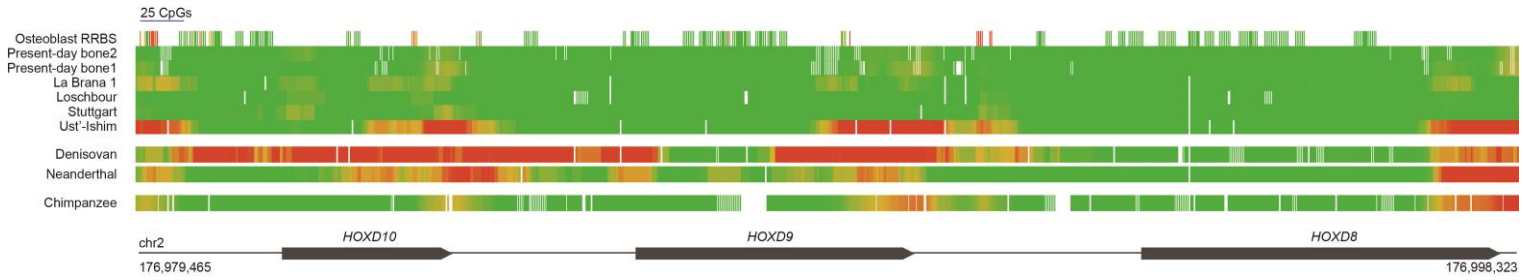
Fig. 3. *NFIX* became down-regulated after the split from archaic humans, and possibly underlies the flattening of the face in AMHs. a. Methylation levels along *NFIX*, color-coded from green (unmethylated) to red (methylated). The top six panels show ancient and present-day AMH samples, where *NFIX* is mostly methylated. The bottom three panels describe the Denisovan, Neanderthal and chimpanzee, where the gene is mostly unmethylated. AMH-derived DMRs across full bone methylomes are shown in dashed rectangles, and AMH-derived DMRs across all skeletal methylomes are shown in plain rectangles. Chimpanzee and present-day samples were smoothed using the same sliding window as in ancient samples to allow easier comparison. **b,c**. Methylation levels in DMRs #360 and #61 vs. expression levels of *NFIX* across 21 tissues. In both DMRs, higher methylation is associated with lower expression of *NFIX*. **d**. Craniofacial features of the Neanderthal, healthy AMH, and AMH with Marshall-Smith or Malan syndromes. *NFIX* controls the upper vs. lower prognathism of the face. Individuals where *NFIX* is partially or completely inactive present phenotypes that are largely the opposite of the Neanderthal facial features. For each facial part we show the phenotype of the Marshall-Smith and Malan syndromes (S), as well as the corresponding Neanderthal (N) phenotype. Phenotypes are compared to a healthy AMH. Opposite phenotypes are marked with dark grey rectangles, and shared phenotypes are marked with light grey rectangles.



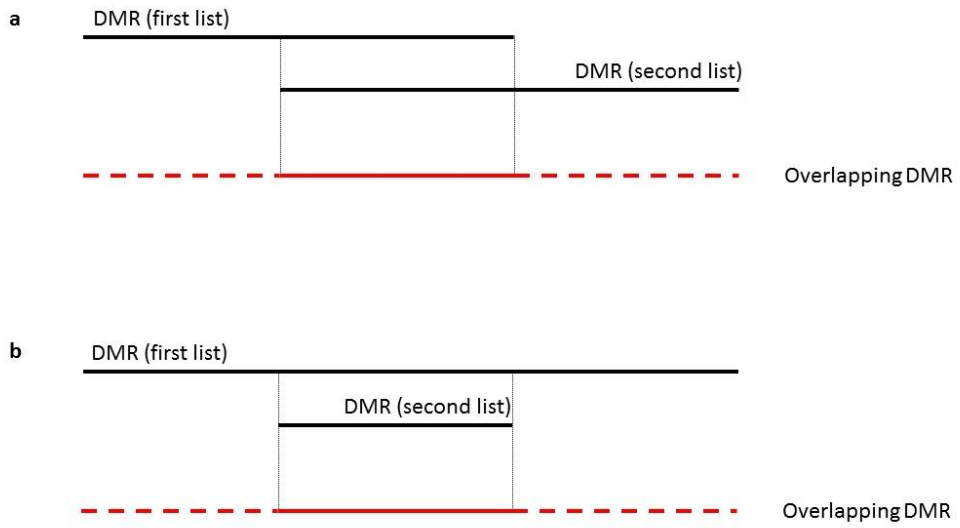
Extended Data Fig. 1. Genes affecting endocrine glands and the skeleton are the most enriched within archaic-derived DMRs. A heat map representing the level of enrichment of each anatomical part within the archaic-derived DMRs. Only body parts that are significantly enriched (FDR < 0.05) are colored. **a.** Enrichment within full bone AMH-derived DMRs. **b.** Enrichment within archaic-derived DMRs with little variability across AMH full bone methylomes. **c.** Enrichment within archaic-derived DMRs with little variability across all AMH bone methylomes. **d.** Enrichment within archaic-derived DMRs with little variability across all AMH skeletal methylomes.



Extended Data Fig. 2. Simulations of cytosine deamination, followed by reconstruction reproduce DNA methylation maps. Deamination was randomly drawn for each position based on its methylation level, read coverage and the observed rate of deamination in each hominin. Then, DNA methylation maps were reconstructed and matched against the original map. The number of DMRs found were used as an estimate of false discovery rate. Three exemplary regions are presented.



Extended Data Fig. 3. The HOXD cluster is hypermethylated in archaic humans, and in the Ust'-Ishim individual. Methylation levels along *HOXD8, HOXD9* and *HOXD10*, color-coded from green (unmethylated) to red (methylated). The top seven panels show ancient and present-day AMH samples, then the Denisovan, Neanderthal and chimpanzee. The promoter region of *HOXD9* is hypermethylated in the Neanderthal and the Denisovan, but not in AMHs. The 3' ends of the three genes are hypermethylated in the Neanderthal, Denisovan, Ust'-Ishim and chimpanzee, but not in other AMH samples. The promoter of HOXD10 is methylated only in the Denisovan.



Extended Data Fig. 4. Intersection procedure for overlapping DMRs. Plain red regions represent overlapping regions. Dashed red lines represent regions that were identified as a DMR in one of the lists, but not the other. These regions were added to the overlapping region if the clustered significantly closer to the reference human.

Methods

Skeletal Methylation Maps

Previously, our ability to identify differentially methylated regions (DMRs) that discriminate between human groups was confined by three main factors: i. We had a single DNA methylation map from a present-day human bone, which was produced using a reduced representation bisulfite sequencing (RRBS) protocol, which provides information for only ~10% of the CpG positions in the genome. Moreover, the fact that the archaic and present-day methylomes were produced using different technologies – computational reconstruction versus RRBS – potentially introduces a bias. ii. The analyses included only one bone methylation map from each of the human groups, which limited our ability to identify fixed differences between the groups. While dozens of maps from additional tissues in present-day humans were included in the analyses, this narrowed the DMRs to represent only human-specific changes that are invariable between tissues. iii. The work did not include a great ape outgroup. Thus, when an AMH-specific change was identified, it was impossible to determine whether it happened on the AMH lineage, or in the ancestor of Neanderthals and Denisovans[4].

To overcome these obstacles, one of the goals of the current study was to significantly extend the span of our skeletal methylome collection, covering as many individuals, sexes, and bone types as we could. This included the generation of many new samples, including the high-coverage sequencing of additional ancient genome, as listed below.

Present-day human bone DNA methylation maps

We generated full DNA methylation maps in two healthy bones from present-day human using whole-genome bisulfite sequencing (WGBS). In addition, we collected more than 50 publically available partial skeletal methylation maps.

WGBS of two modern human bones

DNA Extraction

DNA was extracted from bones using QIAamp® DNA Investigator kit (56504, Qiagen). In brief, bones were cut to thin slices (0.2-0.5 mm) and then thoroughly washed (X5) with PBS, to clean

samples from blood. Bones were crushed with mortar and pestle in liquid nitrogen and 100 mg bone powder was taken to extract DNA according to the protocol "Isolation of Total DNA from Bones and Teeth" of the DNA Investigator kit.**WGBS**

Partial skeletal DNA Methylation maps of modern humans

Osteoblast RRBS map[10], extracted from the limb and rib bones of a 12 year-old female, was downloaded from GEO accession GSE27584. 48 450K methylation array maps, extracted from the femora of adult males and females[11], were downloaded from GEO accession GSE64490. Four 450K methylation array maps, extracted from unspecified bones of adult males and females[12], were downloaded from GEO accession GSE50192.

Chimpanzee bone DNA methylation maps

Overall, we produced six methylation maps from bones of six common chimpanzee (*Pan troglodytes*) individuals. They include one WGBS of a wild chimpanzee, one RRBS of a captive chimpanzee, and four 850K methylation arrays of captive chimpanzees.

Ethics Statement

Chimpanzee tissue samples included in this study were opportunistically collected at routine necropsy of these animals. No animals were sacrificed for this study, and no living animals were used in this study.

RRBS of a Chimpanzee bone

Sample collection

We used two unidentified long bone fragments that belonged to a wild chimpanzee infant who died during a documented infanticide event at Gombe National Park in 2009. The infant was known to be the offspring of a chimpanzee called Eliza. The sample was collected from the ground a day or two after the infant's death and stored in RNAlater solution until arrival to Arizona State University (ASU).

DNA Extraction

Sampling and DNA extractions were conducted at the ASU Ancient DNA Laboratory, a Class 10,000 clean-room facility in a separate building from the Molecular Anthropology Laboratory. Precautions taken to avoid contamination included bleach decontamination and UV irradiation of tools and work area before and between uses, and use of full body coverings for all researchers. The bone sample was pulverized in December 2012 using a SPEX CertiPrep Freezer Mill. Three DNA extractions were conducted using 50-100 mg of bone powder (Extended Data Table 8) and following the extraction protocol by Rohland and Hofreiter[44]. Two extraction blank controls were included to monitor contamination of the extraction process. One µL each of the sample extract and the blank control were used for fluorometric quantification with the Qubit 2.0 Broad Range assay[45]. All extracts were combined for a total volume of 345 µL and approximately 0.652 µg of total DNA.

RRBS

RRBS libraries were generated according to Boyle *et al.*[46]. 100-200 genomic DNA was digested with MspI. Subsequently, the digested DNA fragments were end-repaired and adenylated in the same reaction. After ligation with methylated adapters, samples with different adapters were pooled together and were subjected to bisulfite conversion using the EpiTect Bisulfite kit (QIAGen) per the manufacturer's recommendations with the following modification: after first bisulfite conversion, the converted DNA was treated with sodium bisulfite again to guarantee that conversation rates were no less than 99%. Two third of bisulfite converted DNA was PCR amplified and final RRBS libraries were sequenced in an Illumina Hiseq 2000 sequencer.

850K DNA methylation arrays

Samples collection

Four chimpanzee cadavers from captive colonies at the Southwest National Primate Research Center in Texas were used. Femora were opportunistically collected at routine necropsy of these animals and stored in -20°C freezers at the Texas Biomedical Research Institute after dissection. These preparation and storage conditions ensured the preservation of skeletal DNA methylation patterns.

DNA extraction

DNA was extracted from the femoral trabecular bone using a phenol-chloroform protocol optimized for skeletal tissues[47]. From the distal femoral condyles, trabecular bone was collected using coring devices and pulverized into bone dust using a SPEX SamplePrep Freezer/Mill. Specifically, bone cores were obtained from a transverse plane through the center of the medial condyle on the right distal femur, such that the articular surface remained preserved. Cortical bone was removed from these cores using a Dremel (Extended Data Table 9).

Genome-Wide DNA Methylation Profiling

Genome-wide DNA methylation was assessed using Illumina Infinium MethylationEPIC microarrays. These arrays analyze the methylation status of over 850,000 sites throughout the genome, covering over 90% of the sites on the Infinium HumanMethylation450 BeadChip as well as an additional 350,000 sites within enhancer regions. For each sample, 400ng of genomic DNA was bisulfite converted using the EZ DNA Methylation™ Gold Kit according to the manufacturer's instructions (Zymo Research), with modifications described in the Infinium Methylation Assay Protocol. Following manufacturer guidelines (Illumina), this processed DNA was then whole-genome amplified, enzymatically fragmented, hybridized to the arrays, and imaged using the Illumina iScan system.

Methylation Data Processing

Raw fluorescent data were normalized to account for the noise inherent within and between the arrays themselves. Specifically, we performed a normal-exponential out-of-band (Noob) background correction method with dye-bias normalization[48] to adjust for background fluorescence and dye-based biases and followed this with a between-array normalization method (functional normalization)[49] which removes unwanted variation by regressing out variability explained by the control probes present on the array as implemented in the minfi package in R[50,51] which is part of the Bioconductor project[52]. This method has been found to outperform other existing approaches for studies that compare conditions with known large-scale differences[49], such as those assessed in this study.

After normalization, methylation values (β values) for each site were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities. These β values range from 0 to 1 and represent the average methylation levels at each site across the entire population of cells from which DNA was extracted (0 = completely unmethylated sites, 1 = fully methylated sites).

$$\beta\ \text{Value} = \frac{\text{Methylated Signal}}{(\text{Methylated Signal} + \text{Unmethylated Signal})}$$

Every β value in the Infinium platform is accompanied by a detection p-value, and those with failed detection levels (p-value > 0.05) in greater than 10% of samples were removed from downstream analyses.

The probes on the arrays were designed to specifically hybridize with human DNA, so our use of chimpanzee DNA required that probes non-specific to the chimpanzee genome, which could produce biased methylation measurements, be computationally filtered out and excluded from downstream analyses. This was accomplished using methods modified from [53,54]. Briefly, we used blastn[55] to map the 866,837 50bp probes onto the chimpanzee genome (Assembly: Pan_tro_3.0, Accession: GCF_000001515.7) using an e-value threshold of e^{-10}. We only retained probes that successfully mapped to the genome, had only 1 unique BLAST hit, targeted CpG sites, had 0 mismatches in 5bp closest to and including the CpG site, and had 0-2 mismatches in 45bp not including the CpG site. This filtering retained 622,819 probes.

Additionally, β values associated with cross-reactive probes[56], probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, and probes targeting sites within the sex chromosomes were removed using the minfi package in R[50,51]. This filtering retained a final set of 576,804 probes.

Reconstructing of ancient DNA methylation maps

The Reconstruction procedure

Reconstruction of DNA methylation maps was performed on the genomes of the following individuals: Ust'-Ishim[7], Loschbour[8], Stuttgart[8], La Braña 1, and the Vindija Neanderthal, as well as on the previously published Altai Neanderthal and the Denisovan (Extended Data Table 1).

Additional human ancient genomes have been published to date, however, these were sequenced to a relatively low coverage (<5x), and thus, only crude methylation maps could be reconstructed from them. C→T ratio was computed for every CpG position along the hg19 (GRCh37) human genome assembly, for each of the samples, as previously described[4].

In order to exclude from the analyses positions that potentially represent pre-mortem C→T mutations, rather than post-mortem deamination, the following filters were applied: i. Positions where the sum of A and G reads was greater than the sum of C and T reads were excluded. ii. For genomes that were produced using single-stranded libraries (i.e., Ust'-Ishim, Altai Neanderthal, Denisovan, Vindija Neanderthal and ~1/3 of the Loschbour library), positions where the G→A ratio on the opposing strand was equal or greater than 2/(genome coverage/2) were excluded. This fraction represents a threshold of one sequencing error allowed per position.. For Loschbour, this was performed only on the fraction of reads that came from the single stranded library. iii. For all genomes, positions with a C→T ratio > 0.25 were discarded. For the Vindija Neanderthal, this threshold was raised to 0.5, due to its relatively low coverage (~7x). iv. Finally, a maximum coverage threshold of 100 reads was used to filter out regions that are suspected to be PCR duplicates.

In all genomes, excluding Vindija, a fixed sliding window of 25 CpGs was used smoothing of the C→T ratio. This allowed for an unbiased scanning of differentially methylated regions (DMRs) that is not affected by the size of the window. Due to its relatively low coverage, we extended the sliding window used on the Vindija genome to 50 CpGs. This extended window is not expected to introduce a bias, as this genome was not used for DMR detection, but only for subsequent filtering that was applied equally to all genomes (see later).

As previously described, C→T ratio was translated to methylation percentage using linear transformation determined from two points: mean C→T ratio in completely unmethylated (0% methylation) CpG positions in modern human bone reference (hereinafter μ_0) was set to the value 0% methylation, and mean C→T ratio in completely methylated (100% methylation) CpG positions in modern human bone reference (hereinafter μ_{100}) was set to the value 100% methylation. Positions where C→T ratio > μ_{100} were set to 100% methylation, and positions where C→T ratio < μ_0 were set to 0% methylation. For genomes that were extracted from bones, the modern osteoblast RRBS map was used as reference. For genomes that were extracted from

teeth, there was no available modern reference methylation map, and therefore, we transformed the C→T ratio into methylation percentage based on the assumption that the genome-wide mean methylation is similar to bone tissue. Thus, the genome-wide mean C→T ratio represents 75% methylation, which is the genome-wide mean of measured methylation in the bone references. This was accomplished by setting μ_0 to 0, and setting μ_{100} to 1.33 x mean genome-wide C→T ratio.

DMR detection

DMR-detection algorithm

We developed an algorithm specifically designed to identify DMRs between a deamination map and a full methylome reference. Let i enumerate the CpG positions in the genome. In the deamination map, let t_i be the number of T's at the C position + the number of A's in the opposite strand at the G position, i.e., it counts the total number of T's that appear in a position that is originally C, in the context of a CpG dinucleotide. We similarly use c_i to count the total number of C's that appear in a position that is originally C, in the context of a CpG dinucleotide. The C→T ratio is defined as t_i/n_i, where $n_i = c_i + t_i$. Let φ_i and ψ_i (both between zero and one) be the methylation of this position in the reference genome and in the reconstructed one, respectively. If we denote by π the deamination rate, assumed to be constant throughout the genome, and if we assume that deamination of C into T is a binomial process with probability of success $\pi\psi_i$, we get

$$t_i \sim B(n_i, \pi\psi_i). \tag{1}$$

Our null hypothesis is that the ith CpG is not part of a DMR, namely that $\psi_i = \varphi_i$. The alternative hypothesis states that this CpG is part of a DMR. The definition of this statement is that $|\psi_i - \varphi_i| \geq \Delta$, where Δ is some pre-specified threshold. In other words, under the alternative hypothesis we get that $\psi_i \geq \varphi_i + \Delta$ if the site has low methylation in the reference genome, and $\psi_i \leq \varphi_i - \Delta$ if it has high methylation in the reference genome.

Per-site statistic

Let us start with the first option, testing whether $\psi_i \geq \varphi_i + \Delta$ when φ_i is low. A log-likelihood-ratio statistic would be

$$\ell_i^+ = \ln \frac{\Pr\left(t_i \middle| n_i, \pi(\varphi_i + \Delta)\right)}{\Pr(t_i | n_i, \pi\phi_i)} = t_i \left[\ln\left(1 + \frac{\Delta}{\varphi_i}\right) - \ln \frac{1 - \pi(\varphi_i + \Delta)}{1 - \pi\varphi_i}\right] + n_i \ln \frac{1 - \pi(\varphi_i + \Delta)}{1 - \pi\varphi_i}.$$

Similarly, we can test whether $\psi_i \leq \varphi_i - \Delta$ when φ_i is high using the log-likelihood-ratio statistic

$$\ell_i^- = \ln \frac{\Pr\left(t_i \middle| n_i, \pi(\varphi_i - \Delta)\right)}{\Pr(t_i | n_i, \pi\phi_i)} = t_i \left[\ln\left(1 - \frac{\Delta}{\varphi_i}\right) - \ln \frac{1 - \pi(\varphi_i - \Delta)}{1 - \pi\varphi_i}\right] + n_i \ln \frac{1 - \pi(\varphi_i - \Delta)}{1 - \pi\varphi_i}.$$

We used the value $\Delta = 0.5$ for all samples. The value of π, the deamination rate, was estimated using the overall C→T ratio in CpG positions whose methylation level is 1 in the modern human RRBS methylation map, after exclusion of putative pre-mortem substitutions, as described in the *DNA methylation reconstruction* chapter (Extended Data Table 1).

Detecting DMRs

The statistics ℓ_i^+ and ℓ_i^- quantify how strongly the estimated methylation in position i deviates from φ_i. Next, we use these values to identify DMRs using the cumulative-sum procedure explained below. The process is repeated twice: on the statistic ℓ_i^+ to identify DMRs where the sample has elevated methylation with respect to the reference, and on the statistic ℓ_i^- to identify DMRs where the sample has reduced methylation with respect to the reference.

For convenience, we explain the cumulative-sum procedure in the context of ℓ_i^+, but an essentially identical procedure is used for ℓ_i^-. We define a new vector Q^+ by the recursion

$$Q_0^+ = 0, \qquad Q_i^+ = \max(Q_{i-1}^+ + \ell_i^+, 0).$$

Under the null hypothesis, ℓ_i^+ has a negative expectation which produces a negative drift that keeps Q^+ at zero, or close to zero, levels. Under the alternative hypothesis the expectation is positive, hence the drift over a DMR is positive, leading to an elevation in the values of Q^+. Therefore, our next step is to find all intervals $[a, b]$ such that $Q_{a-1}^+ = 0$, $Q_{b+1}^+ = 0$, and $Q_i^+ > 0$

for $a \leq i \leq b$. Let Q_m^+ be the maximum value of Q^+ in this interval, where m is the position of the maximum. Then, the interval $[a, m]$ would be called a putative DMR.

Filtering DMRs

Of course, Q^+ may increase locally due to randomness, and thus a putative DMR may not reflect a true DMR. To filter out such intervals, we used two strategies. First, we applied a set of filters to assure that the putative DMRs have reasonable biological properties. Second, we cleaned the remaining putative DMRs by applying a false discovery rate (FDR) procedure. In the first strategy, we applied three filters: (i) Putative DMRs whose length was less than a 100 nt were removed. (ii) Putative DMRs that harbor less than 50 CpG positions, thus are shorter than twice the smoothing window size, were removed. (iii) To avoid situations where two consecutive CpG sites whose genomic locations are remote appear on the same DMR, we modify the vector Q_i^+ as follows. Let $d_{i,j}$ be the distance along the genome (in nucleotides) between CpG sites i and j. Then, for every site i such that $d_{i,i-1} > \delta$ we set $Q_i^+ = 0$. We used $\delta = 1000$ nt for all samples.

To further remove putative DMRs that are unlikely to reflect true DMRs, we eliminated all putative DMRs where $Q_m^+ < Q_T^+$. Here, Q_m^+ is the maximum value of Q^+ in the interval as defined earlier, and Q_T^+ is a threshold determined using a false discovery rate (FDR) procedure, see section "filtering out noise" below.

Testing the algorithm

To verify that the approach above results in a low number of false positives, we applied the procedure for deamination maps, when compared to themselves in the form of reconstructed methylomes. As expected, we obtained a negligible number of DMRs, ranging between 0.4% and 1% of the number of DMRs detected between the humans.

Two-way DMR detection

In order to avoid artifacts that could potentially be introduced by comparing DNA methylation maps that were produced using different technologies, our core analysis relied on the comparison of the three reconstructed maps of the Altai Neanderthal, Denisovan and Ust'-Ishim. These are all high-resolution maps that were derived from genomes sequenced to high coverage (Extended

Data Table 1). In particular, the Ust'-Ishim methylome is of exceptional quality due to its high coverage and deamination rate (Extended Data Table 1). Also, going through the same post-mortem degradation processes, the Ust'-Ishim cellular composition is likely to be similar to that of the Neanderthal and Denisovan.

In order for a deamination map to serve as a reference in the comparison, we have transformed its C→T ratio values into methylation values (see "the reconstruction procedure" section above). Therefore, the comparison of three genomes required a total of six two-way comparisons: Ust'-Ishim versus an Altai Neanderthal reference, Ust'-Ishim versus a Denisovan reference, Altai Neanderthal versus an Ust'-Ishim reference, Altai Neanderthal versus a Denisovan reference, Denisovan versus Ust'-Ishim reference, and Denisovan versus Altai Neanderthal reference. In order to remove potential bias that could be introduced through the comparison of a reconstructed methylation map to a deamination map, we ran each two-way comparison twice: once with the methylation map of sample 1 against the deamination map of sample 2, and once with the deamination map of sample 1 against the methylation map of sample 2. Overall, this resulted in twelve two-way comparisons. Because the DNA of these three individuals was extracted from both sexes, the DMR-detection algorithm was only applied to autosomal chromosomes.

Three-way DMR detection

In order to identify DMRs where one group of humans (hereinafter, hominin 1) differs from the other two human groups (hereinafter, hominin 2 and hominin 3), we set out to find those DMRs that were detected both between hominin 1 and 2, and between hominin 1 and 3. To this end, we compare the two lists (hominin 1 vs. hominin 2 and hominin 1 vs. hominin 3) and look for overlapping DMRs, as previously described[4]. An overlapping DMR exists when a DMR from one list partially (or fully) overlaps a DMR from the second list, and is constructed as follows (Extended Data Figure 4). The region of overlap between the two DMRs is taken as the core of the overlapping DMR. For a region that is included within the first DMR (hominin 1 vs. hominin 2) but not within the second DMR (hominin 1 vs. hominin 3), we used *t*-test to check whether methylation in hominin 3 clusters significantly closer to the hominin 2. If it does, the overlapping DMR was extended to include this region. An analogous test was used for regions that are

included in the second DMR but not in the first. *P*-values were adjusted using FDR, and only regions with FDR < 0.05 were taken as part of the overlapping DMR.

Filtering out noise

There are different factors that potentially introduce noise into the reconstruction process. These include the stochasticity of the deamination process, the use of a sliding window to smooth the C→T signal, and variations in read depth. In order to account for these factors and estimate noise levels, we ran simulations that mimic the post-mortem degradation processes of ancient DNA, then reconstructed methylation maps from the simulated deamination maps and finally compared them to the original map and identified DMRs.

The simulation process starts with a methylation map, where the measured or reconstructed methylation at position i is ψ_i and assumed the true methylation. Given that n_i is the coverage at this position, we use the binomial distribution (1) to randomly draw t_i – the number of C's that had become T's through deamination. The resulting t_i's, and their complement c_i's (where $c_i = n_i - t_i$) were then used to compute the C→T ratios for each position, smoothed and filtered using the same sliding window and thresholds used in the original analysis, and linearly transformed to methylation percentages as explained above (hereinafter, simulated methylation map, Extended Data Fig. 2). Any differences in methylation levels between the simulated map and the original reference map stem from noise. Thus, running the same DMR-detection algorithm described above on the simulated map vs. the reference map, enables an estimation of the false discovery rate. We ran these simulations 100 times for each of the three genomes (Altai Neanderthal, Denisovan, Ust'-Ishim) and determined the value of the Q_T^+ and Q_T^- thresholds (see section "filtering DMRs" above) such that the mean number of DMRs that are detected in the simulations is < 0.05 the number of real DMRs detected (i.e., FDR < 0.05).

Polarizing DMRs

DMRs where Ust'-Ishim differs from the Neanderthal and the Denisovan could either arose on the AMH branch, or in the ancestor of Neanderthals and Denisovans. In order to polarize the DMRs, i.e., allocating them to the branch in which the change occurred, we turned to the chimpanzee DNA methylation data.

First, we used the chimpanzee bone WGBS map. For every DMR and for each hominin H, we computed the mean absolute difference in methylation from chimpanzee, $d_{H,C} = \sum_{i \in DMR} |\psi_i^H - \psi_i^C|$. Here, ψ_i^H is the reconstructed methylation at the i'th CpG in hominin H, and ψ_i^C is the measured methylation in the same site in the chimpanzee. Any Ust'-Ishim-specific DMR where both archaic humans were closer to the chimpanzee, the DMR was placed on the AMH branch. If Ust'-Ishim was closer than both archaic humans to the chimpanzee, the DMR was placed on the branch of the ancestor of Neanderthals and Denisovans. Otherwise, the DMR was discarded. Out of 5,111 Ust'-Ishim-specific DMRs, we could place 1,729 DMRs on the AMH branch and 1,106 on the branch of the ancestor of Neanderthals and Denisovans. 1,956 Ust'-Ishim-specific DMRs where discarded due to inconclusive polarization, and 320 had no data in the chimpanzee WGBS map. Any Neanderthal-specific DMR where Ust'-Ishim and Denisova were not found to be closer to the chimpanzee than the Neanderthal were discarded. Out of 3,107 Altai Neanderthal-specific DMRs, 599 were placed on the Neanderthal branch, 2,296 were deemed inconclusive and discarded, and 212 had no data in the chimpanzee WGBS map. Similarly, any Denisovan-specific DMR where Ust'-Ishim and Altai Neanderthal were not found to be closer to the chimpanzee than the Denisovan were discarded. Out of 1,461 Denisovan-specific DMRs, 503 were placed on the Denisovan branch, 851 were deemed inconclusive, and for 107 we had no data in the chimpanzee WGBS map.

We next developed a second, stricter, polarization scheme by also using the chimp 850K DNA methylation arrays datasets. As the probes cover just part of the CpGs in a DMR, we need to adjust the DMR methylation level in order to allow a meaningful comparison of 850K methylation data to full methylation maps. If we mark by j the CpGs in a DMR that are covered by 850K methylation array (which is a subset of all the CpGs in this DMR), and mark their total number by $J = \sum_{j \in DMR} 1$, then the methylation in the DMR as measured by the array is $m = 1/J \cdot \sum_{j \in DMR} \psi_j^{\text{array}}$, where ψ_j^{array} is the methylation level measured at position j in the array. Let $m_I = \sum_{i \in DMR} \psi_i^{\text{WGBS}}$ be the methylation of this DMR as computed from the full methylation map, where ψ_i^{WGBS} is the methylation level measured at position i in the full map. Let $m_J = \sum_{j \in DMR} \psi_j^{\text{WGBS}}$ be the methylation as computed from the full methylation map when limited only to positions j. Then, we correct the array methylation value m to:

$$m' = \min\left(m \cdot \frac{m_I}{m_J}\right). \tag{2}$$

This procedure was applied to DMRs covered by at least one probe (~65% of DMRs). For the remaining ~35%, polarization was done using only the WGBS samples. This left 1,539 AMH-derived DMRs, 1,134 archaic-derived DMRs, 539 Neanderthal-derived DMRs and 460 Denisovan-derived DMRs. This polarization approach was used in parallel with the 450K-based approach used for filtering DMRs (see next chapter).

There are pros and cons to each of these polarization approaches. Using more chimpanzee datasets allow for more informative polarization. However, 850K methylation array probes are distributed unevenly across the genome. Although most DMRs are covered by at least one probe (mean number of probes per DMR: 1.7, median: 1, maximum: 64), many are nonetheless uncovered. On one hand, polarization of DMRs for which we have array data is more robust and less prone to misclassification. On the other hand, DMRs with array data are more likely to be filtered out, as there is more power to detect variability. This could potentially alter the genomic distribution of DMRs. Therefore, we use both approaches throughout the paper. In analyses where it is important to maintain an unbiased distribution of DMRs we only use the chimpanzee WGBS map for polarization, and AMH bone WGBS maps for filtering (see next chapter), whereas in analyses where it is more important to minimize variability, or where we look at specific DMRs, we use the stricter approach. The chimpanzee RRBS data was adjusted using the same technique. However, it was not used for polarization, but rather only as a source for additional information on DMRs. This is because this protocol particularly targets unmethylated CpGs, and is therefore too biased for polarization.

Removing DMRs with high within-group variability

Our three-way DMR detection algorithm above produces a list of DMRs where one of the three hominins (Ust'-Ishim, Altai Neanderthal or Denisovan) is significantly different from the other two. However, such DMRs could stem from variability within any of the groups, and in such cases cannot be regarded as truly differentiating between the human groups. Some variability may be removed during the polarization process, but even DMRs that had been successfully polarized do not necessarily represent fixed methylation changes. To filter out regions that are

variable within any of the human groups, or across all of them, we used three approaches. First, we used the two modern human WGBS maps. DMRs where the Neanderthal or Denisovan methylation levels were found within the range of modern human methylation (i.e., Ust'-Ishim and the two WGBS maps) were discarded. This left 1,667 out of 1,728 Ust'-Ishim-specific DMRs (hereinafter, full bone AMH-derived DMRs), 1,103 out of 1,106 DMRs where the Neanderthal and Denisovan are both derived (hereinafter, archaic-derived DMRs), 597 out of 599 Neanderthal-derived DMRs, and 502 out of 503 Denisovan-derived DMRs.

The second approach adds to this the 52 450K methylation array samples. As described above, using also methylation probes for filtering DMRs provides more power, but can also introduce biases. Thus, this filtering was used for most analyses, except those where unbiased genomic distribution of DMRs is critical. Probe methylation data was corrected as described in equation (2). Within AMH- and archaic-derived DMRs, a DMR was deemed fixed if the Neanderthal and the Denisovan methylation levels both fell outside the range of methylation across all modern humans samples (reconstructed, WGBS and 450K maps). Similarly, within Neanderthal and Denisovan DMRs, a DMR was deemed fixed if the respective hominin fell outside the range of methylation across both modern human samples and the other archaic hominin. This left 1,100 out of 1,667 AMH-derived DMRs (hereinafter referred to as bone AMH-derived DMRs), 976 out of 1,103 archaic-derived DMRs, 513 out of 597 Neanderthal-derived DMRs, and 441 out of 502 Denisovan-derived DMRs.

Finally, we used a third filtering, which also incorporated teeth methylomes, to identify DMRs that are fixed across all skeletal samples. This was done similarly to the above, but also included the Loschbour, Stuttgart and La Braña 1 samples. This left 881 AMH-derived DMRs (hereinafter skeletal AMH-derived DMRs), 838 archaic-derived DMRs, 494 Neanderthal-derived DMRs, and 442 Denisovan-derived DMRs.

The limited number of archaic human methylation maps introduces asymmetry in our ability to determine the level of fixation of DMRs along different lineages. Whereas we used dozens of AMH skeletal samples, we have just a few archaic samples. This provides us with the ability to better estimate the distribution of methylation values within each DMR in AMH, and thus to determine how significantly methylation values in other samples deviate from it. To enhance our ability to estimate variability within archaic human lineages, we added to the analysis the

reconstructed methylation map of the Vindija Neanderthal. The USER-treated portion of this genome (the portion amenable for methylation reconstruction) was sequenced to a depth of x7. Therefore, the methylation map that could be reconstructed from this individual has a considerable lower resolution compared to the other reconstructed maps used in this study (coverage 19x to 52x). Nevertheless, due to the reduced ability to detect variability along the archaic human linages, we employed this map for additional variability filtering along these lineages. DMRs where the Vindija Neanderthal clustered with the other hominins, and not with the Altai Neanderthal (or not with either of the archaic humans in the archaic-derived DMRs) were discarded. The number of DMRs mentioned throughout this chapter already includes this filtering.

A general concern is working with DNA methylation data is that DMRs that are specific to one group do not necessarily represent an evolutionary change, but rather reflect a characteristic such as tissue, sex or age that is shared by individuals in this group and not by others. In this regard, it is important to mention that such a scenario is unlikely in our study: the chimpanzee samples, as well as the modern human samples, include both males and females, juveniles and adults, and samples that come from limbs, skull, rib and teeth. Thus, it is unlikely that the DMRs that differentiate these groups reflect variability that stems from these parameters[57].

Comparison to previous reports

We have previously reported that compared to present-day humans, the HOXD cluster of genes is significantly hypermethylated in the Neanderthal and Denisovan[4]. Using the new methylation maps, we show that this observation holds (Extended Data Fig. 3). Adding chimpanzee data, we see that similarly to AMHs, chimpanzee samples are also hypomethylated compared to archaic humans. This suggests that the hypermethylation arose along the archaic-human lineage. However, we find that the Ust'-Ishim individual is an outlier among modern humans, and that his methylation levels are closer to the Neanderthal than to modern humans, as was also shown by Hanghøj et al[58]. The Neanderthal and Ust'-Ishim individuals are found >2 standard deviations from the mean observed methylation in modern humans. This suggests that although the Neanderthal is hypermethylated compared to most modern humans, she is not found completely outside modern human variation. The Denisovan, on the other hand, is found even further away,

and significantly outside the other populations. Given this, the HOXD DMR was classified as Denisovan-derived (Extended Data Table 2).

Compared to the previously reported DMRs[4], in this study we found four times more AMH- and archaic-derived DMRs (3,442 compared to 891) and roughly twice as many Neanderthal- and Denisovan-derived DMRs (502 and 597 compared to 295 and 307 in the Denisovan and Neanderthal, respectively). We also found that ~20% of the previous list was identified here too. There are several key factors that could underlie the differences in the reported DMRs: i. the current study used stricter thresholds for DMR detection, including a minimum of 50 CpGs in each DMR (compared to 10 CpGs previously), and a requirement for physical overlap in the three-way DMR detection procedure. ii. While in this study the AMH reference is a reconstructed ancient map, in the previous study the AMH reference, as well as the other tissues used for filtering out noise, were mainly cultured cell lines. iii. The previous study focused on DMRs that are invariable across tissues. In this study we focused on DMRs in skeletal tissues. In the previous study we were therefore able to extrapolate and find trends that extend beyond the skeletal system, such as neurological diseases. In this paper, we focus on the skeletal system, hence the different look of the body map (Figure 1b,c).

Computing correlation between methylation and expression

In order to identify regions where DNA methylation is tightly linked with expression levels, we scanned each DMR in overlapping windows of 25 CpGs (the window used for smoothing the deamination signal). In each window we computed the correlation between DNA methylation levels and expression levels across 21 tissues[59]. For each DMR, we picked the window with the best correlation (in absolute value) and computed regression FDR-adjusted *P*-value. DMRs that overlap windows with FDR < 0.05 were considered to be regions where methylation levels are significantly correlated with expression levels. 70 such DMRs were found among the skeletal AMH-derived DMRs, and 57 among the archaic human-derived DMRs, 22 among Neanderthal-derived DMRs, and 12 among Denisovan-derived DMRs.

Studying the function of DMGs

Similarly to sequence mutations, changes in regulation are likely to be unequally distributed across different body systems, owing to negative and positive selection, as well as inherent traits

of the genes affecting each organ. Thus, we turned to investigate which body parts are affected by the DMGs. To this end, we ran the lists of DMGs in Gene ORGANizer (geneorganizer.huji.ac.il, paper submitted), which is a tool that links genes to the organs they affect, through known disease and normal phenotypes. Thus, it allows to investigate directly the phenotypic function of genes, to identify their shared targets and to statistically test the significance of such enrichments. We ran the lists of DMGs in the ORGANize option using the default parameters (i.e., based on *confident* and *typical* gene-phenotype associations).

For the full bone AMH-derived DMGs, the analysis produced 40 significantly enriched body parts, the majority of which are skeletal. The vocal cords were the most enriched body part, with an enrichment of x1.49 (FDR = 0.039), followed by the nails, cheeks, larynx and lips (Extended Data Table 4). When we ran the list of bone AMH-derived DMGs, we found that the epiglottis and vocal cords are the most enriched body parts (x1.70 and x1.69, respectively), albeit the epiglottis was not significant (P = 0.054), and the vocal cords were not significant after FDR correction ($P = 8.8\text{x}10^{-3}$, FDR = 0.131). No other body parts were significantly enriched in this analysis. When we ran the list of skeletal AMH-derived DMRs, we found 14 significantly enriched body parts, with the vocal cords and the larynx being the most enriched parts (x2.18 and x1.74, FDR = 0.010 and FDR = 0.029, respectively). Most other parts belonged to the face (teeth, forehead, lips, eyelid, nose, maxilla, face, jaws, mandible), as well as the pelvis and spinal column (Extended Data Table 4). For archaic-derived DMGs, endocrine glands such as the thyroid and hypothalamus were consistently the most enriched across analyses. The skeleton was enriched as well, particularly the limbs and face (Extended Data Fig. 1, Extended Data Table 4). The Neanderthal-derived and Denisovan-derived DMG lists did not produce any significantly enriched organs.

In order to examine whether such trends could arise randomly from the reconstruction method, we repeated the analysis on the previously described 100 simulations. We ran all simulated DMGs (4,153) in Gene ORGANizer and found that no enrichment was detected, neither for voice-related organs (vocal cords: x0.99, FDR = 0.731, larynx: x.1.02, FDR = 0.966, epiglottis: x1.00, FDR = 0.966), nor for any other organ.

We next wanted to check the possibility that genes affecting the larynx and face tend to be longer than other genes, and are thus more likely to contain DMRs. We found that length of genes could

not be a factor explaining the enrichment within genes affecting the larynx, as these genes tend to be shorter than other genes in the genome (mean: 62.5 kb vs. 73.2 kb, $P = 0.001$, *t*-test). Genes affecting the face, on the other hand, tend to be longer than other genes (mean: 77.1 kb vs. 65.6 kb, $P = 4.6\text{x}10^{-5}$, *t*-test). To examine if this factor may lie behind the enrichment we observe, we repeated the analysis using only DMRs that are found within promoter regions (5 kb upstream to 1 kb downstream of TSS), thus eliminating the gene length factor. We found that the genes where such DMRs occur are still significantly associated with the face ($P = 0.036$, Fisher's exact test).

Gene ontology and expression analyses were conducted using Biological Process and UNIGENE expression tools in DAVID.

Computing the density of changes along the genome

To gain insight into the distribution of methylation changes, we computed the density of derived CpG positions along the genome in two ways. First, we used a 100 kb window centered in the middle of each DMR, and computed the fraction of CpGs in that window which are differentially methylated (i.e., are found within a DMR). Second, for the chromosome density plots, we did not center the window around each DMR, but rather used a non-overlapping sliding 100 kb window starting at position 1 and running the length of the chromosome. In order to avoid biases in genomic distribution of DMRs, in this analysis we used full bone AMH-derived DMRs, see "polarizing DMRs" and "Removing DMRs with high within-group variability" sections.

NFIX

Skeletal phenotypes that are associated with the Marshall-Smith syndrome were extracted from the Human Phenotype Ontology (HPO)[60]. Non-directional phenotypes (e.g., irregular dentition), and phenotypes that are expressed in both directions (e.g., tall stature and short stature) were removed.

Mutations in *NFIX* have also been linked to the Sotos syndrome. However, *NFIX* is not the only gene that was linked to this syndrome; mutations in *NSD1* were also shown to drive similar phenotypes[35]. Therefore, it is less relevant in assessing the functional consequences of general shifts in the activity levels of *NFIX*. Nevertheless, it is noteworthy that in the Sotos syndrome

too, most symptoms are a mirror image of the Neanderthal phenotype (e.g., prominent chin and high forehead).

Though various levels of speech could probably be achieved in other configurations, the 1:1 conformation was shown to be optimal for speech[14,15]. This ratio was attained through two processes: the flattening of the face and the descent of the larynx. Current methods could not determine with certainty the exact location of the Neanderthal larynx, but the flattening of the face was shown to have occurred on the modern human lineage, through a reduction in ECM deposition[41]. The fact that the same network of DMGs affect both facial flattening and the larynx, suggests an intriguing possibility that flattening of the face in AMHs was accompanied by a simultaneous descent of the larynx.