Harnessing epigenetics to study human evolution
Yoav Mathov1,2,3, Daniel Batyrev1,2,3, Eran Meshorer1,2 and Liran Carmel1

Recent advances in ancient DNA extraction and high-throughput sequencing technologies enabled the high-quality sequencing of archaic genomes, including the Neanderthal and the Denisovan. While comparisons with modern humans revealed both archaic-specific and human-specific sequence changes, in the absence of gene expression information, understanding the functional implications of such genetic variations remains a major challenge. To study gene regulation in archaic humans, epigenetic research comes to our aid. DNA methylation, which is highly correlated with transcription, can be directly measured in modern samples, as well as reconstructed in ancient samples. This puts DNA methylation as a natural basis for comparative epigenetics between modern humans, archaic humans and nonhuman primates.

Addresses
1 Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, 9190400, Israel
2 The Edmond and Lily Safra Center for Brain Sciences (ELSC), The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 9190400, Israel
3 These authors have contributed equally.

Current Opinion in Genetics & Development 2020, 62:23–29
This review comes from a themed issue on Genetics of human origin
Edited by Sarah Tishkoff and Joshua Akey

https://doi.org/10.1016/j.gde.2020.05.023
0959-437X/© 2020 Elsevier Ltd. All rights reserved.

Introduction
The emergence of high-quality genomes of archaic humans (Neanderthals and Denisovans) [1–3] opened up the opportunity to look for the genetic underpinning of traits that separate anatomically modern humans (AMHs) from archaic ones [4]. Many such traits are believed to result from changes in gene regulation [5,6], hence the identification of differentially regulated genes across human groups has become a major avenue of research. Direct access to RNA molecules in ancient specimens is typically unattainable due to the relatively short half time of RNA, although ancient RNA has been successfully retrieved from relatively young plant seeds [7] and from exceptionally preserved permafrost specimens, including wolf skins and a 14 300-year-old canid liver [8]. Regardless, RNA from archaic humans has yet to be sequenced, triggering the use of indirect methods to obtain information on their transcriptome.

One strategy is identifying DNA sequence changes in known regulatory elements. Following the high-coverage DNA from the Altai Neanderthal, almost 15 000 genetic variants that separate archaic from modern humans have been suggested to potentially affect regulatory elements [2]. Twenty-five of those were experimentally tested in neurons, and roughly half exerted expression changes [9]. Another approach relies on the fact that ~2% of the genomes of present-day non-African individuals is introgressed from Neanderthals, and that the introgressed regions vary across individuals. Consequently, each introgressed sequence is present in some individuals but absent in others, providing an opportunity to compare the regulatory differences of the Neanderthal and AMH variants. Using 450 individuals for which both genotype and expression data were measured in 48 tissues, introgressed sequences were reported to be enriched in regulatory variants but not in non-synonymous variants, suggesting that their function is mostly conveyed through their effect on gene regulation. These functions are potentially related to adipose tissues, body mass index (BMI), immunity and neurology [10]. Measuring allele-specific expression in 214 introgressed individuals across 52 tissues suggested cis-regulatory effects of roughly one-quarter of the introgressed regions, as well as downregulation in brain and testes [11]. Examination of archaic SNP frequencies in promoters, enhancers and micro-RNA binding sites across 127 tissues of introgressed individuals from the 1000 genome project, revealed that Neanderthal variants tend to affect enhancers much more than other elements. Particularly strong enrichment of Neanderthal variants was found in adipose-related tissues and primary T cells [12]. Using a statistical method to impute the cis-regulatory component of gene expression, more than 750 genes that are not present in introgressed sequences were reported to likely harbor different cis-regulation in Neanderthals and AMHs [13**].

Despite the many insights obtained by these analyses, they are restricted by our limited ability to predict how sequence changes affect gene regulation. Therefore, a second strategy to obtain indirect information on gene
regulation is to study epigenetic differences between the human groups. This review will focus on this strategy. We will explain how epigenetics can be studied in ancient genomes, and what new insights could be gained.

**Evolution of gene regulation through the lens of epigenetics**

Epigenetics is a term that describes a set of regulatory layers that can be modified without altering the DNA sequence and that result in heritable changes in gene expression. Well-known epigenetic marks include DNA methylation, histone modifications, nuclease positions and 3D DNA packaging. Different epigenetic marks are highly co-regulated, and bear ample information on gene expression patterns [14,15].

Two epigenetic marks are currently known to retain information in ancient DNA sequences. Nucleosomes tend to be depleted in active regulatory regions, hence their position bears information on such elements [16]. It has been shown that nucleosome positions along the DNA can be inferred based on patterns of DNA fragmentation. Breaks tend to occur between nucleosomes and not within them, possibly because nucleosomes survive long enough to shield their DNA from damage [17,18]. However, not all ancient samples are suitable for nucleosome fragmentation analysis, and difficulties in reconstructing nucleosome positions are expected in highly fragmented samples.

A richer source of information can be obtained from DNA methylation maps. CpG methylation, especially

---

**Box 1 Regulation of gene expression by DNA methylation.**

Methylation changes in gene promoters alter the binding affinity of different transcription factors, regulating their function.

(a) In the absence of promoter methylation, a transcriptional activator binds the unmethylated promoter, enabling its expression.

(b) Once the promoter is methylated, the transcriptional activator cannot bind the hypermethylated promoter, and the gene is repressed.

(c) and (d) Transcription factors also bind distal enhancers and can regulate gene expression by associating with promoters through three-dimensional (3D) organization of the genome. Topologically associated domains (TADs) are formed through genomic loops that are held together through the binding of the zinc finger protein CCCTC-binding factor (CTCF). CTCF binding to DNA is methylation sensitive and can only bind its target motif if unmethylated. Methylation of CTCF binding sites can locally disrupt the 3D structure of the genome and affect gene regulation inside specific TADs.

(c) When CTCF binds its unmethylated target motifs, it facilitates loop formation, bringing together a distal enhancer and a promoter. A transcriptional activator, which binds the distal enhancer, can thus activate gene expression.

(d) When CTCF motifs are methylated, CTCF does not bind its targets, leading to TAD disruption. Distal enhancer regions, even if unmethylated and bound by transcriptional activators, cannot activate their target genes due to physical separation.
in CpG-dense regions called CpG islands (CGIs), which are mostly found in promoters and enhancers, strongly modulate gene expression [19] (Box 1). Changes in DNA methylation levels are thought to affect gene expression by modulating the affinity of methylation-sensitive DNA binding proteins [20], thus preventing or promoting the action of transcriptional activators and repressors. DNA methylation also regulates the binding of chromatin structural proteins. Among the most prominent of them is CTCF [21–23], which, together with Cohesin and Condensin, shapes the 3D organization of the genome by promoting DNA loop formation and bringing distal regulatory elements to the vicinity of their target genes.

The remaining of this review will describe insights that have been obtained from studying the evolution of DNA methylation in humans, including its implications on the evolution of gene regulation across tissues and species. We will make a distinction between two time scales. First, we will examine comparisons between modern humans and other extant primates, which highlight regulatory changes that shaped the human lineage since we split from chimpanzees and bonobos. Second, we will show how DNA methylation was studied in ancient samples, and examine comparisons between modern and archaic humans, which highlight regulatory changes that discriminate between the different human groups.

Comparing DNA methylation between humans and other primates

Comparative studies of DNA methylation try to characterize how DNA methylation varies between and across primate species, and how differential methylation is related to differential gene expression. First attempts measured the DNA methylation of selected genes (e.g., [24,25]), but here we will mainly look at studies that measured genome-wide DNA methylation maps. Whereas the different studies vary in factors like the technique used to measure DNA methylation, the number of measured CpG sites, the compared species, and the tissues used, several general conclusions have been drawn.

Similar to what has been observed for RNA expression, genome-wide DNA methylation patterns cluster first by tissue and then by species [26–28]. Tissue-specific differentially methylated regions (DMRs) tend to be associated with developmental genes, and to be conserved between species. Genes associated with tissue-specific DMRs that are conserved across species are enriched with regulatory functions related to the same tissue, and depleted with primary metabolism functions [26,27*]. Combined, these results highlight the highly conserved role of DNA methylation in determining cell-type identity during development.

For the same tissue, distances between DNA methylation patterns across species replicate the known primate phylogeny [29,30]. Generally, the correlation between the DNA methylation patterns of different species in the same tissue is high, and just slightly less than the correlation between same-tissue patterns in different individuals from the same species. For example, the correlation between human and chimpanzee sperm DNA methylation is 0.86–0.88, whereas it is 0.89 and 0.91 within human and chimpanzee, respectively [31]. Roughly 9% of CpG position are differentially methylated between species [26,29], and they tend to reside in CpG shores and regions outside of CGIs [29,30], which is not surprising given that hypomethylation of CGIs tend to be conserved even between species as remote as humans and zebrafish [32]. Human-specific DMRs tend to be hypomethylated in blood [30] and brain [33,34*], and to be enriched within promoters and gene bodies of genes related to neurological and developmental functions, including the development of the nervous system, higher cognitive functions, and embryonic morphogenesis [29,34*,35,36]. Moreover, these genes are enriched with a range of diseases including cancer and neurological disorders such as autism and cognitive impairment [29,33,34*,37]. However, these trends might not be general, as a recent study suggested that more than 95% of human-chimpanzee differentially methylated regions in glia cells are hypermethylated in human [28]. Comparative analysis of DNA methylation in femora revealed that species-specific DMRs are enriched in several gene categories, including those involved in limb development and skeletal system development, potentially contributing to differential phenotypes [37]. Overall, these studies establish the links between differential DNA methylation, differential gene expression, and phenotypic variation. It is estimated that 11%–25% of human-chimpanzee differentially expressed genes are affected by changes in promoter methylation, which explain 12%–18% of the differences in expression levels [26,27*].

Comparing DNA methylation between modern and archaic humans

DNA methylation is a stable chemical modification, and was shown to be present in ancient DNA even after tens of thousands of years, suggesting the feasibility of accessing DNA methylation patterns in ancient specimens, and thus obtaining information on patterns of gene expression in these specimens.

Direct measurement of ancient DNA methylation levels was demonstrated to be impeded by technical obstacles that arise due to the damage — mostly fragmentation and deamination — inflicted on the ancient DNA molecules [38–40]. Therefore, DNA methylation was directly measured using bisulfite conversion in only a small number of ancient samples, including a 26 000 year-old bison [41], and 30 native Americans (230–4500 years BP) [42]. DNA
methylation can also be measured using methyl-binding-domains enrichment. This method was used to measure methylation patterns in Egyptian barley grains (800 BCE to 1812 CE) [43], and in extracts from soft and calcified tissues of several samples including woolly mammoths, polar bears, equids, and a human [39]. However, this method was shown to be biased towards large fragments and CpG-rich regions [39]. A key source of error when short DNA fragments are treated by bisulfite is the reduced alphabet of DNA, from ACGT to AGT, limiting the accuracy of their alignment to the reference genome. Recently, a modified bisulfite sequencing technique was developed to circumvent this limitation [40,44]. In this method, hairpin adapters are ligated to the ancient fragments before the bisulfite treatment, keeping both strands connected during sequencing, and thus allowing for a direct identification of C → T conversions and avoiding the need to reduce the DNA alphabet.

An alternative to a direct measurement of DNA methylation is a computational approach that harnesses the fact the deamination of ancient DNA works differently for methylated and unmethylated cytosines [17,45]. Unlike the direct approach, the computational technique provides information at a regional, rather than at a single base-pair, resolution. Using this approach, the methylomes of a paleo-Eskimo [17], a Denisovan, two Neanderthals and five other AMHs were reconstructed [45,46] (Figure 1). An initial list of genes whose methylation differs between modern and archaic humans, but is invariant across modern human tissues, found that they are enriched with disease-related genes, of which a third are associated with neurological and psychiatric disorders [45]. These genes were found to be enriched in GWAS SNPs that are related to schizophrenia, supporting the suggestion that schizophrenia is a by-product of modern human evolution [47]. Using a large collection of modern and ancient bone samples, as well as chimpanzee bones, nearly 3000 high-confident DMRs separating the different human groups were found, representing methylation differences of at least 50% that span at least 50 CpG positions. These DMRs are associated with 588 differentially methylated genes, where the methylation change occurred in modern humans after the split from archaic humans. These genes were shown to be enriched in genes affecting the face and voice box anatomy, suggesting that voice box in modern humans is anatomically different than that of Neanderthals and Denisovans [46]. These results become even more significant when the 3D genome organization was accounted for [48]. In the absence of HiC maps for human bone, topologically associated domains (TADs) were derived from consensus maps in three karyotypically normal human cell lines, obtaining ‘consensus-TADs’, which allowed to assign ~70 additional DMRs to their target genes [48] (Figure 2). The phenotypic

![Figure 1](image-url)

Ancient human samples where DNA methylation has been reconstructed. Sample age, sex and coverage are designated. The UDG-treated part of the Vindija Neanderthal genome has a coverage of only ×7, and consequently its DNA methylation map could not be reconstructed accurately. Rather, its DNA methylation map was used as a validation to DMRs detected using other samples [46].
information within the DMRs that separate the human groups was shown to be sufficient to reconstruct the anatomy of the elusive Denisovan with ~85% accuracy [49].

Conclusions

Evolutionary changes in gene regulation might underlie many human-specific adaptations, but their study during our very recent evolution is obstructed by the rarity of ancient transcriptomics. Epigenetic marks may serve as a proxy for gene activity patterns. As far as we know today, ancient sequences preserve two such marks, nucleosome position and DNA methylation. However, this is unlikely to be a major hurdle, as different epigenetic layers carry strongly overlapping information on gene activity [50].

Epigenetic marks are tissue-specific. Therefore, as of today, reconstruction of DNA methylation mainly provided information on the skeletal system. Whereas changes in this system are central to human evolution, there is much interest in understanding the evolution of other systems, such as the nervous system and the brain. A method to identify bone DMRs that likely extend to other tissues has been offered [38], but its validation awaits future research.

Computational reconstruction of epigenetic marks bears a great potential, yet is limited in different ways. Information on nucleosome positions seems to be present in many samples [17], yet to be absent in others, without a clear understanding of the factors that lead to this variation [51]. The reconstruction of DNA methylation requires high-coverage shotgun sequencing, estimated in at least ×15, which is rare among ancient samples. However, it is likely that pooling together of low-coverage shotgun samples may offer some remedy to this limitation. Moreover, the reconstruction of ancient DNA methylation also requires that the DNA libraries be prepared using UDG treatment [52], which is not always practiced in ancient DNA labs. Inferring the significance of a particular DMR on gene expression or on specific phenotypes is challenging, as is the case for inferring the significance of genetic variations. Usually, such inferences are carried out using enrichment analyses, using simultaneous measurements of expression and methylation, and using experiments on specific loci [46].

As of today, only three methylation maps of archaic humans have been reconstructed. As more high-coverage archaic samples become available, it will ultimately allow inter-population analysis in archaic humans. When conducted on present-day human populations, such studies demonstrated that DNA methylation mirrors genetics in inter-population pattern of divergence [53,54].

Experimental techniques to measure DNA methylation in ancient DNA are also limited. It would be interesting to see whether the recent development of methylation-sensitive long read sequencing tools that do not require PCR-amplification and bisulfite conversion, for example, Nanopore sequencing, may be useful to overcome some of the current limitations.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work has been supported by the Israel Science Foundation (grant 1009/17 to L.C. and 1140/17 to E.M.).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


14. Used DNA sequence analysis to identify cis-regulatory differences between modern human and Neanderthal genes that are not known to be introgressed and thus their relative expression cannot be estimated based on modern individuals.


The first work that reconstructed epigenetics marks in an ancient genome.


Simultaneous whole-genome measurement of DNA methylation and RNA expression in humans, chimpanzees and macaques in four different tissues (heart, liver, kidney and lung).


Whole-genome DNA methylation in the prefrontal cortices of humans, chimpanzees and macaques, revealing differentially methylated regions in the brain.


   The first reconstructed DNA methylation map in a Neandertal and a Denisovan.


