Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Archaic hominin genomics provides a window into gene expression evolution

Stephanie M Yan and Rajiv C McCoy

Differences in gene expression are thought to account for most phenotypic differences within and between species. Consequently, gene expression is a powerful lens through which to study divergence between modern humans and our closest evolutionary relatives, the Neanderthals and Denisovans. Such insights complement biological knowledge gleaned from the fossil record, while also revealing general features of the mode and tempo of regulatory evolution. Because of the degradation of ancient RNA, gene expression profiles of archaic hominins must be studied by indirect means. As such, conclusions drawn from these studies are often laden with assumptions about the genetic architecture of gene expression, the complexity of which is increasingly apparent. Despite these challenges, rapid technical and conceptual advances in the fields of ancient genomics, functional genomics, statistical genomics, and genome engineering are revolutionizing understanding of hominin gene expression evolution.

Address
Department of Biology, Johns Hopkins University, Baltimore, MD, 21218, USA

Corresponding author: McCoy, Rajiv C (rajiv.mccoy@jhu.edu)

Current Opinion in Genetics and Development 2020, **62**:44–49

This review comes from a themed issue on **Genetics of human origin**

Edited by **Joshua Akey** and **Sarah Tishkoff**

For a complete overview see the Issue and the Editorial

Available online 29th June 2020

https://doi.org/10.1016/j.gde.2020.05.014

Introduction

Over the past decade, high-coverage genome sequences from archaic hominins have provided a catalog of genetic changes that distinguish modern humans from our extinct evolutionary relatives, the Neanderthals and Denisovans [1,2,3]. These studies identified ~150 000 single nucleotide substitutions and short insertions or deletions that are fixed (or nearly fixed) in modern humans, but absent from Neanderthal, Denisovan, and non-human great ape genomes. Despite this knowledge, drawing causal links between genetic and phenotypic differences has proven challenging. Few substitutions lie in coding regions of the genome, and of those, even fewer are predicted to consequentially alter protein structure or function. Protein-coding changes are thus unlikely to account for the bulk of phenotypic divergence between modern humans and archaic hominins. Regulatory mutations that influence gene expression provide an appealing alternative, as such mutations are thought to drive most phenotypic differences within [4–6] and between [7,8] species.

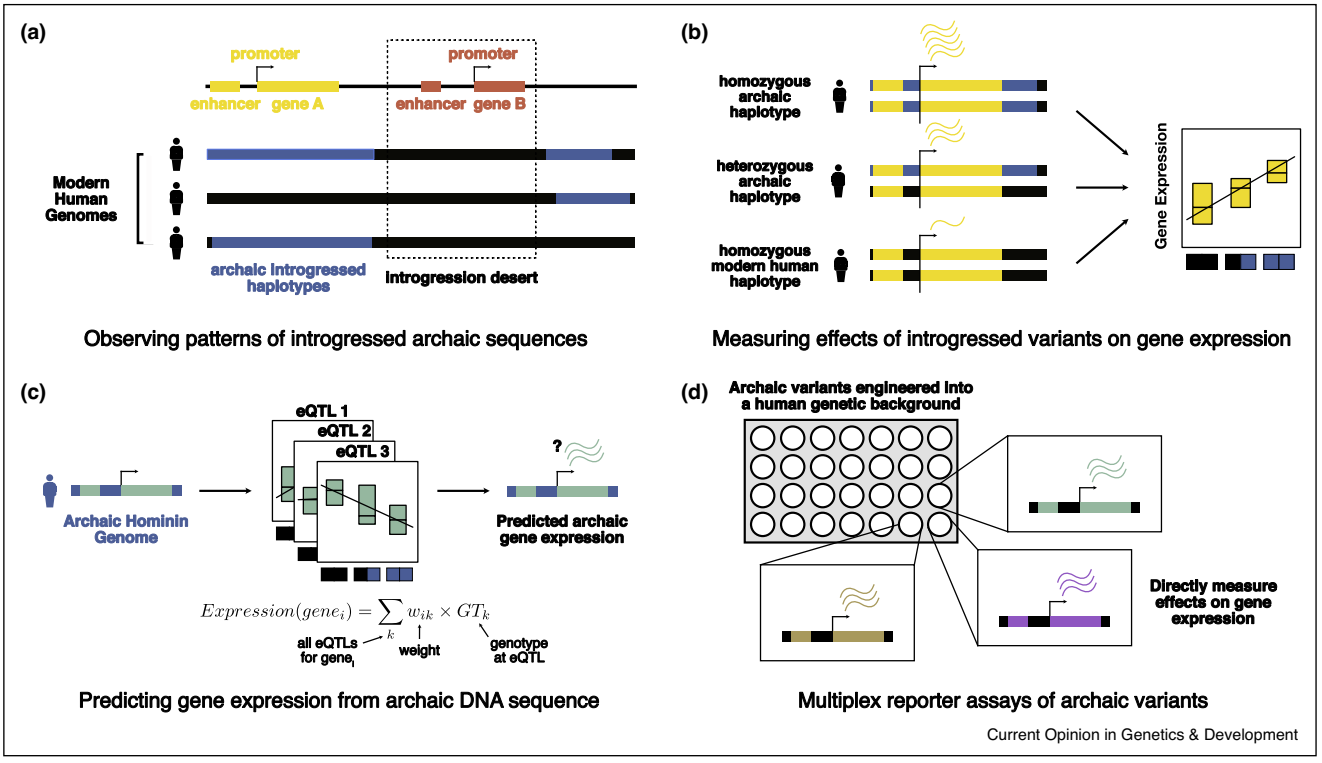
With few exceptions [9], the rapid degradation of ancient RNA prevents researchers from directly assaying differences in gene expression between modern and archaic samples. Several complementary indirect approaches have therefore been developed to leverage functional genomic data to infer these differences, as well as their underlying regulatory mechanisms. Here, we discuss recent approaches that use archaic introgression, genome engineering, and/or statistical methods to study gene expression divergence in the hominin lineage (Figure 1). We examine their unique strengths and limitations in turn.

Archaic introgression in regulatory regions

A variety of statistical methods have been developed to discover Neanderthal and Denisovan haplotypes that persist in modern human genomes [11–15]. Application of these methods to genomic data from large human population reference panels allowed researchers to map the genome-wide distribution of introgressed haplotypes across populations. These haplotypes are reduced or absent in some areas of the genome, termed 'deserts' of introgression, and enriched in others, suggesting that both negative and positive selection influenced the fate of archaic introgressed sequence during the generations following admixture [12,16,17]. Characterizing the heterogeneity of introgressed sequence with respect to annotated genes and regulatory elements provides insight into the functional basis of phenotypic divergence, as well as the evolutionary forces that have shaped the landscape of persisting introgressed haplotypes.

Such studies have revealed that archaic introgressed sequences are depleted in regions under strong selective constraint, including coding and non-coding sequences that are conserved among primates [16,18•]. Fine-grained examination of regulatory elements further demonstrated that archaic introgression is reduced in promoters, miRNAs, and certain classes of enhancers [18•,19•,20•]. Specifically, Telis *et al.* found that enhancers harbor fewer introgressed variants on average, despite having a higher

Figure 1



Approaches for studying archaic gene expression. Modern humans diverged from Neanderthals and Denisovans approximately 500 kya, with subsequent admixture among these lineages <100 kya [10]. Differences in gene expression between modern and archaic hominins likely contributed to their phenotypic differences, as well as the phenotypic impacts of introgressed sequences. Because of the degradation of ancient RNA, assaying the expression of archaic alleles requires indirect methods. **(a)** Heterogeneity in the genomic landscape of introgressed haplotypes in modern human populations serves as a signature of positive, negative, and balancing selection targeting genes and regulatory elements following introgression. **(b)** Associations between introgressed haplotypes and modern human gene expression can be tested directly in contemporary samples. **(c)** Statistical methods can be used to predict archaic gene expression by training models with features such as genotype-expression associations, DNA sequence features, and methylation status. **(d)** Massively parallel reporter assays allow thousands of archaic (and non-archaic) alleles to be introduced episomally or by genome editing into human cell lines or model organisms, then tested simultaneously for effects on gene expression.

tolerance for new mutations than other genomic regions [19•]. In apparent contradiction, Silvert *et al.* found that enhancers are enriched for common archaic introgressed alleles, particularly in adipose tissue, where some enhancers also exhibit signatures of adaptive introgression [20•]. These conflicting results highlight statistical challenges in determining whether particular genomic elements are enriched or depleted of introgression. Interpreting such enrichment tests is complicated by disparate methods of discovering introgressed alleles, defining control sets of variation, annotating regulatory elements, correcting for multiple hypothesis testing, and controlling for confounding factors such as linkage disequilibrium (LD), allele frequency, and expression level. As knowledge of genome function continues to grow, so does the urgency for studies focused on identifying and mitigating relevant biases in enrichment tests as applied to evolutionary processes such as introgression.

One interpretation of the broadly reduced archaic introgression in regulatory regions is that it reflects differences in the efficacy of purifying selection in small Neanderthal and Denisovan populations. Once introgressed into larger modern human populations, these weakly deleterious mutations would be exposed to negative selection [21,22]. Indeed, introgressed Neanderthal ancestry is lowest in regions of low recombination, consistent with historical purifying selection against weakly deleterious introgressed alleles [23]. An alternate explanation for negative selection against Neanderthal alleles invokes regulatory epistasis — the concept that Neanderthal genes and regulatory elements are adapted to function within the Neanderthal genetic background. Rapid evolution of such regulatory incompatibilities has been observed in other systems and is one proposed mechanism contributing to hybrid incompatibility and eventual speciation [24–26]. In modern humans, genes that are

highly expressed in the testes are depleted of archaic ancestry, potentially consistent with male hybrid infertility [16]. In addition, a large proportion of human accelerated regions (HARs) — sequences that are conserved in mammals but divergent in humans — are thought to function as enhancers, suggesting that the corresponding regulatory networks may have evolved rapidly within the hominin lineage [27,28]. While providing circumstantial support for regulatory epistasis, direct tests are hampered by the statistical intractability of pairwise association, as well as the lack of reciprocal data measuring the expression of modern human alleles within archaic genetic backgrounds.

Effects of introgressed variants on gene expression

Introgressed archaic haplotypes allow researchers to assay associations with gene expression, providing insight into the characteristics of gene expression divergence throughout hominin evolution. Like their non-introgressed counterparts, many introgressed variants are expression quantitative trait loci (eQTLs), significantly associated with the expression of nearby genes in *cis*. Quach *et al.* [29], Dannemann *et al.* [30••], and Natri *et al.* [31] identified introgressed eQTLs of genes involved in immune functions. These included multiple eQTLs exhibiting strong allele frequency differentiation in specific populations — a signature of historical positive selection following introgression. McCoy *et al.* developed a complementary approach to examine allele-specific expression (ASE) of Neanderthal haplotypes in heterozygous samples, finding that approximately one quarter of all loci tested exhibited significant ASE [32••]. The authors also discovered anecdotal examples of ASE mediated by changes in patterns of alternative splicing (i.e. isoform-specific gene expression). Such introgressed splicing QTLs (sQTLs) have been characterized in further detail with respect to their impacts on immune response [33], including several examples associated with signatures of adaptive introgression [34,35]. Beyond the context of immune function, the impacts of introgressed sQTLs and their implications for hominin splicing divergence merit further investigation.

Genetic effects on expression are ubiquitous, and a large proportion of modern human variants constitute eQTLs [36]. One question of interest is whether introgressed archaic variants as a group are unique in their functional impacts, due to their unique evolutionary histories. Given the aforementioned hypotheses about genetic load accumulated within small Neanderthal populations, one may predict that Neanderthal-introgressed alleles would be enriched for effects on modern human gene expression. While some studies have confirmed such an enrichment [20•,30••,33] others have not [32••]. As with assessing the enrichment or depletion of introgressed sequences in regulatory regions, these discrepancies likely arise from methodological discrepancies in defining case and control groups and accounting for relevant covariates.

Perhaps more insidiously, because of the nature of LD in human populations, identifying an archaic-introgressed eQTL is not sufficient for establishing its causality. The archaic variant may simply tag one or more causal variants (introgressed or non-introgressed) through LD. Fine-mapping approaches can be used to assess causality [4,37], but have not been widely applied to studies of archaic gene expression. These considerations are especially important in light of recent findings that Neanderthal sequences in modern humans make limited contributions to variation in phenotypes, including gene expression. Rare and singleton variants explain the majority of gene expression heritability in humans, but with a negligible contribution from Neanderthal-introgressed singletons [38]. Many phenotypic associations previously attributed to archaic alleles are better explained by linked non-archaic variation [39•]. Further nuancing this picture, introgressed haplotypes may indirectly contribute to phenotypic variation by reintroducing ancestral alleles that were previously lost in the human lineage [40]. Together, these results suggest that future research into the gene expression impacts of archaic introgression must carefully consider the potential effects of linked alleles with distinct evolutionary histories.

Predicting gene expression from archaic DNA sequence

Even when scaling to large samples, less than half of the Neanderthal and Denisovan genomes are represented by introgressed fragments that persist within modern humans today, thus limiting insights about regulatory divergence that can be deduced from introgressed haplotypes [12]. Recent methodological advances in predicting gene expression from DNA sequences have enabled the study of archaic expression beyond the landscape of introgression.

Colbran *et al.* used the method PrediXcan [41], which imputes the *cis*-regulatory component of gene expression from genotype data, to predict the expression of archaic DNA sequences. They found that Neanderthal and modern human gene expression diverged substantially in genes without introgression, suggesting that this divergence may have served as a barrier to introgression [42••]. Similarly, by inferring DNA methylation from archaic hominin genome sequencing data (obtained from bones) [43], Gokhman *et al.* [44] and Gokhman *et al.* [45] identified differentially methylated regions between modern humans and archaic hominins. Upon extrapolating to downstream effects on expression and organismal phenotypes, the authors recapitulated several known features of archaic hominin anatomy, including a wide anterior mandible in Denisovans, and a low forehead and wide pelvis in both Neanderthals and Denisovans [44,45].

While highly innovative, the indirect nature of these studies poses a challenge for validation. One key concern regards generalizability to divergent hominin populations, as statistical models are generally trained on predictor (genotype or methylation) and response (gene expression or other phenotypes) data from contemporary human populations. Underscoring this concern, the accuracy of polygenic risk scores inferred from European populations decreases by more than 50% in non-European populations [46]. Maintaining strict criteria to filter out false positives is one possible solution to this issue. For example, Colbran *et al.* showed that the polygenic prediction method PrediXcan tends to underestimate divergence in more distant populations, leading to more false negatives than false positives when identifying divergent archaic loci [42••]. The accuracy of these 'imputation' methods can also be evaluated by benchmarking in extant species, such as non-human great apes [45].

Ongoing studies are clarifying the complex genetic architecture of gene expression, including widespread allelic heterogeneity (i.e. multiple causal variants) and abundant but weak *trans* effects. Future statistical models of gene expression evolution will be tasked with the challenge of incorporating these features while maintaining model tractability.

Multiplex assays of archaic gene expression

Recent *in vitro* methods offer promising approaches for disentangling LD and testing the causal effects of individual archaic hominin substitutions on gene expression. For example, in unpublished work, scientists reportedly used CRISPR to introduce the Neanderthal allele of the splicing factor *NOVA1* in a brain organoid cell line, characterizing downstream effects on global gene expression and cell morphology [47]. Reporter assays have also been used to demonstrate that some ancestral hominin alleles reintroduced into humans through Neanderthal introgression have strong effects on gene expression, even when tested independently of other alleles in perfect LD [40].

Beyond such candidate mutations, high-throughput methods such as massively parallel reporter assays (MPRAs) provide hypothesis-free approaches to simultaneously quantify the expression impacts of hundreds of thousands of loci genome wide [48,49,50]. In these assays, individual variants and their surrounding DNA sequence are genetically engineered into reporter constructs, allowing the gene expression effects of each variant to be measured experimentally. MPRAs were recently used to identify HARs that act as enhancers in neuronal cell lines, and whose expression differs significantly between human and chimpanzee sequences [28]. Further application of these methods to assess the effects of fixed archaic hominin substitutions will provide important insight into the role of regulatory changes in hominin evolution.

By necessity, most MPRA studies are performed on human cell lines, and the tested variants are often introduced episomally, outside of their endogenous loci. As a result, these studies cannot account for epistatic interactions, differences in nuclear architecture, or epigenetic modifications. CRISPR-based genome editing, which incorporates mutations into their native genomic contexts, will allow for more comprehensive assays of variant effects [51].

Conclusions

Changes in gene expression are a key driver of phenotypic innovation. Studies of archaic hominin gene expression thus provide a window into the genetic basis of phenotypic divergence in the hominin lineage, as well as the evolution of modern human-specific phenotypes. Hominin evolution may also serve as a useful model for investigating broader questions about regulatory evolution, given the wealth of available data. Such data can be leveraged to address fundamental questions such as the role of positive and purifying selection in shaping gene expression evolution; the mode and tempo of regulatory divergence among genes expressed in different cells, tissues, and developmental contexts; and the relative contributions of protein-coding and regulatory changes to phenotypic diversity and divergence across evolutionary timescales. The growth of ancient DNA, functional genomic, and phenotypic datasets, as well as the development of novel assays and analytical approaches, will fuel further understanding of gene expression evolution over the coming decades.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work is supported by N.I.H. grant R35GM133747 to R.C.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

1. Prüfer K *et al.*: **The complete genome sequence of a Neanderthal from the Altai Mountains**. *Nature* 2014, **505**:43-49.
2. Prüfer K *et al.*: **A high-coverage Neandertal genome from Vindija Cave in Croatia**. *Science* 2017, **358**:655-658.
3. Meyer M *et al.*: **A high-coverage genome sequence from an archaic Denisovan individual**. *Science* 2012, **338**:222-226.
4. Hormozdiari F *et al.*: **Colocalization of GWAS and eQTL signals detects target genes**. *Am J Hum Genet* 2016, **99**:1245-1260.
5. Maurano MT *et al.*: **Systematic Localization of common disease-associated variation in regulatory DNA**. *Science* 2012, **337**:1190-1195.
6. Albert FW, Kruglyak L: **The role of regulatory variation in complex traits and disease**. *Nat Rev Genet* 2015, **16**:197-212.
7. King M, Wilson A: **Evolution at two levels in humans and chimpanzees**. *Science* 1975, **188**:107-116.

8. Wittkopp PJ, Kalay G: **Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence**. *Nat Rev Genet* 2012, **13**:59-69.

9. Smith O, Dunshea G, Sinding M-HS, Fedorov S, Germonpre M: **Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival**. *PLoS Biol* 2019, **28**:e3000166.

10. Nielsen R *et al.*: **Tracing the peopling of the world through genomics**. *Nature* 2017, **541**:302-310.

11. Plagnol V, Wall JD: **Possible ancestral structure in human populations**. *PLoS Genet* 2006, **2**:e105.

12. Vernot B, Akey JM: **Resurrecting surviving Neandertal lineages from modern human genomes**. *Science* 2014, **343**:1017-1021.

13. Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM: **Analysis of human sequence data reveals two pulses of archaic Denisovan admixture**. *Cell* 2018, **173**:53-61.e9.

14. Sankararaman S *et al.*: **The genomic landscape of Neanderthal ancestry in present-day humans**. *Nature* 2014, **507**:354-357.

15. Chen L, Wolf AB, Fu W, Li L, Akey JM: **Identifying and interpreting apparent Neanderthal ancestry in African individuals**. *Cell* 2020, **180**:677-687.

16. Sankararaman S, Mallick S, Patterson N, Reich D: **The combined landscape of Denisovan and Neanderthal ancestry in present-day humans**. *Curr Biol* 2016, **26**:1241-1247.

17. Vernot B *et al.*: **Excavating Neandertal and Denisovan DNA from the genomes of melanesian individuals**. *Science* 2016, **352**:235-239.

18. • Petr M, Pääbo S, Kelso J, Vernot B: **Limits of long-term selection against Neandertal introgression**. *Proc Natl Acad Sci U S A* 2019, **116**:1639-1644

Petr *et al.* used simulations of human demographic history to determine that gene flow between modern human populations leads to the appearance of a constant decline in Neanderthal ancestry in humans over time, which was previously interpreted as ongoing negative selection. Instead, they showed that the majority of deleterious Neanderthal alleles were removed quickly from the human population. They also found that Neanderthal ancestry was significantly reduced in promoters and conserved coding and noncoding regions of the genome.

19. • Telis N, Aguilar R, Harris K: **Selection against archaic DNA in human regulatory regions**. *bioRxiv* 2019 http://dx.doi.org/10.1101/708230

Telis *et al.* found that archaic alleles are less likely to reside in enhancers than non-archaic alleles are, after controlling for allele frequency and level of background selection. Enhancers' tolerance for new mutations was not associated with strength of selection against archaic introgression, suggesting that negative selection against introgression in enhancers occurred as a result of regulatory incompatibilities rather than increased functional restraint.

20. • Silvert M, Quintana-Murci L, Rotival M: **Impact and evolutionary determinants of neanderthal introgression on transcriptional and post-transcriptional regulation**. *Am J Hum Genet* 2019, **104**:1241-1250

Silvert *et al.* sought to identify introgressed Neanderthal alleles in miRNA binding sites, promoters, and enhancers. Enhancers in several tissues, including immune cells, were enriched for Neanderthal introgression, and enhancers in adipose tissue exhibited signatures of adaptive introgression.

21. Harris K, Nielsen R: **The genetic cost of Neanderthal introgression**. *Genetics* 2016, **203**:881-891.

22. Juric I, Aeschbacher S, Coop G: **The strength of selection against Neanderthal introgression**. *PLoS Genet* 2016, **12** e1006340.

23. Schumer M *et al.*: **Natural selection interacts with recombination to shape the evolution of hybrid genomes**. *Science* 2018, **360**:656-660.

24. Mack KL, Nachman MW: **Gene regulation and speciation**. *Trends Genet* 2017, **33**:68-80.

25. Landry CR *et al.*: **Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila**. *Genetics* 2005, **171**:1813-1822.

26. Zhang M *et al.*: **Effects of parental genetic divergence on gene expression patterns in interspecific hybrids of Camellia**. *BMC Genomics* 2019, **20**:828.

27. Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, Pollard KS: **Many human accelerated regions are developmental enhancers**. *Philos Trans R Soc B Biol Sci* 2013, **368** 20130025.

28. Ryu H *et al.*: **Massively parallel dissection of human accelerated regions in human and chimpanzee neural progenitors**. *bioRxiv* 2018 http://dx.doi.org/10.1101/256313.

29. Quach H *et al.*: **Genetic adaptation and Neandertal admixture shaped the immune system of human populations**. *Cell* 2016, **167**:643-656.e17.

30. •• Dannemann M, Prüfer K, Kelso J: **Functional implications of Neandertal introgression in modern humans**. *Genome Biol.* 2017, **18**:61

Dannemann *et al.* identified Neanderthal-introgressed eQTLs by associating genotypes of introgression 'tag SNPs' with the expression of nearby genes, using data from the GTEx project. Archaic-introgressed SNPs were more likely than modern human SNPs to be associated with variation in gene expression. The authors also characterized changes in archaic allele frequency in humans as a proxy for selection, and found that most archaic alleles have decreased in frequency over time.

31. Natri H *et al.*: **Genome-wide DNA methylation and gene expression patterns reflect genetic ancestry and environmental differences across the Indonesian archipelago**. *PLoS Genet* 2020, **16**:e1008749.

32. •• McCoy RC, Wakefield J, Akey JM: **Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression**. *Cell* 2017, **168**:916-927.e12

McCoy *et al.* developed a statistical approach to quantify allele-specific expression of Neanderthal-introgressed sequences in the GTEx dataset. One quarter of all loci tested revealed significant allele-specific expression, but introgressed archaic variants were not more likely than frequency-matched modern human variants to affect gene expression. Neanderthal alleles of brain and testes-expressed genes were significantly downregulated compared to genes expressed in other tissues.

33. Rotival M, Quach H, Quintana-Murci L: **Defining the genetic and evolutionary architecture of alternative splicing in response to infection**. *Nat Commun* 2019, **10**:1671.

34. Sams AJ *et al.*: **Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans**. *Genome Biol* 2016, **17**:246.

35. Gittelman RM *et al.*: **Archaic hominin admixture facilitated adaptation to out-of-Africa environments**. *Curr Biol* 2016, **26**:3375-3382.

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

37. Giambartolomei C *et al.*: **A Bayesian framework for multiple trait colocalization from summary association statistics**. *Bioinformatics* 2018, **34**:2538-2545.

38. Hernandez RD *et al.*: **Ultrarare variants drive substantial cis heritability of human gene expression**. *Nat Genet* 2019, **51**:1349-1355.

39. • Skov L *et al.*: **The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes**. *Nature* 2020, **582**:78-83

Skov *et al.* identified archaic-introgressed haplotypes in the whole-genome sequences of 27 566 Icelandic individuals. Upon accounting for potential effects of linked non-archaic variants, they discovered only five introgressed alleles that were confidently associated with phenotypes in this Icelandic database. The authors concluded that archaic-introgressed alleles contribute only modestly to contemporary human phenotypic diversity, partially by consequence of purifying selection against impactful archaic haplotypes.

40. Rinker DC *et al.*: **Neanderthal introgression reintroduced functional alleles lost in the human out of Africa bottleneck**. *bioRxiv* 2019 http://dx.doi.org/10.1101/533257.

41. GTEx Consortium *et al.*: **A gene-based association method for mapping traits using reference transcriptome data**. *Nat Genet* 2015, **47**:1091-1098.

42. •• Colbran LL *et al.*: **Inferred divergent gene regulation in archaic hominins reveals potential phenotypic differences**. *Nat Ecol Evol* 2019, **3**:1598-1606

Colbran *et al.* used the machine learning method PrediXcan, which predicts the genetic component of gene expression from genotype data, to impute archaic hominin gene expression from their ancient DNA sequences. They discovered hundreds of Neanderthal and Denisovan genes whose expression was predicted to fall outside of the distribution of modern human gene expression. By intersecting this list of 'divergently regulated' genes with phenotype associations, the authors generated hypotheses about phenotypes disproportionately affected by regulatory divergence among hominin lineages.

43. Gokhman D *et al.*: **Reconstructing the DNA methylation maps of the Neandertal and the Denisovan**. *Science* 2014, **344**:523-527.

44. Gokhman D *et al.*: **Genes affecting vocal and facial anatomy went through extensive regulatory divergence in modern humans**. *bioRxiv* 2017 http://dx.doi.org/10.1101/106955.

45. Gokhman D *et al.*: **Reconstructing Denisovan anatomy using DNA methylation maps**. *Cell* 2019, **179**:180-192.e10.

46. Martin AR *et al.*: **Human demographic history impacts genetic risk prediction across diverse populations**. *Am J Hum Genet* 2017, **100**:635-649.

47. Cohen Jon: **Exclusive: Neanderthal 'minibrains' grown in dish**. *Science* 2018 http://dx.doi.org/10.1126/science.aau5384.

48. Tewhey R *et al.*: **Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay**. *Cell* 2016, **165**:1519-1529.

49. Inoue F, Ahituv N: **Decoding enhancers using massively parallel reporter assays**. *Genomics* 2015, **106**:159-164.

50. van Arensbergen J *et al.*: **High-throughput identification of human SNPs affecting regulatory element activity**. *Nat Genet* 2019, **51**:1160-1169.

51. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J: **Saturation editing of genomic regions by multiplex homology-directed repair**. *Nature* 2014, **513**:120-123.