RESEARCH ARTICLE

A signature of Neanderthal introgression on molecular mechanisms of environmental responses

Anthony S. Findley[1], Xinjun Zhang[2], Carly Boye[1], Yen Lung Lin[3], Cynthia A. Kalita[1], Luis Barreiro[3], Kirk E. Lohmueller[2,4], Roger Pique-Regi[1,5], Francesca Luca[1,5]*

1 Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan, United States of America, **2** Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, California, United States of America, **3** Genetics Section, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America, **4** Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, California, United States of America, **5** Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan, United States of America

* fluca@wayne.edu

Abstract

Ancient human migrations led to the settlement of population groups in varied environmental contexts worldwide. The extent to which adaptation to local environments has shaped human genetic diversity is a longstanding question in human evolution. Recent studies have suggested that introgression of archaic alleles in the genome of modern humans may have contributed to adaptation to environmental pressures such as pathogen exposure. Functional genomic studies have demonstrated that variation in gene expression across individuals and in response to environmental perturbations is a main mechanism underlying complex trait variation. We considered gene expression response to *in vitro* treatments as a molecular phenotype to identify genes and regulatory variants that may have played an important role in adaptations to local environments. We investigated if Neanderthal introgression in the human genome may contribute to the transcriptional response to environmental perturbations. To this end we used eQTLs for genes differentially expressed in a panel of 52 cellular environments, resulting from 5 cell types and 26 treatments, including hormones, vitamins, drugs, and environmental contaminants. We found that SNPs with introgressed Neanderthal alleles (N-SNPs) disrupt binding of transcription factors important for environmental responses, including ionizing radiation and hypoxia, and for glucose metabolism. We identified an enrichment for N-SNPs among eQTLs for genes differentially expressed in response to 8 treatments, including glucocorticoids, caffeine, and vitamin D. Using Massively Parallel Reporter Assays (MPRA) data, we validated the regulatory function of 21 introgressed Neanderthal variants in the human genome, corresponding to 8 eQTLs regulating 15 genes that respond to environmental perturbations. These findings expand the set of environments where archaic introgression may have contributed to adaptations to local environments in modern humans and provide experimental validation for the regulatory function of introgressed variants.

Author summary

Humans have populated the entire world thus adapting to live in very different environments. Recent studies have suggested that the presence of Neanderthal DNA sequences in the genomes of modern humans contribute to our ability to respond to pathogens. Here we investigated whether the Neanderthal sequences present in modern human genomes also contribute to our ability to respond to environmental changes. We found that DNA sequences from Neanderthals modify the molecular mechanisms that regulate gene activity in different environments, including in response to stress hormones, caffeine, and vitamin D. We also found an important role of Neanderthal sequences in regulating sugar metabolism. Using experimental data, we provide evidence that Neanderthal sequences modify the activity of several genes in our genomes, including genes important for our ability to respond to a broad set of environmental stimuli.

Introduction

Studies of ancient DNA samples in recent years have discovered that interbreeding occurred between modern humans and our extinct relatives [1,2,3]. Approximately 2% of the genome in modern Europeans and Asians is the result of introgressed Neanderthal sequences [4]. However, Neanderthal ancestry is estimated to be 1.54x the genome average in genomic regions with the lowest density of functional elements. The uneven distribution of these introgressed sequences in the human genome suggests that natural selection may have acted to remove Neanderthal sequences that were deleterious for modern humans [5]. The presence of introgressed archaic sequences in the human genome, instead, raises the question of whether these sequences were maintained in the human genome because they have functional or even beneficial impact in modern humans [6–9]. Accordingly, Neanderthal introgressed haplotypes have been found to occur at higher frequency than expected by drift and to impact genes with important biological functions for human adaptations in different environments [10,11]. These include for example BNC2, a gene associated with skin pigmentation [5,12] and several genes with important immunological functions [13].

Gene regulatory variants in the human genome are associated with complex traits and disease risk, supporting the concept that these variants have important phenotypic consequences [14,15]. Identifying functional non-coding variants that regulate transcriptional processes poses severe challenges because we cannot directly predict function from sequence. Several approaches have been proposed to computationally predict the effect of a nucleotide change on transcription factor binding and gene expression, including methods that are based on sequence content only [16–18] and methods that integrate sequence content information with functional genomic data, like DNase-seq [19–21].

Recent studies have focused on understanding the role of Neanderthal introgression on gene regulation in humans using genetic variants which are associated with gene expression (expression quantitative trait loci, or eQTL) from the multi tissue GTEx consortium [22]. Two independent studies used eQTL mapping and allele-specific expression analysis, respectively, to show that a substantial number of Neanderthal introgressed haplotypes harbor regulatory variants in the human genome [23,24]. Introgressed haplotypes were found to have lower expression compared to modern haplotypes in most tissues [24]. In particular, brain and testis had lower expression of introgressed haplotypes, compared to other tissues. Given the human-specific and reproductive functions of these tissues, these previous observations support the

hypothesis that Neanderthal introgression may have affected modern human fitness through gene regulation. These observations suggest that it is possible to analyze functional patterns of Neanderthal introgressed sequences to identify potential selection effects on overall pathways or molecular mechanisms.

Studies of genetic regulation of gene expression in cells exposed to pathogens confirmed a role for Neanderthal introgression in immune functions important for the response to bacterial and especially viral infections [25,26]. Adaptation to pathogen exposure is part of a broader set of signals detected in the human genome and interpreted as adaptations to local environments in human populations. In addition to pathogen exposures, human populations adapted to climate, diet and high altitude [27]. One of the limitations in studies of adaptations to local environments is the difficulty to collect and utilize data on past environmental exposures.

More generally, dissecting the complex environment that we are exposed to today and in the past is a difficult task that requires some necessary simplifications. For example, population genetics studies have summarized past environments with latitude or historical climate and subsistence data [28,29]; while epidemiological studies generally focus on current measurable lifestyle environments, such as smoking and drinking. An alternative approach to study genetic variants that modulate the response to environmental changes is through *in vitro* systems. Though a simplified version of organismal environments, cultured primary cells exposed to controlled treatments have been pivotal in identifying genetic variants that regulate the response to a variety of different exposures including drugs, hormones, pathogens and other chemical stimuli. We recently demonstrated that this approach can be scaled to investigate hundreds of cellular environments, defined as the combination of specific cell types exposed to different treatments [30]. Approximately 50% of the genes under genetic regulation of transcriptional response are also associated with complex traits variation in humans, thus confirming that the molecular phenotypes measured in this experimental setting (gene expression responses) are important mechanisms in human complex traits variation [30].

Here we used this approach to investigate the role of Neanderthal introgression in modern human response to environmental perturbations. We use a functional annotation of regulatory variants that disrupt transcription factor binding, to predict the regulatory function of Neanderthal introgressed alleles and provide an experimental validation of their regulatory function in human cells (Fig 1).

Results

Role of Neanderthal introgressed variants in transcription factor binding and gene regulation

The availability of large amounts of genetic data from human populations and of a variety of different functional annotations makes it possible to explore several unresolved questions on the role of Neanderthal introgression in human history. To analyze the contribution of Neanderthal introgression to modern human response to environmental perturbations, we first refined the definition of introgressed Neanderthal variants. Specifically, we considered variants in previously defined introgressed regions [2] and also present in 2 of 3 high-quality Neanderthal genomes [31–33], but absent in all African samples from the 1000 Genomes project (S1 Fig). We annotated 177,578 SNPs from the 1000 Genomes project [34] as Neanderthal introgressed variants (N-SNPs).

One of the main mechanisms used by human cells to respond to environmental perturbations is through transcriptional changes, which are largely the result of changes in transcription factor binding. To investigate the role of Neanderthal introgression in this cellular response, we focused on N-SNPs that are more likely to have a regulatory function.
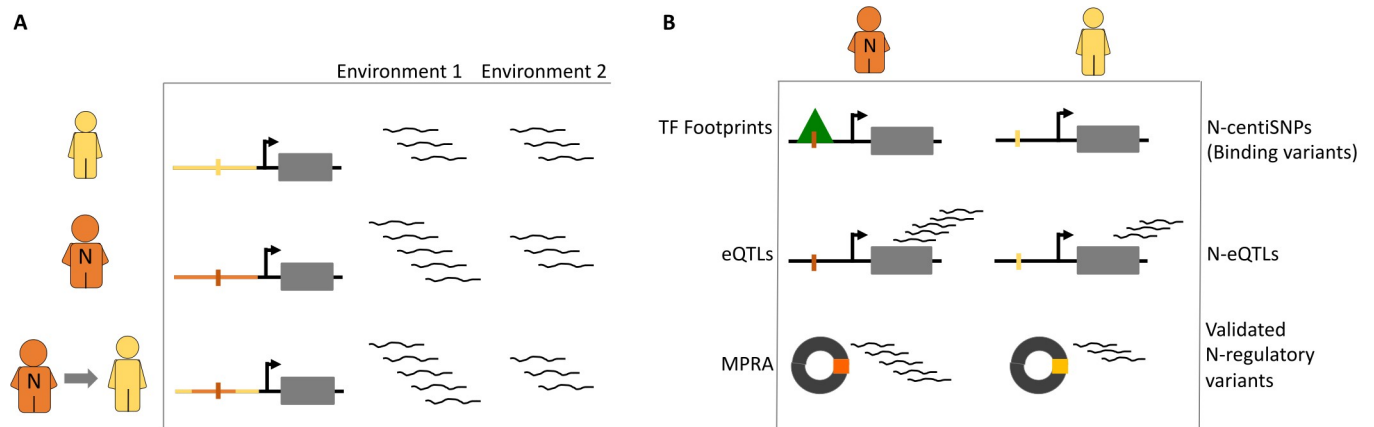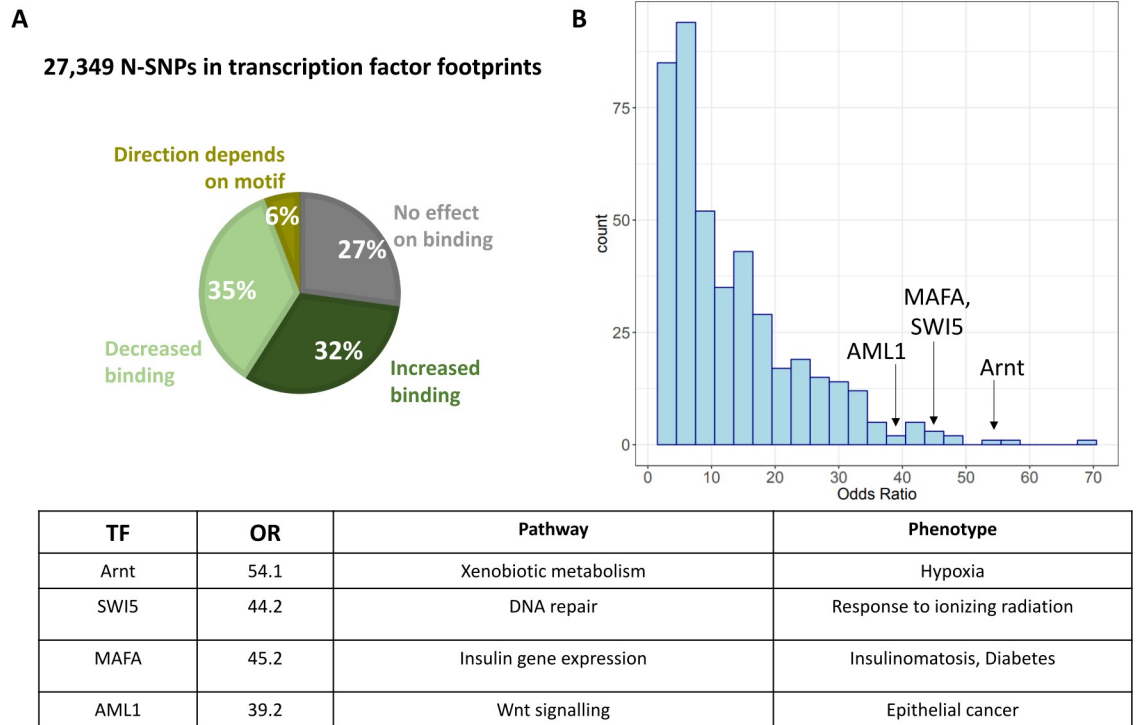
Fig 1. Hypothesis and study design (A) Conceptual framework. We hypothesized that the Neanderthal's transcriptional response to environmental perturbations was different than modern humans, and introgression of Neanderthal alleles modifies modern human's response to environmental perturbations. (B) The three approaches used in this study: analysis of Neanderthal introgression in transcription factor footprints, in eQTLs for differentially expressed genes in response to environmental perturbations, and MPRA to experimentally compare the regulatory function of introgressed alleles to modern alleles.

Specifically, we considered N-SNPs that are in active footprints for transcription factors. To annotate these N-SNPs, we considered the catalog of 1000 Genomes SNPs in active transcription factor footprints generated by [21] using the DNase-seq data from ENCODE and the RoadMap Epigenomics. This catalog annotates 5.8 million SNPs in active transcription factor footprints across 153 tissues and 1,372 motifs. We identified 58,967 N-SNP-motif pairs, corresponding to 27,349 unique N-SNPs in active footprints for a total of 1,255 unique transcription factor motifs. For each transcription factor, we investigated the predicted effect of N-SNPs on binding across all the binding sites in the genome. When considering all N-SNPs in transcription factor binding sites, we found that 19,924 N-SNPs (73%) were computationally predicted to alter binding (centiSNPs), which was significantly higher than the 66% of all SNPs which are centiSNPs ($p < 2.2$ x 10^{-16}). The Neanderthal allele is predicted to decrease binding in 48% of the cases, increase binding in 44% of cases, and both increase and decrease binding depending on the transcription factor in 8% of cases (Fig 2A). N-SNPs in footprints are more likely to disrupt transcription factor binding (centiSNPs) than non N-SNPs (positive enrichment), with a significant enrichment for 434 transcription factor motifs (BH-FDR < 0.05; range of enrichment odds ratios: [1.5, 67.7]) out of 435 that have at least 10 N-SNPs in footprints. Neanderthal introgressed variants may be absent in binding sites for specific transcription factors that are important for modern human-specific functions. To investigate whether any transcription factor is depleted of N-SNPs in footprints, we considered all TFs with at least 100 SNPs in footprints, regardless of the number of N-SNPs. None of the factors considered showed a significant depletion of N-SNPs in footprints.

Several motifs for transcription factors activated in response to various environmental perturbations were among the most enriched in N-centiSNPs genome-wide. ARNT, which is the most enriched motif, is involved in the metabolism of chemical substances that are not produced by the organism (xenobiotic metabolism) and is also a cofactor for hypoxia-inducible factor 1 and NFAT3 (NFATC4) belongs to a family of factors important for T cell activation [35]. SWI5 is a DNA repair protein involved in cellular response to ionizing radiation. Finally, AML1, also known as RUNX1 is involved in hair follicle development and epithelial cancer development. We also found enrichment in footprints for activator protein 1 (AP1), which is known to interact with the glucocorticoid receptor in mediating the response to stress, and for

TF	OR	Pathway	Phenotype
Arnt	54.1	Xenobiotic metabolism	Hypoxia
SWI5	44.2	DNA repair	Response to ionizing radiation
MAFA	45.2	Insulin gene expression	Insulinomatosis, Diabetes
AML1	39.2	Wnt signalling	Epithelial cancer

Fig 2. N-SNPs in transcription factor binding sites (A) N-SNPs were annotated in transcription factor binding sites using the centiSNPs annotation, which includes a computational prediction of the SNP's effect on transcription factor binding. The percentage of N-SNPs in each prediction category is shown in the pie chart. **(B)** Histogram of enrichments of N-SNPs that disrupt transcription factor binding compared to non N-SNPs for 435 transcription factors. **(C)** Phenotypic relevance of the top transcription factor motifs enriched for N-centiSNPs in their binding sites and that are involved in environmental responses.

https://doi.org/10.1371/journal.pgen.1009493.g002

multiple transcription factors (c-Jun, ATF2, and Elk1 [36–38]) that are downstream of the JNK signaling pathway, which is often induced by environmental stress.

The top 15 enriched motifs (OR>38, S1 Table) also include factors associated with glucose metabolism. For example, MAFA activates expression of the insulin gene (INS) [39]. Mutations in this gene result in the autosomal dominant disease Insulinomatosis and Diabetes Mellitus [40]. Furthermore, when we considered 37 transcription factors that mediate the response to glucose and/or insulin [41,42], 16 (43%) were enriched for N-centiSNPs in their footprints genome-wide, suggesting a potential contribution of Neanderthal introgression to modern human metabolism. To further investigate the regulatory function of N-SNPs, we considered a dataset of allele-specific binding for SNPs associated with Type 2 Diabetes risk and assayed through SNP-SELEX [43]. One advantage of this method over other experimental assays is that allele-specific effects are measured in vitro, outside of a cellular context, therefore are independent of any trans-effects and the identity of the transcription factor is known. SNP-SELEX provides a preferential binding score (PBS), which compares the relative binding affinity of two alleles for a specific transcription factor. 15,431 N-SNPs were assayed by SNP-SELEX across 270 factors. We considered significant PBS values, using a cut-off of $p<0.05$. There were 505 N-SNPs that had significant PBS values. We conducted an enrichment analysis via Fisher's exact test and found that N-SNPs were more likely to have a significant PBS, thus affecting binding to a transcription factor, when compared to non-N-SNPs in the SNP-SELEX data set (OR = 1.105626, $p = 0.03861$). We investigated if the distribution of PBS was different between the N-SNPs and non-N-SNPs. We used the Kolmogorov-Smirnov test and found that the distribution of PBS for non-N-SNPs was significantly different from the

distribution of PBS for N-SNPs (D = 0.011458, p = 0.03614). We then repeated the Kolmogo-rov-Smirnov test for every transcription factor tested both computationally and in SNP-SE-LEX to find if the directional effects on binding are similar or different between N-SNPs and non N-SNPs. We filtered our results for an adjusted p-value less than 0.1 and found 12 factors with significantly different PBS distribution between N-SNPs and non N-SNPs: for EGR4, ELF2, ETV2, IRF5, POU2F3, ZBTB26, and ZNF740 the Neanderthal allele decreased binding, while for HOXA6, IRF8, PAX2, TGIF2LX, and ZNF821 the Neanderthal allele increased bind-ing (S3 Fig). We compared our N-centiSNP predictions for the ELF2 transcription factor to the experimental SELEX PBS score and found that both indicated decreased binding for the introgressed variants. Lastly, we conducted a literature search to investigate the potential bio-logical effects of the N-SNPs that alter TF binding in the SNP-SELEX data. Because the vari-ants within the SNP-SELEX data are cis-regulatory variants near type-2 diabetes risk loci, many of the transcription factors were associated with related processes, such as insulin and glucose metabolism or adipogenesis. Some notable examples include IRF5, which when defi-cient can cause increased insulin sensitivity in obese mice [44]. We found that the Neanderthal allele decreased binding for this transcription factor, which may mean Neanderthals experi-enced enhanced insulin sensitivity. Another example is PAX2, which may upregulate expres-sion of the glucagon gene [45]. The Neanderthal allele had increased binding for PAX2, indicating Neanderthals may have increased glucagon gene expression.

Overall our results support previous findings that Neanderthal introgression may play a role in the immune response, but importantly also suggest a novel role in regulatory mecha-nisms for the response to broader environmental stress, and for glucose metabolism.

Neanderthal introgression regulates the transcriptional response to environmental perturbations

Previous studies have shown that Neanderthal alleles play an important role in human gene regulation [23,24]. To directly investigate whether Neanderthal alleles regulate the expression of genes that respond to environmental perturbations, we considered expression quantitative trait loci (eQTLs) from the GTEx database and genes differentially expressed in a panel of 52 cellular environments, resulting from 5 cell types and 26 treatments, including hormones, vita-mins, drugs, and environmental contaminants [30]. We focused on eQTLs for genes differen-tially expressed in each cellular environment considered. Individuals carrying a high expression genotype at an eQTL for a gene that responds to a specific environmental stimulus will have an overall greater response compared to individuals with a low expression genotype in the same environment (Fig 3A). Overlapping our N-SNPs catalog with the GTEx database, we found that 30,464 GTEx eQTLs (16,065 unique SNPs) in any tissue are N-SNPs. When we considered the set of differentially expressed genes in response to treatments, we found that 27,789 N-SNPs are eQTLs for genes differentially expressed in at least one condition, corre-sponding to 15,026 unique SNPs and 1,629 unique genes (S2 Table). For 8 cellular environ-ments we identified a significant enrichment for N-SNP eQTLs in genes differentially expressed, compared to genes that do not respond to the treatment considered (BH-FDR = 0.05) (Fig 3B and 3C). Among these differentially expressed genes regulated by introgressed alleles, we found 239 genes that respond to dexamethasone, a synthetic glucocor-ticoid that regulates gene expression through activation of the glucocorticoid receptor (GR). The GR is ubiquitously expressed and glucocorticoids regulate different key biological pro-cesses depending on the specific body site. For example, in immune cells glucocorticoids act as immuno-suppressors to prevent systemic inflammation and sepsis, while in vascular endothe-lial cells, glucocorticoids regulate angiogenesis and vascular remodeling. Because of the anti-
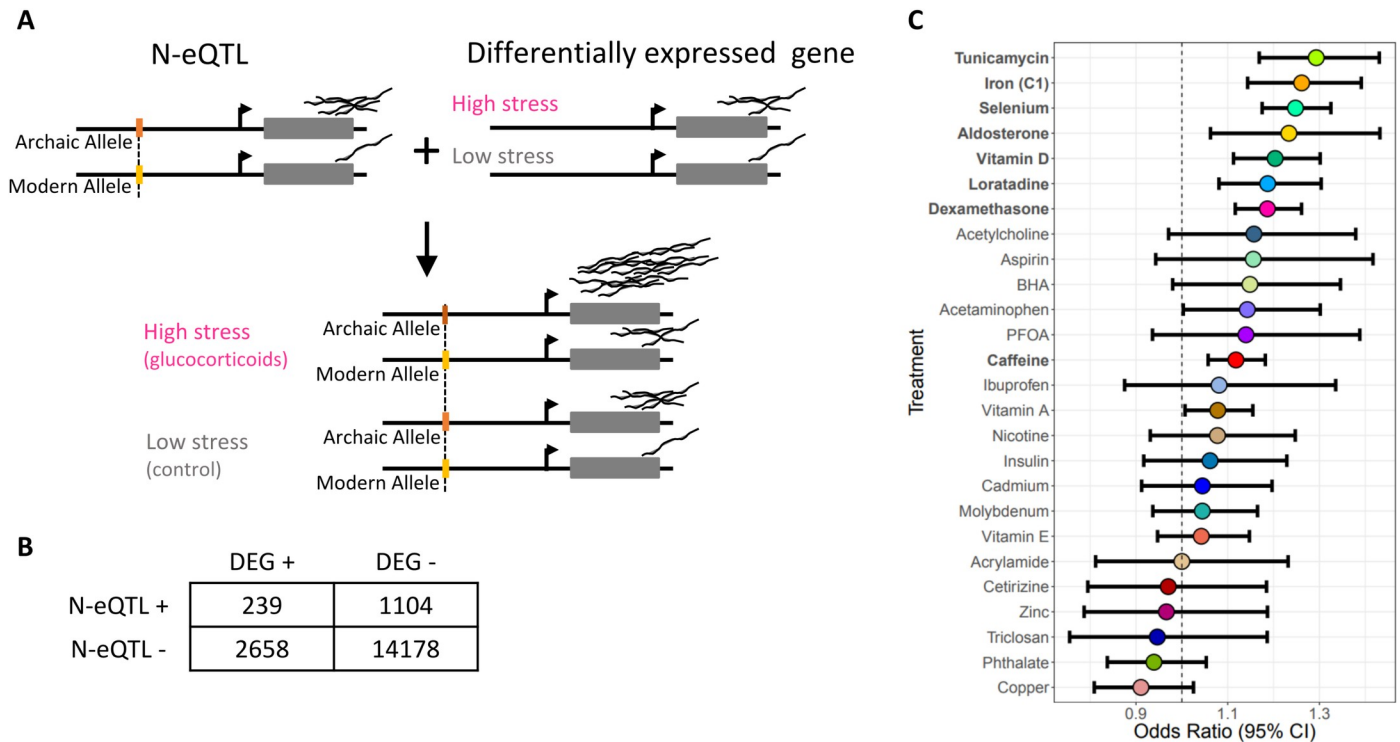
Fig 3. N-SNPs regulate differentially expressed genes (A) Neanderthal introgressed alleles can be linked to the gene they regulate through eQTL mapping signals reported by GTEx. If these genes respond to environmental perturbation, the introgressed variant will contribute to modulating the transcriptional response through additive genetic and environmental effects. **(B)** Example of a contingency table used to test for significant enrichment of N-SNPs for genes that respond to glucocorticoids (dexamethasone treatment). **(C)** Odds ratios and 95% confidence intervals for enrichment of N-SNP eQTLs in genes differentially expressed to each treatment.

https://doi.org/10.1371/journal.pgen.1009493.g003

inflammatory function of glucocorticoids, it is likely that the observed enrichment captures the same signature of Neanderthal introgression on transcriptional regulation of immune response, which was reported in other studies [26,25]. For example, we found a N-eQTL for the gene RAPGEF3, which encodes for a Ras GTPase. This gene is downregulated in response to dexamethasone in PBMCs and was also among the infection eQTLs identified in monocytes treated with the influenza A virus [26].

We observed a 1.2 fold enrichment for N-SNPs that regulate genes responding to Vitamin D (1,435 genes). Vitamin D synthesis depends on UV exposure, with important trade-offs between UV exposure and skin pigmentation and its effect on Vitamin D production. Our observation suggests a role for Neanderthal introgression in the response to UV exposure however the specific adaptive phenotype remains to be defined. For example, BNC2 is downregulated in response to Vitamin D in melanocytes, but upregulated in PBMCs and vascular endothelial cells. In our analysis we found that rs2288634 is a N-eQTL that regulates expression of ATP1B3, a gene down-regulated in response to Vitamin D in PBMCs. This N-eQTL is also associated with hair color in the GWAS catalog. Overall we found a total of 66 N-eQTLs for genes that respond to environmental stimuli that are also significantly associated with a complex trait in the GWAS catalog (S3 Table).

Adaptive introgression of N-SNPs

It is well known that more introgressed segments of Neanderthal DNA have been removed by purifying selection in genomic regions with high selective constraint [46]. Less selective constraint could explain the patterns observed in this study, in terms of excess of N-SNPs

predicted to alter TF binding sites as well as N-eQTLs for responding genes. For example, genes with a transcriptional response to environmental perturbations could be less constrained, therefore more likely to harbor introgressed regulatory N-SNPs. However, the set of N-SNPs considered in this study have similar levels of background selection (KS test $p>0.05$), as measured by the B-statistic from McVicker et al. [47], thus indicating that the observed enrichments are not due to relaxation of selective constraint, but may suggest adaptive introgression in gene regulation (S5 Fig). Selective pressures on introgressed Neanderthal variation can leave a genomic signature indicative of adaptive introgression. To test for possibility of Neanderthal adaptive introgression at the N-eQTLs in Europeans, we computed summary statistics that have been demonstrated to capture the signature of adaptive introgression, including the RD statistic, the Q95 statistic, and the U20 statistic [8]. The RD statistic is defined as the average ratio of the sequence divergence between an individual from the source population and an individual from the admixed population, and the sequence divergence between an individual from the source population and an individual from the non-admixed population. If a region of the genome is adaptively introgressed into a non-African population, the sequence divergence between Neanderthals and non-African populations should be less than the sequence divergence between Neanderthals and African populations, and, as a result, RD would be small. We identified 6,403 N-SNPs in windows that were significant for the RD statistics (lowest 5% of RD values genome-wide), corresponding to 397 genes (S2A Fig). On average, 23.6% of differentially expressed genes in response to all treatments are regulated by an N-SNP in a region with an adaptive introgression signal via RD statistic. For each treatment, the number of genes in regions with evidence of adaptive introgression from the RD statistics is proportional to the number of differentially expressed genes regulated by a N-eQTL (S6 Fig). The U20 statistic is based on allele frequency in the recipient and donor population and identified adaptive introgression variants based on their elevated allele frequency in the recipient population ($>20\%$) compared to the donor population. 6,281 SNPs were in regions of adaptive introgression identified via the U20 statistic, corresponding to 296 genes (S2B Fig). On average, 17.9% of differentially expressed genes in response to all treatments are regulated by an N-SNP in a region with a significant U20 statistic. We also computed the Q95 statistics, that considers "high-frequency" archaic alleles, and represents a summary statistic of the site-frequency spectrum in the recipient population. This statistic should be high when a region contains many alleles at especially high frequencies in the recipient population. 6,917 SNPs had a signal of adaptive introgression (highest 5% of Q95 statistic genome-wide) via the Q95 statistic, corresponding to 408 genes (S2C Fig). On average, 25.4% of differentially expressed genes in response to all treatments are regulated by an N-SNP in a region with an adaptive introgression signal via Q95 statistic.

To identify signatures of adaptive introgression, we considered genomic windows that exceed the critical values on all three adaptive introgression summary statistics (the most extreme 5% quantile value from its genome-wide distribution). We therefore identified 1000 N-SNPs in windows that exceed the critical values on all three adaptive introgression summary statistics, hence candidates of adaptive introgression. Of these adaptive introgressed N-SNPs, 384 are N-eQTLs and 363 regulate genes that respond to treatments, for a total of 132 genes (Fig 4A). The proportion of genes regulated by N-eQTLs in regions of adaptive introgression is similar across all treatments (7.5% on average, Fig 4B) and it is not significantly different for genes that do not respond to a treatment.

Experimental validation of Neanderthal introgressed alleles

Association studies, including eQTLs, are limited in their ability to identify true causal variants in the region of association with the trait of interest. Even considering the lead eQTL for each
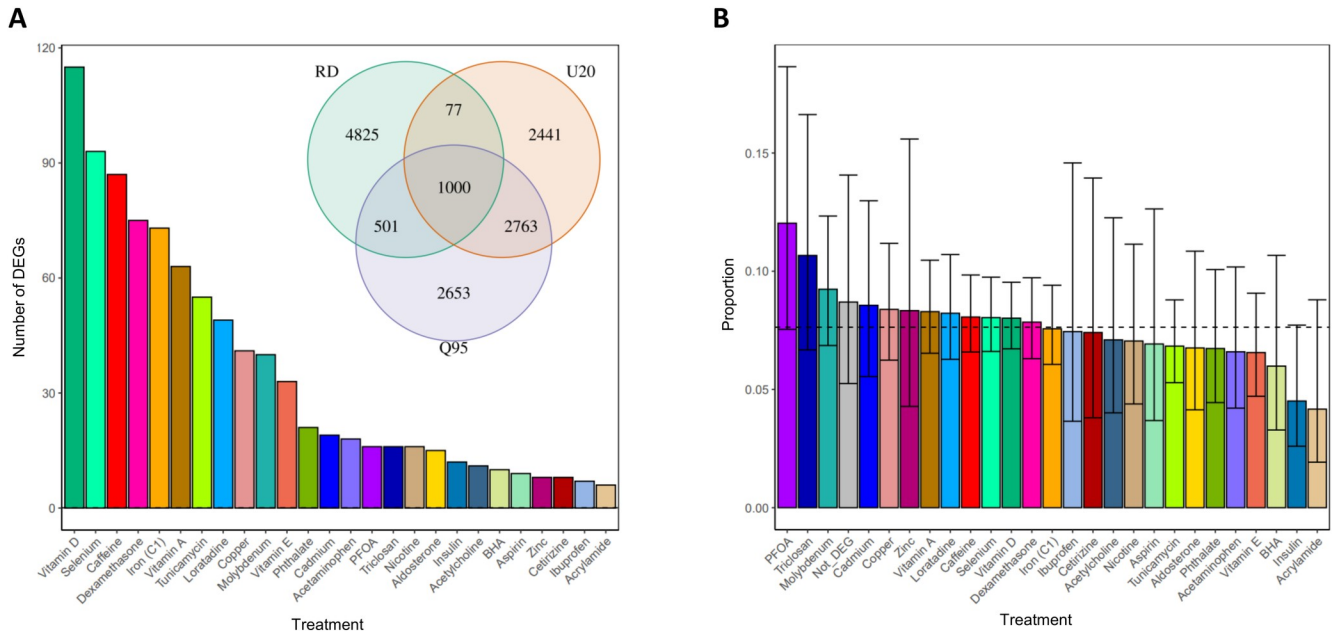
Fig 4. Signals of adaptive introgression in genomic regions containing N-eQTLs (A) Number of genes differentially expressed in response to the treatments and that are regulated by N-eQTLs in regions of adaptive introgression. The inset venn diagram shows the number of N-SNPs which are outliers for each statistic. (**B**) Proportion of genes from A, relative to the number of genes differentially expressed that are regulated by N-eQTLs. The grey bar indicates the proportion of genes regulated by N-eQTLs with adaptive introgression but that are not responding to the treatments.

eGene does not ensure that the true causal variant is being considered. Massively parallel reporter assays (MPRAs) can be used to test the gene regulatory function of DNA sequences carrying individual variants by transfecting them in human cells as part of reporter gene plasmids. To validate N-SNPs that regulate gene expression response, we used BiT-STARR-seq, an MPRA we recently developed to test allele-specific regulatory function for tens of thousands of centiSNPs. We identified 21 N-SNPs (9.3%) with allele-specific regulatory function (BH-FDR<0.1) when considering the 226 N-SNPs included in the library of variants used in [48] (Fig 5A, Table 1 and S4 Table). This validation rate was expected and was not significantly different than the validation rate in the original study, which did not distinguish between N-SNPs and modern SNPs (2,720 out of 43,500, 6.2%). Interestingly, the Neanderthal allele led to increased gene expression for 20 of the 21 N-SNPs, which was a significantly different pattern compared to modern human alleles (binomial test $p = 2.41 \times 10^{-8}$). This result could not be explained by differences in background selection between genomic regions harboring experimentally validated compared to non-validated N-SNPs (S4 Fig). Of these 21 N-SNPs, 13 are centiSNPs, 8 are eQTLs in GTEx and 5 are infection eQTLs in a study of macrophages exposed to listeria and salmonella [25]. The N-centiSNP rs4784812 is in a region with significant evidence for adaptive introgression based on both the RD and the Q95 statistics (RD p-value = 0.046, Q95 p-value = 0.040). The Neanderthal allele at rs4784812 is predicted to alter binding for six transcription factors (NF-Y, CCAAT box, ELK4, or with an ETS DNA binding domain).

We then investigated whether the 8 N-SNPs GTEx eQTLs regulate genes that respond to environmental perturbations. We found that these genes respond to treatments and each of them is differentially expressed in response to several different treatments, as shown in Fig 5B and 5C. While this set of 8 eQTLs is not a random sample of regulatory N-SNPs, nevertheless, each of the genes they regulate is likely to be central to the cellular response to multiple environmental perturbations, rather than to specific stimuli.
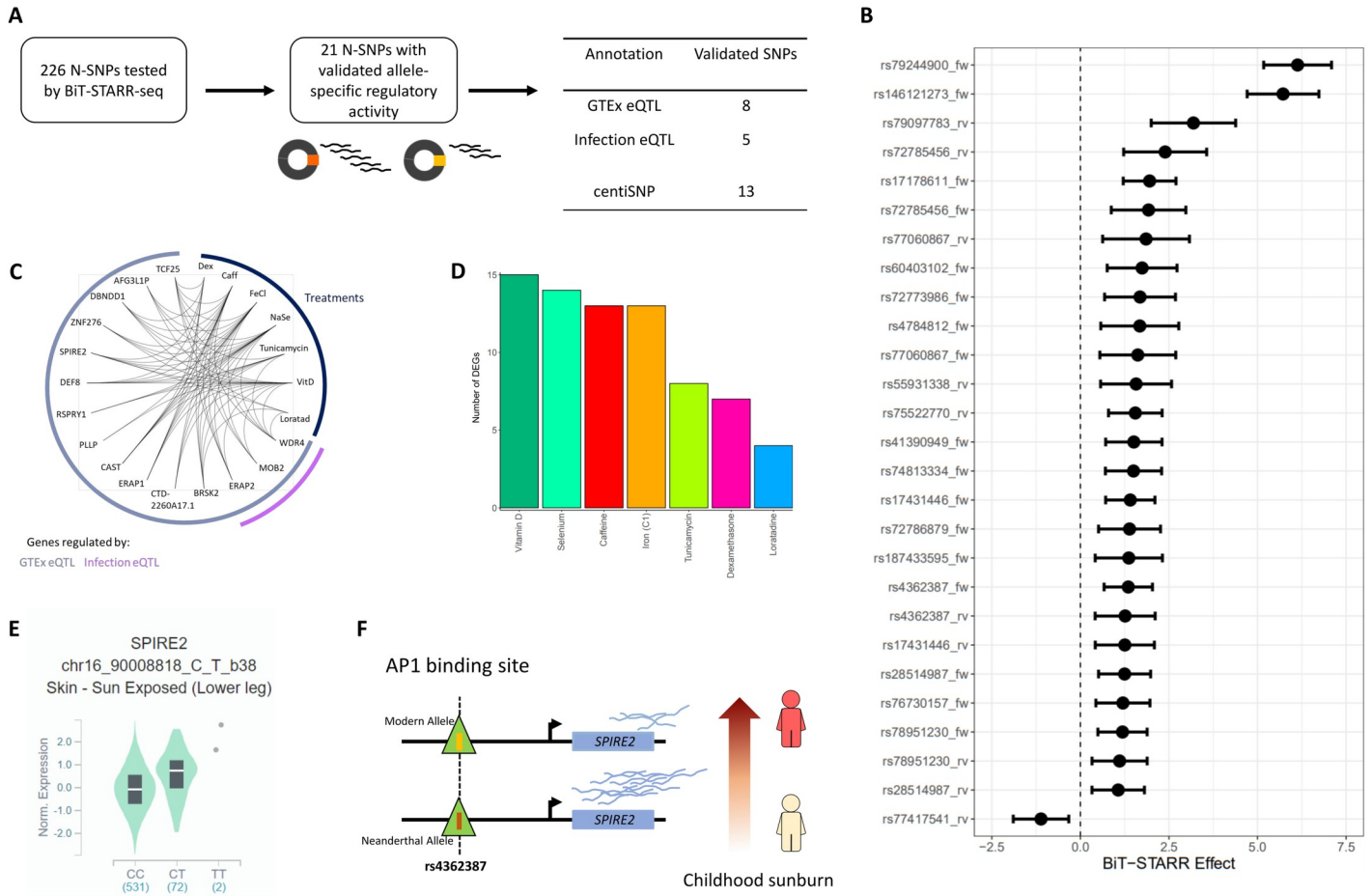
Fig 5. Validation of N-SNP function on gene expression (A) Number of N-SNPs tested and validated by BiT-STARR, with genomic annotations. (B) Forest plot showing the effect of the Neanderthal introgressed allele on gene expression in the BiT-STARR-seq experiment. The allelic imbalance is normalized to the allelic ratio in the DNA library. (C) Network depicting the genes regulated by N-SNPs validated by BiT-STARR and the treatments in which they are differentially expressed. (D) Number of differentially expressed genes per treatment regulated by validated N-SNPs. (E) Violin plot of the eQTL signal in the GTEx data for rs4362387 and SPIRE2. (F) Graphic representation of a likely mechanism connecting rs4362387 to childhood sunburn based on the molecular signals presented in this study.

https://doi.org/10.1371/journal.pgen.1009493.g005

Three notable examples are rs72773986, rs28514987 and rs4362387. rs72773986 is an infection eQTL and a GTEx eQTL for ERAP2, and is predicted to alter binding of EPAS1, a key transcription factor for response to hypoxia. ERAP2 encodes a zinc metalloaminopeptidase involved in antigenic response and polymorphisms at this locus are responsible for differential response to bacterial infection and influenza [49]. ERAP2 was also associated with susceptibility to Crohn's disease [50]. rs28514987 is in a region with significant signals of adaptive introgression by U20 and Q95 statistics (U20 p-value = 0.022, Q95 p-value = 0.031). It is also a GTEx eQTL for BRSK2 and is in a bZIP911 binding site. BRSK2 is expressed in pancreatic islets and negatively regulates insulin secretion [51]. rs4362387 is located within an AP1 binding site and is a GTEx eQTL for SPIRE2 and 6 additional genes in the region. This genomic region also contains the MC1R gene, which is implicated in skin and hair pigmentation and rs4362387 is also an eQTL also for MC1R. The expression of both SPIRE2 and MC1R is highly upregulated by Vitamin D (fold change = 2.6–2.7, FDR = 0) and likely share regulatory elements. In GTEx, a strong eQTL effect (both in the single tissue and multi-tissue analysis) is observed in the skin tissues, where the T allele (Neanderthal allele) is associated with higher expression of both genes (Skin sun exposed; SPIRE2: normalized effect size: 0.52; MC1R:

Table 1. N-SNPs validated by MPRA. Columns 1–12 are: 1) rsID; 2) Chromosome; 3) Position; 4) Modern human allele; 5) Neanderthal allele; 6) Allele frequency among all 1000 Genomes populations; 7) RD p-value; 8) U20 p-value; 9) Q95 p-value; 10) transcription factor motif name; 11) GTEx gene name; 12) Infection eQTL gene name.

rsID	chr	pos1	MH	Nean	AF	RD	U20	Q95	TF Name	GTEx gene name	ieQTL gene name
rs79244900	1	234041317	C	T	0.059	0.363	1	0.040	ATFs,ATF2:c-Jun,CREB,DREB1B,CREM	NA	NA
rs146121273	15	52436165	A	C	0.017	0.548	1	0.418	Elk-1,Ets,NUR77,PPARA	NA	NA
rs79097783	21	44229777	T	G	0.027	0.128	1	0.100	NA	NA	WDR4
rs17178611	21	44229685	A	G	0.027	0.128	1	0.100	NA	WDR4	WDR4
rs75522770	21	44268431	G	A	0.027	0.096	1	0.132	NA	WDR4	WDR4
rs72785456	2	11309644	A	G	0.016	0.088	1	0.123	AP1	NA	NA
rs17431446	12	120884291	C	T	0.041	0.450	1	0.123	StuAp,Pax-3,AML1	NA	NA
rs4362387	16	90075226	C	T	0.057	NA	NA	NA	STB4,AP1	DEF8,SPIRE2,ZNF276, DBNDD1,AFG3L1P,TCF25	NA
rs41390949	16	57447951	G	A	0.050	0.190	1	0.164	NA	NA	NA
rs74813334	16	57298926	G	T	0.110	0.270	1	0.123	NA	PLLP,RSPRY1	NA
rs60403102	16	57297312	G	A	0.114	0.270	1	0.123	NA	RSPRY1,PLLP	NA
rs78951230	9	85198625	G	A	0.028	0.894	1	0.152	SF1,PPARA,NR4A2	NA	NA
rs28514987	11	1487024	C	G	0.049	0.146	0.022	0.031	bZIP911	MOB2,BRSK2	MOB2
rs72773986	5	96159717	G	A	0.070	0.207	1	0.062	HIF-2alpha	AC008906.1,ERAP1,CAST, ERAP2	ERAP2
rs72786879	16	57410998	G	A	0.064	0.640	1	0.164	NA	NA	NA
rs76730157	8	123793459	G	A	0.021	0.196	1	0.272	StuAp,ZF5,MBP1	NA	NA
rs55931338	1	3321941	A	G	0.018	0.738	1	0.272	GBP,TAF-1,GBF,E4F1,CPRF-3,TGA1b, GBF1	NA	NA
rs4784812	16	57536204	A	C	0.081	0.040	1	0.046	NF-Y,CCAAT box,c-ets-1 (ets1),ELK4,/ EMBO10/Fli1, /EMBO10/Gm4881	NA	NA
rs77060867	16	57298163	G	A	0.110	0.270	1	0.123	NA	PLLP,RSPRY1	NA
rs187433595	11	119205315	T	C	0.005	0.710	1	0.418	GABP-alpha	NA	NA
rs77417541	9	93660685	C	T	0.044	0.406	1	0.176	LBP-1	NA	NA

p = 0.0000026, normalized effect size: 0.26). In the MPRA data, the Neanderthal allele increases expression of the reporter gene, thus showing the same direction of effect as the one observed in the eQTL data, and is also predicted to alter AP1 binding.

Discussion

In vitro studies allow researchers to analyze molecular phenotypes in specific environments. While these environments are a simplified version of organismal exposures, they have two key advantages: they are tightly controlled and can simulate environments that are difficult to measure *in vivo*, including ancient environments. To directly investigate whether Neanderthal introgression may contribute to modern humans' ability to respond to a wider range of stressors, beyond pathogens, we used data on transcriptional response in 35 cellular environments that reflect common individual exposures, including dietary components, environmental contaminants, metal ions, over the counter drugs, and hormonal signaling. While some of these compounds clearly represent very recent exposures in contemporary human history, the underlying hypothesis is that these compounds target existing response mechanisms. For example, endocrine disruptors, such as BPA, which are commonly found in plastics, target the estrogen response pathway [52]. Our approach combines functional genomics and paleogenomics, identifying Neanderthal introgressed variants which may have gene regulatory function. We hypothesized that Neanderthal introgression may have contributed to adaptations to

environmental changes by introducing regulatory variants that modify binding of specific transcription factors to a large number of downstream targets. When considering human transcription factors, variation in the binding sites is more common than non-synonymous variants in the transcription factor genes themselves. This is in line with the idea that changes in the function of master regulators would have large pleiotropic effects[53]. Variation in transcription factor binding sites, instead, introduces specific functional consequences but preserves unaltered the overall function of the transcription factor[54,55]. Using computational predictions and experimental data on eQTLs and transcriptional response, we assigned a putative function to 33,248 variants which are likely to contribute to modern human response to the environment. For 21 N-eQTLs we validated their ability to modify gene expression in massively parallel reporter gene assays.

While most studies of archaic introgression in the genome of modern humans have focused on non-African populations, recent results from Chen et al [56] indicated the presence of Neanderthal haplotypes in sub-Saharan African populations due to back migration from non-Africans. Our set of N-SNPs excludes variants where the Neanderthal allele is present in African populations, potentially missing Neanderthal alleles re-introduced in sub-Saharan African genomes. It is thus possible that functionally important N-SNPs have been missed in the present study due to our inclusion criteria. Indeed, when considering N-SNPs that are polymorphic in African populations, we identified 1919 (out of 6,884) that are eQTLs and regulate a total of 1158 genes. Of these genes, 654 are regulated by N-SNPs found in sub-Saharan African populations and would be missed if considering only N-SNPs found in non-Africans. The BiT-STARR assay included 61 of the N-SNPs potentially introgressed in sub-Saharan African populations, 7 of which showed allele-specific regulatory function (S5 Table). While further studies are needed to fully understand the impact of archaic introgression introduced by back-migration to Africa, our results suggest a regulatory function for some of these N-SNPs present in African genomes.

While several studies have highlighted a potential contribution of archaic introgression to human phenotypes, the molecular evidence in most cases is suggestive and does not provide experimental data in support of a putative molecular mechanism linking archaic alleles to specific functions. Assigning a function to non-coding variants is a daunting task, which is further complicated by the limited availability of archaic genomes and the inability to directly measure archaic molecular phenotypes [57]. By focusing directly on introgressed alleles and considering fine scale computational predictions, here we have provided an annotation of Neanderthal introgressed alleles at single nucleotide resolution. One advantage of this annotation is that it pinpoints transcription factors whose function is likely to be affected by these archaic variants. For 434 transcription factor motifs Neanderthal introgressed alleles are predicted to modify DNA binding. These putative molecular mechanisms represent ideal candidates for follow-up studies focused on specific transcription factors and/or variants.

Another advantage of focusing on the specific introgressed nucleotides within a full haplotype is the ability to directly test and validate the gene regulatory function of these non-coding variants. Massively parallel reporter assays have been successful in identifying new regulatory variants and validating computational predictions of regulatory elements and variants [58–63]. One key advantage of this approach is the ability to test tens of thousands of variants in parallel because the assays use a library of plasmid that can be designed to investigate a catalog of variants of interest. At the same time, these assays cannot test genetic variants within their native chromatin context and by design are unable to test long distance interactions. As a consequence, additional introgressed variants that we were not able to validate in this study may indeed regulate gene expression in other environmental contexts. Nevertheless, we provide a proof of concept that applying computational and experimental functional genomics

approaches enables researchers to assay the gene regulatory function of archaic alleles in modern human cells. Recently, MPRA data have been used to show that ancestral alleles reintroduced by Neanderthal introgression have activity levels similar to non-introgressed variants, while Neanderthal derived alleles are depleted for regulatory activity compared to reintroduced ancestral alleles [64]. Interestingly, in the dataset we considered, 95% of the functionally validated Neanderthal alleles led to increased gene expression. If future studies show this to be the case for other introgressed non-coding variants, it would suggest that variants that disrupt function at the gene regulatory level were not tolerated by modern humans and therefore are less likely to be introgressed in our genomes.

One key result of our study is the finding that Neanderthal introgression may contribute not only to modern human's ability to respond to pathogens, as previously reported, but likely contributes to our ability to adapt and respond to a broader set of environmental challenges. We discovered that introgressed alleles are enriched in binding sites for transcription factors that regulate response to environmental stressors, including hypoxia and ionizing radiation, and also for transcription factors involved in glucose metabolism. These do not seem to be incidental findings, as they find support in the results from our complementary approaches. For example, we found an enrichment for Neanderthal alleles that disrupt binding for 16 transcription factors involved in the response to glucose and insulin. Additionally, using SNP-SE-LEX data targeting SNPs involved in glucose metabolism and T2D risk, we identified sets of transcription factors where Neanderthal introgression increases or decreases binding with likely consequences for glucose and insulin metabolism. The transcription factors include IRF5 and PAX2, which are involved in insulin sensitivity and glucagon gene expression respectively. One of the N-eQTLs functionally validated in the BiT-STARR-seq assay regulates the expression of BRSK2, a gene that plays a role in the regulation of insulin secretion in response to elevated glucose levels [51]. This N-eQTL is in a region of adaptive introgression, characterized by significant U20 and Q95 statistics.

Several studies in the last decade have identified genetic variants that modify how cells respond to environmental perturbations. These molecular gene-environment interactions have shown that genetic variants have varying effects on gene expression depending on the specific context considered and that these genetic effects are also important for complex trait variation in humans [25,26,65–76]. The majority of these studies have focused on the response to pathogens thus there is limited availability of eQTL datasets to investigate the role of Neanderthal introgression on GxE in modern humans. However, the response to environmental changes depends not only on interaction effects but also on additive effects between environmental variables and genotypes. In other words, an individual carrying a high expression genotype at an eQTL for a gene that responds to a specific environmental stimulus, will have an overall greater response than an individual with a low expression genotype in the same environment. This is because the gene is induced and consequently the genetic effect is amplified [77]. The observed enrichment of N-eQTLs for responding genes suggests a contribution of Neanderthal introgression to modern humans' transcriptional response to treatments that represent the in vitro counterpart for common exposures, some of which may have represented a selective pressure during human history. Some examples include Vitamin D (a proxy for solar radiation), aldosterone (the main hormone regulating fluid balance), tunicamycin (endoplasmic reticulum stress agent), and glucocorticoids, which exert an anti-inflammatory function in the human body. Glucocorticoids are regulated by the hypothalamus-pituitary adrenal axis in response to stress, thus representing a molecular connection between stress and immune function.

We also found that genes responding to vitamin D, which is produced in response to solar radiation [78], are enriched for Neanderthal eQTLs (OR = 1.2, FDR = 1.9 x 10^{-5}). Previous

studies suggested that Neanderthal introgressed haplotypes contribute some of the genetic variation underlying the red hair phenotype in modern humans [23,79]. A non-synonymous variant in the MC1R gene was identified in a Neanderthal genome but was not seen in subsequently sequenced individuals, thus suggesting that it may represent a polymorphic site in Neanderthal populations. The N-eQTL we functionally validated is significantly associated with red hair in the UK Biobank data ($p = 2.2 \times 10^{-308}$) and is also an eQTL for MC1R and SPIRE, two genes that are differentially expressed in response to Vitamin D. In transcriptome wide association studies, both genes are associated with skin and hair pigmentation phenotypes, including red hair color and childhood sunburn [80] (Fig 5D). While it is difficult to pinpoint a specific phenotypic effect for each of the introgressed sequences, it is likely that the most relevant environmental stressor is solar radiation. Overall, our results raise the possibility that multiple alleles and mechanisms underlie the red hair phenotype not only in humans but also in Neanderthals. As more response eQTL datasets become available, they will enable further analysis of the contribution of Neanderthal and other archaic humans' introgression to gene expression response in modern humans, including for example estimating the portion of gene expression response heritability that can be attributed to archaic introgression.

With increasing availability of ancient DNA sequences, combining paleogenomic and functional genomic tools will allow researchers to explore in greatest depth the functional differences and similarities of modern humans from our extinct relatives. Together with other recent studies, we have demonstrated the value of taking these complementary approaches. Our results uncovered a putative role of Neanderthal introgression to modern humans' ability to adapt and respond to environmental changes, with potential consequences for glucose metabolism and response to solar radiation.

Methods

Identification of introgressed Neanderthal variants

We determined a list of putative Neanderthal introgressed SNPs by screening the biallelic SNPs from the 1000 genomes project (1KGP) [34] based on three criteria. The conservative criteria address the notions that Neanderthal introgression occurred outside of Africa and that the Neanderthal-introgressed haplotypes are not observed at certain regions of the human genome. Briefly, the criteria for a Neanderthal-introgressed SNP are: 1) the SNP is within the genomic regions reported to harbor the Neanderthal introgressed haplotypes [2], we used the combined regions from all the populations tested including Europeans, East Asians, South Asians, and Melanisians. We intended to be more generous when defining the Neanderthal introgressed regions to establish a more general input for samples with different ancestry, and compensate this relaxed stringency with conservative screening steps described below; 2) the SNP is confidently genotyped in at least two of the three high coverage Neanderthal genomes available [31–33] and is fixed among the genotyped alleles; 3) the SNP has all of the 504 sub-Saharan Africans (ESN, GWD, LWK, MSL, YRI) in 1KGP fixed for an allele alternative to the Neanderthal one, and is polymorphic among the non-Africans (present in at least one individual). The rationale of these criteria is to address the notion that much of the Neanderthal introgression occurred outside of Africa. We found a total of 186,871 SNPs that pass these criteria for further analysis. We further removed structural variants and multi-allelic SNPs to end up with 177,578 N-SNPs.

As illustrated by Chen et al [56], sub-Saharan African populations contain Neanderthal introgressed sequences likely due to migration from ancestral Europeans back to Africa. Our requirement that the modern human allele be fixed in sub-Saharan African populations filtered out 6,884 SNPs which could potentially be Neanderthal in origin. As a result, we analyzed

these SNPs separately to determine the number which regulate differentially expressed genes and which were validated by BiT-STARR-seq, as described below.

N-SNPs in transcription factor binding sites

We annotated N-SNPs present in transcription factor binding sites using the centiSNP annotation from Moyerbrailean *et al* [21]. Specifically, we used the centiSNPs Compendium (http://genome.grid.wayne.edu/centisnps/compendium/) which summarizes all 1000 Genomes Project SNPs in 1,372 transcription factor motifs across 153 tissues and estimates the likelihood that a SNP disrupts transcription factor binding. N-SNPs were located within 1,255 transcription factor motifs. N-centiSNPs are defined as those that are predicted to alter binding of at least one transcription factor. For each of the 435 transcription factor motifs with at least 10 N-SNPs within their binding sites, we performed Fisher's exact test to determine whether N-SNPs in transcription factor binding sites were more likely to disrupt transcription factor binding than non-introgressed SNPs within transcription factor binding sites. Multiple-test correction was performed using the Benjamini-Hochberg procedure [81]. Input data and results from the Fisher's exact tests can be found in S1 Table.

To further investigate the regulatory function of N-SNPs in glucose metabolism, we used a data set that analyzed allele-specific binding of SNPs with a minor allele frequency >1% in regulatory sequences within 500kb of type 2 diabetes lead SNPs from GWAS or SNPs in linkage disequilibrium with them ($r^2 > 0.8$) using SNP-SELEX [43]. SNP-SELEX is an in vitro assay that evaluates binding of specific transcription factors to DNA oligos outside of a cellular context. Briefly, to perform SNP-SELEX, oligos matching 40 bp sequences including the SNP of interest are synthesized. To examine binding, in each experiment the oligos are combined with a single transcription factor. Unbound oligos are washed from the plate such that only oligos bound to the transcription factor remain and can be identified through sequencing. The enrichment of each oligo present in the final sample is used to determine binding affinity and to calculate binding effect for each SNP. This relative enrichment is used to calculate the preferential binding score (PBS), which is defined as the difference in relative enrichment between two alleles. PBS were obtained from supplemental data 2 [43]. We intersected the SNPs tested in this data set with our library of N-SNPs. We performed Fisher's Exact Test for enrichment analyses. We then performed the Kolmogorov-Smirnov test on the PBS values to compare the distribution of PBS between N-SNPs and non-N-SNPs. We also repeated this analysis for each individual transcription factor.

N-SNP regulation of differentially expressed genes

To link N-SNPs to genes they regulate, we annotated N-SNPs which were significant eQTLs in any of 48 tissues in GTEx v7 (https://storage.googleapis.com/gtex_analysis_v7/single_tissue_eqtl_data/GTEx_Analysis_v7_eQTL.tar.gz). 16,065 N-SNPs were GTEx eQTLs in any tissue, regulating 1,939 genes. We then annotated N-eQTLs which regulated differentially expressed genes in response to environmental perturbation from Moyerbrailean *et al* [30]. We chose all treatment-cell type combinations with at least 1000 differentially expressed genes and downloaded the gene expression data from http://genome.grid.wayne.edu/gxebrowser/Tables/Supplemental_Table_S6.tar.gz. In total, we considered five cell types (melanocytes, peripheral blood mononuclear cells, lymphoblastoid cell lines (LCLs), smooth muscle cells, and human umbilical vein endothelial cells) and the 26 treatments shown in Fig 3B for a total of 52 treatment-cell type combinations. These experiments were performed in cell lines from European (N = 3), African (N = 3) and African American (N = 9) individuals. The cell types

were chosen by [30] because: they are relevant for the biology of complex traits, they were readily available primary cells (with normal karyotype, as opposed to cancer cell lines), can be passaged and cultured for multiple experiments and finally they can be collected non-invasively, therefore representing potentially ideal cell types for future larger eQTL mapping studies. We identified treatments in which differentially expressed genes were enriched for Neanderthal regulation using Fisher's exact test. We limited our analysis to genes with an eQTL and tested for an enrichment of N-eQTL genes in genes differentially expressed for each treatment-cell type combination separately. For treatments which were tested in multiple cell types, we meta-analyzed the enrichment odds ratios for each treatment using a fixed effects model with inverse variance weights to obtain a single enrichment per treatment. Multiple-test correction was performed using the Benjamini-Hochberg procedure [81].

Experimental validation of N-SNPs from BiT-STARR-seq

To validate N-SNPs which cause changes in gene expression, we used data from a massively parallel reporter assay (MPRA) known as BiT-STARR-seq performed in LCLs [48]. Allele-specific cis-regulatory effects in MPRAs are independent of the ancestry of the cell line used. Potential ancestry-specific trans effects (e.g. expression of a specific TF) would equally affect both alleles and should not create a bias. Results from BiT-STARR-seq were downloaded from https://genome.cshlp.org/content/suppl/2018/10/17/gr.237354.118.DC1/Supplemental_Table_S1_.txt. We also included an annotation for Neanderthal SNPs which were eQTLs in macrophages infected with *Listeria* or *Salmonella* from Nedelec *et al* [25] (downloaded from https://genome.cshlp.org/content/suppl/2018/10/17/gr.237354.118.DC1/Supplemental_Table_S4.txt). 226 N-SNPs were tested by BiT-STARR-seq for allele-specific regulatory differences, and 21 were significant after multiple test correction (Benjamini-Hochberg adjusted p value<0.1). BiT-STARR-seq results for all 226 N-SNPs can be found in S4 Table.

Adaptive Introgression summary statistics

To test for adaptive introgression, we chose a set of statistics developed by Racimo et al [8] and shown to robustly identify adaptive introgression. A description of each statistic is provided below. Because each individual statistic may yield significant results for other demographic or non-adaptive processes independently of adaptive introgression [8,82], here we decided to take a conservative approach and define as adaptive introgressed the N-SNPs in windows that exceed the critical values on all three adaptive introgression summary statistics.

RD statistic is defined as the average ratio of sequence divergence between an individual from the recipient and an individual from the donor population, and the divergence between an individual from the outgroup and an individual from the donor population. The U20 statistic is defined as the number of uniquely shared alleles between the recipient and donor population that are of frequency <1% in the outgroup, 100% in the donor, and >20% in the recipient population. The Q95 statistic is defined as the 95% quantile of the distribution of derived allele frequencies in the recipient population, that are of frequency <1% in the outgroup and 100% in the donor population [8].

In this analysis, we computed the above statistics in non-overlapping 50-kb windows across the autosomes using modern human genome data from Phase 3 of the 1000 Genomes Project [34] as in Zhang et al [82]. We used individuals from the CEU population as the recipient population, individuals from YRI population as the non-introgressed outgroup, and the unphased, high-quality whole genome sequence from the Altai Neanderthal individual [31] as the donor population. We define the statistical values for each eQTL as the AI summary statistic values of

each 50kb window that contains the eQTL SNP. The CEU was chosen as the recipient population to match the ancestry of the functional genomics data from Moyerbrailean et al [30], which were collected in cells from donors with European and African ancestries.

To identify signatures of AI, we adopted an "outlier approach" where we define the critical value for each statistic as the most extreme 5% quantile value from its genome-wide distribution [8,82] (S6 Table). We plotted the genome-wide distribution of the AI statistics as Manhattan plots in S2 Fig, with windows containing N-eQTLs highlighted in red. The solid line shows the critical value of each statistic.

Supporting information

S1 Table. Enrichment of N-centiSNPs to disrupt transcription factor binding. Each row represents a transcription factor. Columns 1–12 are: 1) Motif; 2) Odds ratio; 3) Lower confidence interval; 4) Upper confidence interval; 5) Enrichment p-value; 6) Number of N-SNPs which are predicted disrupt transcription factor binding; 7) Number of non-Neanderthal SNPs which are predicted to disrupt transcription factor binding, 8) Number of N-SNPs in transcription factor binding site which are not predicted to disrupt binding, 9) Number of non-Neanderthal SNPs in transcription factor binding site which are not predicted to disrupt binding, 10) FDR of enrichment, 11) transcription factor motif name.
(TXT)

S2 Table. N-eQTLs. For each N-SNP, we annotated the gene it regulates in GTEx and whether that gene is differentially expressed in Moyerbrailean *et al* [30]. Allele frequencies are from the 1000 Genomes project. Columns 1–47 are: 1) Chromosome; 2) Position; 3) rsID; 4) GTEx SNP ID. "NA" indicates the SNP is not an eQTL in GTEx; 5) Modern human allele; 6) Neanderthal allele; 7) Global allele frequency, 8) East Asian allele frequency, 9) American allele frequency, 10) European allele frequency, 11) South Asian allele frequency, 12) African allele frequency, 13) Ensembl gene id, 14) Gene symbol, 15–46) Treatments from Moyerbrailean *et al* [30]. "1" indicates the gene is differentially expressed in that condition in at least one cell type, 47) "1" indicates the gene is not differentially expressed in response to any treatment.
(TXT)

S3 Table. GWAS catalog overlap with N-eQTLs which regulate genes which respond to the environment. 66 N-eQTLs which regulate DE genes were found in the GWAS catalog. Columns 1–10 are: 1) Chromosome; 2) 0-based coordinate; 3) 1-based coordinate; 4) rsID; 5) Trait; 6) Risk allele; 7) P-value, 8) Odds ratio or beta, 9) 95% confidence interval, 10) PubMed ID.
(TXT)

S4 Table. BiT-STARR results for N-SNPs. Results for the 226 N-SNPs tested by BiT-STARR-seq by Kalita *et al* [48]. The table was downloaded from S1 Table from Kalita *et al* [48]. We added a final column indicating the FDR calculated using the 226 N-SNPs only.
(TXT)

S5 Table. N-eQTLs. For each of the N-SNPs filtered out due to it being polymorphic in Africans, we annotated the gene it regulates in GTEx, whether that gene is differentially expressed in Moyerbrailean *et al* [30], and if the SNP was validated by BiT-STARR-seq. Columns 1–43 are: 1) Chromosome; 2) Position; 3) rsID; 4) GTEx SNP ID. "NA" indicates the SNP is not an eQTL in GTEx; 5) Modern human allele; 6) Neanderthal allele; 7) Global allele frequency, 8) Ensembl gene id, 9) Gene symbol, 10–41) Treatments from Moyerbrailean *et al* [30]. "1" indicates the gene is differentially expressed in that condition in at least one cell type, 42) "1"

indicates the gene is not differentially expressed in response to any treatment 43) BiT-STARR-seq results. NA indicates SNP was not tested, 0 indicates it was not validated, and 1 indicates validation.
(TXT)

S6 Table. Adaptive introgression outliers. 1000 SNPs had outlier values for all three introgression statistics: RD, Q95, and U20. Columns 1–13 are: 1) Chromosome; 2) Position; 3) rsID; 4) GTEx SNP ID. "NA" indicates the SNP is not an eQTL in GTEx; 5) Modern human allele; 6) Neanderthal allele; 7) Neanderthal allele frequency, 8) Ensembl gene id, 9) Gene symbol, 10) 50 kb genomic window where SNP resides, 11) RD statistic, 12) Q95 statistic, 13) U20 statistic.
(TXT)

S1 Fig. Identification of N-SNPs Overview of how N-SNPs were selected.
(TIF)

S2 Fig. Adaptive introgression statistics (A) RD, **(B)** U20, and **(C)** Q95 statistics for 50kb windows genome-wide. The horizontal line indicates the 5th percentile, and windows with $p < 0.05$ with a N-eQTL are colored red.
(TIF)

S3 Fig. SNP-SELEX analysis of N-SNPs that alter transcription factor binding. Density plots for significant Kolmogorov-Smirnov test results comparing the distribution of PBS values for N-SNPs and non-N-SNPs for individual transcription factors. The blue color represents N-SNPs and the red represents non-N-SNPs.
(TIF)

S4 Fig. B-statistics for N-SNPs validated by BiT-STARR-seq vs N-SNPs not validated by BiT-STARR seq.
(TIF)

S5 Fig. B-statistic comparisons (A) Distribution of B statistics for centiSNPs which are N-SNPs and centiSNPs which are not N-SNPs. **(B)** Distribution of B statistics forSNPs in TF footprints with $> = 10$ N-SNPs in footprint and for SNPs in TF footprints with < 10 N-SNPs in footprint. **(C)** Distribution of B statistics for centiSNPs in 16 TF footprints relevant for glucose metabolism that are enriched for N-centiSNPs and for a control set of centiSNPs in 21 TF footprints relevant for glucose metabolism but that are not enriched for N-centiSNPs. **(D)** Distribution of B statistics for N-centiSNPs and non-centi N-SNPs in footprints for the most enriched TFs (ARNT, SWI5, AML1, AP1, c-Jun, ATF2, Elk1).
(TIF)

S6 Fig. Relationship between number of DEGs and significant AI per treatment. The x-axis represents the number of DEGs per treatment, and the y-axis represents the number of genes in regions with evidence of adaptive introgression from the RD statistics. Each dot represents a different treatment.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Xinjun Zhang, Kirk E. Lohmueller, Roger Pique-Regi, Francesca Luca.

Data curation: Anthony S. Findley, Xinjun Zhang, Yen Lung Lin, Cynthia A. Kalita.

Formal analysis: Anthony S. Findley, Xinjun Zhang, Carly Boye.

Funding acquisition: Roger Pique-Regi, Francesca Luca.

Methodology: Xinjun Zhang, Kirk E. Lohmueller, Roger Pique-Regi, Francesca Luca.

Project administration: Francesca Luca.

Resources: Yen Lung Lin, Cynthia A. Kalita, Luis Barreiro, Kirk E. Lohmueller, Roger Pique-Regi, Francesca Luca.

Supervision: Roger Pique-Regi, Francesca Luca.

Visualization: Anthony S. Findley, Xinjun Zhang, Carly Boye, Roger Pique-Regi, Francesca Luca.

Writing – original draft: Anthony S. Findley, Francesca Luca.

Writing – review & editing: Xinjun Zhang, Carly Boye, Luis Barreiro, Kirk E. Lohmueller, Roger Pique-Regi, Francesca Luca.

References

1. Mallick S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206 (2016). https://doi.org/10.1038/nature18964 PMID: 27654912

2. Vernot B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352, 235–239 (2016). https://doi.org/10.1126/science.aad9416 PMID: 26989198

3. Sankararaman S., Mallick S., Patterson N. & Reich D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol 26, 1241–1247 (2016). https://doi.org/10.1016/j.cub.2016.03.037 PMID: 27032491

4. Green R. E. et al. A draft sequence of the Neandertal genome. Science 328, 710–722 (2010). https://doi.org/10.1126/science.1188021 PMID: 20448178

5. Sankararaman S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507, 354–357 (2014). https://doi.org/10.1038/nature12961 PMID: 24476815

6. Vernot B. & Akey J. M. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343, 1017–1021 (2014). URL http://science.sciencemag.org/. https://doi.org/10.1126/science.1245938 PMID: 24476670

7. Ding Q., Hu Y., Xu S., Wang J. & Jin L. Neanderthal Introgression at Chromosome 3p21.31 Was Under Positive Natural Selection in East Asians. Molecular Biology and Evolution 31, 683–695 (2014). URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst260. PMID: 24336922

8. Racimo F., Marnetto D. & Huerta-Sánchez E. Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. Molecular Biology and Evolution 34, msw216 (2016). URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw216.

9. Rotival M. & Quintana-Murci L. Functional consequences of archaic introgression and their impact on fitness. Genome Biology 2020 21:1 21, 1–4 (2020). URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1920-z. https://doi.org/10.1186/s13059-019-1920-z PMID: 31898502

10. Juric I., Aeschbacher S. & Coop G. The Strength of Selection against Neanderthal Introgression. PLoS Genet 12, e1006340 (2016). https://doi.org/10.1371/journal.pgen.1006340 PMID: 27824859

11. Quach H. & Quintana-Murci L. Living in an adaptive world: Genomic dissection of the genus Homo and its immune response. J Exp Med 214, 877–894 (2017). https://doi.org/10.1084/jem.20161942 PMID: 28351985

12. Vernot B. & Akey J. M. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343, 1017–1021 (2014). https://doi.org/10.1126/science.1245938 PMID: 24476670

13. Nedelec Y. et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell 167, 657–669 (2016). https://doi.org/10.1016/j.cell.2016.09.025 PMID: 27768889

14. Maurano M. T. et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337, 1190–1195 (2012). URL http://www.sciencemag.org/cgi/doi/10.1126/science.1222794. PMID: 22955828

15. Pickrell J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. American Journal of Human Genetics 94, 559–573 (2014). URL http://www.cell.com/ajhg/abstract/S0002-9297(14)00106-2 http://www.ncbi.nlm.nih.gov/pubmed/24702953. arXiv:1311.4843v3. https://doi.org/10.1016/j.ajhg.2014.03.004 PMID: 24702953

16. Beer M. A. & Tavazoie S. Predicting gene expression from sequence. Cell 117, 185–198 (2004). https://doi.org/10.1016/s0092-8674(04)00304-6 PMID: 15084257

17. Lee D. et al. A method to predict the impact of regulatory variants from DNA sequence. Nature Genetics 47, 955–961 (2015). URL http://www.nature.com/doifinder/10.1038/ng.3331. PMID: 26075791

18. Alipanahi B., Delong A., Weirauch M. T. & Frey B. J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature biotechnology 33, 831–8 (2015). URL http://dx.doi.org/10.1038/nbt.3300 http://www.nature.com/doifinder/10.1038/nbt.3300 http://arxiv.org/abs/cs/9605103 http://www.ncbi.nlm.nih.gov/pubmed/26213851.9605103. PMID: 26213851

19. Pique-Regi R. et al. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome research 21, 447–55 (2011). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3044858&tool=pmcentrez&rendertype=abstract. https://doi.org/10.1101/gr.112623.110 PMID: 21106904

20. Boyle A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Research 22, 1790–1797 (2012). URL http://genome.cshlp.org/cgi/doi/10.1101/gr.137323.112. PMID: 22955989

21. Moyerbrailean G. G. A. et al. Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding? PLoS genetics 12, e1005875 (2016). URL http://dx.plos.org/10.1371/journal.pgen.1005875 http://www.ncbi.nlm.nih.gov/pubmed/26901046 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4764260. https://doi.org/10.1371/journal.pgen.1005875 PMID: 26901046

22. authors listed N. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020). https://doi.org/10.1126/science.aaz1776 PMID: 32913098

23. Dannemann M. & Kelso J. The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. Am J Hum Genet 101, 578–589 (2017). https://doi.org/10.1016/j.ajhg.2017.09.010 PMID: 28985494

24. McCoy R. C., Wakefield J. & Akey J. M. Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. Cell 168, 916–927 (2017). https://doi.org/10.1016/j.cell.2017.01.038 PMID: 28235201

25. Nédélec Y. et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. Cell 167, 657–669.e21 (2016). URL http://www.ncbi.nlm.nih.gov/pubmed/27768889. https://doi.org/10.1016/j.cell.2016.09.025 PMID: 27768889

26. Quach H. et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell 167, 643–656 (2016). https://doi.org/10.1016/j.cell.2016.09.024 PMID: 27768888

27. Fan S., Hansen M. E., Lo Y. & Tishkoff S. A. Going global by adapting local: A review of recent human adaptation. Science 354, 54–59 (2016). https://doi.org/10.1126/science.aaf5098 PMID: 27846491

28. Nakajima K. et al. (Myocardial perfusion imaging with 99mTc-SQ30217: application of three-headed SPECT system). Kaku Igaku 28, 127–133 (1991). PMID: 2051650

29. Hancock A. M. et al. Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7, e1001375 (2011). https://doi.org/10.1371/journal.pgen.1001375 PMID: 21533023

30. Moyerbrailean G. A. et al. High-throughput allele-specific expression across 250 environmental conditions. Genome research 26, 1627–1638 (2016). URL http://genome.cshlp.org/lookup/doi/10.1101/gr.209759.116 http://www.ncbi.nlm.nih.gov/pubmed/27934696 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5131815 http://genome.cshlp.org/content/26/12/1627.abstract. PMID: 27934696

31. Prüfer K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49 (2014). URL https://www.nature.com/articles/nature12886. https://doi.org/10.1038/nature12886 PMID: 24352235

32. Prüfer K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358, 655–658 (2017). URL http://science.sciencemag.org/. https://doi.org/10.1126/science.aao1887 PMID: 28982794

33. Mafessoni F. et al. A high-coverage neandertal genome from chagyrskaya cave. Proceedings of the National Academy of Sciences of the United States of America 117, 15132–15136 (2020). URL www.pnas.org/cgi/doi/10.1073/pnas.2004944117. PMID: 32546518

34. Auton A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015). URL http://www.nature.com/doifinder/10.1038/nature15393.15334406. PMID: 26432245

35. Macian F. NFAT proteins: key regulators of T-cell development and function. Nat Rev Immunol 5, 472–484 (2005). https://doi.org/10.1038/nri1632 PMID: 15928679

36. Kyriakis J. M. et al. The stress-activated protein kinase subfamily of c-Jun kinases. Nature 369, 156–160 (1994). https://doi.org/10.1038/369156a0 PMID: 8177321

37. Gupta S., Campbell D., Dérijard B. & Davis R. J. Transcription factor ATF2 regulation by the JNK signal transduction pathway. Science 267, 389–393 (1995). https://doi.org/10.1126/science.7824938 PMID: 7824938

38. Whitmarsh A. J., Shore P., Sharrocks A. D. & Davis R. J. Integration of MAP kinase signal transduction pathways at the serum response element. Science 269, 403–407 (1995). https://doi.org/10.1126/science.7618106 PMID: 7618106

39. Zhang C. et al. MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion. Molecular and Cellular Biology 25, 4969–4976 (2005). URL https://pubmed.ncbi.nlm.nih.gov/15923615/. https://doi.org/10.1128/MCB.25.12.4969-4976.2005 PMID: 15923615

40. Iacovazzo D. et al. MAFA missense mutation causes familial insulinomatosis and diabetes mellitus. Proc Natl Acad Sci U S A 115, 1027–1032 (2018). https://doi.org/10.1073/pnas.1712262115 PMID: 29339498

41. Chiefari E. et al. Transcriptional Regulation of Glucose Metabolism: The Emerging Role of the HMGA1 Chromatin Factor. Front Endocrinol (Lausanne) 9, 357 (2018). https://doi.org/10.3389/fendo.2018.00357 PMID: 30034366

42. Sutherland C., O'Brien R. M. & Granner D. K. Insulin Action Gene Regulation. In Madame Curie Bioscience Database (Landes Bioscience, 2013). URL https://www.ncbi.nlm.nih.gov/books/NBK6471/.

43. Yan J. et al. Systematic analysis of binding of transcription factors to noncoding variants. Nature 2021 591:7848 591, 147–151 (2021). URL https://www.nature.com/articles/s41586-021-03211-0. https://doi.org/10.1038/s41586-021-03211-0 PMID: 33505025

44. E, D. et al. Irf5 deficiency in macrophages promotes beneficial adipose tissue expansion and insulin sensitivity during obesity. Nature medicine 21, 610–618 (2015). URL https://pubmed.ncbi.nlm.nih.gov/25939064/. https://doi.org/10.1038/nm.3829 PMID: 25939064

45. B, R.-L., A, E., B, G. & J, P. The paired homeodomain transcription factor pax-2 is expressed in the endocrine pancreas and transactivates the glucagon gene promoter. The Journal of biological chemistry 275, 32708–32715 (2000). URL https://pubmed.ncbi.nlm.nih.gov/10938089/. https://doi.org/10.1074/jbc.M005704200 PMID: 10938089

46. Juric I., Aeschbacher S. & Coop G. The strength of selection against neanderthal introgression. PLOS Genetics 12, e1006340 (2016). URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006340. PMID: 27824859

47. McVicker G., Gordon D., Davis C. & Green P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet 5, e1000471 (2009). URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19424416. https://doi.org/10.1371/journal.pgen.1000471 PMID: 19424416

48. Kalita C. A. et al. High-throughput characterization of genetic effects on DNA-protein binding and gene transcription. Genome research 28, 1701–1708 (2018). URL http://www.ncbi.nlm.nih.gov/pubmed/30254052 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6211638 https://www.biorxiv.org/content/early/2018/03/21/270991 http://genome.cshlp.org/lookup/doi/10.1101/gr.237354.118. PMID: 30254052

49. Ye C. J. et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. Genome Res 28, 1812–1825 (2018). https://doi.org/10.1101/gr.240390.118 PMID: 30446528

50. Jostins L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119–124 (2012). https://doi.org/10.1038/nature11582 PMID: 23128233

51. Chen X. Y. et al. Brain-selective kinase 2 (BRSK2) phosphorylation on PCTAIRE1 negatively regulates glucose-stimulated insulin secretion in pancreatic β-cells. Journal of Biological Chemistry 287, 30368–30375 (2012). URL /pmc/articles/PMC3436288/ /pmc/articles/PMC3436288/?report = abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3436288/. https://doi.org/10.1074/jbc.M112.375618 PMID: 22798068

52. Acconcia F., Pallottini V. & Marino M. Molecular Mechanisms of Action of BPA. Dose Response 13, 1559325815610582 (2015). https://doi.org/10.1177/1559325815610582 PMID: 26740804

53. Wittkopp P. J., Haerum B. K. & Clark A. G. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet 40, 346–350 (2008). https://doi.org/10.1038/ng.77 PMID: 18278046

54. Maurano M. T., Wang H., Kutyavin T. & Stamatoyannopoulos J. A. Widespread site-dependent buffering of human regulatory polymorphism. PLoS Genet 8, e1002599 (2012). https://doi.org/10.1371/journal.pgen.1002599 PMID: 22457641

55. Spivakov M. et al. Analysis of variation at transcription factor binding sites in Drosophila and humans. Genome Biol 13, R49 (2012). https://doi.org/10.1186/gb-2012-13-9-r49 PMID: 22950968

56. L, C., AB, W., W, F., L, L. & JM, A. Identifying and interpreting apparent neanderthal ancestry in african individuals. Cell 180, 677–687.e16 (2020). URL https://pubmed.ncbi.nlm.nih.gov/32004458/. https://doi.org/10.1016/j.cell.2020.01.012 PMID: 32004458

57. Noonan J. P. Neanderthal genomics and the evolution of modern humans. Genome Res 20, 547–553 (2010). https://doi.org/10.1101/gr.076000.108 PMID: 20439435

58. Melnikov A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30, 271–277 (2012). https://doi.org/10.1038/nbt.2137 PMID: 22371084

59. Kwasnieski J. C., Mogno I., Myers C. A., Corbo J. C. & Cohen B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. Proc Natl Acad Sci U S A 109, 19498–19503 (2012). https://doi.org/10.1073/pnas.1210678109 PMID: 23129659

60. Tewhey R. et al. Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell 165, 1519–1529 (2016). https://doi.org/10.1016/j.cell.2016.04.027 PMID: 27259153

61. Vockley C. M. et al. Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res 25, 1206–1214 (2015). https://doi.org/10.1101/gr.190090.115 PMID: 26084464

62. Arnold C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339, 1074–1077 (2013). https://doi.org/10.1126/science.1232542 PMID: 23328393

63. Rabani M. Massively Parallel Analysis of Regulatory RNA Sequences. Methods Mol Biol 2218, 355–365 (2021). https://doi.org/10.1007/978-1-0716-0970-5_28 PMID: 33606245

64. Rinker D. C. et al. Neanderthal introgression reintroduced functional ancestral alleles lost in Eurasian populations. Nat Ecol Evol 4, 1332–1341 (2020). https://doi.org/10.1038/s41559-020-1261-z PMID: 32719451

65. Knowles D. A. et al. Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. eLife 7 (2018). URL https://elifesciences.org/articles/33480 http://www.ncbi.nlm.nih.gov/pubmed/29737278 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6010343. https://doi.org/10.7554/eLife.33480 PMID: 29737278

66. Manry J. et al. Deciphering the genetic control of gene expression following Mycobacterium leprae antigen stimulation. PLOS Genetics 13, e1006952 (2017). URL https://dx.plos.org/10.1371/journal.pgen.1006952. https://doi.org/10.1371/journal.pgen.1006952 PMID: 28793313

67. Alasoo K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nature Genetics 1 (2018). URL http://www.nature.com/articles/s41588-018-0046-7. https://doi.org/10.1038/s41588-018-0046-7 PMID: 29379200

68. Kim-Hellmuth S. et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nature Communications 8, 1–10 (2017). URL http://immunpop.com/kim/eQTL. https://doi.org/10.1038/s41467-016-0009-6 PMID: 28232747

69. Caliskan M., Baker S. W., Gilad Y. & Ober C. Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. PLOS Genetics 11, e1005111 (2015). URL https://doi.org/10.1371/journal.pgen.1005111 PMID: 25874939

70. Lee M. N. et al. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science 343, 1246980–1246980 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24604203 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4124741 http://www.sciencemag.org/cgi/doi/10.1126/science.1246980. PMID: 24604203

71. Fairfax B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science (New York, N.Y.) 343, 1246949 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24604202. https://doi.org/10.1126/science.1246949 PMID: 24604202

72. Maranville J. C. et al. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. PLoS genetics 7, e1002162 (2011). https://doi.org/10.1371/journal.pgen.1002162 PMID: 21750684

73. Mangravite L. M. et al. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. Nature 502, 377–80 (2013). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3933266&tool=pmcentrez&rendertype=abstract. NIHMS150003. https://doi.org/10.1038/nature12508 PMID: 23995691

74. Barreiro L. B. et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proceedings of the National Academy of Sciences of the United States of America 109, 1204–9 (2012). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3268270&tool=pmcentrez&rendertype=abstract. https://doi.org/10.1073/pnas.1115761109 PMID: 22233810

75. Alasoo K. et al. Genetic effects on promoter usage are highly context-specific and contribute to complex traits. eLife 8 (2019). URL http://www.ncbi.nlm.nih.gov/pubmed/30618377 https://elifesciences.org/articles/41673. https://doi.org/10.7554/eLife.41673 PMID: 30618377

76. Huang Q. Q. et al. Neonatal genetics of gene expression reveal the origins of autoimmune and allergic disease risk. bioRxiv 683086 (2019). URL https://doi.org/10.1101/683086.

77. Maranville J. C., Luca F., Stephens M. & Di Rienzo A. Mapping gene-environment interactions at regulatory polymorphisms: insights into mechanisms of phenotypic variation. Transcription 3, 56–62 (2012). https://doi.org/10.4161/trns.19497 PMID: 22414753

78. Wacker M. & Holick M. F. Sunlight and Vitamin D: A global perspective for health. Dermatoendocrinol 5, 51–108 (2013). https://doi.org/10.4161/derm.24494 PMID: 24494042

79. Lalueza-Fox C. et al. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. Science 318, 1453–1455 (2007). https://doi.org/10.1126/science.1147417 PMID: 17962522

80. Pividori M. et al. PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci Adv 6 (2020). https://doi.org/10.1126/sciadv.aba2083 PMID: 32917697

81. Benjamini Y. & Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing (1995). URL https://www.jstor.org/stable/2346101. https://doi.org/10.1016/0306-9877(95)90228-7 PMID: 8748093

82. Zhang X., Kim B., Lohmueller K. E. & Huerta-Sanchez E. The impact of recessive deleterious variation on signals of adaptive introgression in human populations. Genetics 215, 799–812 (2020). URL https://www.genetics.org/content/215/3/799 https://www.genetics.org/content/215/3/799.abstract. https://doi.org/10.1534/genetics.120.303081 PMID: 32487519