

6. H. Zhao, J. B. French, Y. Fang, S. J. Benkovic, *Chem. Commun. (Camb.)* **49**, 4444–4452 (2013).
7. P. C. Havugimana et al., *Cell* **150**, 1068–1081 (2012).
8. F. J. Schendel, Y. S. Cheng, J. D. Otvos, S. Wehrli, J. Stubbe, *Biochemistry* **27**, 2614–2623 (1988).
9. Q. C. Zhang et al., *Nature* **490**, 556–560 (2012).
10. H. Zhao et al., *J. Biol. Chem.* **290**, 6705–6713 (2015).
11. S. An, M. Kyoung, J. J. Allen, K. M. Shokat, S. J. Benkovic, *J. Biol. Chem.* **285**, 11093–11099 (2010).
12. C. Y. Chan et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1368–1373 (2015).
13. Y. Fang, J. French, H. Zhao, S. Benkovic, *Biotechnol. Genet. Eng. Rev.* **29**, 31–48 (2013).
14. F. Verrier et al., *Nat. Chem. Biol.* **7**, 909–915 (2011).
15. R. Fu et al., *Mol. Genet. Metab.* **114**, 55–61 (2015).
16. J. B. French et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2528–2533 (2013).
17. A. M. Robitaille et al., *Science* **339**, 1320–1323 (2013).
18. F. M. Meyer et al., *Metab. Eng.* **13**, 18–27 (2011).
19. P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 305–310 (2007).
20. C. Jose, N. Bellance, R. Rossignol, *Biochim. Biophys. Acta* **1807**, 552–561 (2011).
21. L. J. Reitzer, B. M. Wice, D. Kennell, *J. Biol. Chem.* **254**, 2669–2676 (1979).
22. B. Huang, S. A. Jones, B. Brandenburg, X. Zhuang, *Nat. Methods* **5**, 1047–1052 (2008).
23. B. Huang, W. Wang, M. Bates, X. Zhuang, *Science* **319**, 810–813 (2008).
24. M. J. Rust, M. Bates, X. Zhuang, *Nat. Methods* **3**, 793–796 (2006).
25. S. A. McKinney, C. S. Murphy, K. L. Hazelwood, M. W. Davidson, L. L. Looger, *Nat. Methods* **6**, 131–133 (2009).
26. Y. Deng et al., *J. Biol. Chem.* **287**, 36201–36207 (2012).
27. S. Wang, J. R. Moffitt, G. T. Dempsey, X. S. Xie, X. Zhuang, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8452–8457 (2014).
28. H. Pelicano, D. S. Martin, R. H. Xu, P. Huang, *Oncogene* **25**, 4633–4646 (2006).
29. N. Masson, P. J. Ratcliffe, *Cancer Metab.* **2**, 3 (2014).
30. A. Ramanathan, S. L. Schreiber, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22229–22232 (2009).
31. Y. Fang, *Front. Pharmacol.* **5**, 52 (2014).
32. A. Birmingham et al., *Nat. Methods* **6**, 569–575 (2009).
33. A. Franceschini et al., *Nucleic Acids Res.* **41** (D1), D808–D815 (2013).
34. I. Ben-Sahra, J. J. Howell, J. M. Asara, B. D. Manning, *Science* **339**, 1323–1328 (2013).
35. K. Düvel et al., *Mol. Cell* **39**, 171–183 (2010).
36. M. Shimobayashi, M. N. Hall, *Nat. Rev. Mol. Cell Biol.* **15**, 155–162 (2014).
37. I. G. Ganley et al., *J. Biol. Chem.* **284**, 12297–12305 (2009).
38. C. H. Jung et al., *Mol. Biol. Cell* **20**, 1992–2003 (2009).
39. M. Laplante, D. M. Sabatini, *J. Cell Sci.* **122**, 3589–3594 (2009).

ACKNOWLEDGMENTS

We thank T. Laremore at the Proteomics and Mass Spectrometry Core Facility of the Huck Institutes of the Life Sciences at The Pennsylvania State University for assistance with data collection and analyses. The Orbitrap mass spectrometer was funded by a grant from the Pennsylvania Department of Health Tobacco Settlement Funds. J.B.F. acknowledges the Canadian Institutes of Health Research for fellowship support. This work was funded by the National Institutes of Health grants NIH GM024129 (S.J.B.) and 1R33EB019785-01 (T.J.H. and S.J.B.) as well as the Howard Hughes Medical Institute (X.Z.). J.B.F., S.A.J., Y.F., X.Z., and S.J.B. designed the experiments; J.B.F., S.A.J., H.D., H.H., A.M.P., C.Y.C., D.K., R.J.P., H.Z., and Y.Z. performed the experiments and analyzed the data; J.B.F., S.A.J., A.M.P., R.J.P., and Y.F. prepared the manuscript; J.B.F., Y.F., X.Z., and S.J.B. directed the research; all authors have reviewed and edited the manuscript. The authors declare that they have no competing interests. Additional data reported in this manuscript are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6274/733/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S3

19 May 2015; accepted 23 December 2015
10.1126/science.aac6054

HUMAN GENOMICS

The phenotypic legacy of admixture between modern humans and Neandertals

Corinne N. Simonti,¹ Benjamin Vernot,² Lisa Bastarache,³ Erwin Bottinger,⁴ David S. Carrell,⁵ Rex L. Chisholm,⁶ David R. Crosslin,^{2,5} Scott J. Hebring,⁷ Gail P. Jarvik,^{2,5} Iftikhar J. Kullo,⁸ Rongling Li,⁹ Jyotishman Pathak,^{10*} Marylyn D. Ritchie,^{11,12} Dan M. Roden,^{1,3,13,14} Shefali S. Verma,¹¹ Gerard Tromp,^{15,16} Jeffrey D. Prato,³ William S. Bush,¹⁷ Joshua M. Akey,^{2†} Joshua C. Denny,^{1,3,13†} John A. Capra^{1,3,18,19†}

Many modern human genomes retain DNA inherited from interbreeding with archaic hominins, such as Neandertals, yet the influence of this admixture on human traits is largely unknown. We analyzed the contribution of common Neandertal variants to over 1000 electronic health record (EHR)–derived phenotypes in ~28,000 adults of European ancestry. We discovered and replicated associations of Neandertal alleles with neurological, psychiatric, immunological, and dermatological phenotypes. Neandertal alleles together explained a significant fraction of the variation in risk for depression and skin lesions resulting from sun exposure (actinic keratosis), and individual Neandertal alleles were significantly associated with specific human phenotypes, including hypercoagulation and tobacco use. Our results establish that archaic admixture influences disease risk in modern humans, provide hypotheses about the effects of hundreds of Neandertal haplotypes, and demonstrate the utility of EHR data in evolutionary analyses.

As anatomically modern human (AMH) groups left Africa and began to spread across Europe and Asia ~60,000 years ago, they encountered other archaic hominins. The fossil record suggests that AMHs and

several archaic hominins overlapped in space and time (1), and genomic analyses of modern and ancient humans and of extinct Neandertals and Denisovans have revealed interbreeding between these groups (2, 3). As a result, the genomes of modern Eurasians contain a small fraction (~1.5 to 4%) of DNA inherited from interbreeding with Neandertals around 50,000 years ago (4–6).

The patterns of surviving Neandertal DNA across modern Eurasian genomes indicate that introgressed Neandertal DNA experienced strong selective pressures. Surviving Neandertal lineages are significantly depleted in several genomic regions, such as on the X chromosome and the q arm of chromosome 7, suggesting that there are deleterious consequences of Neandertal DNA at many loci (5, 6). However, some Neandertal alleles are found at higher than expected frequencies and thus may have provided an evolutionary advantage to AMH populations (5–7). Consistent with this hypothesis, Neandertals are believed to have lived out of Africa long enough to adapt to the climatic, dietary, and pathogenic landscapes found at higher latitudes.

Indeed, isolated introgressed loci have been identified with potential roles in human adaptation (7, 8). Furthermore, recent studies of genomic regions enriched in Neandertal alleles have suggested potential effects on skin and hair phenotypes, lipid metabolism, depression, and other traits (5, 6, 9). However, whether introgressed Neandertal alleles have a significant functional effect on these traits in human populations has

¹Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA. ²Department of Genome Sciences, University of Washington, Seattle, WA, USA. ³Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA. ⁴Mount Sinai School of Medicine, New York, NY, USA. ⁵Department of Medicine (Medical Genetics), University of Washington Medical Center, Seattle, WA, USA. ⁶Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. ⁷Center for Human Genetics, Marshfield Clinic, Marshfield, WI, USA. ⁸Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA. ⁹Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. ¹⁰Division of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. ¹¹Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA. ¹²Biomedical and Translational Informatics, Geisinger Health System, Danville, PA, USA. ¹³Department of Medicine, Vanderbilt University, Nashville, TN, USA. ¹⁴Department of Pharmacology, Vanderbilt University, Nashville, TN, USA. ¹⁵Weis Center for Research, Geisinger Health System, Danville, PA, USA. ¹⁶Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Health Science, Stellenbosch University, Tygerberg, South Africa. ¹⁷Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA. ¹⁸Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA. ¹⁹Center for Quantitative Sciences, Vanderbilt University, Nashville, TN, USA.

*Department of Healthcare Policy and Research, Weill Cornell Medical College, New York, NY, USA. †These authors contributed equally to this work. ‡Corresponding author.

E-mail: tony.capra@vanderbilt.edu

not been established, because of the difficulty of confidently identifying Neandertal-derived DNA and the expense of performing tests for trait association between individuals with and without Neandertal ancestry at specific sites.

We addressed these challenges by integrating the phenotype data present in electronic health records (EHRs) with high-resolution maps of Neandertal haplotypes across individual human genomes (Fig. 1A). We performed a large-scale assessment of the functional effects of DNA inherited from Neandertals on health-related traits in modern populations of European ancestry. In particular, we analyzed genotype and phenotype data from the Electronic Medical Records and Genomics (eMERGE) Network, a consortium that unites EHR systems linked to patient genetic data from nine sites across the United States (10). EHRs contain quantitative and qualitative data on individuals' traits; however, algorithms are required to derive consistent phenotypes

appropriate for use in genetic association testing from these records. For the majority of our analyses, we used a set of 1689 hierarchically related phenotypes (including 1087 leaf phenotypes) defined from the use of International Classification of Diseases (ICD-9) billing codes in the EHRs (11). We analyzed a set of 28,416 adults of European ancestry from across the eMERGE sites who had been genotyped on genome-wide arrays and had sufficient EHR data to define phenotypes. These individuals naturally fell into separate discovery and replication cohorts on the basis of their inclusion in the eMERGE Network Phase 1 (E1; $N = 13,686$ individuals) or Phase 2 (E2; $N = 14,730$ individuals) data releases (12).

To identify Neandertal alleles in the genotyping data available from eMERGE, we used a recent genome-wide map of ~6000 Neandertal haplotypes inferred by computing the S^* statistic (13) and refining putative introgressed haplotypes

by comparing sequenced individuals from the 1000 Genomes (1KG) Project (14) with the Altai Neandertal genome (3, 6). We defined ~135,000 high-confidence "Neandertal single-nucleotide polymorphisms (SNPs)" among the introgressed haplotypes by filtering out SNPs whose frequency significantly differed from the overall Neandertal haplotype frequency and removing haplotypes with fewer than four likely Neandertal-derived SNPs (11). This filtering was necessary to remove variants unlikely to derive from Neandertal admixture.

Neandertal variants have been hypothesized to influence many phenotypes in AMHs, including lipid metabolism, immunity, depression, digestion, hair, and skin, on the basis of the enrichment of Neandertal variants in regions of the genome relevant to these traits (3, 5, 6, 9). Accordingly, we first tested these hypotheses using genome-wide complex trait analysis (GCTA) to estimate the phenotypic risk explained by

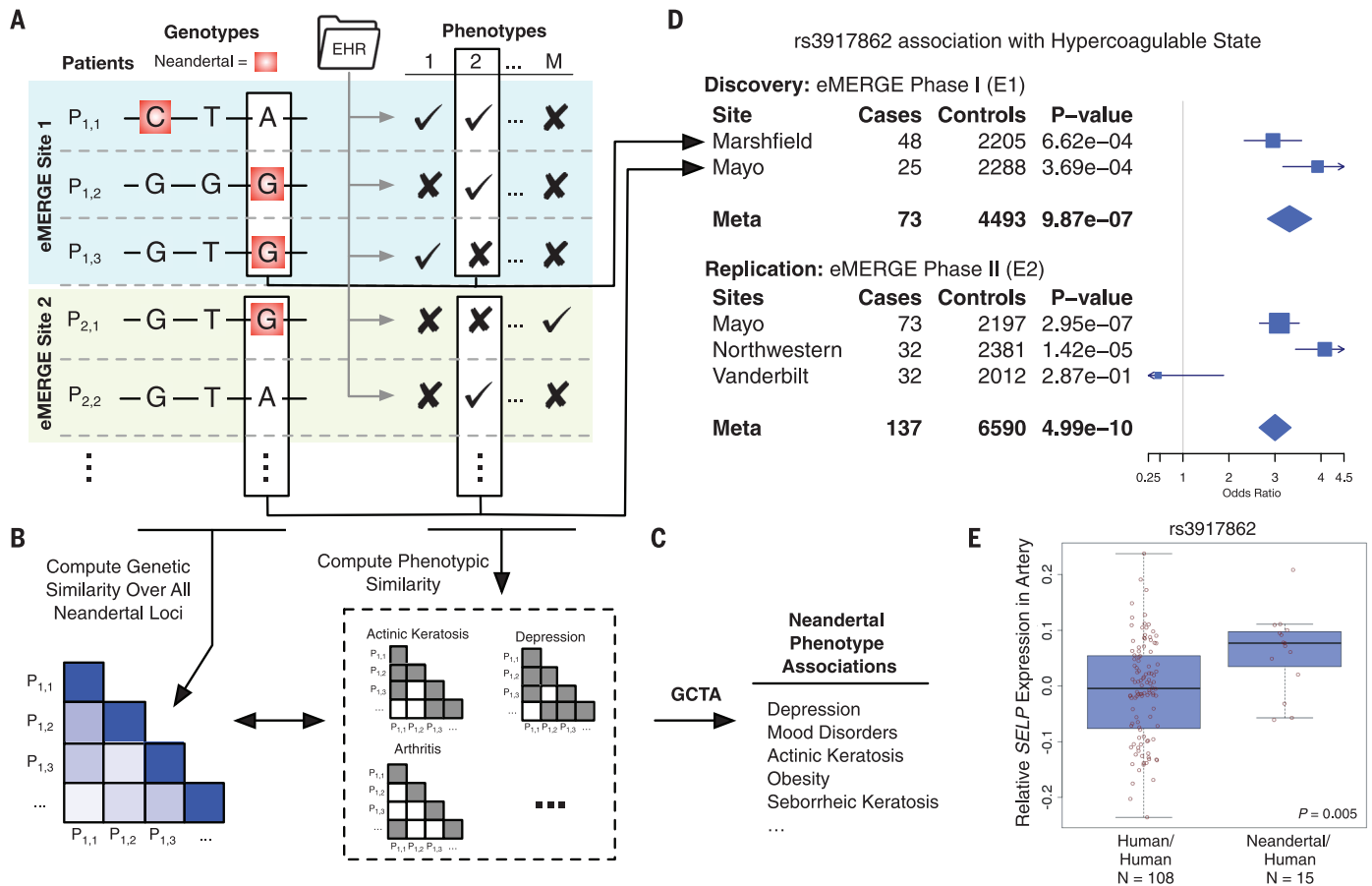


Fig. 1. Analysis of EHRs reveals clinical effects of Neandertal alleles in modern humans.

(A) Thousands of Neandertal alleles were identified in ~28,000 individuals of European ancestry across the eMERGE Network. We derived phenotypes for each individual from data in their EHRs. (B) To test Neandertal alleles in aggregate for phenotype associations, we computed the genetic similarity of all pairs of individuals over 1495 genotyped Neandertal loci and their phenotypic similarity over 46 EHR-derived traits. (C) We estimated the overall variance in risk explained by Neandertal alleles using mixed linear models in GCTA (15) and found that Neandertal alleles explain significant variance in several traits (Table 1). (D) To test individual Neandertal alleles for

trait associations, we performed a discovery meta-analysis across eMERGE E1 sites with sufficient data. We then ran a replication meta-analysis over the independent eMERGE E2 cohort. This approach identified and replicated several associations (Table 2). The example forest plot illustrates the association of Neandertal SNP rs3917862 with hypercoagulable state in each site with >20 cases for the separate discovery and replication analyses. (E) rs3917862 is located in an intron of P-selectin (*SELP*), a gene that mediates leukocyte action at injuries in the early stages of inflammation. The Neandertal allele is significantly associated (linear regression, $P = 0.005$) with increased expression of *SELP* in tibial artery data from GTX.

1495 genotyped common (minor allele frequency > 1%) Neandertal SNPs for a set of 46 high-prevalence phenotypes from the hypothesized categories, using age, sex, and eMERGE site as covariates (Fig. 1, B and C) (15). Neandertal SNPs explained a significant [likelihood ratio test; false discovery rate (FDR) < 0.05 over all phenotype tests] percent of the risk in three traits in the E1 discovery cohort (Table 1): depression (2.03%, $P = 0.0036$), myocardial infarction (1.39%, $P = 0.0026$), and corns and callosities (1.26%, $P = 0.01$). Neandertal SNPs also explained a nominally significant ($P < 0.1$) percent of risk for nine additional traits, including actinic and seborrheic keratosis, coronary atherosclerosis, and obesity (Table 1).

Of the 12 nominally significant associations, 8 were replicated in the independent E2 data set, including actinic keratosis ($P = 0.0059$), mood disorders ($P = 0.018$), depression ($P =$

0.020), obesity ($P = 0.030$), and seborrheic keratosis ($P = 0.045$) at $P < 0.1$ (Table 1; likelihood ratio test). We also tested whether the percent of phenotypic variance explained by Neandertal SNPs remained significant in the context of non-Neandertal SNPs by including an additional genetic relationship matrix (GRM) computed from non-Neandertal SNPs across the rest of the human genome in the mixed linear model (11). Depression ($P = 0.031$), mood disorders ($P = 0.029$), and actinic keratosis ($P = 0.036$) were replicated with these stricter criteria in the independent E2 cohort.

These analyses establish the influence of Neandertal SNPs in concert on the variance in these traits. We estimated individual effects for each SNP by the best linear unbiased predictions (BLUPs); this indicated that a similar number of Neandertal SNPs increased and decreased risk for each associated phenotype (table S1) (11). To

gain insight into the loci driving these associations, we analyzed the genomic distribution of the 10% of SNPs with the highest and lowest BLUPs for actinic keratosis and depression. We found enrichment (FDR < 0.05; hypergeometric test) for many functional annotations: most notably, keratinocyte differentiation and several immune functions for actinic keratosis and regions involved in neurological diseases, cell migration, and circadian clock genes for depression (fig. S1 and table S2) (11).

The significant replicated association of Neandertal SNPs with mood disorders, in particular depression, is intriguing because Neandertal alleles are enriched near genes associated with long-term depression (5), and human-Neandertal DNA and methylation differences have been hypothesized to influence neurological and psychiatric phenotypes (16, 17). Depression risk in modern human populations is influenced by sunlight exposure (18), which differs between high and low latitudes, and we found enrichment of circadian clock genes near the Neandertal alleles that contribute most to this association (11). The replicated nominal association of Neandertal SNPs with actinic keratosis (precancerous scaly skin lesions) further links introgressed alleles in AMHs to a phenotype directly related to sun exposure. It also suggests that the signatures of adaptive introgression and strong enrichment of Neandertal alleles near genes associated with keratin filament formation (5) and keratinocytes (6) reflect the influence of Neandertal alleles on a modern human phenotype. However, further genetic analyses are necessary to resolve the differences in the strength of this association between E1 and E2. These results establish the impact of Neandertal DNA on diseases in AMHs that involve traits potentially influenced by environmental differences experienced by non-African populations.

GCTA quantifies the overall influence of Neandertal SNPs together on traits in AMHs. To identify individual Neandertal loci associated with AMH phenotypes and potentially discover additional biological systems influenced by Neandertal admixture, we performed a phenome-wide association study (PheWAS) of these 1495 Neandertal SNPs with 1152 EHR-derived phenotypes with at least 20 cases in at least one site (Fig. 1D). PheWAS allows for large-scale characterization of the effects of variants of interest (19). We carried out two meta-analyses across the eMERGE Network sites: one over the discovery cohort and one over the replication cohort. We focus on the meta-analyses here (Table 2 and table S3), but a pooled analysis using the eMERGE site as a covariate produced largely consistent results (table S4).

Four Neandertal SNP-phenotype associations passed a locus-wise Bonferroni corrected significance threshold ($P = 3.3 \times 10^{-5}$) in the E1 meta-analysis and were replicated ($P < 0.05$) with the same direction of effect in the independent E2 meta-analysis (Table 2). The strongest signal was a Neandertal SNP (rs3917862, 6.5% European (EUR) 1KG frequency) in an intron of P-selectin

Table 1. Neandertal alleles explain risk for human clinical traits. The eight traits for which Neandertal alleles explained a nominally significant proportion of variance in risk in both the E1 discovery and E2 replication analyses are listed (GCTA, $P < 0.1$). The depression association remained significant after controlling the false discovery rate at 5%. The Neandertal associations with actinic keratosis, mood disorders, and depression were also maintained in a two-GRM model that considered the risk explained by non-Neandertal variants. Phenotypes are sorted by their E2 P value.

Phenotype	Discovery (E1)		Replication (E2)		Replication (E2; two-GRM)	
	Risk explained	P	Risk explained	P	Risk explained	P
Actinic keratosis	0.64%	0.066	3.37%	0.0059	2.49%	0.036
Mood disorders	1.11%	0.0091	0.75%	0.018	0.68%	0.029
Depression	2.03%	0.0023	1.15%	0.020	1.06%	0.031
Obesity	0.59%	0.048	1.23%	0.030	0.39%	0.27
Seborrheic keratosis	0.77%	0.038	0.61%	0.045	0.41%	0.13
Overweight	0.60%	0.037	0.53%	0.052	0.23%	0.24
Acute upper respiratory infections	0.70%	0.043	0.56%	0.062	0.34%	0.18
Coronary atherosclerosis	0.68%	0.04	0.42%	0.098	0.34%	0.15

Table 2. Individual Neandertal SNPs with significant replicating phenotype associations. Four locus-wise Bonferroni significant Neandertal SNP-phenotype associations replicated (with a fixed effect $P < 0.05$ and consistent direction of effect). Nominally significant replicating results can be found in table S3 and in the PheWAS Catalog (<https://phewas.mc.vanderbilt.edu/neanderthal>). Chr, chromosome.

Phenotype	Chr:position (hg19)	SNP	Flanking gene(s)	Discovery		Replication	
				Odds ratio	P	Odds ratio	P
Hypercoagulable state	1:169593113	rs3917862	SELP	3.32	9.9×10^{-7}	3.00	5.0×10^{-10}
Protein-calorie malnutrition	1:234099819	rs12049593	SLC35F3	1.77	2.0×10^{-6}	1.63	5.5×10^{-5}
Symptoms involving urinary system	11:3867350	rs11030043	RHOG, STIM1	1.76	7.4×10^{-6}	1.65	4.3×10^{-2}
Tobacco use disorder	3:10962315	rs901033	SLC6A11	2.19	1.7×10^{-5}	1.75	7.9×10^{-4}

(*SELP*) that was significantly associated with hypercoagulable state in both E1 and E2 (Table 2; Fig. 1D). This haplotype contains several genes directly involved in blood coagulation and inflammation, most notably *SELP*, which encodes a cell adhesion protein expressed on the surface of endothelial cells and platelets that recruits leukocytes to injuries during inflammation. The gene encoding factor V (*F5*), a coagulation cofactor associated with several coagulation defects, is located ~37 kilobases downstream. The Neandertal haplotype overlaps histone modifications suggestive of gene-regulatory activity in blood cells and vein epithelial cells (fig. S2). Using data from the Genotype-Tissue Expression (GTEx) Project (20), we found indications that the Neandertal allele at rs3917862 significantly increased the expression of *SELP* ($P = 0.005$) and *F5* ($P = 0.05$) in arteries (Fig. 1E and fig. S2). The association may be influenced by the *F5* Leiden (*F5L*) mutation associated with hypercoagulable state; however, the Neandertal allele appears to have an additional influence on risk (11). The Neandertal SNP is in low linkage disequilibrium (LD) with *F5L* ($r^2 = 0.07$, $D' = 0.42$), and increases risk for venous thromboembolism, beyond the risk associated with *F5L* (21). Furthermore, manual review of the EHRs for 16 hypercoagulable state cases revealed a diverse set of causes, and only 4 out of the 11 individuals tested for *F5L* had the mutation. Due to the direct interaction of coagulation factors with pathogens, these genes have been common targets of positive selection across vertebrate evolution, and *F5* has experienced positive selection in primates (22). Thus, it is possible that this Neandertal haplotype and the associated hypercoagulability provided an advantage in early AMHs outside of Africa.

The second replicating association was a SNP (rs12049593, 5.0% EUR frequency) in an intron of *SLC35F3*, a putative thiamine transporter, which associates with protein-calorie malnutrition. Thi-

amine is crucial to carbohydrate metabolism for all cells, particularly those with increased energy requirements (23). Variants in high LD with this SNP ($r^2 > 0.8$, $D' = 1$) are found in regions bearing enhancer histone marks in the gastrointestinal (GI) tract, brain, and other tissues. Decreased expression of this transporter in the brain or GI tract could exacerbate malnutrition or its symptoms. It is possible that new dietary pressures may have caused changes in carbohydrate metabolism to be beneficial in early human migrants out of Africa; indeed, there is evidence suggesting that Neandertal introgression probably influenced lipid catabolism in Europeans (9). More recently, the reduction of thiamine present in foods from the grain-refining process, as well as an increased intake of simple carbohydrates, make this a potentially harmful allele, because it could reduce thiamine availability although modern diets increase demand.

Another Neandertal SNP (rs11030043, 9.0% EUR frequency) is upstream of *stromal interaction molecule 1* (*STIM1*) and is associated with a phenotype encompassing incontinence, bladder pain, and urinary tract disorders (fig. S4A). *STIM1* is a ubiquitously expressed gene involved in intracellular calcium signaling. Variants in high LD with the Neandertal SNP are found in regions bearing enhancer histone marks and DNase I hypersensitive sites in the brain. Because of this, we examined whether this SNP was associated with gene expression levels in brain tissues in GTEx. The Neandertal allele is associated with significantly decreased expression of *STIM1* in the caudate basal ganglia ($P = 0.02$; fig. S4B), a region of the brain connected to bladder dysfunction, particularly in those with neurological conditions such as Parkinson's (24).

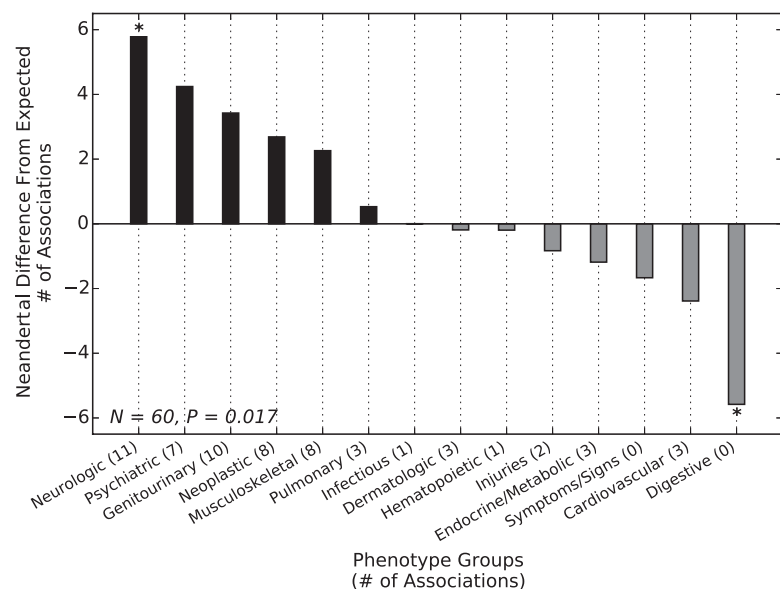
The last replicated association was between rs901033 (0.5% EUR frequency) and tobacco use disorder. This SNP is in an intron of *SLC6A11*, a

solute carrier family neurotransmitter transporter that is responsible for reuptake of the neurotransmitter γ -aminobutyric acid (GABA). Nicotine addiction disrupts GABAergic signaling in the brain and reduces expression of *SLC6A11* (25). This is the second Neandertal SNP to be associated with smoking risk (5).

To test whether Neandertal SNPs were enriched for association with specific classes of phenotypes, we compared the distribution of replicating phenotype associations for a set of 1056 LD-pruned ($r^2 < 0.5$) Neandertal SNPs (table S5) with the associations found in five allele frequency matched non-Neandertal SNP sets (consisting of a total of 5280 SNPs) at a relaxed PheWAS discovery threshold ($P < 0.001$) (11). Overall, the Neandertal SNPs appear to influence significantly different classes of phenotypes (chi-squared test, $P = 0.017$; Fig. 2). In particular, Neandertal SNPs were consistently associated with more neurological and psychiatric phenotypes and fewer digestive phenotypes (Fig. 2) (11).

Given the enrichment for associations with psychiatric and neurological phenotypes, we tested whether Neandertal SNPs were significantly associated with changes in gene expression in previous expression quantitative trait loci (eQTL) analyses of the cerebellum and temporal cortex. Twenty-nine Neandertal SNPs were significant brain *cis*-eQTL in the cerebellum or temporal cortex (FDR < 0.05) (11). This represents significant enrichment for brain eQTL among Neandertal SNPs as compared to the non-Neandertal control SNPs (one-tailed binomial test; $P = 1.68 \times 10^{-4}$ for the cerebellum and $P = 3.49 \times 10^{-5}$ for the temporal cortex) (11). Taken together, the influence of Neandertal SNPs on depression risk (Table 1), the association of individual Neandertal SNPs with diseases with a neurological basis (Table 2), the enrichment for nominal associations with psychiatric and neurological phenotypes (Fig. 2), and the enrichment for

Fig. 2. Neandertal SNPs associate with different phenotypes than matched non-Neandertal SNPs. Each bar gives the difference between the number of replicated Neandertal SNP associations with a phenotype group (at a relaxed discovery threshold of $P < 0.001$) and the number expected from a PheWAS over five sets of non-Neandertal sites matched to the allele frequency of tested Neandertal SNPs. The phenotype distributions were significantly different (chi-squared test, $P = 0.017$), with more Neandertal SNPs associated with neurological and psychiatric phenotypes than expected and fewer digestive phenotypes. The enrichment and depletion were consistent across all five matched non-Neandertal sets (* indicates $P < 0.05$ for all five comparisons; binomial test) (11).



brain eQTL in Neandertal SNPs suggest that Neandertal introgression influenced AMH brain phenotypes.

Our approach establishes a new method for understanding the phenotypic legacy of admixture between AMHs and archaic hominins. Using a large clinical cohort, we discovered functional associations between Neandertal alleles and AMH traits, influencing the skin, the immune system, depression, addiction, and metabolism. Furthermore, several lines of evidence suggest enrichment for associations between Neandertal alleles and neurological and psychiatric phenotypes, as well as the importance of differences in sun exposure between high and low latitudes. It is possible that some Neandertal alleles provided a benefit in early AMH populations as they moved out of Africa, but have become detrimental in modern Western environments.

EHR data, paired with DNA sequencing, hold promise for characterizing the phenotypic impact of regions identified through evolutionary analyses. However, there are currently limitations to this approach. It is difficult to extract nonclinical phenotypes from EHRs, and we were not able to analyze all Neandertal haplotypes due to the limited coverage of the available genotyping data. Nonetheless, as EHRs are increasingly linked to whole-genome sequence data and more sophisticated methods are developed for extracting phenotypes from the rich data stored in EHRs, we anticipate further insights into the functional effects of archaic introgression.

REFERENCES AND NOTES

1. T. Higham *et al.*, *Nature* **512**, 306–309 (2014).
2. R. E. Green *et al.*, *Science* **328**, 710–722 (2010).
3. K. Prüfer *et al.*, *Nature* **505**, 43–49 (2014).
4. J. D. Wall *et al.*, *Genetics* **194**, 199–209 (2013).
5. S. Sankararaman *et al.*, *Nature* **507**, 354–357 (2014).
6. B. Vernot, J. M. Akey, *Science* **343**, 1017–1021 (2014).
7. L. Abi-Rached *et al.*, *Science* **334**, 89–94 (2011).
8. E. Huerta-Sánchez *et al.*, *Nature* **512**, 194–197 (2014).
9. E. E. Khrameeva *et al.*, *Nat. Commun.* **5**, 3584 (2014).
10. A. N. Kho *et al.*, *Sci. Transl. Med.* **3**, 79re1 (2011).
11. L. F. Schroeder, E. Robilotti, L. R. Peterson, N. Banaei, D. W. Dowdy, *J. Clin. Microbiol.* **52**, 489–496 (2014).
12. O. Gottesman *et al.*, *Genet. Med.* **15**, 761–771 (2013).
13. V. Plagnol, J. D. Wall, *PLoS Genet.* **2**, e105 (2006).
14. G. R. Abecasis *et al.*, *Nature* **491**, 56–65 (2012).
15. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **88**, 76–82 (2011).
16. D. Gokhman *et al.*, *Science* **344**, 523–527 (2014).
17. N. Oksenberg, L. Stevion, J. D. Wall, N. Ahituv, *PLoS Genet.* **9**, e1003221 (2013).
18. R. N. Golden *et al.*, *Am. J. Psychiatry* **162**, 656–662 (2005).
19. J. C. Denny *et al.*, *Nat. Biotechnol.* **31**, 1102–1110 (2013).
20. The GTEx Consortium, *Science* **348**, 648–660 (2015).
21. J. A. Heit *et al.*, *J. Thromb. Haemost.* **10**, 1521–1531 (2012).
22. P. M. Rallapalli, C. A. Orengo, R. A. Studer, S. J. Perkins, *Mol. Biol. Evol.* **31**, 3040–3056 (2014).
23. P. R. Martin, C. K. Singleton, S. Hiller-Sturmhöfel, *Alcohol Res. Health* **27**, 134–142 (2003).
24. J. H. Pazo, J. E. Belforte, *Cell. Mol. Neurobiol.* **22**, 645–654 (2002).
25. C. Pickering, V. Bergenheim, H. B. Schiöth, M. Ericson, *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32**, 1521–1526 (2008).

ACKNOWLEDGMENTS

The data reported in this paper are available in the supplementary materials and in the Neandertal PheWAS Catalog (<https://phewas.mc.vanderbilt.edu/neanderthal>). The raw genotype and phenotype data used in this study are subject to the data use and availability requirements of the eMERGE network outlined at <https://emerge.mc.vanderbilt.edu/collaborate>. We thank A. Fish, J. Hall, D. Mortlock, D. Samuels, M. Sivley, and P. Wu for helpful discussions. C.N.S. was supported by NIH grant 5T32EY021453 to Vanderbilt University and a pilot project award from the Vanderbilt Center for Quantitative Sciences. The eMERGE Network was initiated and funded by the National Human Genome Research Institute through the following grants: U01HG004438 to Johns Hopkins University; U01HG004610 and U01HG008657 to Group Health Cooperative and University of Washington, Seattle; U01HG004608 and 1K22LM011938 to the Marshfield Clinic; U01HG04599 and U01HG006379 to the Mayo Clinic; U01HG004609 and U01HG006388 to Northwestern University; U01HG04603 and U01HG006378 to Vanderbilt University; U01HG006385 to the Coordinating Center of the eMERGE Network; U01HG006382 to the Geisinger Clinic; and U01HG006380 to the Mount Sinai School of Medicine. J.M.A. is a paid consultant for Glenview Capital and provides advice on next-generation sequencing technologies. PheWAS method development was supported by grant R01LM010685. J.M.A. was supported in part by NIH grant R01GM110068. S.J.H. was supported in part by NIH grant 1R01GM114128. The authors report no conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6274/737/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S5
References (26–51)

10 August 2015; accepted 8 January 2016
10.1126/science.aad2149

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of February 11, 2016):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

</content/351/6274/737.full.html>

Supporting Online Material can be found at:

</content/suppl/2016/02/10/351.6274.737.DC1.html>

This article **cites 46 articles**, 15 of which can be accessed free:

</content/351/6274/737.full.html#ref-list-1>

This article appears in the following **subject collections**:

Evolution

</cgi/collection/evolution>