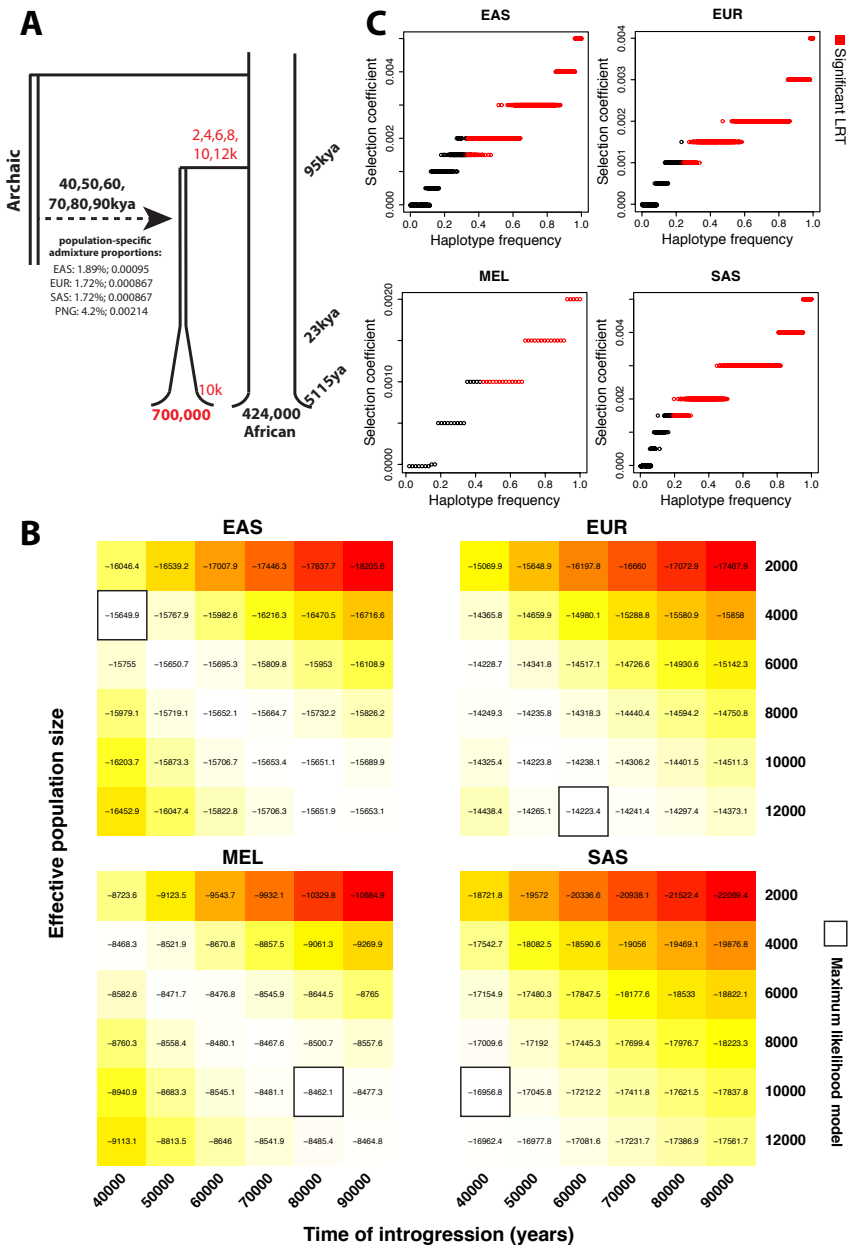

Figure S1, refers to figure 1. Coalescent simulations. A. Demographic model for simulations is shown**.** The base demographic model for the simulation framework is depicted with the grid of parameters used. Text in red indicates effective population sizes in the introgressed population. **B.** Likelihoods for each of the 36 demographic models tested is depicted for each population. Color corresponds to the likelihoods, with more likely models shown in lighter colors. The maximum likelihood model used in subsequent simulations and FDR is highlighted with a black box. **C.** The maximum likelihood selection coefficient is depicted for all frequencies of introgressed haplotypes in each of the four populations studied. Frequencies in which the maximum likelihood selection coefficient is a significantly better fit than the mildly deleterious model are shown in red.

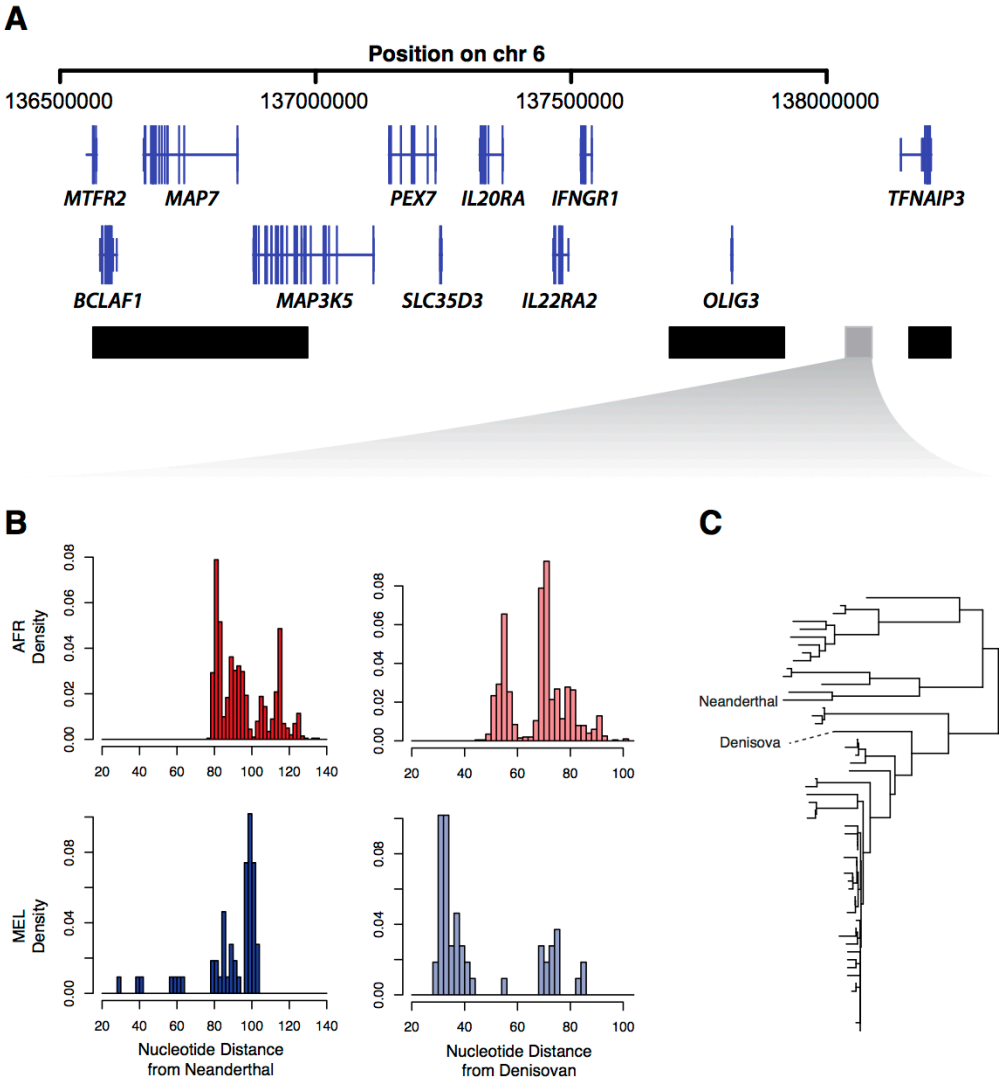

Figure S2, refers to figure 2. A region with both Neandertal and Denisovan ancestry. A. A schematic of the region harboring 4 high frequency Melanesian regions that segregate both Neandertal and Denisovan sequence. The bars indicate the distinct regions, and the grey bar indicates the region that is further characterized in the next panels. **B.** The distribution of absolute genetic distances from Neandertal (left column) and Denisovan (right column) are shown for Africans (top row) and Melanesians (bottom row). **C.** A neighbor-joining tree constructed with sequence from Melanesians, Densiovan, and Neandertal is shown.

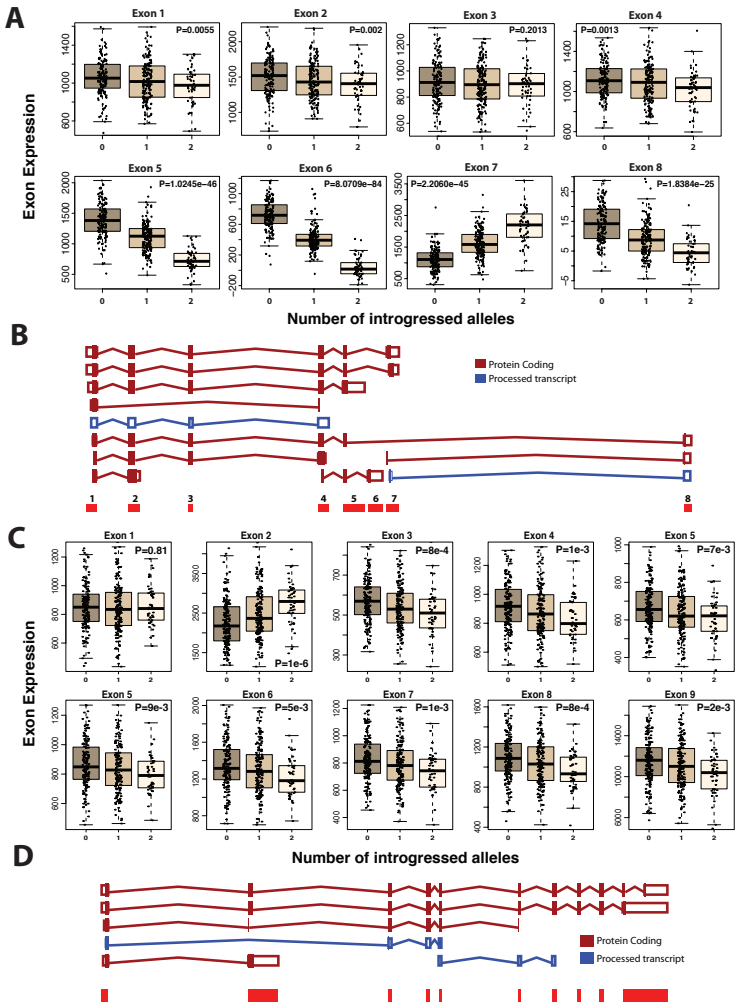

Figure S3. refers to figure 3. Neandertal haplotype association with *OAS1/OAS2* exon expression. A. Gene expression for each exon of *OAS1* is shown stratified by the number of Neandertal alleles each sample has. Data is from Geuvadis project LCLs. **B.** Schematics of observed *OAS1* isoforms are shown. Boxes indicate which exons are included in each transcript. **C.** Gene expression for each exon of *OAS2* is shown stratified by the number of Neandertal alleles each sample has. Data is from Geuvadis project LCLs. **D.** Schematics of observed *OAS2* isoforms are shown. Boxes indicate which exons are included in each transcript.

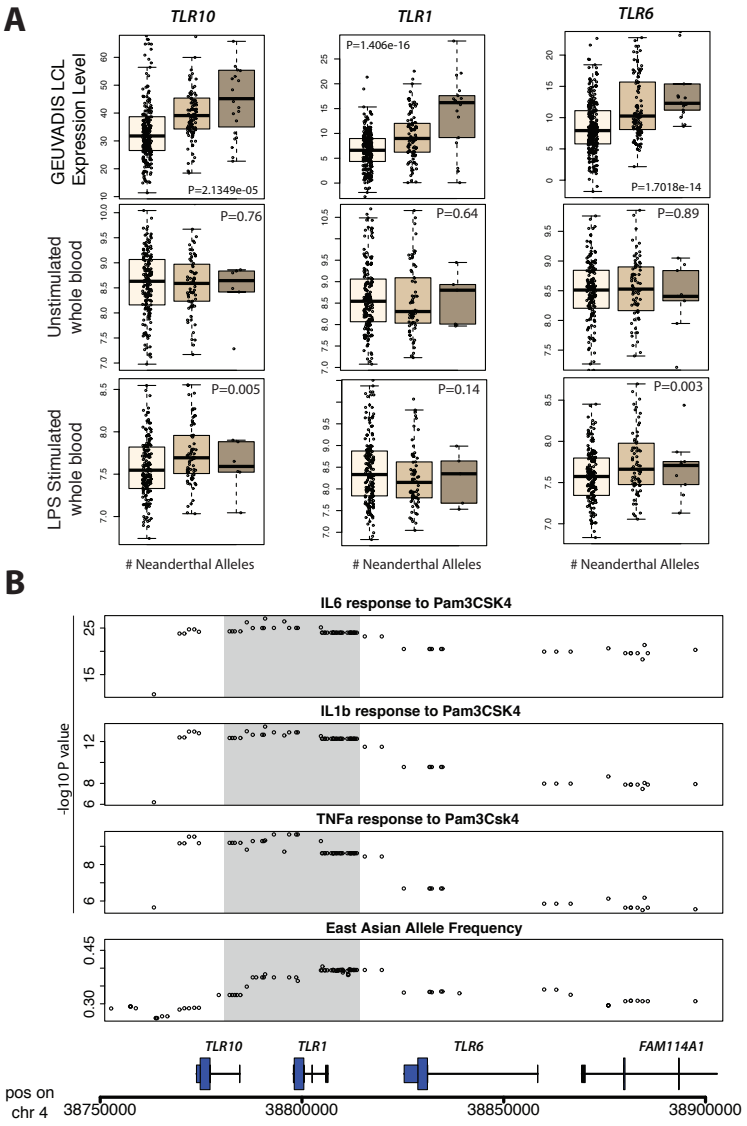

Figure S4, refers to figure 4. Neandertal haplotype encompassing *TlR1/6/10* is associated with gene expression and cellular response to Pam3CSK4. A. Gene expression for *TLR10*/1/6 is shown stratified by the number of Neandertal alleles each sample has. Rows consist of data from a single cell type. P values are indicated for each plot. **B.** The –log10(p value) of association with interleukin response to *TLR1* stimulation by PAM3CSK4 is shown for all variants on the TLR haplotype. On the bottom, allele frequency in East Asians of each variant is shown. They grey box highlights the region of maximal p value and allele frequency.

Supplemental Experimental Procedures

Coalescent simulations

We began by simulating introgression at a single locus with MSMS [S1] given a base demographic model [S2] as follows: a) Ancestral N_e of 10000, b) splitting of archaic and modern human lineages 700,000 years ago, with an archaic N_e of 1500, and modern human N_e of 10000, c) Splitting of Africans and non-Africans at 95,000 years ago, d) a single 500 year pulse of archaic admixture in to the out-of-Africa population, e) population growth in the out of African population starting at 23,000 years ago to an N_e of 10000 at 5,115 years ago, f) exponential population growth starting at 5,115 years ago to a final N_e of 700,000 in the out of Africa population, and 424,000 in the African population. Within this base model we varied several parameters as follows: a) The out of Africa N_e ranged across a grid of N_e = [2000, 4000, 6000, 8000, 10000, 12000], b) The time of introgression ranged across a grid T_I = [40kya, 50kya, 60kya, 70kya, 80kya, 90kya], c) The archaic migration rate in to modern humans varied for each population, with 0.00095 for East Asians, 0.000867 for Europeans, 0.00214 for Melanesians, and 0.000867 for South Asians, d) The number of chromosomes sampled for each population matched the number of chromosomes sampled in empirical data: 1008 for East Asians, 1006 for Europeans, 54 for Melanesians, and 978 for South Asians e) The archaic population harbored a mildly deleterious (s=-0.000021) variant at the time of introgression. The negative selection coefficient was determined based on estimates of the average deleteriousness of introgressed alleles [S3]. Briefly, we conservatively use $-3\text{x}10^{-8}$ as the strength of selection per introgressed exonic base, 70kb as the length of the average introgressed haplotype, and 1% as the fraction of exonic bases in the genome. Then the average selection against an introgressed haplotype is:

$$3\times10^{-8} \times 70000 \times 0.01 = 2.1\times10^{-5}$$

The frequency of this variant in the archaic population is described at the end of this section. This base demographic model is depicted in Figure S1A. We ran 5 million simulations for each of the 36 distinct demographic models and recorded the frequency of introgressed chromosomes in the final Out-of-Africa population, removing all simulations where the introgressed frequency is zero. We then identified the model with the best fit to the observed data by maximizing the following equation:

$$P(D|X) = \prod_{i}^{N} P(D_i|X)$$

Here $P(D_i|X)$ is simply the proportion of simulations from a particular model X at the frequency of the i^{th} haplotype in the observed data (D), where there are N total haplotypes. We noticed that the observed data contains an excess of low frequency haplotypes, likely due to the difficulties of aggregating low frequency tag SNPs into coherent haplotypes. We thus conservatively exclude all haplotypes below 2% frequency from our calculations. Importantly, we note that the chosen parameters may not represent the true demographic history of each population, but are instead the best model given the biases inherent to our dataset, thus allowing us to simulate data that most closely resembles our dataset. Likelihoods for each of the 36 models are depicted for each population in Figure S1B. We then used the simulations from the maximum likelihood demographic model to calculate false discovery rates in the observed data for a given haplotype frequency threshold, using the equation:

$$FDR = \frac{S(\frac{N_O}{N_S})}{O}$$

Where N_O is the total number of observed haplotypes, N_S is the total number of simulated haplotypes, and S and O are the number of simulated and observed haplotypes above a given frequency threshold, respectively.

In the second step, we run additional simulations using the maximum likelihood demographic model, while varying the selection coefficient in across a grid of s = [-0.000021, 0, 0.0005, 0.00075, 0.001, 0.0015, 0.002, 0.003, 0.004, 0.005]. Information on the frequency of the selected allele at the time of introgression can be found in the note at the bottom of this section. We ran 25 million simulations for each selection coefficient, again removing replicates where the introgression frequency is zero in the final population. We next constructed a likelihood landscape for each possible allele frequency by determining the proportion of simulations at that frequency for each selection coefficient. Finally, we conducted a likelihood ratio test between a model with the selection coefficient at the highest likelihood and the "null" model of a mildly deleterious selection coefficient (s=-

0.000021). To determine which tests were significant, we compared test statistics to a chi-square null distribution with one degree of freedom.

Note, in order to accurately simulate the frequency of the selected allele in the archaic population at the time of introgression, we ran an initial set of simulations in the archaic population in which a selected allele arose at a frequency of $1/N_e$, randomly between the time of introgression and 500kya. We then recorded the frequency of the selected allele at the time of introgression, discarding simulations in which the frequency went to zero. For subsequent simulations, we randomly sampled the starting frequencies from these distributions.

Estimates of true positives are robust to FDR threshold

At a FDR threshold of 50%, our simulations suggest the number of true positives is on the order of 10-20 per population. This estimate is generally robust to varying the FDR threshold. For example, at a FDR threshold of 30%, there are 6, 29, 7, and 19 loci significant and thus 4, 19, 5, and 13 true positives in EAS, EUR, MEL, and SAS, respectively. Similarly, at a FDR threshold of 70%, there are 86, 181, 112 AND 107 loci significant and thus 26, 56, 35, and 32 true positives in EAS, EUR, MEL, and SAS, respectively.

Robustness of inferences to assumptions of purifying selection on introgressed sequences

In addition to simulations where introgressed sequence are assumed to be on average mildly deleterious, we also repeated all analyses assuming introgressed archaic sequences were strictly neutral (s = 0). Of the 126 putative archaic haplotypes deemed significant in simulations with purifying selection, 121 were also called significant in simulations where s = 0.

Gene Ontology Enrichments

To obtain a list of genes proximal to, or encompassed by, all high frequency introgressed haplotypes, we downloaded the complete set of UCSC genes from the UCSC genome browser [S4] (https://genome.ucsc.edu/) on 10/14/2014 and took the unique set of the nearest gene to each SNP on a haplotype. We input this list as the “foreground” set of

genes on the WebGestalt gene ontology browser [S5] (http://bioinfo.vanderbilt.edu/webgestalt/), with the following parameters: Enrichment Analysis: GO Analysis; Reference set: hsapiens_genome; Statistical Method: Hypergeometric; Multiple Test Adjustment: BH; Significance Level: 0.05; Minimum Number of Genes: 10.

Coalescent approach to estimate probability of ILS

To provide additional confidence that an archaic haplotype was the result of introgression and not incomplete lineage sorting (ILS) we developed a coalescent likelihood model. Specifically, assume a haplotype spans L bases and contains K mutations relative to the archaic reference sequence. In the following description, we focus on ILS with respect to Neandertals given that most putative adaptively introgressed sequences are Neandertal in origin. To compute the joint probability of L and K, condition on the fragment coalescing with Neandertal time $T = t$ (in coalescent time units). Then, the distribution of L is exponential with rate ρt, where $\rho = 4Nr$, N is the effective population size and r is the per-base pair recombination rate. Given the length $L = l$, the number of differences K is Poisson distributed with mean θlt, where $\theta = 4N\mu$, and μ is the per-base pair mutation rate. Thus, given $T = t$, the joint distribution of K and L is

$$f_{K,L|T}(k, l|t) = \frac{(\theta lt)^k}{k!} e^{-\theta lt} \rho t e^{-\rho t l}.$$

Next, we integrate over the distribution of coalescence times. Assuming that coalescence can begin at some time t^* (e.g. the time of introgression, or the population split time), the distribution of coalescence times is simply a shifted exponential distribution with rate 1. Therefore,

$$f_{K,L}(k, l; t^*) = \int_{t^*}^{\infty} f_{K,L|T}(k, l|t) e^{-(t-t^*)} dt$$

$$= \frac{(\theta l)^k}{k!} \frac{\rho \Gamma(2 + k, \big(1 + l(\rho + \theta)\big)t^*)}{(1 + l(\rho + \theta))^{k+2}} e^{t^*}.$$

We now use Bayes' theorem to compute the probability that a fragment is introgressed by supposing that the introgression proportion is p, and that introgression happened at time t_{GF} in the past while population divergence occurred at time t_D. Implicitly, $t_{GF} < t_D$. So,

$$P(introgressed \mid K, L) = \frac{f_{K,L(k,l;t_{GF})p}}{f_{K,L(k,l;t_{GF})p} + f_{K,L(k,l;t_D)(1-p)}}.$$

We used this equation to calculate $P(introgressed \mid 11{,}29747)$ for the *OCA2* locus. To ensure that our estimate is robust to different demographic models and mutation rates, we randomly sampled parameters to obtain 100 distinct parameters sets and calculated $P(introgressed \mid 11{,}29747)$ with each set. Sampling was from the following uniform distributions: N=[1000-10000]; μ =[5x10^{-9}-5x10^{-8}]; t_{GF} =[40kya-80kya]; t_D =[400kya-600kya], with fixed r=3.19x10^{-8} (the average recombination rate in the region [S4]).

RNA-seq normalization

For GTEx data, we began with un-normalized gene read counts provided with the version 4 of GTEx pilot data and applied a processing pipeline similar to that of the GTEx Project. We removed all samples with <10,000,000 mapped reads and summed expression values for technical replicates. We removed 32 individuals identified as non-Europeans by examining a PCA plot of sample genotypes. We removed any genes in which fewer than 10 individuals had at least 5 reads. We also removed samples that were two or more standard deviations below the mean D statistic [S6], a measure of overall expression similarity to other samples. We then used the R DESeq package [S7] to normalize samples based on library size, and $\log_2$ transformed these data. Finally, we performed outlier detection by mapping expression values for each gene to a standard normal distribution. We ran PEER[8] with these normalized data, using the first two genotype principal components and sex as covariates in order to identify hidden sources of expression heterogeneity. We used the first 15 PEER factors and known covariates in a linear model with the expression data using the lm() function in R [S9], and used the residual expression data as input for association testing.

For the Geuvadis Project data, we downloaded pre-normalized gene RPKM expression data from http://www.ebi.ac.uk/arrayexpress/files/E-GEUV-

1/analysis_results/GD462.GeneQuantRPKM.50FN.samplename.resk10.norm.txt.gz. These data were pre-normalized based on library size and learned PEER factors, and contain only expressed genes. We kept only samples of European origin.

eQTL analyses

We performed eQTL tests using GTEx and Geuvadis data for haplotypes in the top 99th percentile for any population studied. Because these data sets contain only European samples, we only tested haplotypes that reach a frequency of ≥15% in Europeans. Importantly, because SNP content can vary somewhat between the same haplotype in different populations, we also only tested SNPs that were shared between the introgressed haplotypes in Europeans and the population being tested. For GTEx data, we also limited analyses to tissues with 60 or more samples, which included subcutaneous adipose, aorta, tibial artery, transformed fibroblasts, esophagus mucosa, esophagus muscle, left heart ventricle, lung, skeletal muscle, tibial nerve, sun exposed skin, thyroid, and whole blood. Finally, for power considerations, we only ran tests in which there were at least three samples homozygous for each allele.

For each haplotype, we tested for eQTLs across all combinations of SNPs, tissues, and genes within 500kb of the haplotype by building linear models using the lm() function in R [S9]. We retained the minimum *P* value from all SNPs tested as the single test statistic for each tissue-gene combination. Then, to determine which tests were significant, we ran permutations by shuffling genotypes 1000 times and repeating tests. Using these data, we chose the maximum P-value cutoff that gave an FDR ≤ 0.05; which is the cutoff at which the ratio of the number of significant permutation tests to the number of significant tests on the true data is ≤ 0.05. Importantly, we calculated a single cutoff across all haplotypes, rather than controlling for FDR on a per haplotype basis.

Analysis of B Cell eQTLs from Fairfax et al 2012

We normalized raw array expression data using the R packages Lumi, Limma, and a variance stabilizing transformation. We mitigated potential batch effects with SVA and removed a single sample that was an outlier in a principal component analysis performed on the genotype data. We then tested expression probes for each *TLR* gene (one probe for

TLR1, one for *TLR6*, and two from *TLR10*) for association with each of the three SNPs on the Neandertal haplotype that were genotyped in the study, and retained the most significant association for each gene. Tests were performed using standard linear regression models from the lm() function in R [S9].

Analysis of whole blood LPS stimulation cohort

We recruited healthy volunteers from the Seattle metropolitan area. Exclusions to enrollment were active smoking, recent antibiotic use, symptoms consistent with an infection, a history of autoimmune disease, immunodeficiency, cancer, pregnancy, or use of immunosuppressive medications. Volunteers ranged from 18-65 years old and had a 58%/42% gender balance. Additional cohort details have been described previously [S10]. The study was approved by the Human Subjects committee at the University of Washington. We isolated genomic DNA and genotyped subjects using the Illumina Human 1M Beadchip array (San Diego, CA). We imputed genotypes using EUR genotypes from 1000 Genomes as a reference population using the BEAGLE software and poorly imputed SNPs (BEAGLE $R^2 < 0.90$) were removed from further analysis. From the same subjects, whole blood was stimulated with ultrapure LPS from *Salmonella minnesota* R595 (List Biological Laboratories, Inc., Campbell, CA). RNA extraction was done with the AB 6100 Nucleic Acid Prep Station (ABI/Life Technologies, Foster City, CA), and RNA quality was analyzed with Experion Automated Electrophoresis System (Bio-Rad, Hercules, CA). Gene expression was quantified using the Illumina HumanRef-8 v3.0 Gene Expression BeadChip array (Illumina, San Diego, CA). Expression data quality control was performed using GenomeStudio (Illumina). Duplicate arrays were removed and data were log2 transformed in BASE [11]. For this study, 252 subjects were used for the analyses, which were limited to expression for probes specific to TLR 1,TLR10 and TLR6.

We built standard linear regression models using the lm() function in R [9] to test for association between all SNPs on the TLR haplotype that were genotyped (81 SNPs) and all normalized TLR expression probes (two for *TLR6* and *TLR10*, and one for *TLR1*), both before and after stimulation with LPS. We used age, sex, and the first three genotyping principal components as covariates.

Roadmap Epigenomics Data

We downloaded H3K27ac, H3K4ME1, and DNaseI narrowPeak calls for consolidated melanocyte epigenomes E059 and E061 from the Roadmap Epigenomics project [S12]: http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/ and used bedops[13] to determine which variants on the OCA2 haplotype overlapped melanocyte regulatory elements.

Supplemental References

[S1] Ewing, G., and Hermisson, J. (2010) MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics. *26*, 2064–2065.

[S2] Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and Bustamante, C.D. (2011) Demographic history and rare allele sharing among human populations. Proc Natl Acad Sci USa *108*, 11983–11988.

[S3] Juric, I., Aeschbacher, S., Coop, G. (2015) The Strength of Selection Against Neanderthal Introgression. bioRxiv:030148. doi:10.1101/030148.

[S4] Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015) The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. *43*, D670–D681.

[S5] Wang, J., Duncan, D., Shi, Z., Zhang, B. (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res *41*, W77–W83.

[S6] 't Hoen, P.A.C., Friedländer, M.R., Almlöf, J., Sammeth, M., Pulyakhina, I., Anvar, S.Y., Laros, J.F., Buermans, H.P., Karlberg, O., Brännvall, M., et al. (2013) Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol *31*, 1015–1022.

[S7] Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

[S8] Stegle, O., Parts, L., Piipari, M., Winn, J., Durbin, R. (2012) Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc. *7*, 500–507.

[S9] R Core Team. (2013) R: A language and environment for statistical computing. Vienna.

[S10] Wurfel, M.M., Gordon, A.C., Holden, T.D., Radella, F., Strout, J., Kajikawa, O., Ruzinski, J.T., Rona, G., Black, R.A., Stratton, S., et al. (2008) Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. Am J Respir Crit Care Med. *178*, 710–720.

[S11] Saal, L.H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A., Peterson, C. (2002) BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. Genome Biol. *3*, SOFTWARE0003.

[S12] Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature. *518*, 317–330.

[S13] Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012) BEDOPS: high-performance genomic feature operations. Bioinformatics. *28*, 1919–1920.