Disease Heritability Enrichment of Regulatory Elements Is Concentrated in Elements with Ancient Sequence Age and Conserved Function across Species

Margaux L.A. Hujoel,[1,2,*] Steven Gazal,[3,4] Farhad Hormozdiari,[3,4] Bryce van de Geijn,[3,4] and Alkes L. Price[1,3,4,*]

Regulatory elements, e.g., enhancers and promoters, have been widely reported to be enriched for disease and complex trait heritability. We investigated how this enrichment varies with the age of the underlying genome sequence, the conservation of regulatory function across species, and the target gene of the regulatory element. We estimated heritability enrichment by applying stratified LD score regression to summary statistics from 41 independent diseases and complex traits (average N = 320K) and meta-analyzing results across traits. Enrichment of human putative enhancers and promoters was larger in elements with older sequence age, assessed via alignment with other species irrespective of conserved functionality: putative enhancer elements with ancient sequence age (older than the split between marsupial and placental mammals) were 8.8× enriched (versus 2.5× for all putative enhancers; p = 3e−14), and promoter elements with ancient sequence age were 13.5× enriched (versus 5.1× for all promoters; p = 5e−16). Enrichment of human putative enhancers and promoters was also larger in elements whose regulatory function was conserved across species, e.g., human putative enhancers that were enhancers in ≥5 of 9 other mammals were 4.6× enriched (p = 5e−12 versus all putative enhancers). Enrichment of human promoters was larger in promoters of loss-of-function intolerant genes: 12.0× enrichment (p = 8e−15 versus all promoters). The mean value of several measures of negative selection within these genomic annotations mirrored all of these findings. Notably, the annotations with these excess heritability enrichments were jointly significant conditional on each other and on our baseline-LD model, which includes a broad set of coding, conserved, regulatory, and LD-related annotations.

Introduction

Disease-associated variants and disease heritability have been widely reported to be concentrated in regulatory annotations, such as enhancers and promoters.[1–7] These findings have motivated recent studies of how enhancers and promoters evolve across species.[8–11] Vierstra et al.[8] analyzed DNase I hypersensitivity sites (DHSs) in humans and mice and reported that both human-specific DHSs and human DHSs that were conserved in mice were significantly enriched for disease- and trait-associated variants, despite decreased constraint within human-specific DHSs. This implies that both human-specific regulatory elements and regulatory elements that are shared across species are important for disease; however, these analyses were restricted to only one species other than humans, and do not elucidate the relative importance of these two types of regulatory elements. Villar et al.[9] analyzed 20 mammalian species and reported that enhancers evolve more rapidly than promoters, and that enhancers were often species specific whereas promoters were often functionally conserved. Vermunt et al.[10] and Trizzino et al.[11] analyzed 3–6 primate species and reported that regulatory elements were generally functionally conserved across primates, with higher sequence and function conservation for promoters than for enhancers. However, which enhancers and promoters are most important for disease

remains largely unknown. Further investigating which enhancers and promoters are most important for disease would improve our biological understanding of disease architectures.

Here, we characterize the contribution of enhancers and promoters to disease heritability based on sequence age, conserved function across species, and gene function of the target gene. We achieve this by constructing new annotations using enhancers and promoters previously identified in liver tissue using ten high-quality genomes (humans and nine other mammalian species[9]) and applying stratified LD score regression with the baseline-LD model[6,7] to summary association statistics from 41 independent diseases and complex traits (average N = 320K). An overview of the data sources used in our analyses is provided in Figure 1. We find that disease heritability enrichment is concentrated in putative enhancers and promoters with ancient sequence age and conserved function across species, as well as promoters of loss-of-function intolerant genes from the Exome Aggregation Consortium (ExAC).[12] The mean value of several measures of negative selection within these genomic annotations mirrored all of these findings, with larger heritability enrichments for annotations under stronger negative selection. Our findings are consistent with previous studies broadly demonstrating that regions under strong negative selection are enriched for disease heritability and disease-associated variants.[6,7,13–20]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; [2]Division of Biostatistics, Dana-Farber Cancer Institute, Boston, MA 02215, USA; [3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; [4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
*Correspondence: hujoel@g.harvard.edu (M.L.A.H.), aprice@hsph.harvard.edu (A.L.P.)
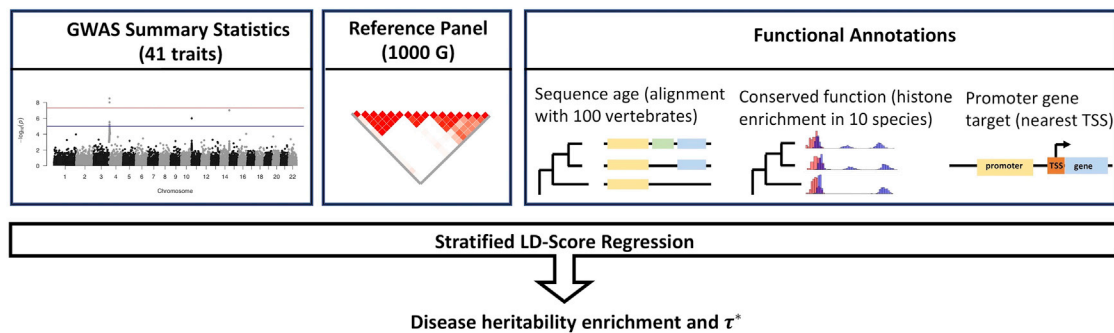https://doi.org/10.1016/j.ajhg.2019.02.008.

Figure 1. Data Sources Used in Analyses
New functional annotations are constructed using a variety of previous research.[9,12,21] By applying stratified LD score regression including both these annotations and the baseline-LD model[6,7] to summary association statistics from 41 independent diseases and complex traits (average N = 320K), we can determine the disease heritability enrichment and standardized effect size for annotations of interest.

Material and Methods

Putative Enhancer and Promoter Annotations

Our goal is to understand the role of human enhancers and promoters in the genetic architecture of diseases and complex traits. We first annotated regions as putative enhancers and promoters using previously reported enhancer and promoter regions that were enriched for histone marks (H3K27ac and H3K4me3) in at least two of four human liver tissue samples.[9] The study identifying these regulatory elements reported that a "sizable majority" of identified enhancers are regulatorily active (based on results of further experimental assays) and that most regions enriched for H3K4me3 (annotated as promoters) lie near transcription start sites.[9] However, we conservatively refer to these enhancer regions as "putative enhancers." We merged any overlapping annotations, resulting in putative enhancer and promoter regions with mean segment lengths of 3.4 kb and 4.3 kb, respectively. In total, 3.3% of common variants lie within putative enhancers and 1.5% within promoters (Table 1). Correlations between the putative enhancer and promoter annotations and various subsets of these annotations (described below) are reported in Figure S1 and Table S1.

Sequence Age Annotations

We constructed genomic annotations based on a previous study which classified sequence age through genome-wide alignments of 100 vertebrates.[21] That study annotated each region of the human genome with an associated score between 1 and 19 based on the number of key ancestral nodes in the tree of vertebrates that it aligned to (1st root = human; 19th root = vertebrates); younger regions were assigned smaller scores, whereas older regions were assigned larger scores. Most regions were assigned a precise age (one score), but some regions were assigned an age interval (range of scores) or an inconsistent age. Regions with inconsistent age were removed.

We annotated SNPs in putative enhancer and promoter regions according to the age of the sequence in the corresponding region of the genome (start location ≤ SNP location < end location). We removed regions in which the alignment at the 19th root was uncertain and assigned the maximum sequence age for SNPs in regions with an age interval. We categorized the ages as post-eutheria split (1–11; young), eutheria (12; intermediate), and pre-eutheria split (13–19; ancient), with approximately one third

of SNPs in putative enhancers and promoters falling in each of these age bins. Pre-eutheria split (or ancient sequence age) means the sequence has an age older than the split between placental and marsupial mammals (>160 million years old[22,23]).

We also analyzed 24 putative regulatory annotations from the baseline model[6] (Table S2). We intersected these annotations with the ancient sequence age annotation, resulting in 24 putative regulatory annotations that have ancient sequence age. Furthermore, we analyzed two chromatin marks (H3K4me3 and H3K27ac) directly measured in various tissues and cell types and defined annotations based on these marks being present in at least 1,10, or 20 tissues/cell types;[5] we also intersected these annotations with the ancient sequence age annotation.

Conserved Function Annotations

We annotated human putative enhancers and promoters according to their conserved function, based on previous work specifying for each human element whether the element was functionally conserved (sequence aligned with histone mark signal conserved across species), mapped (sequence aligned), or missing in analyses of nine other mammalian species (with high-quality genomes).[9] Human-specific and highly conserved putative enhancers and promoters were defined as elements with conserved function in 0 or 9 of the 9 mammals, respectively. We denote conserved putative enhancers and promoters as elements with conserved function in at least 5 of the 9 mammals.

We constructed 6 categorical annotations (3 for promoters and 3 for putative enhancers): each putative enhancer and promoter was annotated with the conservation count (CC) in other species (CC = 0,1,…9; both align and have functional conservation), the mapped count in other species (0,1,…9; align but no functional conservation) and the missing count in other species (0,1,…9). We introduced 20 binary annotations (10 for promoters and 10 for putative enhancers) reflecting the 10 possible values of CC (0,1,…9).

We computed the conservation count, mapped count, and missing count of all elements (prior to merging) and then merged information across overlapping elements. For elements that overlapped, for each of conserved, mapped, and missing count, we computed the union of each count across species; this implies that these small proportions of the genome where two or more elements overlap could get conservation, mapped, and missing counts that add up to more than 9.

Table 1. Annotations Analyzed in Main Analyses

Annotation	Prop. SNPs	Prop. Putative enhancer/ promoter	Mean Segment length (BP)
Putative enhancer	0.0332		3,362
Promoter	0.0152		4,308
Ancient putative enhancer	0.0052	0.1557	191
Ancient promoter	0.0042	0.2765	159
Conserved putative enhancer	0.0055	0.1649	4,962
Conserved promoter	0.0080	0.5251	4,502
Promoter of ExAC gene	0.0025	0.1640	4,549

We report the proportion of common SNPs (MAF $\geq$ 0.05) and mean segment length in base pairs (BP) for each annotation. For the count annotations (putative enhancer conservation count and promoter conservation count), we report here the corresponding binary annotations, conserved putative enhancer and conserved promoter. Mean segment length is computed after merging overlapping elements. Main annotations are publicly available (see Web Resources).

Gene Function Annotations

To assess how the target gene may impact the role of a promoter in disease architecture, we annotated promoters based on whether they were a promoter of an ancient gene (P1–P10 genes[24–26] which emerged before the vertebrates split $\approx$ 500 million years ago[27]), a loss-of-function intolerant gene from ExAC (ExAC gene;[12] 3,230 such genes; Table 1), or a gene with a mouse ortholog (identified in hg38; we assume that gene names remain consistent across builds; see Web Resources).[28]

We obtained the coordinates of all TSSs (see Web Resources) and associated genes.[29] We calculated the mid-point of each merged promoter and determined whether the closest transcription start site (TSS) within 5 kb to the midpoint corresponded to a gene in the specified gene set.

Heritability Enrichment and Standardized Effect Size (τ^*) Metrics

In order to estimate the heritability enrichment of an annotation, we ran stratified LD score regression (S-LDSC)[6,7] using 1000 Genomes as the LD reference panel.[30] Consider C binary or continuous-valued annotations ($a_1, \ldots, a_C$), denote $a_c(j)$ the annotation value of SNP j for annotation c, and assume that the variance of per-normalized-genotype effect sizes linearly depends on the C annotations: $Var(\beta_j) = \sum_c a_c(j)\tau_c$, where τ_c is the per-SNP contribution of one unit of the annotation c to heritability (jointly modeled with all other annotations). S-LDSC estimates τ_c using the summary statistic for a SNP j (χ_j^2) via the following equation:

$$E\left[\chi_j^2\right] = N\sum_c l(j,c)\tau_c + Nb + 1 \qquad \text{(Equation 1)}$$

where N is the sample size of the GWAS, b quantifies confounding biases,[31] and $l(j,c) = \sum_k a_c(k)r_{jk}^2$ is the LD score of SNP j to annotation c where r_{jk} is the correlation between SNPs j and k. S-LDSC estimates two metrics quantifying the role of a functional region in diseases and complex traits. First, it estimates the heritability enrichment of binary annotations, defined as the proportion of heritability explained by SNPs in the annotation divided by the proportion of SNPs in the annotation. The enrichment of annotation c is estimated as

$$Enrichment_c = \frac{\%h^2(c)}{\%SNP(c)} = \frac{h^2(c)/h^2}{|c|/M}, \qquad \text{(Equation 2)}$$

where $h^2(c)$ is the heritability causally explained by common SNPs in annotation c, h^2 is the heritability causally explained by common SNPs, $|c|$ is the number of common SNPs that lie in the annotation, and M is the number common SNPs (in our analyses $M = 5,961,159$ SNPs, see below). A value greater than 1 would indicate a functional annotation is enriched for trait heritability or the proportion of heritability explained is greater than one would expect given the size of the annotation.

Standardized effect size (τ_c^*) was previously defined[7] as the proportionate change in per-SNP heritability associated with a one standard deviation increase in the value of the annotation, conditional on the other annotations in the model; τ_c^* quantifies effects that are unique to the focal annotation, unlike heritability enrichment.[6,7,32] In detail,

$$\tau_c^* = \frac{Msd_c}{h^2}\tau_c, \qquad \text{(Equation 3)}$$

where sd_c is the standard deviation of annotation c.

Regression SNPs (the SNPs used by S-LDSC to estimate τ_c from marginal association statistics) were obtained from the HapMap Project phase 3; these SNPs are considered to be well-imputed SNPs. SNPs with marginal association statistics larger than 80 or $0.001N$ and SNPs that are in the major histocompatibility complex (MHC) region were excluded from all analyses. Reference SNPs (the SNPs used by S-LDSC to compute LD scores) were defined as the set of 9,997,231 biallelic SNPs with minor allele count greater than or equal to five in the set of 489 unrelated and outbred European samples[33] from phase 3 of 1000 Genomes Project (1000G).[30] We note that regression SNPs tag potentially causal reference SNPs via LD scores computed using reference SNPs.[6,7] Heritability SNPs (the SNPs used by S-LDSC to compute h^2, $h^2(c)$, $|c|$ and sd_c) were defined as the 5,961,159 common variants (MAF $\geq$ 0.05) in the set of reference SNPs. Using the LD score for each annotation and the marginal statistics obtained from the trait phenotypes, we computed the heritability enrichment and τ_c^* for each annotation.

In all analyses we included the putative enhancer and promoter annotations[9] as well as a broad set of 75 functional annotations from the baseline-LD (v1.1) model,[7,32] which include functional annotations from the baseline model (i.e., coding, intron, DHS, ...), plus 10 MAF bins and 6 LD-related annotations (Table S2). We note that the inclusion of MAF- and LD-related annotations implies that the expected causal heritability of a SNP is a function of MAF and LD. The 75 functional annotations from the baseline-LD model are included in each analysis to account for LD-dependent architectures and to minimize the risk of model misspecification, which could bias estimates.[6,7] The 75 annotations do not all produce conditionally statistically significant signals, in part due to correlations between annotations that can compromise conditional statistical significance. However, including all of these annotations minimizes the risk of model misspecification when analyzing new annotations. We meta-analyzed results across a previously described set[32] of 41 independent diseases and complex traits (average N = 320K, computed using largest dataset for each trait); for six traits we analyzed two datasets (genetic correlation > 0.9), leading to a total of 47 datasets analyzed

(Table S3; see Web Resources). We performed random-effects meta-analyses across traits using the R package rmeta and the function meta.summaries() (consistent with Finucane et al.[6]). All models tested, including the annotations considered in each model, are listed in Table S4. Correlations between our annotations and annotations from the baseline-LD model are reported in Figure S1 and Table S1.

Reported enrichment estimates are based on a random-effects meta-analysis of enrichment estimates for each trait. The p value for enrichment is computed using a random-effect meta-analysis of $(h^2(C)/|C|) - ((h^2 - h^2(C))/(M - |C|))$ across the traits (we used this quantity because enrichment is not normally distributed[6]) and testing the null hypothesis that this difference is 0 by computing a z-score. For enhancer conservation count and promoter conservation count, we calculate enrichment for bins of this categorical annotation.[7]

For each new annotation, we ran S-LDSC conditional on the putative enhancer and promoter annotations as well as the baseline-LD model. For each annotation type (sequence age, conserved function, and gene function) we derived a joint model by running S-LDSC with the full set of annotations of that type conditional on the enhancer and promoter annotations and the baseline-LD model. (For the sequence age analysis, this set consisted of four annotations: young putative enhancer/promoter and ancient putative enhancer/promoter; we removed the intermediate putative enhancer/promoter to avoid linear dependence between the annotations. For the conserved function model, this set consisted of eight annotations: human-specific putative enhancer/promoter, highly conserved putative enhancer/promoter, putative enhancer/promoter CC, and putative enhancer/promoter missing count; we removed putative enhancer/promoter mapped count to avoid linear dependence between the annotations. For the gene function model, this set consisted of three annotations: promoter of an ancient gene, promoter of ExAC gene, and promoter of a gene with a mouse ortholog.) We then iteratively removed the least statistically significant annotation (excluding annotations in the baseline-LD model as well as the enhancer and promoter annotation) until each remaining annotation was significant (after correction for multiple testing).[7] To produce a combined joint model, we combined the significant sequence age, conserved function, and gene function annotations into a single model and again iteratively removed the least statistically significant annotation until each annotation remained significant (after correction for multiple testing). We note that this model selection technique may result in inflated p values, analogous to winner's curse.[34] However, our assessment of conditional significance addresses this by correcting for the total number of annotations tested. For each model, we performed a secondary analysis in which we additionally included annotations defined by 500 bp flanking regions around each of the new annotations in the model; this helps to guard against bias due to model misspecification.[6]

In order to determine whether a subset of putative enhancers and promoters were particularly enriched as compared to all putative enhancers or promoters, we computed the enrichment difference between an annotation A and a subset a. For each trait, we computed the difference in enrichment between the annotations (Δ) and the standard error for this difference (using block-jackknife) and then meta-analyzed results across 41 traits using random-effects meta-analysis. In order to compute a p value for the difference in enrichment, we computed the normally distributed quantity $\{(h^2(a)/|a|) - ((h^2(A) - h^2(a))/(|A| - |a|))\}$, as well as its standard error for each trait (using block-jackknife), and then meta-analyzed results across traits. We then tested the null hypothesis that this difference is 0; this test assesses whether the per-SNP heritability within annotation A is different within a than outside a. This test is a natural extension of the approach used to assess statistical significance of enrichment.[6]

We computed the proportion of enrichment for an annotation A, attributable to a subset a. The proportion of enrichment for A attributable to a is defined as

$$\frac{(enrichment_a - 1) * \%SNP(a)}{(enrichment_A - 1) * \%SNP(A)}. \qquad \text{(Equation 4)}$$

If $enrichment_a = enrichment_A$, then the proportion of enrichment for A attributable to a is just the proportion of A in subset a. We computed this quantity for each trait, used block-jackknife to compute standard errors, and meta-analyzed results across 41 traits.

Negative Selection Metrics

It has been widely reported that although regions under strong negative selection are depleted of genetic variation, these regions are enriched for disease heritability and disease-associated variants.[6,7,13–20] For example, the 2.6% of SNPs lying in regions that are conserved across 29 mammals (spanning 4.2% of the genome) were reported to explain 24%–35% of disease and complex trait heritability.[6,7,35]

We quantified the strength of negative selection within these annotations by computing the mean value of several measures of negative selection and computing the standard error using block-jackknife with 200 equally sized blocks of adjacent SNPs within the annotations (all measures are annotations in the baseline-LD model; Table S2). First, we computed the proportion of common SNPs with GERP++ rejected substitutions (RS) score ≥ 4 (GERP RS ≥ 4, binary annotation) within the baseline-LD annotations.[7,36] This score is equal to the difference between the neutral and observed substitution rates and reflects the intensity of constraint at a given genomic location, such that a larger score is indicative of stronger negative selection. Second, we computed the mean background selection statistic (BSS = 1-McVicker B statistic[37]) (at common SNPs); a BSS value close to 1 indicates that background selection resulted in near complete removal of diversity whereas a value close to 0 indicates little effect.[7] Third, we computed the proportion of common SNPs conserved across mammals;[6,35] regions that are conserved across mammals are likely to be critical, as mutations were not tolerated. Fourth, we computed the mean MAF-adjusted predicted allele age (at common SNPs); on average, recent variants are more deleterious.[7,38] Fifth, we computed mean nucleotide diversity[39] (at common SNPs); variants that lie in regions with low nucleotide diversity are more likely to be deleterious.[7,40]

Results

Disease Enrichment Is Concentrated in Putative Regulatory Elements with Ancient Sequence Age

We focused our analyses on putative enhancer and promoter elements that were previously annotated based on H3K27ac and H3K4me3 marks assayed in human liver[9] (Table 1). To assess the disease enrichment of these elements, we applied S-LDSC with the baseline-LD model[6,7]
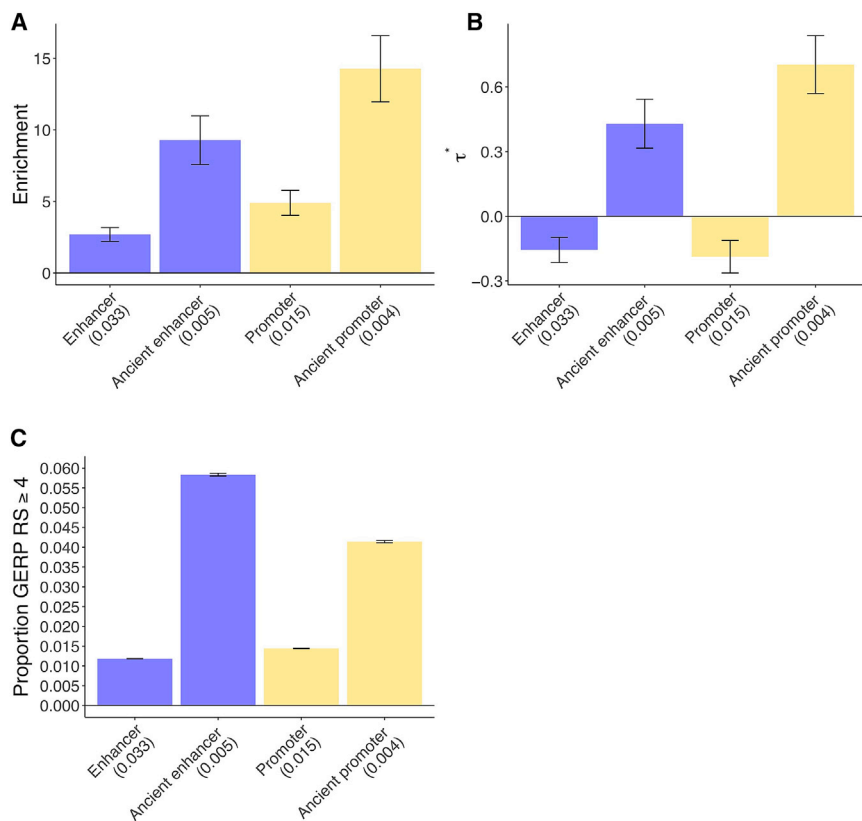
A

B

C

sequence is older than the split between marsupial and placental mammals (>160 million years old[22,23]); 16% of putative enhancer SNPs were annotated as ancient putative enhancer, and 28% of promoter SNPs were annotated as ancient promoter (Table 1). We computed correlations between our annotations and 38 annotations from the baseline-LD model (Table S2): 32 functional annotations and 6 LD-related annotations. The ancient putative enhancer and ancient promoter annotations were only weakly correlated with annotations from the baseline-LD model (Figure S1 and Table S1).

To assess how the disease enrichment of putative enhancers and promoters varies with sequence age, we repeated our S-LDSC analysis with each of the six age-specific annotations (young, intermediate, or ancient; putative enhancer or promoter) included in turn, in addition to baseline-LD + putative enhancer + promoter annotations. We observed the strongest enrichments for ancient putative enhancers and ancient promoters (Table S6). We constructed a joint sequence age model by retaining only the age-specific annotations that remained significant (after correction for multiple testing) when conditioned on the baseline-LD + putative enhancer + promoter annotations;[7] only the ancient putative enhancer and ancient promoter annotations were jointly significant. Ancient putative enhancers were 9.3× enriched, compared to 2.7× for all putative enhancers (p = 4e−15 for difference), and ancient promoters were 14.3× enriched, compared to 4.9× for all promoters (p = 2e−18 for difference) (Figure 2A, Tables S7A and S8). We note that enrichment estimates - which differ from model to model - can change slightly depending on the set of annotations included in the model;[6,7] the enrichment estimates reported in the Abstract are estimates obtained using the combined joint model. Although ancient putative enhancers comprise only 16% of putative enhancers (at the level of common

to summary statistics from 41 independent diseases and complex traits (average N = 320K; Table S3) and meta-analyzed results across traits. We observed significant heritability enrichment for both putative enhancers (2.6×, p = 3e−12) and promoters (4.6×, p = 3e−17) (Table S5A), consistent with previous studies of disease enrichment of regulatory elements.[1–7] Based on significance of regression coefficients, we determined that the promoter annotation (but not the putative enhancer annotation) provides unique information conditioned on the baseline-LD model (p = 0.007; Table S5A), which includes a broad set of regulatory annotations (Table S2). Analyses of highly reproducible putative enhancer and promoter annotations (reproduced in all four tissue samples from Villar et al.[9]) produced similar results (Table S5B).

We annotated putative enhancer and promoter regions according to their underlying sequence age, assessed via genome-wide alignment of 100 vertebrates irrespective of conserved functionality.[21] Each region of the human genome had an associated score between 1 and 19 based on the number of key ancestral nodes in the tree of vertebrates that it aligned to. We classified enhancer and promoter regions as having a young (1–11), intermediate (12), or ancient (13–19) sequence age (see Material and Methods); different regions within the same enhancer or promoter may be assigned different sequence ages, such that ancient enhancers/promoters represent the ancient parts of the enhancers/promoters rather than different enhancers/promoters. Ancient sequence age means the

SNPs), they contribute 59% (SE 5%) of all putative enhancer enrichment. Analogously, although ancient promoters comprise only 28% of promoters, they contribute 82% (SE 4%) of all promoter enrichment.

Both ancient putative enhancers and ancient promoters were uniquely informative for disease heritability conditional on the baseline-LD + putative enhancer + promoter annotations, as quantified by τ^* (the proportionate change in per-SNP heritability associated with an increase in the value of the annotation by one standard deviation, conditional on other annotations included in the model[7]) (Figure 2B and Table S7B). Specifically, we estimated large and highly significant values of τ^* for both ancient putative enhancers ($\tau^* = 0.43$, p = 1e−13) and ancient promoters ($\tau^* = 0.70$, p = 9e−25). In particular, these τ^* values were larger than the analogous τ^* values that we recently estimated for both LD-related annotations[7] and molecular QTL annotations,[32] implying a substantial improvement in our understanding of which regulatory elements contribute to disease heritability. The slightly but significantly negative value of τ^* for (all) putative enhancers and (all) promoters indicates conditional depletion for putative enhancers and promoters that do not have ancient sequence age (Figure 2B and Table S7B).

We quantified the mean strength of negative selection within each of the annotations from Figure 2A. We first calculated the proportion of common SNPs with GERP++ rejected substitutions (RS) score ≥ 4 (GERP RS ≥ 4).[7,36] The GERP RS score reflects the difference between the neutral and observed substitution rates and thus reflects the intensity of constraint at a given genomic location, such that a larger score is indicative of stronger negative selection. We determined that the stronger disease enrichment for ancient putative enhancers and ancient promoters is mirrored by the larger proportion of variants in these annotations with GERP RS ≥ 4, reflecting stronger negative selection (Figure 2C and Table S7C). We note that 1.2% and 1.4% of common SNPs in putative enhancer and promoter annotations (and 5.8% and 4.1% of common SNPs in ancient putative enhancer and promoter annotations) have GERP RS ≥ 4, as compared to 0.81% of all common SNPs, thus regulatory regions enter the regime of strong selection fairly frequently. We note that as sequence age is assessed via sequence conservation across species, we expected the GERP scores in ancient regions to be higher. However, we did not know in advance whether the quantitative pattern of enrichment would closely mirror the quantitative pattern of GERP++ scores. We observed similar patterns for four other measures of negative selection:[7] a background selection statistic (BSS) equal to 1 − McVicker B statistic;[37] sequence conservation across 29 mammals[35]; predicted allele age;[7] and nucleotide diversity[39] (Table S7C). However, as noted above, ancient putative enhancers and ancient promoters were uniquely informative for disease heritability conditional on the baseline-LD model, which includes all of these measures of negative selection.

We performed five secondary analyses to assess the robustness of our results. First, we repeated the analysis of Figures 2A and 2B restricting to three liver-related traits (high cholesterol, HDL, and LDL). Although this analysis is less well-powered, the conditional signal for ancient promoters remained statistically significant (p = 0.0003, Table S9). Second, we repeated the analysis of Figures 2A and 2B by adding a binary annotation defined by ancient sequence age (irrespective to enhancer or promoter status) to the model; this annotation was not conditionally informative for disease heritability as quantified by τ^*, and its addition to the model did not significantly change our results (Table S10). Third, we intersected the ancient sequence age annotation with 24 binary annotations from baseline-LD reflecting putative regulatory elements and ran S-LDSC conditional on the baseline-LD model with each of these intersected annotations included in turn. We obtained similar results, with much stronger enrichments for ancient regulatory elements (Table S11). (We used the putative enhancer and promoter annotations from Villar et al.[9] in our main analyses so that we could integrate annotations based on ancient sequence age and conserved function into a combined joint model; see below.) Fourth, we repeated this analysis using H3K27ac and H3K4me3 annotations from Roadmap,[5] defining annotations based on marks present in 1/10/20 tissues/cell types, respectively. As expected, we found that marks present in more tissues/cell types had higher disease enrichment (Table S12). However, we found that each of these annotations had ~3× greater disease enrichment when restricted to regions of ancient sequence age. This shows that our finding of stronger disease enrichments for liver regulatory elements in regions of ancient sequence age is orthogonal to the number of tissues/cell types with regulatory signal. Fifth, we repeated the analysis of Figures 2A and 2B by including 500 bp flanking regions around each of the annotations from Figures 2A and 2B, to guard against bias due to model misspecification[6] (see Material and Methods). We confirmed that this did not significantly change our results (Table S13).

Disease Enrichment Is Concentrated in Putative Regulatory Elements with Conserved Function

We annotated human putative enhancers and promoters according to their conserved function, assessed via how many of nine other mammalian species assayed by Villar et al.[9] had shared regulatory functionality. Each putative enhancer and promoter was annotated with the conservation count (CC) in other species (CC = 0,1,…9). We constructed both integer-valued "conservation count" (value of CC) and binary "conserved" (CC ≥ 5) annotations (see Material and Methods and Table 1). A large proportion of annotated putative enhancers were functionally human specific (40% human-specific [CC = 0] versus 2% highly conserved [CC = 9]), whereas promoters were more functionally conserved (19% human-specific versus 15% highly conserved) (Table S17). Accordingly, 53% of
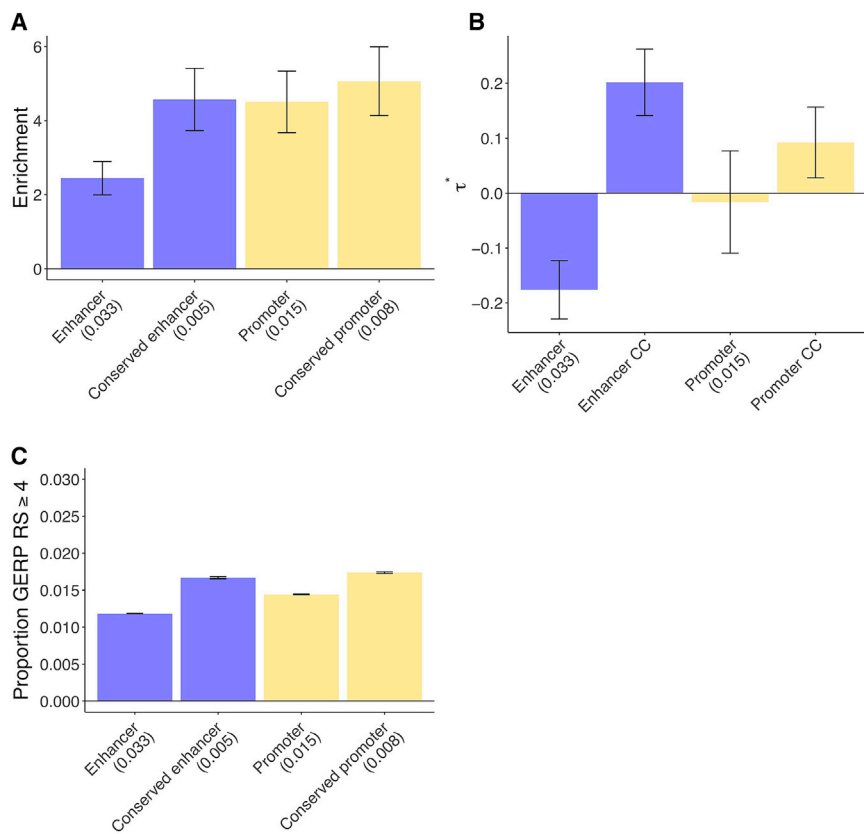
A

B

C

promoters were conserved promoters, whereas only 16% of putative enhancers were conserved putative enhancers (Table 1). The putative enhancer conservation count and promoter conservation count annotations were only weakly correlated with annotations from the baseline-LD model, but moderately correlated with the ancient putative enhancer and ancient promoter annotations (Figure S1 and Table S1).

To assess how the disease enrichment of putative enhancers and promoters varies with conserved function, we performed S-LDSC analyses with each of ten conserved-function-specific annotations (conservation count, highly conserved, human-specific, mapped count, missing count [see Material and Methods]; putative enhancer or promoter) included in turn, in addition to baseline-LD + putative enhancer + promoter annotations. We observed the strongest enrichments for highly conserved putative enhancers and highly conserved promoters, and also observed that while human-specific promoters were enriched, human-specific putative enhancers were not (Table S14). We constructed a joint conserved function model by retaining only the conserved-function-specific annotations that remained significant (after correction for multiple testing) when conditioned on the baseline-LD + putative enhancer + promoter annotations;[7] only the putative enhancer conservation count and promoter conservation count annotations were jointly significant. Because enrichment is not defined for annotations with value 0–9, we estimated the enrichment

of the corresponding binary annotations (conserved putative enhancer and conserved promoter) in the joint model. Conserved putative enhancers were 4.6× enriched, compared to 2.4× for all putative enhancers (p = 3e−12 for difference), and conserved promoters were 5.1× enriched, compared to 4.5× for all promoters (p = 0.022 for difference) (Figure 3A, Tables S15A and S16). We note that enrichment estimates - which differ from model to model - can change slightly depending on the set of annotations included in the model;[6,7] the enrichment estimates reported in the Abstract are estimates obtained using the combined joint model. Although conserved putative enhancers comprise only 16% of putative enhancers, they contribute 35% (SE 2%) of all putative enhancer enrichment. Analogously, although conserved promoters comprise only 53% of promoters, they contribute 59% (SE 2%) of all promoter enrichment.

Both putative enhancer conservation count and promoter conservation count were uniquely informative for disease heritability conditional on the baseline-LD + putative enhancer + promoter annotations, as quantified by τ^* (Figure 3B and Table S15B). Specifically, we estimated significant values of τ^* for both putative enhancer conservation count ($\tau^* = 0.20$, p = 7e−11) and promoter conservation count ($\tau^* = 0.10$, p = 0.005). The significantly negative value of τ^* for (all) putative enhancers indicates conditional depletion for putative enhancers that are not conserved (Figure 3B, Table S15B).

We quantified the mean strength of negative selection within each of the annotations from Figure 3A. We first calculated the proportion of common SNPs with GERP RS $\geq$ 4.[7,36] We determined that the stronger disease enrichments for conserved putative enhancers and conserved promoters is mirrored by the larger proportion of variants in these annotations with GERP RS $\geq$ 4,
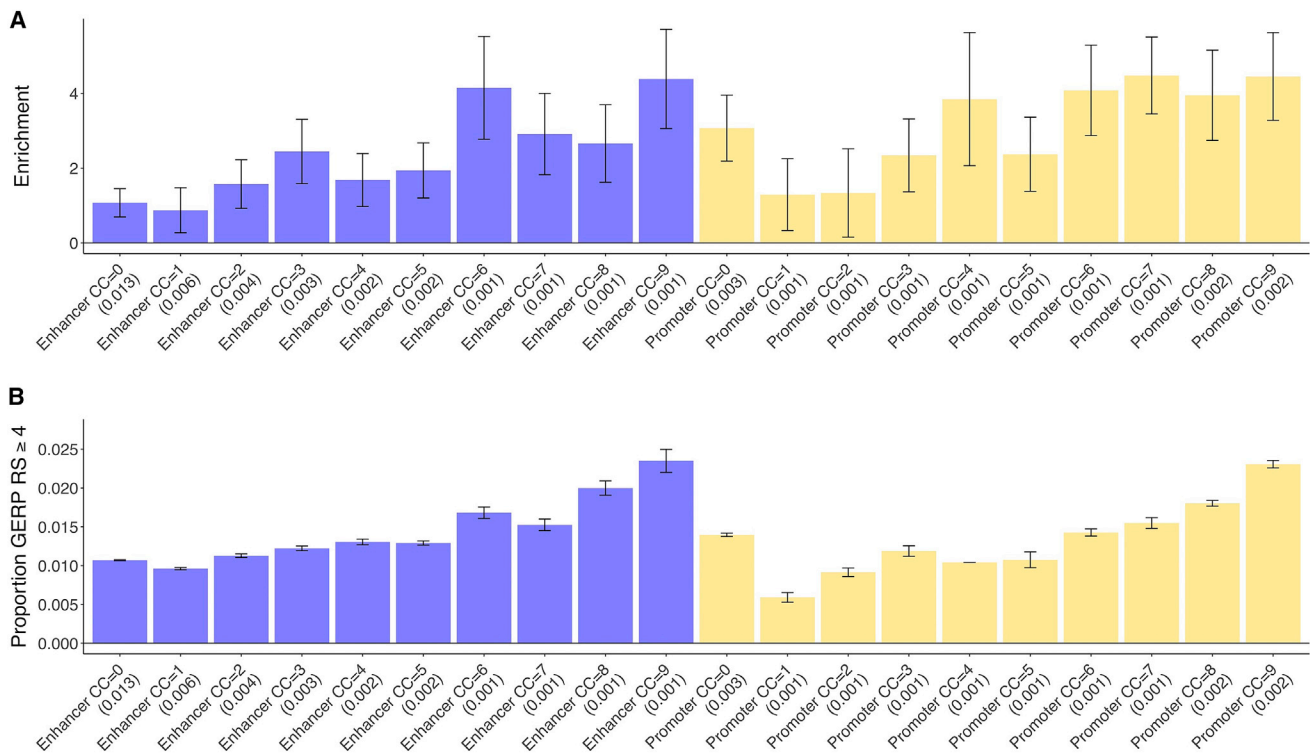
Figure 4. Disease Enrichment of Putative Enhancers and Promoters as a Function of Conservation Count (CC)
(A) Heritability enrichment (± 1.96 standard error); results are meta-analyzed across 41 traits.
(B) Proportion of common SNPs within annotations with GERP RS ≥ 4 [7,36] (± 1.96 standard error). We report the proportion of common SNPs (MAF ≥ 0.05) for each annotation. Numerical results are reported in Table S17, and results for each trait are reported in Table S18.

reflecting stronger negative selection (Figure 3C and Table S15C). We observed similar patterns for four other measures of negative selection (Table S15C). However, as noted above, putative enhancer conservation count and promoter conservation count were uniquely informative for disease heritability conditional on the baseline-LD model, which includes all of these measures of negative selection.

To further assess how the disease enrichment of putative enhancers and promoters varies with conserved function, we repeated our S-LDSC analysis with each of 20 binary conservation count annotations (CC = 0,1,...9; enhancer or promoter) jointly included, in addition to baseline-LD model (the putative enhancer and promoter annotations were excluded to avoid approximate colinearity of the annotations). For putative enhancers, we observed a roughly linear trend whereby putative enhancers conserved in more mammals are progressively more enriched for heritability (Figure 4A, Tables S17A and S18). For promoters, we observed a parabolic trend, similar to the linear trend but with excess heritability for human-specific promoters (Figure 4A and Table S17A).

We quantified the mean strength of negative selection within each of the annotations from Figure 4A. We first calculated the proportion of common SNPs with GERP RS ≥ 4.[7,36] We determined that the linear disease enrichment trend for putative enhancers and parabolic disease enrichment trend for promoters (as conservation count

increases) is mirrored by the proportion of variants in these annotations with GERP RS ≥ 4 (Figure 4B and Table S17B). We observed similar patterns for four other measures of negative selection (Table S17B).

We performed four secondary analyses. First, we repeated the analysis of Figure 3A restricting to the three liver-related traits. Although this analysis is less well powered, the conditional signal for putative enhancer conservation count remained statistically significant (p = 0.001, Table S19). Second, we repeated the analysis of Figure 3A by replacing the putative enhancer conservation count and promoter conservation count annotations in the joint model with binary conserved putative enhancer and conserved promoter annotations, and we confirmed that this did not significantly change our results (Table S20). Third, we repeated the analysis of Table S20 by including 500 bp flanking regions around each of the annotations from Figure 3A (see Material and Methods). This did not significantly change our results; the heritability enrichment for conserved putative enhancer was slightly reduced but remained highly significant (Table S21). Fourth, we repeated the analysis of Figure 3B by including human-specific promoters as an additional annotation. While this new annotation was not conditionally significant, the value of τ^* for the promoter conservation count annotation became larger and more statistically significant (Table S22), consistent with the parabolic trend for promoters in Figure 4A.
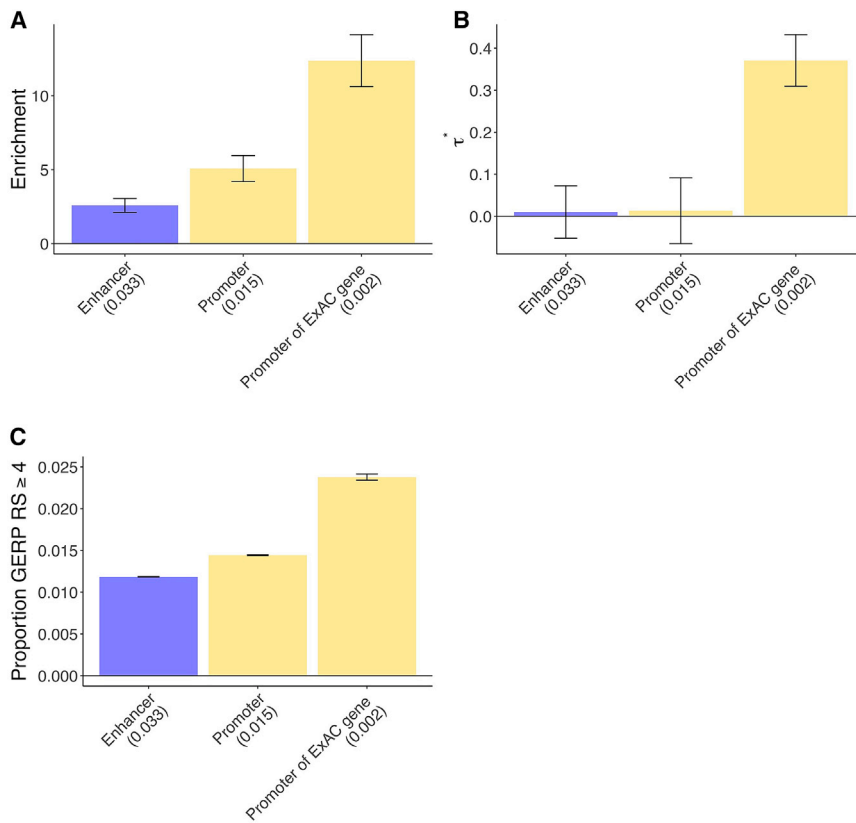
A

B

C

Figure 5. Disease Enrichment of Promoters of ExAC Genes in Gene Function Model

We report results for the gene function annotation that is significant conditional on the baseline-LD model and putative enhancer and promoter annotations (Bonferroni p = 0.05/3 = 0.0167). "ExAC genes" refer to genes annotated as having high pLI in ExAC data.

(A and B) Heritability enrichment (A) and τ^* estimates (± 1.96 standard error) (B); results are meta-analyzed across 41 traits.

(C) Proportion of common SNPs within annotations with GERP RS $\geq$ 4[7,36] (± 1.96 standard error). We report the proportion of common SNPs (MAF $\geq$ 0.05) for each annotation. Numerical results are reported in Table S24, and results for each trait are reported in Table S25.

Disease Enrichment Is Concentrated in Promoters of Loss-of-Function Intolerant Genes

We annotated promoters according to the genes that they regulate (see Material and Methods). In particular, we annotated 16% of promoters as being promoters of the 3,230 ExAC LoF intolerant genes, defined as genes annotated as having a high probability of being LoF intolerant (pLI) in ExAC data[12] (Table 1). The promoter of ExAC gene annotation was only weakly correlated with annotations from the baseline-LD model (Figure S1), but moderately correlated with the ancient promoter and promoter conservation count annotations (Figure S1 and Table S1).

To assess how the disease enrichment of promoters varies with the gene that it regulates, we repeated our S-LDSC analysis with the promoter of ExAC gene annotation included, in addition to baseline-LD + putative enhancer + promoter annotations. We also analyzed promoter of ancient gene and promoter of gene with mouse ortholog annotations in turn (see Material and Methods). The promoter of ExAC gene annotation produced the strongest enrichment (Table S23) and was the only gene function annotation that remained significant (after correction for multiple testing) in a joint analysis conditioned on the baseline-LD + enhancer + promoter annotations.[7] Promoters of ExAC genes were 12.4× enriched, compared to 5.1× for all promoters (p = 9e−16 for the difference) (Figure 5A, Tables S24A and S25). We

note that enrichment estimates - which differ from model to model - can change slightly depending on the set of annotations included in the model;[6,7] the enrichment estimates reported in the Abstract are estimates obtained using the combined joint model. Although promoters of ExAC genes comprise only 16% of promoters, they contribute 39% (SE 2%) of all promoter enrichment.

Promoters of ExAC LoF intolerant genes were uniquely informative for disease heritability conditional on the baseline-LD + putative enhancer + promoter annotations, as quantified by τ^*. Specifically, we estimated a large and highly significant value of τ^* (τ^* = 0.37, p = 2e−32) (Figure 5B and Table S24B).

We quantified the mean strength of negative selection within each of the annotations from Figures 5A and 5B. We first calculated the proportion of common SNPs with GERP RS $\geq$ 4.[7,36] We determined that the stronger disease enrichment for promoters of ExAC genes is mirrored by the larger proportion of variants in these annotations with GERP RS $\geq$ 4, reflecting stronger negative selection (Figure 5C and Table S24C). We observed similar patterns for four other measures of negative selection (Table S24C). However, as noted above, promoters of ExAC genes were uniquely informative for disease heritability conditional on the baseline-LD model, which includes all of these measures of negative selection.

We performed three secondary analyses to assess the robustness of our results. First, we repeated the analysis of Figures 5A and 5B restricting to the three liver-related traits. We observed a non-significant trend toward a conditional signal for promoters of ExAC genes (nominal p = 0.048; not significant after correcting for 3 annotations tested), consistent with the fact that this analysis is less well powered (Table S26). Second, we repeated the analysis
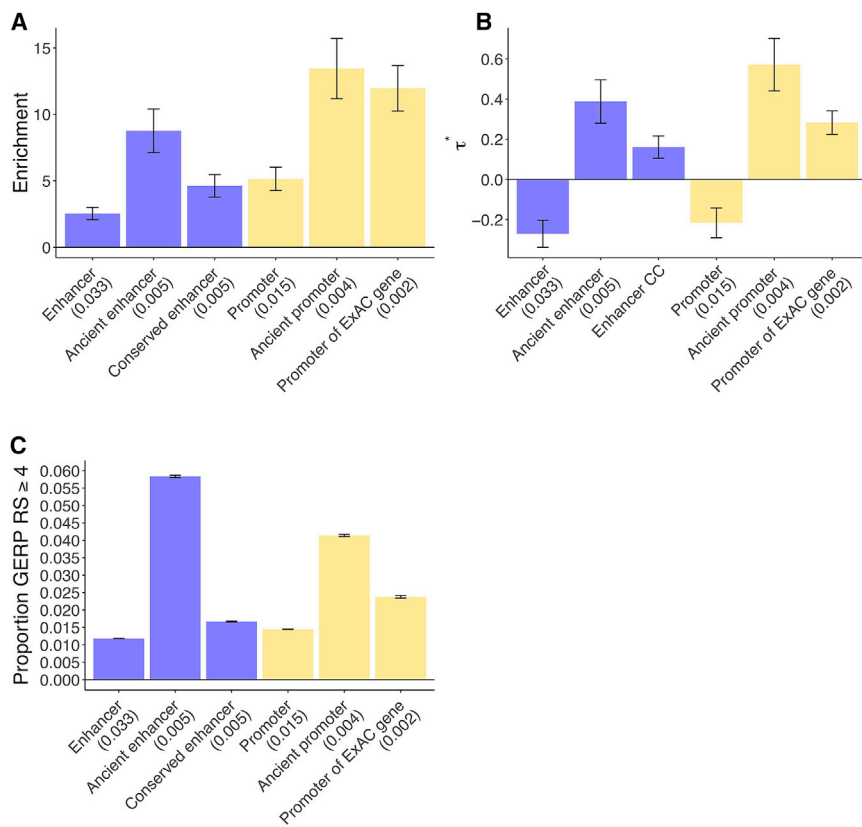
A

B

C

Figure 6. Disease Enrichment of Annotations in Combined Joint Model
We report results for sequence age, conserved function, and gene function annotations that are jointly significant conditional on the baseline-LD model and putative enhancer and promoter annotations (Bonferroni p = 0.05/15 = 0.0033). (A) Heritability enrichment and (B) τ^* estimates (± 1.96 standard error); results are meta-analyzed across 41 traits. CC denotes conservation count. (C) Proportion of common SNPs within annotations with GERP RS ≥ 4[7,36] (± 1.96 standard error). We report the proportion of common SNPs (MAF ≥ 0.05) for each annotation. Numerical results are reported in Table S29, and results for each trait are reported in Table S30.

Ancient putative enhancers were 8.8× enriched, compared to 2.5× for all putative enhancers (p = 3e−14 for difference), and ancient promoters were 13.5× enriched, compared to 5.1× for human promoters (p = 5e−16 for difference) (Figure 6A, Tables S29A and S30); these enrichments differed only slightly from the joint sequence age model (Figure 2A). Conserved putative enhancers were 4.6× enriched (p = 5e−12 for difference versus all human putative enhancers); this enrichment differed only very slightly from the joint conserved function model (Figure 3A). Promoters of ExAC genes were 12.0× enriched (p = 8e−15 for difference versus all promoters); this enrichment differed only very slightly from the joint gene function model (Figure 5A).

In the combined joint model, we estimated highly significant values of τ^* for ancient putative enhancers ($\tau^* = 0.39$, p = 2e−12), ancient promoters ($\tau^* = 0.57$, p = 1e−17), putative enhancer conservation count ($\tau^* = 0.16$, p = 1e−8), and promoters of ExAC genes ($\tau^* = 0.28$, p = 2e−21) (Figure 6B and Table S29B). These τ^* estimates were slightly lower than the corresponding τ^* estimates from the joint sequence age model (Figure 2B), joint conserved function model (Figure 3B), and joint gene function model (Figure 5B), consistent with correlations between these annotations (Figure S1 and Table S1). Notably, the τ^* estimates for ancient enhancers and ancient promoters remained larger than the analogous τ^* values that we recently estimated for LD-related annotations[7] and molecular QTL annotations.[32]

The stronger disease enrichment for ancient putative enhancers, ancient promoters, conserved putative enhancers, and promoters of ExAC genes is mirrored by the larger proportion of variants in these annotations with GERP RS ≥ 4 and four other measures of negative selection, reflecting stronger negative selection (Figure 6C and

of Figures 5A and 5B by including 500 bp flanking regions around each of the annotations from Figures 5A and 5B (see Material and Methods). We confirmed that this did not significantly change our results (Table S27). Third, we repeated the analysis of Figures 5A and 5B by including two annotations based on fine-mapped expression quantitative trait loci (eQTL): the MaxCPP annotation for all genes and the MaxCPP annotation for ExAC LoF genes only.[32] Results were little changed, and promoters of ExAC genes (as well as the MaxCPP (allGenes) and MaxCPP (ExAC) annotations) were still uniquely informative for disease heritability as quantified by τ^* (Table S28).

Combined Joint Model
We constructed a combined joint model by including all jointly significant annotations involving sequence age (Figures 2A and 2B), conserved function (Figure 3B), and gene function (Figures 5A and 5B) and retaining only the annotations that remained significant (after correction for multiple testing) when conditioned both on each other and on the baseline-LD + promoter + putative enhancer annotations.[7] The final joint model included ancient putative enhancer, ancient promoter, putative enhancer conservation count, and promoter of ExAC gene annotations. Because enrichment is not defined for annotations with value 0–9, we estimated the enrichment of conserved putative enhancer in lieu of putative enhancer conservation count, analogous to above.

Table S29C), as we previously determined (Figure 2C and Table S7C; Figure 3C and Table S15C; Figure 5C and Table S24C). However, as noted above, all of these annotations were uniquely informative for disease heritability conditional on the baseline-LD model, which includes all of these measures of negative selection.

We performed six secondary analyses to assess the robustness of our results. First, we repeated the analysis of Figure 6A restricting to the three liver-related traits. Although this analysis is less well powered, the conditional signals for ancient promoters and enhancer conservation count remained statistically significant (Table S31). Second, we repeated the analysis of Figure 6A by replacing the putative enhancer conservation count annotation in the joint model with the binary conserved putative enhancer annotation, and confirmed that this did not significantly change our results (Table S32). Third, we repeated the analysis of Table S32 by including 500 bp flanking regions around each of the annotations from Figure 6A (see Material and Methods). We confirmed that this did not significantly change our results; the enrichment for conserved putative enhancer was slightly reduced but remained highly significant (Table S33). Fourth, we repeated the analysis of Figure 6B by including human-specific promoters and promoter conservation count as additional annotations, in order to investigate whether this might lead to a significant τ^* for the promoter conservation count annotation (as in Table S22) due to the parabolic trend for promoters in Figure 4A. However, the τ^* for both annotations was non-significant (Table S34). Fifth, we repeated the analysis of Figure 6A by including the two fine-mapped eQTL annotations: the MaxCPP annotation for all genes and the MaxCPP annotation for ExAC LoF genes only.[32] Results were little changed, and our new annotations (ancient enhancer, enhancer conservation count, ancient promoter, promoter of ExAC gene) (as well as the MaxCPP [allGenes] and MaxCPP [ExAC] annotations) were still uniquely informative for disease heritability as quantified by τ^* (Table S35A). Sixth, we repeated the analysis from Table S35A, including a new annotation resulting from restricting the MaxCPP of all genes annotation to regions with ancient sequence age. We determined that the conditional eQTL signal is concentrated in the MaxCPP of all genes intersected with ancient sequence age annotation (Table S35B).

Discussion

Our results help elucidate which regulatory elements make the largest contributions to the genetic architecture of diseases and complex traits. We reached three main conclusions. First, disease heritability is concentrated in putative enhancers and promoters with ancient sequence age. Second, disease heritability is concentrated in putative enhancers and promoters with conserved function across species. Third, disease heritability is concentrated in promoters of ExAC LoF intolerant genes. These findings represent unique information about disease heritability conditional on all other available annotations, as quantified by large and highly significant τ^* values (up to 0.57 in combined joint model; Figure 6B), substantially larger than the τ^* values that we reported for other annotations in our recent work.[7,32] In addition to improving our biological understanding of disease architectures, our findings have immediate downstream applications to improve association power,[3,41,42] fine-mapping,[2,43,44] and genetic risk prediction,[45–47] which will provide a means to validate our findings using different methods.

Promoters are known to be functionally conserved more often than enhancers;[9] we determined that conserved putative enhancers, although less common than conserved promoters, are particularly strongly enriched for disease heritability (Figure 6). In addition, previous work reported that human-specific DHSs were significantly enriched for disease- and trait-associated variants, despite decreased constraint; we observed modest enrichment for human-specific promoters but no enrichment for human-specific putative enhancers (Figure 4A). The excess enrichments for putative enhancers and promoters with ancient sequence age raises the question of whether genomic regions with ancient sequence age are broadly important; however, ancient sequence age was not conditionally significant in our analyses (Table S10). Our finding of increased disease enrichment in promoters of ExAC LoF intolerant genes[12] (Figure 6A) is consistent with evidence from eQTL studies;[32] however, our promoter of ExAC gene annotation remains uniquely informative conditional on the fine-mapped eQTL annotations from Hormozdiari et al.[32] (Table S28). We further determined that the conditional eQTL signal is concentrated in MaxCPP (allGenes) intersected with ancient sequence age, suggesting that eQTL integration studies should pay particular attention to whether an eQTL that may be linked to disease lies in a region of ancient sequence age (Table S35B). Our finding of increased disease enrichment in promoters of ancient genes (Table S23) is consistent with previous work showing that genes linked to human disease are more often ancient than recently evolved;[24] however, we determined that the promoter of ancient genes annotation was not uniquely informative once the promoter of ExAC genes annotation was included in our model. Our findings are consistent with previous studies broadly demonstrating that regions under strong negative selection are enriched for disease heritability and disease-associated variants, despite being depleted for genetic variation.[6,7,13–20] (However, analogous to those studies, we are unable to make any statements about lethal mutations that preclude any genetic variation whatsoever.)

We note several limitations of our work. First, we analyzed putative enhancers that were identified using two histone marks in liver tissue,[9] an approach that does not guarantee enhancer functionality. However, that study reported that the majority of the putative

enhancers were regulatorily active (based on results of further experimental assays),[9] implying that our finding of 3.5× stronger disease enrichment for ancient enhancers (and 1.8× stronger disease enrichment for conserved enhancers) cannot arise simply because ancient (or conserved) putative enhancers are more likely to be real enhancers. Nonetheless, the larger disease enrichment for ancient (or conserved) putative enhancers could be due to a combination of ancient (or conserved) enhancers being more strongly enriched and ancient (or conserved) putative enhancers having a higher probability of being truly functional. Second, our main analyses were restricted to putative enhancers and promoters identified in liver tissue.[9] Results involving sequence age were similar for other putative regulatory annotations (Table S11). However, efforts to generalize our results for conserved function are limited by the availability of enhancer and promoter annotations across species in other tissues; one possible solution would be to predict regulatory function across species in other tissues.[48–55] Third, we focused our analyses on common variants by using a 1000 Genomes LD reference panel, but future work could draw inferences about low-frequency variants using larger reference panels.[20] Fourth, inferences about components of heritability can potentially be biased by failure to account for LD-dependent architectures.[7,56–58] All of our analyses used the baseline-LD model, which includes six LD-related annotations.[7] The baseline-LD model is supported by formal model comparisons using likelihood and polygenic prediction methods, as well as analyses using a combined model incorporating alternative approaches;[59] however, there can be no guarantee that the baseline-LD model perfectly captures LD-dependent architectures. Despite these limitations, our results are highly informative for the genetic architecture of diseases and complex traits.

Supplemental Data

Supplemental Data can be found with this article online at https://doi.org/10.1016/j.ajhg.2019.02.008.

Declaration of Interests

The authors declare no competing interests.

Web Resources

Ensembl hg38 build, http://useast.ensembl.org/biomart/martview

Main annotations described in this manuscript, https://data.broadinstitute.org/alkesgroup/LDSCORE/hujoel/

seqMINER, https://sourceforge.net/projects/seqminer/files/Reference%20coordinate

UK Biobank summary statistics, https://data.broadinstitute.org/alkesgroup/UKBB/

References

1. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

2. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. *45*, 124–130.

3. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573.

4. Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

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

6. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al.; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and RACI Consortium (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.

7. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., and Price, A.L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nat. Genet. *49*, 1421–1427.

8. Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

9. Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., et al. (2015). Enhancer evolution across 20 mammalian species. Cell *160*, 554–566.

10. Vermunt, M.W., Tan, S.C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., Kondova, I., Persengiev, S., Bontrop, R., Cuppen, E., et al.; Netherlands Brain Bank (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. Nat. Neurosci. *19*, 494–503.

11. Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G.H., Lynch, V.J., and Brown, C.D. (2017). Transposable elements are the primary source

of novelty in primate gene regulation. Genome Res. *27*, 1623–1633.

12. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

13. Eyre-Walker, A. (2010). Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc. Natl. Acad. Sci. USA *107* (*Suppl 1*), 1752–1756.

14. Agarwala, V., Flannick, J., Sunyaev, S., Altshuler, D.; and GoT2D Consortium (2013). Evaluating empirical bounds on complex disease genetic architecture. Nat. Genet. *45*, 1418–1427.

15. Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al.; NHLBI Exome Sequencing Project (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.

16. Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., and Lander, E.S. (2014). Searching for missing heritability: designing rare variant association studies. Proc. Natl. Acad. Sci. USA *111*, E455–E464.

17. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al.; GERAD1 Consortium; and CRESTAR Consortium (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.

18. Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. *50*, 746–753.

19. Palamara, P.F., Terhorst, J., Song, Y.S., and Price, A.L. (2018). High-throughput inference of pairwise coalescence times identifies signals of selection and enriched disease heritability. Nat. Genet. *50*, 1311–1317.

20. Gazal, S., Loh, P.R., Finucane, H.K., Ganna, A., Schoech, A., Sunyaev, S., and Price, A.L. (2018a). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet. *50*, 1600–1607.

21. Marnetto, D., Mantica, F., Molineris, I., Grassi, E., Pesando, I., and Provero, P. (2018). Evolutionary rewiring of human regulatory networks by waves of genome expansion. Am. J. Hum. Genet. *102*, 207–218.

22. Phillips, M.J., Bennett, T.H., and Lee, M.S.Y. (2009). Molecules, morphology, and ecology indicate a recent, amphibious ancestry for echidnas. Proc. Natl. Acad. Sci. USA *106*, 17089–17094.

23. Luo, Z.-X., Yuan, C.-X., Meng, Q.-J., and Ji, Q. (2011). A Jurassic eutherian mammal and divergence of marsupials and placentals. Nature *476*, 442–445.

24. Domazet-Loso, T., and Tautz, D. (2008). An ancient evolutionary origin of genes associated with human genetic diseases. Mol. Biol. Evol. *25*, 2699–2707.

25. Neme, R., and Tautz, D. (2013). Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics *14*, 117.

26. Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., et al. (2018). Chromatin accessibility landscape in human early embryos and its association with evolution. Cell *173*, 248–259.e15.

27. Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M.K., Turon, X., López-Legentil, S., Piette, J., Lemaire, P., and Douzery, E.J.P. (2018). A phylogenomic framework and time-scale for comparative studies of tunicates. BMC Biol. *16*, 39.

28. Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl. Nucleic Acids Res. *46*, D754–D761.

29. Ye, T., Krebs, A.R., Choukrallah, M.-A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. *39*, e35–e35.

30. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

31. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

32. Hormozdiari, F., Gazal, S., van de Geijn, B., Finucane, H.K., Ju, C.J., Loh, P.R., Schoech, A., Reshef, Y., Liu, X., O'Connor, L., et al. (2018). Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nat. Genet. *50*, 1041–1047.

33. Gazal, S., Sahbatou, M., Babron, M.-C., Génin, E., and Leutenegger, A.L. (2015). High level of inbreeding in final phase of 1000 Genomes Project. Sci. Rep. *5*, 17453.

34. Palmer, C., and Pe'er, I. (2017). Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. PLoS Genet. *13*, e1006916.

35. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; and Genome Institute at Washington University (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

36. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.

37. McVicker, G., Gordon, D., Davis, C., and Green, P. (2009). Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. *5*, e1000471.

38. Maruyama, T. (1974). The age of a rare mutant gene in a large population. Am. J. Hum. Genet. *26*, 669–673.

39. Smith, A.V., Thomas, D.J., Munro, H.M., and Abecasis, G.R. (2005). Sequence features in regions of weak and strong linkage disequilibrium. Genome Res. *15*, 1519–1534.

40. Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.

41. Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S.A., Oddson, A., Másson, G., Holm, H., Kong, A., Thorsteinsdottir,

U., Sulem, P., et al. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nat. Genet. *48*, 314–317.

42. Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. *104*, 65–75.

43. Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. *10*, e1004722.

44. Chen, W., McDonnell, S.K., Thibodeau, S.N., Tillmans, L.S., and Schaid, D.J. (2016). Incorporating functional annotations for fine-mapping causal variants in a bayesian framework using summary statistics. Genetics *204*, 933–958.

45. Shi, J., Park, J.-H., Duan, J., Berndt, S.T., Moy, W., Yu, K., Song, L., Wheeler, W., Hua, X., Silverman, D., et al.; MGS (Molecular Genetics of Schizophrenia) GWAS Consortium; GECCO (The Genetics and Epidemiology of Colorectal Cancer Consortium); GAME-ON/TRICL (Transdisciplinary Research in Cancer of the Lung) GWAS Consortium; PRACTICAL (PRostate cancer AssoCiation group To Investigate Cancer Associated aLterations) Consortium; PanScan Consortium; and GAME-ON/ELLIPSE Consortium (2016). Winner's curse correction and variable thresholding improve performance of polygenic risk modeling based on genome-wide association study summary-level data. PLoS Genet. *12*, e1006493.

46. Hu, Y., Lu, Q., Powles, R., Yao, X., Yang, C., Fang, F., Xu, X., and Zhao, H. (2017). Leveraging functional annotations in genetic risk prediction for human complex diseases. PLoS Comput. Biol. *13*, e1005589.

47. Marquez-Luna, C., Gazal, S., Loh, P.-R., Furlotte, N., Auton, A., 23andMe Research Team, and Price, A. (2018). Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. bioRxiv. https://doi.org/10.1101/375337.

48. Lee, D., Karchin, R., and Beer, M.A. (2011). Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res. *21*, 2167–2180.

49. Ghandi, M., Lee, D., Mohammad-Noori, M., and Beer, M.A. (2014). Enhanced regulatory sequence prediction using gapped k-mer features. PLoS Comput. Biol. *10*, e1003711.

50. Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nat. Methods *12*, 931–934.

51. Whitaker, J.W., Chen, Z., and Wang, W. (2015). Predicting the human epigenome from DNA motifs. Nat. Methods *12*, 265–272, 7, 272.

52. Kelley, D.R., Snoek, J., and Rinn, J.L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res. *26*, 990–999.

53. Hashimoto, T., Sherwood, R.I., Kang, D.D., Rajagopal, N., Barkal, A.A., Zeng, H., Emons, B.J., Srinivasan, S., Jaakkola, T., and Gifford, D.K. (2016). A synergistic DNA logic predicts genome-wide chromatin accessibility. Genome Res. *26*, 1430–1440.

54. Kelley, D.R., Reshef, Y.A., Bileschi, M., Belanger, D., McLean, C.Y., and Snoek, J. (2018). Sequential regulatory activity prediction across chromosomes with convolutional neural networks. Genome Res. *28*, 739–750.

55. Zhou, J., Theesfeld, C.L., Yao, K., Chen, K.M., Wong, A.K., and Troyanskaya, O.G. (2018). Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. Nat. Genet. *50*, 1171–1179.

56. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021.

57. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., Robinson, M.R., Perry, J.R., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al.; LifeLines Cohort Study (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120.

58. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S., Balding, D.J.; and UCLEB Consortium (2017). Reevaluation of SNP heritability in complex human traits. Nat. Genet. *49*, 986–992.

59. Gazal, S., Marquez-Luna, C., Finucane, H.K., and Price, A.L. (2018b). Reconciling S-LDSC and LDAK models and functional enrichment estimates. bioRxiv. https://doi.org/10.1101/256412.

Supplemental Data

Disease Heritability Enrichment of Regulatory

Elements Is Concentrated in Elements with Ancient

Sequence Age and Conserved Function across Species

Margaux L.A. Hujoel, Steven Gazal, Farhad Hormozdiari, Bryce van de Geijn, and Alkes L. Price

Figure S1: **Correlation heat map.** Correlation of the annotations presented in this paper with all functional annotations in baseline-LD model (among common SNPs, MAF $\geq$ 0.05). Numerical values found in Table S1
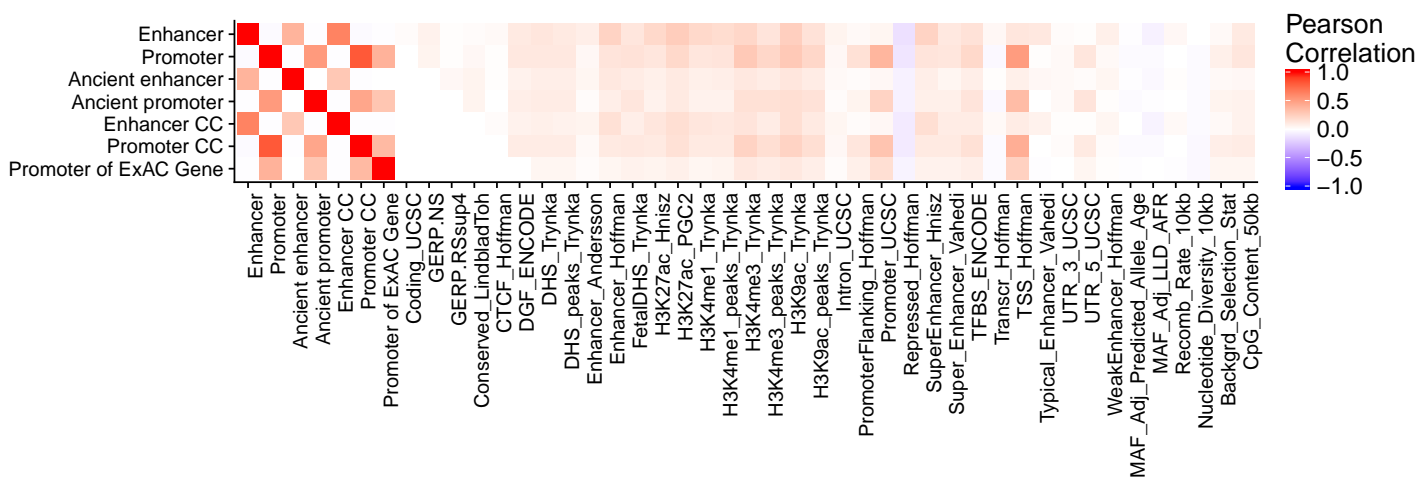
Table S1: **Correlation between main annotations and functional annotations of baseline-LD model.** We report correlation of annotations with all functional annotations in baseline-LD model (among common SNPs, MAF ≥ 0.05).

See attached Excel file

Annotation	Source	Model
All SNPs	Finucane et al.[1]	baseline/baseline-LD
Coding	UCSC Genome Browser[2], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
3' UTR	UCSC Genome Browser[2], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
5' UTR	UCSC Genome Browser[2], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
Promoter	UCSC Genome Browser[2], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
Intron	UCSC Genome Browser[2], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
Transcription factor binding site	ENCODE[4], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
Digital genomic footprint	ENCODE[4], post-processed by Gusev et al.[3], Finucane et al.[1]	baseline/baseline-LD
CTCF	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Promoter-Flanking	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Transcribed	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Transcription start site (TSS)	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Enhancer	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Weak enhancer	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
Repressed	Hoffman et al[5] and Finucane et al.[1]	baseline/baseline-LD
DHS	ENCODE[4]/Roadmap[6] post-processed by Trynka et al[7] and Finucane et al.[1]	baseline/baseline-LD
Fetal DHS	ENCODE[4]/Roadmap[6] post-processed by Trynka et al[7] and Finucane et al.[1]	baseline/baseline-LD
H3K4me1	Roadmap Epigenomics[6], post-processed by Trynka et al[7], Finucane et al.[1]	baseline/baseline-LD
H3K4me3	Roadmap Epigenomics[6], post-processed by Trynka et al[7], Finucane et al.[1]	baseline/baseline-LD
H3K9ac	Roadmap Epigenomics[6], post-processed by Trynka et al[7], Finucane et al.[1]	baseline/baseline-LD
H3K27ac	Roadmap Epigenomics[6], post-processed[8], Finucane et al.[1]	baseline/baseline-LD
H3K27ac	Hnisz et al[9] and Finucane et al.[1]	baseline/baseline-LD
Super-enhancer	Hnisz et al[9] and Finucane et al.[1]	baseline/baseline-LD
Conserved	Lindblad-Toh et al.[10], post-processed by Ward and Kellis[11], Finucane et al.[1]	baseline/baseline-LD
FANTOM5 enhancer	Anderson et al.[12] and Finucane et al.[1]	baseline/baseline-LD
Super-enhancer	Vahedi et al.[13], Gazal et al.[14]	baseline/baseline-LD
Typical enhancer	Vahedi et al.[13], Gazal et al.[14]	baseline/baseline-LD
GERP[15] conserved	GERP++score[15] and Gazal et al.[14]	baseline/baseline-LD
10 MAF bins	Gazal et al.[14]	baseline-LD
Predicted Allele Age	Gazal et al.[14]	baseline-LD
LLD-AFR	Gazal et al.[14]	baseline-LD
Recombination Rate	Gazal et al.[14]	baseline-LD
Nucleotide Diversity	Gazal et al.[14]	baseline-LD
Background Selection Statistics	Gazal et al.[14]	baseline-LD
CpG-Content	Gazal et al.[14]	baseline-LD

Table S2: **75 functional annotations in baseline-LD model.** 59 annotations define the baseline model: 1 annotation for all SNPs, 27 binary functional annotations and one continuous annotation (there exist 2 GERP conserved annotations, one binary and one continuous)[1,14], for each binary annotation (excluding the binary GERP related annotation) a 500-bp window around the category is added; for DHS, H3K4me1, H3K4me3, and H3K9ac, a 100-bp window around the peaks is also added as an additional annotation. The baseline-LD model is the baseline model with 16 additional annotations (ten MAF bins and 6 LD-related annotations)[14].

Table S3: **List of 47 datasets analyzed in this study.** We meta-analyzed all results across a previously chosen collection of 47 datasets[16]. We obtained the summary statistics for each trait from previous published studies where the summary statistics are publicly available. In the case of UK Biobank traits, we computed the summary statistics using BOLT-LMM[17,18]. For some traits we have more than one dataset, thus we have 41 independent traits. However, we utilized all the 47 datasets in our meta-analyses as the number of samples that overlap is low. These traits have been selected based on a heritability z-score > 6.

See attached Excel file

Table S4: **List of all models analyzed in this study.** We report the set of annotations included in each model analyzed.

See attached Excel file

Table S5: **Enhancer and promoter results.** (A) Meta-analysis results from including the annotations for putative enhancer and promoter conditional on baseline-LD. Both putative enhancers and promoters are enriched for heritability, however only promoter provides unique information for trait heritability conditional on the baseline-LD model. (B) Meta-analysis results from including the annotations for highly reproducible putative enhancer and highly reproducible promoter conditional on baseline-LD. Results using highly reproducible elements are similar as those found using all putative enhancers and promoters.

A Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.61(0.24)	2.5e-12	0.008(0.032)	0.79
Promoter	0.015	4.605(0.429)	3.2e-17	0.114(0.043)	0.0074
B Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Highly reproducible enhancer	0.022	2.61(0.249)	3.8e-10	-0.004(0.026)	0.86
Highly reproducible promoter	0.013	4.655(0.444)	4.4e-16	0.118(0.041)	0.0044

Table S6: **Sequence age annotations.** Meta-analysis results from models with each sequence age annotation conditioned on the baseline-LD model and putative enhancer and promoter annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^*	τ^* p-val
Enhancer	0.033	2.584(0.238)	5.6e-12	0.241(0.058)	3.4e-05
Promoter	0.015	4.653(0.432)	1.5e-17	0.127(0.043)	0.003
Young enhancer	0.014	-0.134(0.188)	1.6e-05	-0.333(0.06)	2.5e-08
Enhancer	0.033	2.581(0.239)	6e-12	0.013(0.032)	0.68
Promoter	0.015	4.709(0.436)	1.3e-17	0.462(0.07)	3.7e-11
Young promoter	0.005	-1.395(0.433)	0.0003	-0.484(0.058)	1e-16
Enhancer	0.033	2.61(0.239)	2e-12	0(0.041)	1
Promoter	0.015	4.612(0.43)	3.1e-17	0.114(0.043)	0.0073
Intermediate enhancer	0.013	2.256(0.36)	0.015	-0.094(0.034)	0.0058
Enhancer	0.033	2.613(0.24)	2.3e-12	0.01(0.032)	0.75
Promoter	0.015	4.632(0.43)	3.3e-17	0.228(0.061)	0.00017
Intermediate promoter	0.005	1.742(0.426)	0.26	-0.207(0.041)	3.6e-07
Enhancer	0.033	2.741(0.248)	1.3e-13	-0.184(0.029)	3e-10
Promoter	0.015	4.611(0.429)	2.5e-17	0.135(0.043)	0.0016
Ancient enhancer	0.005	10.102(0.9)	3.1e-19	0.483(0.06)	8.7e-16
Enhancer	0.033	2.57(0.238)	7.3e-12	0.021(0.032)	0.52
Promoter	0.015	4.923(0.446)	6.9e-19	-0.225(0.04)	1.4e-08
Ancient promoter	0.004	15.078(1.219)	4.3e-24	0.76(0.072)	3.6e-26

Table S7: **Joint sequence age model.** Meta-analysis results of joint model with sequence age annotations, the baseline-LD model and putative enhancer and promoter annotations. Only significant sequence age annotations (Bonferroni $p = 0.05/4 = 0.0125$) were retained in the model. (A) Enrichment for annotations as well as the difference of enrichments between annotations and the relevant element annotation (Δ). (B) τ^* value and significance. (C) Measures of negative selection on sequence age annotations. These measures (calculated on common SNPs within the annotation) include: the proportion of GERP RS≥ 4 [14,15], the mean of the background selection statistic (BSS) [14,19], the proportion conserved [1,10], the mean of the MAF adjusted predicted allele age (age) [14], and the mean of nucleotide diversity (10kb) (Nuc. Div.) [14,20]. Standard error is calculated via block-jacknife.

A Category	Prop.	Enr (s.e.)	Enr p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	2.692(0.246)	4.9e-13		
Promoter	0.015	4.903(0.445)	8.3e-19		
Ancient enhancer	0.005	9.283(0.867)	3.4e-17	6.578(0.685)	4.3e-15
Ancient promoter	0.004	14.277(1.179)	1.2e-22	9.176(0.769)	2.1e-18
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	-0.156(0.03)	1.6e-07			
Promoter	-0.188(0.039)	1.1e-06			
Ancient enhancer	0.429(0.058)	1.2e-13			
Ancient promoter	0.704(0.068)	9e-25			
C Category	GERP RS ≥ 4	BSS	Conserved	Age	Nuc. Div.
Enhancer	0.012(1.0e-05)	0.21(1.7e-05)	0.039(2.5e-05)	-0.031(9.9e-05)	4.5(3.3e-04)
Promoter	0.014(2.7e-05)	0.3(5.5e-05)	0.071(5.5e-05)	-0.15(2.1e-04)	4.2(8.6e-04)
Ancient enhancer	0.058(1.8e-04)	0.24(1.4e-04)	0.16(2.3e-04)	-0.13(6.0e-04)	4.1(1.8e-03)
Ancient promoter	0.041(1.5e-04)	0.33(2.0e-04)	0.18(3.2e-04)	-0.29(8.4e-04)	3.9(2.9e-03)

Table S8: **Joint sequence age model enrichment per trait.** Enrichment (standard error) for putative enhancer, promoter, ancient putative enhancer, and ancient promoter for each trait from the joint model with sequence age annotations, the baseline-LD model and putative enhancer and promoter annotations.

See attached Excel file

Table S9: **Joint sequence age model in liver traits only.** Meta-analysis results across liver traits of joint model with sequence age annotations, the baseline-LD model and putative enhancer and promoter annotations. (A) Enrichment for annotations as well as the difference of enrichments between annotations and the relevant element annotation (Δ). (B) τ^* value and significance.

A Category	Prop.	Enr (s.e.)	Enr. p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	6.561(0.977)	0.0015		
Promoter	0.015	17.779(2.833)	4e-06		
Ancient enhancer	0.005	8.441(4.118)	0.12	1.663(4.105)	0.79
Ancient promoter	0.004	38.264(6.649)	0.0029	20.414(5.706)	0.021
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	1.016(0.28)	0.00029			
Promoter	1.485(0.473)	0.0017			
Ancient enhancer	0.137(0.338)	0.68			
Ancient promoter	1.94(0.539)	0.00032			

Table S10: **Sequence age model with ancient sequence age annotation included.** Meta-analysis results from adding the annotation of ancient sequence age (irrespective of putative enhancer/promoter status) to the joint model with sequence age annotations. Regions with ancient sequence age are enriched for heritability, although a portion of this enrichment is explained by ancient putative enhancers and promoters there is still other genomic regions with ancient sequence age that contribute to this enrichment. However, we see that conditional on baseline-LD, putative enhancer, promoter, ancient putative enhancer, and ancient promoter the annotation for ancient sequence age does not provide novel information as quantified via τ^*.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.0330	2.695(0.246)	4.3e-13	-0.142(0.03)	1.9e-06
Promoter	0.0150	4.908(0.445)	7.2e-19	-0.177(0.038)	4.2e-06
Ancient enhancer	0.0050	9.256(0.869)	3.6e-17	0.425(0.06)	1.7e-12
Ancient promoter	0.0040	14.283(1.181)	1.2e-22	0.693(0.071)	2.6e-22
Ancient sequence	0.1040	3.375(0.153)	7.2e-25	0.062(0.051)	0.22

Table S11: **Ancient sequence age results in non-liver specific putative regulatory elements.**
Meta-analysis results of the model with putative regulatory annotations restricted to regions with ancient
sequence age from the baseline-LD model (not liver specific). We find the same main result that putative
enhancers and promoters with ancient sequence age are enriched for trait heritability.

Category	Prop. SNPs	Enrichment (s.e.)	Enrichment_pval
CTCF_Hoffman	0.024	0.384(0.206)	0.035
CTCF_Hoffman.ancient	0.003	4.74(0.703)	0.0011
DGF_ENCODE	0.136	2.551(0.235)	8.8e-13
DGF_ENCODE.ancient	0.024	8.733(0.768)	4.1e-28
DHS_peaks_Trynka	0.111	2.233(0.228)	7.1e-09
DHS_peaks_Trynka.ancient	0.025	7.193(0.709)	8.7e-23
DHS_Trynka	0.166	1.942(0.172)	8.1e-09
DHS_Trynka.ancient	0.036	6.139(0.534)	1.9e-24
Enhancer_Andersson	0.004	3.209(0.873)	0.039
Enhancer_Andersson.ancient	0.001	21.732(2.903)	1.1e-10
Enhancer_Hoffman	0.042	3.029(0.226)	7.1e-14
Enhancer_Hoffman.ancient	0.008	12.197(0.921)	1.9e-26
FetalDHS_Trynka	0.084	2.718(0.215)	6e-13
FetalDHS_Trynka.ancient	0.023	7.203(0.61)	2.7e-23
H3K27ac_Hnisz	0.389	1.639(0.068)	5e-40
H3K27ac_Hnisz.ancient	0.051	5.707(0.29)	9.7e-36
H3K27ac_PGC2	0.269	1.918(0.1)	6.2e-25
H3K27ac_PGC2.ancient	0.041	6.81(0.386)	5.4e-36
H3K4me1_peaks_Trynka	0.170	2.274(0.102)	4.8e-16
H3K4me1_peaks_Trynka.ancient	0.030	7.75(0.466)	7.9e-30
H3K4me1_Trynka	0.424	1.755(0.069)	5.7e-29
H3K4me1_Trynka.ancient	0.066	5.13(0.252)	7.1e-33
H3K4me3_peaks_Trynka	0.042	2.991(0.237)	6.2e-09
H3K4me3_peaks_Trynka.ancient	0.010	12.926(0.849)	7.4e-27
H3K4me3_Trynka	0.133	2.571(0.158)	6.2e-24
H3K4me3_Trynka.ancient	0.026	8.987(0.535)	1.3e-37
H3K9ac_peaks_Trynka	0.038	3.708(0.281)	2.2e-10
H3K9ac_peaks_Trynka.ancient	0.009	13.217(0.851)	5e-26
H3K9ac_Trynka	0.125	2.62(0.201)	1.2e-22
H3K9ac_Trynka.ancient	0.027	8.384(0.539)	3.7e-36
PromoterFlanking_Hoffman	0.008	1.084(0.516)	1
PromoterFlanking_Hoffman.ancient	0.001	12.866(1.248)	2.9e-10
Promoter_UCSC	0.046	2.118(0.172)	5.4e-08
Promoter_UCSC.ancient	0.011	7.809(0.492)	8.2e-22
SuperEnhancer_Hnisz	0.167	2.047(0.098)	1.2e-45
SuperEnhancer_Hnisz.ancient	0.023	7.204(0.401)	1.4e-34
Super_Enhancer_Vahedi	0.021	3.122(0.228)	3.1e-21
Super_Enhancer_Vahedi.ancient	0.003	8.55(0.866)	1.6e-14
TFBS_ENCODE	0.131	2.65(0.199)	9.6e-18
TFBS_ENCODE.ancient	0.024	9.162(0.655)	5.9e-33
Transcr_Hoffman	0.346	1.166(0.033)	0.00015
Transcr_Hoffman.ancient	0.037	4.266(0.202)	9.4e-20
TSS_Hoffman	0.018	5.741(0.492)	6.2e-19
TSS_Hoffman.ancient	0.005	17.709(1.271)	9e-31
Typical_Enhancer_Vahedi	0.022	2.193(0.188)	1.4e-08
Typical_Enhancer_Vahedi.ancient	0.003	8.206(0.866)	1.1e-10
WeakEnhancer_Hoffman	0.021	2.307(0.23)	4.2e-06
WeakEnhancer_Hoffman.ancient	0.004	11.015(0.941)	1.9e-16

Table S12: **Roadmap H3K27ac and H3K4me3 annotations** Meta-analysis results of the model with H3K27ac and H3K4me3 annotations from Roadmap[6] based on marks present in 1/10/20 tissues/cell-types, respectively, as well as restrincting to regions with ancient sequence age.

Category	Prop. SNPs	Enrichment (s.e.)	Enrichment_pval
H3K4me3 $\geq$ 1 tissue-type	0.200	1.851(0.071)	1.6e-15
H3K4me3 $\geq$ 1 tissue-type $\cap$ ancient seq. age	0.033	7.695(0.425)	1.4e-35
H3K4me3 $\geq$ 10 tissue-types	0.029	4.883(0.312)	3e-23
H3K4me3 $\geq$ 10 tissue-types $\cap$ ancient seq. age	0.009	13.752(0.77)	3.2e-32
H3K4me3 $\geq$ 20 tissue-types	0.019	6.368(0.435)	1.1e-22
H3K4me3 $\geq$ 20 tissue-types $\cap$ ancient seq. age	0.007	16.261(0.915)	1.2e-31
H3K27ac $\geq$ 1 tissue-type	0.325	1.843(0.089)	1.2e-25
H3K27ac $\geq$ 1 tissue-type $\cap$ ancient seq. age	0.051	6.231(0.335)	1.9e-36
H3K27ac $\geq$ 10 tissue-types	0.058	3.631(0.277)	7.5e-26
H3K27ac $\geq$ 10 tissue-types $\cap$ ancient seq. age	0.014	11.52(0.834)	4.5e-35
H3K27ac $\geq$ 20 tissue-types	0.026	4.612(0.393)	2.3e-20
H3K27ac $\geq$ 20 tissue-types $\cap$ ancient seq. age	0.007	14.388(1.078)	7.2e-31

Table S13: **Sequence age model with flanking regions.** To guard against bias due to model mis-specification[1], we repeated our sequence age analysis by including 500bp flanking regions around putative enhancer, promoter, and sequence age annotations. In order to facilitate the creation of 500 BP flanking regions we made bed files for the sequence age annotations which were not needed to make the annotations analyzed in the main analyses. The difference of the proportion of SNPs within main analysis annotations and those created using bed files was $\leq 2 * 10^{-6}$.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.266(0.211)	2.5e-08	-0.175(0.033)	1.3e-07
Enhancer flanking	0.010	2.012(0.551)	0.089	0.04(0.049)	0.42
Promoter	0.015	4.933(0.457)	3.2e-17	-0.298(0.041)	3.3e-13
Promoter flanking	0.004	0.175(0.862)	0.41	-0.126(0.053)	0.018
Ancient enhancer	0.005	8.814(0.795)	3.5e-15	0.416(0.055)	3e-14
Ancient enhancer flanking	0.009	1.251(0.351)	0.49	-0.007(0.033)	0.84
Ancient promoter	0.004	12.71(1.137)	6.6e-18	0.673(0.067)	7.4e-24
Ancient promoter flanking	0.005	3.877(0.502)	4.1e-05	0.265(0.043)	7.2e-10

Table S14: **Conserved function annotations.** Meta-analysis results from models with each conserved function annotation conditioned on the baseline-LD model and putative enhancer and promoter annotations. Conservation of regulatory function is assessed using 9 other mammalian species. Binary annotations for enhancers and promoters include highly conserved and human-specific reflecting 9 and 0 other species with function at the given regulatory element, respectively. Categorical annotations include conservation count (CC), mapped count (mappedC), and missing count (missingC) reflecting the number of other mammals for which the element is functionally conserved, the underlying sequence of the element aligns but its function is not conserved, and the underlying sequence of the element does not align (is missing), respectively.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.608(0.24)	2.6e-12	0.008(0.032)	0.81
Promoter	0.015	4.602(0.429)	3.6e-17	0.086(0.041)	0.038
Highly conserved promoter	0.002	5.051(0.637)	4.1e-07	0.017(0.02)	0.39
Enhancer	0.033	2.558(0.237)	8.1e-12	-0.018(0.031)	0.56
Promoter	0.015	4.59(0.428)	4e-17	0.112(0.043)	0.0087
Highly conserved enhancer	0.001	5.142(0.727)	7.8e-06	0.053(0.015)	0.0003
Enhancer	0.033	2.61(0.24)	2.6e-12	0.009(0.032)	0.78
Promoter	0.015	4.602(0.43)	3.6e-17	0.135(0.045)	0.0028
Human-specific promoter	0.003	3.024(0.448)	0.0002	-0.042(0.019)	0.029
Enhancer	0.033	2.55(0.236)	1.1e-11	0.107(0.041)	0.0095
Promoter	0.015	4.549(0.426)	6.9e-17	0.109(0.042)	0.01
Human-specific enhancer	0.013	0.91(0.184)	0.46	-0.186(0.033)	1.4e-08
Enhancer	0.033	2.608(0.24)	2.8e-12	0.009(0.032)	0.77
Promoter	0.015	4.599(0.43)	4.9e-17	-0.043(0.047)	0.35
Promoter CC	N/A	N/A	N/A	0.125(0.033)	0.00014
Enhancer	0.033	2.442(0.23)	1.7e-10	-0.182(0.027)	2.2e-11
Promoter	0.015	4.504(0.424)	1.3e-16	0.107(0.042)	0.011
Enhancer CC	N/A	N/A	N/A	0.212(0.032)	2.1e-11
Enhancer	0.033	2.608(0.24)	2.6e-12	0.008(0.032)	0.79
Promoter	0.015	4.604(0.43)	4e-17	0.112(0.047)	0.016
Promoter mappedC	N/A	N/A	N/A	-0.019(0.019)	0.32
Enhancer	0.033	2.584(0.239)	4.7e-12	0.034(0.045)	0.45
Promoter	0.015	4.585(0.424)	3.4e-17	0.113(0.042)	0.0076
Enhancer mappedC	N/A	N/A	N/A	-0.032(0.034)	0.35
Enhancer	0.033	2.61(0.24)	2.4e-12	0.009(0.032)	0.78
Promoter	0.015	4.575(0.429)	6.4e-17	0.188(0.048)	8.8e-05
Promoter missingC	N/A	N/A	N/A	-0.108(0.029)	0.00017
Enhancer	0.033	2.651(0.243)	1.3e-12	0.098(0.039)	0.011
Promoter	0.015	4.63(0.431)	2.1e-17	0.115(0.043)	0.0072
Enhancer missingC	N/A	N/A	N/A	-0.143(0.024)	5e-09

Table S15: **Joint conserved function model.** Meta-analysis results from joint model with functional count annotations, the baseline-LD model and putative enhancer and promoter annotations. Only significant functional count annotations (Bonferroni $p = 0.05/8 = 0.00625$) were retained in the model. (A) Enrichment for annotations as well as the difference of enrichments between annotations and the relevant element annotation (Δ). The enrichment for the categorical annotations enhancer conservation count (enhancer CC) and promoter conservation count (promoter CC) is shown for a subset of the annotation, namely conserved putative enhancers and promoters (CC $\geq$ 5), but is calculated using the model with conservation count variables. (B) τ^* value and significance. (C) Measures of negative selection on regulatory function annotations. These measures (calculated on common SNPs within the annotation) include: the proportion of GERP RS≥ 4[14,15], the mean of the background selection statistic (BSS)[14,19], the proportion conserved[1,10], the mean of the MAF adjusted predicted allele age (age)[14], and the mean of nucleotide diversity (10kb) (Nuc. Div.)[14,20]. Standard error is calculated via block-jacknife. For categorical annotations (enhancer CC and promoter CC) these measures are shown for conserved enhancers and promoters (CC $\geq$ 5).

A Category	Prop.	Enr (s.e.)	Enr p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	2.448(0.23)	1.5e-10		
Promoter	0.015	4.508(0.425)	1.7e-16		
Conserved enhancer	0.005	4.572(0.429)	3.5e-17	2.054(0.243)	2.6e-12
Conserved promoter	0.008	5.065(0.474)	6.4e-16	0.559(0.133)	0.022
B Category	τ^*	τ^* p-val			
Enhancer	-0.176(0.027)	7.1e-11			
Promoter	-0.016(0.048)	0.73			
Enhancer CC	0.201(0.031)	6.6e-11			
Promoter CC	0.092(0.033)	0.005			
C Category	GERP RS $\geq$ 4	BSS	Conserved	Age	NucDiv
Enhancer	0.012(1.0e-05)	0.21(1.7e-05)	0.039(2.5e-05)	-0.031(9.9e-05)	4.5(3.3e-04)
Promoter	0.014(2.7e-05)	0.3(5.5e-05)	0.071(5.5e-05)	-0.15(2.1e-04)	4.2(8.6e-04)
Conserved enhancer	0.017(7.9e-05)	0.24(1.3e-04)	0.055(1.5e-04)	-0.047(5.1e-04)	4(7.6e-04)
Conserved promoter	0.017(5.4e-05)	0.34(1.0e-04)	0.084(1.2e-04)	-0.15(4.1e-04)	3.8(6.6e-04)

Table S16: **Joint conserved function model enrichment per trait.** Enrichment (standard error) for putative enhancer, promoter, conserved putative enhancer, and conserved promoter for each trait from the joint model with putative enhancer conservation count, promoter conservation count, the baseline-LD model and putative enhancer and promoter annotations.

See attached Excel file

Table S17: **Binary conserved function model.** (A) Meta-analysis results of enrichment and τ^* for putative enhancers and promoters shared in CC other mammals (CC$\in \{0,\ldots,9\}$). (B) Measures of negative selection on common SNPs within the annotations: the proportion of GERP RS≥ 4[14,15], the mean of the background selection statistic (BSS)[14,19], the proportion conserved[1,10], the mean of the MAF adjusted predicted allele age (age)[14], and the mean of nucleotide diversity (10kb) (Nuc. Div.)[14,20]. Standard error is calculated via block-jacknife.

A Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enh CC=0	0.013(39.78%)	1.072(0.192)	0.98	-0.12(0.018)	5.2e-11
Enh CC=1	0.006(17.28%)	0.872(0.307)	0.47	-0.093(0.022)	2.6e-05
Enh CC=2	0.004(11.86%)	1.575(0.331)	0.22	-0.045(0.016)	0.005
Enh CC=3	0.003(8.55%)	2.446(0.438)	0.011	0.007(0.02)	0.7
Enh CC=4	0.002(6.03%)	1.684(0.362)	0.66	-0.03(0.016)	0.066
Enh CC=5	0.002(5.08%)	1.939(0.377)	0.073	-0.023(0.016)	0.14
Enh CC=6	0.001(3.48%)	4.15(0.701)	0.00024	0.047(0.022)	0.034
Enh CC=7	0.001(3.21%)	2.913(0.554)	0.029	0.002(0.016)	0.9
Enh CC=8	0.001(2.58%)	2.661(0.53)	0.015	-0.007(0.014)	0.58
Enh CC=9	0.001(2.15%)	4.388(0.676)	0.00016	0.031(0.016)	0.049
Pro CC=0	0.003(19.17%)	3.072(0.451)	0.00014	-0.001(0.023)	0.98
Pro CC=1	0.001(7.65%)	1.292(0.49)	0.62	-0.04(0.018)	0.028
Pro CC=2	0.001(6.75%)	1.336(0.603)	0.7	-0.048(0.019)	0.013
Pro CC=3	0.001(7.05%)	2.343(0.498)	0.015	-0.024(0.016)	0.14
Pro CC=4	0.001(6.88%)	3.845(0.908)	0.013	0.014(0.027)	0.62
Pro CC=5	0.001(7.48%)	2.372(0.507)	0.0012	-0.027(0.018)	0.13
Pro CC=6	0.001(8.11%)	4.084(0.617)	0.00032	0.023(0.019)	0.22
Pro CC=7	0.001(9.25%)	4.481(0.525)	1.2e-05	0.04(0.019)	0.04
Pro CC=8	0.002(12.95%)	3.95(0.617)	0.00012	0.027(0.027)	0.3
Pro CC=9	0.002(14.72%)	4.451(0.598)	6.9e-06	0.038(0.025)	0.13

B Category	GERP RS ≥ 4	BSS	Conserved	Age	NucDiv
Enh CC=0	0.011(2.8e-05)	0.19(4.9e-05)	0.036(5.7e-05)	-0.033(2.5e-04)	5(1.7e-03)
Enh CC=1	0.0096(7.6e-05)	0.21(1.2e-04)	0.033(1.0e-04)	-0.02(5.4e-04)	4.4(1.1e-03)
Enh CC=2	0.011(1.2e-04)	0.22(1.6e-04)	0.038(1.7e-04)	-0.02(8.2e-04)	4.3(1.4e-03)
Enh CC=3	0.012(1.5e-04)	0.22(2.4e-04)	0.038(2.3e-04)	-0.016(1.2e-03)	4.3(2.1e-03)
Enh CC=4	0.013(1.8e-04)	0.21(3.2e-04)	0.045(2.8e-04)	-0.043(1.4e-03)	4.3(2.4e-03)
Enh CC=5	0.013(1.4e-04)	0.23(3.7e-04)	0.043(3.7e-04)	-0.022(1.8e-03)	4.1(2.9e-03)
Enh CC=6	0.017(3.8e-04)	0.26(6.4e-04)	0.058(7.4e-04)	-0.054(2.7e-03)	4(3.9e-03)
Enh CC=7	0.015(3.8e-04)	0.25(7.3e-04)	0.053(4.6e-04)	-0.028(3.1e-03)	4.1(4.7e-03)
Enh CC=8	0.02(4.7e-04)	0.23(8.8e-04)	0.063(6.3e-04)	-0.056(4.2e-03)	4.1(6.5e-03)
Enh CC=9	0.024(7.6e-04)	0.25(1.1e-03)	0.07(1.2e-03)	-0.11(4.1e-03)	3.8(6.3e-03)
Pro CC=0	0.014(1.1e-04)	0.28(2.8e-04)	0.066(2.8e-04)	-0.11(1.2e-03)	5.1(8.5e-03)
Pro CC=1	0.0059(3.2e-04)	0.23(5.6e-04)	0.042(5.4e-04)	-0.11(2.7e-03)	5.7(2.7e-02)
Pro CC=2	0.0091(2.8e-04)	0.26(8.2e-04)	0.054(6.4e-04)	-0.16(3.4e-03)	4.5(1.4e-02)
Pro CC=3	0.012(3.4e-04)	0.27(7.7e-04)	0.06(5.8e-04)	-0.17(3.0e-03)	4(4.2e-03)
Pro CC=4	0.01(1.1e-17)	0.28(8.7e-04)	0.057(7.0e-04)	-0.19(2.9e-03)	3.9(4.3e-03)
Pro CC=5	0.011(5.2e-04)	0.27(6.8e-04)	0.067(8.1e-04)	-0.13(2.9e-03)	4.1(4.7e-03)
Pro CC=6	0.014(2.3e-04)	0.28(6.1e-04)	0.074(6.8e-04)	-0.14(2.3e-03)	3.9(3.8e-03)
Pro CC=7	0.015(3.5e-04)	0.32(5.7e-04)	0.08(6.3e-04)	-0.15(2.2e-03)	4(4.1e-03)
Pro CC=8	0.018(1.9e-04)	0.34(4.4e-04)	0.087(4.6e-04)	-0.17(1.6e-03)	3.6(2.6e-03)
Pro CC=9	0.023(2.4e-04)	0.4(4.1e-04)	0.097(4.1e-04)	-0.16(1.3e-03)	3.5(2.0e-03)

Table S18: **Binary conserved function model enrichment per trait.** Enrichment (standard error) for putative enhancers and promoters shared in CC other mammals (CC $\in \{0, \ldots, 9\}$) for each trait from the model with 20 binary annotations reflecting putative enhancers and promoters shared in CC other mammals and the baseline-LD model.

See attached Excel file

Table S19: **Joint conserved function model in liver traits only.** Meta-analysis results across liver traits from joint model with functional count annotations, the baseline-LD model and putative enhancer and promoter annotations. (A) Enrichment for annotations as well as the difference of enrichments between annotations and the relevant element annotation (Δ). The enrichment for the categorical annotations enhancer conservation count (enhancer CC) and promoter conservation count (promoter CC) is shown for a subset of the annotation, namely conserved putative enhancers and promoters (CC ≥ 5), but is calculated using the model with conservation count variables. (B) τ^* value and significance.

A Category	Prop.	Enr (s.e.)	Enr. p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	6.268(0.959)	0.0023		
Promoter	0.015	16.327(2.796)	3.8e-07		
Conserved enhancer	0.005	11.846(2.236)	0.00044	5.827(1.777)	0.0017
Conserved promoter	0.008	19.652(3.802)	5.2e-05	2.119(1.765)	0.38
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	0.506(0.208)	0.015			
Promoter	1.758(0.526)	0.00083			
Enhancer CC	0.733(0.226)	0.0012			
Promoter CC	0.549(0.43)	0.2			

Table S20: **Joint conserved function model with binary conserved annotations.** Meta-analysis results of conserved function model with putative enhancer and promoter conservation counts coded as binary annotations reflecting conserved function.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.474(0.232)	8.8e-11	-0.097(0.028)	0.00049
Promoter	0.015	4.503(0.426)	1.8e-16	0.019(0.049)	0.7
Conserved enhancer	0.005	4.483(0.435)	2.9e-14	0.134(0.022)	8e-10
Conserved promoter	0.008	4.981(0.454)	2.3e-15	0.083(0.028)	0.0027

Table S21: **Conserved function model with flanking regions** To guard against bias due to model misspecification[1], we repeated our conserved function analysis with binary conserved function annotations (Table S20) by including 500bp flanking regions around putative enhancer, promoter, conserved putative enhancer, and conserved promoter annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	1.876(0.189)	8.1e-05	-0.137(0.025)	8.1e-08
Enhancer flanking	0.010	2.818(0.606)	0.0056	0.045(0.046)	0.33
Promoter	0.015	4.417(0.431)	1.4e-14	0.054(0.049)	0.26
Promoter flanking	0.004	0.488(0.853)	0.58	-0.136(0.066)	0.041
Conserved enhancer	0.005	3.393(0.352)	6.2e-08	0.1(0.02)	7e-07
Conserved enhancer flanking	0.001	5.037(1.565)	0.063	0.06(0.041)	0.15
Conserved promoter	0.008	4.703(0.47)	2.9e-12	0.049(0.032)	0.12
Conserved promoter flanking	0.002	1.452(1.263)	0.57	0.1(0.071)	0.16

Table S22: **Conserved function model with human-specific promoter annotation added.** Meta-analysis results of adding an annotation for human-specific promoters to our regulatory function model. The value of τ^* for the promoter conservation count annotation became larger and more statistically significant.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.447(0.23)	1.5e-10	-0.176(0.027)	7.8e-11
Promoter	0.015	4.505(0.425)	2e-16	-0.088(0.057)	0.12
Enhancer CC	N/A	N/A	N/A	0.203(0.031)	7.9e-11
Promoter CC	N/A	N/A	N/A	0.138(0.043)	0.0014
Human-specific promoter	0.003	3.113(0.451)	0.00011	0.043(0.026)	0.095

Table S23: **Gene function annotations.** Meta-analysis results from models with each gene function annotation conditioned on the baseline-LD model and putative enhancer and promoter annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.599(0.24)	3.6e-12	0.008(0.032)	0.79
Promoter	0.015	4.779(0.439)	1.3e-17	-0.027(0.042)	0.53
Promoter of ancient gene	0.011	5.24(0.494)	8.2e-16	0.148(0.039)	0.00012
Enhancer	0.033	2.579(0.24)	6.6e-12	0.01(0.032)	0.75
Promoter	0.015	5.076(0.448)	8.7e-20	0.013(0.04)	0.74
Promoter of ExAC gene	0.002	12.365(0.894)	9.3e-22	0.371(0.031)	1.5e-32
Enhancer	0.033	2.598(0.239)	3.4e-12	0.009(0.032)	0.78
Promoter	0.015	4.747(0.437)	1.9e-17	-0.042(0.041)	0.31
Promoter of gene with mouse ortholog	0.011	5.132(0.486)	1.5e-15	0.168(0.041)	4e-05

Table S24: **Joint gene function model.** Meta-analysis results of joint model with gene function annotations, the baseline-LD model and putative enhancer and promoter annotations. Only significant gene functions annotations (Bonferroni $p = 0.05/3 = 0.0167$) were retained in the model. (A) Enrichment for annotations as well as the difference of enrichments between promoter of ExAC gene and (all) promoter annotation (Δ). (B) τ^* value and significance. (C) Measures of negative selection on gene function annotations. These measures (calculated on common SNPs within the annotation) include: the proportion of GERP RS≥ 4[14,15], the mean of the background selection statistic (BSS)[14,19], the proportion conserved[1,10], the mean of the MAF adjusted predicted allele age (age)[14], and the mean of nucleotide diversity (10kb) (Nuc. Div.)[14,20]. Standard error is calculated via block-jacknife.

A Category	Prop.	Enr (s.e.)	Enr p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	2.579(0.24)	6.6e-12		
Promoter	0.015	5.076(0.448)	8.7e-20		
Promoter of ExAC gene	0.002	12.365(0.894)	9.3e-22	6.805(0.491)	9.1e-16
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	0.01(0.032)	0.75			
Promoter	0.013(0.04)	0.74			
Promoter of ExAC gene	0.371(0.031)	1.5e-32			
C Category	GERP RS ≥ 4	BSS	Conserved	Age	Nuc. Div.
Enhancer	0.012(1.0e-05)	0.21(1.7e-05)	0.039(2.5e-05)	-0.031(9.9e-05)	4.5(3.3e-04)
Promoter	0.014(2.7e-05)	0.3(5.5e-05)	0.071(5.5e-05)	-0.15(2.1e-04)	4.2(8.6e-04)
Promoter of ExAC gene	0.024(1.9e-04)	0.38(3.4e-04)	0.12(4.0e-04)	-0.25(1.3e-03)	3.4(1.6e-03)

Table S25: **Joint gene function model enrichment per trait.** Enrichment (standard error) for putative enhancer, promoter, and promoter of ExAC gene for each trait from the joint model with promoter of ExAC gene, the baseline-LD model and putative enhancer and promoter annotations.

See attached Excel file

Table S26: **Joint gene function model in liver traits only.** Meta-analysis results across liver traits of joint model with gene function annotations, the baseline-LD model and putative enhancer and promoter annotations. (A) Enrichment for annotations as well as the difference of enrichments between promoter of ExAC gene and (all) promoter annotation (Δ). (B) τ^* value and significance.

A Category	Prop.	Enr (s.e.)	Enr. p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	6.653(1.003)	0.0015		
Promoter	0.015	17.585(2.789)	2.1e-06		
Promoter of ExAC gene	0.002	26.947(4.795)	0.0056	7.285(4.132)	0.18
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	1.03(0.225)	4.5e-06			
Promoter	2.189(0.471)	3.4e-06			
Promoter of ExAC gene	0.481(0.243)	0.048			

Table S27: **Gene function model with flanking regions.** To guard against bias due to model mis-specification[1], we repeated our gene function analysis by including 500bp flanking regions around putative enhancer, promoter, and promoter of ExAC gene annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.124(0.202)	4.8e-07	-0.043(0.028)	0.12
Enhancer flanking	0.010	2.19(0.568)	0.056	0.057(0.05)	0.26
Promoter	0.015	4.963(0.453)	2.1e-17	0.02(0.039)	0.61
Promoter flanking	0.004	0.614(0.886)	0.68	-0.099(0.055)	0.072
Promoter of ExAC gene	0.002	11.174(0.879)	7.4e-17	0.32(0.035)	9.3e-20
Promoter of ExAC gene flanking	0.001	4.045(2.031)	0.09	0.09(0.051)	0.075

Table S28: **Gene function model with functional annotations based on molecular quantitative trait loci.** Meta-analysis results when including the annotations for Max CPP (allGenes) and MaxCPP (ExAC), where these latter two annotations are found with a fixed effect meta-analysis (FE-Meta-Tissue) of GTEx data[16,21] to the gene function model. Even conditional on both Max CPP and MaxCPP of ExAC Genes we find that promoter of an ExAC LoF gene is still conditionally significant although the τ^* value is slightly smaller than when we exclude these two annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.6(0.24)	3.9e-12	0.017(0.032)	0.59
Promoter	0.015	4.605(0.434)	1.2e-16	0.002(0.039)	0.96
Promoter of ExAC gene	0.002	10.971(0.848)	7.1e-19	0.313(0.029)	3.7e-27
MaxCPP (allGenes)	N/A	N/A	N/A	0.406(0.047)	4.3e-18
MaxCPP (ExAC)	N/A	N/A	N/A	0.352(0.041)	1.3e-17

Table S29: **Final joint model.** Meta-analysis results of final model with regulatory and sequence conservation as well as gene function annotations, the baseline-LD model and putative enhancer and promoter annotations. (A) Enrichment for annotations and the difference of enrichments between annotations and (all) putative enhancer and (all) promoter annotations (Δ). The enrichment for putative enhancer conservation count (CC) is shown for a subset of the annotation, namely conserved putative enhancers (CC $\geq$ 5), but is calculated using the model with putative enhancer conservation count annotation. (B) τ^* value and significance. (C) Measures of negative selection on annotations within final joint model. These measures (calculated on common SNPs within the annotation) include: the proportion of GERP RS$\geq$ 4[14,15], the mean of the background selection statistic (BSS)[14,19], the proportion conserved[1,10], the mean of the MAF adjusted predicted allele age (age)[14], and the mean of nucleotide diversity (10kb) (Nuc. Div.)[14,20]. Standard error is calculated via block-jacknife. For the categorical annotation enhancer CC these measures are shown for conserved enhancers (CC $\geq$ 5).

A Category	Prop.	Enr (s.e.)	Enr p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	2.534(0.238)	3.1e-11		
Promoter	0.015	5.139(0.451)	5.1e-20		
Ancient enhancer	0.005	8.761(0.835)	4.9e-16	6.243(0.662)	2.7e-14
Ancient promoter	0.004	13.451(1.158)	5.1e-21	8.201(0.755)	5.2e-16
Conserved enhancer	0.005	4.614(0.433)	2.7e-17	2.012(0.244)	5.2e-12
Promoter of ExAC gene	0.002	11.958(0.872)	4.5e-21	6.39(0.481)	8.3e-15
B Category		τ^* (s.e.)	τ^* p-val		
Enhancer		-0.27(0.034)	2.6e-15		
Promoter		-0.216(0.038)	9.3e-09		
Ancient enhancer		0.388(0.055)	2.1e-12		
Ancient promoter		0.572(0.067)	1e-17		
Enhancer CC		0.161(0.028)	1.3e-08		
Promoter of ExAC gene		0.283(0.03)	2.3e-21		
C Category	GERP RS $\geq$ 4	BSS	Conserved	Age	Nuc. Div.
Enhancer	0.012(1.0e-05)	0.21(1.7e-05)	0.039(2.5e-05)	-0.031(9.9e-05)	4.5(3.3e-04)
Promoter	0.014(2.7e-05)	0.3(5.5e-05)	0.071(5.5e-05)	-0.15(2.1e-04)	4.2(8.6e-04)
Ancient enhancer	0.058(1.8e-04)	0.24(1.4e-04)	0.16(2.3e-04)	-0.13(6.0e-04)	4.1(1.8e-03)
Ancient promoter	0.041(1.5e-04)	0.33(2.0e-04)	0.18(3.2e-04)	-0.29(8.4e-04)	3.9(2.9e-03)
Conserved enhancer	0.017(7.9e-05)	0.24(1.3e-04)	0.055(1.5e-04)	-0.047(5.1e-04)	4(7.6e-04)
Promoter of ExAC gene	0.024(1.9e-04)	0.38(3.4e-04)	0.12(4.0e-04)	-0.25(1.3e-03)	3.4(1.6e-03)

Table S30: **Final joint model enrichment per trait.** Enrichment (standard error) for putative enhancer, promoter, conserved putative enhancer, ancient putative enhancer, ancient promoter, and promoter of ExAC gene for each trait from the joint model with ancient putative enhancer, putative enhancer conservation count, ancient promoter, promoter of ExAC gene, the baseline-LD model and putative enhancer and promoter annotations.

See attached Excel file

Table S31: **Final joint model in liver traits only.** Meta-analysis results across liver traits of final model with regulatory and sequence conservation as well as gene function annotations, the baseline-LD model and putative enhancer and promoter annotations. Only significant annotations (Bonferroni $p = 0.05/15 = 0.0033$) were retained in the model. (A) Enrichment for annotations and the difference of enrichments between annotations and (all) enhancer and (all) promoter annotations (Δ). The enrichment for putative enhancer conservation count (CC) is shown for a subset of the annotation, namely conserved putative enhancers (CC $\geq$ 5), but is calculated using the model with putative enhancer conservation count annotation. (B) τ^* value and significance.

A Category	Prop.	Enr (s.e.)	Enr. p-val	Δ (s.e.)	Δ p-val
Enhancer	0.033	6.065(0.939)	0.0024		
Promoter	0.015	17.823(2.791)	6e-06		
Ancient enhancer	0.005	6.403(4.166)	0.3	0.16(4.179)	0.93
Ancient promoter	0.004	36.732(6.644)	0.0028	19.001(5.745)	0.022
Conserved enhancer	0.005	11.584(2.177)	0.00065	5.785(1.751)	0.0016
Promoter of ExAC gene	0.002	26.32(4.812)	0.0061	6.389(4.107)	0.23
B Category	τ^* (s.e.)	τ^* p-val			
Enhancer	0.538(0.257)	0.037			
Promoter	1.442(0.471)	0.0022			
Ancient enhancer	-0.03(0.344)	0.93			
Ancient promoter	1.772(0.55)	0.0013			
Enhancer CC	0.748(0.227)	0.00098			
Promoter of ExAC gene	0.264(0.261)	0.31			

Table S32: **Final joint model with binary regulatory function annotations.** Meta-analysis results of final model with putative enhancer conservation count coded as a binary annotation reflecting conserved function.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.563(0.24)	2e-11	-0.212(0.031)	9.1e-12
Promoter	0.015	5.159(0.453)	4e-20	-0.215(0.038)	1.1e-08
Ancient enhancer	0.005	8.811(0.84)	5.8e-16	0.392(0.056)	1.7e-12
Ancient promoter	0.004	13.503(1.165)	4.2e-21	0.574(0.067)	1.2e-17
Conserved enhancer	0.005	4.398(0.432)	7.8e-14	0.092(0.018)	5.9e-07
Promoter of ExAC gene	0.002	11.991(0.875)	3.9e-21	0.283(0.03)	2.4e-21

Table S33: **Final joint model with flanking regions.** To guard against bias due to model misspecification[1], we repeated our final joint with binary conserved putative enhancer rather than putative enhancer conservation count including 500bp flanking regions around all introduced binary annotations (Table S32).

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.036(0.198)	4.2e-06	-0.219(0.034)	1.3e-10
Enhancer flanking	0.010	2.37(0.579)	0.029	0.017(0.044)	0.7
Promoter	0.015	5.085(0.458)	6.8e-18	-0.3(0.041)	4.5e-13
Promoter flanking	0.004	0.47(0.878)	0.6	-0.163(0.056)	0.0038
Ancient enhancer	0.005	8.38(0.753)	2.1e-14	0.385(0.052)	1.2e-13
Ancient enhancer flanking	0.009	1.06(0.348)	0.75	-0.027(0.034)	0.42
Ancient promoter	0.004	12.001(1.12)	1.3e-16	0.55(0.065)	4.4e-17
Ancient promoter flanking	0.005	4.192(0.501)	7e-06	0.232(0.043)	9.4e-08
Conserved enhancer	0.005	3.456(0.358)	3e-08	0.076(0.02)	0.00014
Conserved enhancer flanking	0.001	3.772(1.466)	0.17	0.04(0.041)	0.33
Promoter of ExAC gene	0.002	10.749(0.855)	3.5e-16	0.227(0.035)	8.6e-11
Promoter of ExAC gene flanking	0.001	4.378(2.029)	0.069	0.097(0.051)	0.056

Table S34: **Final joint model with additional promoter annotations.** Meta-analysis results are shown for the final joint model incorporating both promoter conservation count (CC) and human specific promoter annotations.

Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.535(0.238)	3e-11	-0.27(0.034)	2e-15
Promoter	0.015	5.148(0.453)	6.4e-20	-0.232(0.052)	6.4e-06
Ancient enhancer	0.005	8.782(0.834)	4e-16	0.39(0.055)	1.5e-12
Ancient promoter	0.004	13.452(1.16)	5.4e-21	0.576(0.067)	6.9e-18
Enhancer CC	N/A	N/A	N/A	0.161(0.028)	1.3e-08
Promoter CC	N/A	N/A	N/A	0.004(0.042)	0.93
Promoter of ExAC gene	0.002	11.971(0.874)	4.6e-21	0.286(0.03)	3.4e-21
Human-specific promoter	0.003	3.749(0.489)	3.9e-06	-0.007(0.026)	0.78

Table S35: **Final joint model with functional annotations based on molecular quantitative trait loci.** (A) Meta-analysis results when including the annotations for Max CPP (allGenes) and MaxCPP (ExAC), where these latter two annotations are found with a fixed effect meta-analysis (FE-Meta-Tissue) of GTEx data[16,21] to the final joint model. Even conditional on both Max CPP and MaxCPP of ExAC Genes we find that all our annotations remain conditionally significant. (B) Meta-analysis results when including the annotations for Max CPP (allGenes), MaxCPP (ExAC), as well as Max CPP (allGenes) restricted to regions with ancient sequence age to the final joint model. Conditional on these 3 additional annotations, we find that all our annotations remain conditionally significant. The conditional eQTL signal is concentrated in MaxCPP (allGenes) intersected with ancient sequence age.

A Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.547(0.238)	2.3e-11	-0.267(0.034)	7.8e-15
Promoter	0.015	4.666(0.437)	7.2e-17	-0.218(0.037)	5.2e-09
Ancient Enhancer	0.005	8.632(0.831)	1.2e-15	0.377(0.055)	6.7e-12
Ancient Promoter	0.004	12.536(1.128)	3.5e-19	0.547(0.066)	1.1e-16
Enhancer CC	N/A	N/A	N/A	0.169(0.029)	4.3e-09
Promoter of ExAC Gene	0.002	10.596(0.826)	3.2e-18	0.231(0.028)	5.1e-16
MaxCPP (allGenes)	N/A	N/A	N/A	0.393(0.046)	1.8e-17
MaxCPP (ExAC)	N/A	N/A	N/A	0.347(0.041)	2.2e-17
B Category	Prop.	Enr (s.e.)	Enr p-val	τ^* (s.e.)	τ^* p-val
Enhancer	0.033	2.55(0.238)	2.2e-11	-0.255(0.034)	9.5e-14
Promoter	0.015	4.668(0.437)	6.9e-17	-0.194(0.037)	1.3e-07
Ancient Enhancer	0.005	8.53(0.827)	2.4e-15	0.336(0.053)	1.9e-10
Ancient Promoter	0.004	12.37(1.128)	7.9e-19	0.479(0.066)	2.6e-13
Enhancer CC	N/A	N/A	N/A	0.17(0.029)	3.5e-09
Promoter of ExAC Gene	0.002	10.553(0.824)	4.1e-18	0.234(0.028)	1.4e-16
MaxCPP (allGenes)	N/A	N/A	N/A	0.17(0.044)	0.0001
MaxCPP (ExAC)	N/A	N/A	N/A	0.329(0.04)	4.1e-16
MaxCPP (allGenes) ∩ ancient sequence age	N/A	N/A	N/A	0.484(0.066)	1.5e-13

References

[1] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics, 47(11): 1228–1235, 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3404. URL http://www.nature.com/articles/ng.3404.

[2] W. James Kent, Charles W. Sugnet, Terrence S. Furey, et al. The human genome browser at UCSC. Genome Research, 12:996–1006, 2002.

[3] Alexander Gusev, S.A Hong Lee, Gosia Trynka, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. The American Journal of Human Genetics, 95(5):535–552, 2014. ISSN 00029297. doi: 10.1016/j.ajhg.2014.10.004. URL http://linkinghub.elsevier.com/retrieve/pii/S0002929714004261.

[4] The ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. Nature, 489:57–74, 2012.

[5] Michael M. Hoffman, Jason Ernst, Steven P. Wilder, et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Research, 41(2):827–841, 2013. ISSN 1362-4962, 0305-1048. doi: 10.1093/nar/gks1284. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks1284.

[6] Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, et al. Integrative analysis of 111 reference human epigenomes. Nature, 518(7539):317–330, 2015. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14248. URL http://www.nature.com/articles/nature14248.

[7] Gosia Trynka, Cynthia Sandor, Buhm Han, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nature Genetics, 45(2):124–130, 2013. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2504. URL http://www.nature.com/articles/ng.2504.

[8] Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511:421–427, 2014.

[9] Denes Hnisz, Brian J. Abraham, Tong Ihn Lee, et al. Super-enhancers in the control of cell identity and disease. Cell, 155(4):934–947, 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.09.053. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867413012270.

[10] Kerstin Lindblad-Toh, Manuel Garber, Or Zuk, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature, 478(7370):476–482, 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10530. URL http://www.nature.com/articles/nature10530.

[11] Lucas D. Ward and Manolis Kellis. Evidence of abundant purifying selection in humans for recently-acquired regulatory functions. Science, 337(6102):1675–1678, 2012.

[12] Robin Andersson, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, et al. An atlas of active enhancers across human cell types and tissues. Nature, 507(7493):455–461, 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12787. URL http://www.nature.com/articles/nature12787.

[13] Golnaz Vahedi, Yuka Kanno, Yasuko Furumoto, et al. Stretch-enhancers delineate disease-associated regulatory nodes in t cells. Nature, 520(7548):558–562, 2015.

[14] Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. Nature Genetics, 49(10):1421–1427, 2017. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3954. URL http://www.nature.com/doifinder/10.1038/ng.3954.

[15] Eugene V. Davydov, David L. Goode, Marina Sirota, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Computational Biology, 6(12): e1001025, 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1001025. URL http://dx.plos.org/10.1371/journal.pcbi.1001025.

[16] Farhad Hormozdiari, Steven Gazal, Bryce van de Geijn, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nature Genetics, 50(7):1041–1047, 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0148-2. URL http://www.nature.com/articles/s41588-018-0148-2.

[17] Po-Ru Loh, George Tucker, Brendan K. Bulik-Sullivan, et al. Efficient bayesian mixed model analysis increases association power in large cohorts. Nature Genetics, 47(3):284–290, 2015.

[18] Po-Ru Loh, Gleb Kichaev, Steven Gazal, et al. Mixed-model association for biobank-scale datasets. Nature Genetics, 50:906–911, 2018.

[19] Graham McVicker, David Gordon, Colleen Davis, and Phil Green. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genetics, 5(5):e1000471, 2009. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000471. URL http://dx.plos.org/10.1371/journal.pgen.1000471.

[20] Albert V. Smith, Daryl J. Thomas, Heather M. Munro, and Goncalo R. Abecasis. Sequence features in regions of weak and strong linkage disequilibrium. Genome Research, 15(11):1519–1534, 2005.

[21] GTEx consortium. Genetic effects on gene expression across human tissues. Nature, 550(7675):204–213, 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24277. URL http://www.nature.com/doifinder/10.1038/nature24277.