A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids

Ramdas, S., Judd, J., Graham, S. E., Kanoni, S., Wang, Y., Surakka, I., Wenz, B., Clarke, S. L., Chesi, A., Wells, A., Bhatti, K. F., Vedantam, S., Winkler, T. W., Locke, A. E., Marouli, E., Zajac, G. J. M., Wu, K.-H. H., Ntalla, I., Hui, Q., ... Million Veterans Program (2022). A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. *The American Journal of Human Genetics*, *109*(8), 1366-1387. https://doi.org/10.1016/j.ajhg.2022.06.012

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

1 **A multi-layer functional genomic analysis to understand noncoding genetic**
2 **variation in lipids**
3
4 Shweta Ramdas*, Jonathan Judd*, Sarah E Graham*, Stavroula Kanoni*, et al.

5

6 *equal contribution (co-first authors)

7

8 Corresponding Authors:

9 Xiang Zhu, PhD

10 Department of Statistics

11 Huck Institutes of the Life Sciences

12 The Pennsylvania State University

13 University Park, PA 16802

14 xiangzhu@psu.edu

15

16 Christopher D Brown, PhD

17 Department of Genetics

18 Perelman School of Medicine

19 University of Pennsylvania

20 Philadelphia PA 19104

21 chrbro@upenn.edu

22
23
24
25
26
27
28

29 **Abstract**

30 A major challenge of genome-wide association studies (GWAS) is to move from
31 phenotypic associations to biological insight. Here, we integrate a large trans-ancestry
32 GWAS on blood lipids with a wide array of functional genomic datasets to discover
33 regulatory mechanisms underlying lipid associations. We first identify lipid-associated
34 genes with expression quantitative trait locus - colocalizations, and then add chromatin
35 interaction data to narrow the search for functional genes. Polygenic enrichment
36 analysis across tissues and cell types confirms the central role of the liver in lipids, and
37 highlights the selective enrichment of adipose-specific chromatin marks in high-density
38 cholesterol and triglycerides. Overlapping transcription factor (TF) binding sites with
39 lipid-associated loci, we identify TFs relevant in lipid biology. Finally, we present an
40 integrative framework to prioritize causal variants at GWAS loci, producing a
41 comprehensive list of candidate causal genes and variants with multiple layers of
42 functional evidence. Two prioritized genes, *CREBRF* and *RRBP1,* show convergent
43 evidence across functional datasets supporting their roles in lipid biology.

44

45 **Introduction**

46

47 Most GWAS findings have not directly led to mechanistic interpretations, largely
48 because 90% of GWAS associations map to non-coding sequences [1,2]. Mechanistic
49 interpretations in GWAS have proven challenging because the strongest signals
50 identified in GWAS typically contain many variants in strong linkage disequilibrium (LD) [3]
51 and functional mechanisms including genes of action are often not clear from GWAS
52 data alone [4].

53

54 Linking trait-associated variants to genome function has emerged as a promising model
55 for mechanistic interpretation of non-coding findings in GWAS. This 'variant-to-function'
56 model is premised on recent observations that non-coding variants often affect a trait
57 of interest through the regulation of genes and processes in trait-relevant cell types or
58 tissues [2]. Implementing this functional model in GWAS has become more feasible as
59 large-scale functional genomic resources, such as epigenomic [5,6] and transcriptomic
60 catalogues [5] have been systematically generated across a wide range of human cell
61 types and tissues. Indeed, the integration of functional genomics with GWAS has
62 identified regulatory mechanisms in variants associated with obesity [7] and
63 schizophrenia [8], yielding important functional insights in genetic architecture of human
64 complex traits.

65

66 The history of the human genetics of lipids mirrors the success and challenge of GWAS.
67 Increasing sample size and genetic diversity has significantly boosted the power of
68 discovery: the first lipid GWAS in 2008 with 8,816 European-descent individuals

69   identified 29 lipid-associated loci[9] ; the latest study of 1.6 million individuals across five

70   ancestries [Graham et al 2021] found 941. Despite the dramatic increase in the number

71   of associations, our biological understanding of many of these genetic discoveries

72   remains limited. The causal gene has been confidently assigned at only a small fraction

73   of these loci [2], and the regulatory mechanism connecting variant to phenotype has been

74   conclusively characterized only for a handful of genes [10]. Furthermore, systematic

75   mapping of lipid-associated variants to their biological functions has been missing in

76   literature at the time of this study.

77

78   Here we conduct a genome-scale integrative analysis on the largest GWAS to-date of

79   five lipid phenotypes (LDL, or low density cholesterol; HDL, or high density cholesterol;

80   TC, or total cholesterol; nonHDL, or non-high density cholesterol; and TG, or

81   triglycerides) involving 1.6 million individuals from five ancestries [Graham et al 2021].

82   Combining the lipid GWAS with a wide array of functional genomic resources in diverse

83   human tissues and cell types, we identify regulatory mechanisms of noncoding genetic

84   variation in lipids with a full suite of computational approaches. Further, we develop a

85   generalizable framework to understand how tissue-specific gene regulation can explain

86   GWAS findings, and demonstrate its real-world value on lipid-associated loci.

87

88   **Results**

89   *Figure 1: Starting with GWAS summary statistics for five lipid phenotypes, we integrate these*

90   *with eQTL and chromatin interaction data to identify potential genes mediating the GWAS*

91   *association, and use epigenomic annotations from ChIP-seq and ATAC-seq data both to*

92   *identify regulatory mechanisms at these loci, and to arrive at genome-level insights into lipid*

93   *biology, such as tissue relevance.*

94
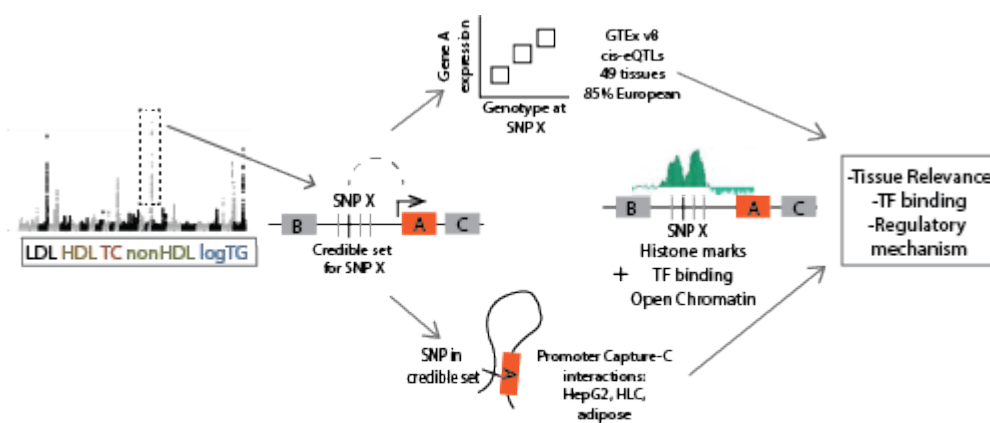
96   We systematically integrated lipid GWAS results [Graham et al 2021] with multiple layers

97   of functional genomic data from diverse tissues and cell types to understand regulatory

98   mechanisms at lipid-associated loci (Figure 1). Specifically, we overlaid GWAS loci with

99   expression quantitative trait loci (eQTL) and chromatin-chromatin interactions to

100   identify causal genes. We further assessed polygenic enrichments of tissue-specific

101   histone marks to prioritize relevant tissues and examined GWAS loci at transcription

102   factor (TF) binding sites to detect lipid-relevant TFs. Finally, we combined all these layers

103   to provide a holistic view of gene regulation at lipid loci in relevant tissue and cell types.

104

105   *Co-localization with eQTLs identifies candidate lipid-relevant genes*

106

107   We identify shared association signals between lipid levels and expression of nearby

108   genes as a first step given that most GWAS signals are presumed to influence complex

109   traits through their impact on gene expression [11]. To do so, we tested for colocalization

110   of each of the significant lipid GWAS association signals (1,750 loci considering the five

111   traits examined) with significant cis-eQTL data across 49 human tissues from the GTEx

112   consortium [5]. Here, we defined GWAS signals as 1,750 loci reaching genome-wide

113   significance and corrected for shadow signals (Methods) in a trans-ancestry meta-

114   analysis for at least one of five lipid traits.

115

116   We then restricted our analysis to those loci likely mediated through regulatory

117   mechanisms as opposed to coding variation. In particular, we excluded all loci with

118   credible sets containing at least one missense variant (369 of 1,750 loci, 21% of credible

119   sets (Paper2). Of the remaining 1,381 GWAS loci, 696 significantly colocalized with eQTLs

120   (the ratio of posterior probability of a shared signal to the posterior probability of two

121   signals being > 0.9 [12]; Methods) in at least one of 49 tissues for at least one lipid

122   phenotype. This resulted in 1,076 co-localized eGenes (an eGene is any gene with a

123   significant eQTL as defined by GTEx); (range of 1 to 16 genes per locus; Table S1). Since

124   with eQTL data alone it is difficult to disentangle a single functional gene from multiple

125   functional (and likely coregulated) genes at a locus [13] we performed all downstream

126   analyses with all 1,076 colocalized genes, to further prioritize functional genes at loci

127   with multiple eGenes.

128

129   To acquire additional functional insights of colocalized genes, we assessed their

130   enrichments across a wide range of existing biological and clinical gene sets. Colocalized

131   genes showed enrichments in 20 KEGG pathways [14] at FDR 5% (Table S2), including

132   known lipid-related processes such as cholesterol metabolism, PPAR signaling and bile

133   secretion. These genes were enriched in 33 Mendelian genes from ClinVar [15] associated

134   with lipid-related ICD codes (ICD code E78), (3.5 fold enrichment, including *APOB, LPL*,

135   and *APOE*), suggesting the shared genetic basis of Mendelian and complex lipid

136   phenotypes [16]. These genes were also enriched in 15 genes with rare-variant burden for

137   lipid phenotypes from UK Biobank (11-fold enrichment, including genes *APOB, LPL, LIPG*

138 and *ANGPTL4*; Figure 2), confirming shared mechanisms of rare and common variation
139 underlying lipid traits[16,17]. Altogether the results demonstrate the biological relevance of
140 candidate functional genes prioritized by our approach.
141
142 *Figure 2: Enrichment of eQTL/Capture-C overlap, and enrichment in 33 ClinVar gold standard*
143 *genes. A. Numbers of genes identified by two approaches: eqtl colocalization (top panel) and*
144 *promoter capture-c interactions (bottom panel) B. The intersection of eQTL and Capture-C*
145 *datasets shows an enrichment beyond what is expected by chance, assuming both genesets*
146 *are independent. C. The overlap between our list of prioritized genes with three sets of genes*
147 *previously associated with lipid biology. Capture-C prioritized genes (on the left) show no*
148 *enrichment, whereas colocalized genes (right) show a much higher overlap than expected by*
149 *chance.*
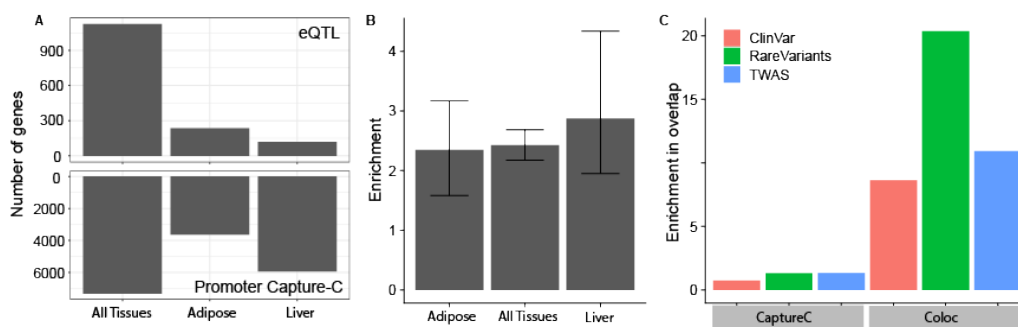
151
152 *Chromatin-chromatin interactions improve eQTL-based colocalization*
153
154 Our eQTL-based colocalization analysis uses a linear sequence of DNA, and ignores
155 physical interaction between non-adjacent DNA segments, another regulatory layer
156 underlying complex human traits [18]. To add this layer to our analysis, we generated
157 promoter-focused Capture-C (henceforth called Capture-C) data from HepG2 liver
158 carcinoma cells (denoted as HepG2.1) and hepatocyte-like cells (HLC) derived from
159 differentiating iPSCs (the latter is described in [19]), as well as publicly-available Capture-C
160 datasets from HepG2 [17,20] (denoted as HepG2.2) and adipose tissue [21]. We defined a
161 GWAS-relevant interaction as any Capture-C interaction between any gene and the 95%
162 credible set for a GWAS locus[22]. Credible set sizes ranged from 1 to 417 variants at the
163 1,750 examined loci, with a median size of 5 variants per credible set. In total, 1,079
164 GWAS loci had at least one variant in the credible set with a physical interaction with a
165 gene promoter and 3,543 of 26,621 genes with promoter-interactions had promoters
166 physically interacting with at least one GWAS credible set variant (Table S3). Unlike
167 eQTL-colocalized genes, these genes interacting with their credible sets showed limited
168 enrichment in relevant KEGG pathways and lipid-related genes from ClinVar (Figure 2B).
169 We observed a similar lack of enrichment when we restricted the physical interaction
170 analysis to protein-coding genes.
171

172   Genes physically interacting with GWAS loci helped shortlist functional genes from eQTL
173   colocalization despite their lack of enrichments in known gene sets. Of 1,079 credible
174   sets with promoter interactions, 224 also colocalized with eQTLs for the same gene
175   (Figure 2A). Among these loci with concordant eQTL colocalizations and Capture-C
176   interactions, only 39% of them mapped to a single gene using eQTL data alone, whereas
177   adding Capture-C information increased this fraction to 80%. At the gene level, 233
178   genes were implicated in both eQTL colocalization and Capture-C interactions,
179   representing an enrichment of 2.4. These results showcase the potential value of
180   combining eQTLs with physical chromatin interactions to prioritize functional genes at
181   GWAS loci.
182
183   Since eQTLs are likely to reside in the same TADs as the genes they regulate [23], we
184   examined topologically associated domain (TAD) structure from independent datasets
185   at lipids GWAS loci with eQTL colocalizations. Of eQTL-GWAS colocalizations in which the
186   sentinel variant resided within a liver TAD [24], the colocalized gene resided in the same
187   liver TAD 84.8% of the time (enrichment P < 0.001 with 1000 permutations; Methods).
188   When we restricted colocalizations to those supported by Capture-C data in any cell
189   type, 91.2% fall in the same TAD. These results add to the existing evidence for TAD
190   boundaries being regulatory insulators in the cell [cite a recent review] and confirm our
191   integration of chromatin interactions with eQTL colocalizations as an effective strategy
192   to hone in on functional genes.
193
194   *Tissue-specific enrichment of GWAS signals differentiates lipid traits*
195
196   Regulatory variants often affect complex traits in a tissue-specific manner [25], as shown
197   in our eQTL colocalization analysis. Specifically, by computing the ratio of the number of
198   colocalizations in a tissue to eQTL sample size in that tissue, we identified that the liver
199   was universally enriched for colocalized eGenes with respect to sample size across all
200   lipid traits whereas adipose was selectively enriched in HDL and TG only (Figure S1).
201   Motivated by these findings, we leveraged systematic approaches and additional data to
202   identify relevant tissues and cell types for each lipid trait.
203
204   We implemented stratified LD score regression (S-LDSC), a polygenic approach not
205   restricted only to genome-wide significant variants, on tissue-specific transcriptomic
206   and epigenomic annotations to identify relevant tissues for each lipid trait (Methods).
207   Consistent with previous studies and our eQTL-based analysis, liver-related tissues
208   (tissue-groupings are defined in Table S4) showed strong enrichments across all lipid
209   traits (S-LDSC enrichment p-values ranging from 1e-3 in TG to 1e-04 in TC), for both
210   expression (Fig 3A) and chromatin annotations (Figure 3B). This result was further
211   confirmed by two other approaches (DEPICT [26]: Figure S2; RSS-NET [27]: Table S5). To

212 assess the robustness of our S-LDSC results based on trans-ancestry GWAS data, we
213 applied S-LDSC to two population-specific GWAS in GLGC (European and East Asian)
214 together with population-specific LD scores (Methods), and we obtained similar results
215 (Table S6).
216
217 *Figure 3: Tissue relevance of lipid loci. Partitioning heritability of summary statistics on gene*
218 *expression (top panel) and active regulatory marks (bottom panel) across tissues. Each*
219 *plotted point represents each tested dataset for enrichment of heritability; multiple*
220 *annotation datasets are tested for the same tissue group. Each color represents a single*
221 *tissue group, and the y-axis represents P-value of enrichment of heritability. Liver-related*
222 *tissues, in teal, consistently show strongest enrichment of heritability.*
223
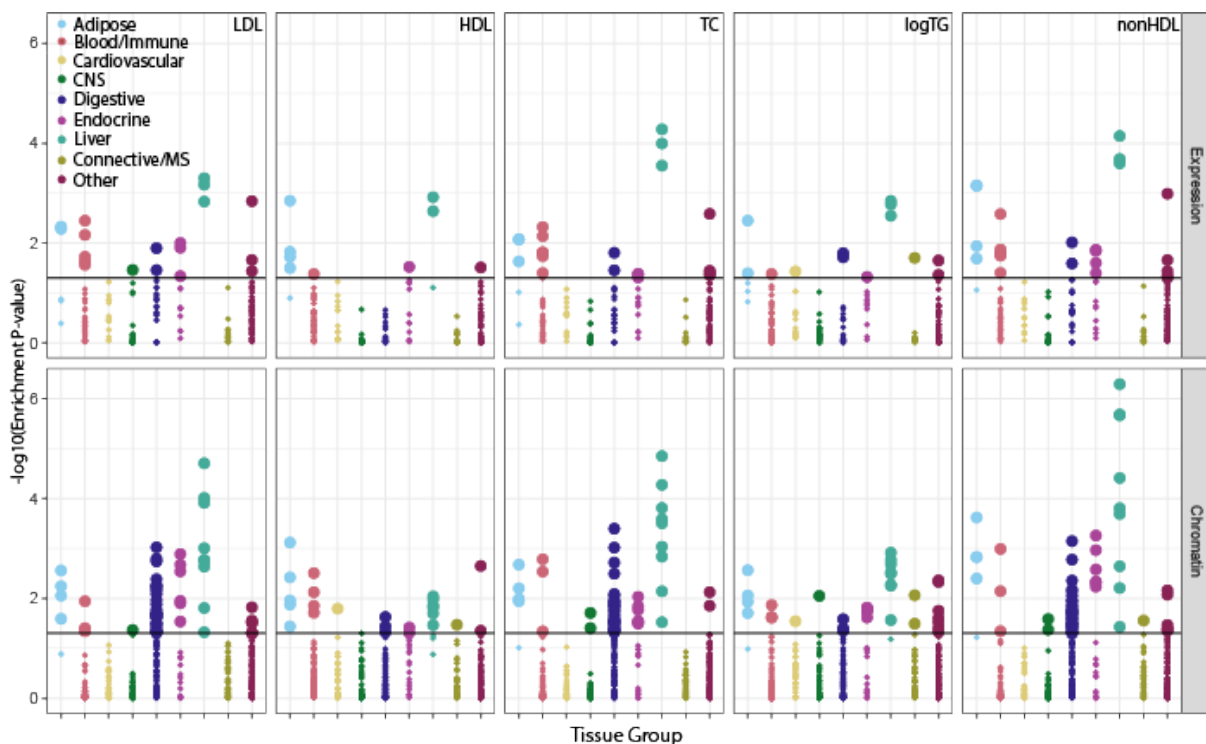
225
226
227
228 The S-LDSC results also highlighted tissues selectively enriched in certain lipid traits
229 similar to the eQTL-based analysis. The most enriched category for HDL using
230 chromatin annotation is 'Adipose H3K4me1'; for TG, enrichment in liver-related tissues
231 is similar to enrichment in adipose. For LDL, TC and non-HDL, enrichment P-values for
232 the liver were much more significant than for any other tissues (Figure 3B). We observed
233 the same pattern in S-LDSC results based on gene expression (Figure 3A). This finding is
234 consistent with the known influence of adipose on plasma HDL levels [28], and the role of
235 adipose as TG deposits. These results were corroborated by eQTL colocalizations
236 stratified by phenotype (Figure S1) and DEPICT analysis on gene expression [26] (Figure

237 S2). Together, these results strongly implicate the liver as the tissue of action for all five
238 lipid traits, but additionally implicate adipose tissue as playing a role in HDL and TG.
239
240 Given the importance of the liver and adipose in modulating lipid levels, we further
241 identified the relevant cell types within these tissues. Using published single-cell data
242 from adipose and liver, we performed gene-set enrichment analysis [29] to identify cell-
243 type clusters enriched for genes colocalized with any lipid trait. Out of 11 identified cell
244 types in 20 clusters in the liver, only hepatocytes were enriched at P < 0.05 (Figure S3),
245 consistent with previous results[20]. In adipose, only adipocyte clusters and macrophage-
246 monocyte clusters showed suggestive enrichment in colocalized genes (Figure S4). Of
247 note, the enrichment in adipocytes was significant when we restricted this analysis to
248 genes that were colocalized only with HDL and TG (FDR-corrected P < 0.05), consistent
249 with the selective enrichments of adipose in HDL and TG (but not the other lipid traits)
250 from our S-LDSC analysis. Evaluations at cellular resolution are required to understand
251 the cell-type specific mechanisms underlying lipid GWAS loci, but our results could form
252 a useful basis for future studies.
253
254 *Overlapping GWAS signals with binding sites highlights lipid-relevant TFs*
255
256 TFs have been implicated as a key mediator of linking genetic variation to complex traits
257 [30]. To understand lipid GWAS in the context of TF activity, we assessed enrichment of
258 genome-wide significant variants at TF binding sites using GREGOR [31] and performed
259 polygenic enrichment analysis of TF binding sites using S-LDSC.
260
261 Using ChIP-Seq data from 161 TFs across 91 cell types from the ENCODE project [6], 32.7%
262 of lipid credible sets overlapped with at least one TF binding site. Using GREGOR [31], we
263 identified 137 TFs whose binding sites were significantly enriched in GWAS lead SNPs for
264 at least one lipid phenotype (enrichment > 2; FDR adjusted P-value < 0.05, Figure S5,
265 Table S7). Among these 137 enriched TFs, 69 of them (50%) showed significant
266 enrichments across all five lipid phenotypes, suggesting a potential core regulatory
267 circuit shared by all lipid traits (Figure S5). The TF with the strongest enrichment in all
268 phenotypes was *ESRRA* (Estrogen-related receptor alpha), a nuclear receptor active in
269 metabolic tissues; ESRRA has been implicated in adipogenesis and lipid metabolism,
270 and ESRRA-null mice display an increase in fat mass and obesity [32].
271
272 The GREGOR analysis also highlighted 68 TFs significantly enriched in specific subsets of
273 (but not all five) lipid phenotypes (Figure S8). For example, we found 4 TFs (FOXM1,
274 PBX3, ZKSCAN1, ZEB1) enriched in HDL and TG only, 4 TFs (EZH2, NFE2, NFATC1,
275 KDM5A) enriched in HDL only and 11 TFs (FOSL1, IRF3, JUN, MEF2C, NANOG, PRDM1,
276 RUNX3, SIRT6, SMC3, STAT3, ZNF217) enriched in TG only. Of these TFs, the central role

277   of *ZEB1* in adiposity [33] and fat cell differentiation has been demonstrated [34]. Taken
278   together, these TF-centric findings corroborate the selective enrichments of adipose in
279   HDL and TG (but not the other lipid traits) identified in our previous tissue prioritization
280   analyses.
281
282   Similar to tissue prioritization, we also performed polygenic enrichment analysis of TF
283   binding sites using S-LDSC (Table S8), which differed from GREGOR analysis by looking
284   at not only the genome-wide significant associations but also the polygenic signal
285   irrespective of GWAS P-values. On the same 161 ENCODE TFs, this polygenic analysis
286   identified 25 TFs whose binding sites were significantly enriched in heritability for at
287   least one lipid phenotype (Fig S6), and reassuringly, 24 of 25 TFs are also significant in
288   GREGOR analysis. Among these enriched TFs, eight of them (34%) were significantly
289   enriched in all five lipid traits (CEBPB, CEBPD, FOXA2, HDAC2, HNF4G, NFYA, RXRA, SP1;
290   enrichment P < 0.05). Of those TFs significant in both analyses, Retinoid X receptor
291   alpha (*RXRA*) is also a colocalized gene near a GWAS hit (chr9:137,268,682). *RXRA* is a
292   ligand-activated transcription factor that forms heterodimers with other receptors
293   (including *PPARG*) and is involved in lipid metabolism [35] and homeostasis. While *RXRA*
294   has been implicated as the causal gene at its GWAS locus [36], our study is the first to
295   demonstrate its role in lipid biology through its regulatory influence on other lipid
296   GWAS genes.
297
298   *Multi-layer functional integration reveals regulatory mechanisms at GWAS loci*
299
300   Motivated by our finding that integrating chromatin interaction improved eQTL
301   colocalizations, we further collated multiple lines of functional evidence at each GWAS
302   locus for mechanistic inference. We started with the list of genes with evidence for both
303   eQTL colocalization in the liver or adipose and credible set physical interactions. We
304   next annotated each variant in the 95% credible set with various indicators of regulatory
305   function, including its open chromatin status in liver or adipose-related cell types, its
306   proximity to a promoter or an enhancer, and its RegulomeDB regulation probability [37]
307   (see Table S9 for the complete list of annotations used). To account for complexities of
308   regulatory mechanisms and limitations of functional datasets, we combined evidence
309   across these datasets to prioritize variants at GWAS loci (Figure 4A). Specifically, we
310   prioritized variants with at least three independent lines of functional evidence
311   (chromatin openness, physically interaction with target genes, and promoter/enhancer
312   status in liver or adipose), with at least two being in the same tissue with colocalization
313   with the target gene, and with a RegulomeDB score > 0.5. Applying this simple
314   procedure to lipid GWAS we identified 14 candidate loci, each with the strongest multi-
315   layer evidence pointing to a single functional variant (Table 1). Below we used *RRBP1*
316   and *CREBRF* to highlight key features of this multi-layer integration framework.

317

318 The first example *RRBP1* (Ribosomal binding protein 1) could be identified from eQTL

319 colocalization alone, but our multi-layer integration approach strengthened the

320 conclusion via convergent evidence from various sources (Figure 4B). The *RRBP1* eQTL

321 signals in the liver colocalize with LDL, TC and nonHDL. The 'T' allele of the single lead

322 variant (chr20:17,844,684, hg19) decreases *RRBP1* expression levels and increases LDL,

323 TC and nonHDL levels. This lead variant at *RRBP1* is in open chromatin in HLC, and also

324 physically interacts with the *RRBP1* promoter (250kb away) in adipose and HepG2. All

325 these data based on the lead variant consistently point to *RRBP1* as the functional gene

326 underlying this locus, which is further supported by external data. RRBP1 specifically

327 tethers the endoplasmic reticulum to the mitochondria in the liver--an interaction that is

328 enriched in hepatocytes--and regulates very low density lipoprotein (vLDL) levels [38]. Rare

329 variants in *RRBP1* are associated with LDL in humans [39] and silencing *RRBP1* in liver

330 affects lipid homeostasis in mice [38].

331

332 *Figure 4. A. Variant annotation and prioritization scheme at each credible set. B. Evidence for*

333 *gene RRBP1 from functional genomics data. The LDL GWAS locus at this region is an eQTL for*

334 *gene RRBP1 in the liver (second panel). Variants in the credible set of this locus interact with*

335 *the gene promoter in both adipose and HepG2 Capture-C data. The interacting variant is also*

336 *in an open chromatin peak in three liver-related cell types. C. Multiple sources of functional*

337 *genomics data support CREBRF as a gene contributing to HDL levels. The HDL GWAS locus at*

338 *this region is an eQTL for gene CREBRF in Adipose (second panel). Variants in the credible set*

339 *at this locus interact with the CREBRF promoter in adipose. The interacting variant is also in*

340 *open chromatin in liver-related cell types.*
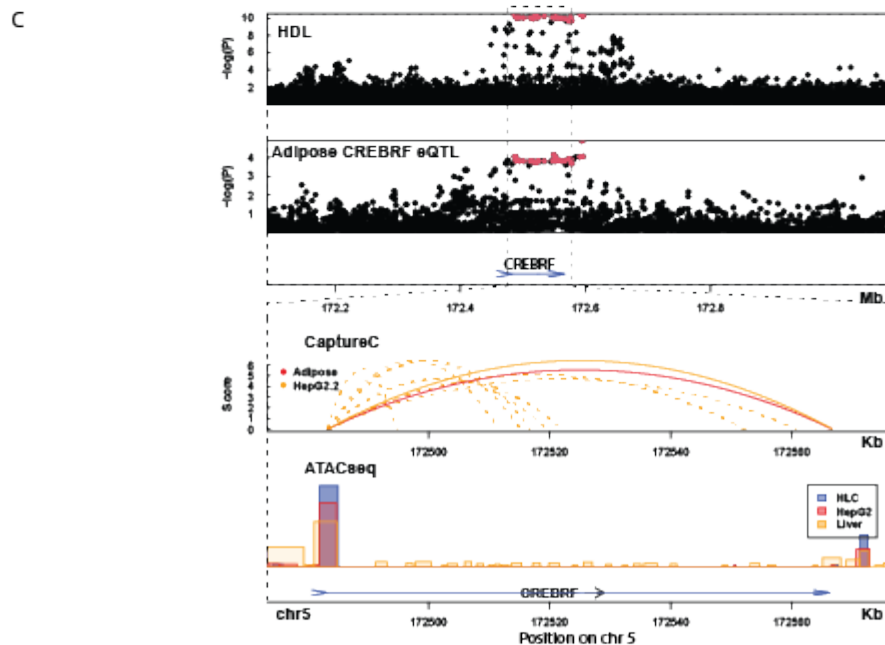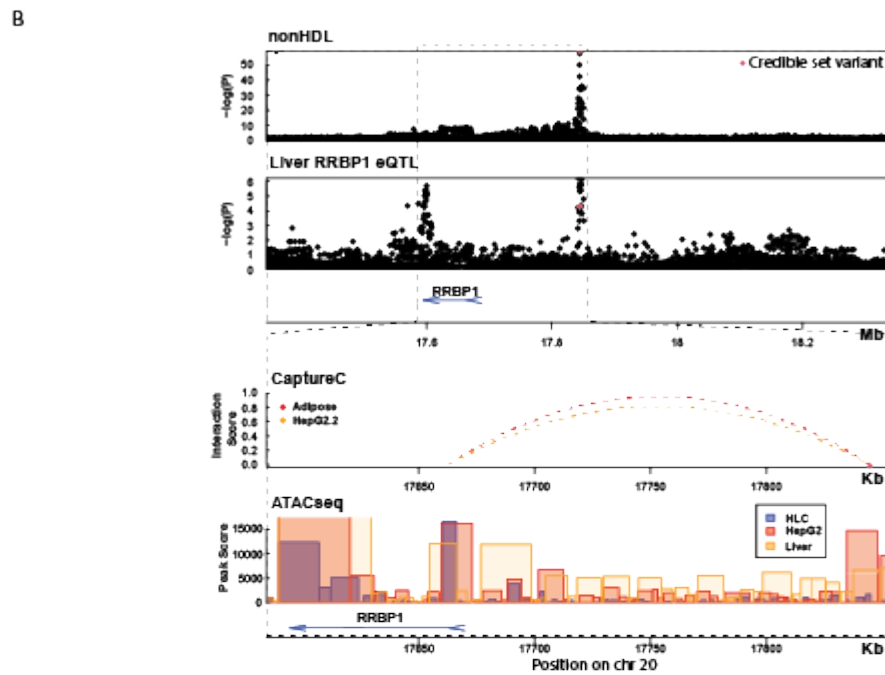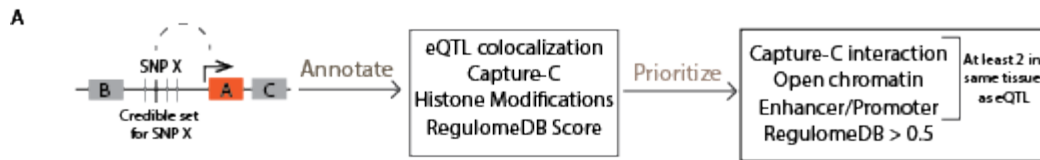
344   The second example *CREBRF* (CREB3 regulatory factor) demonstrates the power of our

345   multi-layer integration framework in prioritizing functional variants (Figure 4C). The

346   eQTL signals of *CREBRF* colocalized with a GWAS locus for HDL with 30 candidate

347   variants. In contrast, our multi-layer approach identified a single candidate variant

348   (chr5:172,566,698) at this locus that physically interacts with the *CREBRF* promoter in

349   adipose, was predicted to be a regulatory element (RegulomeDB score=0.91).

350   Consistent with the index variant (chr5:172,591,337), the allele 'A' at this functional

351   variant increased HDL levels and increased *CREBRF* expression in adipose. Missense

352   variants in *CREBRF* have been linked to body mass index, and the gene has been linked

353   to obesity risk in Samoans [40][41].

354

355   **Discussion**

356

357   Here we integrate the largest trans-ancestry lipid GWAS to date with a wide array of

358   functional genomic resources to understand how noncoding genetic variation affects

359   lipids through gene regulation. Specifically, we identify 1,076 genes whose eQTL signals

360   colocalize with lipid GWAS signals and demonstrate how physical chromatin interaction

361   can improve standard eQTL-based colocalization. We assess tissue-specific enrichments

362   of lipid GWAS signals and demonstrate the selective importance of adipose in HDL and

363   triglyceride biology. We examine binding site enrichments of 137 TFs in lipid GWAS and

364   expand our understanding of lipid GWAS loci (e.g., *RXRA*) in the context of TF activity.

365   Finally, we build a simple and interpretable prioritization framework that automatically

366   combines multiple lines of evidence from orthogonal datasets, pinpointing a single

367   functional variant at each of 14 lipid-associated loci (e.g., *RRBP1* and *CREBRF*). While

368   there are studies that interpret lipid GWAS associations [20,42,43], the size of our trans-

369   ancestry GWAS and multi-layer functional integration represent a comprehensive effort

370   and an important step forward in this direction.

371

372   Our multi-layer analysis has two key strengths. First, despite a large array of functional

373   genomic resources being embedded, our analysis produces results with high

374   consistency. For example, the selective enrichment of adipose in HDL and TG identified

375   by S-LDSC is confirmed by our eQTL-based colocalization and TF binding site overlap.

376   Another example is the prioritization of *RRBP1*, which can be identified from eQTL-based

377   colocalziation alone and it is further validated by chromatin openness and interaction.

378   Such convergent evidence from various sources improves the confidence of our

379   findings. Second, our analysis highlights that combining multiple layers of regulatory

380   information can improve sensitivity to prioritize functional genes and variants. For

381   example, we refined eQTL colocalized genes (1,076) to a smaller set of functional genes

382   (233) through integration with promoter Capture-C data. Another example is *CREBRF*,

383   where eQTL-based colocalization implicates 30 candidate variants and adding other

384   regulatory layers points to a single functional variant. Moving forward, we expect these

385   findings will serve as useful guidelines for future integrative genomic analyses of other

386   traits.

387

388 Our results rely on the breadth and accuracy of functional genomic datasets used in our
389 analyses. First, unlike our lipid GWAS, current functional datasets [44] are limited both in
390 sample size and ancestral diversity, which can affect discovery and replication of
391 regulatory mechanisms in diverse populations. Second, some functional datasets are
392 generated at limited resolution. For example, our colocalizations are based on eQTLs
393 from bulk tissue RNA-seq [5,45], which may miss detailed cell types and biological
394 processes in which lipid-associated SNPs regulate gene expression [45]. Third, some
395 functional datasets are not available across the full spectrum of human tissues and cell
396 types. For example, our chromatin-chromatin interaction analysis only examines a few
397 cell types in two known lipid-related tissues, producing results that may be biased
398 towards known lipid biology. As more comprehensive and accurate functional genomic
399 resources are becoming publicly available in diverse cellular contexts and ancestry
400 groups, the resolution and power of integrative analyses like ours we applied here will
401 be markedly increased.
402
403 Other limitations of this study stem from computational methods embedded in our
404 framework. First, the colocalization approach 'coloc' assumes one causal variant per
405 locus, whereas recent studies suggest extensive allelic heterogeneity [46] consistent with a
406 model of a milieu of related transcription factors binding within a single locus.
407 Accounting for allelic heterogeneity in summary statistics-based colocalization typically
408 requires modeling of LD matrix [47], which is computationally intensive in large-scale
409 analyses derived from many cohorts with diverse ancestries, like the trans-ancestry
410 GWAS examined here. Second, due to restricted access to individual genotypes of 201
411 cohorts, we cannot produce trans-ethnic LD scores within GLGC but have to use
412 European-based LD scores in all S-LDSC analyses. This approach, though less rigorous in
413 principle, provides robust results in practice (as confirmed by our ancestry-specific
414 analysis), largely because 79% of cohorts in GLGC are of European descent [Graham et
415 al 2021]. That said, we caution that the same approach might fall short in ancestrally
416 diverse studies with few European individuals [48]. Third, our multi-layer variant
417 prioritization framework is built on a series of simple rules that are easy to implement
418 on large datasets. This approach could possibly be formalized as statistical models (e.g.,
419 priors in Bayesian methods [27], but certainly simplify computation and improve
420 scalability of our framework. Despite the technical limitations, our approach here can
421 serve as a useful benchmark for future development of methods with improved
422 statistical rigor and computation efficiency.

423 In summary, mapping noncoding genetic variation of complex traits to biological
424 functions can benefit greatly from thorough integration of multiple layers of functional
425 genomics, as demonstrated in the present study. Although tested on lipids only, our
426 integrative framework is straightforward to implement more broadly on many other
427 phenotypes, likely yielding functional insights of heritable traits and diseases in humans.

428 **Methods**

429 *GWAS*

430

431 We performed GWAS for five blood lipid traits (LDL, HDL, TC, TG, and nonHDL) in 1.65
432 million individuals from five ancestry groups \cite{Graham2021} (AFR: African and
433 African-admixed; EAS: East Asian; EUR: European; HIS: Hispanic; SAS: South Asian; at 91
434 million variants imputed primarily from the Haplotype Reference Consortium or 1000
435 Genomes Phase 3. The individual GWAS and meta-analyses (described in Graham et al,
436 2021) were performed using the hg19 version of the human reference genome. We
437 used MR-MEGA [49] for meta-analysis across cohorts.

438

439 We defined 'sentinel variants' as lead variants representing independent trait-
440 associated loci in the genome. These windows are the greater of 500kb or 0.25cM
441 around the sentinel variant; genetic distances were defined using reference maps from
442 HapMap 3. We performed a second round of conditional analysis conditioning on the
443 lead variants to identify and remove any significant windows that are actually shadow
444 signals (or dependent on) of a neighboring locus to enforce true independence of
445 associated loci.

446

447 *Co-localization with gene expression*

448

449 We performed statistical colocalization with eQTLs obtained from GTEx v8. These
450 summary statistics were in GRCh38, so we first lifted over the GWAS summary statistics
451 (in hg19) from the trans-ethnic summary statistics to GRCh38 using UCSC liftOver
452 executable [50]. For each of the five lipid traits, we used the same 'sentinel variants'
453 defined in the previous section to represent approximately independent GWAS-
454 associated windows (also removing shadow signals as described before).

455

456 For each such window, we ran an eQTL colocalization using GTEx v8 eQTL summary
457 statistics [5]. For each of 49 GTEx tissues, we first identified all genes within 1Mb of the
458 sentinel SNP, and then restricted analysis to those genes with eQTLs ('eGenes') in that
459 tissue (FDR < 0.05). We used the R package 'coloc' (run on R version 3.4.3, coloc version
460 3.2.1) [51] with default parameters to run co-localization between the GWAS signal and the
461 eQTL signal for each of these cis-eGenes, using as input those SNPs in the defined

462   window, i.e. all SNPs present in both datasets. A colocalization posterior probability of
463   (PP3+PP4) > 0.8 was used to identify loci with enough colocalization power, and PP4/PP3
464   > 0.9 was used to define those loci that show significant colocalization, described
465   previously [12].
466
467   *Overlap with promoter Capture-C data*
468
469   We used four promoter Capture-C datasets from three cell/tissue types to capture
470   physical interactions between gene promoters and their regulatory elements. We
471   employed three biological replicates of HepG2 liver carcinoma cells [52], another HepG2
472   dataset described in Selvarajan et al [12,20], hepatocyte-like cells (HLC) produced by
473   differentiating three biological replicates of iPSCs, which in turn were generated from
474   peripheral blood mononuclear cells using a previously published protocol [19], and an
475   adipose dataset obtained from Pan et al [21] that was produced using primary human
476   white adipocytes.
477
478   The detailed protocol to prepare HepG2 or HLC cells for the promoter Capture-C
479   experiment is described in [52]. Briefly, for each dataset, 10 million cells were used for
480   promoter Capture-C library generation. Custom capture baits were designed using an
481   Agilent SureSelect library design targeting both ends of DpnII restriction fragments
482   encompassing promoters (including alternative promoters) of all human coding genes,
483   noncoding RNA, antisense RNA, snRNA, miRNA, snoRNA, and lincRNA transcripts,
484   totaling 36,691 RNA baited fragments shows the custom bait map boundaries in hg19
485   coordinates). Each library was then sequenced on an Illumina NovoSeq (HLC), or
486   Illumina HiSeq 4000 (HLC), generating 1.6 billion read pairs per sample (50 base pair
487   read length.) HiCUP [53] was used to process the raw FastQ files into loop calls; we then
488   used CHiCAGO [53,54] to define significant looping interactions; a default score of 5 was
489   defined as significant.
490
491   Starting with Capture-C maps processed as described above, we re-annotated the baits
492   to gene IDs from Gencode v19 [55] to ensure uniformity of gene annotations with the rest
493   of our pipeline. For each bait, we identified any gene whose transcription start site (TSS)
494   from any transcript in Gencode v19 was within 175 base pair distance from the bait to
495   account for differing bait designs for external datasets which may not directly overlap
496   canonical TSS. From GRanges (version 1.42.0 run on R 4.0.2)[56], the findOverlaps function
497   was used for this annotation with the 'maxgap' input set to 175. The final result
498   annotated each bait to a unique gene name and Ensembl ID. The Ensembl IDs were
499   formatted to remove the ID suffix, which included all numeric values after the ".". All
500   datasets were additionally filtered to only include interactions in which the interacting
501   end was not another bait.

502

Overlap between promoter Capture-C data and GWAS credible sets

503

504

505 To identify genetic variants associated with any of the five lipid traits that physically

506 interact with locations in the genome that may influence lipid biology, we continued to

507 use the R package 'Genomic Ranges' [56] to find overlap between previously defined

508 credible sets for each traits' GWAS and the previously annotated promoter Capture-C

509 data, which we refer as Capture-C/GWAS interactions. For all individual variants within

510 all GWAS-associated loci for the five lipid traits, we identified which variants overlapped

511 any interacting end of the four previously annotated promoter Capture-C data.

512

513 *Enrichment of colocalized gene-sentinel variant pairs in topologically associated domains*

514

515 To compute enrichment of colocalized gene-sentinel pairs in the same TAD, we used

516 publicly-available TADs from the liver [24]. We compared the number of colocalizations

517 with the sentinel variant and colocalized gene in the same TAD divided by all

518 colocalizations in which the sentinel variant lies in a TAD.

519

520 *Enrichment in single cell data from liver and adipose*

521

522 We overlapped our list of colocalized genes with publicly available single cell RNA-

523 sequencing data from cells in the liver [57] and 38,408 cells from the adipose [58]. For both

524 datasets, we downloaded normalized TPM data and existing tSNE cluster annotations

525 for each cell. For each cluster, we defined median expression for each gene across all

526 cells in that cluster. Then for each cluster, we calculated the enrichment P-value for our

527 list of colocalized genes using the 'fgsea' R package, which looks for overrepresentation

528 of our gene list in ranked genes for each cluster [59], implemented in R 3.4.3.

529

530 *Pathway Enrichment*

531

532 We used ClusterProfiler v3.6.0 [60] to look for pathways over-represented in each gene list

533 (genes with eQTL colocalization, and genes interacting with GWAS credible sets).

534 Specifically, we used the enrichKEGG function to look for pathway enrichment in KEGG

535 pathways. We first re-mapped gencode IDs to gene symbols using the Gencode v24

536 annotation and then used the biomaRt package [53,54,61] in R to convert gene symbols to

537 Entrez IDs, and then ran enrichKEGG to identify enriched pathways significant at a

538 Benjamini-Hochberg threshold of 0.05.

539

540 *Stratified LD score regression for prioritizing tissues*

541

542 We used LDSC version 1.0.1[62] to estimate the enrichment of heritability of our summary
543 statistics in different epigenetic and transcriptomic annotations (including gene
544 expression from GTEx, and histone epigenetic marks from Roadmap), using python
545 2.7.9. We first converted the summary statistics for each phenotype to LDSC-formatted
546 summary statistics using 'munge_sumstats.py'. Then, we used 'ldsc.py' using the
547 baseline_v1.2 baseline model and 'Multitissuechromatin1000Gv3' and
548 'Multitissuegeneexpr1000Gv3' annotations to estimate enrichment of heritability, using
549 active chromatin marks and gene expression regularly. Links to downloaded files are in
550 Supplementary Information. For primary analyses, we used trans-ethnic GWAS
551 summary statistics, and ld scores from 1000Genomes European samples.
552

References

554 1. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association
555 to Function. Am. J. Hum. Genet. *102*, 717–730.

556 2. Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional
557 Genomics to Identify the Mechanisms Underlying Complex Diseases. Front. Genet. *11*,
558 424.

559 3. Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to
560 candidate causal variants by statistical fine-mapping. Nat. Rev. Genet. *19*, 491–504.

561 4. Smemo, S., Tena, J.J., Kim, K.-H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas,
562 I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated
563 variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–
564 375.

565 5. Consortium, T.G., and The GTEx Consortium (2020). The GTEx Consortium atlas of
566 genetic regulatory effects across human tissues. Science *369*, 1318–1330.

567 6. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in
568 the human genome. Nature *489*, 57–74.

569 7. Loos, R.J.F., and Yeo, G.S.H. (2021). The genetics of obesity: from discovery to biology.
570 Nat. Rev. Genet.

571 8. Huo, Y., Li, S., Liu, J., Li, X., and Luo, X.-J. (2019). Functional genomics reveal gene
572 regulatory mechanisms underlying schizophrenia risk. Nat. Commun. *10*, 1–19.

573 9. Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C.,
574 Timpson, N.J., Najjar, S.S., Stringham, H.M., et al. (2008). Newly identified loci that
575 influence lipid concentrations and risk of coronary artery disease. Nat. Genet. *40*, 161–
576 169.

577 10. Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li,

578   X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2010). From noncoding variant to
579   phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714–719.

580   11. Neumeyer, S., Hemani, G., and Zeggini, E. (2020). Strengthening Causal Inference for
581   Complex Disease Using Molecular Quantitative Trait Loci. Trends Mol. Med. *26*, 232–241.

582   12. Çalışkan, M., Manduchi, E., Rao, H.S., Segert, J.A., Beltrame, M.H., Trizzino, M., Park,
583   Y., Baker, S.W., Chesi, A., Johnson, M.E., et al. (2019). Genetic and Epigenetic Fine
584   Mapping of Complex Trait Associated Loci in the Human Liver. Am. J. Hum. Genet. *105*,
585   89–107.

586   13. (2014). Obesity and FTO: Changing Focus at a Complex Locus. Cell Metab. *20*, 710–
587   718.

588   14. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG
589   as a reference resource for gene and protein annotation. Nucleic Acids Res. *44*, D457–
590   D462.

591   15. Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang,
592   W., Kaur, K., Liu, C., et al. (2020). ClinVar: improvements to accessing data. Nucleic Acids
593   Res. *48*, D835–D844.

594   16. Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H.,
595   Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S., et al. (2013). A nondegenerate
596   code of deleterious variants in Mendelian loci contributes to complex disease risk. Cell
597   *155*,.

598   17. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki,
599   M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical
600   and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

601   18. (2014). The 3D Genome in Transcriptional Regulation and Pluripotency. Cell Stem
602   Cell *14*, 762–775.

603   19. Pashos, E.E., Park, Y., Wang, X., Raghavan, A., Yang, W., Abbey, D., Peters, D.T.,
604   Arbelaez, J., Hernandez, M., Kuperwasser, N., et al. (2017). Large, Diverse Population
605   Cohorts of hiPSCs and Derived Hepatocyte-like Cells Reveal Functional Genetic Variation
606   at Blood Lipid-Associated Loci. Cell Stem Cell *20*, 558–570.e10.

607   20. Selvarajan, I., Toropainen, A., Garske, K.M., López Rodríguez, M., Ko, A., Miao, Z.,
608   Kaminska, D., Õunap, K., Örd, T., Ravindran, A., et al. (2021). Integrative analysis of liver-
609   specific non-coding regulatory SNPs associated with the risk of coronary artery disease.
610   Am. J. Hum. Genet. *108*, 411–430.

611   21. Pan, D.Z., Garske, K.M., Alvarez, M., Bhagat, Y.V., Boocock, J., Nikkola, E., Miao, Z.,
612   Raulerson, C.K., Cantor, R.M., Civelek, M., et al. (2018). Integration of human adipocyte
613   chromosomal interactions with adipose gene expression prioritizes obesity-related
614   genes from GWAS. Nat. Commun. *9*, 1512.

615 22. Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison
616 with P-values. Genet. Epidemiol. *33*, 79–86.

617 23. Yu, J., Hu, M., and Li, C. (2019). Joint analyses of multi-tissue Hi-C and eQTL data
618 demonstrate close spatial proximity between eQTLs and their target genes. BMC Genet.
619 *20*, 43.

620 24. Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.-A., Lin, S.,
621 Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes
622 across human tissues. Nature *518*, 350–354.

623 25. Hekselman, I., and Yeger-Lotem, E. (2020). Mechanisms of tissue and cell-type
624 specificity in heritable traits and diseases. Nat. Rev. Genet. *21*, 137–150.

625 26. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C.,
626 Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-
627 wide association studies using predicted gene functions. Nat. Commun. *6*, 5890.

628 27. Zhu, X., Duren, Z., and Wong, W.H. (2021). Modeling regulatory network topology
629 improves genome-wide analyses of complex human traits. Nat. Commun. *12*, 2851.

630 28. Zhang, T., Chen, J., Tang, X., Luo, Q., Xu, D., and Yu, B. (2019). Interaction between
631 adipocytes and high-density lipoprotein:new insights into the mechanism of obesity-
632 induced dyslipidemia and atherosclerosis. Lipids Health Dis. *18*, 223.

633 29. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
634 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment
635 analysis: a knowledge-based approach for interpreting genome-wide expression
636 profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

637 30. Degtyareva, A.O., Antontseva, E.V., and Merkulova, T.I. (2021). Regulatory SNPs:
638 Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases.
639 Int. J. Mol. Sci. *22*,.

640 31. Schmidt, E.M., Zhang, J., Zhou, W., Chen, J., Mohlke, K.L., Eugene Chen, Y., and Willer,
641 C.J. (2015). GREGOR: evaluating global enrichment of trait-associated variants in
642 epigenomic features using a systematic, data-driven approach. Bioinformatics *31*, 2601–
643 2606.

644 32. Tripathi, M., Yen, P.M., and Singh, B.K. (2020). Estrogen-Related Receptor Alpha: An
645 Under-Appreciated Potential Target for the Treatment of Metabolic Diseases. Int. J. Mol.
646 Sci. *21*,.

647 33. Saykally, J.N., Dogan, S., Cleary, M.P., and Sanders, M.M. (2009). The ZEB1
648 Transcription Factor Is a Novel Repressor of Adiposity in Female Mice. PLoS One *4*,
649 e8460.

650 34. Gubelmann, C., Schwalie, P.C., Raghav, S.K., Röder, E., Delessa, T., Kiehlmann, E.,
651 Waszak, S.M., Corsinotti, A., Udin, G., Holcombe, W., et al. (2014). Identification of the

652   transcription factor ZEB1 as a central component of the adipogenic gene regulatory
653   network. Elife *3*, e03346.

654   35. Neuschwander-Tetri, B.A. (2015). Retinoid X receptor: the forgotten partner in
655   regulating lipid metabolism? Am. J. Clin. Nutr. *102*, 5–6.

656   36. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S.,
657   Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of
658   loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.

659   37. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M.,
660   Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional
661   variation in personal genomes using RegulomeDB. Genome Research *22*, 1790–1797.

662   38. Anastasia, I., Ilacqua, N., Raimondi, A., Lemieux, P., Ghandehari-Alavijeh, R., Faure, G.,
663   Mekhedov, S.L., Williams, K.J., Caicci, F., Valle, G., et al. (2021). Mitochondria-rough-ER
664   contacts in the liver regulate systemic lipid homeostasis. Cell Rep. *34*, 108873.

665   39. Jurgens, S.J., Choi, S.H., Morrill, V.N., Chaffin, M., Pirruccello, J.P., Halford, J.L., Weng,
666   L.-C., Nauffal, V., Roselli, C., Hall, A.W., et al. (2020). Rare Genetic Variation Underlying
667   Human Diseases and Traits: Results from 200,000 Individuals in the UK Biobank.

668   40. Loos, R.J.F. (2016). CREBRF variant increases obesity risk and protects against
669   diabetes in Samoans. Nat. Genet. *48*, 976–978.

670   41. Minster, R.L., Hawley, N.L., Su, C.-T., Sun, G., Kershaw, E.E., Cheng, H., Buhule, O.D.,
671   Lin, J., Reupena, M.S., Viali, S. 'itea, et al. (2016). A thrifty variant in CREBRF strongly
672   influences body mass index in Samoans. Nat. Genet. *48*, 1049–1054.

673   42. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon,
674   D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among
675   ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–
676   1523.

677   43. Li, Z., Votava, J.A., Zajac, G.J.M., Nguyen, J.N., Leyva Jaimes, F.B., Ly, S.M., Brinkman,
678   J.A., De Giorgi, M., Kaul, S., Green, C.L., et al. (2020). Integrating Mouse and Human
679   Genetic Data to Move beyond GWAS and Identify Causal Genes in Cholesterol
680   Metabolism. Cell Metab. *31*, 741–754.e5.

681   44. Varshney, A., VanRenterghem, H., Orchard, P., Boyle, A.P., Stitzel, M.L., Ucar, D., and
682   Parker, S.C.J. (2019). Cell Specificity of Human Regulatory Annotations and Their Genetic
683   Effects on Gene Expression. Genetics *211*, 549–562.

684   45. van der Wijst, M.G.P., de Vries, D.H., Groot, H.E., Trynka, G., Hon, C.C., Bonder, M.J.,
685   Stegle, O., Nawijn, M.C., Idaghdour, Y., van der Harst, P., et al. (2020). Science Forum: The
686   single-cell eQTLGen consortium.

687   46. Arvanitis, M., Tayeb, K., Strober, B.J., and Battle, A. (2021). Redefining tissue
688   specificity of genetic regulation of gene expression in the presence of allelic

689   heterogeneity. medRxiv 2021.06.28.21259545.

690   47. Zhu, X., and Stephens, M. (2017). BAYESIAN LARGE-SCALE MULTIPLE REGRESSION
691   WITH SUMMARY STATISTICS FROM GENOME-WIDE ASSOCIATION STUDIES. Ann. Appl.
692   Stat. *11*, 1561–1592.

693   48. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland,
694   H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse
695   populations improves discovery for complex traits. Nature *570*, 514–518.

696   49. Mägi, R., Horikoshi, M., Sofer, T., Mahajan, A., Kitajima, H., Franceschini, N.,
697   McCarthy, M.I., COGENT-Kidney Consortium, T2D-GENES Consortium, and Morris, A.P.
698   (2017). Trans-ethnic meta-regression of genome-wide association studies accounting for
699   ancestry increases power for discovery and improves fine-mapping resolution. Hum.
700   Mol. Genet. *26*, 3639–3650.

701   50. Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and
702   associated tools. Brief. Bioinform. *14*, 144–161.

703   51. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C.,
704   and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic
705   association studies using summary statistics. PLoS Genet. *10*, e1004383.

706   52. Chesi, A., Wagley, Y., Johnson, M.E., Manduchi, E., Su, C., Lu, S., Leonard, M.E., Hodge,
707   K.M., Pippin, J.A., Hankenson, K.D., et al. (2019). Genome-scale Capture C promoter
708   interactions implicate effector genes at GWAS loci for bone mineral density. Nat.
709   Commun. *10*, 1260.

710   53. Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and
711   Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. *4*,
712   1310.

713   54. Cairns, J., Freire-Pritchett, P., Wingett, S.W., Várnai, C., Dimond, A., Plagnol, V.,
714   Zerbino, D., Schoenfelder, S., Javierre, B.-M., Osborne, C., et al. (2016). CHiCAGO: robust
715   detection of DNA looping interactions in Capture Hi-C data. Genome Biol. *17*, 127.

716   55. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F.,
717   Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference
718   human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

719   56. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan,
720   M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges.
721   PLoS Comput. Biol. *9*, e1003118.

722   57. MacParland, S.A., Liu, J.C., Ma, X.-Z., Innes, B.T., Bartczak, A.M., Gage, B.K., Manuel, J.,
723   Khuu, N., Echeverri, J., Linares, I., et al. (2018). Single cell RNA sequencing of human liver
724   reveals distinct intrahepatic macrophage populations. Nat. Commun. *9*, 4383.

725   58. Single Cell Portal.

726   59. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev,
727   A. Fast gene set enrichment analysis.

728   60. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for
729   Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative
730   Biology *16*, 284–287.

731   61. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for
732   the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat.
733   Protoc. *4*, 1184–1191.

734   62. Finucane, A., Carduff, E., Meade, R., Doyle, S., Fenning, S., Cumming, S., Hekerem, D.,
735   Rahman, F., Lugton, J., Johnston, B., et al. (2021). Palliative care research promotion in
736   policy and practice: a knowledge exchange process. BMJ Support. Palliat. Care.

737