Supplemental Methods

Identification of outlier ATAC-seq libraries using principal component analysis

We tested for outlying ATAC libraries using principal component analysis (PCA) of ATAC read counts within a previously defined set of liver tissue consensus ATAC peaks (Currin et al. 2021). We converted the coordinates of the previously published consensus peaks from GRCh37 to GRCh38 using liftOver from UCSCTools (Hinrichs et al. 2006) with the parameter -minMatch=0.75, counted the number of reads in each ATAC library overlapping these peaks using featureCounts (Liao et al. 2014), and adjusted peak counts for library size and performed variance-stabilization using DESeq2 (Love et al. 2014). We performed PCA using the prcomp function in R (R Core Team 2015). We identified two libraries from one individual that were PC1 outliers, which we removed from further analysis.

VerifyBamID additional details

We selected a set of autosomal genotypes with MAF > 5%, genotype missingness < 5%, and that overlapped previously defined liver tissue consensus ATAC peaks (Currin et al. 2021) using BEDTools (Quinlan 2014) and PLINK (Purcell et al. 2007). We harmonized genotypes to the 1000 Genomes Phase 3 reference panel (The 1000 Genomes Project Consortium 2015) using the HRC-1000G-check-bim-NoReadKey.pl script from the McCarthy Group Tools (https://www.well.ox.ac.uk/~wrayner/tools/) with parameters -g -p ALL and added allele frequency information to the resulting VCF file using the BCFtools fill-tags extension (Danecek et al. 2021). We converted genotype positions from GRCh37 to GRCh38 coordinates using Picard LiftoverVcf v2.17.8 (http://broadinstitute.github.io/picard), resulting in 35,495 genotypes to use with verifyBamID. We ran verifyBamID separately on each high-quality ATAC library with parameters –best –precise –maxDepth 1000 and considered the best matching ATAC library to a set of genotypes to be a confident match if chipmix < 0.02. We corrected sample swaps and removed ATAC libraries from further analysis if they did not confidently match to genotypes from any of the 224 individuals.

Heterozygosity and comparing reported sex to sex predicted from genotypes

We calculated autosomal heterozygosity with the PLINK –het command and compared reported sex to inferred sex based on X Chromosome genotypes using the PLINK –check-sex command. For both the heterozygosity and sex check analyses, we excluded genotypes within regions of unusually long-range LD (https://genome.sph.umich.edu/wiki/Regions_of_high_linkage_disequilibrium_(LD)) and selected a set of variants in approximate linkage equilibrium using PLINK with the parameters --indep-pairwise 50 5 0.2.

Comparison of consensus ATAC peaks to previously identified regulatory elements

We compared ATAC peaks from the current study to ATAC peaks from our pilot liver caQTLs study (Currin et al. 2021) and to liver tissue ATAC peaks from ENCODE(The ENCODE Project Consortium et al. 2020). For the pilot study, we used liftOver (Hinrichs et al. 2006) with -minMatch=0.75 to convert peak coordinates to GRCh38. For ENCODE, we downloaded pseudoreplicated peak files generated from ATAC-seq on eight human liver donors and generated consensus peaks by first merging peaks across all donors using BEDTools (Quinlan 2014) and then selecting merged peaks that overlapped a peak in at least three donors. We considered a peak from the current study to replicate a peak from one of the other two studies if either 50% of a peak from the other study was covered by a peak in the current study or if 50% of a peak from the current study was covered by a peak from the other study using BEDTools (Quinlan 2014) intersect with parameters -u -f 0.5 -F 0.5 -e. We also determined the number of

peaks in the current study that did not share a single base with a peak from the other two studies using BEDTools intersect with the -v parameter.

Genotype imputation additional details
Prior to imputation, we removed variants that were palindromic, not on autosomes, or that had genotype missingness > 5% across the 157 individuals using PLINK (Purcell et al. 2007). Using the TOPMed Imputation Server (Taliun et al. 2021; Das et al. 2016) quality check tool, we identified 278,648 variants on the opposite strand of the TOPMed reference panel. Because the TOPMed quality check tool returns GRCh38 variant coordinates and the pre-imputation variants were in GRCh37 coordinates, we made a key linking the two coordinate systems using the liftOver tool from UCSCTools (Hinrichs et al. 2006) with the option -bedPlus=3; four strand-flipped variants were excluded because they failed to lift over. We used PLINK (Purcell et al. 2007) to correct strand flips. We then performed imputation on 556,767 variants that passed pre-imputation quality control using the TOPMed Imputation Server (Taliun et al. 2021; Das et al. 2016), selecting Eagle v2.4 (Loh et al. 2016) for phasing and Minimac4 v1.0.2 (Das et al. 2016) for imputation. We subsequently filtered the imputation results to contain only the 138 individuals that also passed ATAC quality control and 10,974,995 single nucleotide polymorphisms (SNPs) or indels with imputation r^2 > 0.3 and minor allele frequency (MAF) > 0.01 (calculated in the 138 individuals) using BCFtools v1.13 (Danecek et al. 2021).

Preparation of VCF files for ancestry classifications and genotype principal components
For the 138 individuals that passed genotype and ATAC quality control, we used PLINK (Purcell et al. 2007) and BEDTools (Quinlan 2014) to select autosomal variants with call rate >= 5% and MAF >= 0.05, exclude genotypes within regions of unusually long-range LD (https://genome.sph.umich.edu/wiki/Regions_of_high_linkage_disequilibrium_(LD)), and perform LD pruning (PLINK --indep-pairwise 50 5 0.2). We harmonized genotypes to the 1000 Genomes Phase 3 reference panel (The 1000 Genomes Project Consortium 2015) using the HRC-1000G-check-bim-NoReadKey.pl script from the McCarthy Group Tools (https://www.well.ox.ac.uk/~wrayner/tools/) with parameters -g -p ALL –noexclude. We used a combination of PLINK (Purcell et al. 2007), BEDTools (Quinlan 2014), and tabix (Danecek et al. 2021) to make a VCF file containing the 138 liver donors and the 1000G individuals, which contained 99,206 variants shared between the two studies.

Calculation of ATAC principal components for caQTL covariates
To prepare ATAC peak counts for caQTL mapping, we removed ATAC mapping bias using the WASP mapping pipeline (Geijn et al. 2015) and made a sample-by-consensus ATAC peak count matrix using featureCounts (Liao et al. 2014). We calculated the GC content of consensus peaks using BEDTools nuc (Quinlan 2014), calculated peak count offsets to adjust for the effects of GC bias using full quantile normalization implemented in EDASeq (Risso et al. 2011), integrated EDASeq GC bias offsets and DESeq2 (Love et al. 2014) size factors to adjust for GC bias and library size, and performed variance stabilization using DESeq2 (Love et al. 2014). To generate ATAC PCs to control for unknown sources of variation, we used the prcomp function in R (R Core Team 2015) to perform PCA on variance-stabilized peak counts adjusted for sex and the first two genotype PCs using limma (Ritchie et al. 2015).

Replication of liver caQTL
To determine how many of the 3,123 caQTLs from our previous 20-donor study (Currin et al. 2021) were replicated, we used liftOver (Hinrichs et al. 2006) with -minMatch=0.75 to convert the previous caPeak coordinates to GRCh38 and considered a caPeak replicated if a caPeak from the current study overlapped by at least 50% of bases (BEDTools (Quinlan 2014) intersect

parameters -u -f 0.5 -F 0.5 -e) and if the lead variants from each study were in high LD (r^2>=0.8, TOPMed Europeans (Taliun et al. 2021; Huang et al. 2022)

Incorporating GENCODE v41 gene models into plotgardener

We used the following steps to incorporate GENCODE v41 (Frankish et al. 2019) protein coding and lncRNA genes into Plotgardener. First, we imported the GENCODE GTF file into R as a Granges object using rtracklayer (Lawrence et al. 2009) and made a TxDb object using the GenomicFeatures (Lawrence et al. 2013) makeTxDbFromGRanges function. Second, we made an OrgDb object from GENCODE genes using the AnnotationForge (https://bioconductor.org/packages/AnnotationForge) makeOrgPackage function. Finally, we generated a Plotgardener assembly object using these customized TxDb and OrgDb objects along with the BSgenome.Hsapiens.NCBI.GRCh38 genome object (https://bioconductor.org/packages/BSgenome).

Defining GWAS loci using distance-based clumping

For the two studies that did not report conditionally distinct signals (Graham et al. 2021; Said et al. 2022), we defined loci from the GWAS marginal summary statistics using distance-based clumping by selecting the most significant variant with $p<5\times10^{-8}$ in a region and removed all other variants within 500 kb. For the CRP data (Said et al. 2022), we performed distance-based clumping with Swiss v1.1.1. For the lipids datasets (Graham et al. 2021), the p-values for many variants were smaller than the numerical precision that Swiss could recognize, so we performed clumping using the following approach. We converted variant p-values to negative $\log_{10}$ scale using the formula -log(mantissa)/log(10) - exponent and then performed distance-based clumping in R (R Core Team 2015). Following clumping, we merged regions if lead variants were within 1 Mb of each other into a single locus.

References

Currin KW, Erdos MR, Narisu N, Rai V, Vadlamudi S, Perrin HJ, Idol JR, Yan T, Albanus RD, Broadaway KA, et al. 2021. Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am J Hum Genet* **108**: 1169–1189.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: giab008.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. 2016. Next-generation genotype imputation service and methods. *Nat Genet* **48**: 1284–1287.

Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766–D773.

Geijn B van de, McVicker G, Gilad Y, Pritchard JK. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**: 1061–3.

Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, et al. 2021. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**: 675–679.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, et al. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**: D590-598.

Huang L, Rosen JD, Sun Q, Chen J, Wheeler MM, Zhou Y, Min Y-I, Kooperberg C, Conomos MP, Stilp AM, et al. 2022. TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *Am J Hum Genet* **109**: 1175–1181.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**: 1841–1842.
Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**: e1003118.
Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.
Loh P-R, Palamara PF, Price AL. 2016. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet* **48**: 811–816.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575.
Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**: 11.12.1-34.
R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/.
Risso D, Schwartz K, Sherlock G, Dudoit S. 2011. GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics* **12**: 480.
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47.
Said S, Pazoki R, Karhunen V, Võsa U, Ligthart S, Bodinier B, Koskeridis F, Welsh P, Alizadeh BZ, Chasman DI, et al. 2022. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nat Commun* **13**: 2198.
Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, et al. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**: 290–299.
The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74.
The ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, et al. 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**: 699–710.