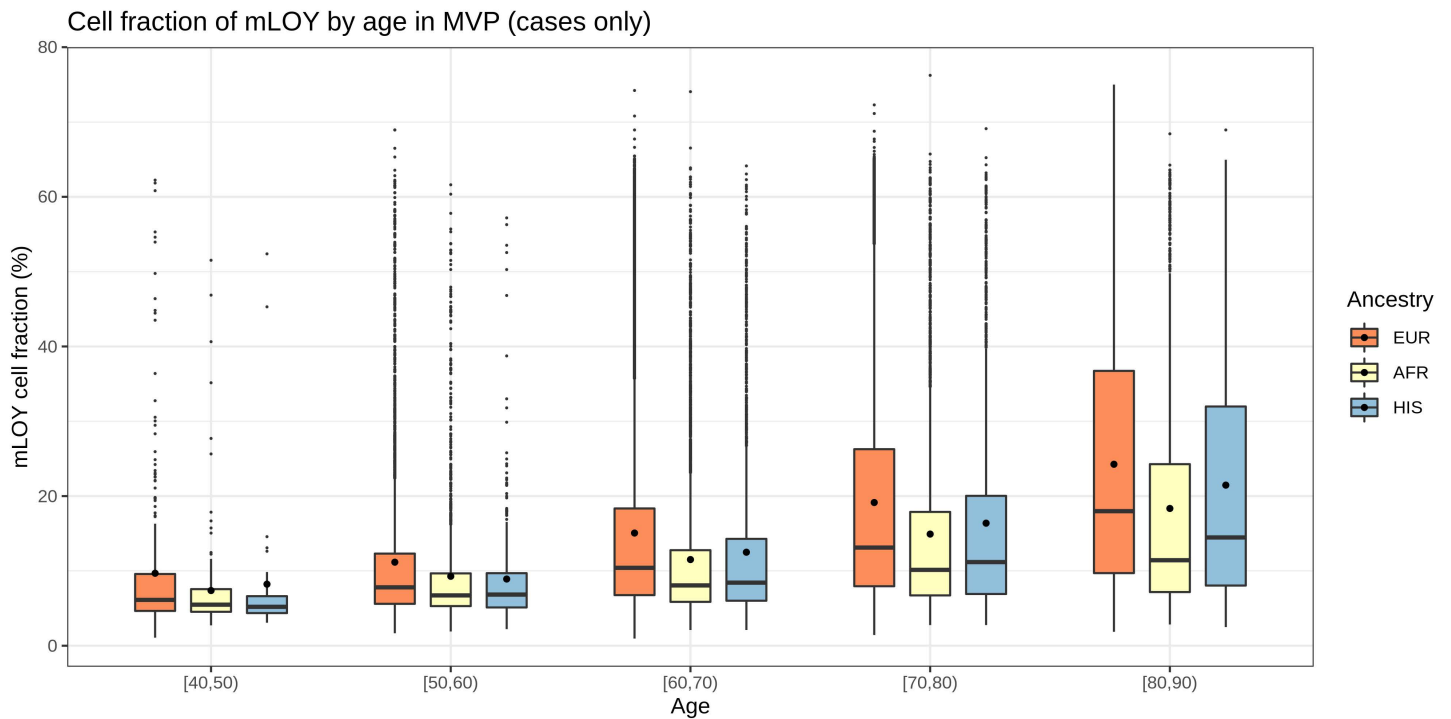
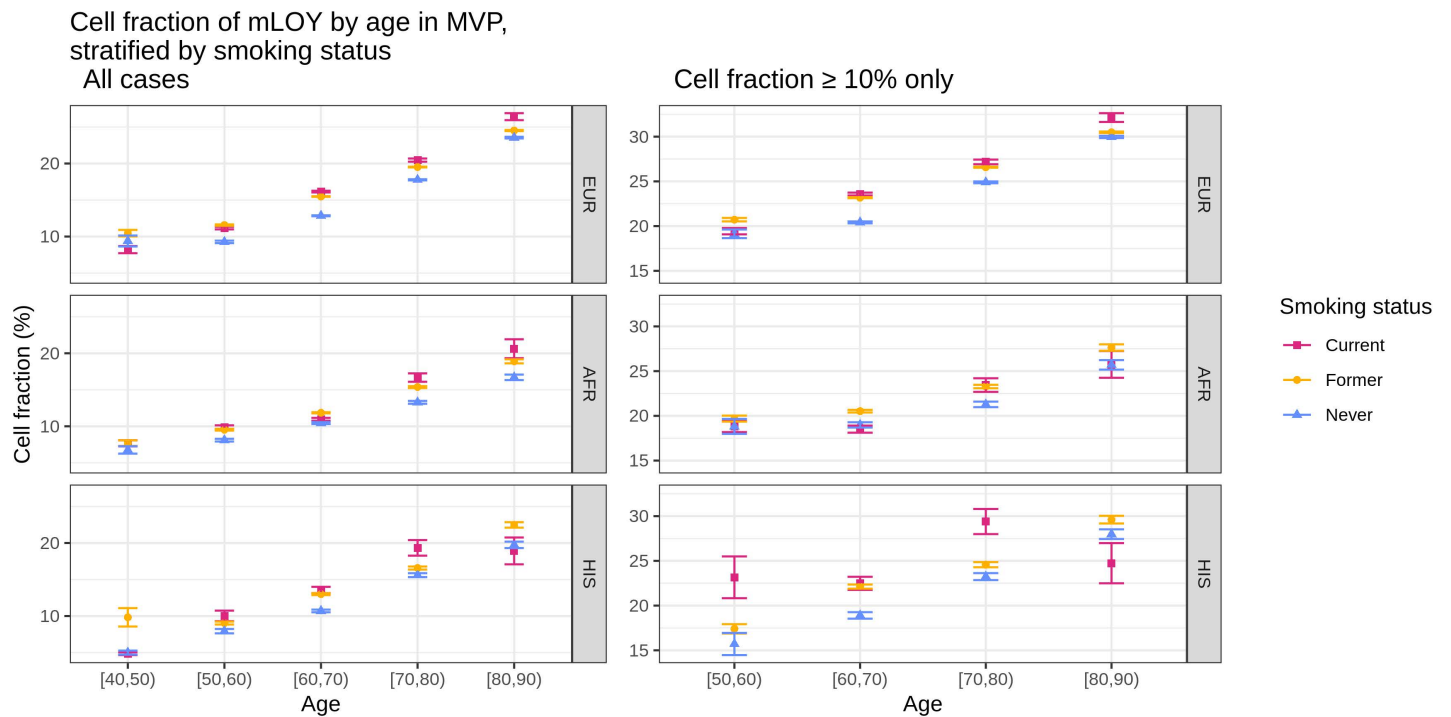
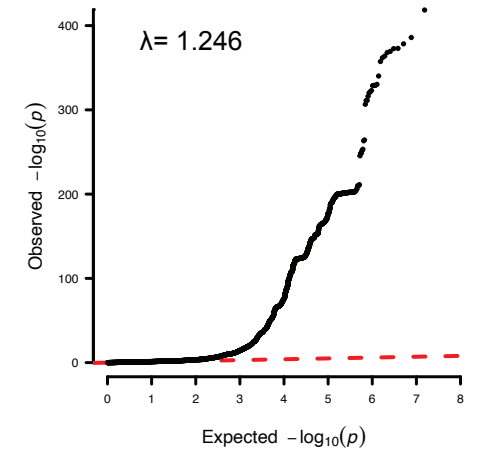
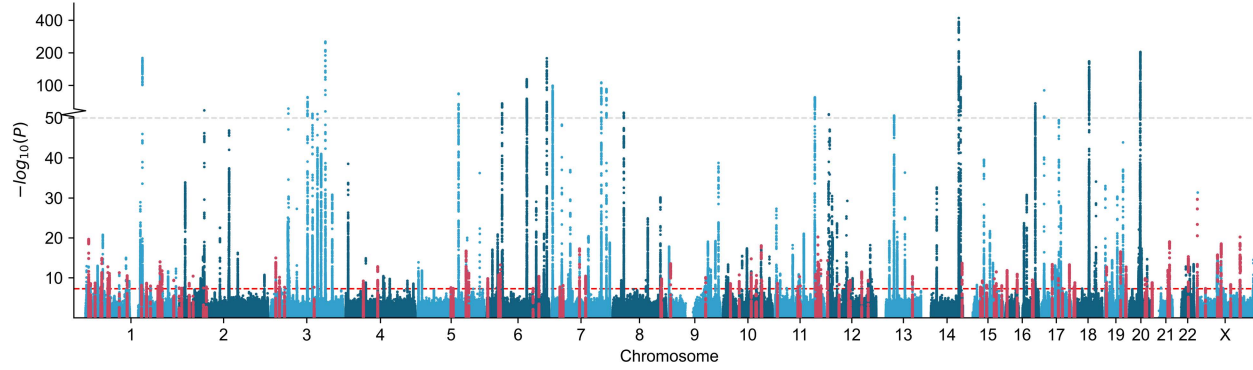


**Supplementary Fig. 1. Overview of genome-wide association study for mosaic loss of Y chromosome (mLOY) in Million Veteran Program (MVP) participants.** Three MVP individual ancestries were meta-analyzed, then replicated in two external cohorts. BMI, body mass index; mtCOJO, multi-trait-based conditional & joint association analysis; PGS, polygenic score; eQTL, expression quantitative trait loci; GSEA, gene set enrichment analysis; mLRR-Y, mean logarithm of R ratio in chromosome Y; MVMR multivariable Mendelian randomization.

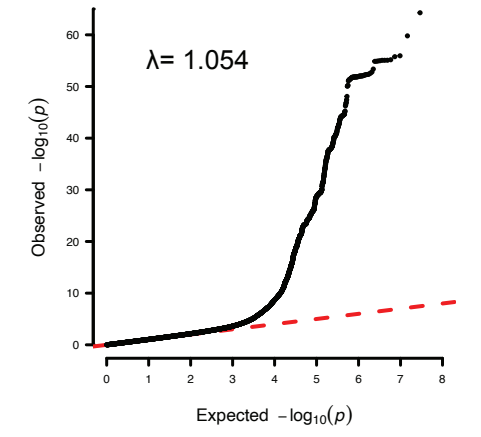
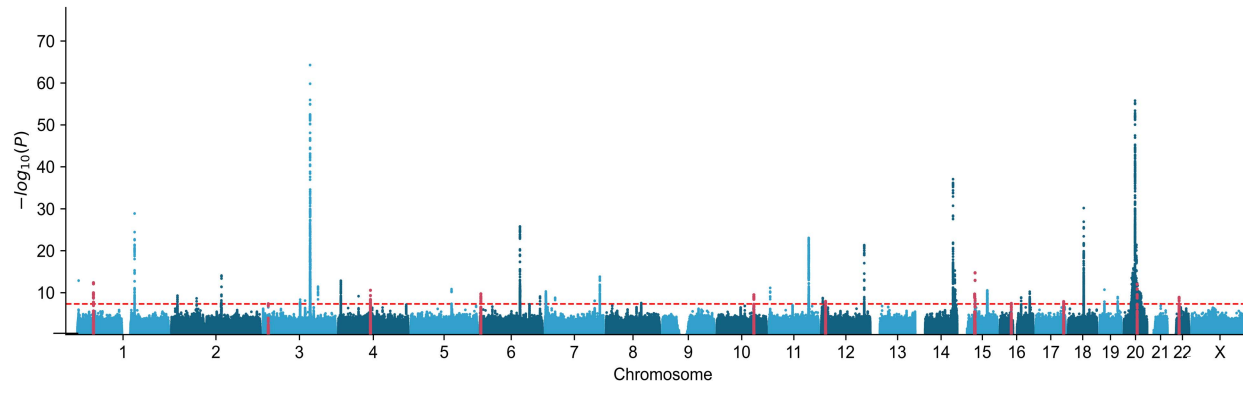
**a****b**

**Supplementary Fig. 2. Cell fraction of mLOY across groups. (a)** Cell fraction of mLOY in Million Veteran Program participants by ten-year age bins, stratified by ancestry. Boxplots show first decile, first quartile, median, third quartile, and last decile; dots indicate mean. **(b)** Cell fraction of mLOY, stratified ancestry and by current, former, and never smoking status. Error bars indicate 95% confidence intervals.

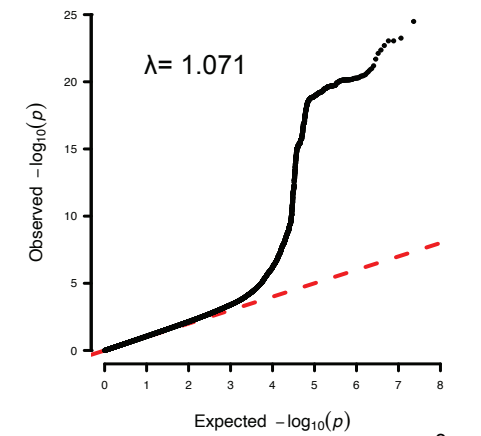
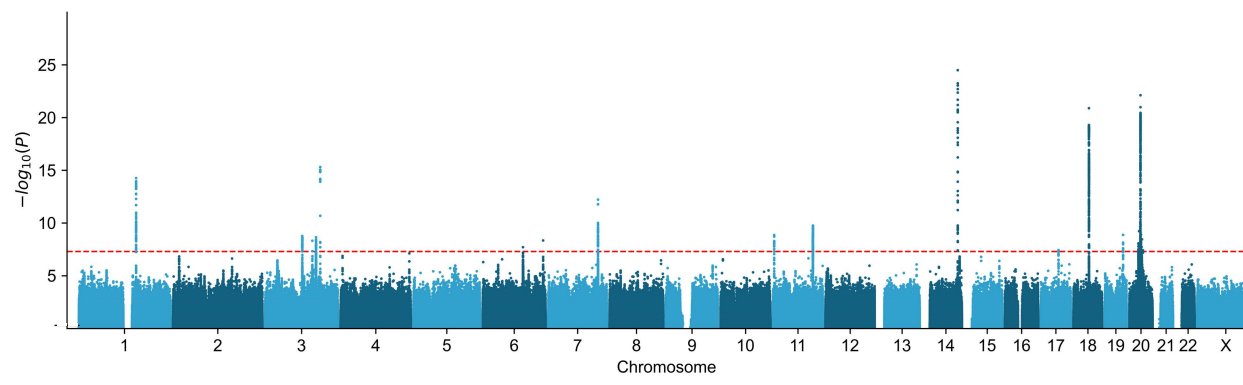
### S3a European ancestry



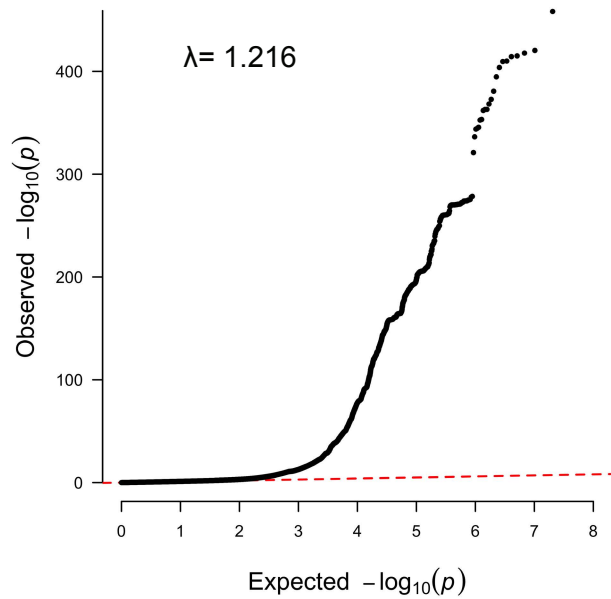
### b African ancestry



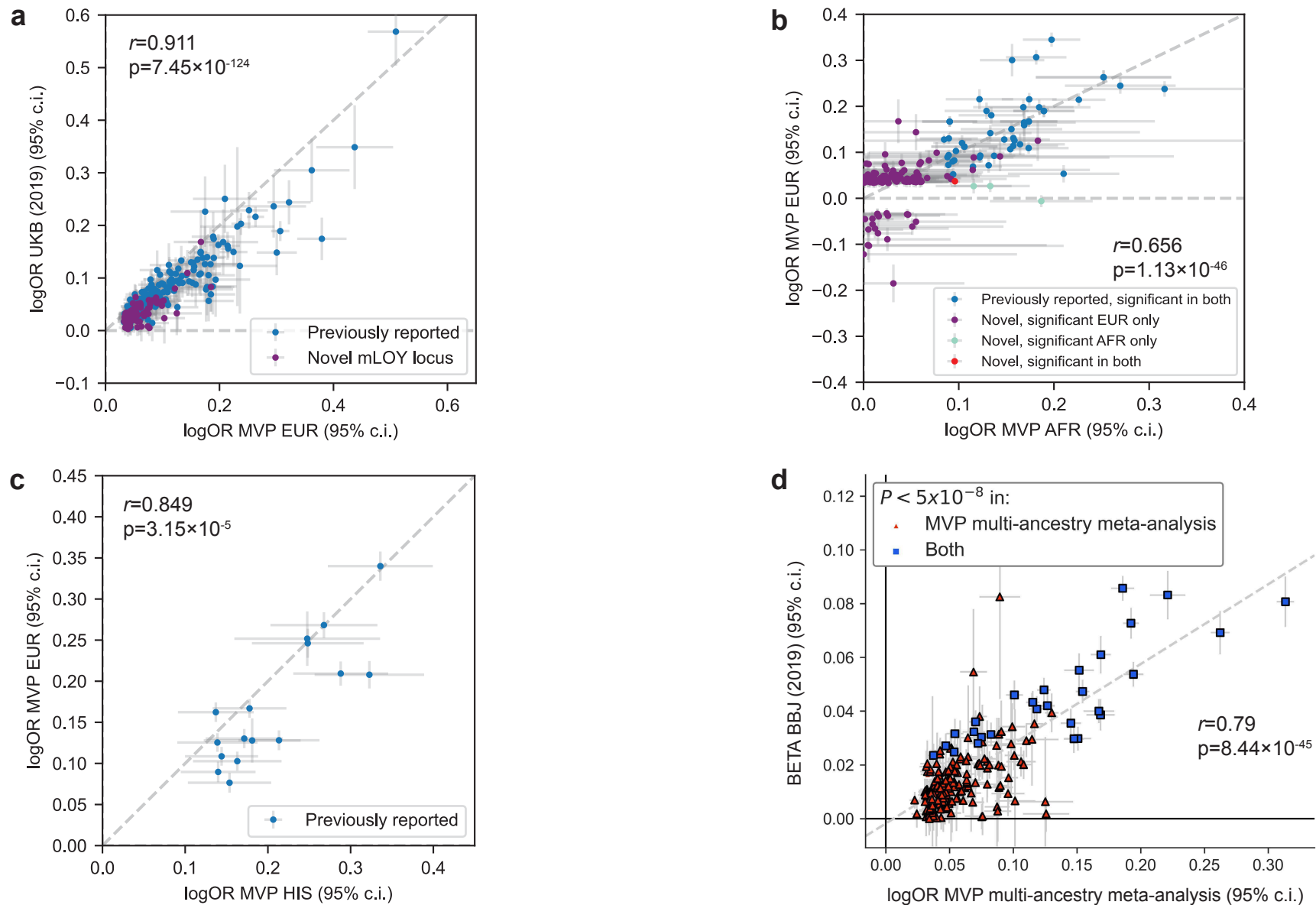
### c Hispanic ancestry



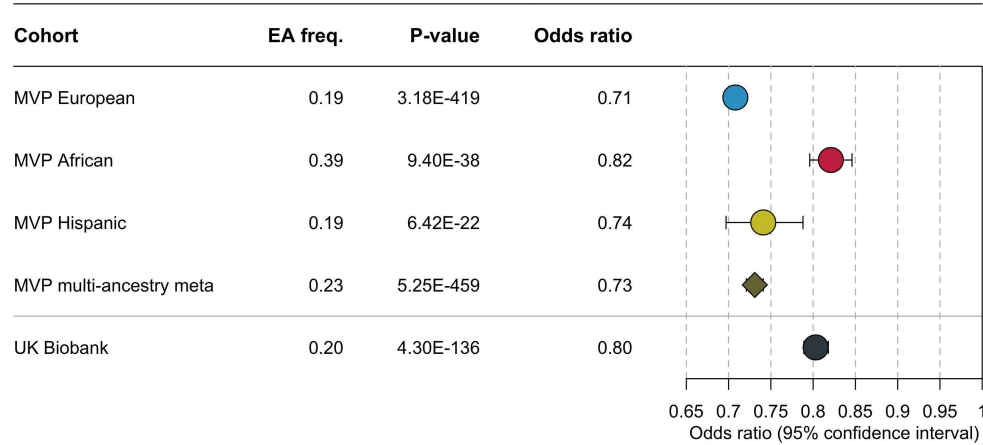
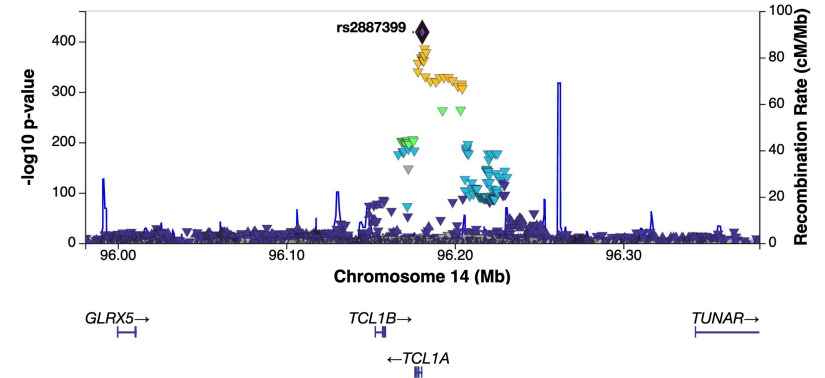
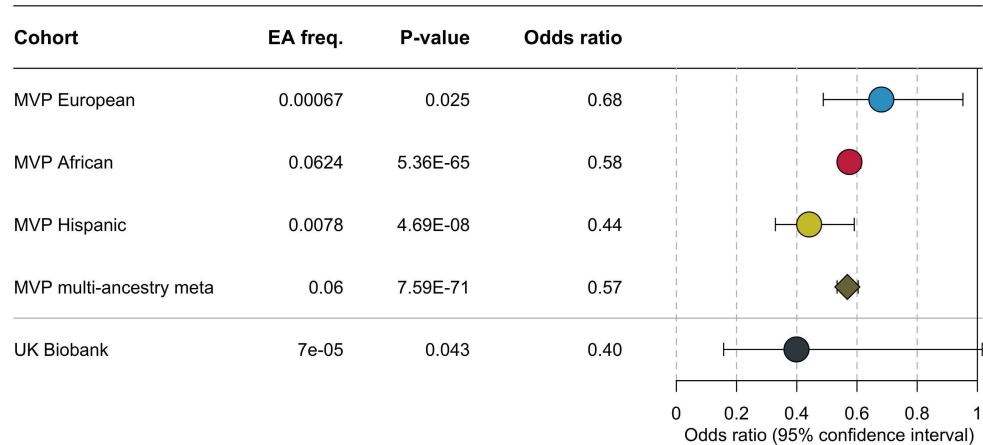
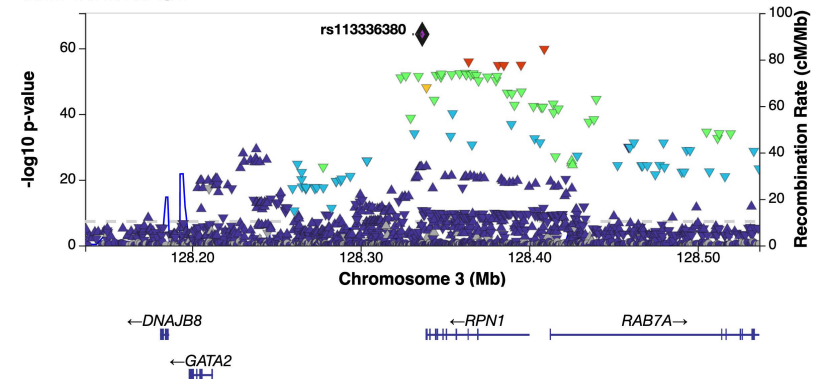
S3d



**Supplementary Fig. 3. Manhattan and quantile-quantile plots of Million Veteran Program participants by ancestry.** Manhattan plots (left) show the  $-\log_{10}(P)$  for associations of genetic variants with mLOY across 22 autosomal chromosomes plus chromosome X for **(a)** European ancestry, **(b)** African ancestry, and **(c)** Hispanic ancestry. Novel mLOY index variants and variants within  $\pm 50$  Kb are highlighted in red. The red line indicates the genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ). Grey dotted lines represent a transition from linear to log-scale on the y-axis. Quantile-quantile (QQ) plots (right) show observed versus expected distributions of  $P$ -values. **(d)** QQ plot of the multi-ancestry meta-analysis (Manhattan plot in Fig. 1d).

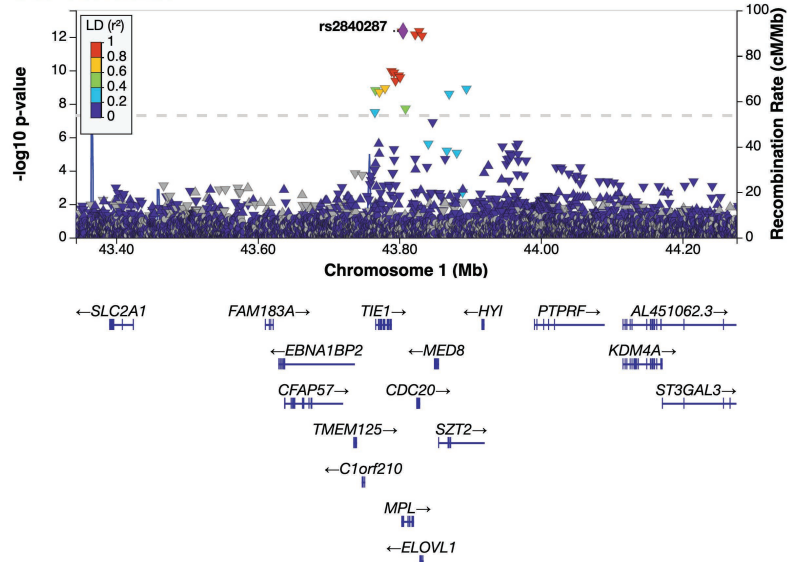


**Supplementary Fig. 4. Effect comparisons between cohorts.** Scatter plots comparing effect sizes (logarithm of odds ratios and 95% confidence interval) of genome-wide significant conditionally independent hits. **(a)** UK Biobank (Thompson et al. 2019) versus Million Veteran Program (MVP) European (EUR), indicating previously reported loci (blue) and novel loci (pink). **(b)** MVP-EUR versus MVP African ancestry (AFR), indicating previously reported loci, novel loci, and novel loci exclusive to EUR or AFR. **(c)** MVP-EUR versus MVP Hispanic ancestry (HIS), where all HIS signals were also found in EUR. **(d)** MVP multi-ancestry meta-analysis versus Biobank Japan (Terao et al. 2019), indicating variants that were genome-wide significant in both studies or only in MVP.

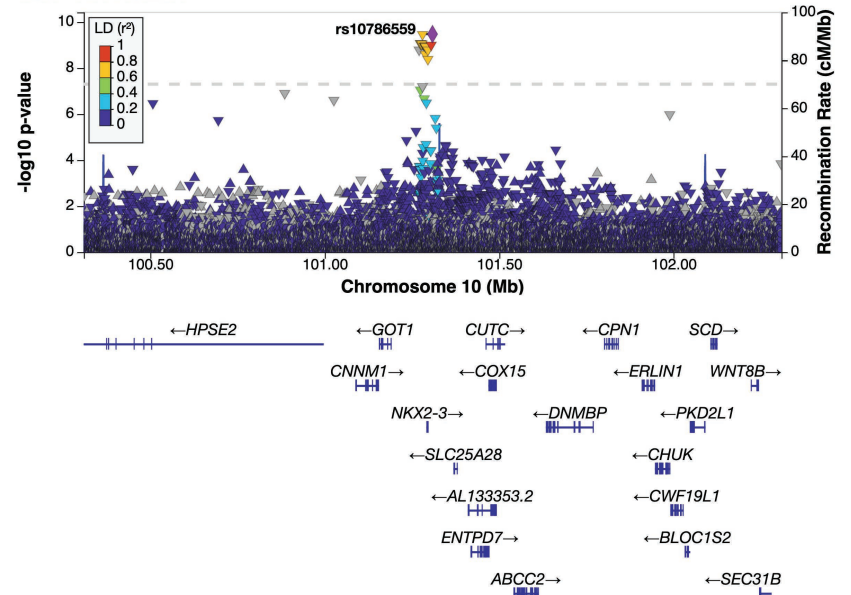
**a****rs2887399 (TCL1A)****MVP EUR mLOY****b****rs113336380 (RPN1)****MVP AFR mLOY**

**Supplementary Figure 5. Forest plots and regional Manhattan plots for most significant mosaic loss of Y associations. a,** rs2887399, the top SNP at *TCL1A*; **b,** rs113336380, the top SNP at *RPN1*. Left: forest plots showing frequency of effect allele, *P*-value, and Odds ratio for each Million Veteran Program (MVP) cohort and multi-ancestry meta-analysis, plus replication cohort UK Biobank. Right: Regional Manhattan plots for the MVP ancestry in which the SNP was most significant.

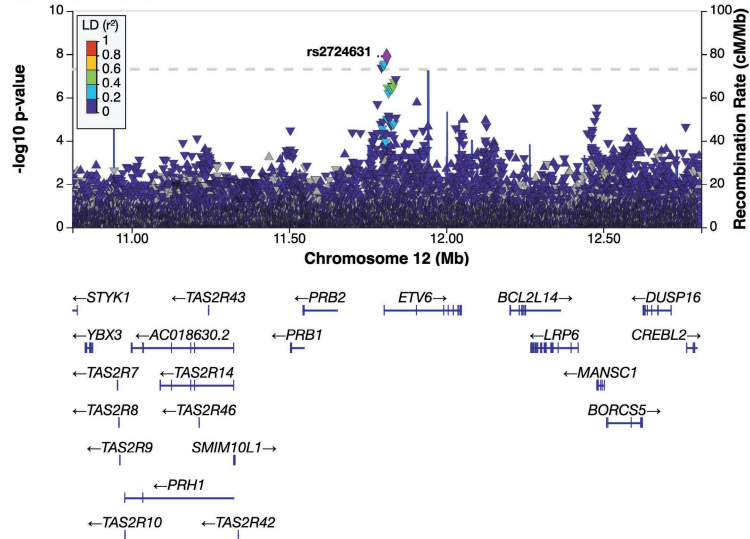
**S6a MVP AFR mLOY**



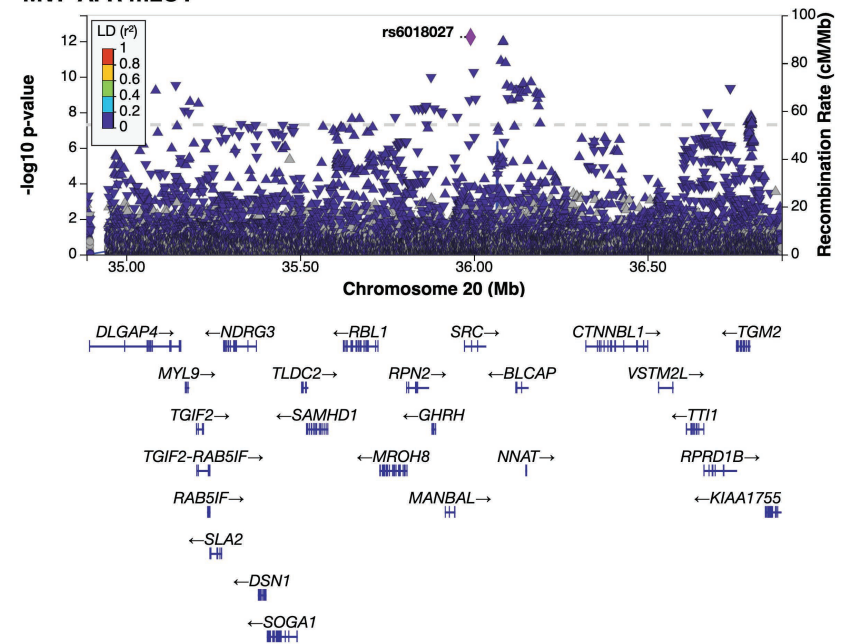
**b MVP AFR mLOY**



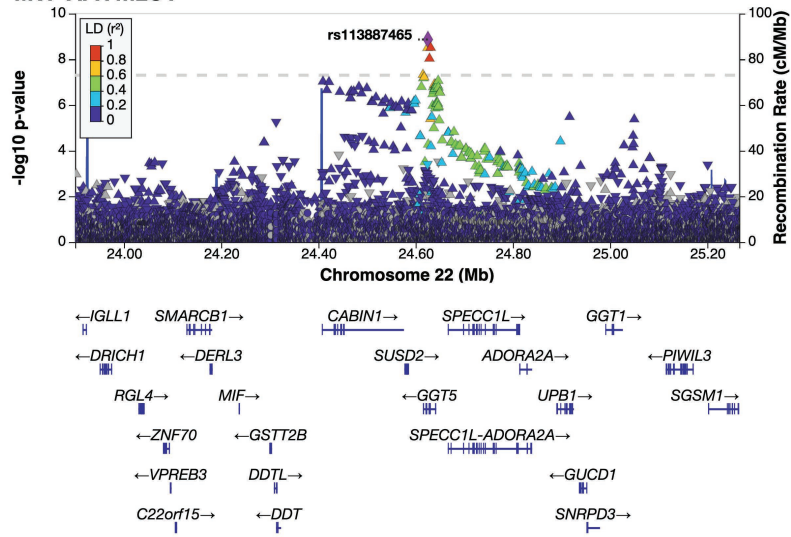
**c MVP AFR mLOY**



**d MVP AFR mLOY**



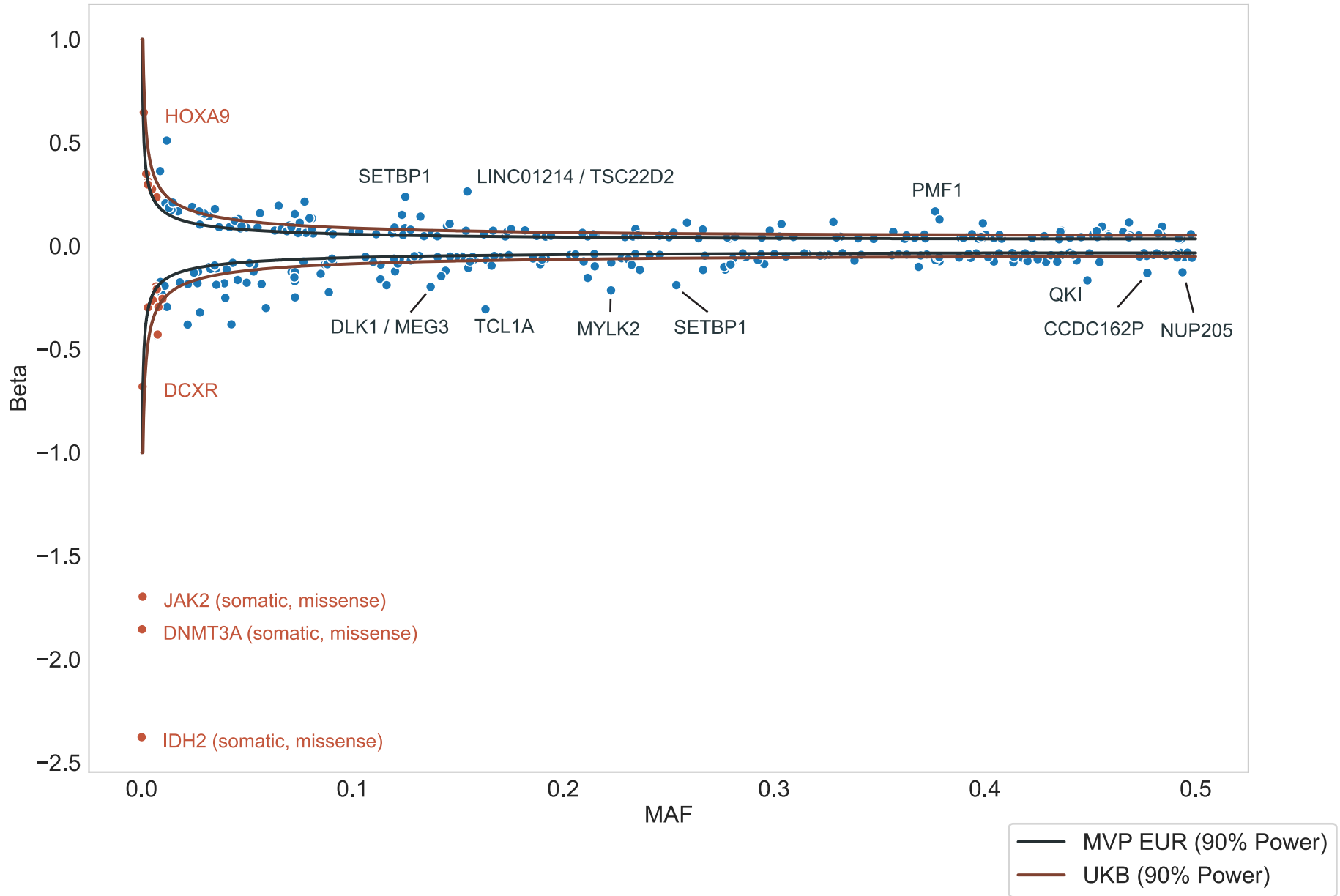
**S6e MVP AFR mLOY**



**Supplementary Fig. 6. AFR-specific novel loci.** Regional Manhattan plots showing variant-level associations with mLOY, and local linkage disequilibrium structure, of five genome-wide significant loci that appeared in the Million Veteran Program African cohort and not in European: **(a)** *MPL*, **(b)** *NKX2-3*, **(c)** *ETV6*, **(d)** *BLCAP*, and **(e)** *GGT5/CABIN1*.

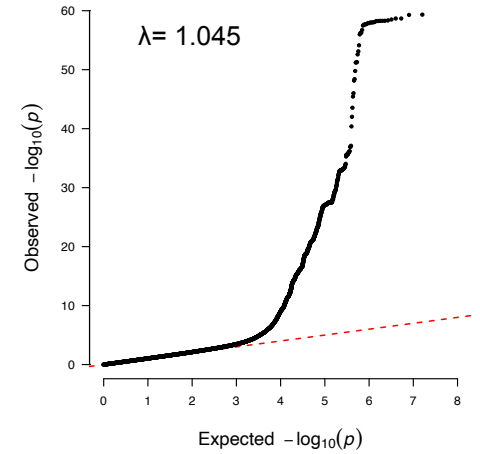
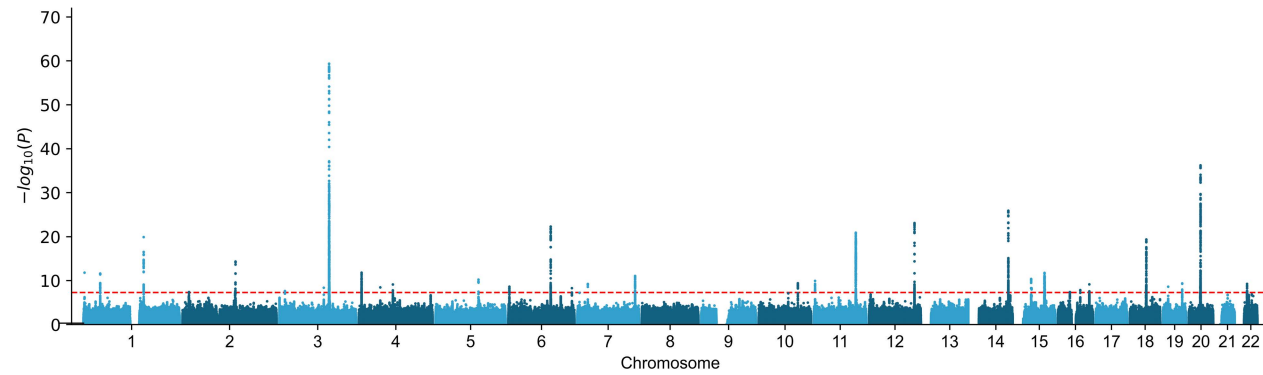


MVP EUR Cohort: Index and Rare SNPs

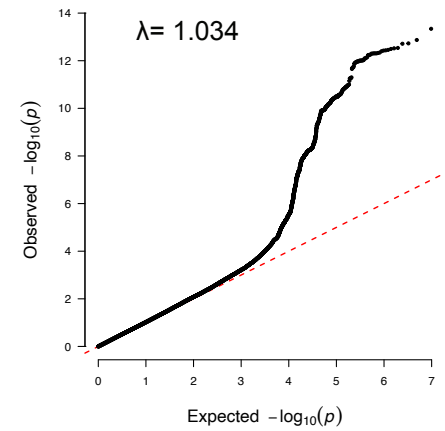
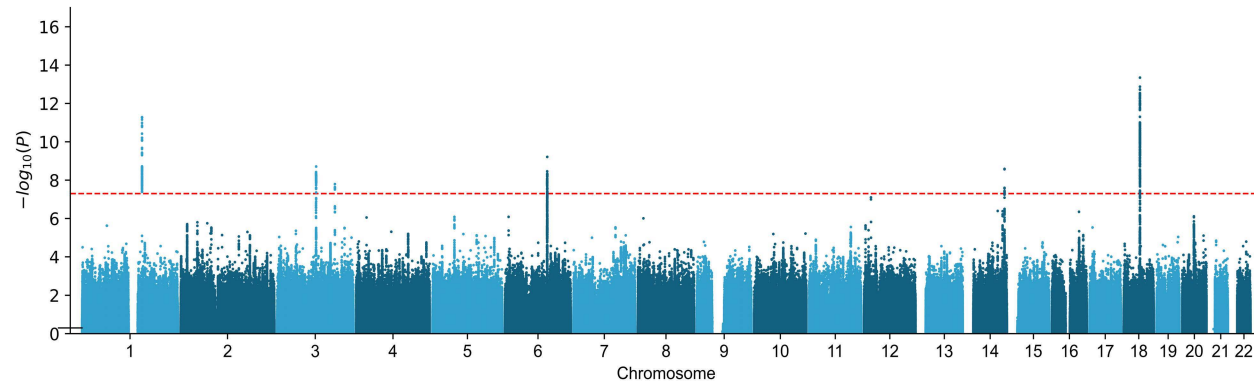


**Supplementary Fig. 7. Relationship between minor allele frequency and effect size.** For 327 genome-wide significant conditionally independent SNPs, plotting effect size Beta (in log odds) versus minor allele frequency (MAF). Simulated power curves are shown for MVP and UKB MAFs representing the expected magnitude of Beta detectable at  $P < 5 \times 10^{-8}$ . Significant rare variants (MAF < 0.001 in MVP) are shown in orange. Gene names provided for SNPs with  $P < 1 \times 10^{-100}$ .

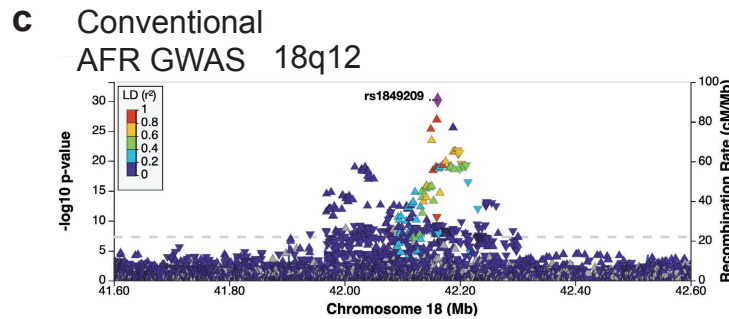
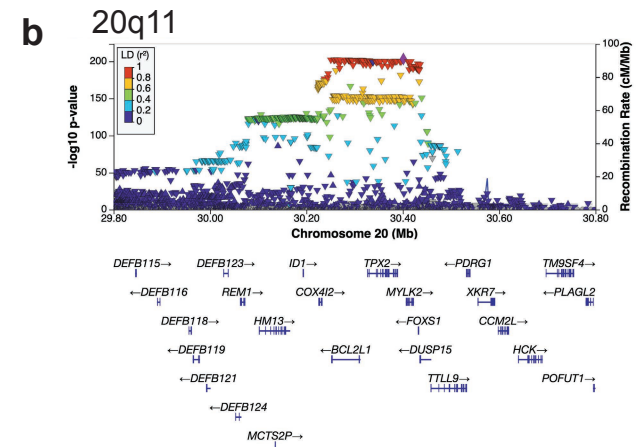
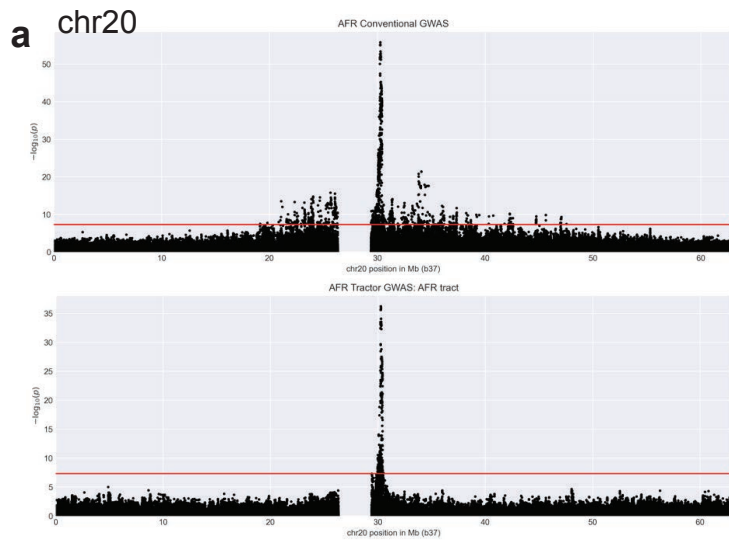
**a** Admixed African ancestry: African tract



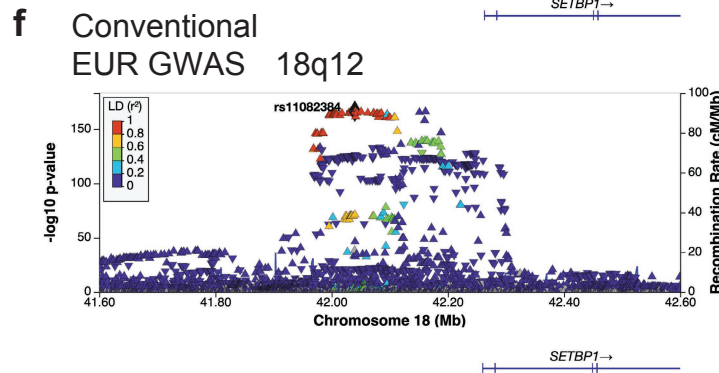
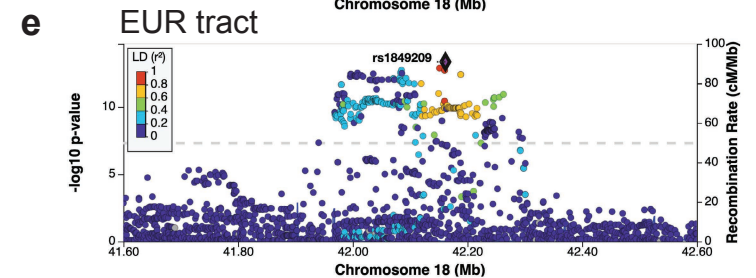
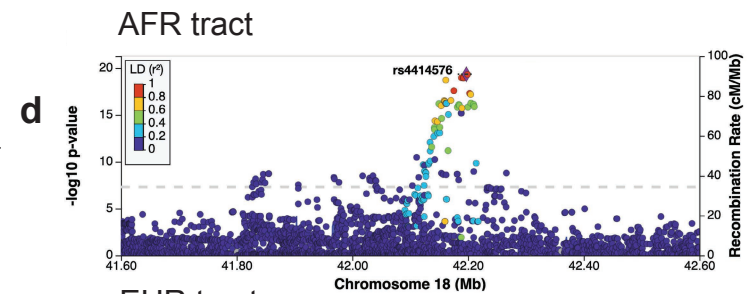
**b** Admixed African ancestry: European tract



**Supplementary Fig. 8. Admixed African individual ancestral tract Manhattan and quantile-quantile (QQ) plots.** Manhattan plots (left) show the  $-\log_{10}(P)$  for associations of genetic variants for **a**, African tract and **b**, European tract. QQ plots (right) show observed versus expected distributions of  $P$ -values.

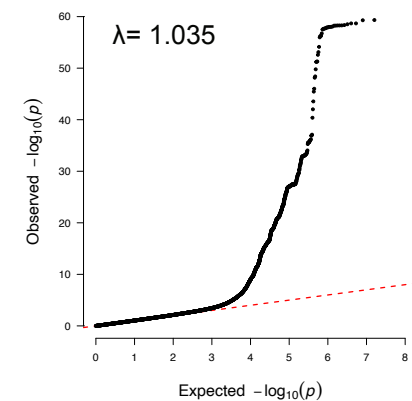
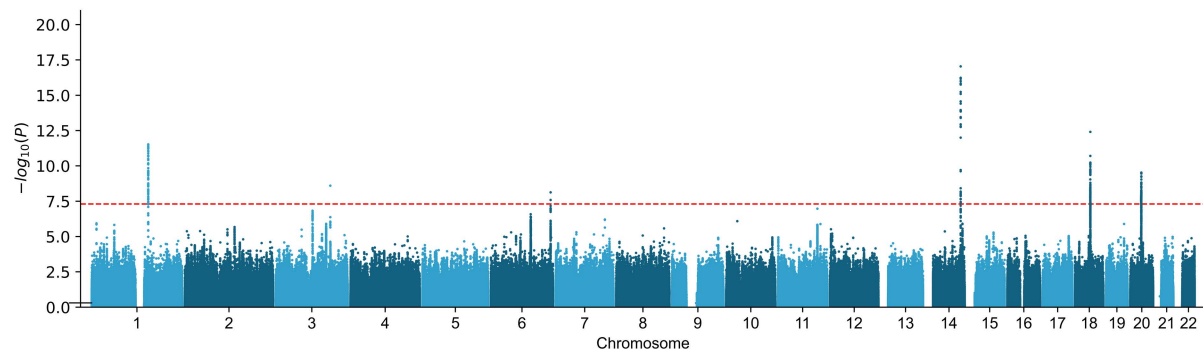


Local ancestry inference + tract-based GWAS

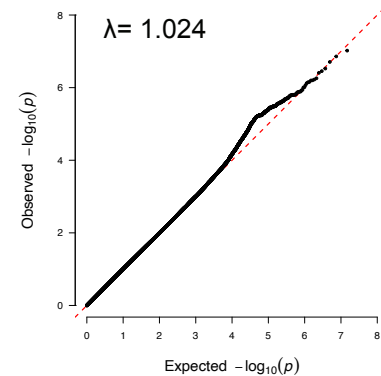
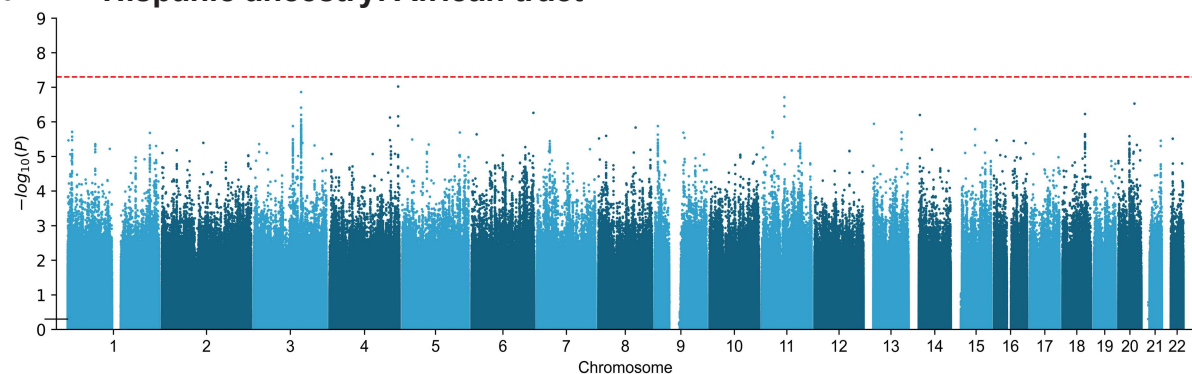


**Supplementary Fig. 9. Supporting evidence for local ancestry association analysis.** **a**, Admixed African (AFR) chromosome 20 Manhattan plot for conventional GWAS (top) and AFR tract (bottom). **b**, EUR regional Manhattan plot for *BCL2L1* in conventional GWAS. **c**, At 18q12 (*SETBP1*), we observed complex LD structure. This complex structure was resolved by separating the **d**, AFR and **e**, EUR tracts. **f**, Regional Manhattan plot for *SETBP1* in conventional EUR genome-wide association study (GWAS).

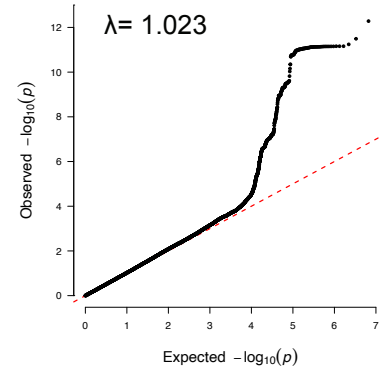
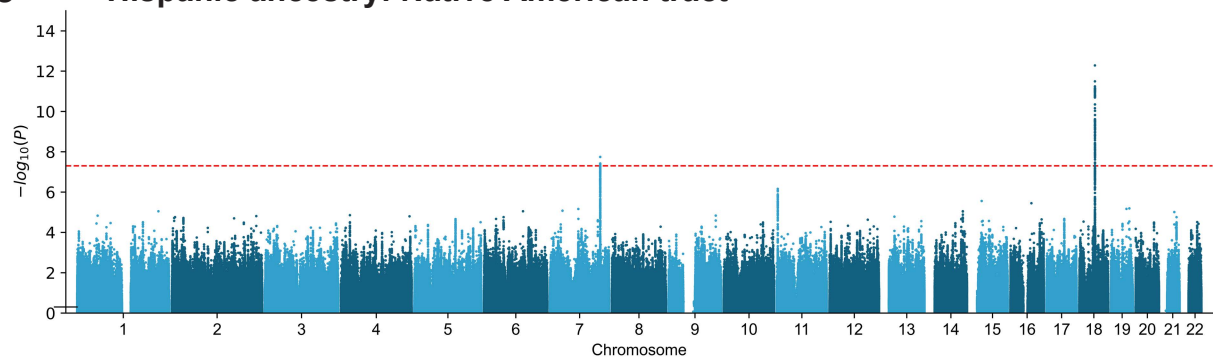
**a** Hispanic ancestry: European tract



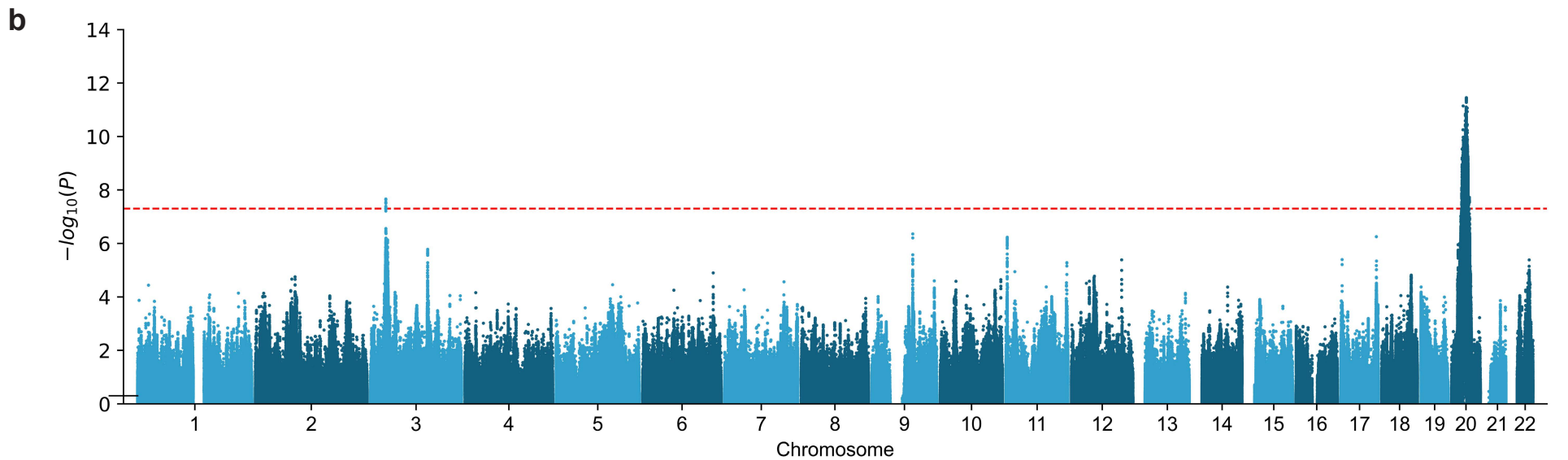
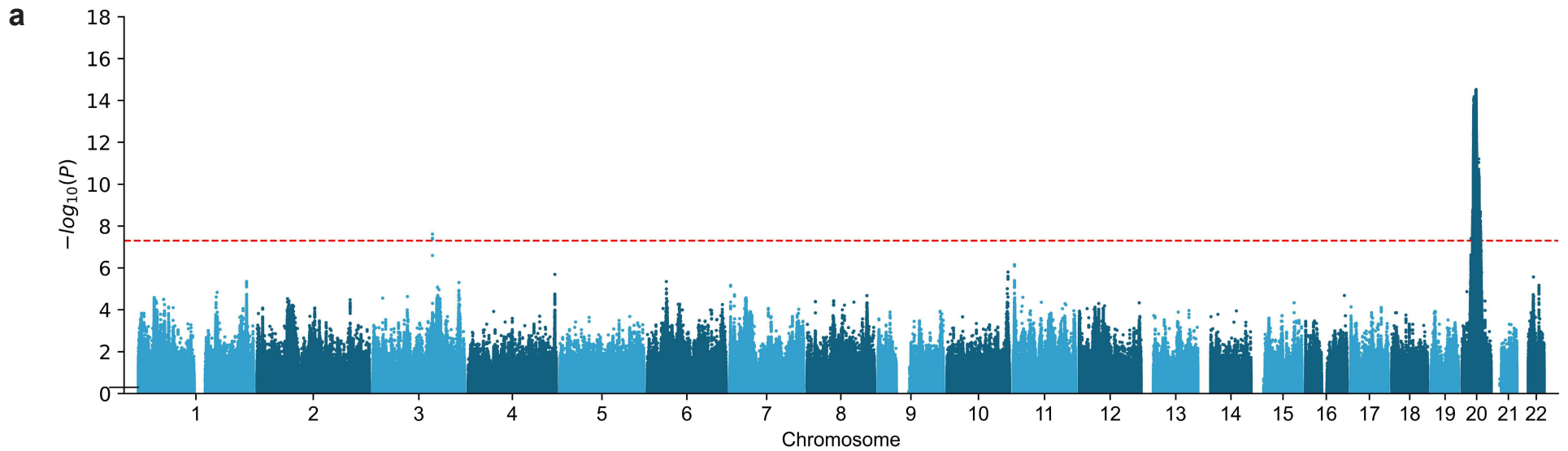
**b** Hispanic ancestry: African tract



**c** Hispanic ancestry: Native American tract

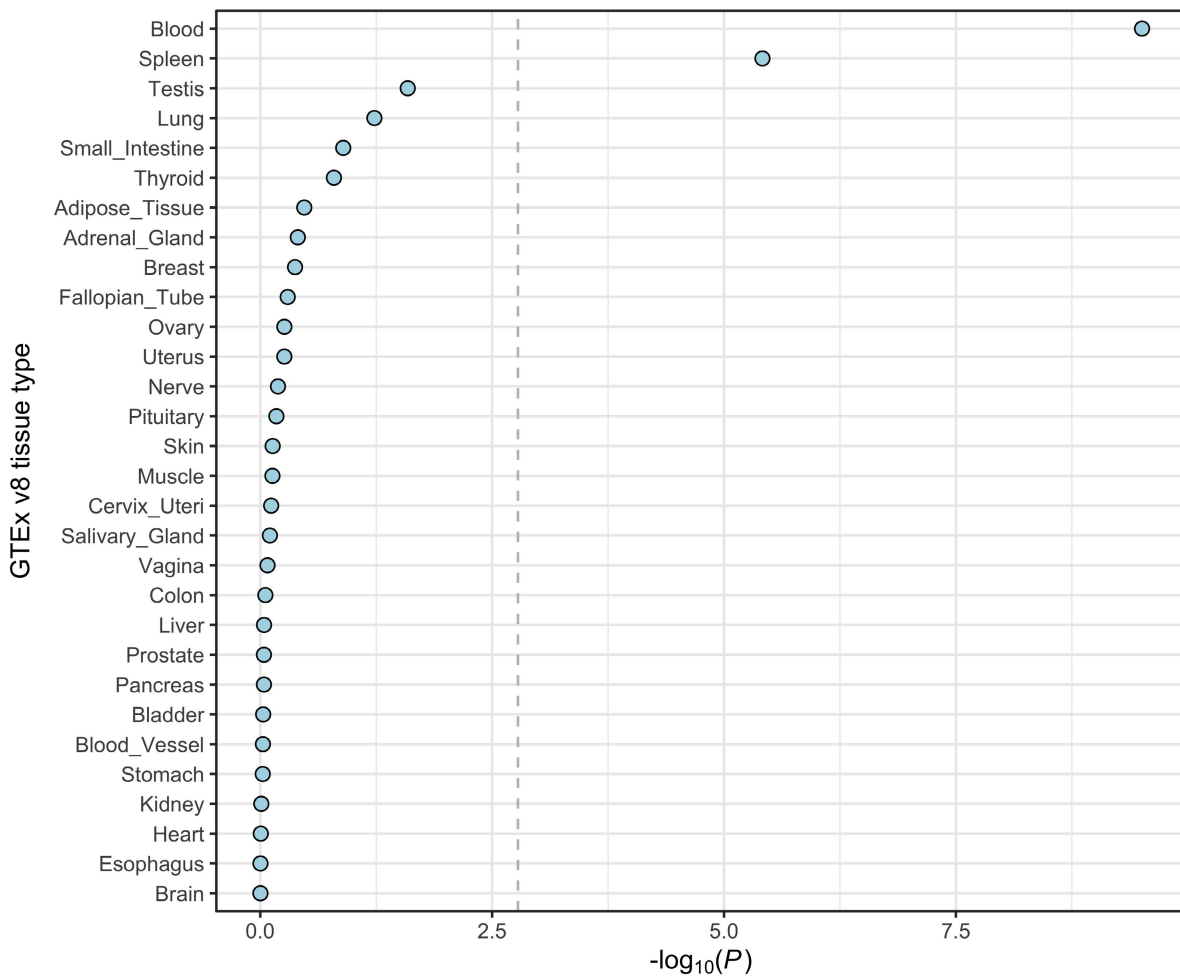


**Supplementary Fig. 10. Hispanic individual individual ancestral tract Manhattan and quantile-quantile (QQ) plots.** Manhattan plots (left) show the  $-\log_{10}(P)$  for associations of genetic variants for **a**, European tract, **b**, African tract, and **c**, Native American tract. QQ plots (right) show observed versus expected distributions of  $P$ -values..

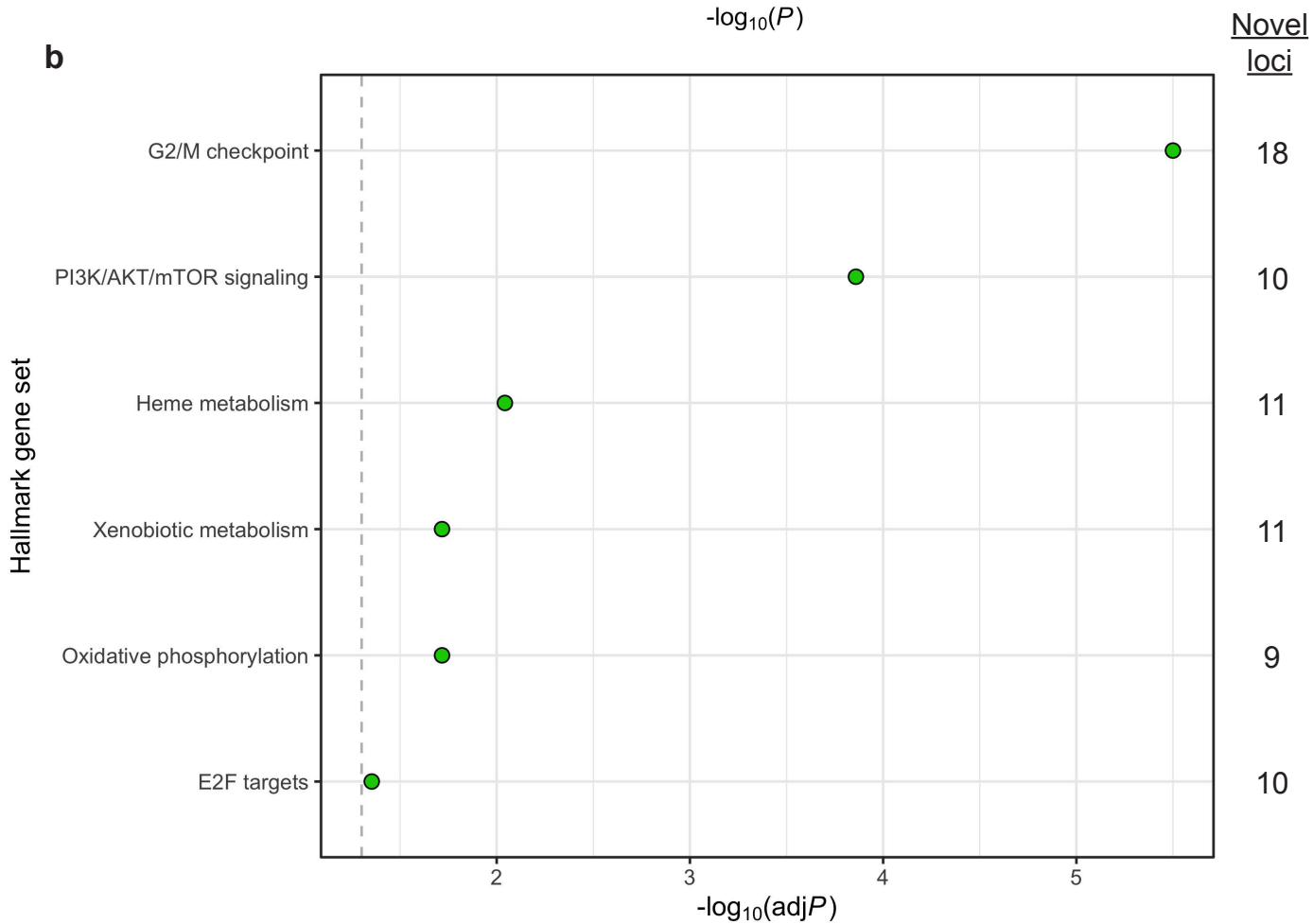


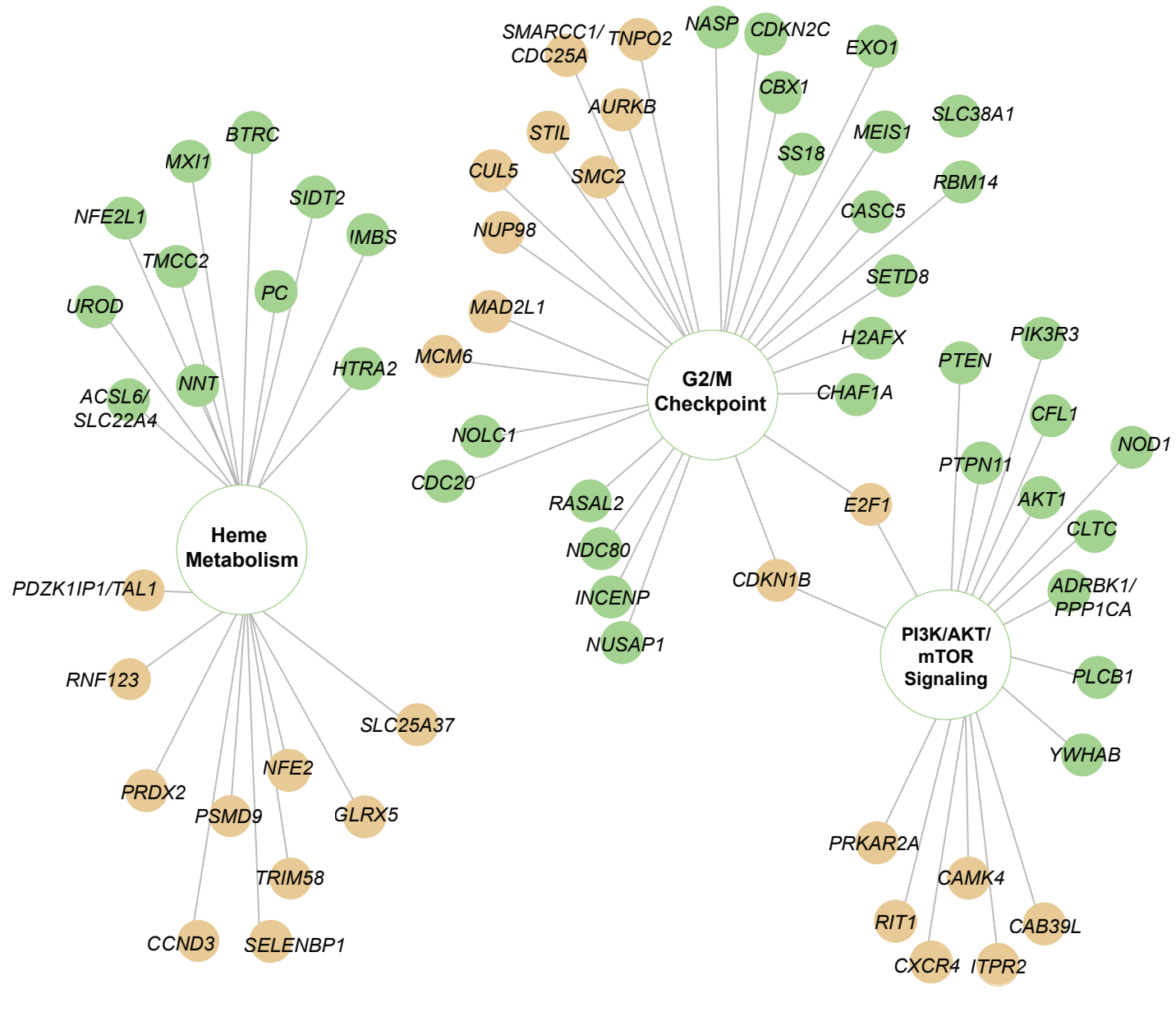
**Supplementary Fig. 11. Manhattan plots of Hispanic haplotype dosage enrichment from ancestral-tract-based model. a,** Manhattan plot of  $P$ -values for haplotype dosage enrichment of EUR with reference to AFR and NAT haplotype dosage in Million Veteran Program (MVP) admixed Hispanic (HIS) participants. **b,** Manhattan plot of  $P$ -values for haplotype dosage enrichment of NAT with reference to EUR and AFR haplotype dosage in MVP HIS (from the same multivariate model as a).

**S12a**

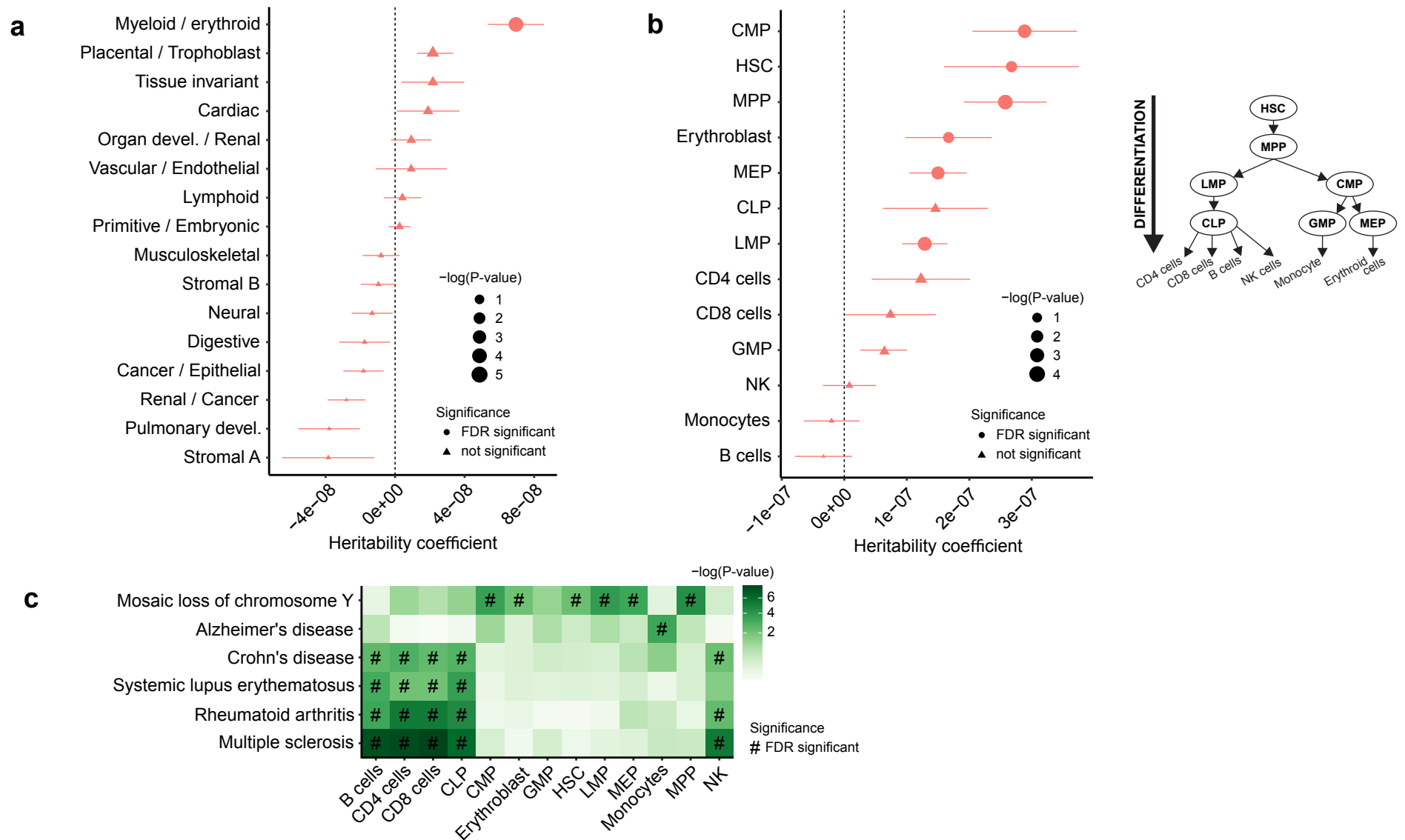


**b**



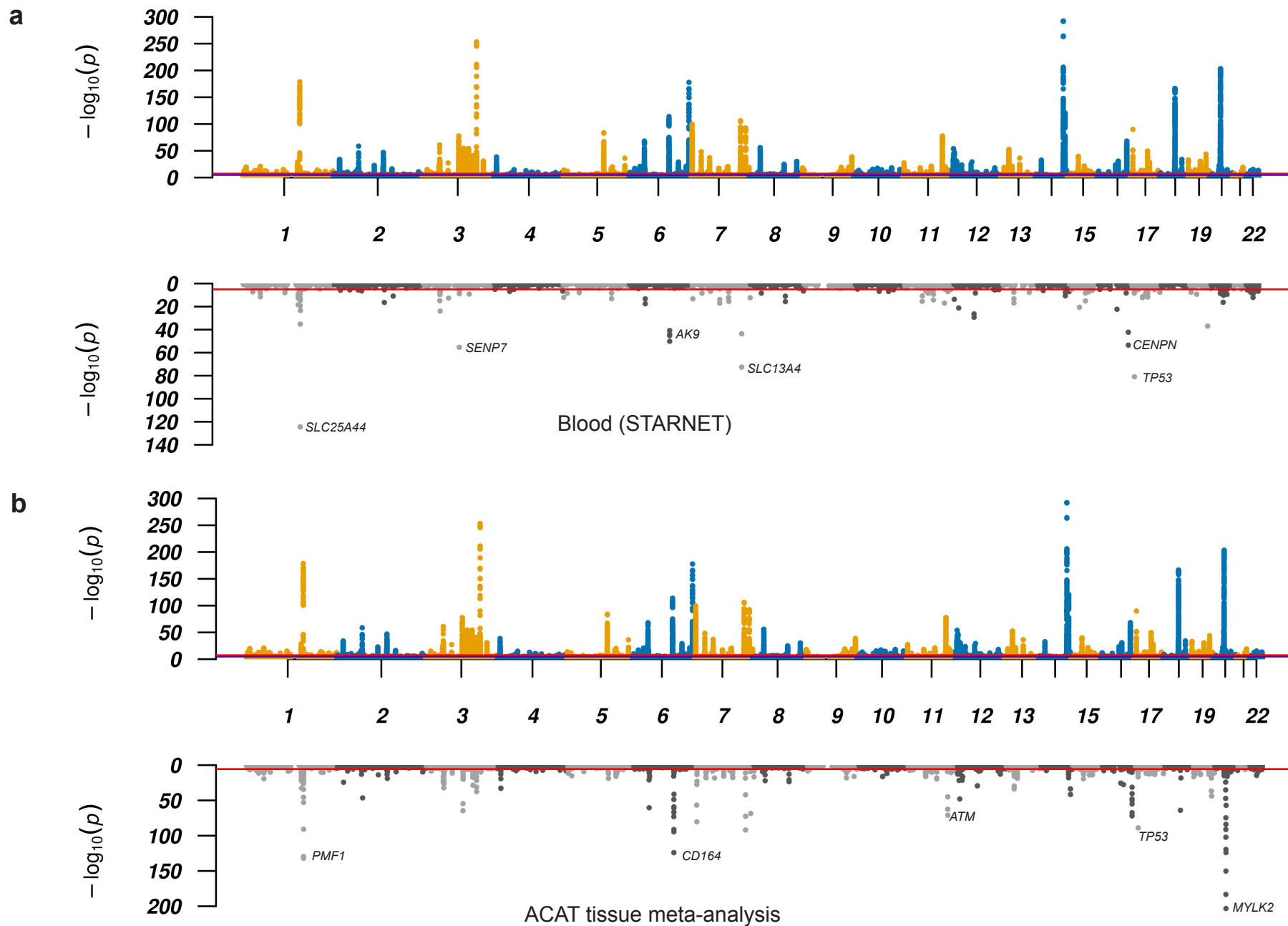


**Supplementary Fig. 12. Gene set enrichment of mLOY multi-ancestry GWAS meta-analysis.** (a) For all genes positionally mapped to significant mLOY multi-ancestry meta-analysis loci,  $-\log_{10}(P)$  for enrichment of GTEx tissue types. (b) For all genes positionally mapped to significant and novel mLOY multi-ancestry meta-analysis loci, FDR adjusted  $-\log_{10}(P)$  for enrichment of Hallmark gene sets (c) For the three most significantly enriched Hallmark gene sets for novel mLOY associations, gene graph showing novel and known loci.

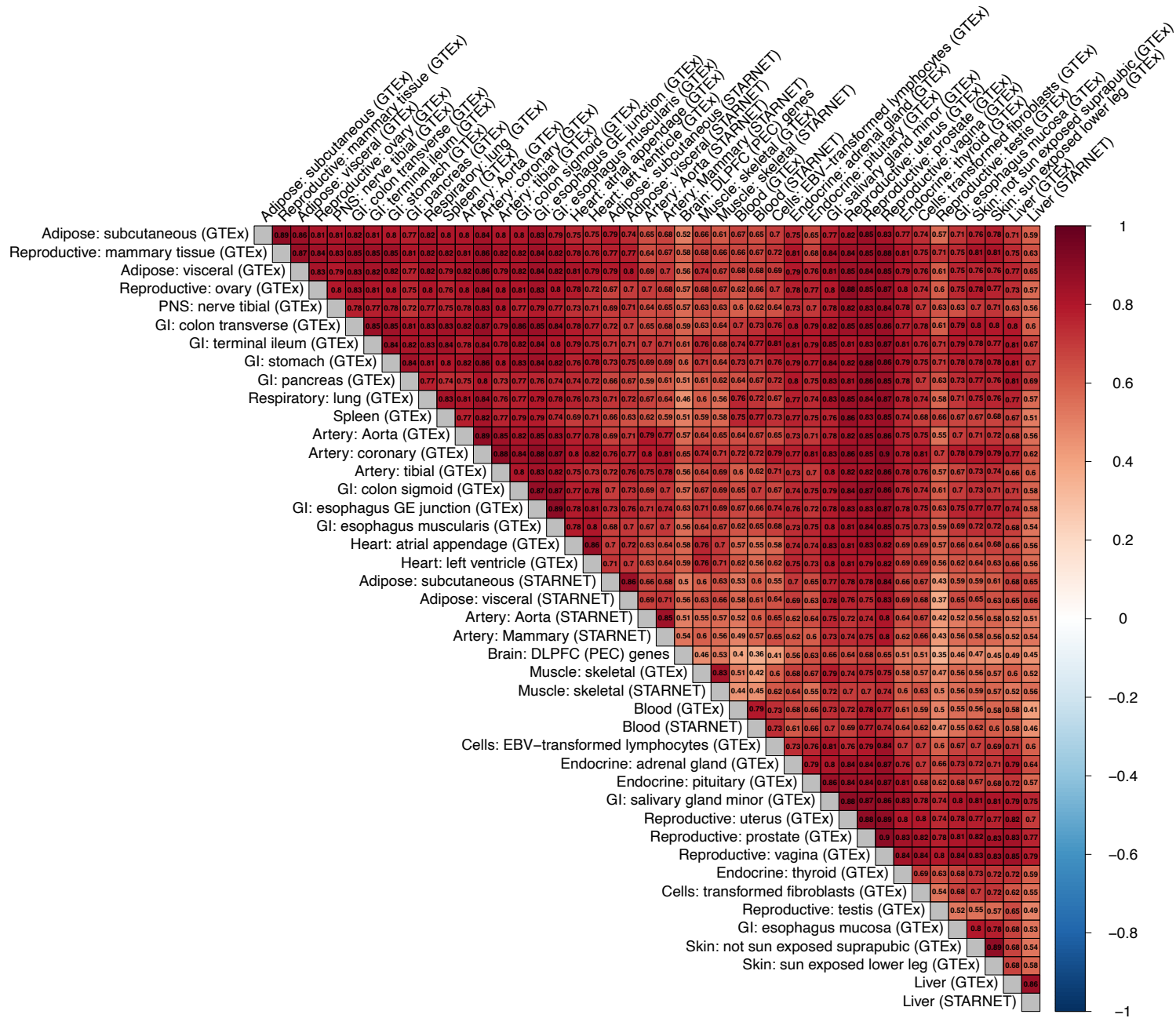


**Supplementary Fig. 13. Genomic enrichment of mLOY risk variants in open chromatin regions across human cell types. (a-b)** Enrichment of mLOY risk variants in cell-specific open chromatin regions from **(a)** main human body lineages and **(b)** blood lineages. Symbols (circle/triangle) represent the LD score coefficient and horizontal bars reflect standard error. The positive LD score coefficient signifies enrichment in heritability (horizontal bars reflect standard error). Symbol size reflects the value of LD score regression. **(Right)** Diagram illustrating the hierarchical structure of the human hematopoietic system, highlighting the blood lineages from the atlas. **(c)** Comparison of enrichment of blood lineages across a panel of diseases. CLP, common lymphoid progenitor cells; CMP, common myeloid progenitor cells; GMP, granulocyte-macrophage progenitor cells; HSC, hematopoietic stem cells; LMP, lymphoid-primed progenitor cells; MEP, megakaryote-erythroid progenitor cells; MPP, multipotent progenitor cells; NK, natural killer cells.

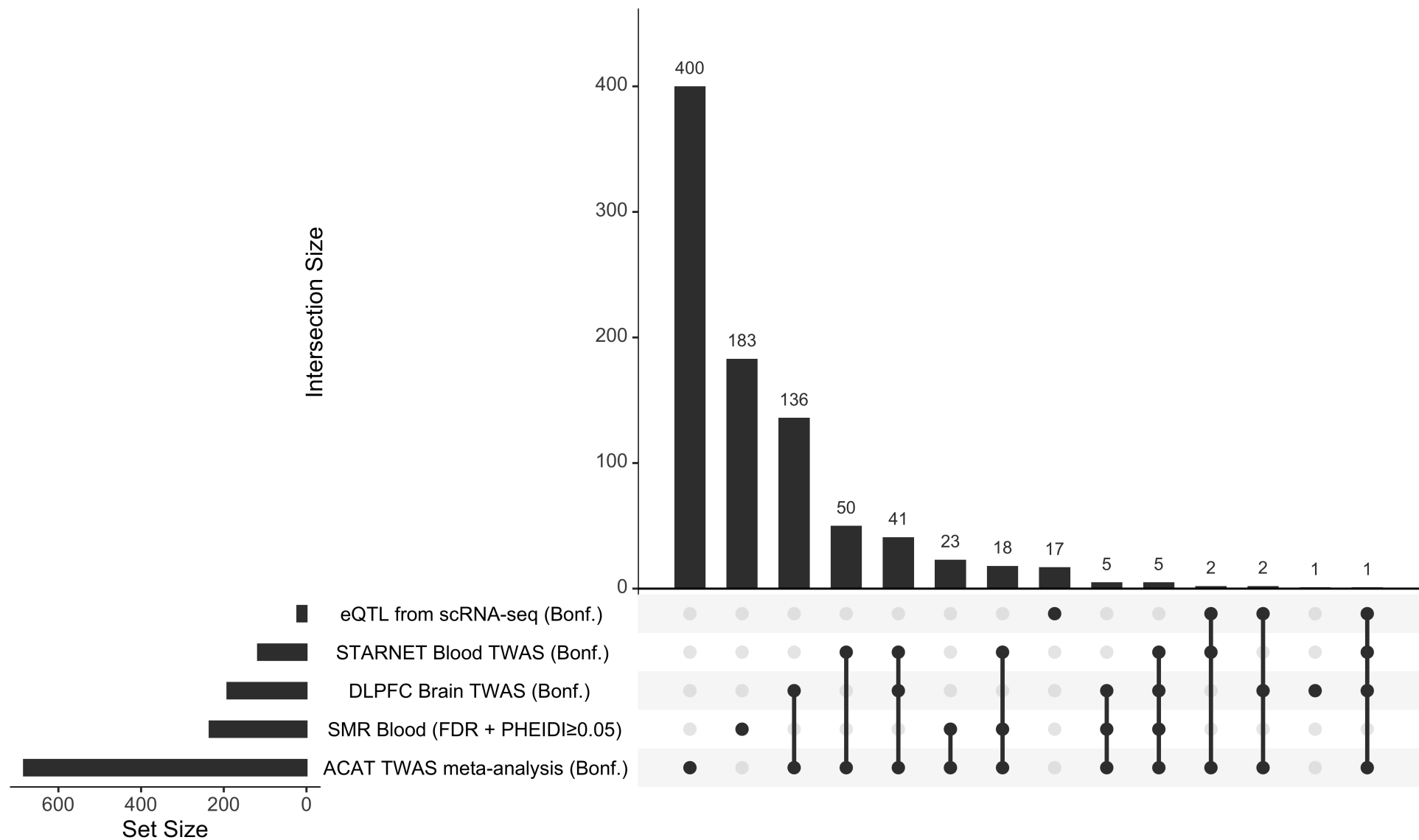




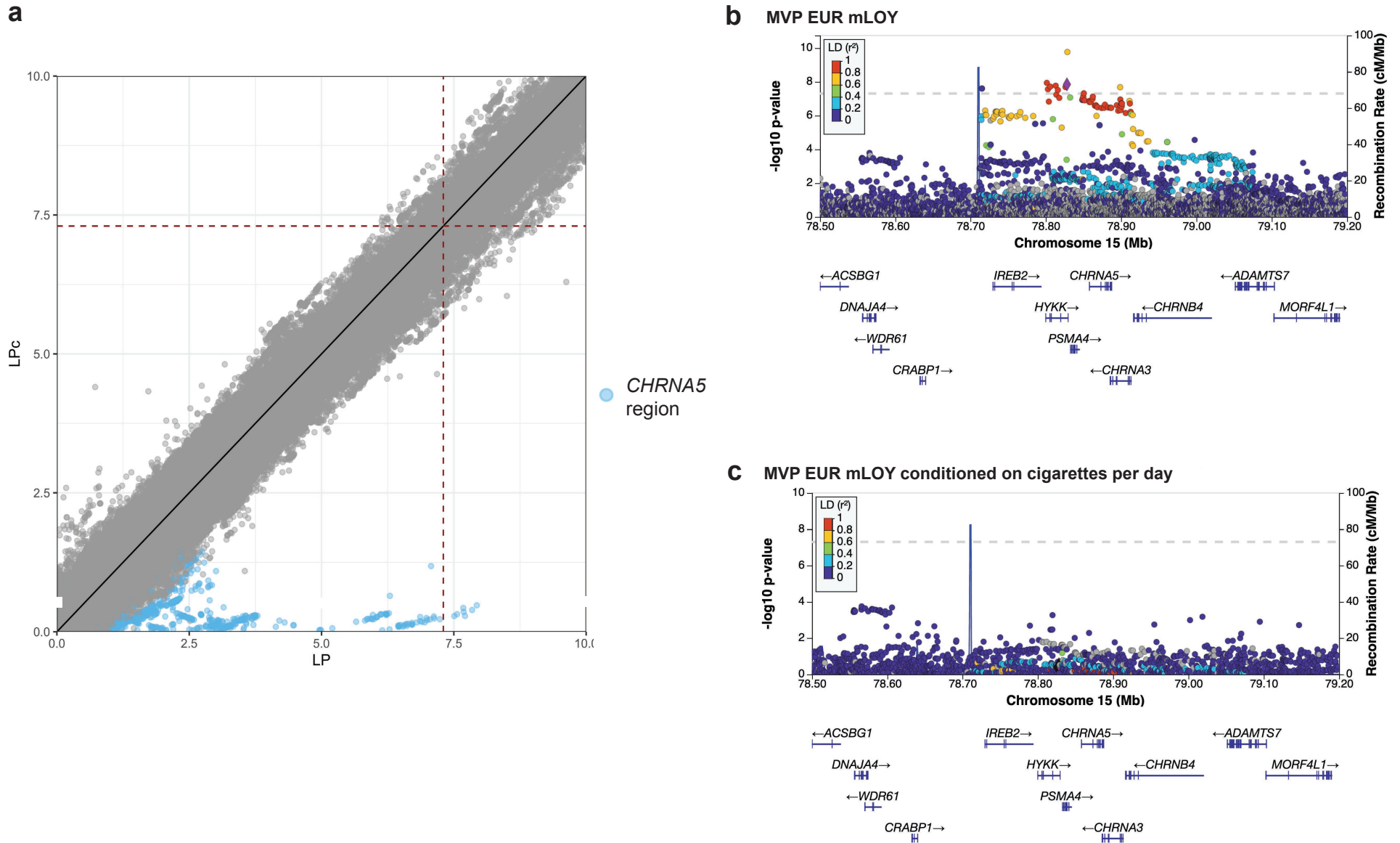
**Supplementary Fig. 14. Global Miami plots for transcriptome-wide association study.** Miami plots with Million Veteran Program European participants  $-\log_{10}(P)$  for associations of genetic variants from GWAS (top) and  $-\log_{10}(P)$  for associations of gene features from TWAS (bottom). Plots show TWAS results for **a**, Blood (STARNET) and **b**, ACAT tissue meta-analysis.



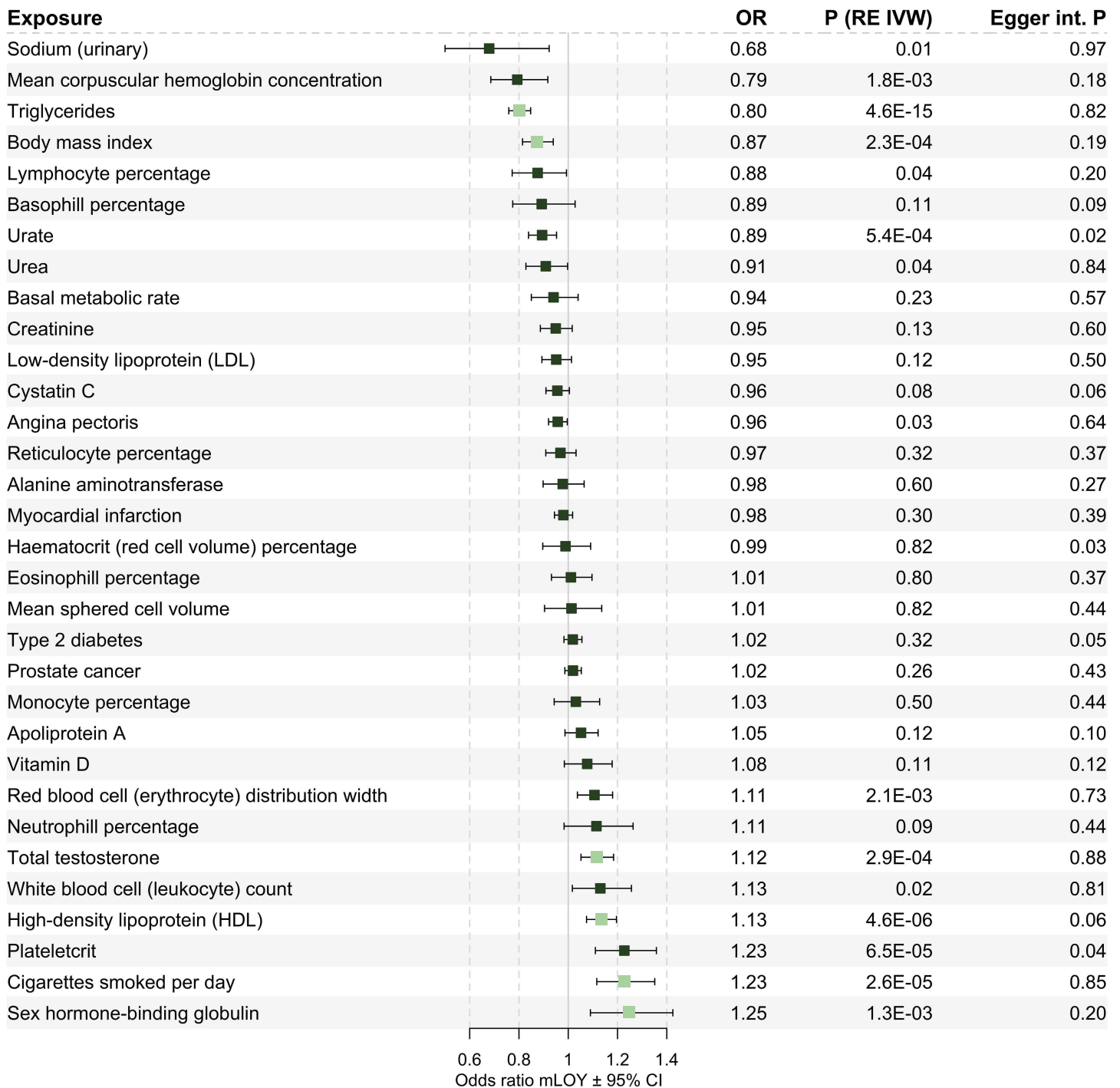
Supplementary Figure 15. Correlation of gene expression across tissues from transcriptome-wide association study (TWAS). Values represent Pearson correlation coefficient for imputed expression changes.



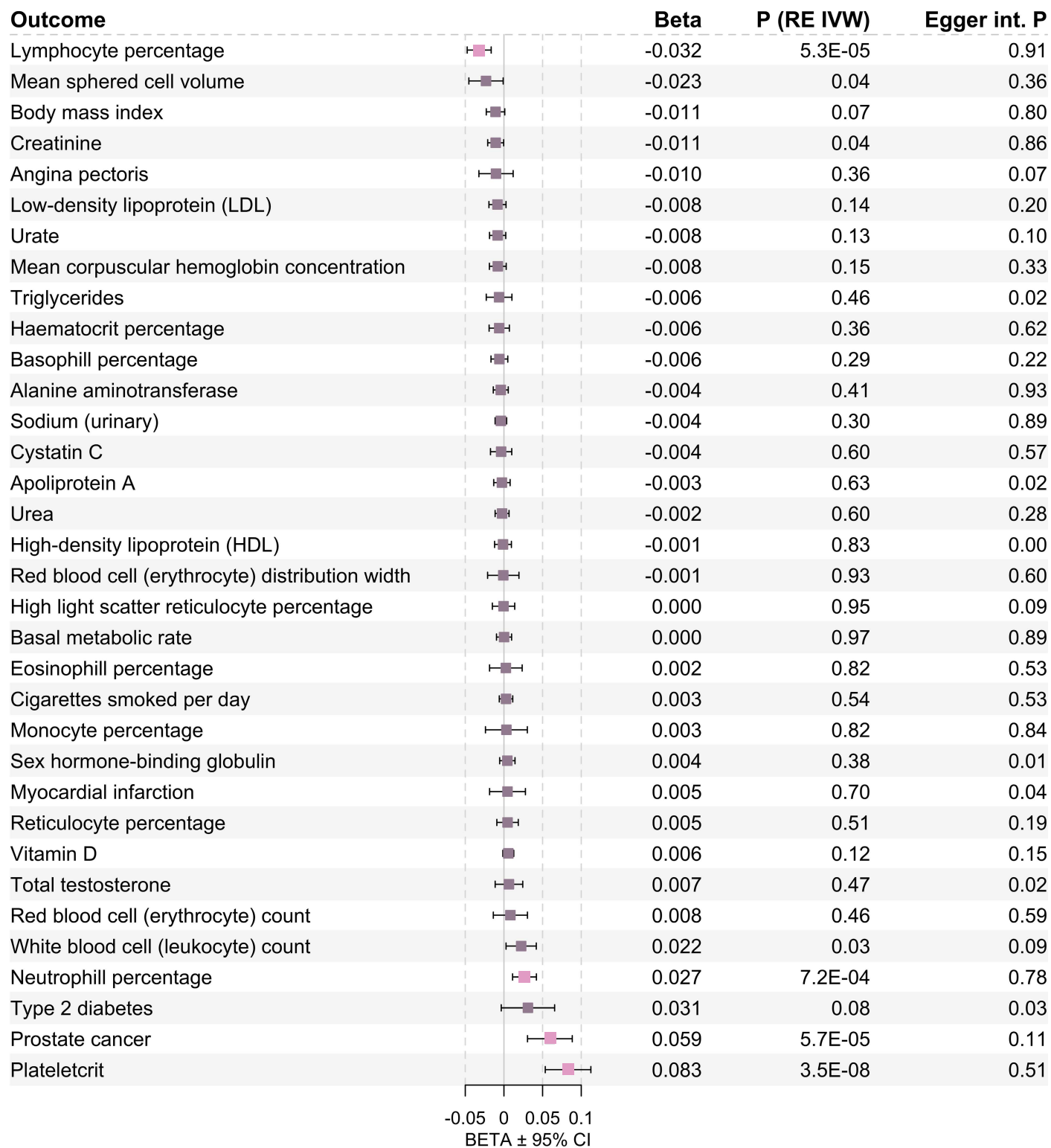
**Supplementary Fig. 16. UpSet plot showing overlap between gene transcript association methods.** Genes are derived from eQTLs with single cell RNA-seq (scRNA-seq) data across different immune cell populations, STARNET Blood transcriptome-wide association study (TWAS), dorsolateral prefrontal cortex TWAS, Summary Mendelian Randomization (SMR), and ACAT TWAS meta-analysis. Bonf., Bonferroni-corrected significance; FDR, false discovery rate.



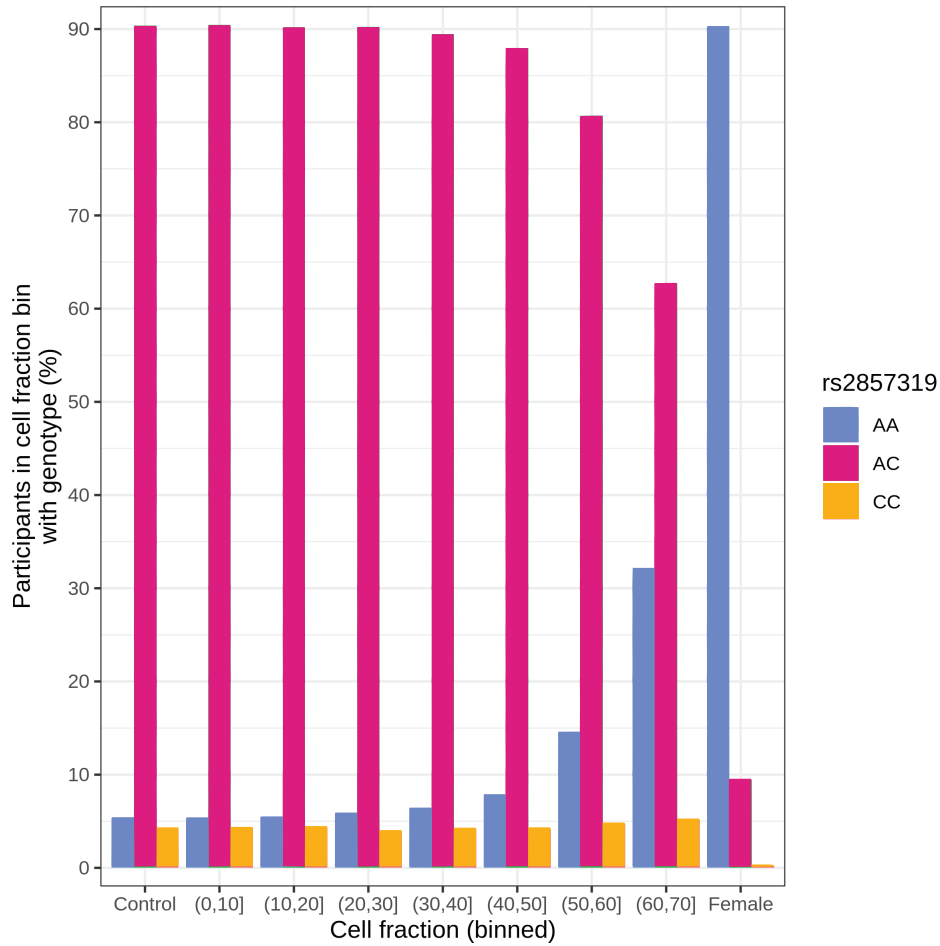
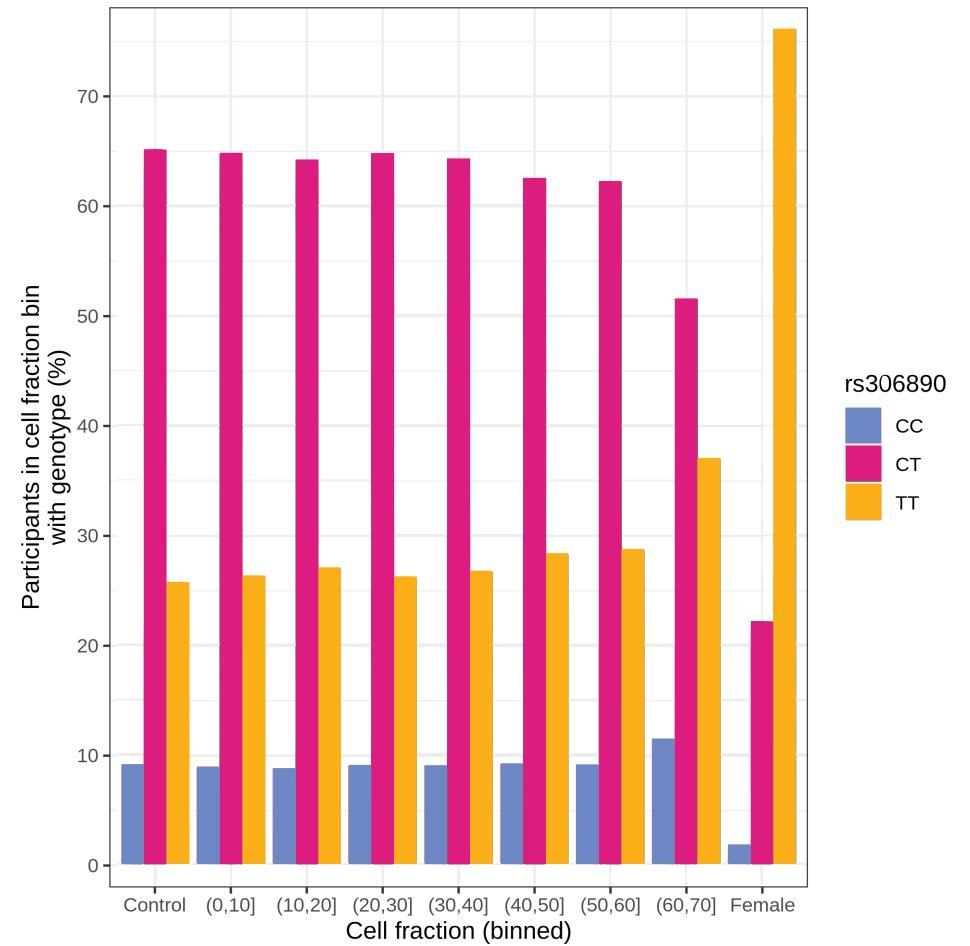
**Supplementary Fig. 17. Multi-trait meta-analysis (mtCOJO) conditional analysis on smoking.** (a) Scatterplot of  $-\log(P)$  from conditioning on cigarettes-per-day with mtCOJO (LPC) vs.  $-\log(P)$  from standard mLOY GWAS (LP) in MVP European ancestry (EUR) demonstrates the region containing the *CHRNA5* gene (chr15:78,711,000-79,258,000), a known smoking locus, was primarily affected by conditioning. Red dashed lines indicate genome-wide significance at  $P < 5 \times 10^{-8}$ . (b) Regional Manhattan plot for *CHRNA5* in MVP EUR. (c) Regional Manhattan plot for *CHRNA5* in MVP EUR after conditioning mLOY on cigarettes-per-day.



**Supplementary Fig. 18. Forward Mendelian Randomization (MR) for all exposure traits.** Forest plots showing random effects inverse-variance weighted (RE IVW) MR for 32 exposure traits on mLOY outcome in Million Veteran Program Europeans. Lighter box color denotes  $P$  (RE IVW) < 0.05/32 and  $P$  MR-Egger intercept (int.)  $\geq$  0.05. OR, odds ratio; CI, confidence interval.



**Supplementary Fig. 19. Reverse Mendelian Randomization (MR) for mLOY exposure on all trait outcomes.** Forest plots showing random effects inverse-variance weighted (RE IVW) MR for mLOY exposure on 32 trait outcomes in Million Veteran Program Europeans. Lighter box color denotes  $P < 0.05/32$  and  $P$  MR-Egger intercept (int.)  $\geq 0.05$ . CI, confidence interval.

**a****b**

**Supplementary Fig. 20. Rejected chromosome X associations.** Two genome-wide significant variants on chromosome X, **(a)** rs2857319 and **(b)** rs306890, near the borders of PAR1 and PAR2, respectively, were found to be likely false positives. Plotting the percentage of participants with each genotype across cell fraction bins reveals a heterozygous dropout in higher cell fractions, suggesting genotyping error.