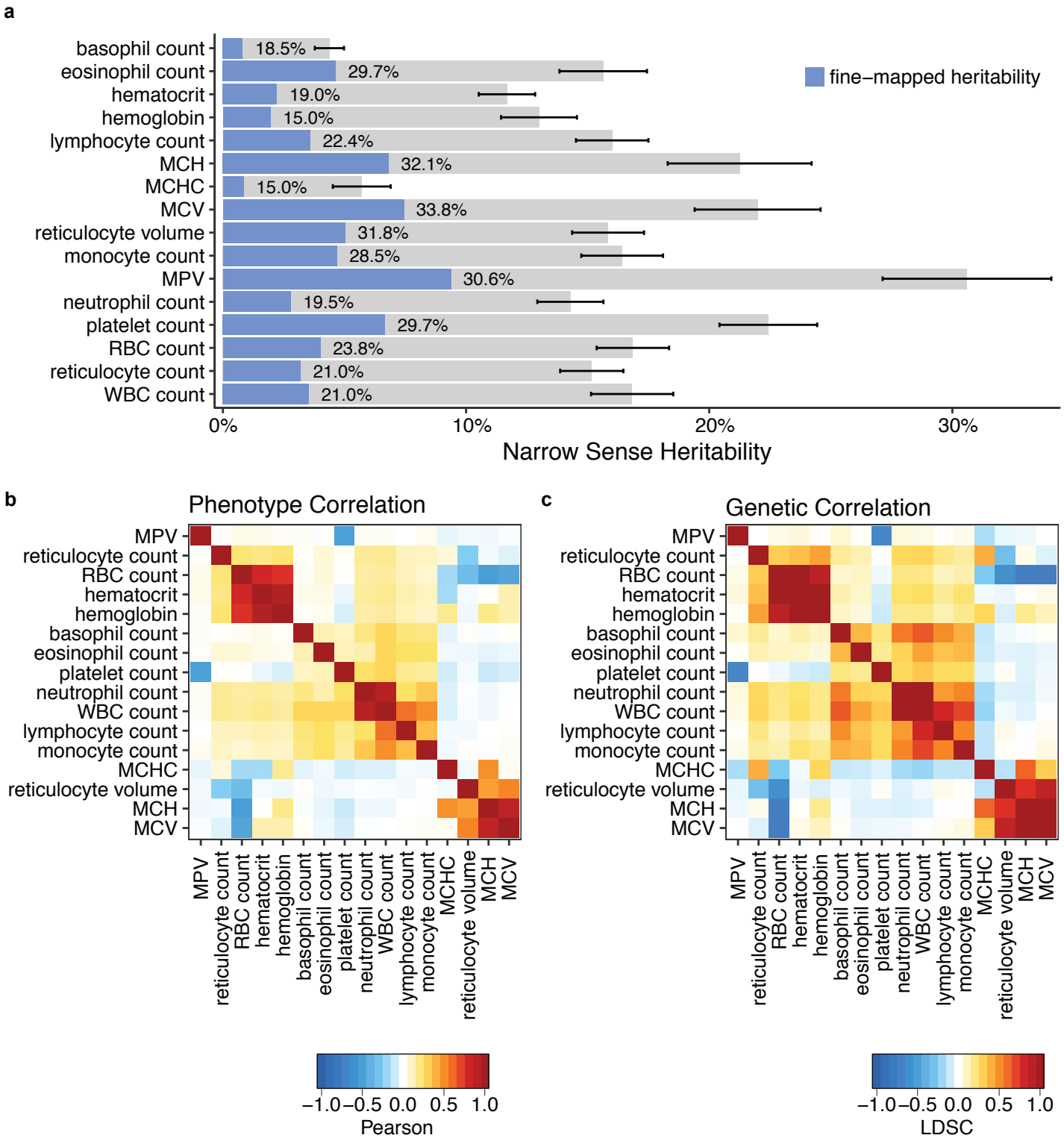
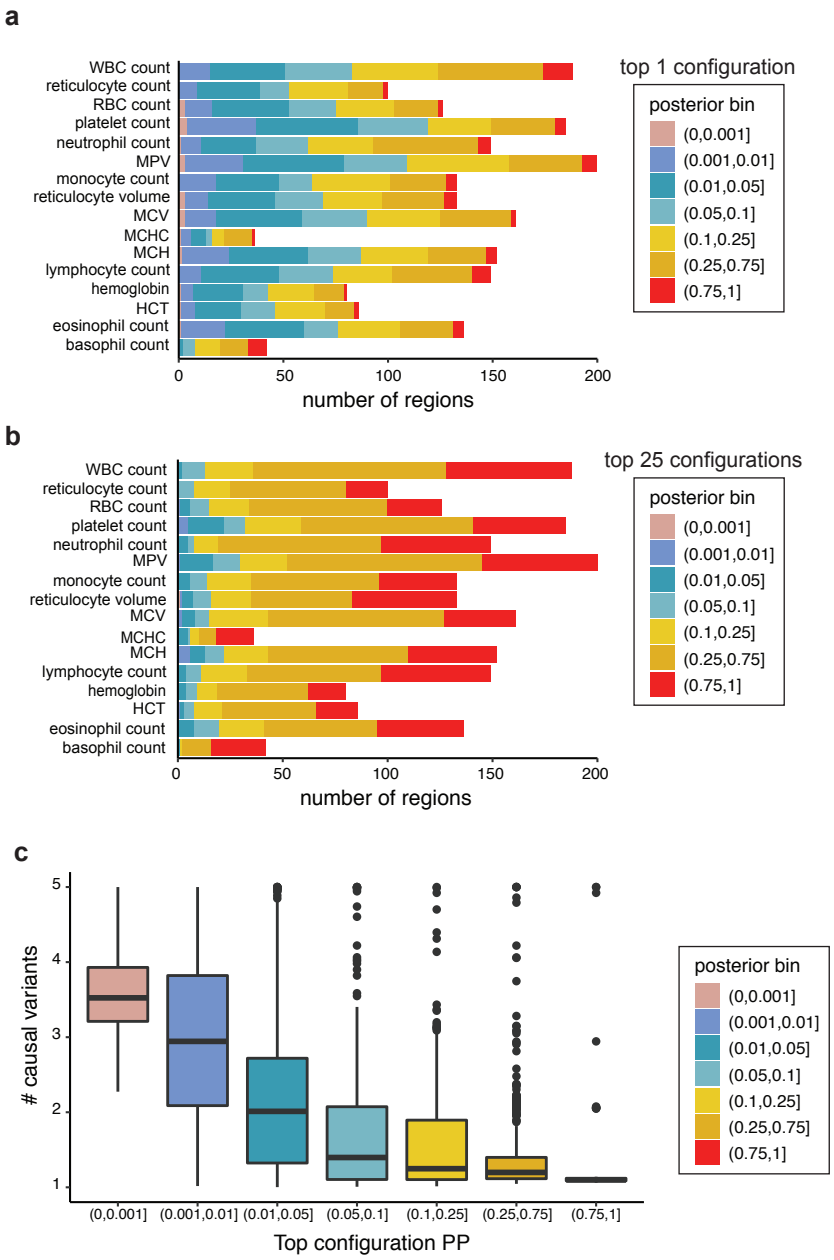
Supplementary information for: Interrogation of human hematopoiesis at single-cell and single-variant resolution

January 12th, 2018

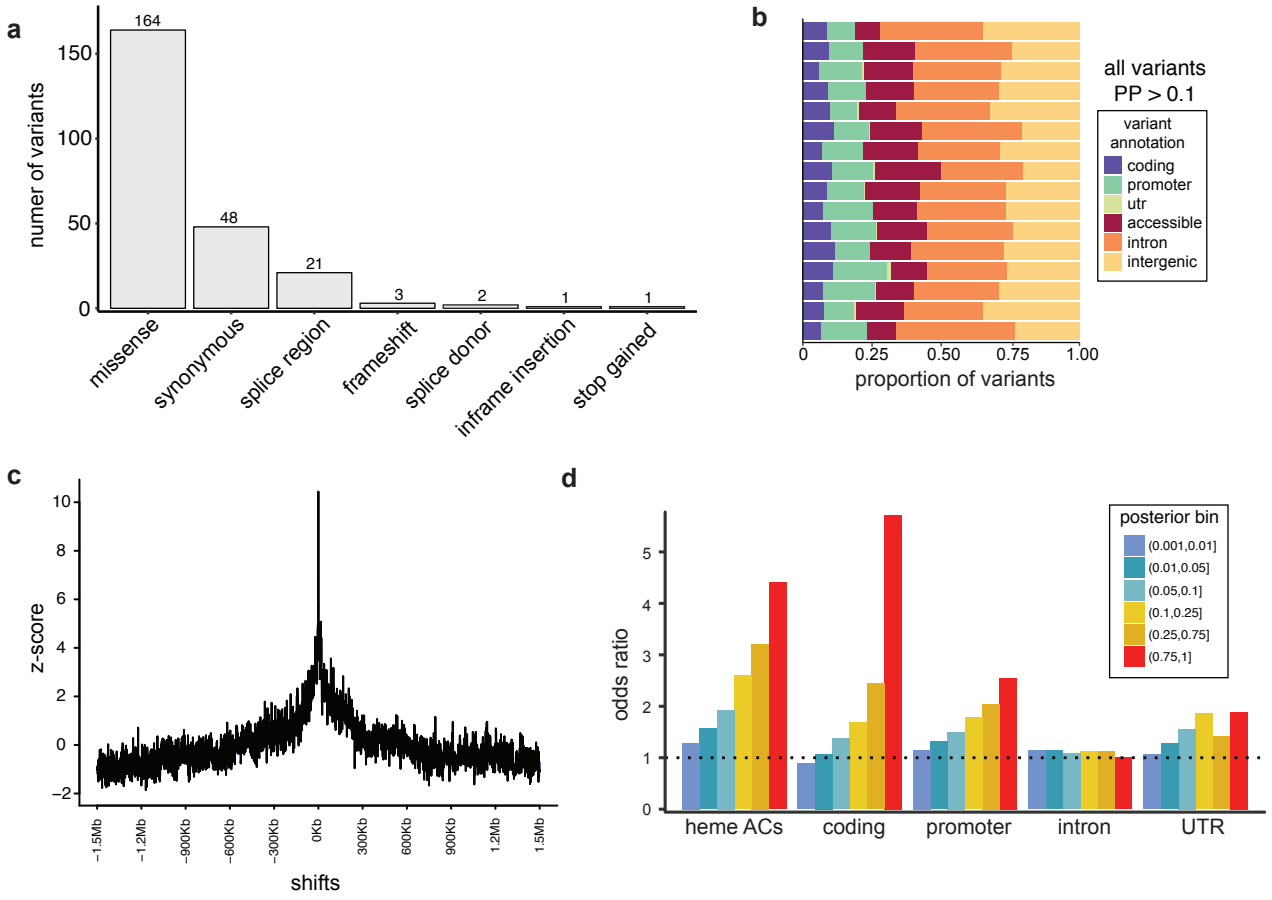
Contents



Supplementary Fig. 1. Narrow-sense heritability and genetic / phenotypic correlations between blood traits. (**a**) Estimates of narrow-sense SNP heritability are plotted in gray with corresponding standard errors. Heritability estimates for all variants with fine-mapped PP > 0.001 are plotted in blue for each trait (n = 114,910- 116,667 individuals), and the proportions of total narrow-sense heritability captured by these fine-mapped variants (blue bar / gray bar) are indicated by the numbered labels. (**b**) Phenotypic and (**c**) genetic correlations across the 16 traits examined (n = 114,910- 116,667 individuals).



Supplementary Fig. 2. Fine-mapped configurations. Total PPs for the top (**a**) 1 and (**b**) 25 configurations per region across all traits. (**c**) The estimated number of causal variants in all fine-mapped regions, stratified by the PP of the top configuration in each region (n = 2,056 regions). The observed association is likely a reflection of our combined ability to fine-map each independent association, which is predominately determined by each association's strength and LD structure. Boxplots represent median and interquartile range.



Supplementary Fig. 3. Additional fine mapping diagnostics. (**a**) Distribution of all coding variants with posterior probability (PP) > 0.10, as annotated by Variant Effect Predictor (VEP).[1] (**b**) Proportion of fine-mapped variants with PP > 0.10 which fall into coding, promoter, UTR, hematopoietic chromatin accessible, intron, or intergenic regions. Overall, 95.0% of variants with PP > 0.10 are non-coding. (**c**) Local z-scores for enrichment of hematopoietic chromatin-accessible regions in the set of fine-mapped variants with PP > 0.10. (**d**) Local shifting enrichments excluding all variants with high correlation (R^2 > 0.6) to the sentinel variants.

a

UKBB Blood Finemap Viewer | Variants | PheWAS | Molecular Function | Target Genes | About

UKBB Blood Traits Data Browser

Aryee, Buenrostro, and Sankaran Labs

b

Highlight Trait

MONO_COUNT

Select Region Coordinates

chr16:85916328-86116328

Plot Region

Show Gene in Region

KIT

Plot Gene

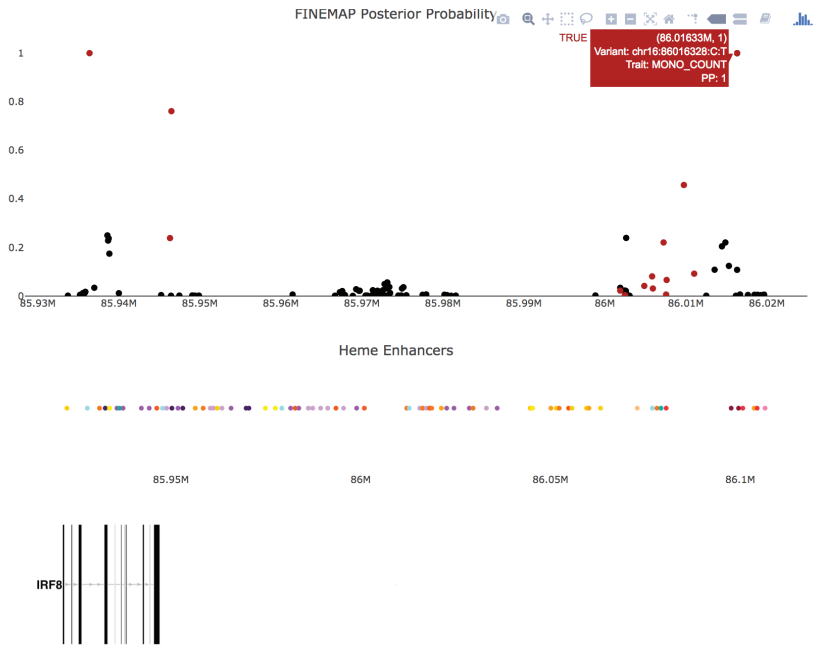
Show SNP in Region

rs11642657

Plot SNP

Zoom In | Zoom Out | Clear

<< | < | > | >>



c

Show 10 entries | Search: 86016328

	chr	variantPos	PP	Trait	EnhancerStart	EnhancerEnd	Gene	CellType	Value
79052	chr16	86016328	1	MONO_COUNT	86012163	86022549	COX4I1	Mon	7.38969053653433
79053	chr16	86016328	1	MONO_COUNT	86012163	86022549	EMC8	Mon	7.38969053653433
79066	chr16	86016328	1	MONO_COUNT	86012163	86022549	IRF8	Mon	10.6669028354115
102671	chr16	86016328	0.1082	NEUTRO_COUNT	86012163	86022549	COX4I1	Mon	7.38969053653433
102672	chr16	86016328	0.1082	NEUTRO_COUNT	86012163	86022549	EMC8	Mon	7.38969053653433
102697	chr16	86016328	0.1082	NEUTRO_COUNT	86012163	86022549	IRF8	Mon	10.6669028354115
147006	chr16	86016328	0.9997	WBC_COUNT	86012163	86022549	COX4I1	Mon	7.38969053653433
147007	chr16	86016328	0.9997	WBC_COUNT	86012163	86022549	EMC8	Mon	7.38969053653433
147013	chr16	86016328	0.9997	WBC_COUNT	86012163	86022549	IRF8	Mon	10.6669028354115
227829	chr16	86016328	1	MONO_COUNT	86012163	86022549	IRF8	Mac0	13.9055756998611

Showing 1 to 10 of 51 entries (filtered from 2,683,717 total entries) | Previous 1 2 3 4 5 6 Next

Download promoter capture Hi-C Associations

d

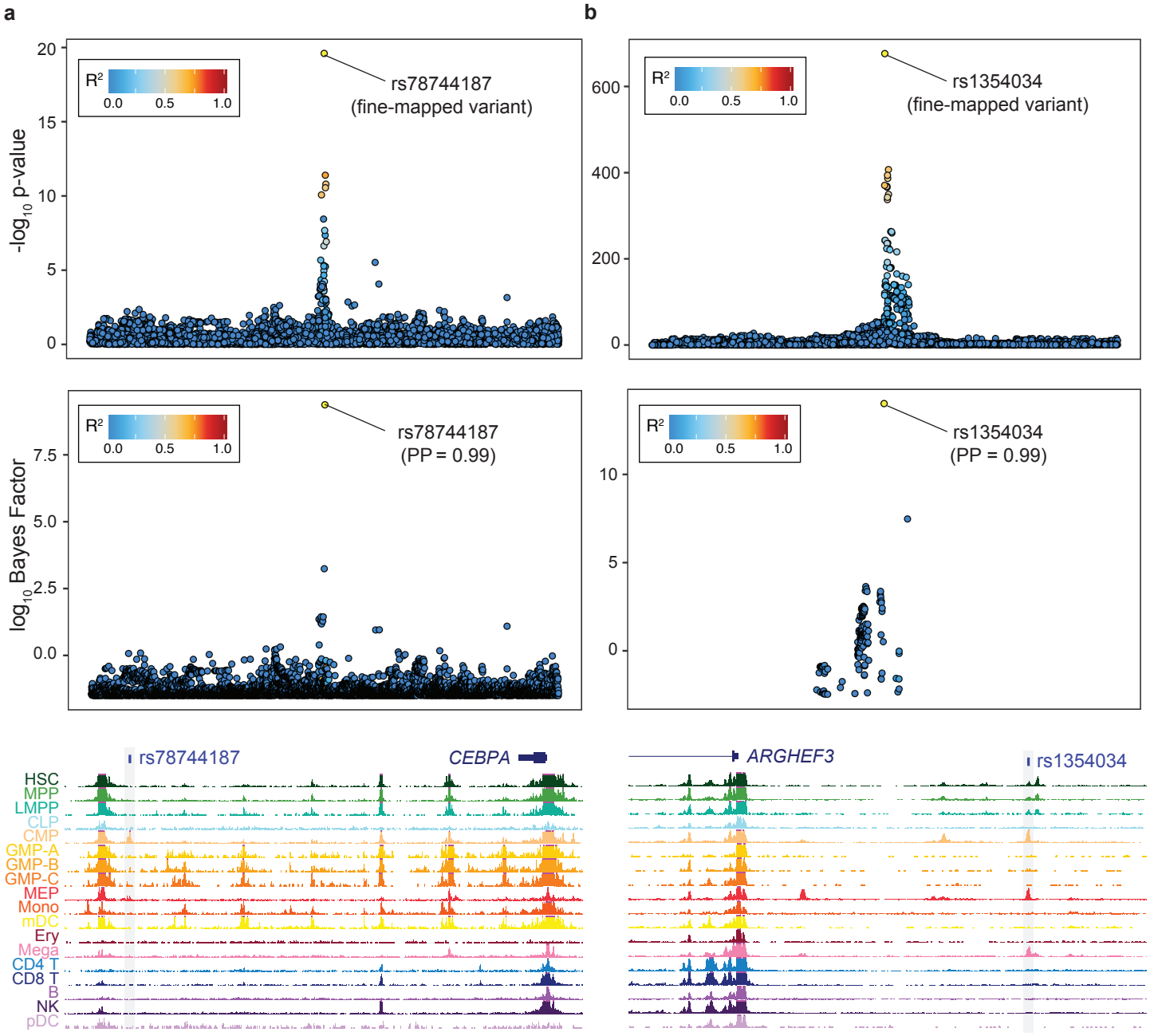
Show 10 entries | Search: 86016328

	ICD10	chr	pos	PP	pval	V2	trait
146342	A87	chr16	86016328	1	0.00274305	Diagnoses - main ICD10: A87 Viral meningitis	MONO_COUNT
148170	D12	chr16	86016328	1	0.00335234	Diagnoses - main ICD10: D12 Benign neoplasm of colon, rectum, anus and anal canal	MONO_COUNT
155628	J98	chr16	86016328	1	0.00451412	Diagnoses - main ICD10: J98 Other respiratory disorders	MONO_COUNT
162322	N97	chr16	86016328	1	0.00897047	Diagnoses - main ICD10: N97 Female infertility	MONO_COUNT
200989	A87	chr16	86016328	0.1082	0.00274305	Diagnoses - main ICD10: A87 Viral meningitis	NEUTRO_COUNT
202224	D12	chr16	86016328	0.1082	0.00335234	Diagnoses - main ICD10: D12 Benign neoplasm of colon, rectum, anus and anal canal	NEUTRO_COUNT
209177	J98	chr16	86016328	0.1082	0.00451412	Diagnoses - main ICD10: J98 Other respiratory disorders	NEUTRO_COUNT
215881	N97	chr16	86016328	0.1082	0.00897047	Diagnoses - main ICD10: N97 Female infertility	NEUTRO_COUNT
293046	A87	chr16	86016328	0.9997	0.00274305	Diagnoses - main ICD10: A87 Viral meningitis	WBC_COUNT
294987	D12	chr16	86016328	0.9997	0.00335234	Diagnoses - main ICD10: D12 Benign neoplasm of colon, rectum, anus and anal canal	WBC_COUNT

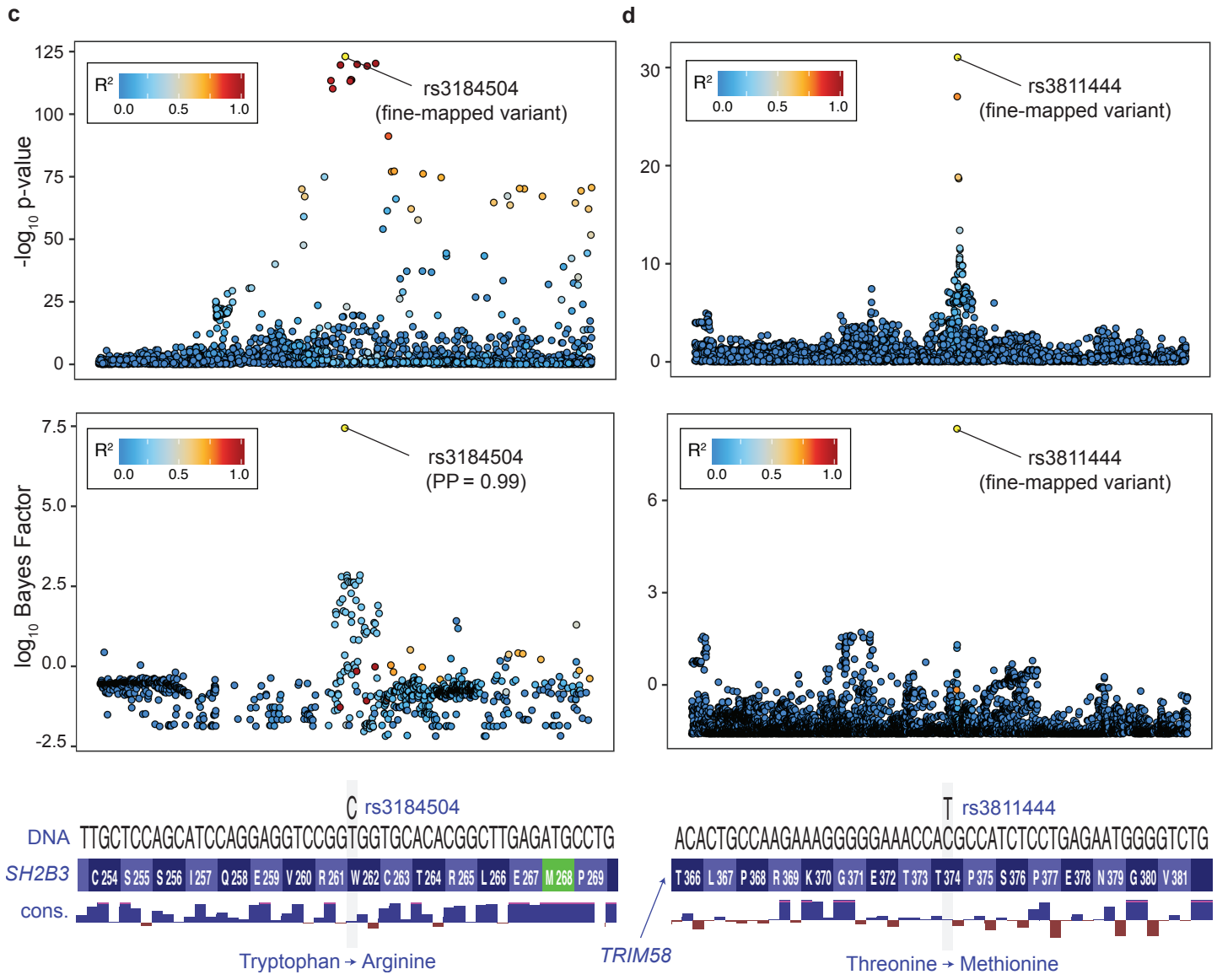
Showing 1 to 10 of 12 entries (filtered from 318,284 total entries) | Previous 1 2 Next

Download PheWAS Data

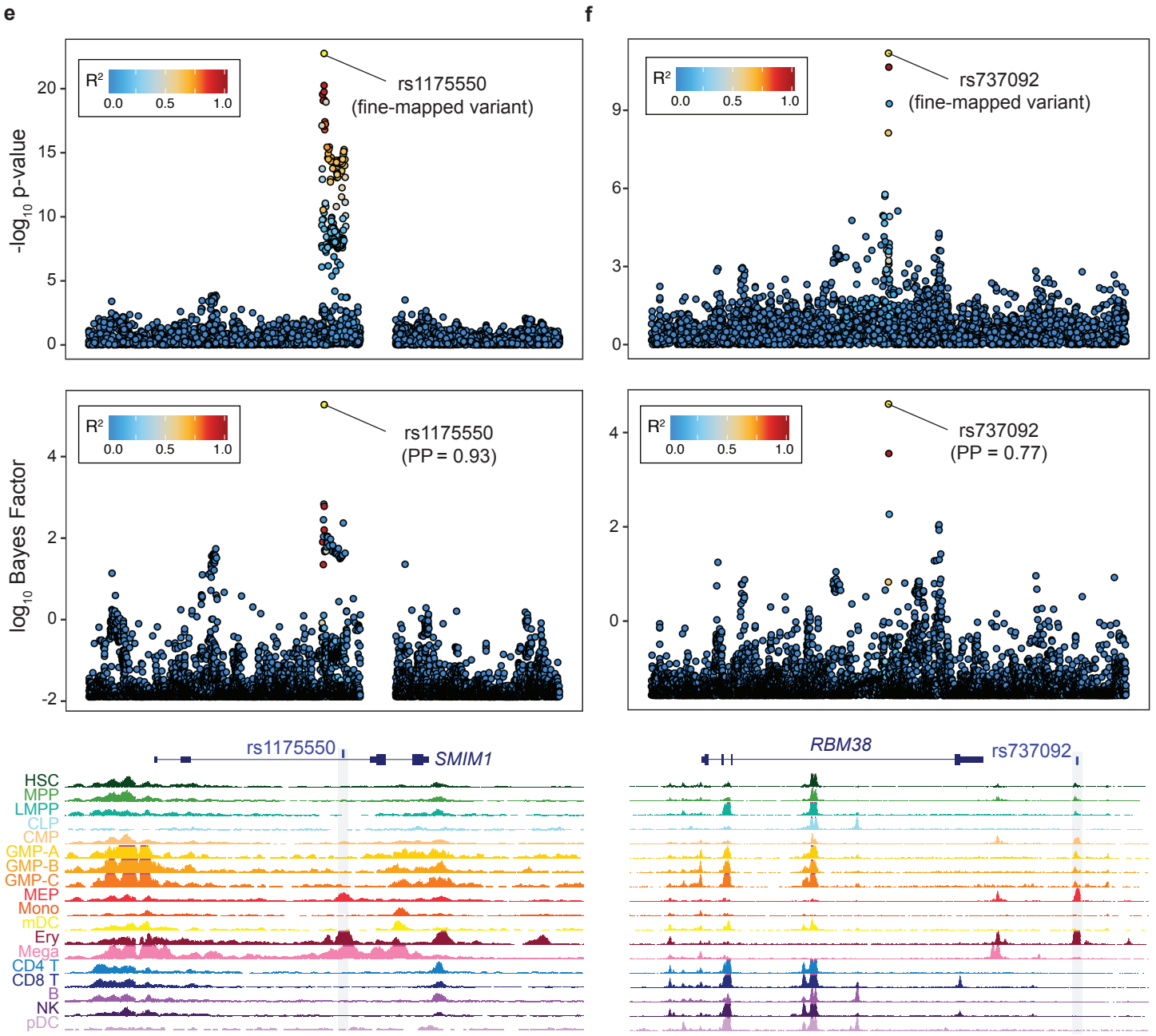
Supplementary Fig. 4. Overview of web resource for fine-mapped variants. (**a-d**) This interactive resource provides users with an integrated experience linking variants identified in this study with disease-relevant traits, putative target genes, and plausible molecular functions.



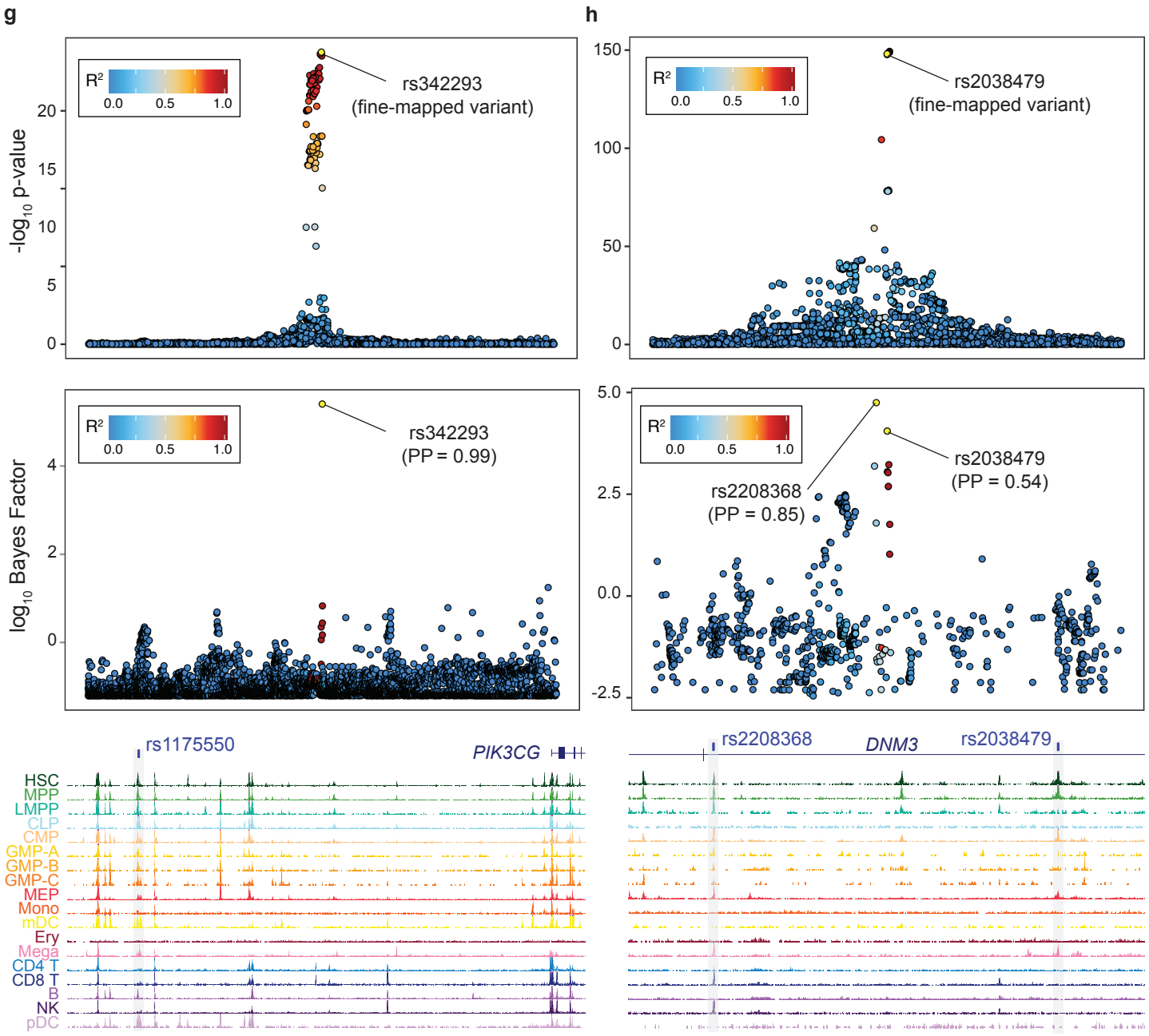
Supplementary Fig. 5. Previously identified causal variants corroborated by fine-mapping results and genomic annotations. (**a**) rs78744187 is associated with basophil count (n = 116,482 individuals; BOLT-LMM p-values) and was shown in Guo et al.[2] to lie in a CMP-specific enhancer that regulates *CEBPA* expression to regulate basophil development and was investigated using reporter assays and genome editing. Although this locus is imputed and poorly tagged by genotyped variants, it is readily resolved by our fine-mapping method. (**b**) rs1354034 is associated with platelet traits (n = 116,663 individuals; BOLT-LMM p-values) and was shown in Zou et al.[3] to lie within a megakaryocyte enhancer and is associated (by eQTL) with expression of *ARHGEF3*; *Arhgef3* KO mice were then shown to have larger platelets than normal. This variant is predicted to disrupt a GATA motif and GATA factors are observed to be bound here by ChIP-seq in several hematopoietic cells.



(**c**) rs3184504 is a nonsynonymous variant in *SH2B3* associated with multiple blood cell traits (n = 116,666 individuals; BOLT-LMM p-values); *SH2B3* is a negative regulator of several signaling pathways, and its perturbation by either genome editing or naturally occurring loss of function mutations has been shown to have a role in erythroid differentiation.[4] (**d**) rs3811444 is a nonsynonymous variant in *TRIM58* associated with red cell traits (n = 116,667 individuals; BOLT-LMM p-values); knockdown of *TRIM58* by RNAi previously showed a role for this gene in erythroblast enucleation through the regulation of dynein degradation.[5]

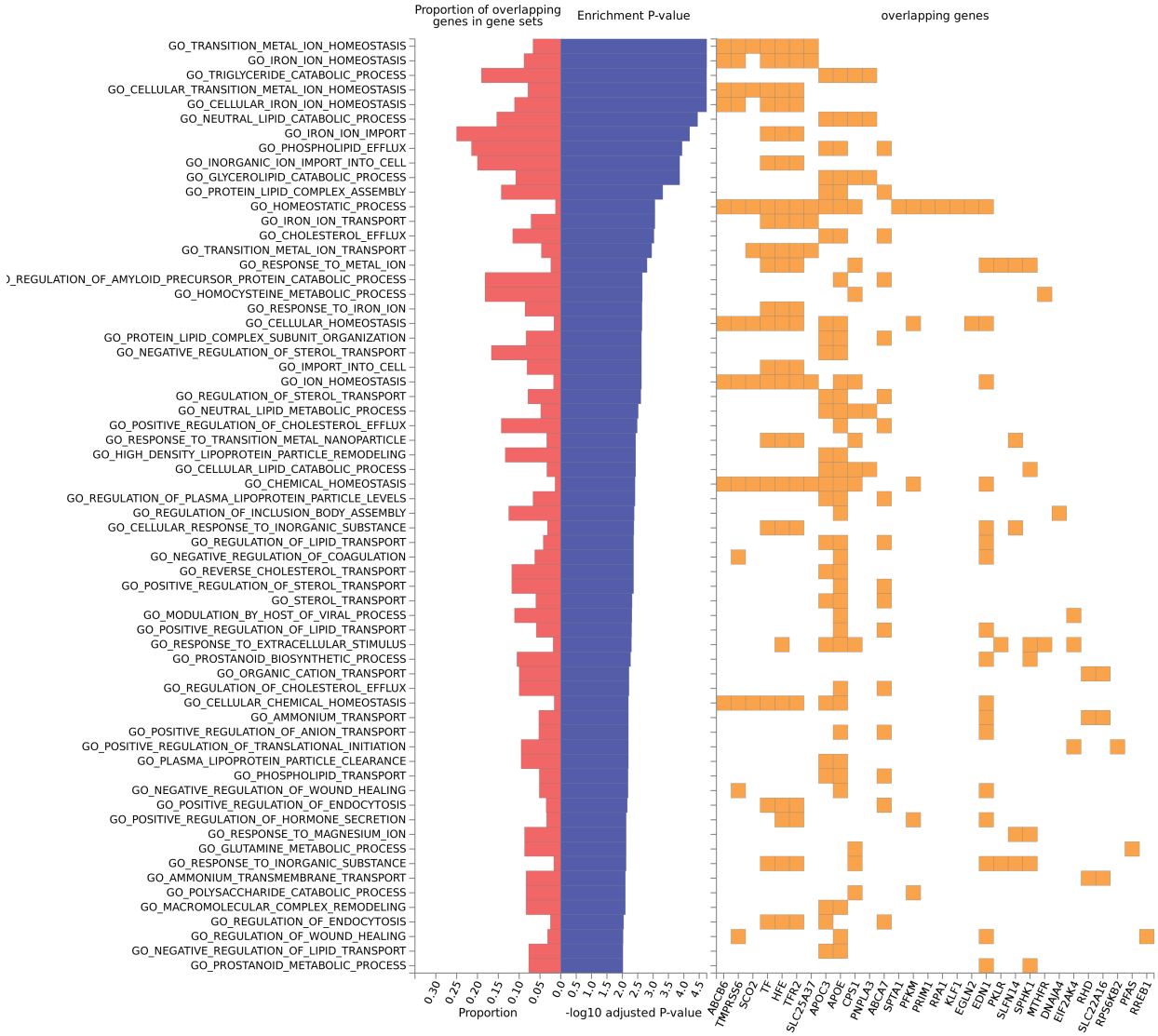


(**e**) rs1175550 is an erythroid enhancer variant associated with red cell traits (n = 114,910 individuals; BOLT-LMM p-values) that is associated with *SMIM1* expression by genome editing and *SMIM1* protein (Vel antigen) by eQTL.[6,7] The Vel antigen is a blood group, and Vel-negative humans often have transfusion reactions to Vel-positive blood. (**f**) rs737092 is an erythroid enhancer variant associated with red cell traits that regulates *RBM38* expression (by genome editing).[7] *RBM38* knockdown by RNAi has been shown to regulate RNA splicing and erythroid maturation.

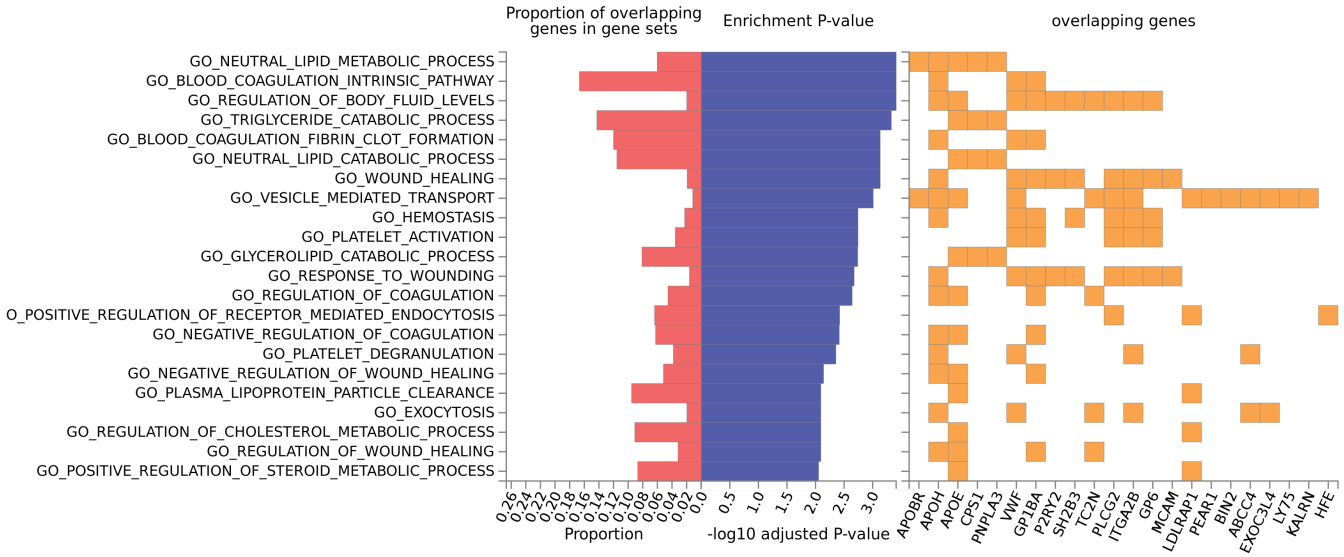


(**g**) rs342293 is a megakaryocyte enhancer variant associated with platelet traits (n = 116,663 individuals; BOLT-LMM p-values) that is associated with *PIK3CG* expression (by eQTL); *PIK3CG* knockout mice exhibited dysfunction platelet function.[8] (**h**) rs2038479 is a megakaryocyte enhancer variant associated with platelet traits (n = 116,666 individuals; BOLT-LMM p-values) that is associated with *DNM3* expression by eQTL and exhibited regulatory function in a reporter assay.[9] Inhibition of *DNM3* activity in mouse resulted in reduced platelet formation. In addition to rs2038479, our fine-mapping identified rs2208368, which was not marginally associated with platelet traits (n = 116,666 individuals; BOLT-LMM $p < 10^{-51}$) but was strongly associated after conditioning on rs2038479. rs2208368 was similarly localized to a megakaryocyte NDR in an intron of *DNM3*.

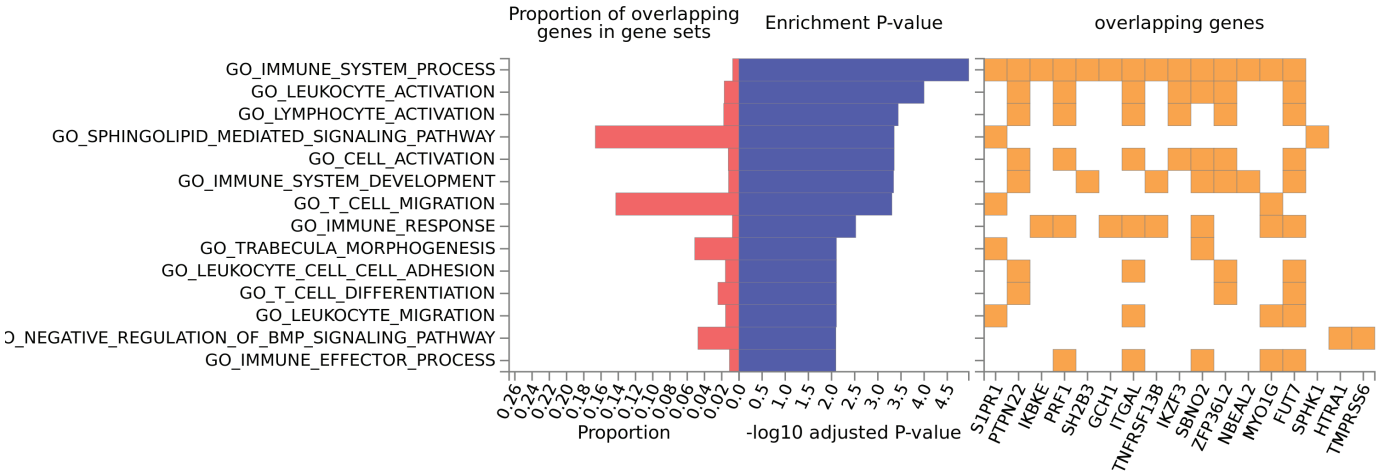
a Red blood cell traits



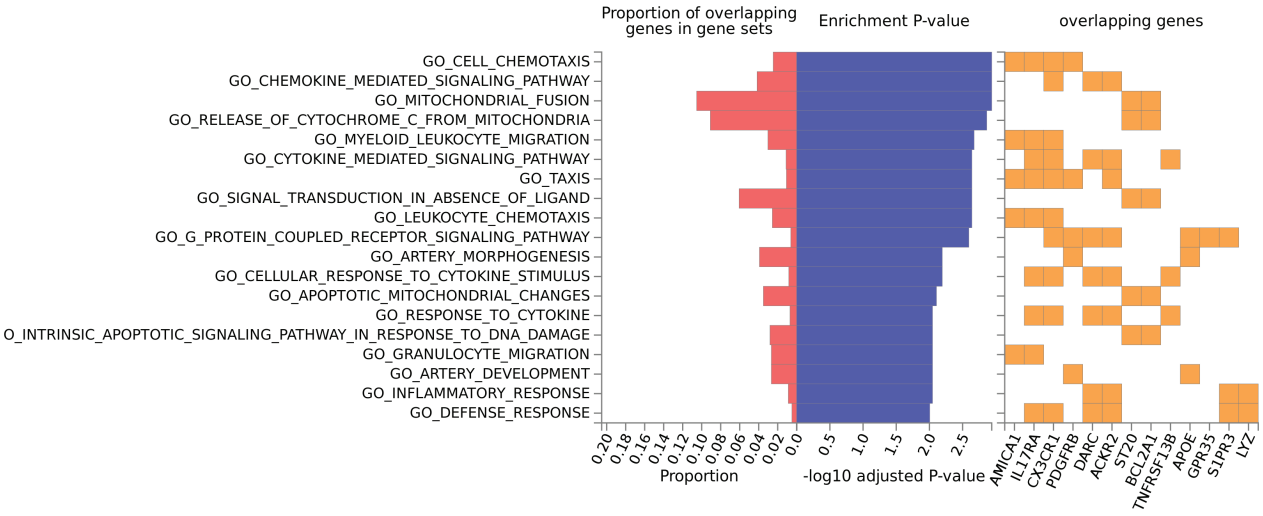
b Platelet Traits



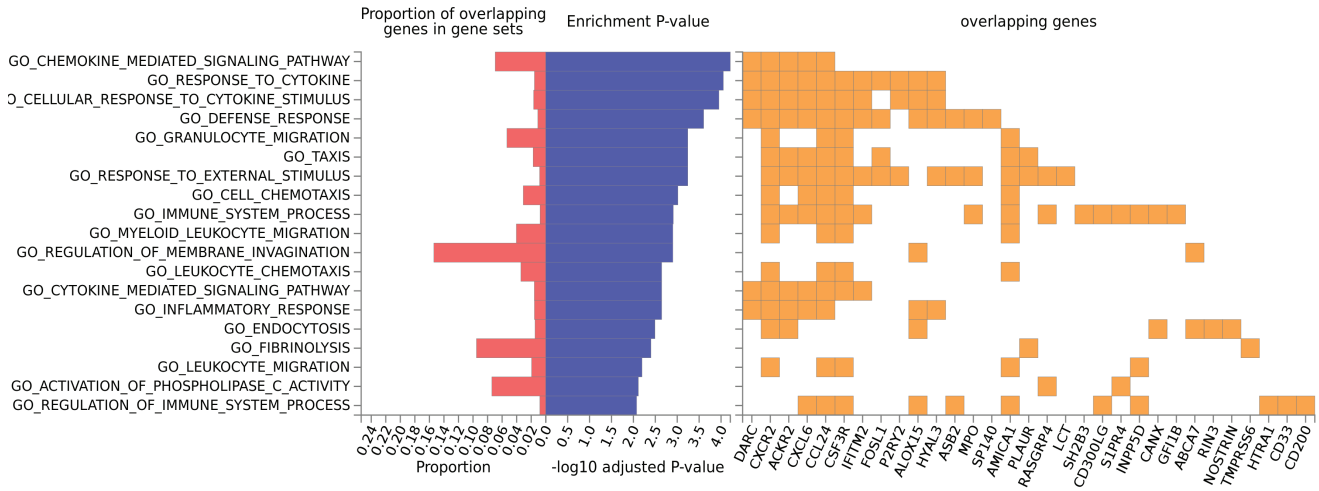
c Lymphoid Traits



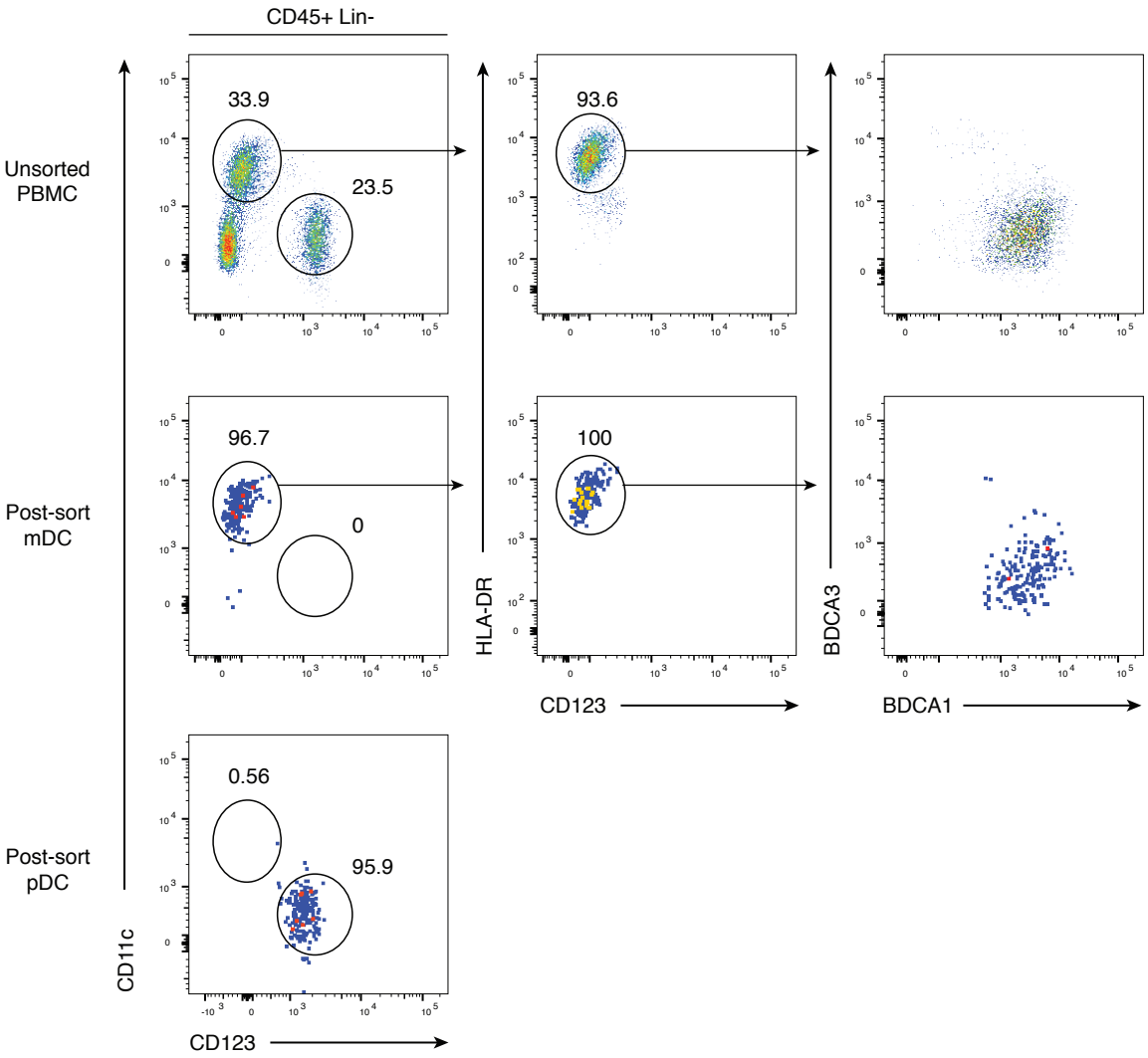
d Monocyte traits



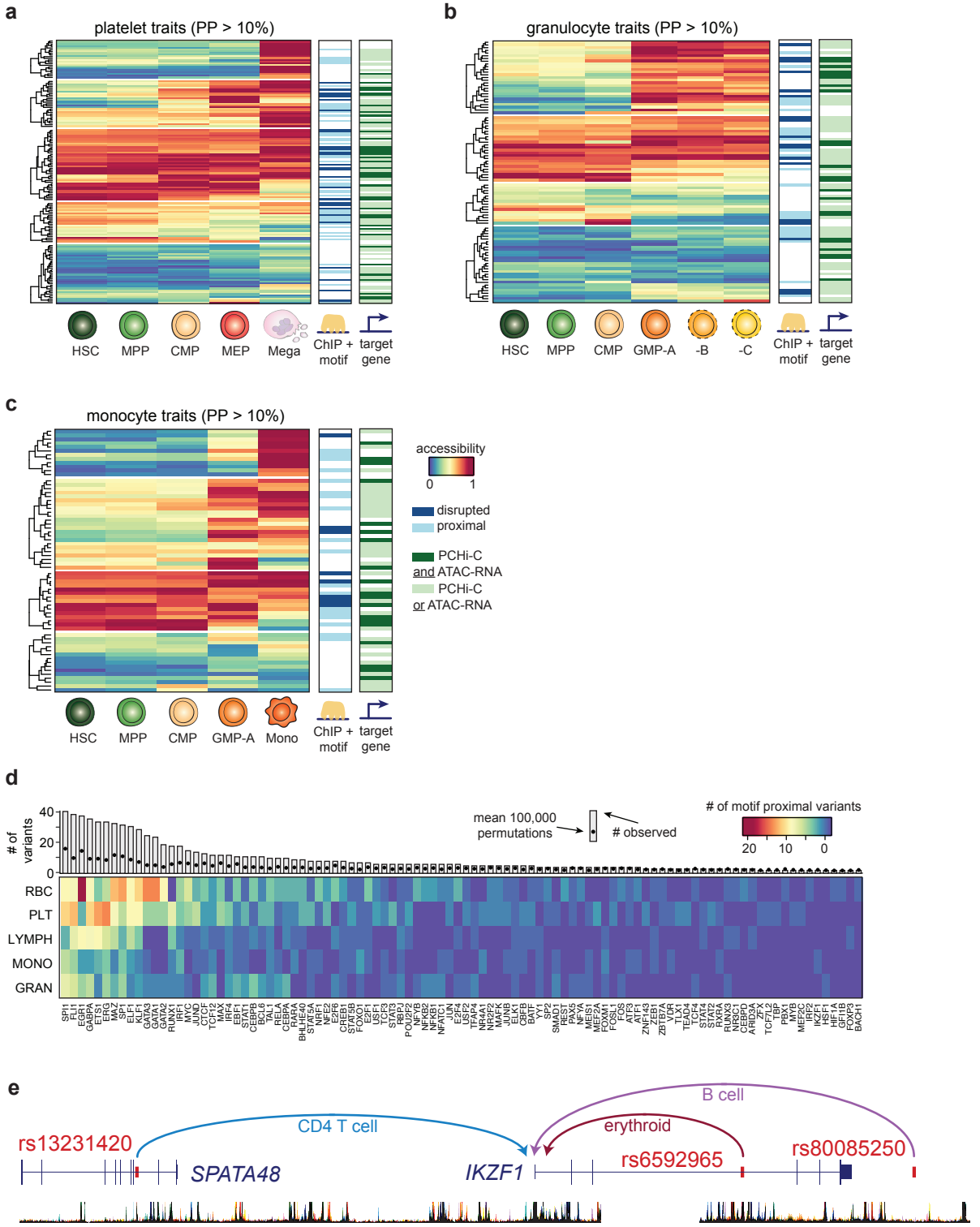
e Granulocyte traits



Supplementary Fig. 6. Gene set enrichments of fine-mapped coding variants with PP > 0.10, calculated using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA).[10] All protein-coding genes were used as background model. Fine-mapped coding variants with PP > 0.10 were divided into five separate lineages depending on their associated trait: (**a**) red blood cell traits (HCT, HGB, MCH, MCHC, MCV, mean reticulocyte volume, RBC count, reticulocyte count; n = 77 genes), (**b**) platelet traits (platelet count, mean platelet volume; n = 59 genes), (**c**) lymphoid traits (lymphocyte count; n = 28 genes), (**d**) monocyte traits (monocyte count; n = 20 genes), and (**e**) granulocyte traits (neutrophil count, basophil count, eosinophil count; n = 46 genes). The most highly enriched Gene Ontology (GO) biological processes are shown, requiring a minimum overlap of two genes and a hypergeometric test FDR < 0.01.

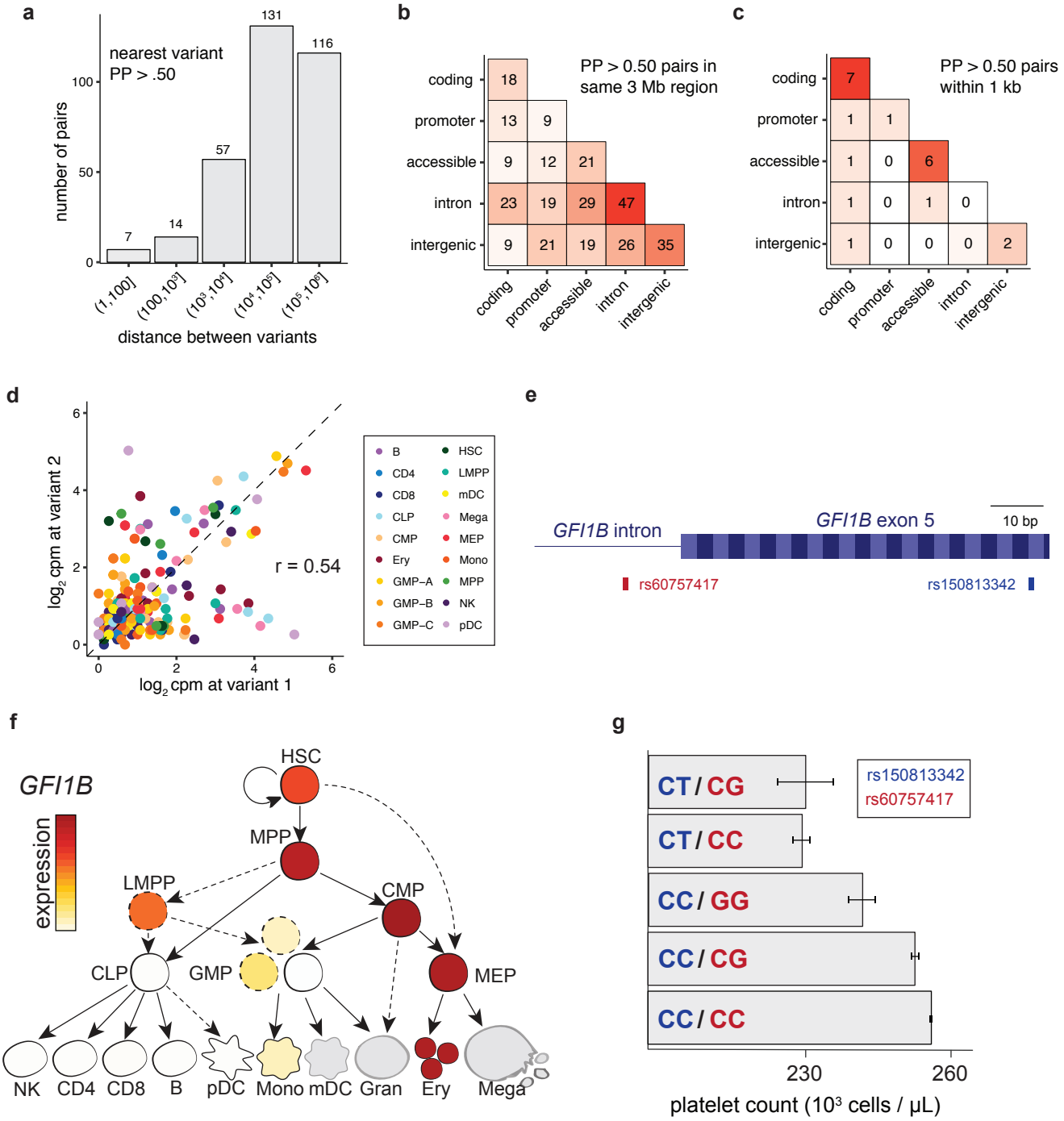


Supplementary Fig. 7. Fluorescence-activated cell sorting (FACS) strategy for mDCs and pDCs reported in this work. Cells derived from this gating strategy were profiled using FAST-ATAC.



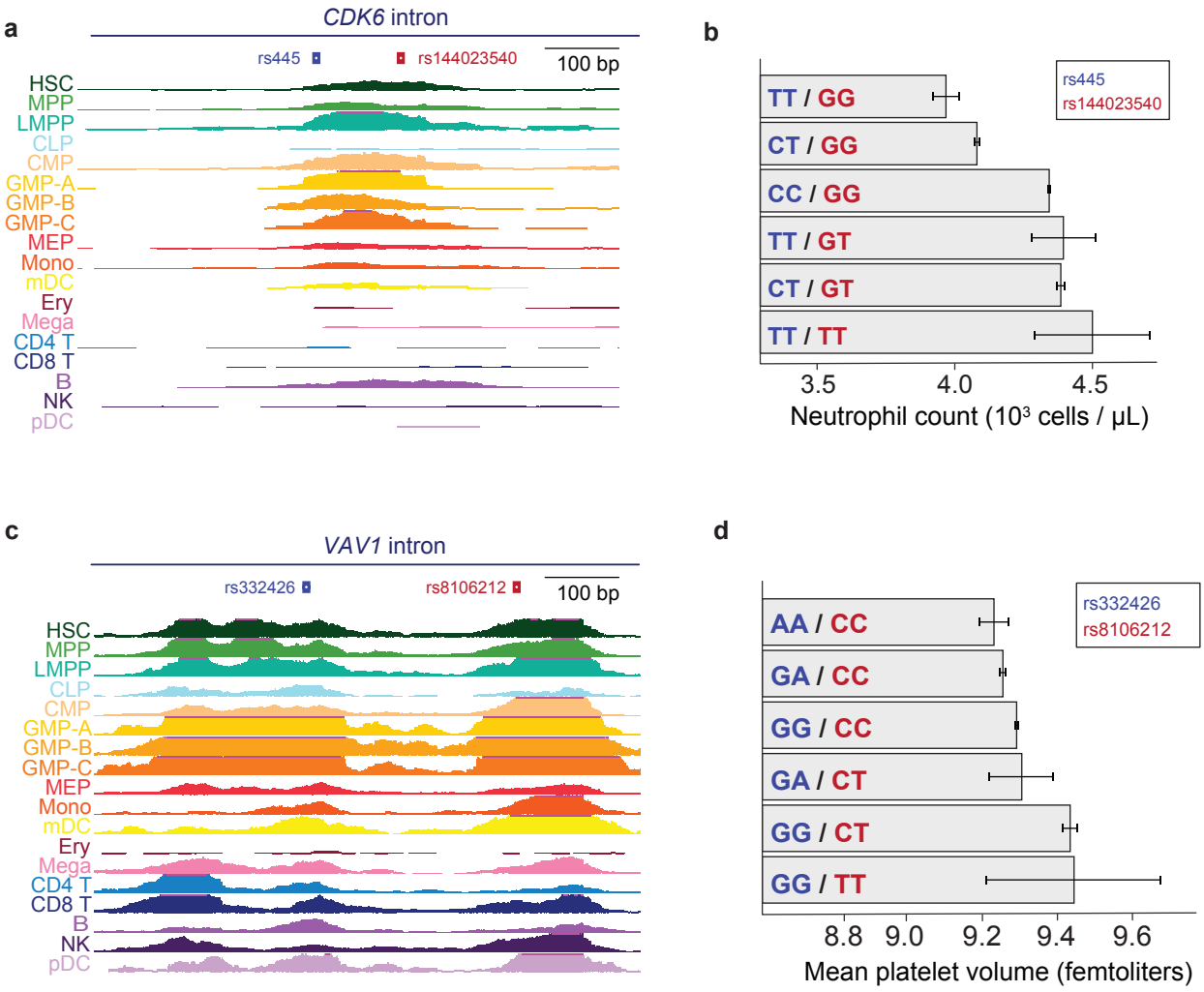
Supplementary Fig. 8. Mechanisms of core gene regulation in blood production across other lineages (similar to Fig. 2B). Heatmaps depicting chromatin accessibility for (**a**) platelet trait-associated variants (PP > 0.10), (**b**) granulocyte-associated variants (PP > 0.10), and (**c**) monocyte count-associated variants (PP > 0.10) across their respective lineages. Each row marks a fine-mapped variant, each column denotes a cell type within the relevant lineage, and the color denotes relative chromatin accessibility along the lineage at each variant (blue = least open chromatin, red = most open chromatin). Putative target genes (predicted by ATAC-RNA correlation and/or PCHi-C) and disrupted TFs (predicted by ChIP-Seq occupancy and motif disruption) are indicated to the right.

(**d**) Fine-mapped variants in or proximal to (+/- 20 bp) TF motifs across groups of hematopoietic traits. Each row represents a different set of traits where the TF motifs support the factor binding from ChIP-seq. The unique margin sums across each lineage are shown in the bar plot for each TF. The expected number of variants with ChIP + motif disruption across all PPs is estimated using 100,000 permutations and is shown as a single point. In total, fine-mapped variants were closer to 50 distinct TFs in AC than expected at an FDR < 10%. (**e**) Characterization of 3 distinct variants near the *IKZF1* locus with a PP > 0.75 (rs6592965, rs13231420, rs80085250) for traits spanning the hematopoietic lineage. An MCV-associated variant was within an erythroid-specific PCHi-C loop (red), a lymphocyte count associated variant was within a B and T cell-specific loop (blue), and a white blood cell count associated variant was specific to a B cell loop (purple).

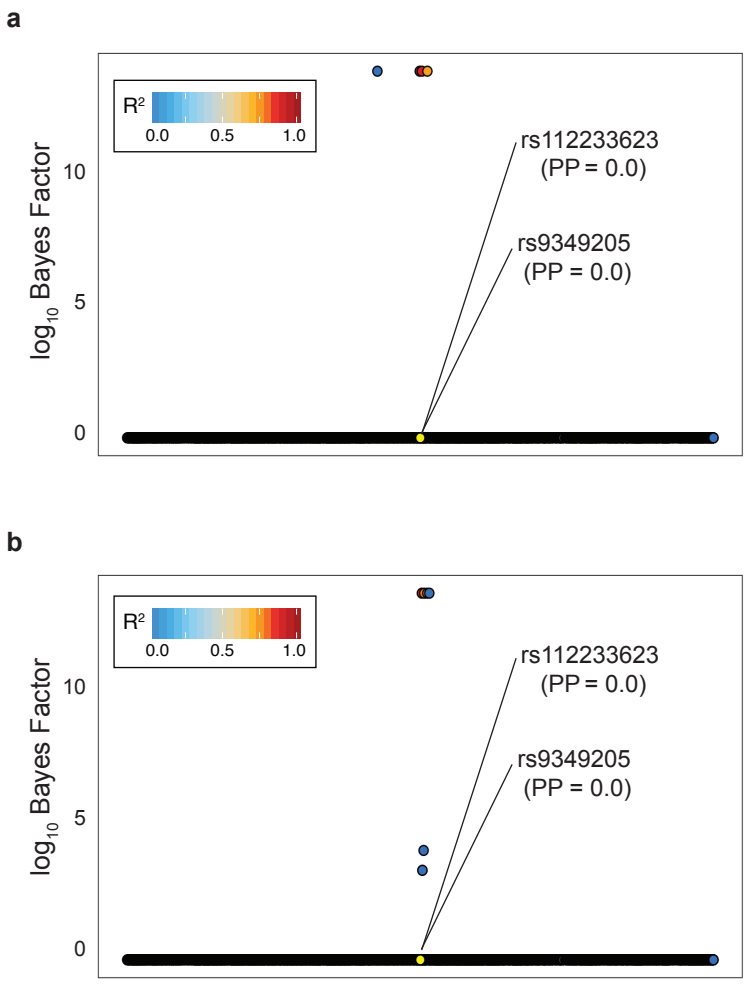


Supplementary Fig. 9. Additional information for fine-mapped variant pairs. (**a**) For each variant with PP > 0.50, the distance to the closest variant also with PP > 0.50 in the same trait was calculated, and this distribution is shown as a barplot. (**b-c**) Genomic annotations for variant pairs with PP > 0.50 either across each (**b**) 3 Mb region (from FINEMAP) or (**c**) restricted to variant pairs within 1 kb. (**d**) Chromatin accessibility for variant pairs within 1 Mb that are not in the same specific AC peak. AC for each hematopoietic population are correlated (Pearson r) across variant pairs (n = 12 pairs), although there are several clear outliers. (**e**) An example of variant pairs where one variant is located in an intron (rs60757417) and another variant within an exon (rs150813342) of *GFI1B*.

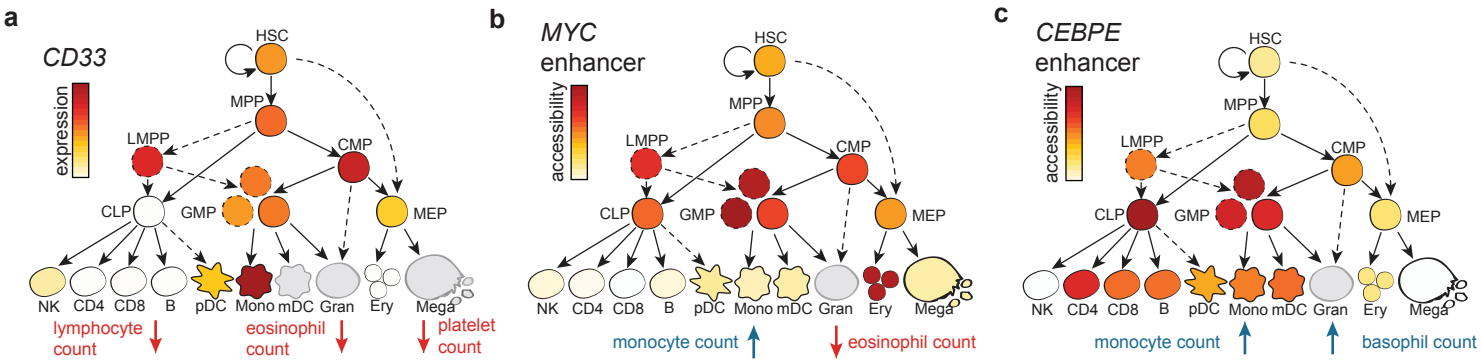
(**f**) RNA-seq relative log2 counts-per-million (min-max normalized) of *GFI1B*. Gray populations did not have expression data available. (**g**) Observed phenotypic effects of these two variants on neutrophil count across diplotypes (n = 116,482 individuals). Mean and standard error are indicated.



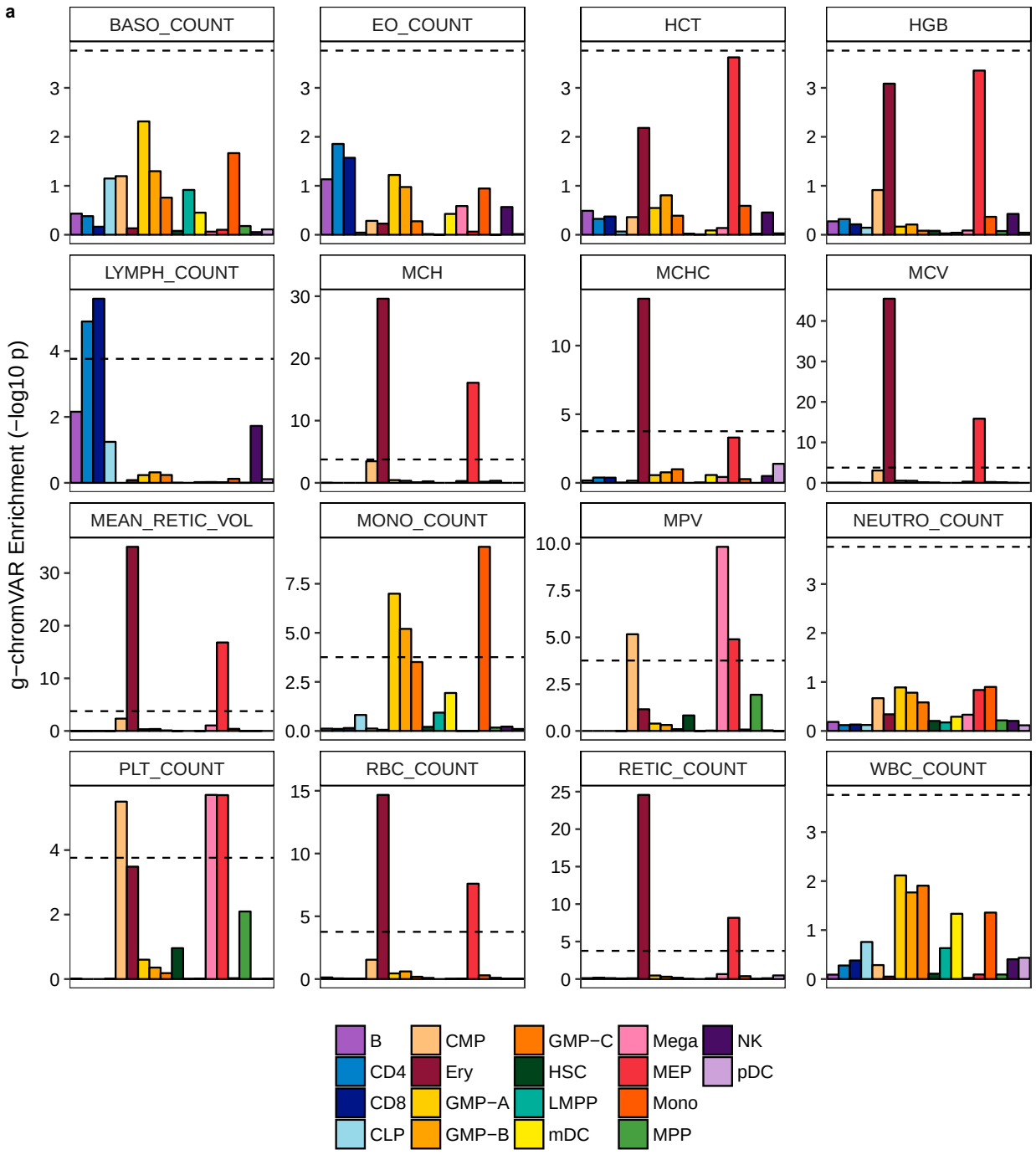
Supplementary Fig. 10. Additional fine-mapped regulatory regions with multiple causal variants. (**a**) Track plot of two variants (rs445, rs144023540) associated with neutrophil count located in an intronic AC region of *CDK6*. (**b**) Observed phenotypic effects of these two variants on neutrophil count across diplotypes (n = 116,482 individuals). Mean and standard error are indicated. (**c**) Track plot of two variants (rs332426, rs106212) associated with platelet traits, located in proximal, but distinct, AC regions within a *VAV1* intron. (**d**) Observed phenotypic effects of these two variants on mean platelet volume across diplotypes (n = 116,666 individuals). Mean and standard error are indicated.

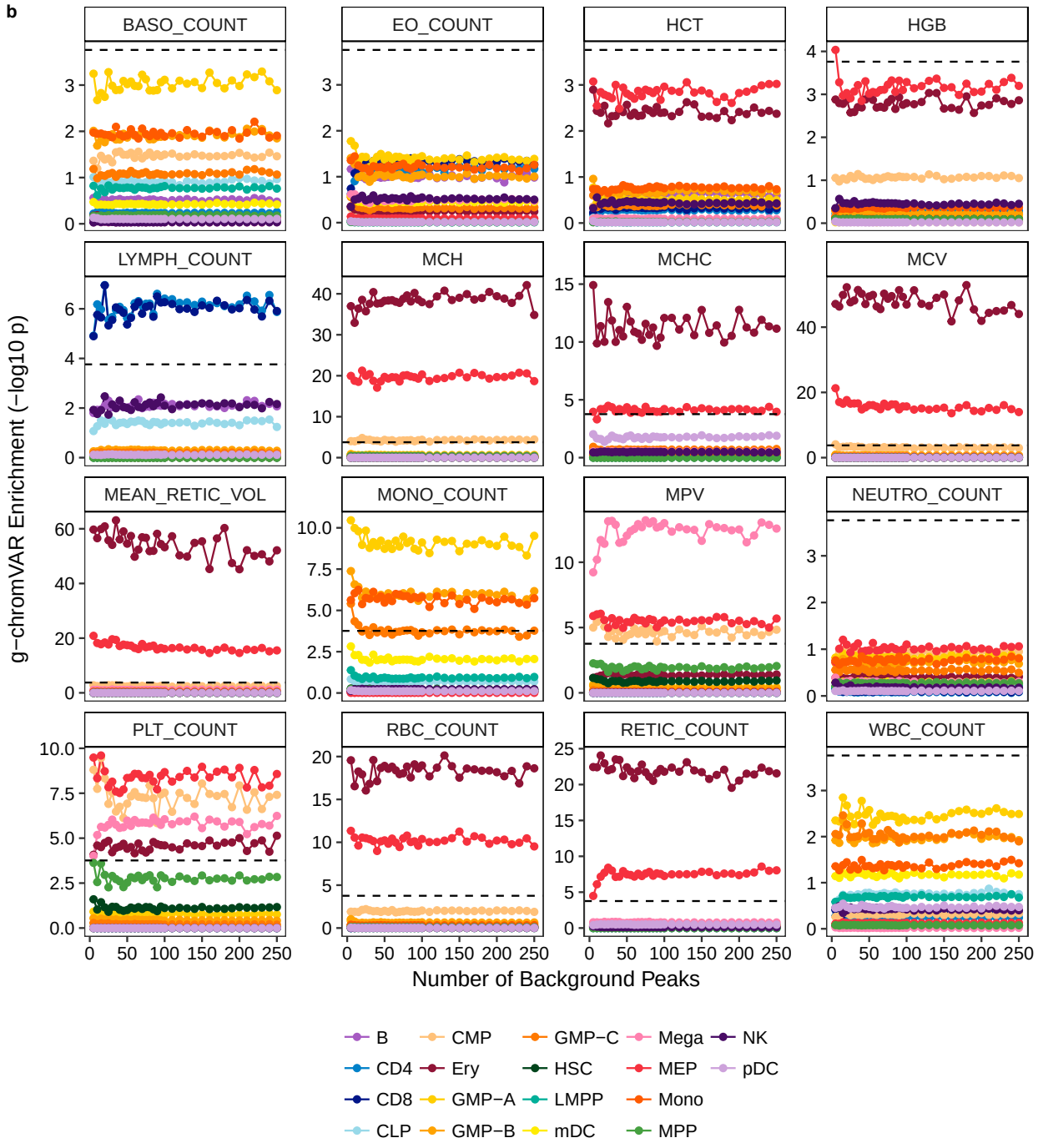


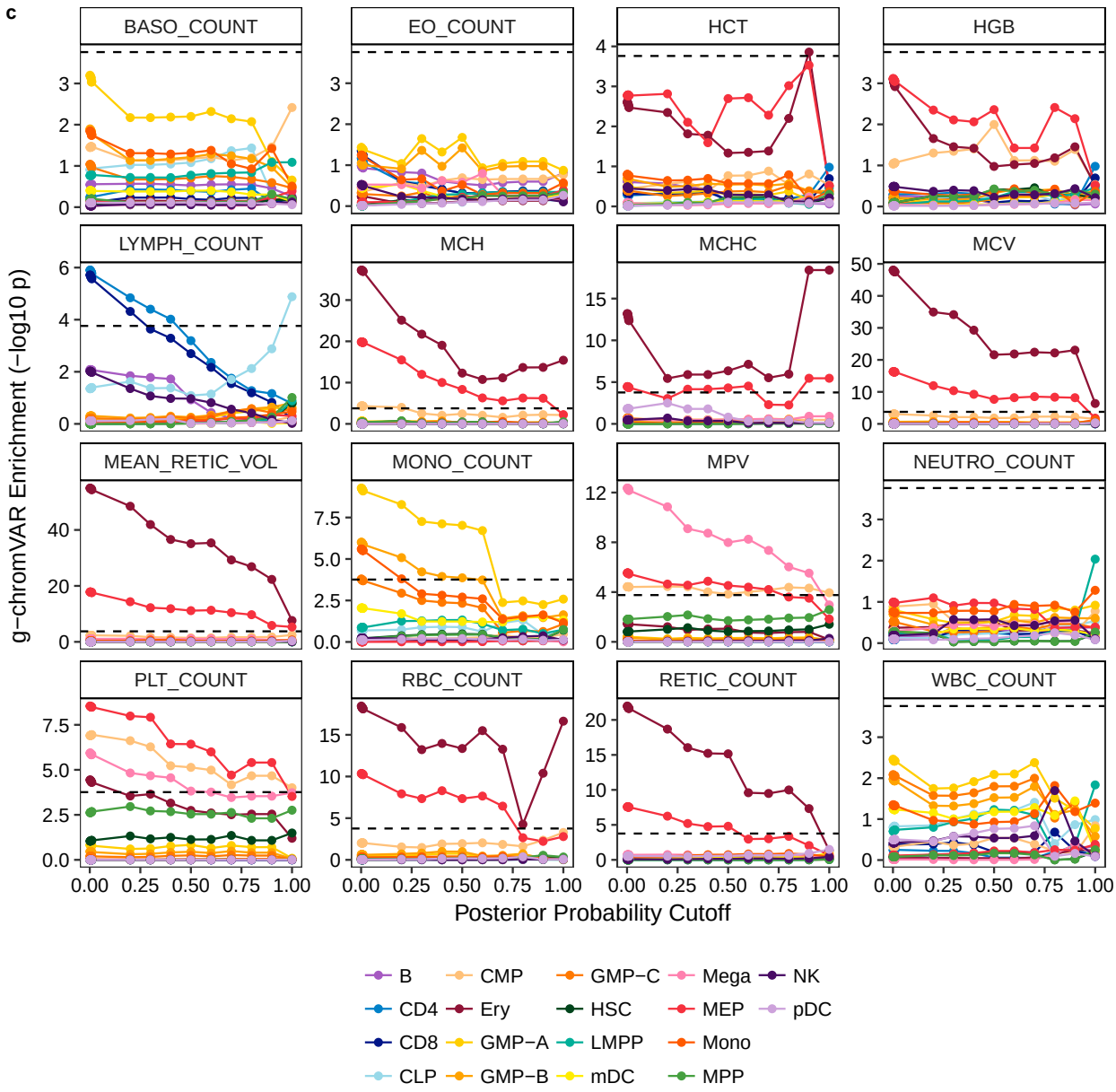
Supplementary Fig. 11. *CCND3* locus fine-mapping across alternative LD backgrounds. Fine-mapped $\log_{10}$(Bayes factor) values for *CCND3* variants when estimating LD with (**a**) hard-called genotypes rather than dosage imputed genotypes, and with (**b**) smaller UK10K reference panel (n = 3,677) instead of the larger UK biobank panel (n = 120,086). Variants discussed in **Fig. 2A** are highlighted.

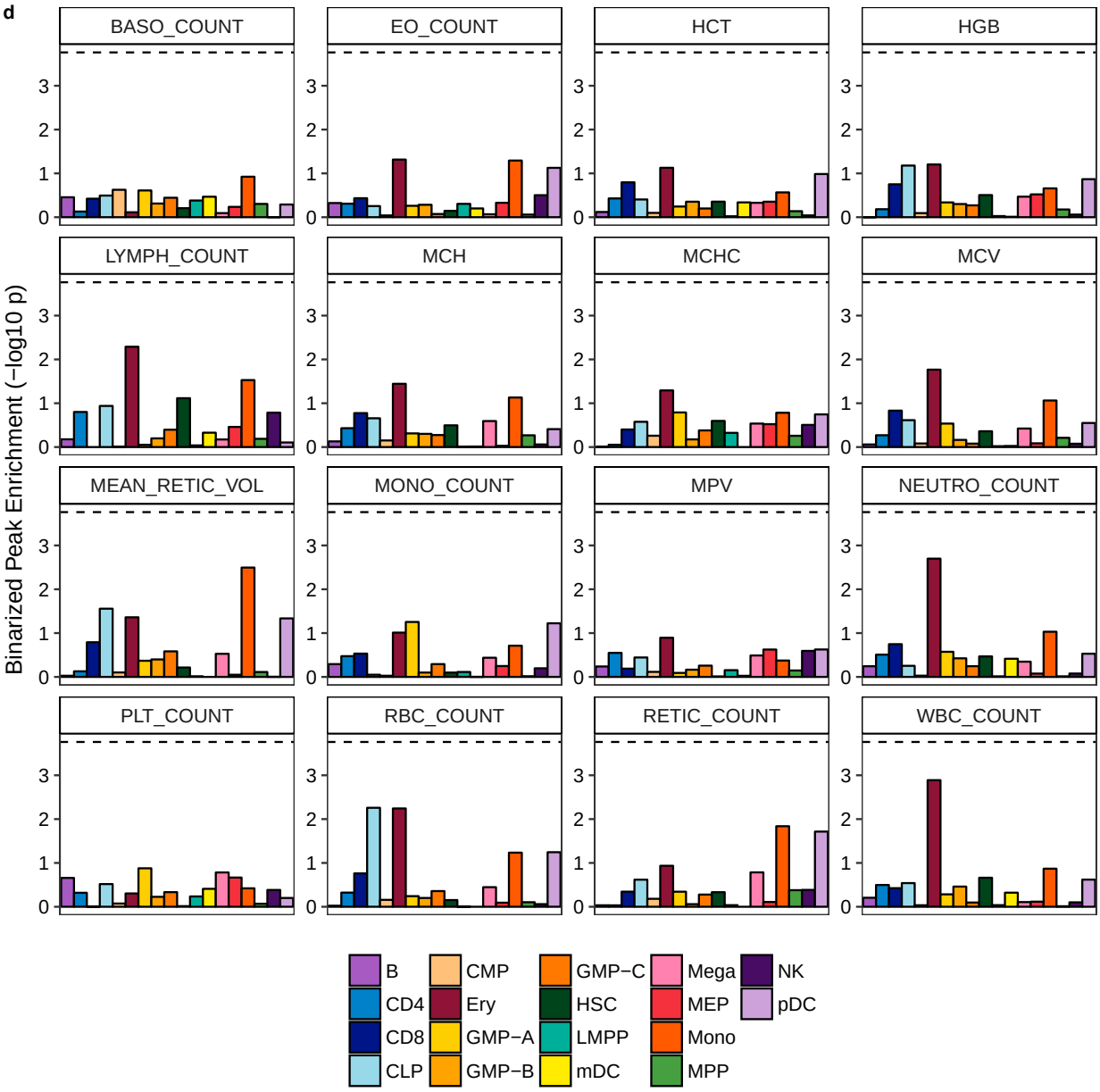


Supplementary Fig. 12. Additional examples of fine-mapped pleiotropic variants. (**a**) A coding variant in *CD33* (rs12459419) tunes eosinophil count (PP = 0.11), lymphocyte count (PP = 0.28), and platelet count (PP = 0.30). RNA-seq across hematopoietic lineages is shown in $\log_2$ counts-per-million (min-max normalized). Gray populations did not have RNA-seq data available. (**b**) A switch variant (rs562240450) located in a regulatory element of *MYC* that is associated with eosinophil count (PP = 0.91) and monocyte count (PP = 0.97). RNA-seq across hematopoietic lineages is shown in $\log_2$ counts-per-million (min-max normalized). (**c**) A variant (rs8017228) that tunes basophil count (PP = 0.85) and monocyte count (PP = 1.00) is located in a regulatory element near *CEBPE*.

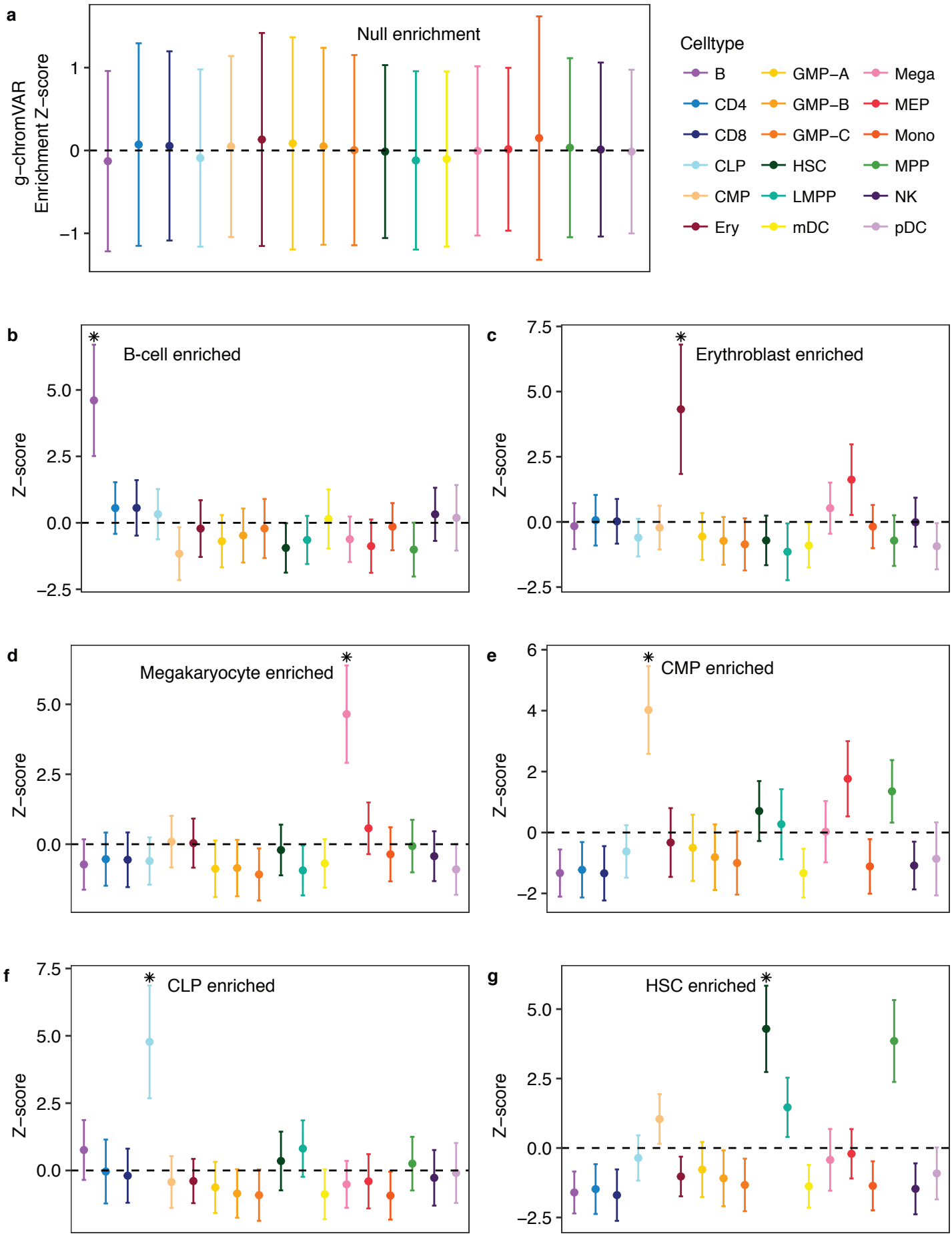

a
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
g−chromVAR Enrichment (−log10 p)
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC


b
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
g−chromVAR Enrichment (−log10 p)
Number of Background Peaks
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC

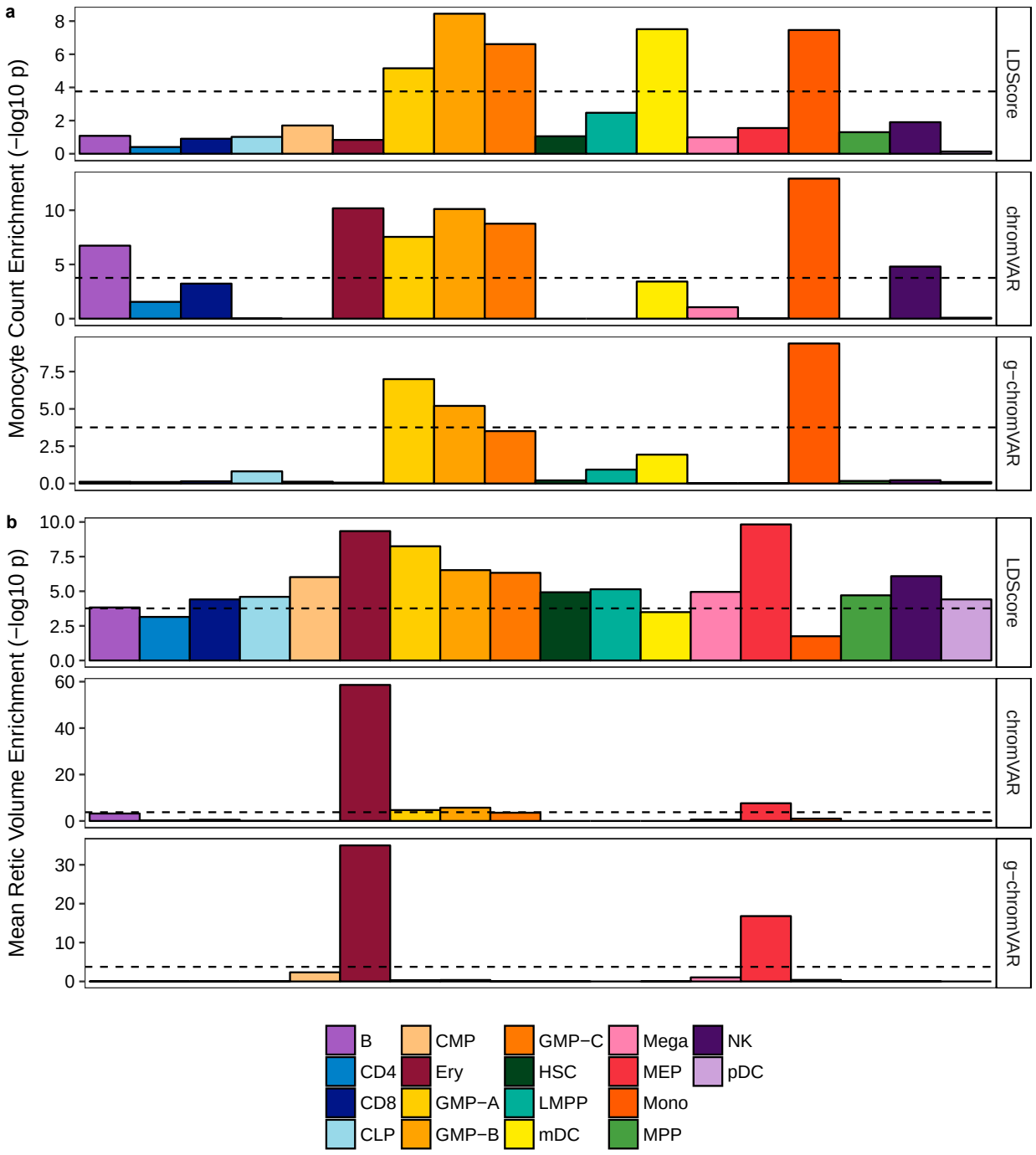

c
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
g−chromVAR Enrichment (−log10 p)
Posterior Probability Cutoff
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC

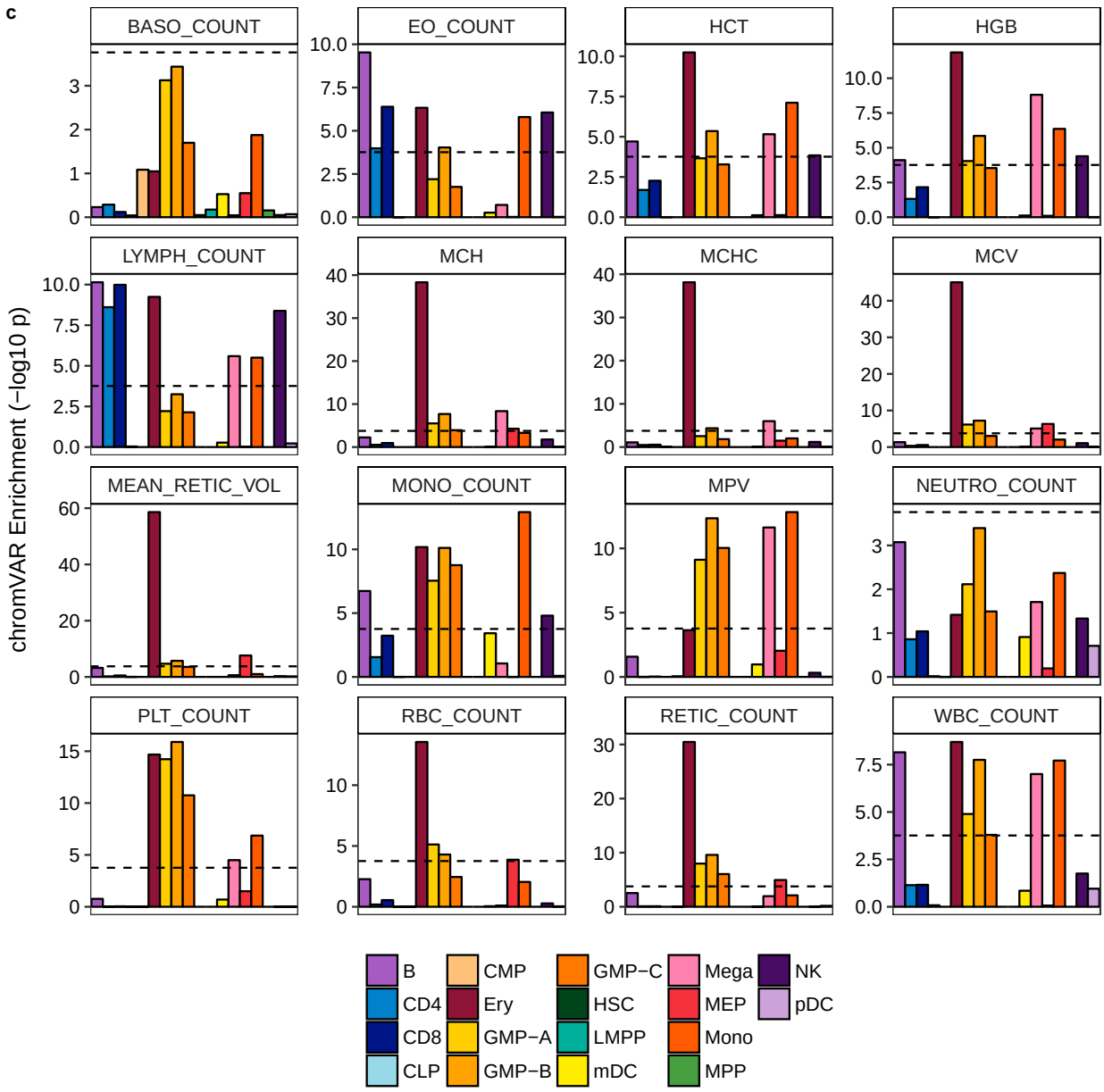


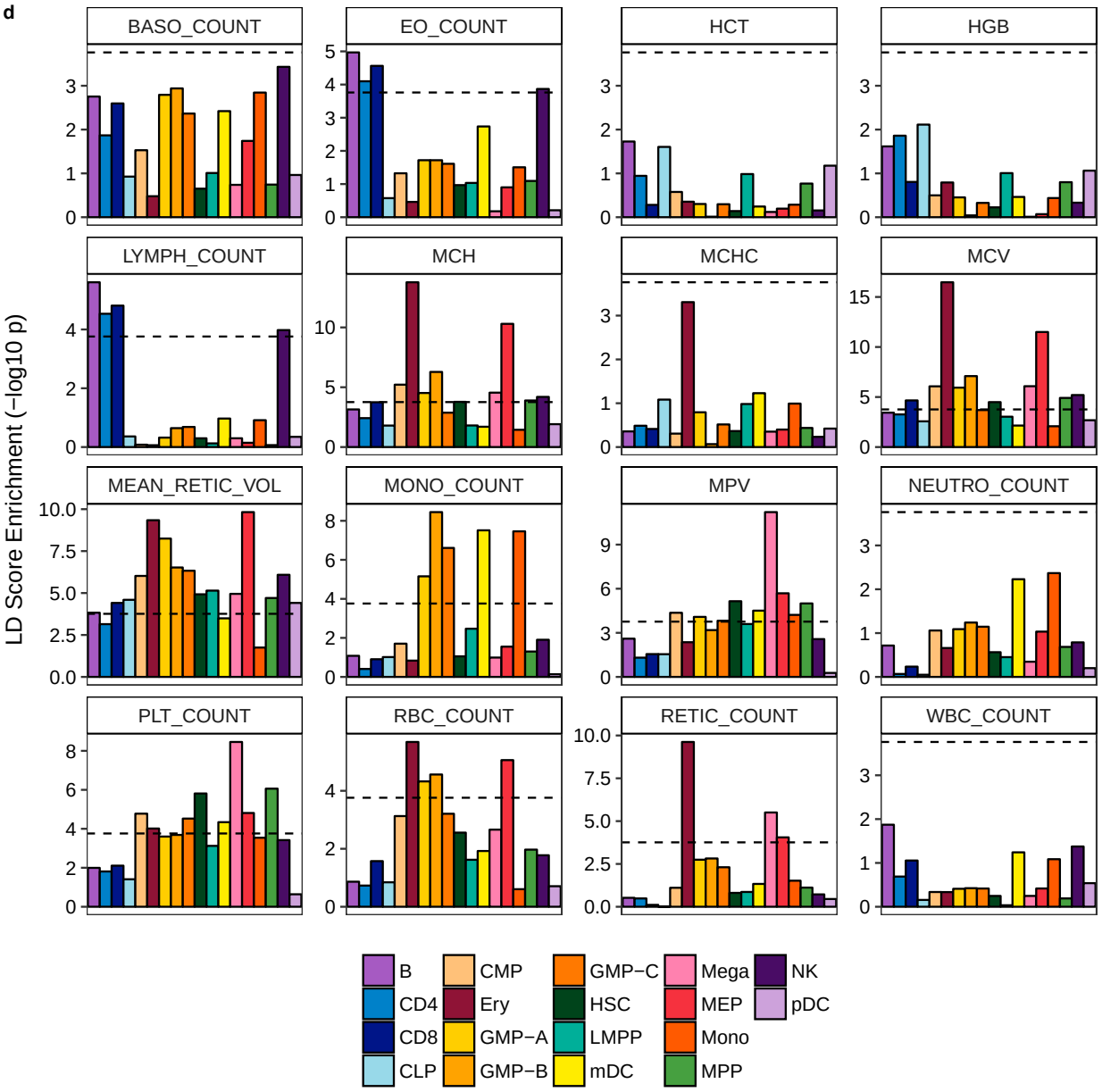
Supplementary Fig. 13. Application of g-chromVAR and parameterization characteristics. All trait / cell type pairs (n = 288 cells) scored by (**a**) g-chromVAR with the Bonferroni-adjusted significance level (one sided *z*-test) indicated by the dotted line. (**b**) Characterization of variable background peak numbers in g-chromVAR. By default, 50 background peaks are used per analysis peak. (**c**) Enrichments from g-chromVAR using varied PP cutoffs. Our analysis used PP > 0.001 for all computations. (**d**) Results from running g-chromVAR on binarized peak counts (0/1) shows a drastic decrease in power, and no pairs passed Bonferroni-adjusted significance.

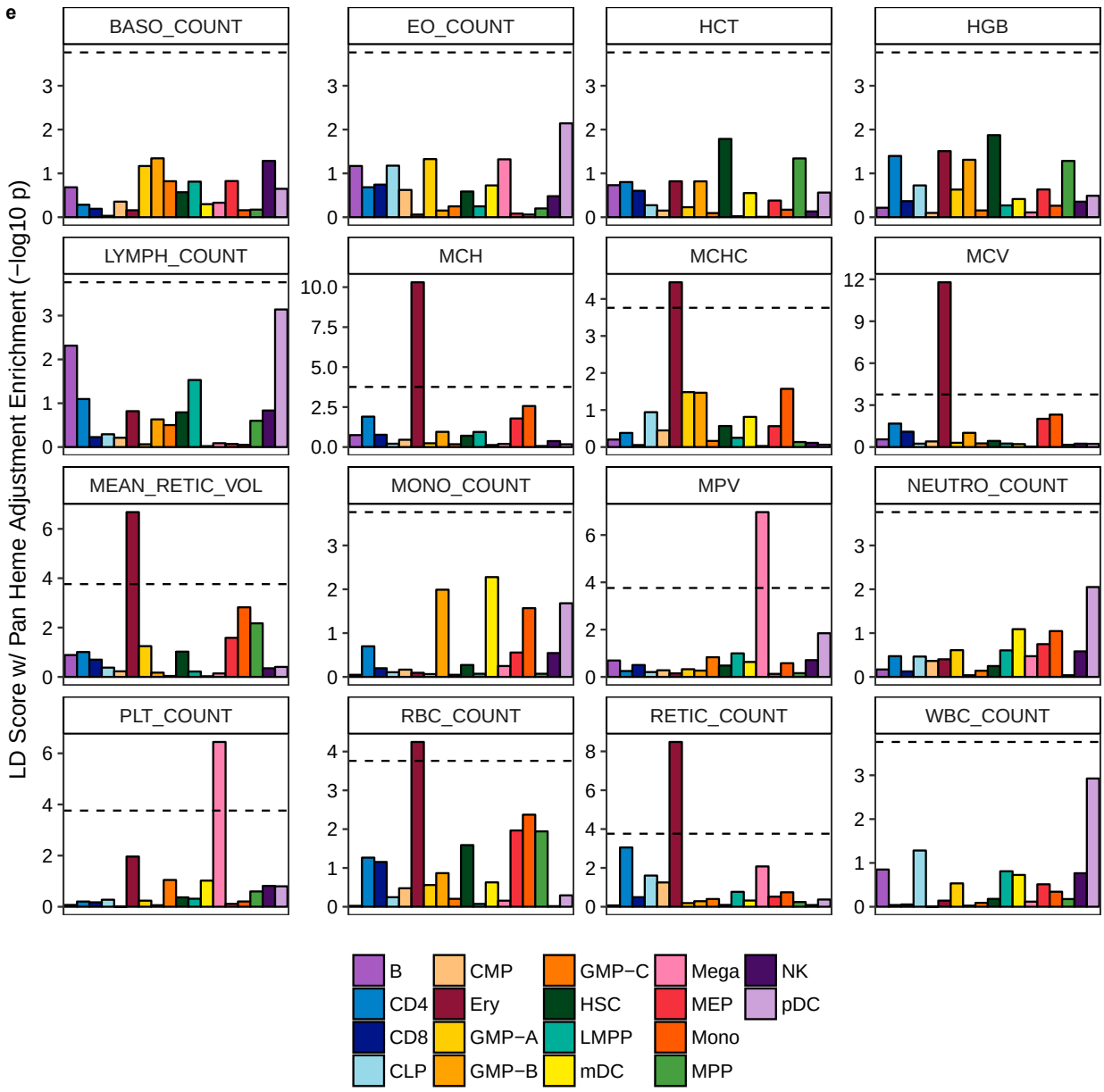

a
Null enrichment
g–chromVAR
Enrichment Z–score
1
0
−1
Celltype
B
CD4
CD8
CLP
CMP
Ery
GMP–A
GMP–B
GMP–C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC
b
B-cell enriched
Z–score
5.0
2.5
0.0
−2.5
c
Erythroblast enriched
7.5
d
Megakaryocyte enriched
e
CMP enriched
6
4
2
f
CLP enriched
g
HSC enriched

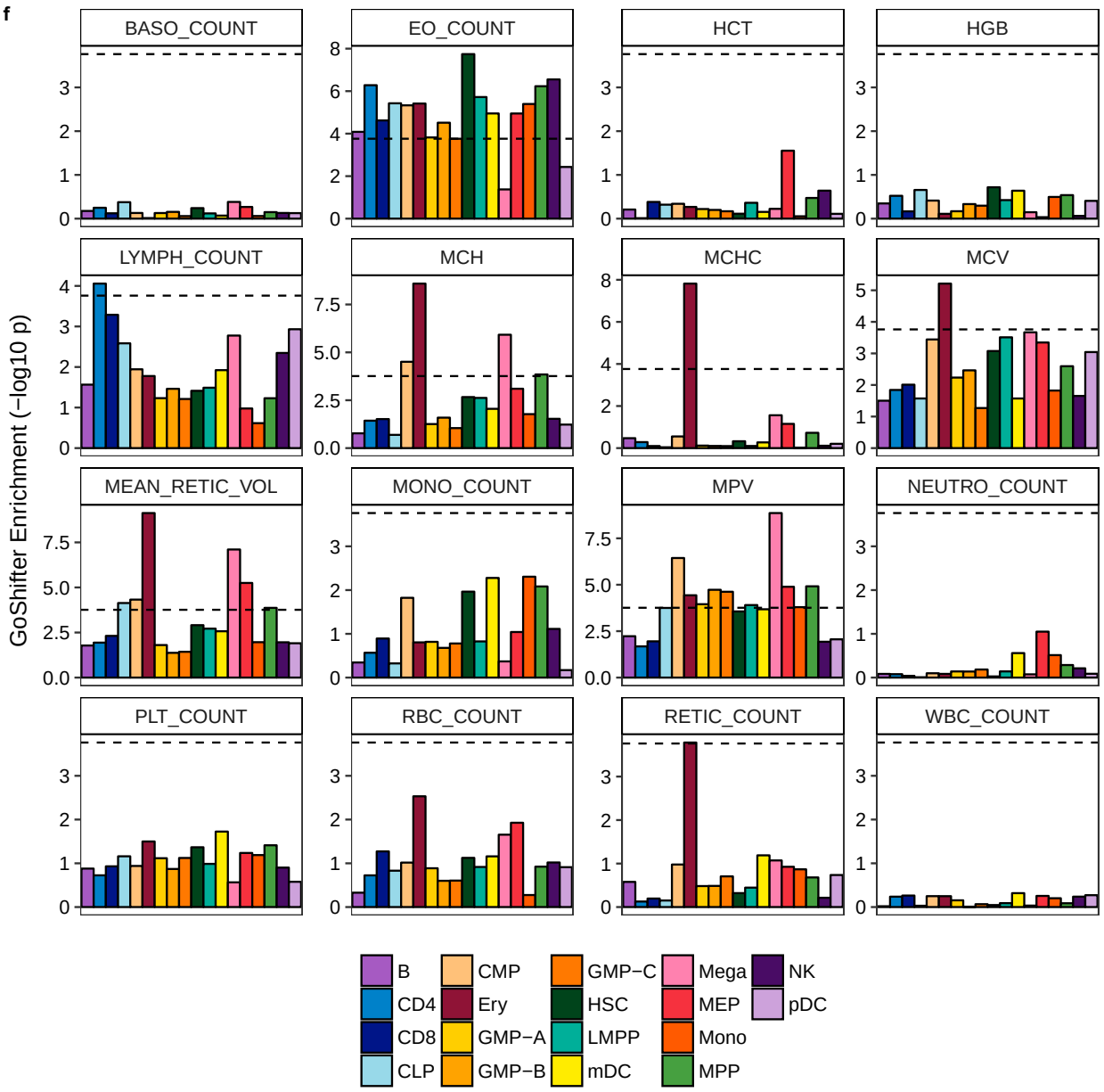
Supplementary Fig. 14. Simulation diagnostics for g-chromVAR. 100 simulations were performed for each panel. (**a**) Mean and standard deviation of enrichments (z-score) under a null phenotype simulation where no cell types are enriched for the simulated phenotype. (**b-g**) Mean and standard deviation of enrichments (z-score) when an arbitrary phenotype is simulated to as enriched for a designated cell type indicated by the asterisk (*). In general, true enrichments for terminally differentiated cell types (**b-d**) and late-stage progenitors (**e,f**) can be reliably detected using g-chromVAR. For early progenitors such as HSCs in (**g**), MPPs also appear enriched due to the striking similarity of the chromatin accessibility profiles of these two cell types.


a
Monocyte Count Enrichment (−log10 p)
LDScore
chromVAR
g−chromVAR
b
Mean Retic Volume Enrichment (−log10 p)
LDScore
chromVAR
g−chromVAR
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC

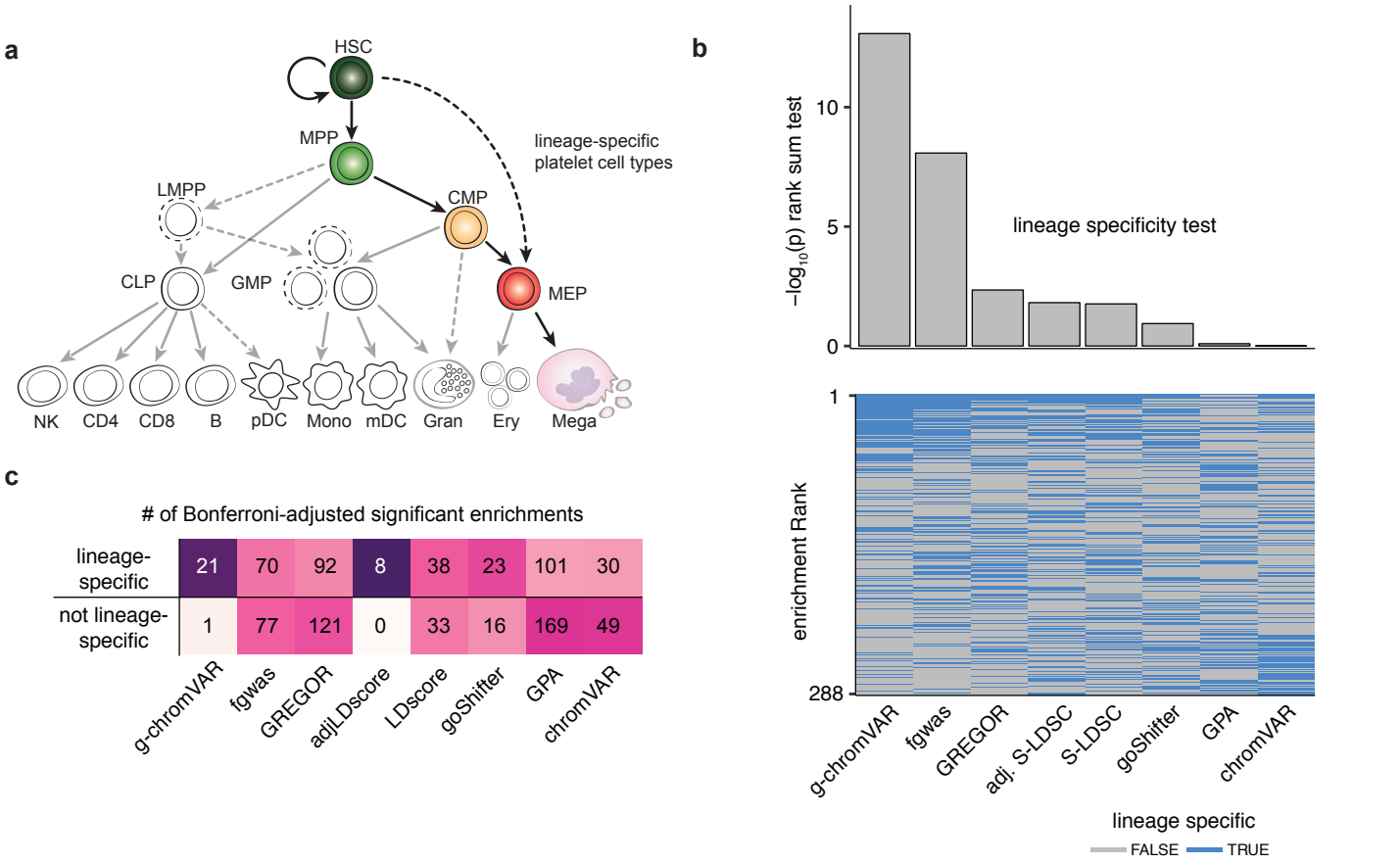

c
chromVAR Enrichment (−log10 p)
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC


d
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
LD Score Enrichment (−log10 p)
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC

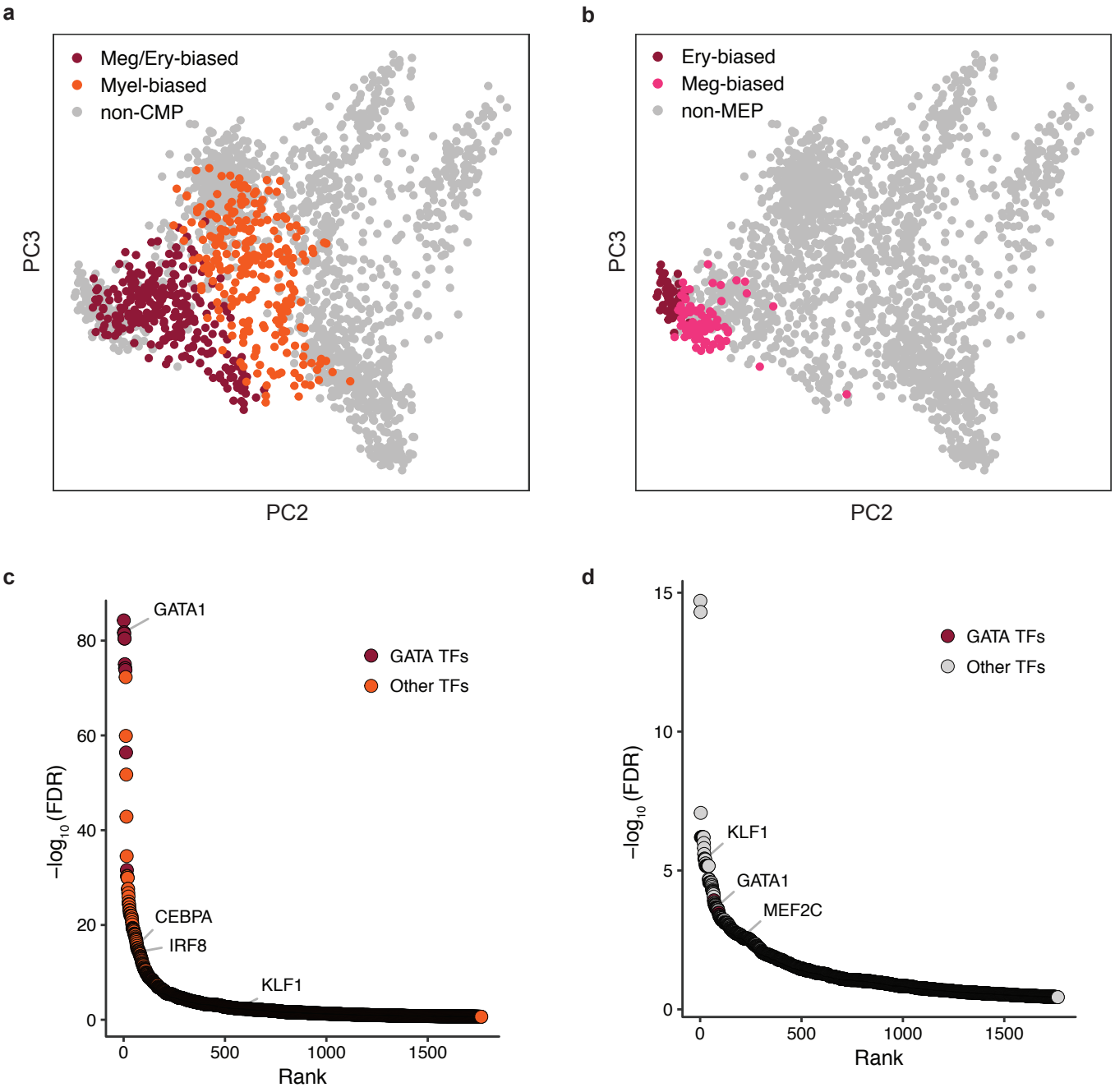

e
LD Score w/ Pan Heme Adjustment Enrichment (−log10 p)
BASO_COUNT
EO_COUNT
HCT
HGB
LYMPH_COUNT
MCH
MCHC
MCV
MEAN_RETIC_VOL
MONO_COUNT
MPV
NEUTRO_COUNT
PLT_COUNT
RBC_COUNT
RETIC_COUNT
WBC_COUNT
B
CD4
CD8
CLP
CMP
Ery
GMP−A
GMP−B
GMP−C
HSC
LMPP
mDC
Mega
MEP
Mono
MPP
NK
pDC



Supplementary Fig. 15. Comparison of g-chromVAR to S-LDSC, GoShifter, and chromVAR. (**a,b**) Selected traits (monocyte count and mean reticulocyte respectively) for all hematopoietic enrichments for g-chromVAR, chromVAR, and stratified LD score regression. Bonferroni-adjusted significance level indicated by the dotted line for all figure panels. Complete visual representation of all trait / cell type pairs (n = 288 pairs) scored by (**c**) chromVAR, (**d**) LD Score Regression, (**e**) LD Score Regression using an additional covariate for all hematopoietic peaks, and (**f**) GoShifter are shown in a consistent manner.



Supplementary Fig. 16. Depiction of the lineage specificity test and enrichment results across 8 methods. (**a**) Example of the biological lineage for platelet traits (mean platelet volume and platelet count). The five colored populations (HSC, MPP, CMP, MEP, megakaryocyte) are defined as “lineage specific”, whereas the other non-lineage specific populations are shown in gray. (**b**) (Top) Results from the lineage specificity rank-sum test for the 8 enrichment methods. (Bottom) Graphical representation of results from the lineage specificity test (two-sided Mann-Whitney U test). Each column shows the relative rank of all pairs of trait/cell type pairs examined (n = 288 pairs). Blue indicates a lineage-specific pair, whereas gray is a population that is not lineage-specific for that trait. (**c**) Bonferroni-adjusted counts of lineage and non-lineage enrichments across all trait/cell type pairs examined for the 8 enrichment methods. Darker color indicates a higher proportion of enrichments (scaled across each method).



Supplementary Fig. 17. Further diagnostics of scATAC clustering and TF variability. Two-dimensional representation (PC2 and PC3) of scATAC samples highlighting defined clusters of (**a**) CMPs and (**b**) MEPs. All variable TFs (n = 1,764 TFs, χ^2 test) between scATAC clusters are shown for (**c**) CMPs and (**d**) MEPs.

Supplementary Table Descriptions

Supplementary Table 1. Summary statistics and information for all fine-mapped variants with PP > 0.001. Summary statistics (Beta, SE) were calculated via BOLT-LMM v2.2[11] linear mixed model association analysis. Fine-mapped posterior probabilities (PP) were calculated via FINEMAP v1.1[12] (see Online Methods). The sample sizes for each trait analyzed were as follows: basophil count 116482; eosinophil count, 116482; hematocrit, 116667; hemoglobin, 116666; lymphocyte count, 116482; mean corpuscular hemoglobin, 116666; mean cellular hemoglobin concentration, 116666; mean cell volume, 116667; mean reticulocyte volume,114910; monocyte count, 116482; mean platelet volume, 116663; neutrophil count, 116482; platelet count, 116666; red blood cell count, 116667; reticulocyte count, 114910; white blood cell count, 116667.

Supplementary Table 2. Summary of top fine-mapped configurations in each region.

Supplementary Table 3. Summary of fine-mapped coding variants.

Supplementary Table 4. Summary statistics for bulk ATAC-seq libraries.

Supplementary Table 5. Summary of motif disrupting variants occupied by corresponding transcription factors.

Supplementary Table 6. Summary of putative gene targets for variants mapping to PCHi-C interactions. Variants with FINEMAP posterior probability (PP) > 0.10 were overlapped with PCHi-C interactions across 14 primary human hematopoietic cell types. PCHi-C interactions were assigned confidence scores via the CHiCAGO pipeline as described in Javierre et al, 2016[13], and only interactions with a CHiCAGO score > 5 were used for this analysis.

Supplementary Table 7. Summary of putative gene targets for variants mapping to ATAC-RNA correlations. Variants with FINEMAP posterior probability (PP) > 0.10 were included in this analysis. ATAC-RNA correlations were calculated by correlating ATAC peak counts to gene RNA counts across n = 16 hematopoietic cell types. Then, two-sided p-values were generated from the Pearson correlation. The qvalue column represents the estimated FDR.

Supplementary Table 8. Fine-mapped variants with PP > 0.5 identified in the same 3 Mb region.

Supplementary Table 9. Pleiotropic variants (PP > 0.1) for blood cell count traits.

Supplementary Table 10. g-chromVAR results for 39 predominately immune-related disorders previously fine-mapped with PICS to 18 chromatin accessibility profiles. A one-sided z-test was used to convert g-chromVAR z-scores to p-values.

Supplementary Table 11. Application of g-chromVAR to DNase 1 hypersensitivity data of 53 tissues from Roadmap Epigenomics. A one-sided z-test was used to convert g-chromVAR z-scores to p-values.

Supplementary Table 12. Top differentially enriched TFs between CMP and MEP sub-clusters. All variable TFs (n = 1,764) between scATAC clusters are listed for CMPs and MEPs. Two-tailed *t*-tests were used for each comparison.

Supplementary Note

Fine-mapping configuration probabilities

In addition to individual variants, fine-mapping also provided posterior probabilities (PPs) for causal configurations, which contain all predicted causal variants in a region. Interestingly, the top 1 and 25 configurations per region demonstrated > 0.75 PP for containing the correct causal configuration in only 3.9% and 28.9% of regions, respectively (**Supplementary Fig. 2; Supplementary Table 2**), likely reflecting the difficulty of predicting exact configurations in regions with multiple causal variants. Indeed, configurations for regions with a greater number of expected causal variants had lower PPs (**Supplementary Fig. 2**).

Additional examples of variants disrupting hematopoietic transcription factor binding

In total, we identified 145 distinct fine-mapped non-coding variants that were predicted to both disrupt a transcription factor (TF) motif and show occupancy by that specific TF in a hematopoietic tissue or cell line (**Fig. 2C**). These variants most commonly disrupted the binding sites of key transcriptional regulators of hematopoietic lineage commitment and differentiation. Several compelling variants include the platelet trait associated variant rs74340846 and the lymphocyte count associated variant rs79716587, which are occupied by and predicted to disrupt RUNX1 binding,[14,15] the RBC trait associated variants rs10758656 and rs66480687, which are occupied by and predicted to disrupt GATA1 binding,[7] the monocyte count associated variant, rs4970966, which is occupied by and predicted to disrupt IRF1 and IRF4 binding,[16] and the platelet trait associated variant, rs75522380, which is occupied by and predicted to disrupt MEF2A and MEF2C binding (**Fig. 2D**)[17]. In each case, the disrupted TF has been previously reported to be involved in the regulation of the corresponding blood cell lineage.

Analysis of fine-mapped variants disrupting proximal transcription factor motifs

In light of previous studies which found that many functional variants are proximal to but not predicted to strongly disrupt the canonical binding sites of relevant TFs[7,18], we extended our analysis and investigated whether the same 426 TF binding motifs were proximal (within 20 bp) to fine-mapped regulatory variants occupied by that TF in accessible chromatin (AC). Overall, we observed similar patterns, but with some key differences (**Supplementary Fig. 8**). For example, we only observed three fine-mapped, RBC trait associated variants predicted to disrupt the binding of GATA1, a key erythroid TF, but a total of 16 fine-mapped variants that were occupied by GATA1 and proximal to a canonical GATA1 binding site (**Supplementary Fig. 2C**, **Supplementary Fig. 8**). These findings build upon our previous functional work[7] suggesting that the majority of common genetic regulatory variation acts by fine-tuning, rather than abolishing, TF binding and activity. Overall, the molecular mechanisms identified here using fine-mapping, relevant AC overlap, and TF occupancy with matching motif disruption are of higher confidence than those based upon only AC overlap[19], motif disruption[18], or TF occupancy[20] alone.

Additional examples of regions with multiple causal variants

We conducted a closer examination of the 785 trait-associated regions with multiple independent causal signals. As an example of a pair of variants in which one is non-coding while the other is coding, we identified MPV-associated rs150813342 and rs60757417 (PP > 0.99; 75 bp apart) in *GFI1B*, which encodes for a transcriptional regulator important for erythroid and megakaryocytic development (**Supplementary Fig. 9E-G**). rs150813342 is a rare synonymous variant that we have previously shown to affect splicing of *GFI1B* isoforms[21], whereas rs60757417 is annotated as a non-coding variant in AC. As rs60757417 is 12 base pairs from the intron-exon junction of exon 5, we hypothesize that this variant may similarly be cryptically involved in regulating *GFI1B* mRNA splicing.

Besides the example of CCND3 in the main text, other variant pairs in the same AC region or within an AC cluster include eosinophil count-associated rs445 and rs8 (PP > 0.99; 39 bp apart), a pair within an AC region in the intron of *CDK6*, a gene that encodes for a cell cycle regulator involved in granulocyte production (**Supplementary Fig. 11A-B**),[22] and MPV-associated rs8106212 and rs332426 (PP > 0.61; 603 bp apart), a pair within an AC region in the intron of *VAV1*, a gene that encodes for a Rho guanine nucleotide exchange factor involved in platelet function (**Supplementary Fig. 11C-D**).[23] Overall, our statistical fine-mapping results have verified that multiple independent causal variants can not only occur in the same LD block but also within the *same regulatory element*. Together, these results indicate that other fine-mapping methodologies that assume one causal variant per locus likely miss true independent effects.

Local annotation shifting

To calculate genomic enrichments (similar to GoShifter), we calculated the overlap between the fine-mapped variant set of each trait (16 total) with each of the 5 genomic annotations. To define the null distribution of annotation overlap, we performed 10,000 locally shifting permutations; with every permutation, we shifted the genomic coordinates of the fine-mapped variant set by a random distance between -1.5 Mb and 1.5 Mb (this approach is equivalent to shifting the annotations). This was performed using the permTest function of the regioneR package[24]. The final odds ratio was calculated by dividing the number of overlaps between the original fine-mapped variant set and a genomic annotation by the mean number of overlaps between the 10,000 permuted sets and the same genomic annotation. To test if the association between a fine-mapped variant set and a genomic annotation (e.g. hematopoietic AC) was highly dependent on their exact positions, we used the localZScore function to calculate enrichment scores after various increments of shifting the fine-mapped variant set.

Extended methodological details of g-chromVAR

The bias-corrected enrichment statistics for *T* traits and a set of *S* samples (chromatin cell type profiles) with *P* peaks computed by g-chromVAR is a generalization of the chromVAR method.[25] Intuitively, our implementation of g-chromVAR relaxes the requirement in chromVAR that trait-peak annotations be binary, allowing for uncertainty in annotations such as transcription factor binding or in our case, localization of GWAS variants. Specifically, the chromVAR implementation requires a binarized matrix $\boldsymbol{M}$ (dimension *P* by *S*) where $m_{i,k}$ is 1 if annotation k is present in peak i and 0 otherwise. For example, in our examination of chromVAR (**Fig. 5**), $\boldsymbol{M}$ represents a binary matrix

where $m_{i,k} = 1$ if a genome-wide significant variant for trait k was present in peak i. However, our application of chromVAR to variant association data for our 16 hematopoietic traits revealed inflated summary statistics and poor lineage-specific enrichments without modeling the posterior confidence of variants (**Fig. 5B-D, Supplementary Fig. 16**). We note that if FINEMAP identified only 1 causal variant per region with a posterior probability of 1, g-chromVAR and chromVAR would yield identical results.

Instead, our methodology, g-chromVAR, uses a matrix of variant posterior probabilities $\boldsymbol{G}$, where $g_{i,k}$ is the sum of the posterior probabilities of the variants contained in the genomic coordinates of peak i for each trait k. Using the matrix of fragment counts in peaks $\boldsymbol{X}$, where $x_{i,j}$ represents the number of fragments from peak i in sample j, a matrix multiplication $\boldsymbol{X^T} \cdot \boldsymbol{G}$ yields the total number of fragments weighted by the fine-mapped variant posterior probabilities for *S* samples (rows) and *T* traits (columns). To compute a raw weighted accessibility deviation, we compute the expected number of fragments per peak per sample in $\boldsymbol{E}$, where $e_{i,j}$ is computed as the proportion of all fragments across all samples mapping to the specific peak multiplied by the total number of fragments in peaks for that sample:

$$e_{i,j} = \frac{\sum_j x_{i,j}}{\sum_j \sum_i x_{i,j}} \Sigma_i x_{i,j}$$

Analogously, $\boldsymbol{X^T} \cdot \boldsymbol{E}$ yields the expected number of fragments weighted by the fine-mapped variant posterior probabilities for *S* samples (rows) and *T* traits (columns). Using the $\boldsymbol{G}$, $\boldsymbol{X}$, and $\boldsymbol{E}$ matrices, we then compute the raw weighted accessibility deviation matrix $\boldsymbol{Y}$ for each sample j and trait k ($y_{j,k}$) as follows:

$$y_{j,k} = \frac{\sum_{i=1}^{P} x_{i,j} g_{i,k} - \sum_{i=1}^{P} e_{i,j} g_{i,k}}{\sum_{i=1}^{P} e_{i,j} g_{i,k}}$$

To correct for technical confounders present in assays (differential PCR amplification or variable Tn5 tagmentation conditions), g-chromVAR borrows the strategy implemented in chromVAR by generating a background set of peaks intrinsic to the set of epigenetic data examined. We note that other GWAS enrichment tools such as S-LDSC or GoShifter ignore biases prevalent in epigenomic assays that are explicitly corrected by g-chromVAR. In particular, variance in PCR or Tn5 tagmentation quality can lead to substantial differences in the number of observed fragment counts between cells based on an individual peak's GC content or average accessibility,[25] leading to errant GWAS-cell type enrichments. To correct for these technical confounders, each peak is assigned a background set of peaks that are matched in mean nucleotide GC content and average fragment accessibility between the sums of the cell types. An inverse Cholesky transformation is applied to a *P* by 2 matrix containing these variables to generate two uncorrelated dimensions describing the per-peak confounding. This two-dimensional space is divided into a pre-defined number of equally spaced bins where bin *i* is indicated β_i. Each peak q is assigned a bin from the shortest Euclidean distance between the bin's centroid and the individual peak in this transformed space. The

probability that a peak q' in bin j is selected as a background peak for peak q is proportional to the distance between bins *i* and *j* over the total number of peaks in bin *j* $|\beta_j|$:

$$P(q' \in \beta_j | \, q \in \beta_i) \propto \frac{d(\beta_i, \beta_j)}{|\beta_j|}$$

where the distance function d contains hyperparameters which, along with the total number of bins, have been previously discussed.[25]

By default, the framework samples a background set of 50 background elements per peak, which we've verified to be robust (**Supplementary Fig. 13B**). The matrix $\boldsymbol{B^{(b)}}$ encodes this background peak mapping where $b_{i,j}^{(b)}$ is 1 if peak i has peak j as its background peak in the b background set ($b \in \{1,2,\ldots,50\}$) and 0 otherwise. The matrices $\boldsymbol{B^{(b)}} \cdot \boldsymbol{X}$ and $\boldsymbol{B^{(b)}} \cdot \boldsymbol{E}$ thus give an intermediate for the observed and expected counts also of dimension *P* by *S*. For each background set b, sample j, and trait k, the elements $y_{j,k}^{(b)}$ of the background weighted accessibility deviations matrix $\boldsymbol{Y^{(b)}}$ are computed as follows:

$$y_{j,k}^{(b)} = \frac{\sum_{i=1}^{P} (\boldsymbol{B^{(b)}} \cdot \boldsymbol{X})_{i,k} g_{i,k} - \sum_{i=1}^{P} (\boldsymbol{B^{(b)}} \cdot \boldsymbol{E})_{i,k} g_{i,k}}{\sum_{i=1}^{P} (\boldsymbol{B^{(b)}} \cdot \boldsymbol{E})_{i,k} g_{i,k}}$$

After the background deviations are computed over the 50 sets, the bias-corrected matrix $\boldsymbol{Z}$ for sample j and trait k ($z_{j,k}$) can be computed as follows:

$$z_{j,k} = \frac{y_{j,k} - \text{mean}(y_{j,k}^{(b)})}{\text{sd}(y_{j,k}^{(b)})}$$

where the mean and variance of $y_{j,k}^{(b)}$ is taken over all values of b ($b \in \{1,2,\ldots,50\}$). Sample-trait p-values can then be computed from the one-tailed normal distribution of these z-scores using the pnorm function in R. Our implementation of g-chromVAR utilizes efficient matrix operations for each step and can compute pair-wise trait-cell type enrichments in ~1 minute on a standard laptop computer.

g-chromVAR simulations

We developed a new approach called genetic-chromVAR (**g-chromVAR**), a generalization of the recently described chromVAR method[25] to measure the enrichment of regulatory variants in each cell state using uncertainties in fine-mapped genetic variants and quantitative measurements of regulatory activity. Briefly, this method weights chromatin features by variant posterior probabilities and computes the enrichment for each cell type versus an empirical background matched for GC content and feature intensity (g-chromVAR is thus intuitively a *competitive* method across cell types based on top loci or "core gene" information).

To verify that enrichments computed by g-chromVAR were well-calibrated in our system, we devised a general simulation framework that computes enrichments for the

18 bulk hematopoietic cell types for an arbitrary simulated phenotype. Using the same matrix of fragment counts in peaks $\boldsymbol{X}$ as described in the *g-chromVAR* section of the Online Methods, where $x_{i,j}$ represents the number of fragments from peak i in sample j, we simulated a causal relationship between one of the accessibility samples j by performing a weighted draw of observed variant posterior probabilities $\boldsymbol{G}$, where $g_{i,k}$ is the sum of the posterior probabilities of the variants contained in the genomic coordinates of peak i for each trait k.

Specially, we first perform a counts-per-million (CPM) transformation of the fragment counts in peaks matrix to account for uneven sequencing depth between samples. Next, we z-transform the CPM-normalized matrix row wise to yield a matrix termed $\boldsymbol{X}^*$ where $x^*_{i,j}$ represents the amount of open chromatin observed in sample j in peak i relative to other samples. Intuitively, elements of the z-score matrix $\boldsymbol{X}^*$ yield larger positive numbers for cell type-specific peaks in specific samples (values $x^*_{i,j}$ range from -3.46 to 4.01). This matrix $\boldsymbol{X}^*$ serves as a basis for determining the cell type specificity of an individual regulatory element.

g-chromVAR simulation framework

To generate simulated elements of $\boldsymbol{G}$, we define a sorted vector v of length *T* * *P* (where 99.7% of values were zero) from the observed elements of $\boldsymbol{G}$ for our *T* = 16 hematopoietic traits and *P* = 451,283 regulatory peak elements. This vector v thus represents empirically derived values from the hematopoietic system studied that serve as input into g-chromVAR. Then, for a fixed causal cell-type j, we generate matrix $\boldsymbol{S}$ of $q \in (1, 2, \ldots, 100)$ simulated traits, where entries are defined as follows:

$$s_{i,q} \sim Unif\left[\, f\left(\Phi(x^*_{i,j})\right), 1 \,\right]$$

Here, f is a linear function that maps the normal cumulative distribution function (CDF) transformation of the $x^*_{i,j}$ z-score to a (0, 1) real number and is calibrated to yield phenotypic values similar to those observed empirically (matched mean column sum of $\boldsymbol{G}$). The $s_{i,q}$ value thus is a randomly generated (0, 1) real number skewed toward 1 when peak i contains cell type-specific chromatin for the fixed cell-type j. A final transformation of $\boldsymbol{S}$ maps these (0, 1) real number values to observed weights (elements of $\boldsymbol{G}$ or equivalently v) using the inverse CDF of v to index values. This final matrix, which serves as the input for g-chromVAR, is simulated to be enriched for cell-type j and null for all others. For the fully null simulation, elements of $s_{i,q}$ were populated from random draws of a $Unif[\, 0 \,, 1 \,]$.

Validation of g-chromVAR

We compared g-chromVAR to three state-of-the-art methods: S-LDSC,[26] which calculates the enrichment for genome-wide heritability using binary annotations after accounting for LD and overlapping annotations, GoShifter,[27] which calculates the enrichment of tight LD blocks containing sentinel GWAS single nucleotide variants for binary annotations, and chromVAR, which calculates enrichments similarly to g-chromVAR but only accepts binary annotations for variants, rather than continuous fine-mapped PPs. Using a Bonferroni correction, g-chromVAR identified 22 trait-tissue

enrichments, S-LDSC identified 71, GoShifter identified 39, and chromVAR identified 79 (**Fig. 5C, S15, S16**).

In order to compare the performance of these enrichment tools, we leveraged our knowledge of the hematopoietic system and devised a *lineage specificity test*. For any measured cell trait, we identified all possible upstream progenitor populations that could be passed through before terminal differentiation (**Fig. 1A**). For example, the differentiation of a platelet is thought to begin at the hematopoietic stem cell (HSC) and progress through multipotent progenitor (MPP), common myeloid progenitor (CMP), and megakaryocyte erythroid progenitor (MEP) before reaching the megakaryocyte stage (**Supplementary Fig. 16A**). The *lineage specificity test* is a nonparametric rank-sum test that measures the relative ranking of *lineage* specific trait-cell type pairs relative to the *non-lineage* specific traits for each of the compared methodologies. Using this metric for specificity, we found that g-chromVAR outperformed all three other methods (**Fig. 5D**). When we extended this comparison to additional cell type enrichment methods, g-chromVAR exhibited the highest lineage specificity among the 8 tested methods (**Supplementary Fig. 16**).

Finally, we assessed the generalizability of g-chromVAR by (**1**) using alternative fine-mapping methods and (**2**) applying it across larger epigenomic datasets. In each scenario, g-chromVAR identified known cell type enrichments as well as several novel and compelling associations, such as an enrichment for genetic variants associated with C-reactive protein level in myeloid dendritic cells[28,29] and variants associated with platelet traits in lung[30] (**Supplementary Table 10, Supplementary Table 11**)[31]

Application of g-chromVAR to single-cell pseudotime trajectories

We applied g-chromVAR to 2,034 single bone-marrow derived hematopoietic progenitor cells profiled using scATAC-seq.[32] In order to model the relatedness and heterogeneity of single cell measurements, we inferred pseudotime trajectories for the megakaryocyte and erythroid lineage (Meg/Ery), the myeloid and monocyte lineage (Myel), and the lymphoid lineage (Lymph) (**Fig. 6C**). We then used local regression to investigate the timing of blood trait GWAS enrichments during lineage commitment. Interestingly, we found that along these trajectories we could reconstruct our observations from bulk data with finer granularity. For example, we found that platelet count showed enrichment early along Meg/Ery differentiation with a sharp increase at firmly committed MEPs (**Fig. 6D**). This enrichment along the Meg/Ery commitment path coincided with negative enrichments along the alternative Myel and Lymph paths (**Fig. 6D**). This suggests that although the majority of variants act in committed progenitors, a subset of regulatory variants act in multipotential or heterogeneous progenitor populations, consistent with our earlier finding that many distinctly fine-mapped variants only overlapped with multipotential progenitor populations (**Fig. 2A-B**).

Supplementary Note References

1. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* **17**, 122 (2016).
2. Guo, M.H. *et al.* Comprehensive population-based genome sequencing provides insight into hematopoietic regulatory mechanisms. *Proc Natl Acad Sci U S A* **114**, E327-E336 (2017).
3. Zou, S. *et al.* SNP in human ARHGEF3 promoter is associated with DNase hypersensitivity, transcript level and platelet function, and Arhgef3 KO mice have increased mean platelet volume. *PLoS One* **12**, e0178095 (2017).
4. Giani, F.C. *et al.* Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. *Cell Stem Cell* **18**, 73-78 (2016).
5. Thom, C.S. *et al.* Trim58 degrades Dynein and regulates terminal erythropoiesis. *Dev Cell* **30**, 688-700 (2014).
6. Cvejic, A. *et al.* SMIM1 underlies the Vel blood group and influences red blood cell traits. *Nat Genet* **45**, 542-545 (2013).
7. Ulirsch, J.C. *et al.* Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell* **165**, 1530-1545 (2016).
8. Paul, D.S. *et al.* Maps of open chromatin guide the functional follow-up of genome-wide association signals: application to hematological traits. *PLoS Genet* **7**, e1002139 (2011).
9. Nurnberg, S.T. *et al.* A GWAS sequence variant for platelet volume marks an alternative DNM3 promoter in megakaryocytes near a MEIS1 binding site. *Blood* **120**, 4859-68 (2012).
10. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
11. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
12. Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-501 (2016).
13. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384 e19 (2016).
14. Elagib, K.E. *et al.* RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood* **101**, 4333-41 (2003).
15. Blyth, K. *et al.* Runx1 promotes B-cell survival and lymphoma development. *Blood Cells Mol Dis* **43**, 12-9 (2009).
16. Chistiakov, D.A., Myasoedova, V.A., Revin, V.V., Orekhov, A.N. & Bobryshev, Y.V. The impact of interferon-regulatory factors to macrophage differentiation and polarization into M1 and M2. *Immunobiology* **223**, 101-111 (2018).
17. Gekas, C. *et al.* Mef2C is a lineage-restricted target of Scl/Tal1 and regulates megakaryopoiesis and B-cell homeostasis. *Blood* **113**, 3461-71 (2009).
18. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).

19. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
20. Oki, S. *et al.* Integrative analysis of transcription factor occupancy at enhancers and disease risk loci in noncoding genomic regions. *bioRxiv* (2018).
21. Polfus, L.M. *et al.* Whole-Exome Sequencing Identifies Loci Associated with Blood Cell Traits and Reveals a Role for Alternative GFI1B Splice Variants in Human Hematopoiesis. *Am J Hum Genet* **99**, 785 (2016).
22. Fujimoto, T., Anderson, K., Jacobsen, S.E., Nishikawa, S.I. & Nerlov, C. Cdk6 blocks myeloid differentiation by interfering with Runx1 DNA binding and Runx1-C/EBPalpha interaction. *EMBO J* **26**, 2361-70 (2007).
23. Pearce, A.C. *et al.* Vav1, but not Vav2, contributes to platelet aggregation by CRP and thrombin, but neither is required for regulation of phospholipase C. *Blood* **100**, 3561-9 (2002).
24. Gel, B. *et al.* regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289-91 (2016).
25. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).
26. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
27. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet* **97**, 139-52 (2015).
28. Frenzel, H. *et al.* Decreased migration of myeloid dendritic cells through increased levels of C-reactive protein. *Anticancer Res* **27**, 4111-5 (2007).
29. He, W., Ren, Y., Wang, X., Chen, Q. & Ding, S. C reactive protein and enzymatically modified LDL cooperatively promote dendritic cell-mediated T cell activation. *Cardiovasc Pathol* **29**, 1-6 (2017).
30. Lefrancais, E. *et al.* The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. *Nature* **544**, 105-109 (2017).
31. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
32. Buenrostro, J.D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548.e16 (2018).