RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens

John A. Morris, Christina Caragine, Zharko Daniloski, Júlia Domingo, Timothy Barry, Lu Lu, Kyrie Davis, Marcello Ziosi, Dafni A. Glinos, Stephanie Hao, Eleni P. Mimitou, Peter Smibert, Kathryn Roeder, Eugene Katsevich, Tuuli Lappalainen*, Neville E. Sanjana*

INTRODUCTION: Genome-wide association studies (GWASs) have identified thousands of human genetic variants associated with diverse diseases and traits, and most of these variants map to noncoding loci with unknown target genes and function. Current approaches to understand which GWAS loci harbor causal variants and to map these noncoding regulators to target genes suffer from low throughput. With newer multi-ancestry GWASs from individuals of diverse ancestries, there is a pressing and growing need to scale experimental assays to connect GWAS variants with molecular mechanisms.

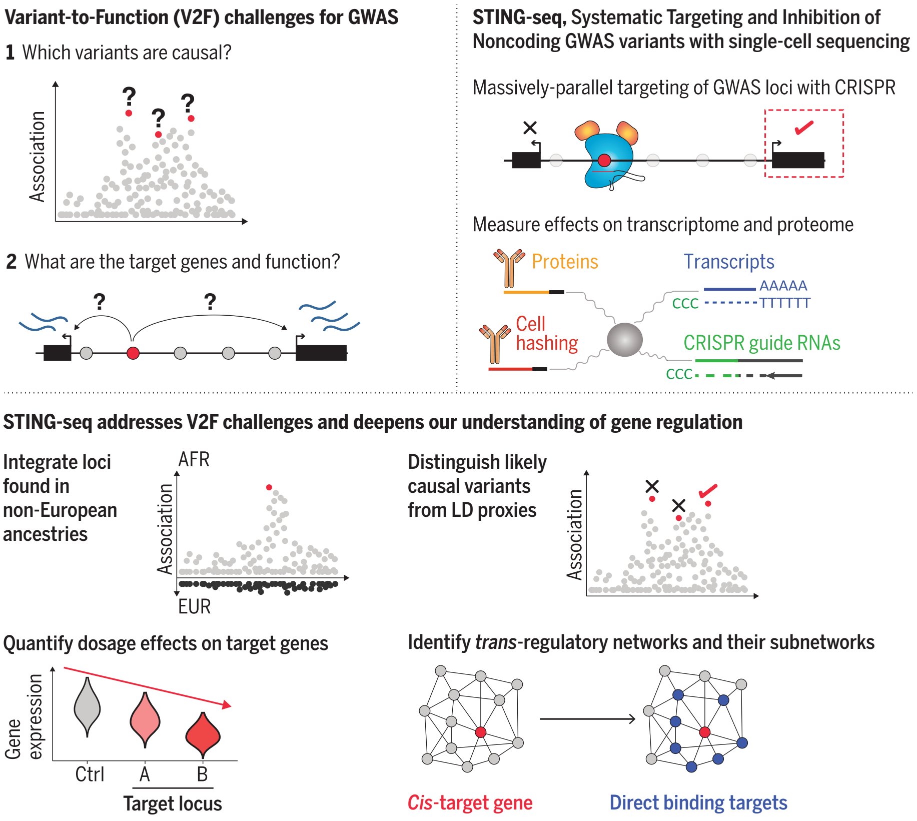
Here, we combined biobank-scale GWASs, massively parallel CRISPR screens, and single-cell sequencing to discover target genes of noncoding variants for blood trait loci with systematic targeting and inhibition of noncoding GWAS loci with single-cell sequencing (STING-seq).

RATIONALE: Blood traits are highly polygenic, and GWASs have identified thousands of noncoding loci that map to candidate *cis*-regulatory elements (CREs). By combining CRE-silencing CRISPR perturbations and single-cell readouts, we targeted hundreds of GWAS loci in a si assay, revealing target genes in *cis* and in *tr* For select CREs that regulate target genes, we performed direct variant insertion. Although silencing the CRE can identify the target gene, direct variant insertion can identify magnitude and direction of effect on gene expression for the GWAS variant. In select cases in which the target gene was a transcription factor or microRNA, we also investigated the gene-regulatory networks altered upon CRE perturbation and how these networks differ across blood cell types.

RESULTS: We inhibited candidate CREs from fine-mapped blood trait GWAS variants (from ~750,000 individual of diverse ancestries) in human erythroid progenitors. In total, we targeted 543 variants (254 loci) mapping to candidate CREs, generating multimodal single-cell data including transcriptome, direct CRISPR gRNA capture, and cell surface proteins.

We identified target genes in *cis* (within 500 kb) for 134 CREs. In most cases, we found that the target gene was the closest gene and that specific enhancer-associated biochemical hallmarks (H3K27ac and accessible chromatin) are essential for CRE function. Using multiple perturbations at the same locus, we were able to distinguish between causal variants from noncausal variants in linkage disequilibrium. For a subset of validated CREs, we also inserted specific GWAS variants using base-editing STING-seq (beeSTING-seq) and quantified the effect size and direction of GWAS variants on gene expression. Given our transcriptome-wide data, we examined dosage effects in *cis* and *trans* in cases in which the *cis* target is a transcription factor or microRNA. We found that *trans* target genes are also enriched for GWAS loci, and identified gene clusters within *trans* gene networks with distinct biological functions and expression patterns in primary human blood cells.

CONCLUSION: In this work, we investigated noncoding GWAS variants at scale, identifying target genes in single cells. These methods can help to address the variant-to-function challenges that are a barrier for translation of GWAS findings (e.g., drug targets for diseases with a genetic basis) and greatly expand our ability to understand mechanisms underlying GWAS loci. ■


Variant-to-Function (V2F) challenges for GWAS
1 Which variants are causal?
Association
?
?
?
2 What are the target genes and function?
?
?
STING-seq, Systematic Targeting and Inhibition of Noncoding GWAS variants with single-cell sequencing
Massively-parallel targeting of GWAS loci with CRISPR
×
✓
Measure effects on transcriptome and proteome
Proteins
Transcripts
AAAAA
CCC
TTTTTT
Cell hashing
CRISPR guide RNAs
CCC
STING-seq addresses V2F challenges and deepens our understanding of gene regulation
Integrate loci found in non-European ancestries
AFR
Association
EUR
Distinguish likely causal variants from LD proxies
Association
×
×
✓
Quantify dosage effects on target genes
Gene expression
Ctrl
A
B
Target locus
Identify *trans*-regulatory networks and their subnetworks
Cis-target gene
Direct binding targets

Identifying causal variants and their target genes with STING-seq. Uncovering causal variants and their target genes or function are a major challenge for GWASs. STING-seq combines perturbation of noncoding loci with multimodal single-cell sequencing to profile hundreds of GWAS loci in parallel. This approach can identify target genes in *cis* and *trans*, measure dosage effects, and decipher gene-regulatory networks.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: tlappalainen@nygenome.org (T.L.); neville@sanjanalab.org (N.E.S.)
Cite this article as J. A. Morris *et al.*, *Science* **380**, eadh7699 (2023). DOI: 10.1126/science.adh7699

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adh7699

RESEARCH ARTICLE

HUMAN GENOMICS

Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens

John A. Morris[1,2], Christina Caragine[1,2], Zharko Daniloski[1,2], Júlia Domingo[1], Timothy Barry[3], Lu Lu[1,2], Kyrie Davis[1,2], Marcello Ziosi[1], Dafni A. Glinos[1], Stephanie Hao[4], Eleni P. Mimitou[4], Peter Smibert[4], Kathryn Roeder[3,5], Eugene Katsevich[6], Tuuli Lappalainen[1,7]*, Neville E. Sanjana[1,2]*

Most variants associated with complex traits and diseases identified by genome-wide association studies (GWAS) map to noncoding regions of the genome with unknown effects. Using ancestrally diverse, biobank-scale GWAS data, massively parallel CRISPR screens, and single-cell transcriptomic and proteomic sequencing, we discovered 124 *cis*-target genes of 91 noncoding blood trait GWAS loci. Using precise variant insertion through base editing, we connected specific variants with gene expression changes. We also identified *trans*-effect networks of noncoding loci when *cis* target genes encoded transcription factors or microRNAs. Networks were themselves enriched for GWAS variants and demonstrated polygenic contributions to complex traits. This platform enables massively parallel characterization of the target genes and mechanisms of human noncoding variants in both *cis* and *trans*.

A major goal for the study of common diseases is to identify causal genes, which can clarify biological mechanisms and inform drug targets for these diseases. To this end, genome-wide association studies (GWASs) have identified thousands of genetic variants associated with disease outcomes and disease-relevant phenotypes. However, because these associations are nearly always found in noncoding regions, their target genes and functions often remain elusive. This is commonly referred to as the variant-to-function (V2F) problem (*1*, *2*).

Recent studies have used statistical fine-mapping to identify plausibly causal GWAS variants and functional genomics to find candidate *cis*-regulatory elements (cCREs) and their putative target genes (*3*–*6*). Other studies have performed CRISPR-based silencing or mutagenesis screens of noncoding regulatory elements to identify target genes (*7*–*9*). Here, we combined these approaches in a modular workflow, systematic targeted inhibition of noncoding GWAS loci coupled with single-cell sequencing (STING-seq), to identify target genes at noncoding GWAS loci using single-cell pooled CRISPR screens. We first prioritized cCREs by functional annotation and overlap with fine-mapped GWAS variants. We then tested for gene-regulatory function using pooled CRISPR inhibition (CRISPRi) and single-cell RNA-sequencing and cell surface protein measurements (Fig. 1A). For a subset of validated CREs, we also inserted specific GWAS variants using base editing STING-seq (beeSTING-seq), which couples base editing with single-cell multiomics. We demonstrate the utility of these approaches in blood cell traits by targeted perturbation of ~500 cCREs at noncoding GWAS loci, identifying target genes in *cis* and *trans* for 134 of these CREs, and further explore the effects of 46 fine-mapped noncoding C-to-T variants using precise variant insertion.

Results

Fine-mapping multiancestry blood trait GWAS to identify candidate CREs

We elected to study blood cell traits because of their high polygenicity, links to multiple common diseases, and the large number of genotyped individuals available in ancestrally diverse biobank-scale data repositories with measured blood traits (*10*–*12*). We examined 29 blood trait GWASs in the UK Biobank (UKBB) and 15 traits from the Blood Cell Consortium (BCX) (*11*), including traits from platelets, red blood cells (RBCs), and white blood cells (WBCs) (table S1A). The UKBB GWASs include 361,194 participants with European ancestries. The BCX multiancestry GWASs include 746,667 participants (76% European, 20% Asian, 2% African, 1% Hispanic/Latino, and 1% South Asian ancestries) with both multiancestry and individual population analyses. We performed statistical fine-mapping for the 29 UKBB blood trait GWASs, identifying a median of 469 conditionally independent signals and 3328 fine-mapped variants per trait (*13*, *14*). Multiancestry BCX meta-analyses identified a median of 384 conditionally independent signals and 3586 fine-mapped variants per trait. Across all BCX population-specific GWASs, excluding European ancestries, there were 42 conditionally independent signals and 418 fine-mapped variants per trait (table S1, A and B). In all cases, we found that >90% of fine-mapped variants were in noncoding regions of the genome.

For our study, we targeted cCREs from different GWASs (543 variants in 254 loci) by intersecting fine-mapped noncoding variants with biochemical hallmarks of enhancer activity, such as chromatin accessibility [assay for transposase-accessible chromatin sequencing (ATAC-seq) and DNase I hypersensitivity] and canonical histone modifications [H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq)] from the human erythroid progenitor cell line K562. K562 cells are an established and well-characterized model for blood traits. In these cells, reporter assays have identified genetic variants with erythroid-specific effects (*15*), transcription factor (TF) occupancy is strongly conserved with human proerythroblasts (*16*), gene expression and open chromatin profiles are similar to human erythrocyte progenitors (*17*), and promoter-interacting regions defined from Hi-C data are enriched for blood trait GWAS variants (*18*). The integration of functional genomic data yielded a large set of targetable variants from UKBB and BCX GWASs (table S1, C and D). The variants that we selected were often the highest-probability variant in a fine-mapped GWAS locus (294 variants) or among the 10 most probable variants (249 variants). We also prioritized variants from non-European ancestries. In total, we selected variants from BCX multiancestry analyses (339 variants), BCX non-European ancestries (118 variants), and UKBB European ancestries (86 variants) (Fig. 1B and table S1, C to E).

Optimized dual-repressor CRISPRi system

To perturb the selected cCREs, we designed (table S1F) a dual-repressor KRAB-dCas9-MeCP2 system (*19*) that yielded 50 to 60% greater gene repression when targeting transcription start sites (TSSs) or previously described enhancer loci (*7*) than a single-repressor (KRAB-dCas9) system (Fig. 1, C and D; fig. S1; and table S2). We further characterized the dual-repressor CRISPRi using a pooled library of ~2000 CRISPR guide RNAs (gRNAs) that target sites at different distances from the TSSs of ~250 essential genes. We found that dual-repressor CRISPRi had a focused activity window with minimal repression beyond 1 kb, and that most of the active gRNAs were located between −400 and +850 nucleotides (nt) from the TSS (fig. S2) (*20*).

[1]New York Genome Center, New York, NY 10013, USA. [2]Department of Biology, New York University, New York, NY 10003, USA. [3]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [4]Technology Innovation Lab, New York Genome Center, New York, NY 10013, USA. [5]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [6]Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. [7]Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 171 65 Solna, Stockholm, Sweden.
*Corresponding author. Email: tlappalainen@nygenome.org (T.L.); neville@sanjanalab.org (N.E.S.)

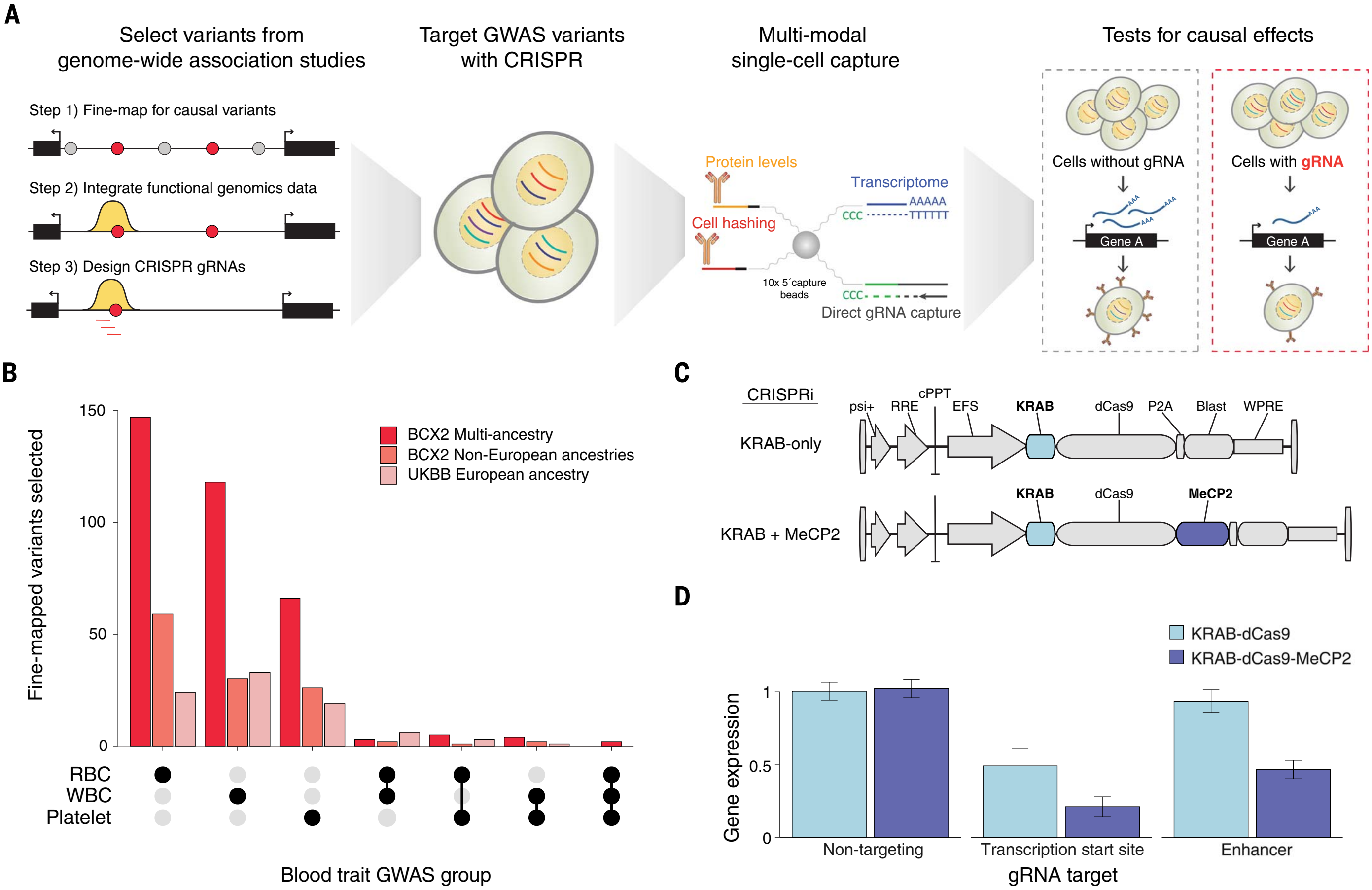


Fig. 1. Overview of STING-seq. (**A**) STING-seq pipeline for perturbation and single-cell analysis of human genetic variants from GWASs. First, plausibly causal variants are identified using statistical fine-mapping of GWAS. After further refinement of candidate cCREs using key molecular hallmarks of regulatory elements, CRISPR gRNAs are designed to target cCREs and lentivirally transduced at a high MOI into human cells. Using multimodal single-cell sequencing, target genes for GWAS variants are identified using differential transcript or protein expression. (**B**) The number of targeted GWAS variants mapping to cCREs across 29 blood traits in UKBB (n = 361,194 participants) and 15 blood traits in the BCX (n = 746,667 participants). (**C**) Lentiviral CRISPRi vector with a single-effector domain (KRAB-dCas9) or dual-effector domains (KRAB-dCas9-MeCP2). (**D**) Mean digital PCR gene expression in human erythrocyte cells (K562) by targeting the TSSs and known enhancers of three genes (*MRPS23*, *SLC25A27*, and *FSCN1*) with either single-effector KRAB-dCas9 or dual-effector KRAB-dCas9-MeCP2 CRISPRi. Error bars indicate SEM.

A massively parallel assay to perturb CREs and find their target genes

We designed STING-seq gRNA libraries to target each blood trait cCRE with up to three gRNAs using the dual-repressor CRISPRi (KRAB-dCas9-MeCP2). These gRNAs were optimized for minimal off-target activity (*21*, *22*). We also embedded into the STING-seq library several control gRNAs: negative (nontargeting) controls (*23*), positive controls (targeting highly-expressed genes at TSSs), and, to estimate the average number of perturbations per cell through flow cytometry, multiple gRNAs targeting a gene encoding a ubiquitously expressed cell surface protein (*CD55*) (table S3A).

We transduced K562 cells with pooled library virus at a high multiplicity of infection (MOI), which we verified by flow cytometry for CD55 (fig. S3). We then simultaneously captured four different modalities from single cells: CRISPR gRNAs, transcriptomes, cell surface proteomes through oligo-tagged antibodies, and cell hashing (table S3B) (*24*, *25*). We recovered 46,583 single cells with a median of 13 gRNAs per cell and with each cCRE targeted in a median of 978 cells (fig. S4, A and B, and table S3C). To perform differential expression testing, we recently developed a conditional resampling approach (SCEPTRE) that yields state-of-the-art calibration on CRISPR single-cell datasets to connect perturbations with changes in gene and protein expression (*26*). Using SCEPTRE, we grouped together gRNAs targeting each cCRE, performing 4627 pairwise tests with a median of seven genes tested per cCRE within 500 kb for *cis* effects (*27*). We observed good calibration for positive and negative controls: Nontargeting gRNAs had no effect, and control genes had decreased expression or protein levels at a 5% false discovery rate (FDR) (Fig. 2A; fig. S5; and table S3, C to E). In most cases, target genes in *cis* for GWAS variants were more likely to be identified when both H3K27ac and open chromatin peaks were present (Fig. 2B).

Of 539 targeted cCREs (from 254 loci), 134 (from 91 loci) had a target gene within 500 kb (Fig. 2C and table S3F). When examining gRNAs that target the same CRE, the number of cells was most directly responsible for statistical power, and not distance between gRNAs or predicted off-target effects (fig. S6). We found minimal differences in target gene identification when looking at potential *cis* effects within a smaller (100 kb) or larger (1 Mb) window surrounding the targeted cCRE (table S3F) (*28*–*30*).

Most *cis*-target genes were also the closest gene to the variant; however, there were 10 *cis*-target genes that were the second closest and eight that were farther away (Fig. 2D). We identified a single *cis*-target gene for 116 CREs and identified 18 CREs with two or more

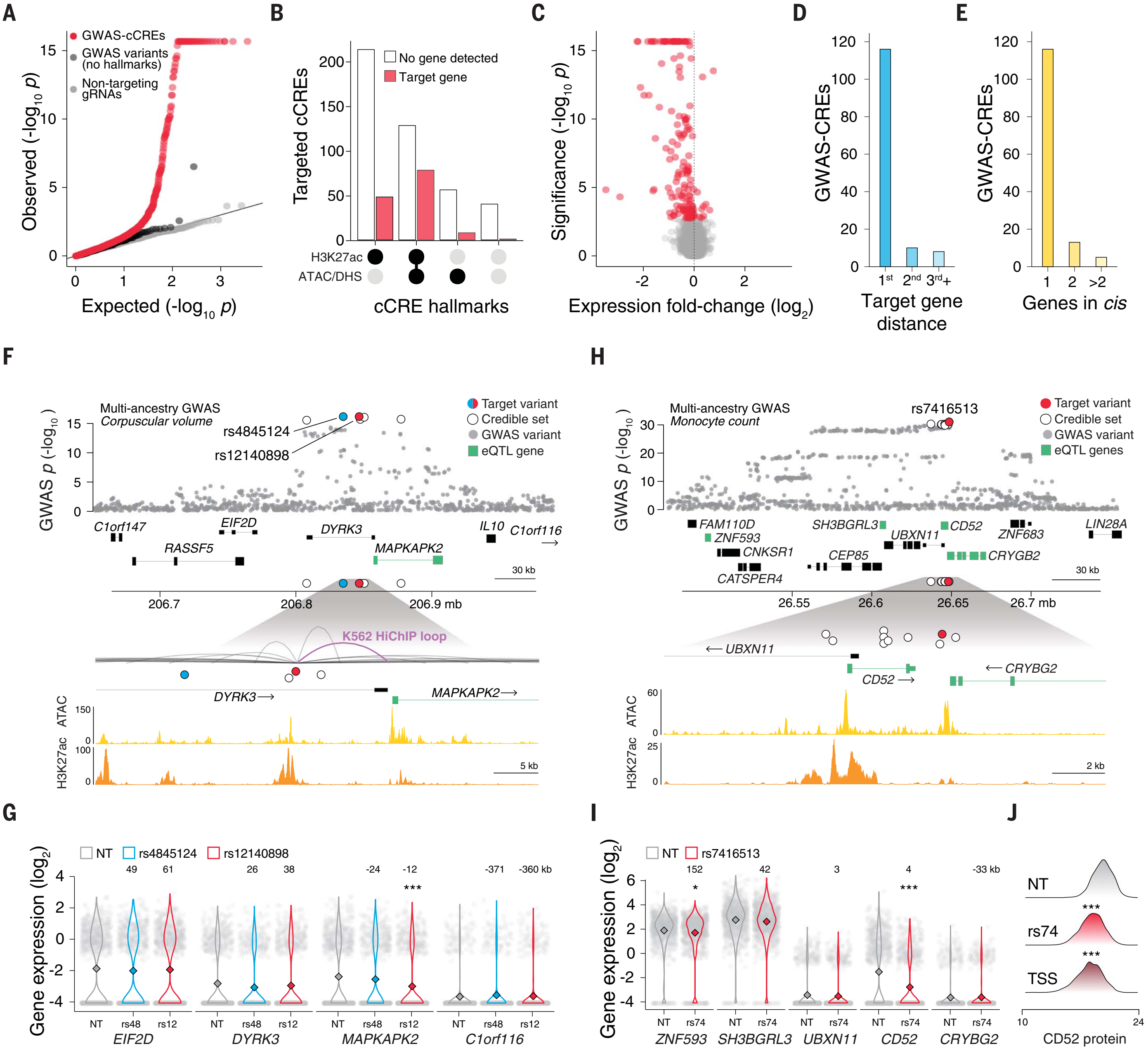


Fig. 2. Mapping *cis*-regulatory target genes for blood trait GWAS variants. (**A**) Quantile-quantile plot of *cis* effects (within 500 kb) of 531 cCREs (defined as regions with regulatory hallmarks ATAC/DHS or H3K27ac) overlapping 535 GWAS variants (GWAS-cCREs), 41 GWAS variants without CRE hallmarks, and 32 nontargeting (NT) gRNAs. Genes for NT tests were randomly sampled from the set of genes in *cis* for targeting gRNAs. We identified 154 pairs of target genes and CREs for GWAS variants with CRE hallmarks, one target gene-CRE pair for GWAS variants without CRE hallmarks, and no target genes with NT gRNAs significant at a 5% FDR (Benjamini-Hochberg–adjusted SCEPTRE *P* value). (**B**) Targeted GWAS-cCREs with and without target genes detected and their functional hallmarks of enhancer activity (ATAC/DHS or H3K27ac) in K562 cells. (**C**) Volcano plot of *cis*-regulatory effects. Significant pairs of genes and GWAS-CREs are indicated in red. (**D**) Distance to gene rank for GWAS-CREs and target genes, where genes were ranked according to closest TSS to a given GWAS-CRE. (**E**) Number of target genes in *cis* per GWAS-CRE. (**F**) Top: For a multiancestry corpuscular volume locus, two fine-mapped variants were targeted, the lead variant, rs4845124 (blue), and rs12140898 (red). *MAPKAPK2* (green) was nominated as the target gene by fine-mapped blood cell eQTLs for both variants. Bottom: rs12140898 mapped to a K562 HiChIP loop connecting its GWAS-CRE to the *MAPKAPK2* promoter. (**G**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs (rs4845124 or rs12140898) or NT. Only rs12140898 had a target gene within 500 kb, *MAPKAPK2*. (**H**) For a multiancestry monocyte count locus, one fine-mapped variant was targeted, the lead variant, rs741613 (red). *ZNF593*, *SH3BGRL3*, *CD52*, and *CRYBG2* (green) were nominated as target genes by fine-mapped blood cell eQTLs. (**I**) Single-cell gene expression for cells with gRNAs targeting the GWAS-cCRE rs741613 or NT. *CD52* and *ZNF593* were both identified as target genes. (**J**) Single-cell protein expression for cells with gRNAs targeting the GWAS-cCRE rs741613, the *CD52* TSS, or NT. Asterisks denote significant *q* values, Benjamini-Hochberg–adjusted SCEPTRE *P* values (**q* < 0.05, ***q* < 0.01, ****q* < 0.001).

cis-target genes (Fig. 2E). We also targeted 41 variants that were the most plausibly causal variants at their respective loci but did not overlap biochemical hallmarks of enhancers. From the 41 variants we targeted that did not overlap called peaks for biochemical hallmarks of enhancers, there was one variant (rs106585 for WBC counts) with a significant target gene, *LTBR* [$\log_2$ fold-change (FC) = −0.38, SCEPTRE $P = 3.1 \times 10^{-7}$] (Fig. 2A, fig. S7, and table S3G). Upon further inspection, we found a weak enhancer-associated histone modification (H3K27ac) at this locus despite the lack of a called peak, suggesting that biochemical hallmarks of enhancer activity are required and that spurious signals from inactive chromatin are rare (fig. S8).

We next sought to characterize concordance between *cis*-target genes identified using STING-seq and other methods such as physical contact mapping and allele-specific expression. To identify gene promoters anchored in three-dimensional space to H3K27ac-bound chromatin, we generated H3K27ac HiChIP libraries in K562 cells. Of the 134 STING-seq CREs and their 124 target genes, we observed 32 CREs in which the same gene was identified with H3K27ac HiChIP contacts, 27 CREs in which the same gene was identified through expression quantitative trait loci (eQTL) mapping of the same fine-mapped variant ([31]), and 73 CREs in which the same gene was identified through a transcriptome-wide association study (TWAS) of a blood trait ([32]). Although the sensitivity of TWAS for target gene identification is reasonably high (54%), we and others have found that specificity can be low using this approach ([33]). Additionally, 54 CREs with fine-mapped GWAS variants had allele-specific effects on enhancer activity or TF binding ([34], [35]), suggesting that these variants are causal at their respective CREs (table S3F).

Identification of causal variants and their impact on gene and protein expression

In the STING-seq dataset, we identified examples in which multiple lines of orthogonal evidence converged to explain how a CRE regulates a *cis*-target gene. For example, the lead variant (rs4845124) at a locus associated with mean corpuscular volume in multiancestry meta-analyses (GWAS $P = 6.9 \times 10^{-17}$) was fine-mapped as plausibly causal (in the 95% credible set with posterior probability ≥ 1%); however, upon CRISPR inhibition of the cCRE, there was no target gene (Fig. 2, F and G). Fine-mapping of this locus nominated a second plausibly causal variant mapping to a cCRE (rs12140898), and its inhibition identified *MAPKAPK2* as the target gene ($\log_2$ FC = −0.64, SCEPTRE $P = 2.2 \times 10^{-16}$). Both variants were fine-mapped eQTLs for *MAPKAPK2* in neutrophils. However, only rs12140898 had predicted allele-specific effects on SPI1 binding and mapped to a HiChIP contact domain for the *MAPKAPK2* promoter. Therefore, although eQTL studies nominated the correct target gene for this locus, it was through experimental CRE to gene mapping that we pinpointed the most likely causal GWAS variant. Most of the targeted GWAS variants did not have supporting evidence from eQTL data but were within proximity (500 kb) of a TWAS gene, demonstrating that we can uncover genes that may be underpowered by eQTL mapping and refine TWAS results that may have high false-positive rates (table S3F) ([33]).

To disentangle loci with multiple target genes in *cis*, we can combine targeted CRE inhibition and gene inhibition. For example, the lead variant (rs7416513) at a locus associated with monocyte count in multiancestry meta-analyses (GWAS $P = 3.8 \times 10^{-32}$) was fine-mapped as plausibly causal (Fig. 2H). This variant maps to an intergenic region between the gene bodies of *CRYBG2* and *CD52*, and the gene with the closest TSS is *UBXN11*. Given this, it is unclear which of these genes, if any, might be the target gene. The variant is also a fine-mapped blood cell eQTL for multiple genes in the locus (*CD52*, *CRYBG2*, *SH3BGRL3,* and *ZNF593*), further obscuring the target gene. Upon inhibiting the rs7416513-CRE, we detected *CD52* as the most significantly altered gene ($\log_2$ FC = −1.6, SCEPTRE $P = 2.2 \times 10^{-16}$) (Fig. 2I), and *ZNF593* also had a weak change in expression ($\log_2$ FC = 0.04, SCEPTRE $P = 1.3 \times 10^{-3}$), with no effect on *SH3BGRL3* or *CRYBG2*. Directly targeting *CD52* does not influence *ZNF593* (SCEPTRE $P = 0.65$) expression, suggesting the rs7416513-CRE has a pleiotropic regulatory effect on multiple genes.

Using single-cell proteomics, we also detected a significant decrease in cell surface CD52 protein expression upon rs7416513-CRE inhibition ($\log_2$ FC = −0.1, SCEPTRE $P = 1.2 \times 10^{-15}$) (Fig. 2J), demonstrating that CREs with GWAS variants modulate not only *cis*-target gene expression but also protein expression. CD52 protein can be targeted with alemtuzumab to improve clinical outcomes in patients with myelodysplastic syndrome, suggesting that this may be the causal gene for the monocyte count GWAS association ([36]). The rs7416513-derived C allele is associated with increased monocyte count in multiancestry meta-analyses (GWAS effect = 0.025, $P = 3.8 \times 10^{-32}$) ([11]) and also with increased CD52 expression in monocytes (eQTL estimate = 0.71, $P = 4.5 \times 10^{-31}$) ([37]), highlighting the power of STING-seq to connect variants to druggable genes and identify those variants that may affect response to drugs such as alemtuzumab.

Target gene discovery in STING-seq using non-European and multiancestry GWASs

Historically, most GWAS loci have been identified using individuals of European ancestry ([38]). Recent efforts to use non-European ancestries and to combine multiple ancestries for GWASs have yielded numerous new associations ([11], [39], [40]). By leveraging ancestry-specific and multiancestry GWASs, we increased the discovery space of CREs and target genes for STING-seq. We identified 16 CREs with *cis*-target genes from GWAS variants in non-European ancestries. For example, we identified *ATP1A1* as the target gene for a locus associated with neutrophil counts exclusively in African ancestries (fig. S9, A and B). The lead variant (rs6674304) was fine-mapped as plausibly causal in individuals with African ancestries (GWAS $P = 3.4 \times 10^{-44}$) but not in individuals with European ancestries (GWAS $P = 0.58$). Although rs6674304 did not map to any cCREs, statistical fine-mapping nominated three additional variants that did map to cCREs (rs6660743, rs12087680, and rs7544679) (fig. S9A). We targeted all three variants using STING-seq and found that targeting the rs12087680-CRE revealed the *cis*-target gene *ATP1A1* ($\log_2$ FC = −0.35, SCEPTRE $P = 2.0 \times 10^{-10}$) (fig. S9B). ATP1A1 maintains electrochemical gradients of sodium and potassium ions, and prior work has linked both *ATP1A1* and neutrophil counts with hypertension ([41]–[43]). As the *ATP1A1* CRE demonstrates, STING-seq using non-European and multiancestry GWAS can identify new trait genes.

A pleiotropic CRE in the APOE *and* APOC1 *locus*

In a minority of STING-seq CREs, we identified multiple *cis*-target genes that may occur through direct regulation of multiple genes or indirect effects on other nearby genes driven by a single *cis*-target gene. These outcomes can be difficult to distinguish without additional perturbations or a known gene-regulatory network. For example, we found that rs1065853 was the lead variant and fine-mapped as plausibly causal for an immature RBC trait (high light scatter reticulocyte percentage) at its locus (GWAS $P = 5.8 \times 10^{-48}$) (fig. S9C). This variant mapped to an intergenic region between the gene bodies of *APOE* and *APOC1*, with *APOE* being the closest gene, and was also associated with high- and low-density lipoprotein levels ([44]). Upon inhibiting the rs1065853-CRE, we observed significant decreases in expression for both *APOE* ($\log_2$ FC = −0.63, SCEPTRE $P = 2.8 \times 10^{-6}$) and *APOC1* ($\log_2$ FC = −0.27, SCEPTRE $P = 3.5 \times 10^{-6}$) (fig. S9D). Previous studies have shown that *APOE* and *APOC1*, which encode apolipoproteins E and C1, respectively, influence blood lipids and diverse ailments including cardiovascular disease and Alzheimer's disease ([45], [46]). To help distinguish direct and indirect regulation, we used a prior genome-wide Perturb-seq (GWPS) study in the same cell line (K562) to determine whether *APOE* or *APOC1* regulate one another ([47]). *APOC1* expression was unchanged upon *APOE* inhibition (GWPS z = 0.02), but *APOE* expression was decreased upon *APOC1* inhibition (GWPS z = –1.4). *APOE* and *APOC1* direct inhibition suggests that rs1065853-CRE may target either *APOC1* alone (even though *APOE* is the closest gene) or both *APOC1* and *APOE*. Because these genes work in a coordinated fashion to regulate lipid metabolism (*48*), the co-regulation of these genes is a notable observation of regulatory pleiotropy that may contribute to trait associations.

Targeting multiple CREs in the PTPRC *locus reveals nonfunctional linkage disequilibrium proxies*

We also examined loci with several fine-mapped variants near a single gene. At the *PTPRC* locus, we targeted nine variants that were fine-mapped variants for 10 traits (fig. S10A and table S1E) and mostly not in strong linkage disequilibrium (LD), as quantified by pairwise R^2 from 1000 Genomes (*49*) (fig. S10B). The nine variants mapped to distinct cCREs: One was 5 kb before the *PTPRC* TSS and the remaining eight were in the first intron, from 2 to 42 kb after the TSS (fig. S10C). We observed modulation in *PTPRC* when targeting six of the cCREs (fig. S10D). For the cCREs with no effect, we found that two variants were in high LD ($R^2 \geq 0.95$) with variants mapping to *PTPRC* CREs, suggesting that these may be nonfunctional variants in LD with functional variants (i.e., nonfunctional LD proxies). For all CREs, *PTPRC* was the only significant target gene and thus is very likely the causal GWAS gene (table S1E).

The high allelic heterogeneity, which is driven by multiple independent regulatory variants in distinct CREs modulating *PTPRC* expression, and the 10 blood trait associations suggest that the CREs may have cell type–specific activity. That is, different CREs may regulate *PTPRC* in different contexts, given that the 10 trait associations include RBCs, WBCs, and platelet traits (fig. S10A).

We found that experimental evidence (e.g., STING-seq) is required to link these CREs to *PTPRC* expression. None of the targeted variants were fine-mapped blood eQTLs, and only a single targeted variant, rs1326279, showed evidence of allele-specific effects on TF binding (*31*, *35*). Thus, in silico methods that use eQTL data are insufficient to measure the impact of the CREs on *PTPRC* expression.

Direct GWAS variant insertion with beeSTING-seq

Next, we sought to expand the STING-seq approach to precise insertion of fine-mapped GWAS variants with base editing. We fused a cytosine base editor (FNLS-BE3) to a protospacer-adjacent motif (PAM)–flexible Cas9 variant (SpRY) (table S1F) and validated activity using gRNAs designed to disrupt splice junctions in *CD46*, which encodes a ubiquitously expressed cell surface protein, in an arrayed fashion (Fig. 3, A and B, and table S3H) (*50*, *51*). We observed up to ~70% knockdown of CD46 when targeting splice sites with diverse PAM sequences and an average knockdown of 27% (n = 12 target sites), similar to prior pooled screens using base editing (*52*, *53*), (fig. S11 and table S3H). We then performed a single-cell pooled base editing screen (beeSTING-seq) targeting 46 C>T fine-mapped GWAS variants mapping to 42 STING-seq–identified CREs with three gRNAs each (table S3I). We tested for direct effects on known target genes and found that 32 of 46 had at least two gRNAs with concordant effects and that all three gRNAs had concordant effects for 17 variants (Fig. 3C and table S3, J and K). We identified three sets of beeSTING-seq gRNAs with *cis*-regulatory effects on the same target genes identified using STING-seq (5% FDR) with no enrichment of nontargeting (negative control) gRNAs (Fig. 3D and table S3, L and M).

In one case, beeSTING-seq gRNAs targeted the lead variant (rs142122062) at a locus associated with RBC volume in multiancestry meta-analyses (GWAS $P = 8.2 \times 10^{-11}$) (Fig. 3E and table S3M). Targeted inhibition of the rs142122062-CRE decreased *APPBP2* expression ($\log_2$ FC = –0.46, SCEPTRE $P = 2.5 \times 10^{-4}$) and identified it as the target gene for this locus (Fig. 3F). For beeSTING-seq, we were able to design multiple gRNAs capable of inserting the same single-nucleotide edit by capitalizing on the targeting flexibility of SpRY Cas9 (*51*). With direct insertion of the rs142122062-T allele with two independent gRNAs, we observed a significant increase in *APPBP2* expression (combined $\log_2$ FC = 0.74, SCEPTRE $P = 7.6 \times 10^{-5}$) (Fig. 3G), demonstrating the ability of beeSTING-seq to identify GWAS variants that act to increase expression. Both gRNAs exclusively edit the GWAS variant, because it is the only C nucleotide within the editing window (*50*). Using TWAS, we found that amyloid precursor protein, which APPBP2 binds, had the strongest association with RBC counts (*54*), suggesting a possible mechanism of how altered *APPBP2* expression affects RBC traits. In this manner, beeSTING-seq can more precisely determine the effect of GWAS variants, moving beyond CRE inhibition to reveal the impact of specific alleles on target gene expression.

CRE-driven, dosage-dependent, transcriptome-wide changes in gene expression

To understand the impact of GWAS-CREs on gene expression across the genome, we performed transcriptome-wide differential expression tests. We applied a strict (1%) FDR to identify target genes in *trans* and again found good calibration with nontargeting gRNAs (Fig. 4A and table S3C). We observed *trans* effects for CREs that targeted in *cis* the TFs *GFI1B*, *NFE2*, *IKZF1*, *HHEX*, and *RUNX1* and the host genes of microRNAs (miRNAs) *miR-142* and *miR-144/451* (Fig. 4A and tables S3F and S4A). These TFs and miRNAs are known to play key roles in hematopoietic stem cell differentiation (*55*–*61*).

For *GFI1B*, we identified two independent CREs with *trans* effects. One variant (rs524137), associated with monocyte percentage and basophil counts, maps to an intergenic CRE 11.5 kb downstream of *GFI1B* (Fig. 4B). The other variant (rs73660574), associated with several RBC traits (mean sphered corpuscular volume, immature reticulocyte fraction, mean reticulocyte volume, and mean corpuscular hemoglobin), maps to a CRE in an intron of *GFI1B* (Fig. 4B). These CREs exhibited independent dosage effects on *GFI1B* expression, with the rs524137-CRE having an ~70% stronger effect than the rs73660574-CRE. Thus, perturbing either rs73660574- or rs524137-CREs led to changes in the expression of *GFI1B* (Fig. 4C) and its target genes. To better understand the *trans* effects of these two *GFI1B* CREs, we examined gene-expression changes in all 1161 differentially expressed genes identified from the rs524137-CRE (Fig. 4D). For these genes, we observed a high correlation between perturbations targeting each CRE (r = 0.84), even though many of the gene expression changes were more modest when perturbing the rs73660574-CRE. We found a linear dosage relationship between the *trans*-regulatory effects for the CREs that agreed with the difference in their effect on *cis* (*GFI1B*) expression (~1.3-fold) (Fig. 4, C and D). Using single-cell proteomics in the same cells, we observed changes in protein levels for nine of the genes in the *GFI1B* network; for these, changes in transcript expression and protein levels were highly correlated (r = 0.9) (fig. S12). This example demonstrates how GWAS variants mapping to CREs perturb regulatory networks, and that these changes at the RNA level also alter protein expression.

In addition to *GFI1B*, we also observed CRE dosage effects on target gene expression and regulatory networks for *NFE2* (rs79755767, associated with hematocrit and red cell distribution width, and rs35979828, associated with eosinophil count, mean corpuscular hemoglobin, and monocyte count) (Fig. 4E). When targeting these variants, we observed dosage effects on *NFE2* expression (rs79755767-CRE $\log_2$ FC = –1.1, SCEPTRE $P = 2.2 \times 10^{-16}$; rs35979828-CRE $\log_2$ FC = –0.6, SCEPTRE $P = 2.2 \times 10^{-16}$) (Fig. 4F) and on a 343-gene regulatory network (r = 0.78) (Fig. 4G). These results reinforce our findings that fine-mapped GWAS variants at independent CREs have independent effects, not only on target gene

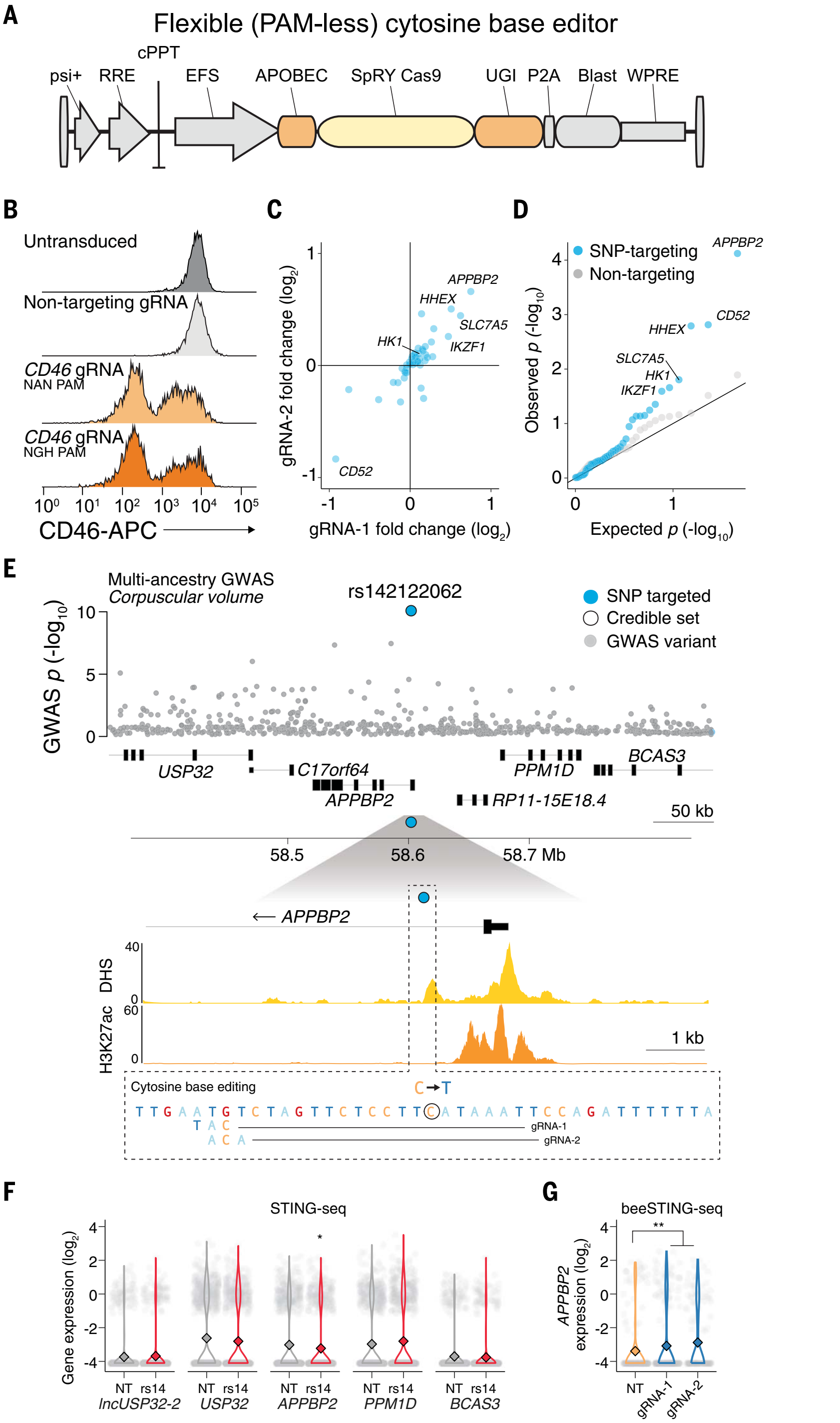


Fig. 3. Precise GWAS variant editing with beeSTING-seq. (**A**) Lentiviral CRISPR base editor (FNLS-BE3) with a relaxed PAM SpCas9 variant, SpRY, for beeSTING-seq. (**B**) Flow cytometry of CD46 cell surface protein after base editing at *CD46* splice donor sites. CD46 knockdown was compared with untransduced and NT controls. (**C**) Target gene fold change for the two gRNAs with the most concordant effects for each variant. (**D**) Quantile-quantile plot of NT gRNAs and gRNAs targeting 46 fine-mapped GWAS variants mapping to STING-seq GWAS-CREs with *cis*-effect genes. Genes for NT tests were randomly sampled from the set of genes in *cis* for targeting gRNAs. (**E**) Top: For a multiancestry corpuscular volume locus, one fine-mapped variant was targeted, the lead variant, rs142122062 (blue). Bottom: Base editing by gRNA-1 and gRNA-2 changes the rs142122062 allele from reference to alternative; for both gRNAs, this is the only cytosine in the base editing window. (**F**) Single-cell gene expression for cells with gRNAs targeting the GWAS-cCRE rs142122062 or NT. *APPBP2* was identified as a *cis*-target gene. (**G**) beeSTING-seq of rs142122062 increases *APPBP2* expression with two independent gRNAs with positions shown in (E). Asterisks denote significant *q* values, Benjamini-Hochberg–adjusted SCEPTRE *P* values (*$q < 0.05$, **$q < 0.01$, ***$q < 0.001$).

expression, but also on entire regulatory networks in *trans*.

A limitation of many GWAS functional interpretation approaches is that they focus on nearby protein-coding genes and overlook relevant noncoding RNAs. With STING-seq, we also identified regulatory networks for miRNAs, which can have a broad impact on gene regulation. For example, STING-seq at the CRE for rs2526377, the most plausibly causal variant for a locus associated with platelet count locus, revealed no protein-coding *cis*-target genes (fig. S13A). However, when examining noncoding transcripts, we found a differentially expressed noncoding transcript, *AC004687.1*, which is also known as the *miR-142* host gene ($\log_2$ FC = –1.8, SCEPTRE $P = 2.2 \times 10^{-16}$) (fig. S13B). This finding is further supported by prior work in the context of Alzheimer's disease showing that the risk allele decreases *miR-142* host gene promoter activity (*62*, *63*).

For STING-seq perturbation of rs2526377, we detected a 119 gene *trans*-regulatory network (fig. S13C). The top up-regulated genes within the rs2526377 *trans*-regulatory network (*WASL* and *CFL2*) were also the top up-regulated genes in miR-142 knockout mice (*60*). This lends further support that the *trans*-regulatory effects of rs2526377 perturbation are caused by *cis* effects on miR-142, as found in STING-seq. This *cis*-target miRNA and its regulatory network can be easily missed when considering only protein-coding genes for target gene annotation.

We also analyzed *trans* effects with direct variant insertion using beeSTING-seq. We could detect changes in regulatory network expression in the expected direction upon inserting the rs12784232-A allele (associated with lymphocyte percentage) and the rs6592965-A allele (corpuscular hemoglobin), which mapped to the *HHEX* and *IKZF1* GWAS-CREs, respectively. In contrast to GWAS-CRE inhibition, which decreased expression of *HHEX* and *IKZF1* (Fig. 4A), direct variant insertion resulted in increased expression of the *cis*-target genes and, accordingly, *trans* effects for genes tended to switch directions in differential expression compared with STING-seq. Specifically, we observed that 60 to 70% of *HHEX* and *IKZF1* network genes had reversed directions of effect, demonstrating that GWAS variants that act to increase expression can affect networks in discordant directions from CRE silencing.

Enrichment of cis*-target binding sites and GWAS genes in* trans*-regulatory networks*

To better characterize how CREs with target genes in *trans* alter blood cell phenotypes, we examined genome-wide binding for GFI1B, NFE2, IKZF1, and RUNX1 (ChIP-seq) (*64*, *65*) and sequence-based predicted targets of miR-142 and miR-144/451 (TargetScan) (*66*, *67*).

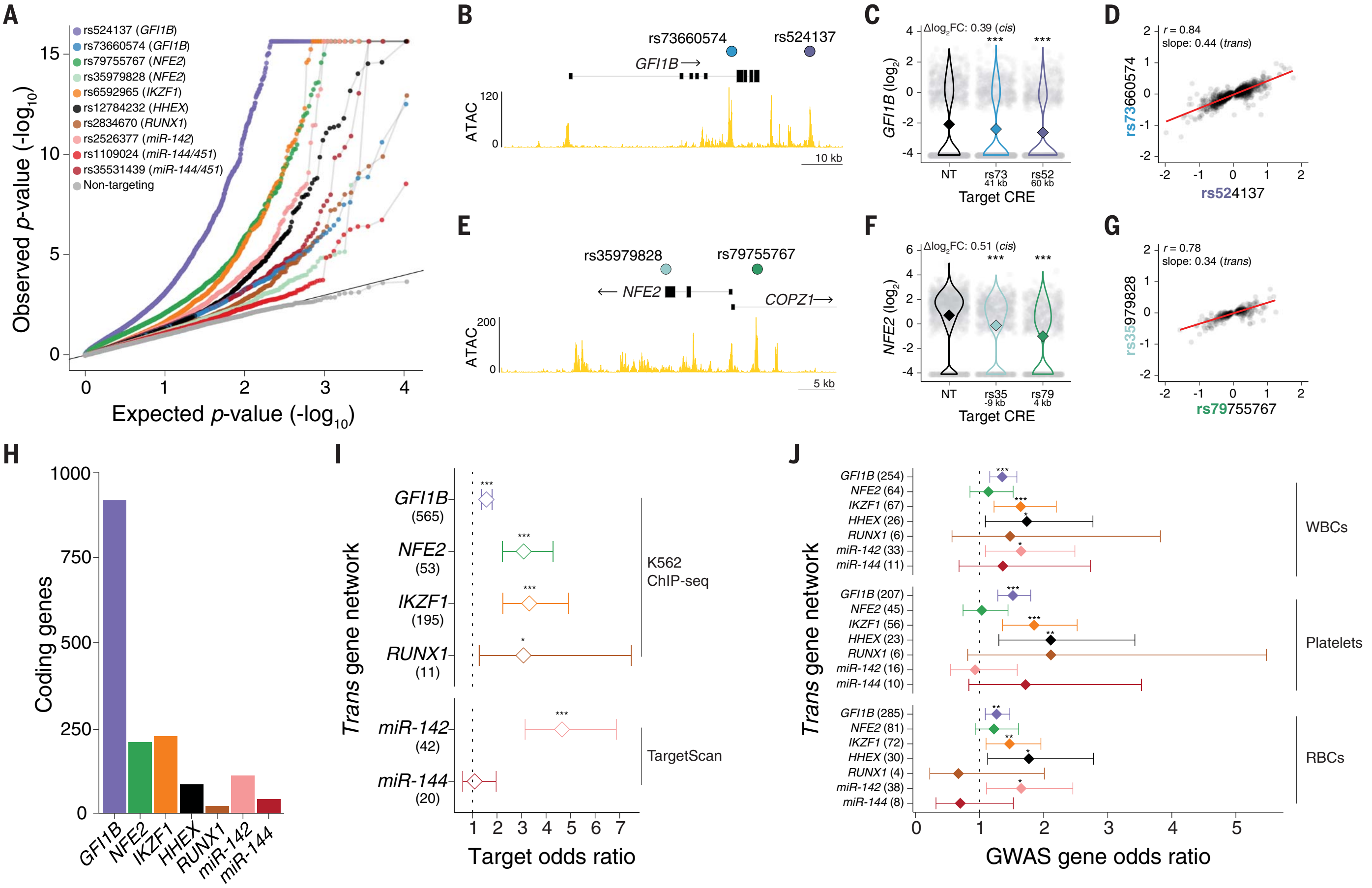


Fig. 4. *Trans*-regulatory network discovery of genes that affect diverse blood cell traits. (**A**) Quantile-quantile plots of *trans* effects (whole transcriptome) of GWAS-CREs and NT gRNAs. We identified significant genes at a 1% FDR (Benjamini-Hochberg–adjusted SCEPTRE *P* value) for GWAS-CREs with TFs (*GFI1B*, *NFE2*, *IKZF1*, *HHEX*, and *RUNX1*) and miRNAs (*miR-142* and *miR-144/451*) as *cis*-regulatory target genes. No *trans* effects were found for NT gRNAs. (**B**) Two GWAS-CREs targeted at the *GFI1B* locus, rs524137 and rs79755767. (**C**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs at the *GFI1B* locus (rs524137 or rs79755767) or NT. (**D**) Expression of rs524137-CRE significant *trans*-target genes in cells with perturbation of either GWAS-CRE at the *GFI1B* locus (rs524137 or rs79755767) (n = 1161 genes at a 1% FDR). (**E**) Two GWAS-CREs targeted at the *NFE2* locus, rs79755767 and rs35979828. (**F**) Single-cell gene expression for cells with gRNAs targeting GWAS-cCREs at the *NFE2* locus (rs79755767 or rs35979828) or NT. (**G**) Expression of rs79755767-CRE significant *trans*-target genes in cells with perturbation of either GWAS-CRE at the *GFI1B* locus (rs79755767 or rs35979828) (n = 343 genes at a 1% FDR). (**H**) Protein-coding genes with changes in expression for *trans*-regulatory networks. (**I**) Gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) for TF and miRNA targets within each *trans*-regulatory network. Targets for TFs were given as the closest gene to each TF-specific ENCODE K562 ChIP-seq peak and for miRNAs as TargetScan-predicted targets on the basis of sequence. (**J**) For each *trans*-regulatory network, gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) of closest genes to fine-mapped variants from WBC, platelet, and RBC GWASs from 29 UKBB blood traits and 15 BCX blood traits. Asterisks: (C) and (F), significant Benjamini-Hochberg–adjusted SCEPTRE *P* values; (I) and (J), logistic regression *P* values (*P < 0.05, **P < 0.01, ***P < 0.001).

We investigated whether the closest genes to each ChIP-seq peak or predicted miRNA target genes were enriched in STING-seq *trans*-regulatory networks (Fig. 4H and table S4B). We observed enrichments of predicted target genes for *GFI1B*, *NFE2*, *IKZF1*, *RUNX1*, and *miR-142* (odds ratio = 2.4 ± 1.9, mean ± SEM) (Fig. 4I and table S4C). Thus, perturbing CREs can reveal second-order interactions for regulatory networks driven by TFs or miRNAs.

A related and pertinent question is whether the genes in the *trans*-regulatory networks identified by STING-seq may also play a role in blood traits and if they also harbor *cis*-regulatory genetic variants. To answer this question, we constructed a set of putatively causal genes for each of the 29 UKBB and 15 BCX GWASs by selecting the closest protein-coding genes to fine-mapped variants of GWAS loci. We then grouped them by cell type, generating gene sets for platelets, RBCs, and WBCs that were mostly distinct (fig. S14 and table S4B). For nearly all *trans*-regulatory networks, we found enrichments for blood cell GWAS genes (Fig. 4J and table S4C). These blood cell trait GWAS loci enrichments indicate that the known roles of these genes in hematopoiesis and cell differentiation are mediated by their effects on regulatory networks. Furthermore, identification of the *trans* genes with STING-seq pinpointed regulatory networks for which polygenic perturbation by distinct variants across the genome appears to contribute to the GWAS signal. This suggests a mechanistic importance for networks themselves, for which we do not need to functionally determine V2F per locus if we know the pathway through which they are likely to act, similar to recent work focusing on perturbation of target genes ([68]).

Trans-regulated genes reveal biological mechanisms and cell types of trait associations

Given these relationships between *trans*-regulated genes and GWAS loci, we analyzed the structures of these regulatory networks to better

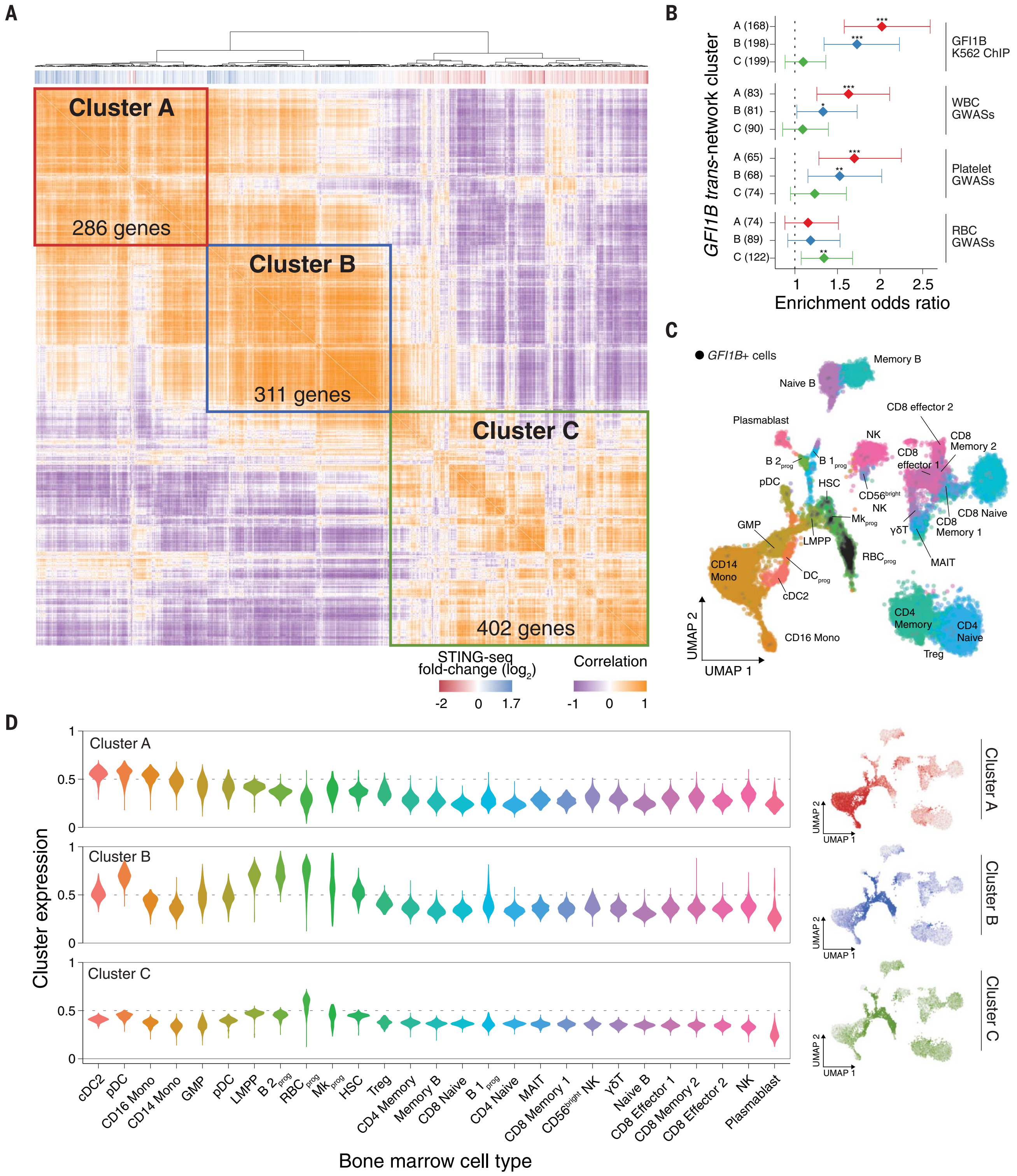


Fig. 5. Subnetworks of *GFI1B* target genes are expressed in specific hematopoietic progenitors and differentiated cells. (**A**) Coexpression matrix of rs524137-CRE *GFI1B* network genes in K562 with hierarchical clustering. Three clusters (A, B, and C) are indicated. The vertical bars below the dendrogram indicate if genes had increased (blue) or decreased (red) expression upon inhibiting the *GFI1B* CRE. (**B**) For each *trans*-regulatory *GFI1B* subnetwork (cluster), gene set enrichment odds ratios (diamonds) and 95% confidence intervals (lines) of closest genes to GFI1B K562 ChIP-seq peaks (top) and fine-mapped variants (bottom) from WBC, platelet, and RBC GWASs from 29 UKBB blood traits and 15 BCX blood traits. Asterisks denote logistic regression *P* values ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). (**C**) UMAP of human bone marrow cell gene expression from 35 Human Cell Atlas donors. Labels and colors indicate cell types. B 2_{prog}, progenitor B-2 cells; RBC_{prog}, RBC progenitors; DC_{prog}, dendritic cell progenitors (for full list, see table S4D). The black dots denote cells expressing *GFI1B*. *GFI1B* is most highly expressed in RBC progenitors, megakaryocyte progenitors, and hematopoietic stem cells. (**D**) Expression of genes from clusters A, B, and C in each human bone marrow cell type (left) and in each cell in the UMAP space from (C) (right).

understand the mechanistic roles of specific genes in blood traits. Using single-cell gene coexpression and clustering, we identified coexpressed gene clusters for each of the loci (Fig. 5A and fig. S15). For the *trans*-acting gene *GFI1B*, we identified two clusters (A and B) of genes with increased expression upon *GFI1B* CRE repression with STING-seq. These clusters were the most strongly enriched for *GFI1B* binding sites (Fig. 5B and table S4, B and C). A third cluster (C) consisted primarily of genes with decreased expression, which were not enriched for GFI1B-binding sites. Clusters A and B were enriched for genes from platelet and WBC GWASs, whereas cluster C was only enriched for genes from RBC GWASs.

To further refine and validate the individual cell types involved with different clusters of coregulated genes, we integrated the *GFI1B* coexpression network with primary cells from the Human Cell Atlas, which includes progenitors and/or differentiated cell types for platelets, WBCs, and RBCs. Specifically, we used single-cell RNA sequencing from 35 bone marrow donors (69, 70), because bone marrow includes a rich sample of multipotent progenitor cells crucial for hematopoiesis. We first confirmed that *GFI1B* was expressed in hematopoietic stem cells and progenitor cells for RBCs and megakaryocytes, consistent with *GFI1B*'s well-established role as a transcriptional repressor in early and lineage-specific progenitors (Fig. 5C) (55, 71–73). As expected, *GFI1B* was not expressed in granulocytes and lymphocytes (73, 74). Genes from cluster A were highly enriched for GFI1B-binding sites and had increased expression upon inhibiting *GFI1B*, suggesting that these genes are actively repressed in cells in which *GFI1B* is expressed (Fig. 5B). We next observed that genes from cluster A were highly expressed in granulocyte-monocyte progenitors and differentiated WBC types, including monocytes and dendritic cells (Fig. 5D and table S4D). For example, *CD33* is a well-known marker for myeloid cells that is commonly used to diagnose acute myeloid leukemia, and its expression increases upon inhibiting the *GFI1B* CRE (fig. S12) (75, 76). GFI1B directly binds the promoter of *CD33* (fig. S16A) and, upon inhibiting *GFI1B*, we found that *CD33* transcript and protein expression were both increased (fig. S16, B and C). *CD33* is expressed in myeloid progenitors and differentiated cells such as dendritic cells or monocytes (fig. S16D). Overall, cluster A is composed of genes that *GFI1B* directly represses, along with their downstream targets, to prevent differentiation of hematopoietic stem cells into WBCs.

Like cluster A, genes in cluster B were also enriched for GFI1B-binding sites and had increased expression upon inhibiting *GFI1B* (Fig. 5, A and B). However, genes in cluster B were not expressed in differentiated WBCs, but rather in a broad set of progenitor cell types (Fig. 5D), suggesting that these may be genes that are repressed in hematopoietic stem cells to maintain a multipotent cell state. Cluster C differed from clusters A and B in that it was not enriched for GFI1B-binding sites and had decreased expression upon inhibiting *GFI1B*. Genes in cluster C were expressed most highly in RBC progenitors, suggesting that these genes are secondary targets of *GFI1B* that act in a lineage-specific manner to differentiate hematopoietic stem cells into erythrocytes. These findings are supported by this cluster being enriched for RBC GWAS genes (Fig. 5B), and pathway analysis identifying these genes as part of the heme biosynthesis pathway (table S4E). The identification of these *trans*-regulatory networks in a homogeneous blood progenitor–like cell type (K562) demonstrates the utility of STING-seq in studying diverse effects of CREs on target genes.

Trade-offs between CRE effect sizes, number of cells, and sequencing depth in STING-seq

Given the large number of GWASs performed over the past 15 years, with numbers of trait-associated loci per GWAS ranging from tens to thousands (44), we wanted to understand the scale of cells needed to perform STING-seq under various settings. By performing statistical down-sampling experiments on the *cis*-regulatory effects identified by STING-seq, we computed the number of cells required for nominal significance (SCEPTRE $P < 10^{-3}$) for target genes with different expression levels, different CRE perturbation effect sizes, and different per-cell sequencing depths (fig. S17). For CREs with large effects, STING-seq requires as few as 100 cells and 5000 reads per cell, comparable to methods such as Perturb-seq and ECCITE-seq which target genes directly (47, 68, 77, 78). For CREs with moderate effects, STING-seq requires about 400 cells per gRNA or, if the cell number is fixed at 100 cells, 15,000 reads per cell. This down-sampling analysis provides a useful set of guidelines for estimating the resources required for applying STING-seq to other GWASs beyond blood traits.

Discussion

We have developed an approach for the characterization of functional effects of GWAS loci that takes noncoding human genetic variants and integrates fine-mapping, pooled CRISPR screens and single-cell RNA and protein sequencing to identify target genes in *cis* and *trans*. We have demonstrated the utility of STING-seq to identify target genes of CREs overlapping GWAS variants and described complex regulatory architectures of CREs. We found that 77% of blood trait GWAS loci have at least one fine-mapped variant overlapping an enhancer region and can be targeted with STING-seq. We identified target genes for 25% of tested cCREs and 36% of tested loci, a high yield over previous studies on the regulatory effects of noncoding genomic loci (7, 8). We also found that CRE activity is needed for CRISPRi-based target detection, and that spurious signals from inactive chromatin are rare. Additionally, we identified CREs with GWAS variants for TFs and miRNAs and, through their perturbation, identified *trans*-regulatory network clusters with distinct biological functions. The enrichment of genes in independent blood cell trait GWAS loci in these networks implies a polygenic contribution to the cellular functions that underlie diverse blood cell traits. We also identified target genes for non-European associations for which functional genomics data are typically sparse. For example, we nominated *ATP1A1* as a causal gene for neutrophil counts by targeting a locus identified exclusively in African ancestries. Targeting loci identified from ancestry-specific GWASs in cell models is ancestry agnostic if the GWAS variant maps to a candidate regulatory element and can lead to target gene identification.

We also performed direct variant insertion with beeSTING-seq, identifying noncoding GWAS variants with causal effects on target gene expression. Given the incomplete editing efficiencies [with many studies reporting ~30% (79)], the fact that the biological effect of individual GWAS variants is expected to be small, and that single-cell transcriptome data are sparse, it was not unexpected that we were only able to identify few loci, and future work is needed to further optimize base editors for studying the effects of GWAS variants. Targeted enrichment panels will have utility in improving the sparsity of single-cell sequencing; however, further innovation will be necessary to improve base-editing efficiency through directed evolution of existing base editors and the discovery of additional ones. However, the trade-off between higher yield from blunt perturbations such as CRISPRi versus highly precise base editing with smaller functional effects is likely to persist, and the ideal approach depends on the goals and design of each study.

A key feature of recent CRISPRi screens of cCREs (7, 8), including STING-seq, is the introduction of multiple perturbations per cell. This substantially increases the number of loci that can be feasibly analyzed. Although this is feasible for immortalized cell lines, expanding multiple perturbations (using either high MOI transduction or innovative vector designs) to other cell lines and primary cells will be instrumental for the next stage of target gene identification and characterization for diverse GWAS traits. However, caution is warranted in study designs in which a large proportion of gRNAs are likely to have *trans* effects, because their potential interactions may complicate interpretation of the data. In these cases, reducing the number of perturbations per cell may be necessary.

Our results demonstrate the power of single-cell sequencing for sensitive and scalable readout of regulatory effects of GWAS loci in *cis* and *trans*. Although we have a high yield in *cis* target gene discovery, identification of a *cis* gene alone with STING-seq does not prove its mechanistic causal role driving the GWAS association, nor does it exclude other potential causal variants, CREs, and genes, including in other cell types. Indeed, our observation of multiple CREs with highly correlated *cis* and *trans* effects but GWAS associations for different blood traits suggests that they might have distinct additional effects in other cellular contexts. In loci in which *cis* effects are coupled with *trans*-network effects, STING-seq can be highly informative of potential cellular mechanisms, which also provides strong support for the causal role of the *cis*-target gene. Given these network enrichments, we suggest that GWAS loci that putatively target TFs or miRNAs should be high-priority targets for STING-seq given the wealth of information that we could gain. Furthermore, integration of STING-seq with cellular phenotype screens (80–82) will be an invaluable next step to connecting genetic variants with cellular mechanisms driving GWAS associations.

The STING-seq workflow provides a roadmap to addressing V2F challenges and identifying target genes for GWAS loci in a high-throughput fashion, enabling a deeper understanding of human noncoding genome function and translation of these insights into new therapies.

Materials and Methods

UKBB genome-wide association studies of blood cell traits

UKBB data were used upon ethical approval from the Northwest Multi-Centre Research Ethics Committee, and informed consent was obtained from all participants before participation. We used GWAS summary statistics for 29 blood cell traits from 361,194 white British UKBB participants: WBC (leukocyte) count, RBC (erythrocyte) count, hemoglobin concentration, hematocrit percentage, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, RBC (erythrocyte) distribution width, platelet count, platelet crit, mean platelet (thrombocyte) volume, platelet distribution width, lymphocyte count, monocyte count, neutrophil count, eosinophil count, basophil count, lymphocyte percentage, monocyte percentage, neutrophil percentage, eosinophil percentage, basophil percentage, reticulocyte percentage, reticulocyte count, mean reticulocyte volume, mean sphered cell volume, immature reticulocyte fraction, high light scatter reticulocyte percentage, and high light scatter reticulocyte count (table S1A). Each GWAS was performed by fitting the following covariates to inverse normal transformed traits with linear regression models: Principal components 1 through 20, sex, age, age^2, sex and age interaction, and sex and age^2 interaction. The summary statistics were generated by the Neale Lab (www.nealelab.is/uk-biobank).

Statistical fine-mapping of UKBB blood cell traits

The 29 UKBB GWASs of blood cell traits were uniformly processed with a statistical fine-mapping pipeline. First, each GWAS was analyzed with GCTA-COJO v.1.93.1 (13, 14) to identify conditionally independent lead variants (COJO $P < 6.6 \times 10^{-9}$) and define 1-Mb regions for statistical fine-mapping. All variants within 500 kb of a lead variant were analyzed with FINEMAP v.1.3.1 (83), a Bayesian fine-mapping method that assigns each variant a Bayes factor for being plausibly causal. Both GCTA-COJO and FINEMAP require population-matched covariance matrices, so we generated these with PLINK v.2.0 (84), QCTOOL v.2.0.2, BGENIX v.1.1.5 (85), and LDstore v.1.1 (86) using a subset of 50,000 UKBB white UK participants (UKBB accession code 47976). FINEMAP allows for a maximum number of causal configurations to test for each input set of variants, so we set the maximum to 10 causal configuration variants per fine-mapped region and excluded cases for which FINEMAP failed to converge. We then retained noncoding variants with a high Bayes factor ($\log_{10}$ BF ≥ 2) and that were at least 1% likely to be causal for a set of causal variants. Fine-mapped variants that had more than one Bayes factor because they were within 500 kb of multiple lead variants had their highest value retained. Across all 29 GWASs, we identified 827 loci, separated by at least 500 kb, and 57,531 fine-mapped variants. The Variant Effect Predictor (VEP) tool (87) was used to identify 53,874 noncoding variants.

Fine-mapped BCX blood cell trait GWAS

The BCX generated GWAS summary statistics and fine-mapped 95% credible sets for 15 blood traits from 746,667 participants from five global populations (European ancestries, South Asian ancestries, Hispanic ancestries, East Asian ancestries, and African ancestries): RBC count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, MCH concentration, RBC distribution width, WBC count, neutrophils, monocytes, lymphocytes, basophils, eosinophils, platelet count, and mean platelet volume (11). Each GWAS was performed within each global population by fitting linear mixed models, adjusting for cohort-specific covariates, to generate population-specific GWAS summary statistics. Population-specific GWAS were fine-mapped using an approximate Bayesian approach (88) to construct 95% credible sets from all variants within 250 kb of a lead variant. The 95% credible sets were generated by ordering marginal variant posterior probabilities from highest to lowest and retaining variants until the probabilities summed 95%. Population-specific GWASs for each trait were then meta-analyzed using a multiancestry meta-analysis method (89) that also generates marginal variant posterior probabilities, from which multiancestry 95% credible sets were generated. We additionally required that variants were at least 1% likely to be causal. Across all 15 multiancestry meta-analyzed GWASs, we identified 1191 loci, separated by at least 500 kb, and 62,494 fine-mapped variants. VEP (87) was used to identify 58,573 noncoding variants.

Functional annotation of causal noncoding single-nucleotide polymorphisms

We integrated multiple functional genomics datasets for K562 cells. Specifically, we used DNase I–hypersensitive sites (DHS) from ENCODE (65), H3K27ac ChIP-seq peak calls from ENCODE, and ATAC-seq peak calls that we generated previously (81) to identify candidate cCREs. We used bedtools v.2.25.0 (90) and bedops v.2.4.3 (91) to identify variants mapping directly to cCREs. We also required variants to be farther than 1 kb from any gene TSS. We analyzed the UKBB and BCX GWAS variants separately. For UKBB GWASs, we identified 10,628 distinct variants mapping cCREs in 629 loci. We then selected 88 variants from 56 loci for targeting on the basis of whether a variant was targetable and more plausibly causal than others for a given GWAS and locus by ranking FINEMAP $\log_{10}$ Bayes factors and manual inspection of loci. For the 88 selected variants, 32 were the most probable variant for at least one GWAS locus, and 52 were in the top-10 most probable variants. For the 56 loci, there was a median of 10.5 (± 8.6) targetable single-nucleotide polymorphisms (SNPs). Elements of manual inspection included selecting variants that mapped to intergenic regions between gene TSSs or selecting multiple variants that map proximal to the same gene. For BCX GWASs, we identified 10,446 variants mapping to 886 loci. We selected 507 variants mapping to 265 loci for targeting, including 41 variants mapping to closed chromatin. Of the cCRE-mapping variants, we targeted 137 that were the sole variant within the 95% credible set and 239 variants that were composed of all targetable 95% credible set variants for 112 loci. The remaining 131 variants were selected because they were identified by GWASs from non-European ancestries and either fine-mapped in a population-specific GWAS or in the multiancestry meta-analysis. K562 DHS peaks and H3K27ac, RUNX1, IKZF1, and NFE2 ChIP-seq peaks are available from the ENCODE Project (www.encodeproject.org). K562 ATAC-seq peaks are available from the Gene Expression Omnibus (GEO) under accession number GSE161002; K562 GFI1B ChIP-seq peaks are available from GEO under accession number GSE117944.

Plasmid cloning for lentiviral CRISPRi, cytosine base editor, and modified gRNA scaffold vectors

To generate the KRAB-dCas9 (lentiCRISPRi (v1)-Blast) and KRAB-dCas9-MeCP2 [lentiCRISPRi(v2)-Blast] plasmids, KRAB and dCas9 were polymerase chain reaction (PCR) amplified from pCC_09 (Addgene 139094) (92), and the MeCP2 effector domain was synthesized as a gBlock (IDT). KRAB and MeCP2 were linked to dCas9 with flexible glycine-serine linkers and cloned into lentiCas9-Blast (Addgene 52962) (23). To generate the FNLS-BE3-SpRY (lentiBE3-SpRY-Blast) plasmid, we used Gibson cloning to replace the puromycin resistance gene in pLenti-FNLS-P2A-Puro (Addgene 110841) with blasticidin resistance from lentiCRISPRi(v2)-Blast. We then used Gibson cloning to replaced SpCas9(D10A) with the SpRY nickase from pCAG-CBE4max-SpRY-P2A-EGFP (Addgene 139999) (51). To generate the gRNA vector (lentiGuideFE-Puro), we digested pCC_09 with NheI and KpnI to isolate the U6 promoter and Cas9 guide RNA scaffold with the F+E scaffold modification (93). After gel extraction (Qiagen 28706), we ligated this piece into NheI- and KpnI-digested pLentiRNAGuide_001 (Addgene 138150) vector using T4 ligase (NEB M0202M) (94). Primer sequences for Gibson cloning reactions are available in table S1F.

Cell culture and monoclonal cell line generation

Human embryonic kidney (HEK) 293FT cells were acquired from Thermo Fisher Scientific (R70007). HEK293FT (human) cells were maintained at 37°C with 5% CO_2 in D10 medium: Dulbecco's modified Eagle's medium (DMEM) with high glucose and stabilized L-glutamine (Caisson DML23) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher 16000044). K562 cells were acquired from ATCC (CCL-243) and were maintained at 37°C with 5% CO_2 in R10 medium: RPMI with stabilized L-glutamine (Thermo Fisher 11875119) supplemented with 10% FBS (Thermo Fisher 16000044). Cells were regularly passaged and tested for the presence of mycoplasma contamination with the MycoAlert Plus Mycoplasma Detection Kit (Lonza).

Lentivirus was produced by polyethylenimine linear MW 25000 (Polysciences 23966) transfection of HEK293FT cells with the transfer plasmid containing a Cas9 effector, or gRNA library, packaging plasmid psPAX2 (Addgene 12260) and envelope plasmid pMD2.G (Addgene 12259). At 72 hours after transfection, cell medium containing lentiviral particles was harvested and filtered through 0.45-mm filter Steriflip-HV (Millipore SE1M003M00). K562 cells were transduced with lentiCRISPRi(v1)-Blast, lentiCRISPRi(v2)-Blast, or lentiBE3-SpRY–Blast at a low MOI (<1). Transduced K562 cells were selected with 10 μg/μl blasticidin (Thermo A1113903) for 10 days to enrich for expression of the Cas9 effector proteins. To isolate individual clones, K562 polyclonal lines were serially diluted to 50 cells per 10 ml of medium. We then plated 100 μl of this cell-medium mixture in 96-well round bottom plates (~0.5 cells/well).

Digital PCR for CRISPRi gene repression

We compared the single-repressor CRISPRi (KRAB-dCas9) and dual-repressor CRISPRi (KRAB-dCas9-MeCP2) systems by targeting the transcription start sites and known enhancers of three genes (*MRPS23*, *SLC25A37*, and *FSCN1*) with two gRNAs per targeted region. We synthesized gRNAs as top- and bottom-strand oligos (IDT) and cloned them into BsmBI-digested lentiGuideFE-Puro. We transduced the cells in biological triplicate with gRNA lentiviruses at a low MOI and after 24 hours selected for cells with gRNAs using puromycin (1 μg/μl, Thermo Fisher A1113803). We harvested the cells 10 days after transduction and extracted RNA using TRIzol (ThermoFisher 15596026). We quantified RNA concentration by spectrophotometry (NanoDrop). To measure gene expression, we performed digital PCR (Formulatrix Consellation) with Cy5/Iowa Black RQ target gene probes (IDT), FAM/ZEN/Iowa Black FQ for the actin normalizer (IDT), and Luna Universal One-Step RT qPCR Master Mix kit (NEB E3005L) and Tween-20 (Sigma-Aldrich P1379). We first normalized the target gene expression by actin expression per sample and then normalized this ratio to the ratio from cells transduced with nontargeting control gRNAs.

KRAB-dCas9-MeCP2 CRISPRi pooled screen for essential gene gRNA depletion

We performed CRISPRi pooled screens to quantify the KRAB-dCas9-MeCP2 inhibitory effect window in HCT116 and MCF7 cell lines. Both lines were acquired from ATCC (CCL-247 and and HTB-22, respectively) maintained in the appropriate medium (McCoy's 5A medium and DMEM, respectively) supplemented with 10% serum and 1% penicillin–streptomycin. These cell lines were cultured at 37°C, 5% CO_2, and ambient oxygen levels. Monoclonal HCT116 KRAB-dCas9-MeCP2 and MCF7 KRAB-dCas9-MeCP2 cell lines were generated as previously described for K562 cells. Expression was confirmed using Western blot.

For screening, HEK293 cells were plated in DMEM + 10% FBS (D10) in a 15-cm dish so that the following day cells were 90% confluent. Half of the medium was removed from the flask, and cells in each flask were transfected with 13.8 μg of a cCRE/TSS–targeting library specific to HCT116 and MCF7, 6.6 μg of pMD2.G (envelope plasmid), and 9.6 μg of psPAX2 (packaging plasmid) using 1.2 ml of Opti-MEM and 112.5 μl of polyethylenimine linear 25K (Polysciences 23966). The following morning, the medium was removed and fresh D10 + 1% bovine serum albumin (BSA) was added. Then, 48 hours later, we collected the viral supernatant and put it immediately on ice. We concentrated the supernatant by centrifugation at 100,000*g* (Thermo Sorvall LYNX) for 2 hours at 4°C. The resulting pellet was resuspended in cold DMEM and stored at –80°C until use.

We determined the appropriate titer of virus before the experimental transduction. We transduced 3M cells with a standard spinfection protocol with different dilutions of virus in a 12-well plate and in a no-virus control well. After adding virus, we spun the cells at 2000 rpm for 1 hour at 37°C (Beckman Coulter Allegra X-14R) and incubated overnight. The next day, we plated half of the cells in each well into two new wells of a six-well plate. In one set of wells, we added the appropriate puromycin concentration (1.5 μg/ml for HCT116 and 3 μg/ml for MCF7). After all the cells in the no-virus well had died, cells in the corresponding wells (with puromycin) were counted to determine the viral volume that results in 20 to 40% cell survival, corresponding to a MOI of 0.2 to 0.5.

We cultured each cell line in the appropriate medium and transduced 2 × 10^8 of them with the CRISPR lentiviral library using spinfection with the viral volume determined from the previous spinfection. As before, after adding virus, we spun cells at 2000 rpm for 1 hour at 37°C and incubated them overnight. The following day, cells were plated at 30% confluence and selected in the appropriate puromycin concentration for 3 days. After selection, we passaged cells in 15-cm dishes for 21 days and split at ~80% confluence. We isolated genomic DNA from cells using a modified salting-out precipitation. The gRNA readout was performed using two rounds of PCR. For PCR1, we used 10 μg of gDNA in each 100-μl reaction. We pooled the PCR1 products and used the mixture for a second PCR, adding on Illumina sequencing adaptors and barcodes. We performed PCR1 reactions using TaqB polymerase (Enzymatics P7250L) and PCR2 reactions with Q5 (NEB M0491). We pooled and purified PCR2 reactions with Illumina Purification Beads. We quantified the concentration of the gel-extracted PCR products using Qubit dsDNA HS Assay Kit (Thermo Fisher Q32851), and then diluted and sequenced it on NextSeq 500 high-output (Illumina). We demultiplexed the samples using bcl2fastq v2.20.0.422 (Illumina), trimmed off adapters, and aligned to our guides with bowtie v.1.1.2 (95). We library normalized the resulting reads (each read divided by the total number of reads). We then used the robust rank aggregation algorithm (96) and estimataed log$_2$ fold changes as log$_2$(day 21/day 1). We targeted ± 5 kb of the TSS essential genes (DepMap Chronos scores < –1) (97–100). In total, we screened 1992 gRNAs targeting 263 essential human genes. As negative controls, we embedded 1000 nontargeting gRNAs into this library.

Flow sorting for near PAM-less base editing

We verified cytosine base editing by designing 12 gRNAs targeting *CD46* splice sites using SpliceR v1.2.0 (101). SpliceR designed gRNAs that were predicted to disrupt *CD46* splice sites through C>T nucleotide changes. These included gRNAs that would recognize a diverse set of noncanonical PAMs, such as NGN, NAN, NCN, and NTN (table S3H). We also used four nontargeting gRNAs from the GeCKOv2 library (23) as negative controls. We synthesized gRNAs as top- and bottom-strand oligos (IDT) and cloned them into BsmBI-digested lentiGuideFE-Puro. We transduced the cells with gRNA lentiviruses at a low MOI in an arrayed fashion and after 24 hours selected for cells with gRNAs using puromycin (1 μg/μl, Thermo Fisher A1113803). After 6 days of selection, we proceeded to flow cytometry to measure CD46 protein. For flow cytometry, 1×10^6 cells per condition were harvested and washed with phosphate-buffered saline (PBS) after selection. The cells were stained for 5 min at room temperature with LIVE/DEAD Fixable Violet Dead Stain Kit (Thermo Fisher L34864). Subsequently, the cells were stained with antibodies for 20 min on ice with 1 μl of CD46-APC (clone TRA-2-10) (BioLegend 352405). Cells were washed with PBS to remove unbound antibodies before sorting. Cell acquisition and sorting was performed using a Sony SH800S cell sorter. Sequential gating was performed as follows: exclusion of debris on the basis of forward and side scatter cell parameters followed by dead cell exclusion. The sorting gates were set such that 90% of live K562 cells would be considered CD46 positive.

CRISPR inhibition and base-editing library design and cloning

Two individual CRISPR inhibition libraries were designed and cloned, called STING-seq v1 and STING-seq v2, and one base-editing library was designed and cloned, called beeSTING-seq. For STING-seq v1, we designed 20-nt gRNAs to target within 200 base pairs (bp) of the 88 selected plausibly causal noncoding variants from UKBB GWASs of blood traits. We used FlashFry v1.10.0 (22) to retain gRNAs with the lowest predicted off-target activity, as estimated by the Hsu-Scott score (21). Each variant was targeted by two different gRNAs. In addition, we included in our library 12 nontargeting gRNAs from the GeCKOv2 library (23) as negative controls and 12 gRNAs targeting the TSSs of six nonessential genes as positive controls. The six nonessential genes (*CD46*, *CD52*, *HSPA8*, *NMU*, *PPIA*, and *RPL22*) were identified by a CRISPR knock-out screen in K562 cells using the PICKLES database (102). We additionally included 10 gRNAs targeting the CD55 TSS for our fluorescence-activated cell sorting (FACS)–based MOI estimator, bringing the total number of gRNAs to 210. For STING-seq v2, we designed 20-nt gRNAs to target within 200 bp of the 507 selected plausibly causal variants from the BCX multiancestry and ancestry-specific blood trait GWASs. We again retained gRNAs with the lowest predicted off-target activity, and each variant was targeted by three different gRNAs. In addition, we included 30 nontargeting gRNAs from the GeCKOv2 library and 32 groups of three TSS-targeting gRNAs for positive controls. We additionally included 45 CD55 TSS-targeting gRNAs for FACS-based MOI estimation. For beeSTING-seq, we designed three sets of gRNAs for each of 46 C>T select GWAS variants mapping to CREs with *cis*-target genes. We followed recommended gRNA design instructions, and positioned the target nucleotide within a 5-nt window (103). We also included 28 nontargeting gRNAs from the GeCKOv2 library.

To clone the STING-seq v1 gRNA library, top- and bottom-strand oligos (IDT) were resuspended in water at 100 μM and then mixed at 1:1 ratio for each gRNA. Then, 1 μl of the oligo mix was added to a master mix containing 1× T4 ligase buffer (NEB M0202M) and 0.5 μl of T4 PNK (NEB M0201L) and water to a final concentration of 10 μl per reaction. For oligo annealing, we incubated the oligo mix at 37°C for 30 min, then 95°C for 5 min with a temperature change of 1°C every 5 s until reaching 4°C. To create the oligo pool, we pooled together 3 μl of each annealed oligo. The oligo pool was diluted 1:10 with water and then cloned in the lentiGuideFE-Puro, which was linearized with BsmBI (Thermo Fisher ER0451) and dephosphorylated. The ligation was performed in 11 reactions, with each reaction consisting of 5 μl of Rapid Ligation Buffer (Enzymatics B101), 0.5 μl of T7 ligase (Enzymatics L602L), digested plasmid at 25 ng per reaction, 1 μl of diluted oligo mix and double-distilled water to final volume of 10 μl. The ligation was performed at room temperature for 15 min.

Next, 100 μl of the combined ligation reactions were mixed with 100 μl of isopropanol, 1 μl of GlycoBlue (Thermo Fisher AM9515), and 2 μl of 5 M NaCl (50 mM final concentration), incubated for 15 min at room temperature, and spun at 12,000*g* for 15 min. The pellet was washed twice with prechilled 70% ethanol, air dried for 15 min or until dried completely, and resuspended in 5 μl of 1× TE buffer (Sigma). Next, 2 μl of library ligation was added to 50 μl of Endura cells (Lucigen) and then electroporated, recovered, and plated. The following day, bacterial colonies were scraped, plasmids were isolated using a maxi prep (Qiagen 12965), and library representation was determined by MiSeq (Illumina).

The STING-seq v2 and beeSTING-seq pooled gRNA libraries were synthesized as single-stranded oligonucleotide pools (Twist Biosciences) and diluted to 0.5 ng/μl in molecular-grade water. Then, 2 μl of the diluted pooled oligos were added to a master mix containing forward and reverse primer mixes (10 μM) and NEBNext High-Fidelity 2X PCR Master Mix (M0541S). We then PCR purified the product and Gibson cloned it in pLentiGuideFE-Puro, which was linearized as described above. We used 500 ng of the digested vector, maintained a 1:10 molar ratio of library, and incubated at 50°C for 1 hour. We concentrated DNA using isopropanol precipitation, washed and resuspended the DNA, and then transformed 1 μl of library in 25 μl of Endura cells (Endura 60242-2) according to protocol specifications. We then plated the transformed cells on Luria broth–ampicillin plates to get at least 100 to 500 colonies per gRNA.

The quality of all pooled libraries was verified by sequencing with a MiSeq (Illumina) to estimate the 90:10 quantile ratio. To generate and concentrate all pooled libraries, lentivirus was generated as described above. Briefly, we seeded 10 × 225 cm^2 flasks with HEK293FT cells and, at 70% confluency, we cotransfected the pooled gRNA library, psPAX2, and pMD2.G. Lentivirus was collected 72 hours after transfection and filtered using a 0.45-μm filter. The supernatant was then ultracentrifuged for 2 hours at 100,000*g* (Sorvall Lynx 6000), and the pellet was resuspended overnight at 4°C in PBS with 1% BSA.

MOI estimation using flow cytometry

When transducing cells at a high MOI, it is not possible to estimate the MOI by traditional methods (e.g., survival after drug selection) or without the time and cost of single-cell sequencing. By including multiple gRNAs that target the *CD55* TSS (10 gRNAs for STING-seq v1, 45 gRNAs for STING-seq v2), we were able to estimate the number of gRNAs per cell (MOI) using flow cytometry for CD55 cell surface protein knockdown over a range of viral transduction volumes. We performed two transductions for STING-seq v1 with concentrated lentivirus (4 and 6 μl) and, after 48 hours, we selected with puromycin for 10 days. We performed five transductions for STING-seq v2 with concentrated lentivirus (1, 5, 10, 20, and 30 μl) and, after 48 hours, we selected with puromycin for 10 days. We included three positive control transductions with different *CD55* TSS-targeting gRNAs and three negative control transductions with three different nontargeting gRNAs for both experiments. For beeSTING-seq, we performed five transductions with concentrated lentivirus (1, 5, 10, 25, and

50 μl), and, after 48 hours, we selected with puromycin for 10 days. We used the most viable cell culture for beeSTING-seq for sequencing (10 μl) with MACS dead cell removal kit (Miltenyi Biotec 130-090-101) because we observed high cell death at higher lentivirus concentrations.

For flow cytometry, 1 × 10^6 cells per condition were harvested and washed with PBS after selection. The cells were stained for 5 min at room temperature with LIVE/DEAD Fixable Violet Dead Stain Kit (Thermo Fisher L34864). Subsequently, the cells were stained with antibodies for 20 min on ice with 1 μl of CD55-FITC (clone JS11) (BioLegend 311306). Cells were washed with PBS to remove unbound antibodies before sorting. Cell acquisition and sorting was performed using a Sony SH800S cell sorter. Sequential gating was performed as follows: exclusion of debris on the basis of forward and side scatter cell parameters followed by dead cell exclusion. The sorting gates were set such that 90% of live K562 cells would be considered CD55 positive. From this estimation, we can estimate MOI using $X = 1 - N^Y$, where X is the proportion of cells with CD55 targeting gRNAs, N is the inverse of the number of CD55 targeting gRNAs divided by the total library size, leaving X as the predicted MOI. For the STING-seq v1 library, $N = 1 - (10/210)$, and for the STING-seq v2 library, $N = 1 - (45/1695)$. We estimated that 6 μl of STING-seq v1 viral volume yielded an MOI of ~13.5 and 30 μl of STING-seq v2 viral volume yielded an MOI of ~30, and elected to use these conditions for our STING-seq assay (fig. S2).

Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes

For the Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes sequencing (ECCITE-seq) and the STING-seq v1 experiment, we ran one lane of a 10× Genomics 5′ kit (Chromium Single Cell Immune Profiling Solution v1.0, 1000014, 1000020, and 1000151) with superloading and recovered 15,285 total cells (including multiple cells per droplet counts or "multiplets"). Cell hashing was performed as described in a previously published protocol using four hashtag-derived oligonucleotides (HTOs) using hyperconjugation (24). Gene expression (cDNA), hashtags (HTOs), and gRNA (guide-derived oligos, GDOs) libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, and GDO libraries with two NextSeq 500 high-output runs (Illumina). For the ECCITE-seq and the STING-seq v2 experiment, we ran four lanes of a 10x Genomics 5′ v2 kit (Chromium Next GEM Single Cell 5′ Kit v2 1000265) with superloading. We recovered 82,339 total cells (including multiplets). Cell hashing was performed using eight HTOs followed by staining with a 188 antibody-tagged oligonucleotides (ADTs) panel (BioLegend) (table S3B). The cDNA, HTO, ADT, and GDO libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, ADT, and GDO libraries with one NovaSeq 6000 S1 run and two NovaSeq 6000 S2 runs (Illumina). For the ECCITE-seq and beeSTING-seq experiment, we ran three lanes of a 10x Genomics 5′ v2 kit with superloading and recovered 39,049 total cells, including multiplets. Cell hashing was performed using nine HTOs. The cDNA, HTO, and GDO libraries were constructed by following 10x Genomics and ECCITE-seq protocols. We sequenced the cDNA, HTO, and GDO libraries with one NextSeq 500 mid-output run, one NovaSeq 6000 SP run, and one NovaSeq 6000 S1 run (Illumina).

Single-cell data processing

UMI count matrices were generated for all single-cell sequencing libraries with 10x Cell Ranger v.6.0.0 (104). We generated outputs using the Gene Expression Output, Antibody Capture Output, and CRISPR Guide Capture Output functions. We then analyzed the UMI count matrices in R v.4.0.2 with Seurat v.4.0.0 (105) and tested for differential gene expression and protein levels within the SCEPTRE framework (26). The distributions of cDNA, GDO, HTO, and ADT UMIs were inspected manually for each lane of 10× sequenced. Custom thresholds were set to remove outliers for total cDNA count, unique genes detected, mitochondrial percentage, total gRNA count, unique gRNAs detected, total HTO count, unique HTOs detected, total ADT count, and unique ADTs detected. Lanes were merged for STING-seq v2 and beeSTING-seq only after quality control was completed. For STING-seq v1, we processed cDNA UMI count matrices and retained cells between the 15th to 99th percentiles for unique gene count, between the 20th and 99th percentiles for total cDNA UMI count, and between the 5th and 90th percentile for mitochondrial percentage. Next, we center-log-ratio (CLR) transformed the HTO UMI counts and demultiplexed cells by their transformed HTO counts to identify singlets. We used the HTODemux function implemented in Seurat v.4.0.0 to maximize the number of singlets detected. We used then processed the GDO UMI count matrix, keeping cells between the 1st and 99th percentiles for total GDO count and used 10x Cell Ranger predicted GDO thresholds per cell, but required at least three UMIs per GDO to assign a GDO to a given cell. This resulted in a high-confidence set of 7667 single cells for the STING-seq v1 experiment. For STING-seq v2, we uniformly processed all four cDNA UMI count matrices and retained cells between the 1st and 99th percentile for unique gene count, between the 10th and 99th percentile for total cDNA UMI count, and between the 1st and 90th percentile for mitochondrial percentage. Next, we CLR transformed the HTO UMI counts and maximized singlet count using the HTODemux function. We then processed the GDO UMI count matrices, keeping cells between the 1st and 99th percentiles for total GDO count and again used the 10x Cell Ranger predicted GDO thresholds per cell, but required at least three UMIs per GDO. This resulted in a high-confidence set of 38,916 cells for differential expression testing. We further applied quality control filters for ADTs, retaining cells with between the 1st and 99th percentiles for total ADT count. This resulted in 38,133 cells for differential protein testing. For beeSTING-seq, we uniformly processed all three cDNA UMI count matrices and retained cells between the 10th and 90th percentiles for unique gene count, between the 10th and 90th percentiles for total cDNA count, and between the 10th and 90th percentiles for mitochondrial percentage. We then CLR transformed the HTO counts and used the HTODemux function to maximize singlets and retained cells between the 1st and 99th percentiles for total GDO counts. 10x Cell Ranger set most UMI thresholds to 1, so we generated a series of GDO UMI count matrices with thresholds from 1 to 5 to iteratively test optimal GDO thresholds for each gRNA. This resulted in a series of UMI count matrices for each GDO threshold. We had sets of 12,068 cells (GDO threshold = 1), 11,235 cells (GDO threshold = 2), 9739 (GDO threshold = 3), 7869 (GDO threshold = 4), and 5896 (GDO threshold = 5) for differential expression testing.

Differential gene expression– and protein-level testing with SCEPTRE

We used the processed UMI count matrices for gene expression or protein levels and gRNA expression, along with accompanying single-cell metadata to use as covariates in model fitting (table S3B). For STING-seq analyses, we defined for each cCRE targeted by two to three gRNAs a list of genes within 500 kb to be tested for differential expression in *cis*. For each gene per set of gRNAs, we extracted that gene's UMI counts and labeled the cells with the given gRNAs. We then tested for differential outcomes within the SCEPTRE framework (26), adjusting for the following single-cell covariates for expression tests: total gene expression UMIs, unique genes, total gRNA expression UMIs, unique gRNAs, percentage of mitochondrial genes, and 10x lane (for STING-seq v2 and beeSTING-seq). For protein tests, we adjusted for total ADT count, total HTO count, total gRNA expression UMIs, unique gRNAs, and ADTs for four mouse-specific antibody controls to represent nonspecific binding. We developed SCEPTRE as a statistical framework to analyze high-MOI CRISPR screens in single cells with state-of-the-art calibration. First, SCEPTRE fits a negative binomial distribution to measure the effect of a single gRNA on a given gene using the Z score. Then, the distribution of gRNAs to cells is randomly sampled to build a gRNA-specific null distribution, recomputing a negative binomial Z score. A skew-t distribution is fit to compare the test Z score and the null distribution, and a two-sided P value is derived, allowing for significance tests of increased or decreased gene expression or protein levels (26). To test for differential expression in *trans*, we defined for each set of gRNAs a list of all genes detected in at least 5% of cells and repeated the test above. Nontargeting gRNAs were tested against all genes used in the *cis* and *trans* settings discussed previously and randomly sampled to match the number of *cis* and *trans* tests displayed on QQ-plots. For each set of gRNAs with a *cis*-effect target gene we then performed marginal gRNA-gene pair testing, observing that the number of cells bearing gRNAs is the main driver behind statistical power (fig. S6). To determine significance for multiple hypotheses (genes) tested in *cis*, SCEPTRE P values were adjusted with the Benjamini-Hochberg procedure. For beeSTING-seq differential expression tests, we tested each gRNA against its known target gene from STING-seq analyses. We used for each gRNA the lowest GDO UMI threshold that resulted in at least 100 cells per gRNA and repeated this strategy for all nontargeting gRNAs against the same set of known *cis*-effect target genes. We then repeated differential expression testing, grouping together GWAS-CRE targeting gRNAs if they shared concordant effects and UMI thresholds and evaluating their combined effects on target gene expression.

To report significant results for STING-seq analyses, we identified *cis*-target genes if they were significant at a 5% FDR (Benjamini-Hochberg–adjusted SCEPTRE $P < 0.05$). We defined *trans*-target genes of each GWAS-CRE as those significant at a stricter 1% FDR. For beeSTING-seq analyses, we identified *cis*-target genes if they were significant at a 5% FDR. We examined all STING-seq genes significant at a 10% FDR and beeSTING-seq genes with SCEPTRE $P < 0.05$ to compare the *trans*-regulatory network effects from perturbing *HHEX* and *IKZF1* GWAS-CREs with direct variant insertion.

Fine-mapped eQTL credible set integration

We examined 31 fine-mapped eQTL studies from the eQTL Catalogue (31) specific to blood traits. Specifically, we used eQTLs identified from human macrophages (106, 107), monocytes (37, 108, 109), neutrophils (37), lymphoblastoid cell lines (110–113), whole blood (110, 114, 115), induced pluripotent stem cells (116–118), T cells (37, 109, 112), B cells (109), and natural killer cells (109). We then retained eQTL variants that were at least 1% plausibly causal and investigated whether our fine-mapped GWAS variants were in these data. eQTL summary statistics are available from the eQTL Catalog (www.ebi.ac.uk/eqtl).

K562 HiChIP for H3K27ac-interacting promoters

AQuA-HiChIP cell libraries were prepared as described previously (119). Briefly, NIH3T3 cells (mouse) and K562 cells were grown in the appropriate medium. Cells were fixed in 1% formaldehyde for 10 min and quenched to a final concentration of 125 nM glycine. Two million fixed mouse cells were mixed with 10 million fixed K562 cells. The cells were lysed in 0.5% SDS, quenched with 10% Triton X-100, and digested with MboI (NEB R0147M). The DNA overhangs were blunted, biotinylated (Thermo Fisher 19524016), and ligated. Nuclei were spun down, resuspended in nuclear lysis buffer, and sonicated using a Covaris LE220 with the following conditions: fill level 10, PIP 450, duty factor 30, and CPB 200. The sheared DNA was incubated with Dynabeads Protein A (Thermo Fisher 10001D) for 2 hours at 4°C. The tubes were placed on a magnet and the supernatant was kept. Immunoprecipitation was performed with a cross-species reactive H3K27ac antibody (Active Motif 39133). The samples were incubated with the antibody overnight at 4°C. The samples were then washed, eluted, and treated with Proteinase K. The samples were purified using Zymo DNA Clean & Concentrator. Biotin capture was performed with Dynabeads M-280 Streptavidin (Thermo Fisher 11205D), followed by library preparation. The amplified libraries were purified with Illumina Sample Purification Beads. The libraries were sequenced using paired-end reads with a NovaSeq 6000 S2 (Illumina) to generate 100 to 200 million read pairs per sample.

HiChIP paired end reads were mapped to the hg19 genome using HiC-Pro v.2.10.1 (120). Default settings were using to remove duplicate reads, identify valid interactions, and generate contact maps. Statistically significant contacts were identified using FitHiChIP v.9.1 (121) at a 5% FDR. H3K27ac ChIP-seq data (65) were used as a reference set of peaks in the FitHiChIP pipeline.

Trans-regulated network gene set enrichments

We used chromatin immunoprecipitation sequencing (ChIP-seq) datasets in K562 cells to identify *GFI1B* (64), *NFE2*, *IKZF1*, and *RUNX1* (65) TF-binding sites. There we no publicly available *HHEX* K562 ChIP-seq datasets. We assigned the closest protein-coding gene to each ChIP-seq peak with bedtools v2.25.0 (90). For predicted miRNA targets we used the TargetScan database (66, 67). To test for enrichment of ChIP-seq peak or TargetScan genes in *trans*-regulatory gene sets, we fit logistic regression models adjusting for K562 expression (gene expression counts from scRNA-seq data) and computed odds ratios with 95% confidence intervals. To construct GWAS-identified sets of genes, we used all fine-mapped SNPs from the 29 UKBB GWASs and 15 BCX GWASs previously described (categorized by cell type) with a high Bayes factor for being plausibly causal ($\log_{10}$ BF ≥ 2) and that were at least 1% plausibly causal. GWAS gene enrichment was performed in a similar fashion as for ChIP-seq peaks.

Gene coexpression analyses and bone marrow single-cell gene expression

To compute coexpression matrices for each *trans*-regulatory network, we used cDNA UMI count matrices with missing genes per cell imputed with the MAGIC algorithm (122). As a measure of coexpression, the biweight midcorrelation, a weighted correlation analysis, was calculated for each pair of genes (123). Genes were then clustered on the basis of their coexpression patterns by hierarchical clustering. TF-binding site, direct miRNA target, and GWAS gene enrichment was performed as described above. We used Human Cell Atlas single-cell RNA-sequencing from 35 bone marrow donors (69) and identified 27 cell types as described previously (70). Single-cell data were processed with Seurat v.4.0.0 to generate uniform manifold approximation and projection (UMAP) plots and heatmaps. To visualize entire *trans*-regulatory network clusters on a UMAP plot, we plotted the mean expression of all cluster genes within each cell.

STING-seq power estimations

We down-sampled 136 *cis* effects of gRNAs targeting CREs on their target genes across two key conditions for experimental design: sequencing read depth per cell and the number of cells per gRNA. We sequenced all STING-seq libraries to a depth of ~55,000 to 65,000 reads per cell and thus repeated the entire STING-seq quality control and differential expression testing pipeline with 5000, 15,000, 25,000, 35,000, 45,000 and 55,000. Sequencing reads were down-sampled to generate cDNA UMI count matrices with DropletUtils v.1.18.0 (124, 125) and repeated 10 times with different seed numbers. For each set of 10 randomly down-sampled UMI count matrices at each read depth, we repeated differential expression testing with SCEPTRE. We required at least 500 cells bearing each set of gRNAs, then at each set of 10 randomly down-sampled UMI count matrices at each read depth, we randomly down-sampled the number of cells bearing each set of gRNA from at least 500 cells to 50 and repeated this process 10 times at each stage. We averaged the SCEPTRE skew fit t test P values within replicates at each to compute precise measurements for each stage in the down-sampling procedures. We then divided all genes by their expression level and *cis* effects by their $\log_2$-fold changes into tertiles to examine at what number of cells and read depth could nominal significance (skew fit t test $P < 0.0001$) be attained.

REFERENCES AND NOTES

1. J. B. Wright, N. E. Sanjana, CRISPR screens to discover functional noncoding elements. *Trends Genet.* **32**, 526–529 (2016). doi: 10.1016/j.tig.2016.06.004; pmid: 27423542
2. T. Lappalainen, D. G. MacArthur, From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021). doi: 10.1126/science.abi8207; pmid: 34554789
3. J. C. Ulirsch *et al*., Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019). doi: 10.1038/s41588-019-0362-6; pmid: 30858613
4. J. A. Morris *et al*., An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019). doi: 10.1038/s41588-018-0302-x; pmid: 30598549
5. J. Nasser *et al*., Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021). doi: 10.1038/s41586-021-03446-x; pmid: 33828297
6. Q. Sun *et al*., From GWAS variant to function: A study of ∼148,000 variants for blood cell traits. *HGG Adv.* **3**, 100063 (2021). doi: 10.1016/j.xhgg.2021.100063; pmid: 35047852
7. M. Gasperini *et al*., A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390.e19 (2019). doi: 10.1016/j.cell.2018.11.029; pmid: 30612741
8. S. Xie, D. Armendariz, P. Zhou, J. Duan, G. C. Hon, Global analysis of enhancer targets reveals convergent enhancer-driven regulatory modules. *Cell Rep.* **29**, 2570–2578.e5 (2019). doi: 10.1016/j.celrep.2019.10.073; pmid: 31775028
9. F. Wünnemann *et al*., Multimodal CRISPR perturbations of GWAS loci associated with coronary artery disease in vascular endothelial cells. *PLOS Genet.* **19**, e1010680 (2023). doi: 10.1371/journal.pgen.1010680; pmid: 36928188
10. W. J. Astle *et al*., The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016). doi: 10.1016/j.cell.2016.10.042; pmid: 27863252
11. M.-H. Chen *et al*., Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020). doi: 10.1016/j.cell.2020.06.045; pmid: 32888493
12. D. Vuckovic *et al*., The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**, 1214–1231.e11 (2020). doi: 10.1016/j.cell.2020.08.008; pmid: 32888494
13. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: 10.1016/j.ajhg.2010.11.011; pmid: 21167468
14. J. Yang *et al*., Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375, S1–S3 (2012). doi: 10.1038/ng.2213; pmid: 22426310
15. V. G. Sankaran *et al*., Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes Dev.* **26**, 2075–2087 (2012). doi: 10.1101/gad.197020.112; pmid: 22929040
16. J. C. Ulirsch *et al*., Altered chromatin occupancy of master regulators underlies evolutionary divergence in the transcriptional landscape of erythroid differentiation. *PLOS Genet.* **10**, e1004890 (2014). doi: 10.1371/journal.pgen.1004890; pmid: 25521328
17. J. C. Ulirsch *et al*., Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell* **165**, 1530–1545 (2016). doi: 10.1016/j.cell.2016.04.048; pmid: 27259154
18. J. Wen *et al*., Super interactive promoters provide insight into cell type-specific regulatory networks in blood lineage cell types. *PLOS Genet.* **18**, e1009984 (2022). doi: 10.1371/journal.pgen.1009984; pmid: 35100265
19. N. C. Yeo *et al*., An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018). doi: 10.1038/s41592-018-0048-5; pmid: 30013045
20. L. A. Gilbert *et al*., Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014). doi: 10.1016/j.cell.2014.09.029; pmid: 25307932
21. P. D. Hsu *et al*., DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013). doi: 10.1038/nbt.2647; pmid: 23873081
22. A. McKenna, J. Shendure, FlashFry: A fast and flexible tool for large-scale CRISPR target design. *BMC Biol.* **16**, 74 (2018). doi: 10.1186/s12915-018-0545-0; pmid: 29976198
23. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014). doi: 10.1038/nmeth.3047; pmid: 25075903
24. M. Stoeckius *et al*., Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017). doi: 10.1038/nmeth.4380; pmid: 28759029
25. E. P. Mimitou *et al*., Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019). doi: 10.1038/s41592-019-0392-0; pmid: 31011186
26. T. Barry, X. Wang, J. A. Morris, K. Roeder, E. Katsevich, SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis. *Genome Biol.* **22**, 344 (2021). doi: 10.1186/s13059-021-02545-2; pmid: 34930414
27. R. Andersson *et al*., An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014). doi: 10.1038/nature12787; pmid: 24670763
28. E. B. Fauman, C. Hyde, An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. *BMC Bioinformatics* **23**, 169 (2022). doi: 10.1186/s12859-022-04706-x; pmid: 35527238
29. D. Yao *et al*., Multi-center integrated analysis of non-coding CRISPR screens. bioRxiv 520137 [Preprint] (2022); https://doi.org/10.1101/2022.12.21.520137.
30. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
31. N. Kerimov *et al*., A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.* **53**, 1290–1299 (2021). doi: 10.1038/s41588-021-00924-w; pmid: 34493866
32. B. Rowland *et al*., Transcriptome-wide association study in UK Biobank Europeans identifies associations with blood cell traits. *Hum. Mol. Genet.* **31**, 2333–2347 (2022). doi: 10.1093/hmg/ddac011; pmid: 35138379
33. B. Li *et al*., Tissue specificity-aware TWAS (TSA-TWAS) framework identifies novel associations with metabolic, immunologic, and virologic traits in HIV-positive adults. *PLOS Genet.* **17**, e1009464 (2021). doi: 10.1371/journal.pgen.1009464; pmid: 33901188
34. M. T. Maurano *et al*., Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015). doi: 10.1038/ng.3432; pmid: 26502339
35. S. Abramov *et al*., Landscape of allele-specific transcription factor binding in the human genome. *Nat. Commun.* **12**, 2751 (2021). doi: 10.1038/s41467-021-23007-0; pmid: 33980847
36. K. Blatt *et al*., Identification of campath-1 (CD52) as novel drug target in neoplastic stem cells in 5q-patients with MDS and AML. *Clin. Cancer Res.* **20**, 3589–3602 (2014). doi: 10.1158/1078-0432.CCR-13-2811; pmid: 24799522
37. L. Chen *et al*., Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016). doi: 10.1016/j.cell.2016.10.026; pmid: 27863251
38. A. R. Martin *et al*., Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019). doi: 10.1038/s41588-019-0379-x; pmid: 30926966
39. C.-Y. Chen *et al*., Analysis across Taiwan Biobank, Biobank Japan and UK Biobank identifies hundreds of novel loci for 36 quantitative traits. medRxiv 2021.04.12.21255236 [Preprint] (2021); https://doi.org/10.1101/2021.04.12.21255236.
40. W. Zhou *et al*., Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics* **2**, 100192 (2022). doi: 10.1016/j.xgen.2022.100192; pmid: 36777996
41. M. M. Shull, J. B. Lingrel, Multiple genes encode the human Na+,K+-ATPase catalytic subunit. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4039–4043 (1987). doi: 10.1073/pnas.84.12.4039; pmid: 3035563
42. N. Glorioso *et al*., Association of ATP1A1 and dear single-nucleotide polymorphism haplotypes with essential hypertension: Sex-specific and haplotype-specific effects. *Circ. Res.* **100**, 1522–1529 (2007). doi: 10.1161/01.RES.0000267716.96196.60; pmid: 17446437
43. P. Araos, S. Figueroa, C. A. Amador, The role of neutrophils in hypertension. *Int. J. Mol. Sci.* **21**, 8536 (2020). doi: 10.3390/ijms21228536; pmid: 33198361
44. A. Buniello *et al*., The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47** (D1), D1005–D1012 (2019). doi: 10.1093/nar/gky1120; pmid: 30445434
45. V. Leduc, S. Jasmin-Bélanger, J. Poirier, APOE and cholesterol homeostasis in Alzheimer's disease. *Trends Mol. Med.* **16**, 469–477 (2010). doi: 10.1016/j.molmed.2010.07.008; pmid: 20817608
46. S. Suchindran *et al*., Genome-wide association study of Lp-PLA(2) activity and mass in the Framingham Heart Study. *PLOS Genet.* **6**, e1000928 (2010). doi: 10.1371/journal.pgen.1000928; pmid: 20442857
47. J. M. Replogle *et al*., Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022). doi: 10.1016/j.cell.2022.05.013; pmid: 35688146
48. E. V. Fuior, A. V. Gafencu, Apolipoprotein C1: Its pleiotropic effects in lipid metabolism and beyond. *Int. J. Mol. Sci.* **20**, 5939 (2019). doi: 10.3390/ijms20235939; pmid: 31779116
49. A. Auton *et al*., A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: 10.1038/nature15393; pmid: 26432245
50. M. P. Zafra *et al*., Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018). doi: 10.1038/nbt.4194; pmid: 29969439
51. R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020). doi: 10.1126/science.aba8853; pmid: 32217751
52. R. E. Hanna *et al*., Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021). doi: 10.1016/j.cell.2021.01.012; pmid: 33606977
53. R. Cuella-Martin *et al*., Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081–1097.e19 (2021). doi: 10.1016/j.cell.2021.01.041; pmid: 33606978
54. N. Mancuso *et al*., Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017). doi: 10.1016/j.ajhg.2017.01.031; pmid: 28238358
55. S. H. Orkin, L. I. Zon, Hematopoiesis: An evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008). doi: 10.1016/j.cell.2008.01.025; pmid: 18295580
56. J. J. Gasiorek, V. Blank, Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. *Cell. Mol. Life Sci.* **72**, 2323–2335 (2015). doi: 10.1007/s00018-015-1866-6; pmid: 25721735
57. K. L. Davis, Ikaros: Master of hematopoiesis, agent of leukemia. *Ther. Adv. Hematol.* **2**, 359–368 (2011). doi: 10.1177/2040620711412419; pmid: 23556102
58. R. J. Chan, R. Hromas, M. C. Yoder, The role of Hex in hemangioblast and hematopoietic development. *Methods Mol. Biol.* **330**, 123–134 (2006). doi: 10.1385/1-59745-036-7:123; pmid: 16846021
59. M. Ichikawa *et al*., A role for RUNX1 in hematopoiesis and myeloid leukemia. *Int. J. Hematol.* **97**, 726–734 (2013). doi: 10.1007/s12185-013-1347-3; pmid: 23613270
60. E. Chapnik *et al*., miR-142 orchestrates a network of actin cytoskeleton regulators during megakaryopoiesis. *eLife* **3**, e01964 (2014). doi: 10.7554/eLife.01964; pmid: 24859754
61. T. Wang, F. Wu, D. Yu, miR-144/451 in hematopoiesis and beyond. *ExRNA* **1**, 16 (2019). doi: 10.1186/s41544-019-0035-8
62. M. Ghanbari *et al*., A functional variant in the miR-142 promoter modulating its expression and conferring risk of Alzheimer disease. *Hum. Mutat.* **40**, 2131–2145 (2019). doi: 10.1002/humu.23872; pmid: 31322790
63. C. Bellenguez *et al*., New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022). doi: 10.1038/s41588-022-01024-z; pmid: 35379992
64. P. Shooshtarizadeh *et al*., Gfi1b regulates the level of Wnt/β-catenin signaling in hematopoietic stem cells and megakaryocytes. *Nat. Commun.* **10**, 1270 (2019). doi: 10.1038/s41467-019-09273-z; pmid: 30894540

65. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
66. V. Agarwal, G. W. Bell, J.-W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015). doi: 10.7554/eLife.05005; pmid: 26267216
67. S. E. McGeary *et al.*, The biochemical basis of microRNA targeting efficacy. *Science* **366**, eaav1741 (2019). doi: 10.1126/science.aav1741; pmid: 31806698
68. G. R. Schnitzler *et al.*, Mapping the convergence of genes for coronary artery disease onto endothelial cell programs. bioRxiv 514606 [Preprint] (2022); https://doi.org/10.1101/2022.11.01.514606.
69. A. Regev *et al.*, Science Forum: The Human Cell Atlas. *eLife* **6**, e27041 (2017). doi: 10.7554/eLife.27041; pmid: 29206104
70. T. Stuart *et al.*, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
71. T. Möröy, L. Vassen, B. Wilkes, C. Khandanpour, From cytopenia to leukemia: The role of Gfi1 and Gfi1b in blood formation. *Blood* **126**, 2561–2569 (2015). doi: 10.1182/blood-2015-06-655043; pmid: 26447191
72. E. Anguita, F. J. Candel, A. Chaparro, J. J. Roldán-Etcheverry, Transcription factor GFI1B in health and disease. *Front. Oncol.* **7**, 54 (2017). doi: 10.3389/fonc.2017.00054; pmid: 28401061
73. H. Beauchemin, T. Möröy, Multifaceted actions of GFI1 and GFI1B in hematopoietic stem cell self-renewal and lineage commitment. *Front. Genet.* **11**, 591099 (2020). doi: 10.3389/fgene.2020.591099; pmid: 33193732
74. M. Osawa *et al.*, Erythroid expansion mediated by the Gfi-1B zinc finger protein: Role in normal hematopoiesis. *Blood* **100**, 2769–2777 (2002). doi: 10.1182/blood-2002-01-0182; pmid: 12351384
75. F. Garnache-Ottou *et al.*, Expression of the myeloid-associated marker CD33 is not an exclusive factor for leukemic plasmacytoid dendritic cells. *Blood* **105**, 1256–1264 (2005). doi: 10.1182/blood-2004-06-2416; pmid: 15388576
76. M. S. De Propris *et al.*, High CD33 expression levels in acute myeloid leukemia cells carrying the nucleophosmin (NPM1) mutation. *Haematologica* **96**, 1548–1551 (2011). doi: 10.3324/haematol.2011.043786; pmid: 21791474
77. A. Dixit *et al.*, Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016). doi: 10.1016/j.cell.2016.11.038; pmid: 27984732
78. B. Adamson *et al.*, A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016). doi: 10.1016/j.cell.2016.11.048; pmid: 27984733
79. S.-Y. Yu *et al.*, Increasing the targeting scope of CRISPR base editing system beyond NGG. *CRISPR J.* **5**, 187–202 (2022). doi: 10.1089/crispr.2021.0109; pmid: 35238621
80. Z. Daniloski *et al.*, Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell* **184**, 92–105.e16 (2021). doi: 10.1016/j.cell.2020.10.030; pmid: 33147445
81. N. Liscovitch-Brauer *et al.*, Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021). doi: 10.1038/s41587-021-00902-x; pmid: 33927415
82. M. Legut *et al.*, A genome-scale screen for synthetic drivers of T cell proliferation. *Nature* **603**, 728–735 (2022). doi: 10.1038/s41586-022-04494-7; pmid: 35296855
83. C. Benner *et al.*, FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016). doi: 10.1093/bioinformatics/btw018; pmid: 26773131
84. S. Purcell *et al.*, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi: 10.1086/519795; pmid: 17701901
85. G. Band, J. Marchini, BGEN: a binary file format for imputed genotype and haplotype data. bioRxiv 308296 (2018).
86. C. Benner *et al.*, Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017). doi: 10.1016/j.ajhg.2017.08.012; pmid: 28942963
87. W. McLaren *et al.*, The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016). doi: 10.1186/s13059-016-0974-4; pmid: 27268795
88. J. B. Maller *et al.*, Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012). doi: 10.1038/ng.2435; pmid: 23104008
89. R. Mägi *et al.*, Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017). doi: 10.1093/hmg/ddx280; pmid: 28911207
90. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). doi: 10.1093/bioinformatics/btq033; pmid: 20110278
91. S. Neph *et al.*, BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012). doi: 10.1093/bioinformatics/bts277; pmid: 22576172
92. M. Legut *et al.*, High-throughput screens of PAM-flexible Cas9 variants for gene knockout and transcriptional modulation. *Cell Rep.* **30**, 2859–2868.e5 (2020). doi: 10.1016/j.celrep.2020.02.010; pmid: 32130891
93. B. Chen *et al.*, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013). doi: 10.1016/j.cell.2013.12.001; pmid: 24360272
94. H.-H. Wessels *et al.*, Massively parallel Cas13 screens reveal principles for guide RNA design. *Nat. Biotechnol.* **38**, 722–727 (2020). doi: 10.1038/s41587-020-0456-9; pmid: 32518401
95. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
96. R. Kolde, S. Laur, P. Adler, J. Vilo, Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 (2012). doi: 10.1093/bioinformatics/btr709; pmid: 22247279
97. R. M. Meyers *et al.*, Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017). doi: 10.1038/ng.3984; pmid: 29083409
98. J. M. Dempster *et al.*, Extracting biological insights from the Project Achilles genome-scale CRISPR screens in cancer cell lines. bioRxiv 720243 [Preprint] (2019); https://doi.org/10.1101/720243.
99. J. M. Dempster *et al.*, Chronos: a CRISPR cell population dynamics model. bioRxiv 432728 [Preprint] (2021); https://doi.org/10.1101/2021.02.25.432728.
100. C. Pacini *et al.*, Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021). doi: 10.1038/s41467-021-21898-7; pmid: 33712601
101. M. G. Kluesner *et al.*, CRISPR-Cas9 cytidine and adenosine base editing of splice-sites mediates highly-efficient disruption of proteins in primary and immortalized cells. *Nat. Commun.* **12**, 2437 (2021). doi: 10.1038/s41467-021-22009-2; pmid: 33893286
102. W. F. Lenoir, T. L. Lim, T. Hart, PICKLES: The database of pooled in-vitro CRISPR knockout library essentiality screens. *Nucleic Acids Res.* **46** (D1), D776–D780 (2018). doi: 10.1093/nar/gkx993; pmid: 29077937
103. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi: 10.1038/nature17946; pmid: 27096365
104. G. X. Y. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: 10.1038/ncomms14049; pmid: 28091601
105. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021). doi: 10.1016/j.cell.2021.04.048; pmid: 34062119
106. K. Alasoo *et al.*, Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018). doi: 10.1038/s41588-018-0046-7; pmid: 29379200
107. Y. Nédélec *et al.*, Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669.e21 (2016). doi: 10.1016/j.cell.2016.09.025; pmid: 27768889
108. H. Quach *et al.*, Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016). doi: 10.1016/j.cell.2016.09.024; pmid: 27768888
109. B. J. Schmiedel *et al.*, Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018). doi: 10.1016/j.cell.2018.10.022; pmid: 30449622
110. A. Buil *et al.*, Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015). doi: 10.1038/ng.3162; pmid: 25436857
111. T. Lappalainen *et al.*, Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013). doi: 10.1038/nature12531; pmid: 24037378
112. M. Gutierrez-Arcelus *et al.*, Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**, e00523 (2013). doi: 10.7554/eLife.00523; pmid: 23755361
113. E. Theusch, Y. I. Chen, J. I. Rotter, R. M. Krauss, M. W. Medina, Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics* **21**, 555 (2020). doi: 10.1186/s12864-020-06966-4; pmid: 32787775
114. F. Aguet *et al.*, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: 10.1126/science.aaz1776; pmid: 32913098
115. K. Lepik *et al.*, C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLOS Comput. Biol.* **13**, e1005766 (2017). doi: 10.1371/journal.pcbi.1005766; pmid: 28922377
116. H. Kilpinen *et al.*, Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017). doi: 10.1038/nature22403; pmid: 28489815
117. A. D. Panopoulos *et al.*, iPSCORE: A resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Reports* **8**, 1086–1100 (2017). doi: 10.1016/j.stemcr.2017.03.012; pmid: 28410642
118. E. E. Pashos *et al.*, Large, diverse population cohorts of hiPSCs and derived hepatocyte-like cells reveal functional genetic variation at blood lipid-associated loci. *Cell Stem Cell* **20**, 558–570.e10 (2017). doi: 10.1016/j.stem.2017.03.017; pmid: 28388432
119. B. E. Gryder, J. Khan, B. Z. Stanton, Measurement of differential chromatin interactions with absolute quantification of architecture (AQuA-HiChIP). *Nat. Protoc.* **15**, 1209–1236 (2020). doi: 10.1038/s41596-019-0285-9; pmid: 32051612
120. N. Servant *et al.*, HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015). doi: 10.1186/s13059-015-0831-x; pmid: 26619908
121. S. Bhattacharyya, V. Chandra, P. Vijayanand, F. Ay, Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019). doi: 10.1038/s41467-019-11950-y; pmid: 31530818
122. D. van Dijk *et al.*, Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018). doi: 10.1016/j.cell.2018.05.061; pmid: 29961576
123. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: 10.1186/1471-2105-9-559; pmid: 19114008
124. J. A. Griffiths, A. C. Richard, K. Bach, A. T. L. Lun, J. C. Marioni, Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9**, 2667 (2018). doi: 10.1038/s41467-018-05083-x; pmid: 29991676
125. A. T. L. Lun *et al.*, EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019). doi: 10.1186/s13059-019-1662-y; pmid: 30902100

ACKNOWLEDGMENTS

We thank the staff of the Sanjana and Lappalainen laboratories for support and advice; the New York Genome Center for flow cytometry resources; E. Hoelzli for support with digital PCR; T. Stuart and H. Wessels for advice on single-cell sequencing quality control; and S. Kasela for help with library design. This research was conducted using the UKBB Resource under application no. 47976. **Funding:** J.A.M. is supported by a Canadian Institutes of Health Research Banting Postdoctoral Fellowship and the National Institutes of Health (NIH)/National Human Genome Research Institute (NHGRI) (grant K99HG012792). Z.D. is supported by an American Heart Association Postdoctoral Fellowship (20POST35220040). J.D. is supported by a European Molecular Biology Organization Postdoctoral Fellowship (ALTF 345-2021). K.R. is supported by the NIH/National Institute of Mental Health (NIMH) (grant R01MH123184). E.K. is supported by the National Science Foundation (NSF) (grant DMS-2113072) and the Wharton

Data Science and Business Analytics Fund. T.L. is supported by the NIH/NIMH (grant R01MH106842), the NIH/NHGRI (grant UM1HG008901), and the NIH/National Institute of General Medical Sciences (NIGMS) (grant R01GM122924). N.E.S. is supported by the NIH/NHGRI (grant DP2HG010099, including a supplemental award for STING-seq), the NIH/National Cancer Institute (NCI) (grants R01CA279135 and R01CA218668), the NIH/National Institute of Allergy and Infectious Diseases (NIAID) (grant R01AI176601), the Simons Foundation for Autism Research, the MacMillan Center for the Study of the Non-Coding Cancer Genome, New York University, and New York Genome Center funds. **Author contributions:** J.A.M. and N.E.S. conceived the project. J.A.M., D.A.G., and N.E.S. performed GWAS statistical fine-mapping, functional annotation, and STING-seq library design. J.A.M., Z.D., M.Z., L.L., and K.D. developed the lentiviral constructs and monoclonal K562 cells. J.A.M., Z.D., and L.L. designed and performed lentiviral transductions. C.C. performed the essential gene pooled screens and physical contact mapping experiments. J.A.M., S.H., and E.P.M. performed single-cell sequencing. J.A.M., T.B., and E.K. performed differential gene expression– and protein-level testing. J.A.M. and J.D. performed coexpression analyses and human bone marrow single-cell data analyses. P.S., K.R., E.K., T.L., and N.E.S. supervised the work. J.A.M., T.L., and N.E.S. wrote the manuscript with input from all authors. **Competing interests:** N.E.S. is an adviser to Qiagen and is a cofounder and adviser of OverT Bio. T.L. is an adviser to Goldfinch Bio and GSK and, with equity, Variant Bio. The remaining authors declare no competing interests. **Data and materials availability:** Plasmids (lentiCRISPRi(v1)-Blast, lentiCRISPRi(v2)-Blast, lentiGuideFE-Puro, and lentiBE3-SpRY-Blast) have been deposited with Addgene (plasmid nos. 170067, 170068, 170069, 199303). Raw and processed single-cell sequencing files are available from Gene Expression Omnibus (GEO) under accession number GSE171452. SCEPTRE and related tutorials are available on Github (www.github.com/Katsevich-Lab/sceptre).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh7699
Figs. S1 to S17
Tables S1 to S4
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 13 March 2023; accepted 20 April 2023
Published online 4 May 2023
10.1126/science.adh7699

Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens

John A. Morris, Christina Caragine, Zharko Daniloski, Jlia Domingo, Timothy Barry, Lu Lu, Kyrie Davis, Marcello Ziosi, Dafni A. Glinos, Stephanie Hao, Eleni P. Mimitou, Peter Smibert, Kathryn Roeder, Eugene Katsevich, Tuuli Lappalainen, and Neville E. Sanjana

Science, **380** (6646), eadh7699.
DOI: 10.1126/science.adh7699

Editor's summary

Genome-wide association studies (GWASs) identify links between individual gene variants and various traits and diseases. Unfortunately, the findings from these studies cannot be used to determine whether the gene variants associated with a disease directly cause the condition or just happen to be located near biologically relevant genes or regulatory regions. Most of the variants identified through GWASs are located in noncoding regions of the genome, further increasing the difficulty of interpretation. A workflow developed by Morris *et al*. addresses this problem by using CRISPR-based editing to directly introduce variants of interest and then assessing their effects on gene expression in individual cells, thereby identifying their contributions to specific bood cell traits. —Yevgeniya Nusinovich

View the article online
https://www.science.org/doi/10.1126/science.adh7699
Permissions
https://www.science.org/help/reprints-and-permissions

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.