Title: CRISPRi perturbation screens and eQTLs provide complementary and distinct insights into GWAS target genes

Authors:

Samuel Ghatan[1], Jasper Panten[2], Winona Oliveros[1,3], Neville E Sanjana[1,4], John A Morris[1,5], Tuuli Lappalainen[1,2*]

Affiliations:

1. New York Genome Center, New York, USA
2. Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Solna, Sweden
3. Department of Life Sciences, Barcelona Supercomputing Center (BCN-CNS), Barcelona, Catalonia, 08034, Spain
4. Department of Biology, New York University, New York, NY, USA
5. Donnelly Centre for Cellular and Biomolecular Research, Department of Molecular Genetics, University of Toronto, Toronto, Canada

*Corresponding author. Email: tlappalainen@nygenome.org

Abstract

Most genetic variants associated with human traits and diseases lie in noncoding regions of the genome [1], and a key challenge is determining which genes they affect [2,3]. A common approach has been to leverage associations between natural genetic variation and gene expression to identify expression quantitative trait loci (eQTLs) in the population [4,5]. At the same time, a novel approach uses pooled CRISPR interference (CRISPRi) perturbations of noncoding loci with single-cell transcriptome sequencing [6,7]. Here, we systematically harmonized and compared the results from these approaches across hundreds of genomic regions associated with blood cell

traits. We find that while the two approaches sometimes identify the same target genes, there are considerable differences that affect biological inferences made from the data. CRISPRi preferentially maps highly proximal, constraint-enriched genes, whereas eQTLs recover multiple, often distal targets. By benchmarking against 1,075 gold-standard CRE–gene pairs linked to blood traits, we show that the two approaches identify largely distinct targets; when combined, they achieve a balance between accuracy and completeness of gene discovery. Our results offer guidance for improved design of CRISPRi and eQTL studies and highlight their joint potential as a powerful toolkit for interpreting disease-associated loci.

Main text:

Introduction

Genome-wide association studies (GWAS) have discovered hundreds of thousands of loci that associate with complex traits and diseases, but elucidating the functional mechanism of these variants remains challenging [2,3]. Notably, ~90% of GWAS-implicated variants lie in noncoding regions [1], making identifying causal genes difficult. Expression quantitative trait loci (eQTL) studies were the first genome-wide approach for associating variants with gene regulation. Nevertheless, despite their popularity, an emerging body of work highlights the limitations of relying exclusively on eQTL data to link GWAS variants to their putative target genes [8,9]. Firstly, we still lack eQTL associations for many GWAS variants [4,5]. When an eQTL does exist for a GWAS variant, it often finds genes unlikely to be causal to the trait [4,8], and its reliability in identifying the true causal gene is relatively poor compared to simple metrics such as assigning the closest gene as the causal gene [10,11]. Recently, single-cell pooled CRISPR interference (CRISPRi) screens have emerged as a potentially powerful strategy for pinpointing causal genes at hundreds of GWAS loci [6,7]. However, CRISPRi inhibits entire enhancers [12], producing effects that may not be analogous to single-nucleotide variation, and the studies are typically conducted in non-primary human cell lines, raising questions about the transferability of

biological effects into primary cells and disease associations. The sensitivity or specificity of the CRISPRi approach for identifying likely causal GWAS genes has not been systematically evaluated, and the extent to which eQTL and CRISPRi approaches converge or diverge in identifying GWAS gene targets has been unknown.

GWAS target gene discovery

We analyzed and fine-mapped GWAS loci from 44 blood trait studies from the UK Biobank and the Blood Cell Consortium (BCX) (data S1). To identify *cis*-eQTL target genes (eGenes) for GWAS variants, colocalization was performed with fine-mapped eQTL variants from 51 studies based on RNA-sequencing of whole blood, bulk RNA-seq of isolated blood cell types, and single cell RNA-sequencing (scRNA-seq) of peripheral blood mononuclear cells (PBMCs) (data S2) [5,13–15]. To identify CRISPRi cis target genes (cGenes), we utilized and/or re-analyzed 12 CRISPRi studies targeting enhancers in K562 erythroleukemia cells (data S3) [6,7,16] and identified fine-mapped GWAS variants intersecting gRNA loci (Fig. 1a, mean PIP = 0.20). Overall, we identified 882 GWAS loci in cis-regulatory elements (GWAS CREs) that were targeted by a CRISPRi gRNA. Of these, 145 CREs (16%) had a cGene (FDR < 0.10, fig. S1a), and 578 (66%) had a colocalized eGene (PP.H4 > 0.5, fig. S1b) within 1 Mb (Supplementary Tables 4 & 5). In 104 CREs, both eGenes (n = 321) and cGenes (n = 130) were identified, with 81 overlapping CRE–gene target pairs shared between the two approaches (Fig. 1b).

CRISPRi and eQTL power shape the discovery of GWAS target genes

We first investigated how these discoveries were affected by various gene, CRE, and study features related to statistical power by analyzing patterns in the empirical data, complemented by simulations of both eQTL and CRISPRi discovery in each locus (Methods, Supplementary Tables 6 & 7, Extended Data Fig 2). . Only 3 of the overlapping genes were uniquely identified in scRNA-seq eQTLs, pointing to limited power in current single-cell eQTL data in identifying

77 new GWAS target genes, particularly for less abundant cell types (Fig. 1c, fig. S1c). Based on
78 our simulations, the eQTL studies had a high mean power of 0.90 to detect associations with
79 moderate effects of $\beta = 0.25$, but much larger samples would be needed to detect smaller
80 effects (fig. S3c). While we observed additional GWAS colocalizations with eQTL sample size,
81 this increase scaled sublinearly with sample size: INTERVAL is 8.3× larger than GTEx yet has
82 only 2.7× more cis-eQTLs (56,959 vs 20,818) and 2.5× more colocalized loci (10,662 vs 4,292)
83 (fig. S3d). Even though we analyzed only colocalized eQTLs (PP.H4 > 0.5), GWAS variants with
84 eGenes that intersected cGenes had a higher colocalization probability (PP.H4) ($P = 7.8 \times 10^{-10}$,
85 Fig. 1d), even after controlling for eQTL and GWAS P-values. These results suggest that when
86 cell type isolation is technically feasible, bulk cell type eQTL studies hold advantages over sc-
87 eQTL and bulk tissue studies in providing both good statistical power and cell type specificity,
88 and that stricter criteria for eQTL colocalization may yield more reliable target inference.
89
90 Some of the differences between eGene and cGene discovery are simply due to cell type
91 specific gene expression: in the 104 CREs with both cGenes and eGenes, 98% of cGenes were
92 expressed in eQTL datasets, whereas only 76% of eGenes were expressed in K562 cells,
93 underscoring the limitations of inferring causal genes from a single cell line. Correlation in rank
94 concordance of trait associations was moderate (Spearman $\rho = 0.58$, $P = 0.001$; Fig. S4).
95 Consistent with the erythroid lineage of K562, CRISPRi mapped genes for red-blood-cell GWAS
96 loci more effectively (within-trait rank test, Δrank = rank_CRISPRi − rank_eQTL; median = −2.0;
97 one-sided Wilcoxon $P = 0.04$), whereas eQTL studies tended to capture more genes for white-

98 blood-cell traits (median = 5.0; P = 0.07). This highlights the importance of matching studied cell

99 types to the studied GWAS traits.

100

101 cGenes were identified for GWAS variants at lower minor allele frequencies (MAFs) than

102 eGenes (P = 0.04, Fig. 1e), a pattern recapitulated in simulations (fig. S3a). This is expected, as

103 eQTL discovery power, similarly to other variant association studies, declines for low-frequency

104 variants. In contrast, CRISPRi-based mapping is unaffected by allele frequency. Conversely,

105 eGenes have significantly lower expression than cGenes in K562 cells ($P = 1.9. \times 10^{-3}$, Fig. 1f)

106 and GTEx whole blood ($P = 2.2 \times 10^{-8}$, Fig. 1g), with simulations further indicating that CRISPRi

107 discovery power is more affected by lower expression than eQTLs, likely deriving from sparsity

108 of scRNA-seq data (fig. S3b). Altogether, eGene discovery power was similar for genes that

109 intersect cGenes and for those that do not (P = 0.43, Fig. 1h), suggesting that the eQTL

110 analysis is not lacking power to capture these target genes. The opposite was true for cGene

111 discovery power, which was significantly higher for intersecting target genes compared to non-

112 intersecting ones ($P = 3.8 \times 10^{-5}$, Fig. 1i), indicating that higher powered CRISPRi data would

113 result in better alignment with eQTL results.

114

115

116

117

118

119

120

121

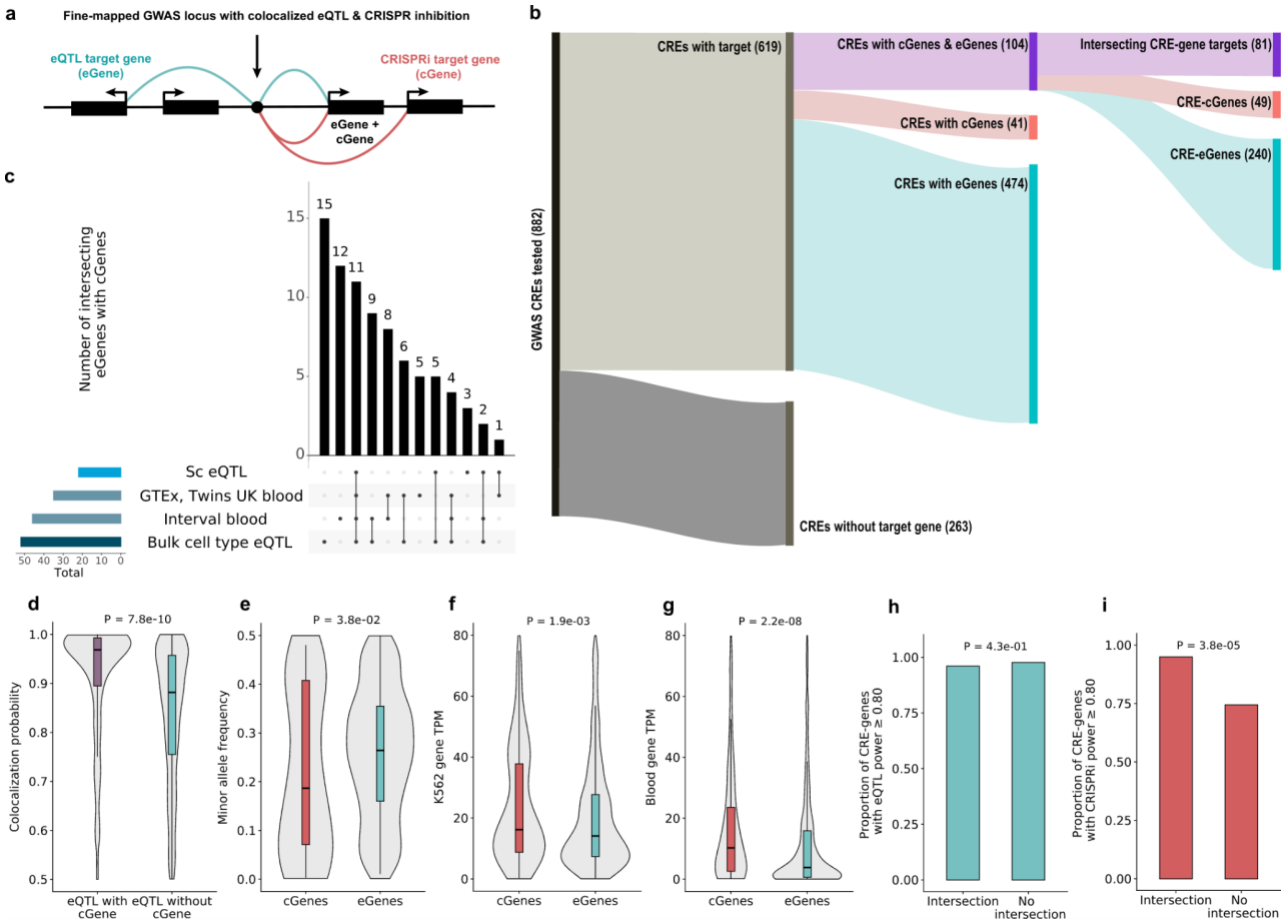


***Figure 1. Discovery of eQTL and CRISPRi target genes for GWAS loci. a**, Illustration of the comparison of gene targets for GWAS variants identified through eQTL and CRISPRi approaches. **b**, The number of CREs and genes discovered by eQTL and CRISPRi. **c**, The number of eGenes that intersect with cGenes across different eQTL study types. **d**, Comparison of GWAS–eQTL colocalization probabilities between eQTLs whose eGenes intersect a cGene and those that do not. **e**, Minor allele frequencies (MAFs) for GWAS variants uniquely mapped to eGenes or cGenes. **f-g**, Gene expression*

135 *(transcripts per million; TPM) for cGenes and eGenes in K562 erythroleukemia cells* ***(f)*** *and in GTEx*
136 *whole blood* ***(g)****.* ***h-i****, Simulated discovery power for eGene at aFC of 0.25* ***(h)*** *and cGene at effect size*
137 *0.25* ***(i)*** *between CRE-gene targets that intersect between CRISPRi and eQTL methods versus those that*
138 *do not, at GWAS-associated CREs with both target types.*
139
140 **Biological properties of eQTL and CRISPRi discoveries**
141 Next, we examined the biological properties of the discovered eGenes and cGenes. While most
142 cGenes (81%) and eGenes (68%) lie within 100 kb of the GWAS CRE (Fig. 2a), cGenes are
143 more proximal: 41% are within 10 kb, compared to only 23% of eGenes. As a baseline, 21% of
144 CREs without a cGene or eGene have the closest gene within 10 kb (Fig. 2b). The proximity of
145 CRE–cGene pairs is notable given that CRISPRi screens specifically targeted regions more
146 than 1 kb from transcription start sites, and repression by KRAB/MeCP2 is generally limited to
147 local chromatin [6,7,17]. Thus, we do not expect CRISPRi to cause gene repression by inhibition of
148 gene promoters. eQTL mapping more often identifies genes beyond the 2nd closest gene (64%
149 vs. 23% for cGene-linked CREs, $P = 1.0 \times 10^{-10}$, Fig. 2c). Additionally, CRISPRi identifies a
150 smaller number of genes, with 82% of cGene-linked CREs and 33% of eGene-linked CREs
151 being associated with a single gene ($P = 2.0 \times 10^{-26}$, Fig. 2d). This difference reflects not only
152 the power of eQTL mapping but also the broader range of eQTL studies and cell types
153 analyzed, with a median of two distinct eQTL studies identifying unique genes per CRE (fig.
154 S5a). These results indicate that eQTL mapping better captures weaker regulatory effects that
155 span larger genomic distances. This is important for identifying more distal target genes as
156 more distal CRE-promoter interactions are, on average, weaker [18] ($P = 7.5 \times 10^{-05}$, fig. S5b) and
157 thus harder to detect. Larger future CRISPRi studies with more diverse cell types might bring
158 the two approaches closer to alignment.

159 It has been proposed that, due to natural selection, eQTLs that rely on natural genetic variation
160 are biased against selectively constrained effects [19]. Indeed, we observed that cGenes were

161 more haplosensitive, measured by Episcore predictions ($P = 2.5 \times 10^{-3}$, Fig. 2e), pHaplo scores
162 ($P = 0.01$), and pLI scores ≥ 0.9 (P = 0.04, Fig. 2f). This suggests that CRISPRi is better at
163 identifying genes that do not tolerate dosage variation. Furthermore, cGenes had a higher
164 number of associated enhancers ($P = 2.4 \times 10^{-149}$, Fig. 2g) and super-enhancers ($P = 3.4 \times$
165 10^{-4}), suggesting a more complex regulatory architecture and tighter regulatory control than
166 eGenes. These results highlight the potential of experimental perturbations such as CRISPRi to
167 discover genes under selective and regulatory control, which are more likely to be relevant for
168 diseases and traits [20]. However, it remains unclear how often highly constrained genes undergo
169 more buffering [21] to the extent that their expression changes are harder to detect even after
170 CRISPR perturbations.

171

172 From ATAC-seq data spanning 18 blood cell types (Supplementary Methods), we observed that
173 CREs with only eGenes had more cell-type specific open chromatin than those with also or
174 exclusively cGenes (Fig. 2h; Kruskal-Wallis rank sum test, $P = 0.03$). This suggests that cell
175 type differences between primary cell types in eQTL studies and the K562 line contribute to the
176 sharing of detected regulatory effects, and that experiments in K562s tend to pick up regulation
177 that is shared between many primary cell types. This is potentially due to K562 being a
178 progenitor cell type, or because distal enhancers that are shared across cell types are also
179 stronger on average, and thus we have more power to detect their associated genes. Notably,
180 12% of eGenes were located in regions of closed chromatin across all blood cell types
181 (excluding K562), compared with only 5% of cGenes (P = 0.02), supporting the idea that
182 CRISPRi-based mapping is especially sensitive to classical biochemical enhancer features [6,7].
183 Hi-C data in K562 cells revealed that enhancer-gene targets from physical chromatin contacts

184 were more likely to overlap with cGenes than eGenes ($P = 1.8 \times 10^{-5}$, fig. S5c), although 53% of
185 the GWAS CREs tested had no annotated Hi-C contacts.

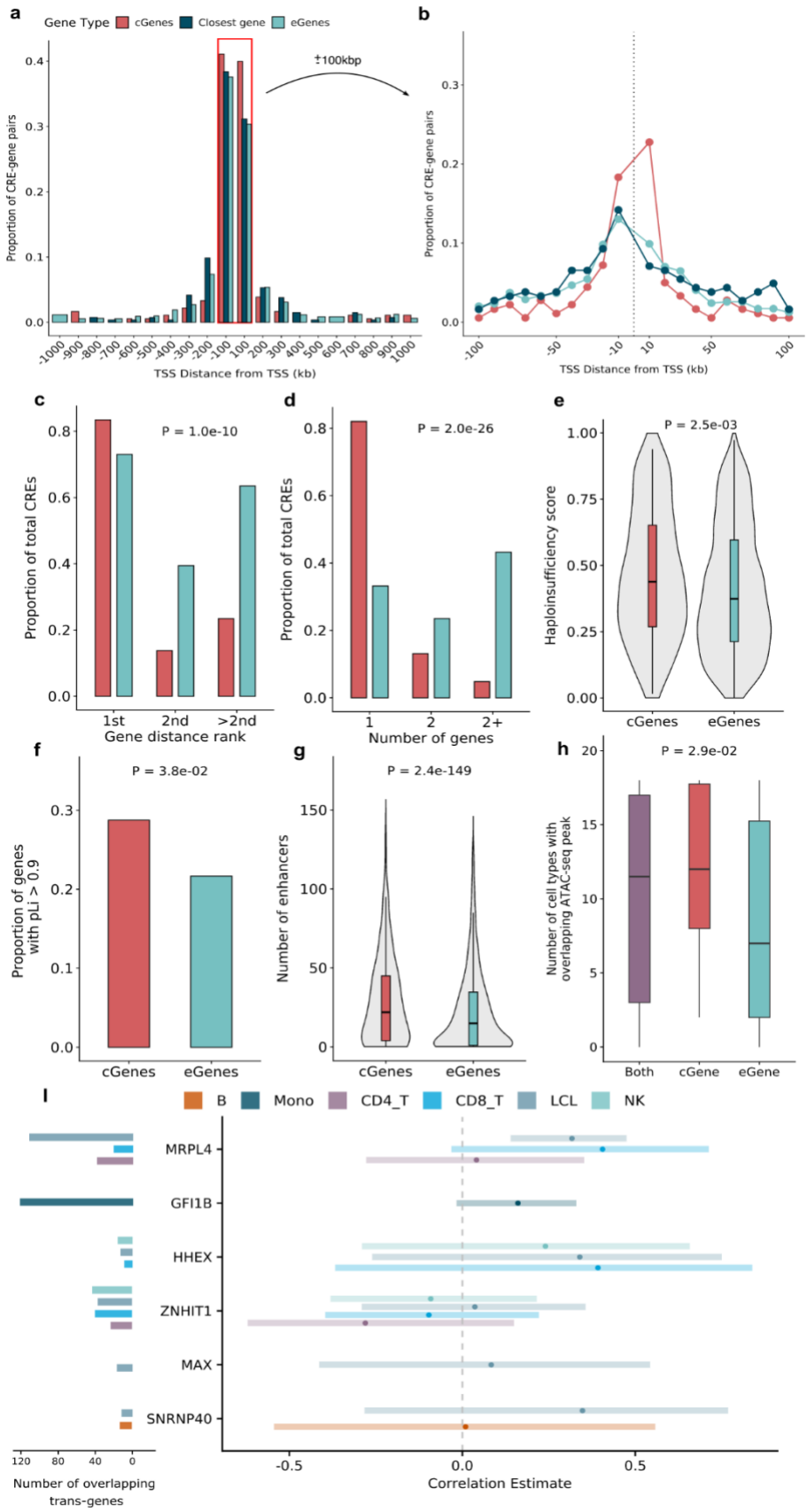
186

187 **Shared cis-gene anchors reveal variable *trans*-regulatory landscapes across datasets**

188 We next examined how CRISPRi-identified *trans*-regulatory genes (*trans*-cGenes) compared
189 with eQTL-based *trans* genes (*trans*-eGenes) for the same GWAS CREs acting through a
190 shared *cis*-gene. We detected significant *trans*-cGene effects for 16 GWAS CREs (FDR < 0.10;
191 72–3,870 trans-cGenes per CRE; fig. S6a, Supplementary Tables 8 & 9). At two of these loci
192 (*GFI1B* and *HHEX*), trans-cGenes were detected by both Morris et al. and Gasperini et al.[6,7],
193 with high correlation in effect sizes (*GFI1B*: r = 0.88, $P < 2 \times 10^{-16}$, n = 921; *HHEX*: r = 0.92, P =
194 9.6×10^{-8}, n = 18, Extended Data Figs. 6b & c). To identify *trans*-eGenes for these loci, we
195 analyzed *trans*-eQTLs (FDR < 0.05) from single-cell PBMC data from OneK1K [15] and a large
196 meta-analysis of lymphoblastoid cell lines (LCLs) (n = 3,892) [22]. We discovered *trans*-eGenes
197 for 11 out of 16 CREs (data S10), of which 6 had overlapping *trans*-genes (fig. S6d). The
198 correlation between trans-eGene and trans-cGene effects was modest for most loci and often
199 variable between cell types (r = -0.28–0.41, Fig. 2i), and significant only for the CRE regulating
200 *MRPL4* in LCLs (r = 0.32, FDR = 9.9×10^{-3}, data S11). Additionally, the overlap of trans-cGenes
201 and trans-eGenes was not significant for any CREs or cell types (fig. S6e, data S11). These
202 data indicate the potential for robust detection of trans-regulatory networks with both
203 approaches, but also the need for further work to fully understand the drivers of shared and
204 distinct patterns across a larger number of loci.

205

206


a
Gene Type cGenes Closest gene eGenes
±100kbp
Proportion of CRE-gene pairs
TSS Distance from TSS (kb)
b
Proportion of CRE-gene pairs
TSS Distance from TSS (kb)
c
P = 1.0e-10
Proportion of total CREs
1st 2nd >2nd
Gene distance rank
d
P = 2.0e-26
Proportion of total CREs
1 2 2+
Number of genes
e
P = 2.5e-03
Haploinsufficiency score
cGenes eGenes
f
P = 3.8e-02
Proportion of genes with pLi > 0.9
cGenes eGenes
g
P = 2.4e-149
Number of enhancers
cGenes eGenes
h
P = 2.9e-02
Number of cell types with overlapping ATAC-seq peak
Both cGene eGene
i
B Mono CD4_T CD8_T LCL NK
MRPL4
GFI1B
HHEX
ZNHIT1
MAX
SNRNP40
Number of overlapping trans-genes
Correlation Estimate

228
229
230
231
232
233
234
235 ***Figure 2. Biological properties of eQTL and CRISPRi discoveries**, **a,** The distance between GWAS-*
236 *associated CREs and target gene TSS, comparing cGenes, eGenes, and the closest gene to GWAS*
237 *CREs where no significant association was identified, with **b** showing the region within 100 kb of TSS. **c,***
238 *The proportion of GWAS-associated CREs linked to their closest, second closest, or further away genes.*
239 *CREs may be counted in multiple categories if they are associated with multiple genes. **d**, The proportion*
240 *of GWAS-associated CREs linked to one, two, or more genes. **e**, Haploinsufficiency (predicted by*
241 *Episcore) for cGenes and eGenes. **f**, The proportion of cGenes and eGenes with loss-of-function*
242 *intolerance metric pLI > 0.9. **g**, The number of enhancers for cGenes and eGenes. **h**, Number of*
243 *blood-cell types in which a CRE has open chromatin for CREs with both eGenes and cGenes, only*
244 *cGenes, and only eGenes. **i**, Trans-gene analysis, with the y-axis indicating the cis-genes through which*
245 *trans-effects occur. The left-hand side of the plot shows the number of overlapping trans eGenes and*
246 *cGenes, and the correlation and its confidence interval of the effect sizes are shown on the right-hand*
247 *side.*
248

249 **Detection of gold-standard GWAS genes by eQTL and CRISPRi**

250 To evaluate the ability of each method to identify genes that are likely to be true causal genes
251 for blood trait phenotypes, we constructed a “gold-standard" gene list of 421 protein-coding
252 genes in the studied loci (data S12). This included genes identified through loss-of-function
253 (LoF) burden tests for 29 blood cell traits in the UK Biobank [23] and a curated set of genes linked

254 to Mendelian blood disorders (Methods). eQTL mapping identified 65 of these, and 23 were
255 discovered as cGenes, corresponding to 11% (65/578) of all eGenes and 22% (23/107) of all
256 cGenes in the loci containing gold-standard genes (Fig. 3a). Only 10 genes were detected by
257 both approaches. When a cGene overlapped a gold-standard gene, it was the closest gene to
258 the GWAS CRE in 74% of cases, compared to 54% for eGenes, indicating the further potential
259 of eQTLs to identify more distal target genes that are more difficult to identify as putative targets
260 of GWAS loci. Interestingly, gold-standard eGenes had a similar eQTL effect size distribution as
261 other eGenes, indicating that even weak eQTL effects will pinpoint true target genes (Extended
262 Data Fig 7) [9,24]. However, the patterns of gene target discovery are highly variable between the
263 loci (Fig. 3d, Extended Data Figs. 8-16).

264

265 Next, we analyzed loci at which the gold-standard gene was not identified to understand why it
266 was missed. In eQTL analysis, lack of power is an infrequent explanation, as only 7% of the loci
267 had simulated statistical power <0.8. Again, other eGenes were detected in 84% of the loci,
268 which may lead to a misleading inference. Conversely, in CRISPRi analysis, as much as 43% of
269 gold-standard genes were either underpowered in CRISPRi study simulations (15%) or not
270 expressed in K562 cells (29%), and in 37% of the loci, a different cGene was identified (Fig. 3b).
271 We note that both eQTL and CRISPRi power calculations assume uniform effect sizes across
272 CRE–gene pairs; however, distal or more constrained genes likely exhibit weaker regulatory
273 effects, leading to overestimation of power for these targets. These results indicate distinct
274 shortcomings in both approaches in the identification of causal genes for complex traits.

275

276 Assuming gold-standard genes represent true causal targets within a locus, we established a
277 set of 1,075 causal CRE–gene pairs among 15,699 candidates located within ±1 Mb of each

278 GWAS variant (Methods) to analyze the performance across target gene mapping strategies.
279 cGenes achieved the highest precision (24%) but with the lowest recall (3%, Fig. 3c). eQTL
280 colocalization identified more causal genes (recall = 8%) at lower precision (11%), and
281 tightening the threshold to PP.H4 > 0.9 raised the precision to 16% while reducing recall to 3%.
282 Assigning each GWAS variant to its nearest gene provided the greatest recall (9 %) with good
283 precision (18%). Hi-C interactions and activity-by-contact (ABC) scores exhibited intermediate
284 precision–recall trade-offs, with TWAS having a similar performance (Fig. 3c). Combining both
285 cGenes and eGenes (PP.H4 > 0.9) offered increased recall (6%), while maintaining good
286 precision (18%).The respective importance of precision or recall depends on the downstream
287 applications. However, the largely non-overlapping subsets of gold-standard genes identified by
288 the two methods point to their orthogonal nature in discovering candidate causal genes for
289 GWAS loci.

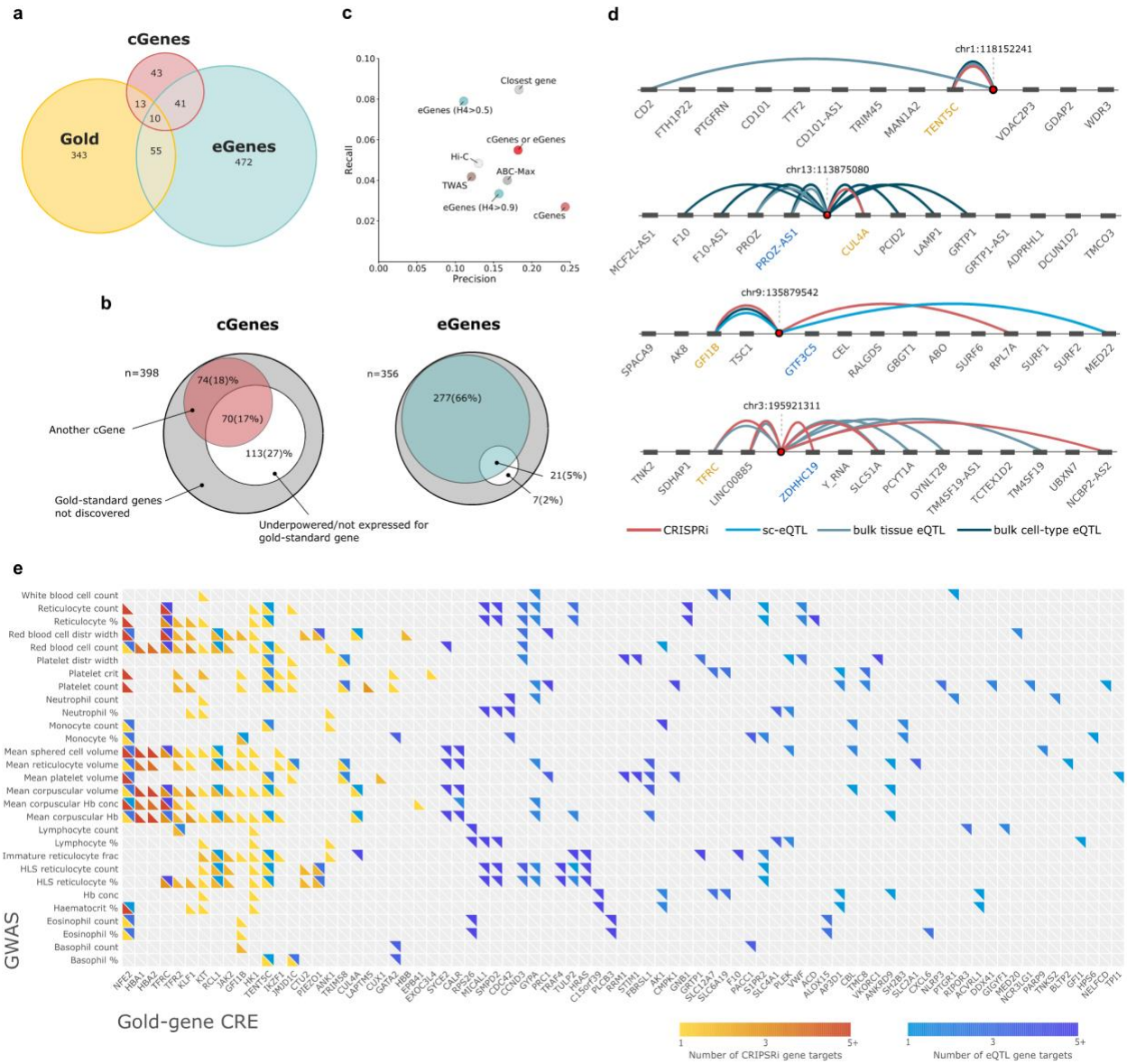


Figure 3. Discovery of gold-standard genes for blood traits. a, *The overlap between cGenes, eGenes, and gold-standard genes.* ***b,*** *The gold-standard genes that were not identified as either cGenes or eGenes, with the proportion of undiscovered gold-standard genes for which eQTL and CRISPRi studies suggest a different causal gene at the same locus, as well as the proportion that remain undetected due to power limitations or lack of gene expression.* ***c,*** *Precision-recall for mapping GWAS variants to gold-standard causal genes (n = 1,075 CRE-gene pairs) within 1 Mb (n total = 15,699 CRE-gene pairs). Precision is the proportion of predicted positives that are true positives, while recall is the proportion of true positives that are correctly identified.* ***d****, Example loci showing the position of a GWAS*

299 *variant and its gene associations as identified by different study types. Gold labels indicate the gold-*
300 *standard gene in each locus, while blue labels mark the closest gene to the variant. If no blue label is*
301 *present, the gold-standard gene is also the closest gene.* ***e,*** *Heatmap showing for each gold-standard*
302 *gene intersecting a cGene and/or eGene (x-axis) and GWAS trait that the CRE mapping to that gene is*
303 *associated to (y-axis), the number of genes located at the same locus identified by either CRISPRi*
304 *perturbation screens (warm colours) or eQTL analyses (cool colours).*

305

306 **Discussion**

307 We present the first systematic evaluation of the performance of the increasingly common
308 CRISPRi approach for elucidating gene targets of GWAS loci, and contrast it to eQTL analysis.
309 Our results revealed that the gene targets and GWAS CREs captured by each approach are
310 often different. Some discordance is expected due to the differences in the designs of the
311 analyzed studies, which rely on CRISPRi silencing in a cancer cell line versus genetic variation
312 in primary cells. Furthermore, this work compares the results of the first generation of CRISPRi
313 studies to results from more than a decade of eQTL work. The discoveries of the first eQTL
314 studies were not too dissimilar from the patterns of CRISPRi results here: limited by discovery
315 power, biased towards strong regulatory effects close to the promoter, detected in a single cell
316 type [25,26]. However, the major investments in both CRISPRi and eQTL studies to identify *cis*
317 target genes for GWAS loci creates a critical need to understand the yield from each approach
318 to pursue optimal study designs and practical applications, as well as make accurate biological
319 inferences.

320

321 Our results show that while the eQTL data analyzed here was highly sensitive to detect proximal
322 and distal cis-target genes of GWAS loci, important gaps remain. While sc-eQTL analysis has
323 recently emerged as an attractive opportunity to capture genetic regulatory effects across

324 diverse cellular contexts, further investments to increase power are needed for efficient GWAS
325 target gene identification. eQTL analyses also struggle to detect regulatory effects of GWAS
326 variants at lower allele frequencies and at low effect sizes. Many gold-standard GWAS genes
327 being picked up by eQTLs with small effect sizes indicates continued opportunity for discovery
328 in this realm. This is critical, as disease-relevant genes are more dosage-sensitive, constrained
329 and thus harbor mostly minute regulatory effects and/or rare variants[9,19]. This, however, creates
330 a double-edged sword: analyses that are sufficiently powered to detect small disease-relevant
331 effects will readily capture many other genes as well, leading to loss of specificity, also
332 demonstrated by our data. The real strength of eQTL studies is the study of primary cells and
333 tissues, which inherently capture regulatory effects in a context that is closer to the in vivo state
334 than model systems. Our results show how matching the analyzed cell types for the different
335 blood cell traits contributes to the yield in target discovery, motivating ongoing investments in
336 analysis genetic regulatory effects across disease-relevant cellular contexts.
337
338 The studied CRISPRi data sets detected a smaller number of target genes, mostly proximal to
339 the GWAS locus, with statistical power limiting the discovery of weaker effects that predominate
340 distal regulatory effects. This is an important gap in the ability to capture novel insights and
341 unexpected GWAS target genes, as genes within tens of kilobases of a GWAS locus would be
342 the first candidates based on gene annotation alone. However, our results show that CRISPRi
343 target genes capture gold-standard GWAS genes with high specificity, and are better at
344 detecting target genes that are more selectively constrained and tightly regulated. Ongoing
345 expansion of CRISPR studies to accessible primary cell types and to variant-level editing will be
346 critical to capture directly analogous disease biology and to understand to which extent this can
347 be imputed from more scalable model systems and perturbations. Furthermore, CRISPR
348 activation can complement CRISPRi by enhancing the power to detect lowly expressed genes.
349 While we anticipate further discoveries as the CRISPR field grows and matures, the

350 improvements in sensitivity, resolution, and diversity are likely to come with some loss of
351 specificity.
352
353 Multiple gene targets of CREs across different cell types is a core feature of genome function [27–
354 29], and accurate identification of which of the many target of a GWAS locus is the true driver of
355 disease biology, and is likely not going to be achieved by cataloguing CRE targets alone.
356 Rather, it will require further gene prioritization based on their functional properties. However,
357 we note that while the identification of causal GWAS genes is a major research focus due to
358 their role as potential drug targets[30], this is not the only outcome or objective for eQTL and
359 CRISPRi studies that also provide invaluable, generalizable information about the biology of
360 gene regulatory mechanisms.
361
362 Our work is the first empirical demonstration of how both the eQTL and CRISPRi approaches
363 have substantial limitations and orthogonal strengths. This highlights the importance of
364 systematic evaluation of analytical approaches to understand their current yield and future
365 opportunities. The complementarity of approaches that rely on population variation and
366 experimental perturbations is an exciting opportunity for joint biological discovery that would be
367 difficult to achieve with any approach alone, providing inspiration for building a diverse and
368 comprehensive toolkit for GWAS interpretation.
369
370 **Methods**
371 **Data**
372 **GWAS data**
373 We utilized the genome-wide association study (GWAS) results for 29 different blood cell traits
374 from 361,194 white British participants in the UK Biobank and trans-ethnic meta-analyses for 15
375 hematological traits in 746,667 participants from the BCX consortium [31]. These traits included

measures such as white blood cell count, red blood cell count, hemoglobin concentration, hematocrit percentage, and various others listed in data S1.

eQTL data

A total of 51 eQTL studies [5,13–15,32], focusing on blood and/or blood cell traits were analyzed (data S2). Summary statistics and log Bayes factors for most studies were obtained from the eQTL Catalogue [13]. Summary statistics for GTEx [5], the Multi-Ancestry Analysis of Gene Expression (MAGE) study [14], the INTERVAL Study [32], and Meta LCL [22] were acquired from their respective study data sources listed in data S2. OneK1K association summary statistics were generated through re-analysis of the individual-level data (see Supplementary Methods).

CRISPRi perturbation data

CRISPRi data were obtained from Morris et al. [6], Gasperini et al. [7], and Gschwind et al [16]. The Morris and Gasperini studies involved high-MOI CRISPRi screens targeting enhancers in K562 cells and were reanalyzed using SCEPTRE [33] with the default “union” setting, which groups gRNAs targeting the same element for association testing. Gschwind et al. compiled a harmonized gold-standard dataset of 10,411 element–gene pairs in K562 cells by reanalyzing CRISPR-based perturbation data from multiple studies. The largest number of CRE-gene targets within the Gschwind et al. study consisted of data from Schraivogel et al 2020., which utilized TAP-seq, and Nasser et al 2021., which utilized CRISPR FlowFISH. The different CRISPR datasets applied different candidate element selection strategies, leading to different characteristics, however, the vast majority of the data (>95% of CRE-gene targets) were data comprised from Morris et al. and Gasperini et al. which share a similar experimental design. To account for CREs that were targeted by multiple studies and the tiling approach utilized in

400 flowFISH experiments we combined the targets of gRNA that were within 2kb of each other.

401

402 **Capture Hi-C data data**

403 We obtained capture Hi-C data in K562 cells from Zhigulev et al. 2024 [34]. HiCapTools [35] was
404 used to call interactions in all samples. A detailed description of how the interactions were called
405 is reported elsewhere. Briefly, at least four pairs supporting each interaction were required.
406 Three Bonferroni-adjusted *P*-value cutoffs (0.1, 0.01, 0.001) were used for filter interactions. For
407 consistency with the eQTL and CRISPRi datasets, we used the 0.1 threshold. To identify GWAS
408 target genes, we intersected fine-mapped credible sets with interacting regions using bedtools.
409 All HiCap results were mapped to the GRCh37 genome assembly.

410

411 **Determination of CRISPRi target genes**

412 To link CRISPRi perturbations to GWAS variants, we obtained the SNP coordinates of the
413 genomic regions targeted by gRNAs from Morris et al. For Gasperini et al., gRNA targeting
414 coordinates were determined by aligning spacer sequences to the GRCh37 reference genome
415 and identifying their positions within the targeted enhancers. Next, we intersected these
416 CRISPRi-targeted regions with fine-mapped GWAS variants within ±2 kb to determine which
417 credible set variants were perturbed by a particular CRISPRi experiment.

418

419 **Gold-standard gene list**

420 Summary statistics for pLoF burden tests on 29 blood cell traits were obtained from Genebass
421 [36]. Protein-coding genes with a SKAT p-value $< 5 \times 10^{-7}$ and located within 1 Mb of GWAS
422 CREs targeted by CRISPRi gRNAs were selected for analysis. In addition to these burden test
423 genes, curated gene lists associated with rare inherited Mendelian blood disorders [37,38] that

424 were also in cis (within 1 Mb) of a GWAS CRE were also included. Fisher's exact tests were
425 performed to assess the enrichment of overlap between Mendelian genes, burden test genes,
426 cGenes, eGenes, and Hi-C gene targets. Hi-C, genes were filtered to protein-coding genes that
427 were associated with a CRE within 1 Mb of a gold-standard gene.
428

429 **Activity-by-contact gene predictions**

430 Activity-by-Contact (ABC) enhancer–gene predictions were overlapped with the genomic
431 coordinates of our CRISPR-targeted candidate regulatory elements (CREs) to identify putative
432 target genes for each guide RNA. We first downloaded the ABC prediction file
433 (AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt.gz; Nasser et al., 2021) from
434 the Broad Institute FTP server. A subset of immune-relevant and hematopoietic cell types was
435 specified, including K562 (Roadmap), B cells (ENCODE and BJAB), $CD4^+$ and $CD8^+$ αβ T cells
436 (ENCODE and Corces 2016), $CD14^+$ monocytes under baseline and stimulated conditions
437 (ENCODE and Novakovic 2016), LPS-treated dendritic cells (Garber 2017), megakaryocyte–
438 erythroid progenitors (Corces 2016), NK cells (Roadmap and Corces 2016), GM12878
439 (Roadmap), Jurkat (Engreitz), U937 (Engreitz), THP-1 monocytes/macrophages (Van
440 Bortle 2017), and OCI-LY7 (ENCODE). The CRE-guide coordinate file was intersected with
441 each cell-type-specific ABC file using bedtools intersect resulting in 466 gRNAs that intersected
442 a ABC enhancer. Each gRNA coordinate was assigned to the gene with the max ABC score
443 (ABC-Max) across all the cell types specified.

444

445 **TWAS gene predictions**

446 To identify the most significant transcriptome-wide association study (TWAS) gene in cis for
447 each GWAS CRE test, we used results from the Rowland et al. who performed a TWAS for 29

448 blood-cell traits from the UK biobank [39]. We retained only protein-coding genes and for each
449 GWAS CRE, we defined a cis window as ±1 Mb around the gRNA's genomic position. Within
450 this window, we extracted all TWAS associations whose start or end coordinates overlapped the
451 interval on the matching chromosome. If one or more associations were found, we retained the
452 gene with the most significant association (i.e., the highest log10 p-value). TWAS associations
453 were only retained if they passed a marginal significance threshold (as defined in the TWAS
454 dataset). The resulting dataset consisted of the most significant cis TWAS gene for each
455 CRISPR perturbation, and was used to evaluate the overlap between TWAS predictions and
456 CRISPR-based regulatory inferences.

457

458 **Statistical analysis**

459 **Statistical fine-mapping of UK Biobank blood cell traits**

460 For each of the 29 UK Biobank blood trait GWASs, statistical fine-mapping was conducted using
461 Sum of Single Effects regression (SuSiE), as implemented in the susieR R package [40]. We
462 defined 899 disjoint fine-mapping regions (Supplementary Methods) containing genome-wide
463 significant variants ($P < 6.6 \times 10^{-9}$). We generated linkage disequilibrium (LD) matrices from a
464 subset of 50,000 UK Biobank white British participants (UK Biobank accession code: 47976)
465 using PLINK. We extracted all variants in the 95% credible set with a posterior inclusion
466 probability (PIP) > 0.01.

467

468 **Colocalization with UKBB GWAS variants**

469 We performed colocalization between 29 UK Biobank blood trait GWASs and 49 eQTL studies
470 of European ancestry. Pairwise colocalization was conducted using SuSiE COLOC [41] between

471 each blood trait GWAS and eQTL study. When available, we used fine-mapped eQTL data from
472 the eQTL Catalogue; otherwise, eQTL data were fine-mapped before colocalization using in-
473 sample LD matrices. To ensure consistency in genomic positions, all data were mapped to the
474 GRC37 using bcftools liftover [42]. Colocalization was performed within regions meeting the
475 thresholds of $p < 6.6 \times 10^{-9}$ for GWAS associations and $p < 1 \times 10^{-5}$ for eQTL signals. SuSiE
476 COLOC was conducted with default settings, and if convergence failed at a particular locus,
477 colocalization was performed using COLOC [43] under the assumption of a single causal variant.
478 Additionally, we conducted colocalization between 15 blood cell traits from the BCX GWAS [31]
479 consortium and the multi-ethnic eQTL study MAGE [14]. For this analysis, COLOC was performed
480 under the single causal variant assumption without requiring additional LD information.

481

482 **Statistical testing**

483 To compare the number of enhancers between gene sets, we used Poisson regression.
484 Differences in Episcore, pHaplo, gene expression levels in K562 and GTEx whole blood, and
485 MAF were assessed with Gaussian regression, while differences in pLI were assessed with
486 binomial regression. All gene-level models were adjusted for gene length.

487 We applied the Wilcoxon rank-sum test to compare eQTL colocalization probabilities and
488 Fisher's exact test to evaluate the number of genes with estimated CRISPRi or eQTL power
489 greater than 0.8. Chi-squared tests were used to assess differences in categorical distributions,
490 including the number of genes per locus and gene–CRE distances. All statistical tests were two-
491 sided. Associations between log-fold change and TSS distance were tested using linear
492 regression, adjusting for the number of cells bearing each gRNA.

493 **CRISPRi power calculation**

494 We estimated the power to detect reductions in gene expression for each of the perturbed

495 CRE–target gene pairs using a modified version of the approach published by Gschwind et al.
496 2023[16]. For the Morris et al. and Gasperini et al. datasets, we first applied the
497 estimateDispersions function (DESeq2, v1.46.2) to estimate mean and overdispersion
498 parameters across cells for each gene. Using these estimates, we sampled expression values
499 from a negative binomial distribution, introducing a fixed reduction in expression (15%) for each
500 perturbed gene linked to a given CRE. We calculated power for effect sizes 10%, 15%, 20%,
501 25%, 30%, 40%, 50% reduction in expression. An effect size of 25% was determined based on
502 the number of significant empirical CRISPRi CRE-gene associations observed within the data,
503 in comparison to the number significant empirical eQTL variant gene associations. We sampled
504 the same number of cells for each perturbation as the original datasets and aggregated them to
505 ensure that perturbed and non-perturbed cell counts remained consistent. We next ran
506 SCEPTRE in high-MOI mode—without any extra quality filtering—to assess each CRE–gene
507 pair for a significant decrease in expression, correcting P-values across all pairs within each
508 dataset with the Benjamini–Hochberg procedure. The entire simulation was repeated 10 times,
509 and a CRE–gene pair was considered powered for detection if a significant expression change
510 was observed at FDR < 10% in at least 80% of simulations. For the remaining datasets, power
511 estimates produced by Gschwind et al. 2023 were used [16].
512
513 **Estimation of equivalent eQTL effect size for CRISPRi perturbations**
514 Empirical power analyses indicated that approximately 80% of significantly perturbed CREs
515 exhibited ≥80% statistical power at an estimated 25% reduction in gene expression relative to
516 non-targeting control (NTC) cells. To translate this level of repression into an equivalent eQTL
517 effect size, we used allelic fold-change (aFC) estimates from GTEx v10 Whole Blood eQTLs.

518 eGenes were restricted to those with q-values ≤0.05. We identified variants whose rounded aFC

519 values approximated log2(0.75), corresponding to a 25% decrease in expression, while

520 excluding variants whose 95% confidence intervals encompassed zero. The mean β-coefficient

521 among these variants was 0.27, which we used as the eQTL effect size comparable to the

522 CRISPRi perturbation magnitude.

523

524 **eQTL power calculation**

525 For genes with a mean count greater than 5, power is estimated analytically using an F-test.

526 Power in this context depends on sample size n, the significance threshold α ($P = 1 \times 10^{-3}$), and

527 the proportion of variance explained (q^2) by the variant in the population. Assuming genotypes

528 are in Hardy-Weinberg equilibrium, q^2 is given by:

529

530 $q^2 = 2 \times MAF \times (1 - MAF) \times \beta^2$ [44]

531

532 The minor allele frequency (MAF) refers to the GWAS variant, with β set to 0.25 and the

533 phenotype (gene expression) expressed in standard deviation units.. For lowly expressed

534 genes, power is determined through simulations using scPower [45]. In sc-eQTL studies, the

535 overall power is the product of eQTL power and expression probability, which quantifies the l

536 likelihood of a gene being detected in a given cell type [45]. Expression probability is

537 determined

538 by sample size, the number of cells per cell type, cell type frequencies, and the proportion of

539 individuals in which a gene is expressed. Read depth is implicitly accounted for through UMI

540 counts. When comparing power across studies and methods, we used 13,874 protein-coding
541 genes that passed QC in the INTERVAL study as the reference set, assumed to be
542 expressed
543 across all studies. For a detailed description of power calculations, refer to the
544 Supplementary
545 Methods.
546
547 **Trans-eQTL calling**
548 To identify GWAS CREs with trans-eQTLs, we selected *cis*-eQTL variants for genes that
549 displayed significant trans-effects in CRISPRi studies and were cis-targets of GWAS CREs.
550 Trans-eQTL analysis was then performed for these cis-eQTL variants across Azimuth level 1
551 predicted cell types in the OneK1K dataset using tensorQTL [46]. Trans associations within a ±5
552 Mb window that met a significance threshold of FDR < 0.05 were retained. Additionally, trans-
553 eQTL results from a meta-analysis of Lymphoblastoid Cell Lines (LCLs) in MetaLCL [22] were
554 incorporated. Using the same approach, we identified cis-eQTL variants for genes with
555 CRISPR-derived trans networks that were also GWAS targets based on MAGE data [14]. We then
556 extracted trans associations for these cis-eQTLs from the MetaLCL results and retained those
557 with FDR < 0.05 to further assess their trans-regulatory effects.
558
559 **Trans-eQTL comparison**
560 To assess concordance of trans-effects between methods, we first identified significant trans-
561 cGene associations (FDR < 0.10) from Morris et al. and Gasperini et al. using SCEPTRE,
562 filtered to cis-genes linked to GWAS CREs. Trans-eGene effect sizes were then aligned to the
563 allele associated with decreased expression of the cis-gene. We then tested the correlation of
564 effect sizes across overlapping trans-genes using Pearson correlation, adjusting p-values for
565 multiple testing with the Benjamini-Hochberg procedure. Enrichment analyses were conducted

566 using a permutation-based Fisher's exact test to estimate odds ratios and confidence intervals,
567 with multiple testing correction applied similarly.
568
569 **Gene properties**
570 Enhancer and super-enhancer counts per gene were derived from GeneHancer v5 [47]
571 [47], with filtering applied to retain enhancers with a score ≥ 0.7 and gene associations
572 with a "link_score" ≥ 7. Haploinsufficiency (pHaplo) scores and Episcore values were
573 obtained
574 from Collins et al [48]. pLI scores were retrieved from gnomAD v3, and gene
575 expression levels (TPM) for whole blood were obtained from GTEx version 8. Hi-C data for
576 enhancer-promoter interactions in K562 was obtained from Zhigulev et al [34].
577
578 **Calculation of Precision and Recall for gene mapping approaches**
579 To assess the sensitivity and specificity of CRISPRi and eQTLs in identifying the correct
580 target
581 gene, we evaluated all genes within ±1 Mb of each GWAS CRE that contained a gold-
582 standard
583 gene and at least one significant cGene or eGene. Any gold-standard gene at each locus
584 was
585 treated as the true causal gene. 497 GWAS CREs contained at least one gold-standard gene
586 within 1 Mbp, resulting in 1,075 causal CRE-gene connections were determined out of a total
587 of
588 13,000 CRE-gene links. We limited our analysis to only protein coding genes as the majority
589 of

590 the gold-standard genes were protein coding. A CRE–gene match to the gold-standard was
591 counted as a true positive (TP); identification of a different gene as a false positive (FP);
592 failure
593 to identify the gold-standard as a false negative (FN); and all correctly excluded non-gold
594 genes
595 as true negatives (TN). Recall was calculated as TP / (TP + FN), and precision as TP / (TP +
596 FP).
597
598
599
600
601
602
603 **References**
604

605 1. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds,
606 A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common
607 disease-associated variation in regulatory DNA. Science *337*, 1190–1195.
608 10.1126/science.1222794.

609 2. Lappalainen, T., and MacArthur, D.G. (2021). From variant to function in human disease
610 genetics. Science *373*, 1464–1468. 10.1126/science.abi8207.

611 3. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs:
612 illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.
613 10.1016/j.ajhg.2013.10.012.

614 4. Barbeira, A.N., Bonazzola, R., Gamazon, E.R., Liang, Y., Park, Y., Kim-Hellmuth, S.,
615 Wang, G., Jiang, Z., Zhou, D., Hormozdiari, F., et al. (2021). Exploiting the GTEx resources
616 to decipher the mechanisms at GWAS loci. Genome Biol. *22*, 49. 10.1186/s13059-020-
617 02252-4.

618 5. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across
619 human tissues. Science *369*, 1318–1330. 10.1126/science.aaz1776.

620 6. Morris, J.A., Caragine, C., Daniloski, Z., Domingo, J., Barry, T., Lu, L., Davis, K., Ziosi, M.,
621 Glinos, D.A., Hao, S., et al. (2023). Discovery of target genes and pathways at GWAS loci

622 by pooled single-cell CRISPR screens. Science *380*, eadh7699. 10.1126/science.adh7699.

623 7. Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson,
624 D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for
625 Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377-390.e19.
626 10.1016/j.cell.2018.11.029.

627 8. Tambets, R., Kolde, A., Kolberg, P., Love, M.I., and Alasoo, K. (2024). Extensive co-
628 regulation of neighboring genes complicates the use of eQTLs in target gene prioritization.
629 HGG Adv. *5*, 100348. 10.1016/j.xhgg.2024.100348.

630 9. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in
631 discovery of genetic effects on gene expression and complex traits. Nat. Genet. *55*, 1866–
632 1875. 10.1038/s41588-023-01529-1.

633 10. Stacey, D., Fauman, E.B., Ziemek, D., Sun, B.B., Harshfield, E.L., Wood, A.M.,
634 Butterworth, A.S., Suhre, K., and Paul, D.S. (2019). ProGeM: a framework for the
635 prioritization of candidate causal genes at molecular quantitative trait loci. Nucleic Acids
636 Res. *47*, e3. 10.1093/nar/gky837.

637 11. Weeks, E.M., Ulirsch, J.C., Cheng, N.Y., Trippe, B.L., Fine, R.S., Miao, J., Patwardhan,
638 T.A., Kanai, M., Nasser, J., Fulco, C.P., et al. (2023). Leveraging polygenic enrichments of
639 gene features to predict genes underlying complex traits and diseases. Nat. Genet. *55*,
640 1267–1276. 10.1038/s41588-023-01443-6.

641 12. Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013).
642 CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat.
643 Protoc. *8*, 2180–2196. 10.1038/nprot.2013.132.

644 13. Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča,
645 M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., et al. (2021). A compendium of uniformly
646 processed human gene expression and splicing quantitative trait loci. Nat. Genet. *53*,
647 1290–1299. 10.1038/s41588-021-00924-w.

648 14. Taylor, D.J., Chhetri, S.B., Tassia, M.G., Biddanda, A., Yan, S.M., Wojcik, G.L., Battle, A.,
649 and McCoy, R.C. (2024). Sources of gene expression variation in a globally diverse human
650 cohort. Nature *632*, 122–130. 10.1038/s41586-024-07708-2.

651 15. Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S.,
652 Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. (2022). Single-cell eQTL mapping
653 identifies cell type-specific genetic control of autoimmune disease. Science *376*, eabf3041.
654 10.1126/science.abf3041.

655 16. Gschwind, A.R., Mualim, K.S., Karbalayghareh, A., Sheth, M.U., Dey, K.K., Jagoda, E.,
656 Nurtdinov, R.N., Xi, W., Tan, A.S., Jones, H., et al. (2023). An encyclopedia of enhancer-
657 gene regulatory interactions in the human genome. BioRxiv. 10.1101/2023.11.09.563812.

658 17. Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy,
659 T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by
660 CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*,
661 1143–1149. 10.1038/nmeth.3630.

662 18. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska,
663 M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through
664 enhancer-promoter interactions. Nature *604*, 571–577. 10.1038/s41586-022-04570-y.

665 19. O'Connor, L.J., Schoech, A.P., Hormozdiari, F., Gazal, S., Patterson, N., and Price, A.L.
666 (2019). Extreme polygenicity of complex traits is explained by negative selection. Am. J.
667 Hum. Genet. *105*, 456–476. 10.1016/j.ajhg.2019.07.003.

668 20. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins,
669 R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint
670 spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
671 10.1038/s41586-020-2308-7.

672 21. Hartman, J.L., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic
673 variation. Science *291*, 1001–1004. 10.1126/science.1056072.

674 22. Freimann, K., Brümmer, A., Warmerdam, R., Rupall, T.S., Hernández-Ledesma, A.L.,
675 Chiou, J., Holzinger, E.R., Maranville, J.C., Nakic, N., Ongen, H., et al. (2024). *USP18*
676 modulates lupus risk via negative regulation of interferon response. medRxiv.
677 10.1101/2024.07.15.24310442.

678 23. Backman, J.D., Li, A.H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M.D., Benner, C.,
679 Liu, D., Locke, A.E., Balasubramanian, S., et al. (2021). Exome sequencing and analysis of
680 454,787 UK Biobank participants. Nature *599*, 628–634. 10.1038/s41586-021-04103-z.

681 24. Rosen, J.D., Broadaway, K.A., Brotman, S.M., Mohlke, K.L., and Love, M.I. (2025). Higher
682 eQTL power reveals signals that boost GWAS colocalization. bioRxiv.

683 25. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M., and Burdick, J.T.
684 (2005). Mapping determinants of human gene expression by regional and genome-wide
685 association. Nature *437*, 1365–1369. 10.1038/nature04244.

686 26. Stranger, B.E., Forrest, M.S., Clark, A.G., Minichiello, M.J., Deutsch, S., Lyle, R., Hunt, S.,
687 Kahl, B., Antonarakis, S.E., Tavaré, S., et al. (2005). Genome-wide associations of gene
688 expression variation in humans. PLoS Genet. *1*, e78. 10.1371/journal.pgen.0010078.

689 27. Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function
690 of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154. 10.1038/nrm3949.

691 28. Meuleman, W., Muratov, A., Rynes, E., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D.,
692 Neri, F., Teodosiadis, A., et al. (2020). Index and biological spectrum of human DNase I
693 hypersensitive sites. Nature *584*, 244–251. 10.1038/s41586-020-2559-3.

694 29. Donohue, L.K.H., Guo, M.G., Zhao, Y., Jung, N., Bussat, R.T., Kim, D.S., Neela, P.H.,
695 Kellman, L.N., Garcia, O.S., Meyers, R.M., et al. (2022). A cis-regulatory lexicon of DNA
696 motif combinations mediating cell-type-specific gene regulation. Cell Genomics *2*.
697 10.1016/j.xgen.2022.100191.

698 30. Trajanoska, K., Bhérer, C., Taliun, D., Zhou, S., Richards, J.B., and Mooser, V. (2023).
699 From target discovery to clinical drug development with human genetics. Nature *620*, 737–
700 745. 10.1038/s41586-023-06388-8.

701 31. Chen, M.-H., Raffield, L.M., Mousas, A., Sakaue, S., Huffman, J.E., Moscati, A., Trivedi, B.,
702 Jiang, T., Akbari, P., Vuckovic, D., et al. (2020). Trans-ethnic and Ancestry-Specific Blood-
703 Cell Genetics in 746,667 Individuals from 5 Global Populations. Cell *182*, 1198-1213.e14.
704 10.1016/j.cell.2020.06.045.

705 32. Tokolyi, A., Persyn, E., Nath, A.P., Burnham, K.L., Marten, J., Vanderstichele, T.,
706 Tardaguila, M., Stacey, D., Farr, B., Iyer, V., et al. (2025). The contribution of genetic
707 determinants of blood gene expression and splicing to molecular phenotypes and health
708 outcomes. Nat. Genet. *57*, 616–625. 10.1038/s41588-025-02096-3.

709 33. Barry, T., Wang, X., Morris, J.A., Roeder, K., and Katsevich, E. (2021). SCEPTRE
710 improves calibration and sensitivity in single-cell CRISPR screen analysis. Genome Biol.
711 *22*, 344. 10.1186/s13059-021-02545-2.

712 34. Zhigulev, A., Norberg, Z., Cordier, J., Spalinskas, R., Bassereh, H., Björn, N.,
713 Pradhananga, S., Gréen, H., and Sahlén, P. (2024). Enhancer mutations modulate the
714 severity of chemotherapy-induced myelosuppression. Life Sci. Alliance *7*.
715 10.26508/lsa.202302244.

716 35. Anil, A., Spalinskas, R., Åkerborg, Ö., and Sahlén, P. (2018). HiCapTools: a software suite
717 for probe design and proximity detection for targeted chromosome conformation capture
718 applications. Bioinformatics *34*, 675–677. 10.1093/bioinformatics/btx625.

719 36. Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G., Lu, W., Riley-
720 Gillis, B.M., Tsai, E.A., Kim, H.I., Zheng, X., et al. (2022). Systematic single-variant and
721 gene-based association testing of thousands of phenotypes in 394,841 UK Biobank
722 exomes. Cell Genomics *2*, 100168. 10.1016/j.xgen.2022.100168.

723 37. Westbury, S.K., Turro, E., Greene, D., Lentaigne, C., Kelly, A.M., Bariana, T.K., Simeoni, I.,
724 Pillois, X., Attwood, A., Austin, S., et al. (2015). Human phenotype ontology annotation and
725 cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and
726 platelet disorders. Genome Med. *7*, 36. 10.1186/s13073-015-0151-5.

727 38. Greene, D., NIHR BioResource, Richardson, S., and Turro, E. (2016). Phenotype similarity
728 regression for identifying the genetic determinants of rare diseases. Am. J. Hum. Genet. *98*,
729 490–499. 10.1016/j.ajhg.2016.01.008.

730 39. Rowland, B., Venkatesh, S., Tardaguila, M., Wen, J., Rosen, J.D., Tapia, A.L., Sun, Q.,
731 Graff, M., Vuckovic, D., Lettre, G., et al. (2022). Transcriptome-wide association study in
732 UK Biobank Europeans identifies associations with blood cell traits. Hum. Mol. Genet. *31*,
733 2333–2347. 10.1093/hmg/ddac011.

734 40. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to
735 variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc.
736 Series B Stat. Methodol. *82*, 1273–1300. 10.1111/rssb.12388.

737 41. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple
738 causal variants. PLoS Genet. *17*, e1009440. 10.1371/journal.pgen.1009440.

739 42. Genovese, G., Rockweiler, N.B., Gorman, B.R., Bigdeli, T.B., Pato, M.T., Pato, C.N.,
740 Ichihara, K., and McCarroll, S.A. (2024). BCFtools/liftover: an accurate and comprehensive

741 tool to convert genetic variants across genome assemblies. Bioinformatics *40*.
742 10.1093/bioinformatics/btae038.

743 43. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C.,
744 and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic
745 association studies using summary statistics. PLoS Genet. *10*, e1004383.
746 10.1371/journal.pgen.1004383.

747 44. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J.
748 (2017). 10 years of GWAS discovery: biology, function, and translation. Am. J. Hum. Genet.
749 *101*, 5–22. 10.1016/j.ajhg.2017.06.005.

750 45. Schmid, K.T., Höllbacher, B., Cruceanu, C., Böttcher, A., Lickert, H., Binder, E.B., Theis,
751 F.J., and Heinig, M. (2021). scPower accelerates and optimizes the design of multi-sample
752 single cell transcriptomic studies. Nat. Commun. *12*, 6625. 10.1038/s41467-021-26779-7.

753 46. Taylor-Weiner, A., Aguet, F., Haradhvala, N.J., Gosai, S., Anand, S., Kim, J., Ardlie, K.,
754 Van Allen, E.M., and Getz, G. (2019). Scaling computational genomics to millions of
755 individuals with GPUs. Genome Biol. *20*, 228. 10.1186/s13059-019-1836-7.

756 47. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N.,
757 Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: genome-wide integration of
758 enhancers and target genes in GeneCards. Database (Oxford) *2017*.
759 10.1093/database/bax028.

760 48. Collins, R.L., Glessner, J.T., Porcu, E., Lepamets, M., Brandon, R., Lauricella, C., Han, L.,
761 Morley, T., Niestroj, L.-M., Ulirsch, J., et al. (2022). A cross-disorder dosage sensitivity map
762 of the human genome. Cell *185*, 3041-3055.e25. 10.1016/j.cell.2022.06.036.

763

764 **Acknowledgments**

765 This research has been conducted using the UK Biobank Resource under Application Number
766 47976. The authors would like to thank Angli Xue for valuable discussions and for providing the
767 OneK1K raw VCF file and cell-type classifications.

768 This work was supported by funding from the National Institutes of Health R01MH106842,
769 24HG012090, R01HL168247 and R01HG012790, the European Research Council (ERC) under
770 the European Union's Horizon 2020 research and innovation programme (Grant agreement no.
771 101043238), and Knut and Alice Wallenberg Foundation grant KAW 2023.0337. A part of the
772 computations were enabled by resources in project (SNIC 2022-5-560 and NAISS 2023-5-517)
773 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at

774 UPPMAX, funded by the Swedish Research Council through grant agreement no. 721 2022-
775 06725.

776

777 **Author contributions**

778 S.G., J.A.M., and T.L. conceived and designed the project. S.G. performed most of the data
779 analysis. J.P. and S.G. performed power analyses and advised on analysis methods. W.O.
780 recalled eQTLs from OneK1K. J.A.M., S.G., and T.L. provided feedback on the study design
781 and data analysis. J.A.M. and T.L. supervised the work. S.G. and T.L. wrote the manuscript with
782 contributions from other authors. All authors read, provided feedback, and approved the
783 manuscript.

784

785 **Competing interests**

786 T.L. is an advisor with equity at Variant Bio and has received speaker fees from AbbVie. N.E.S.
787 is an advisor to Qiagen and a co-founder and advisor of OverT Bio and TruEdit Bio. The
788 remaining authors declare no competing interests.

789

790 **Data availability**

791 GWAS CRE and eGene and cGene target results are available within the supplementary
792 tables
793 of this manuscript. The fine-mapped credible sets and log-Bayes factors for the UKBB,
794 GTEx,
795 and OneK1K will be available for download upon publication at:
796 https://drive.google.com/drive/folders/1UdnlUs-
797 xtutWk_hSofk1iX0tnS64P9dO?usp=drive_link.

798 The summary statistics of the recalled eQTLs of the OneK1K dataset will also be available
799 upon
800 publication at:
801 https://drive.google.com/drive/folders/1Q3Zz4EFvvgn4gYlIBJQl-
802 q8605AYOhSO?usp=drive_link.
803 Reanalyzed CRISPRi data URLs can be found within data S3. We obtained
804 summary statistics and log-Bayes factors for eQTL studies from the eQTL catalog [13].
805 GWAS summary statistics for 44 blood cell traits were downloaded from the Neale Lab
806 (http://www.nealelab.is/uk-biobank/) and Lettre Lab
807 (http://www.mhi-humangenetics.org/en/resources/). The datasets used in this study are
808 publicly
809 available from various sources. GeneHancer v5.10 provides enhancer size and the
810 number of enhancers per gene, this dataset is available upon request from GeneCards:
811 https://www.genecards.org/Download/File?file=GeneHancer_v5.10.gff.
812 GTEx v8 transcript-level TPM data was obtained from GTEx Analysis v8 and can be
813 accessed
814 at
815 https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEx_Analysis_2017-06-
816 05_v8
817 _RSEMv1.3.0_transcript_tpm.gct.gz. The pHaplo and pTriplo scores, which estimate the
818 probability of a gene being haploinsufficient or triplosensitive, along with Episcore values,
819 were
820 obtained from https://ars.els-cdn.com/content/image/1-s2.0-S0092867422007887-
821 mmc7.xlsx.
822 The pLI and LOEUF scores were retrieved from gnomAD v3 and are available at
823 https://gnomad.broadinstitute.org/. Hi-C data for enhancer-promoter interactions in blood cell

824 lines, including K562, was obtained from

825 https://kth-my.sharepoint.com/:f:/g/personal/pelinak_ug_kth_se/Eq0b5SU6LGVFj4My7hiQi

826 wUB5gmQuHEL-unZRrBCRa19ig?e=MinXnm.

827

828 **Code availability**

829 CRISPRi power calculations are available at:

830 https://github.com/jasperpanten/sc_crispr_simulations

831 All other analysis codes are available at: https://github.com/PandaPowell/CRISRPi-eQTLs-

832 Manuscript

833