

RESEARCH

Open Access



Insights from pooled CRISPRi single-cell screens in K562 cells reveal gene functions, regulatory networks, and highlight opportunities and limitations

Hong Zhang¹, Peifen Zhang², Eric Bindels³ and Eskeatnaf Mulugeta^{1*}

Abstract

Pooled CRISPR screening combined with single-cell RNA sequencing (scRNA-seq) has emerged as a powerful strategy for dissecting gene function and reconstructing gene regulatory networks (GRNs) in complex biological systems. This approach enables high-throughput, parallel perturbation of multiple genes while providing transcriptome-wide readouts at single-cell resolution, overcoming many limitations of traditional arrayed screens. However, its broader application remains limited by technical challenges, including variable perturbation efficiency and difficulties in accurately identifying perturbed cells.

In this study, we adapted and applied a modified CRISPR droplet sequencing (CROP-seq) protocol using CRISPR interference (CRISPRi) in K562 cells to knockdown six transcription factors (TFs): LMO2, TCF3, LDB1, MYB, GATA2, and RUNX1. Our modified approach, which allows direct capture of sgRNAs from the cDNA library without a separate enrichment step, significantly improved sgRNA assignment per cell. We successfully achieved reproducible knockdown of three TFs (MYB, GATA2, and LMO2), captured the impact of these perturbations on the TF target genes, and enabled us to reconstruct their GRNs and identify key regulons and transcriptional targets. These networks revealed both previously established (such as LMO2 GATA2 interaction) and novel regulatory interactions, which we independently validated, providing new insights into hematopoietic transcriptional control. To assess the efficiency of CRISPRi based pooled perturbation, we additionally analyzed publicly available Perturb-seq CRISPRi datasets and found that only ~40–50% of targeted genes led to effective knockdown, underscoring the variability in perturbation efficiency across experiments.

Together, our results demonstrate both the potential and the current technical limitations of pooled CRISPRi-based single-cell screens. While this integrated approach holds great promise for high-resolution functional genomics, further optimization and standardized benchmarking are essential to improve its reliability, scalability, and reproducibility.

Keywords Pooled CRISPR screening, CRISPR interference (CRISPRi), CROP-seq, single-cell RNA sequencing, gene regulatory networks, transcription factors, hematopoietic regulation, functional genomics

*Correspondence:
Eskeatnaf Mulugeta
e.mulugeta@erasmusmc.nl

¹Department of Internal Medicine, Erasmus University Medical Center
Rotterdam, Rotterdam 3015 CN, The Netherlands

²Department of Developmental Biology, Erasmus University Medical
Center Rotterdam, Rotterdam 3015 CN, The Netherlands

³Department of Hematology, Erasmus University Medical Center
Rotterdam, Rotterdam 3015 CN, The Netherlands



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Genetic screening is a powerful technique for systematically analyzing gene function, reconstructing gene regulatory networks (GRNs), and understanding biological mechanisms [1, 2]. These screens are typically conducted in either an individual (“arrayed”) format, where each genetic perturbation is introduced and analyzed separately, or a pooled format, where all perturbations occur simultaneously within a population [2–5]. While both approaches have distinct advantages, recent advances in pooled screening have greatly enhanced the ability to evaluate gene function at a genome-wide scale in mammalian cells [1–3, 6–11]. Arrayed screens allow for more detailed molecular readouts, such as RNA sequencing, but at the cost of lower throughput and a highly labor-intensive process, as individual gRNAs must be physically separated [12–14]. In contrast, pooled screening is more efficient, scalable and overcomes the shortcomings of arrayed screens, and had been instrumental [5, 15, 16]. However, a key limitation of pooled screening is its reliance on relatively simple phenotypic readouts, such as cell viability or the expression of a few exogenous reporters, which provide population-averaged measurements rather than cell-specific insights [17–19]. Additionally, pooled screens are not well suited for complex molecular analyses like transcriptome profiling, which is one of the most comprehensive measures of cellular response [20, 21]. Balancing the strengths of both screening approaches is crucial for refining genetic screening methodologies and advancing our understanding of gene function in complex biological systems.

In recent years, with the advancements of CRISPR techniques and single-cell level quantification approaches, pooled CRISPR screening has emerged as a powerful and widely used method, addressing the shortcomings of the traditional pooled screening approach. These approaches allow detecting each cell’s perturbation status along with the corresponding single-cell phenotype at different omics layers [2, 3, 6, 7]. The initial pooled CRISPR screenings approach combines CRISPR-based perturbation to gene expression signatures for individual gene perturbation in a complex pool of cells [2, 3, 7]. These revolutionary experiments have made significant advances in comprehending the functions of hundreds of genes simultaneously, and helped in uncovering their cellular impact, deciphering gene regulatory networks, and shedding light on several biological mechanisms. Recent developments are now being enhanced to encompass all major omics techniques, including chromatin accessibility, protein concentrations, and imaging approaches that provide novel insights into cell phenotypes [22–27].

Although the initial approaches were based on CRISPR knockout (CRISPR KO) techniques, recent advancements in CRISPR approaches offer a range of options, including

gene knockdown (CRISPR interference, CRISPRi) and activation (CRISPRa) of the genes of interest [28–30].

CRISPR KO generates permanent gene disruption by introducing small insertions or deletions (indels) that cause a frameshift mutation, and is widely used in genome-wide screens to identify essential genes and studying loss-of-function mutations. In contrast, CRISPRi reduces but does not eliminate gene expression, making it ideal for studying essential genes where a complete knockout would be lethal and for reversible gene modulation. In addition, CRISPRi is advantageous in avoiding genetic variability during indel formation and non-specific toxicity due to DNA cutting [31, 32]. However, despite this advantage, CRISPRi has also its own shortcomings as it requires continuous expression of dCas9 (dead Cas9) and guide RNAs to maintain gene inactivation. These CRISPR-based methods are now being integrated with single-cell approaches to capture cell states influenced by specific perturbations [33–35]. CRISPRi has been successfully applied in combination with single-cell omics approaches. The first pooled CRISPRi perturbation, Perturb-seq [2], used barcodes to identify cells expressing specific gRNAs. Subsequently, other approaches emerged that utilized sgRNA capture and used them as barcodes (CROP-seq) [3]. Despite this development, similar to other pooled perturbations approaches, the field is still emerging. Several improvements and standardizations from experimental to analytical points are still needed to be implemented.

Here, we coupled CROP-seq with CRISPRi to target six transcription factors (TFs) — LMO2, TCF3, LDB1, MYB, GATA2 and RUNX1 in K562 cells. We chose CRISPRi over CRISPR knockout because many TFs, such as GATA2, are essential genes. To investigate the role of GATA2 in cells, complete KO of GATA2 is not an appropriate option; KD is a more suitable approach. We used a modified but well-established CROP-seq-CRISPRi vector (pBA950) [28], which improves CRISPRi repression efficiency by including an optimized sgRNA(E+F) scaffold. Importantly, this vector is still fully compatible with direct sgRNA capture by the 10x Genomics 5′ CRISPR Capture kit. In addition, we employed dCas9-KRAB-ZIM3 to further enhance repression efficiency. Using this approach, we achieved efficient and reproducible knockdown for only three TFs (MYB, GATA2, and LMO2). This enabled us to map their transcriptional targets and reconstruct gene regulatory networks (GRNs), uncovering both known and novel regulatory relationships, with key regulons independently validated. To assess the generalizability of our approach, we analyzed publicly available pooled CRISPRi single-cell screen datasets and found that a substantial fraction of sgRNAs failed to induce effective knockdown, highlighting the variability in perturbation efficiency.

Our findings demonstrate that CRISPRi-CROP-seq is a promising strategy for dissecting gene function and regulatory circuitry at single-cell resolution. However, improvements in perturbation consistency, sgRNA design, and experimental optimization are still necessary to enhance its robustness and scalability for broader applications in functional genomics.

Results

Pooled CRISPRi screen with direct 5' CRISPR capture sequencing

In order to assess the effectiveness of pooled CRISPR-based gene inactivation (CRISPRi) and capture the subsequent impact on phenotype at the single-cell transcriptome level, we used a modified version of CROP-seq [28, 36]. This modified CROP-seq (CRISPRi-CROP-seq) contains a scaffold sequence (sgRNA^{E+F}, which combines an A-U flip with a 5 bp hairpin extension) that can improve the knockdown (KD) efficiency and can be directly captured using the 10X genomics “Single Cell 5' CRISPR Screening Assay” (Figure 1A). In the final cDNA library of this assay, two distinct cDNA libraries are simultaneously produced: the mRNA library and the

sgRNA library. This direct capture approach is suggested to enhance sgRNA capture sensitivity, as it enables direct capture of both polymerase II and III transcripts, thereby enhancing the accuracy of the results. For this experiment, we selected 7 transcription factors (TFs) known to play key role in the pathogenesis of acute myeloid leukemia (AML) based on: the analysis of a patient cohort [37], matching with TCGA data [38], and a thorough literature search. These TFs also show high expression levels in AML patients. The selected TF are LMO2, TCF3, LDB1, MYB, GATA2, and RUNX1 which are all expressed in K562 cells. Additionally, FLI1, is included as a negative control, since it is not expressed in K562 cells. To conduct a comprehensive and thorough investigation of the CRISPRi-CROP-seq, we decided to limit the number of genes in the pool to these seven genes and generate a detailed output.

Aiming to inactivate these genes in a pool, we selected 26 sgRNAs: 3 gRNA per gene from the Dolcetto library [39], and 5 non-targeting sgRNAs from literature [3]. After cloning the sgRNA pool into a lentiviral vector, we confirmed the presence of each gRNA by sequencing the lentiviral vector plasmid library (Supplementary Figure

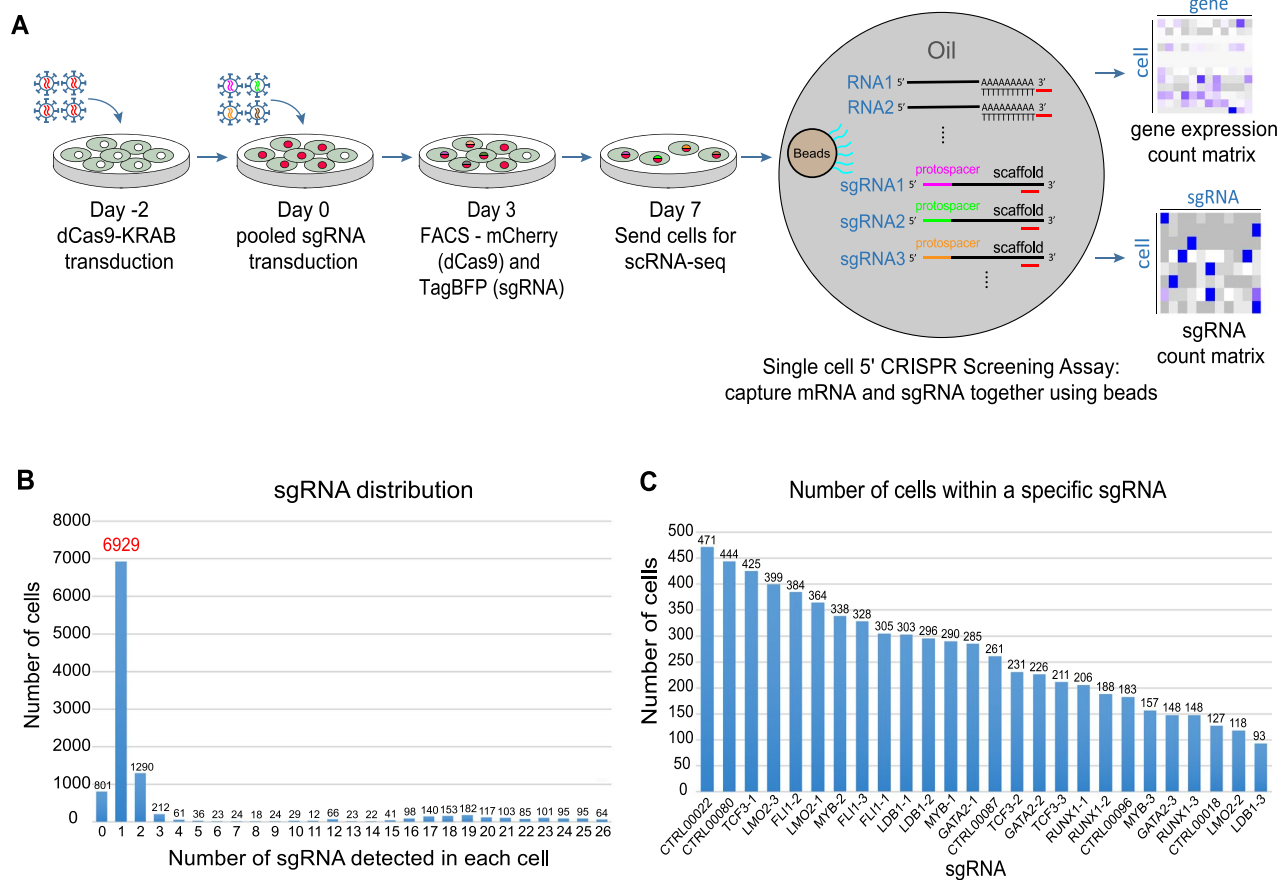


Fig. 1 Overview and result of direct 5' CRISPR capture sequencing. **A** Workflow of the pooled CRISPRi screen with Single Cell 5' CRISPR sequencing in K562 cells. **B** Distribution of the total number of sgRNAs detected per cell. **C** Number of cells assigned to each sgRNA

1 A). Subsequently, the sgRNAs were packaged into lentiviruses, transduced into K562 cells, and the presence of each gRNA was confirmed by sequencing amplified sgRNA sequences extracted from the genomic DNA (Supplementary Figure 1B). Next, we transduced these pooled sgRNA, after packaging lentivirus vector, into K562 cells at day 0 (D0), which had been transduced with dCas9 -KRAB-ZIM3 lentivirus two days earlier (D-2). After Fluorescence-activated cell sorting (FACS), sorting double-positive cells at D3, using TagBFP (for cells with sgRNA) and mCherry (for dCas9-KRAB), we performed single cell sequencing at D7 using the 10X genomics platform (Figure 1A).

After sequencing, we recovered 10848 cells for sequencing with two count matrices: the mRNA matrix and the sgRNA matrix. For the gene expression count matrix, we generated a total of 417,3 M reads, with an average of 38,477 reads per cell and a median of 4,623 genes per cell. For the sgRNA matrix (capturing sgRNA), we generated 94,9 M reads, and around 85% of them were usable (reads with a recognized protospacer sequence; a valid unique molecular identifier (UMI); and a cell-associated barcode). The median sgRNA UMIs per cell is 3535. After assigning sgRNA to cells, we found 6929 cells (63.9%) harboring only one sgRNA (Figure 1B), 1290 cells (11.9 %) carrying two sgRNA, and 212 cells (2%) with three sgRNA. Additionally, we also observed a relatively few numbers of cells carrying more than 3 sgRNA, and 801 (7.4%) cells without any sgRNA (with protospacer sequence identified as background). When we assessed the number of cells that are associated with a specific sgRNA, we observed that the highest number was 471, while the minimum count was 93 (Figure 1C), which is sufficient for subsequent analyses. Furthermore, these results demonstrate that the “CRISPR Screening Assay” enables the capture and sequencing of sgRNA with high quality and efficiency.

To conclude, the modified CROP-seq (CRISPRi-CROP-seq) enables the cloning and transduction of gRNAs in a pool. Moreover, it allows direct capture of sgRNA from the cDNA library without the need for a separate sgRNA enrichment step. Furthermore, in the cell line we tested (K562), after sequencing, we recovered a significantly high number of cells with a single sgRNA, which is crucial for subsequent analysis steps.

Efficiency of gene inactivation in single cell experiments is not efficient

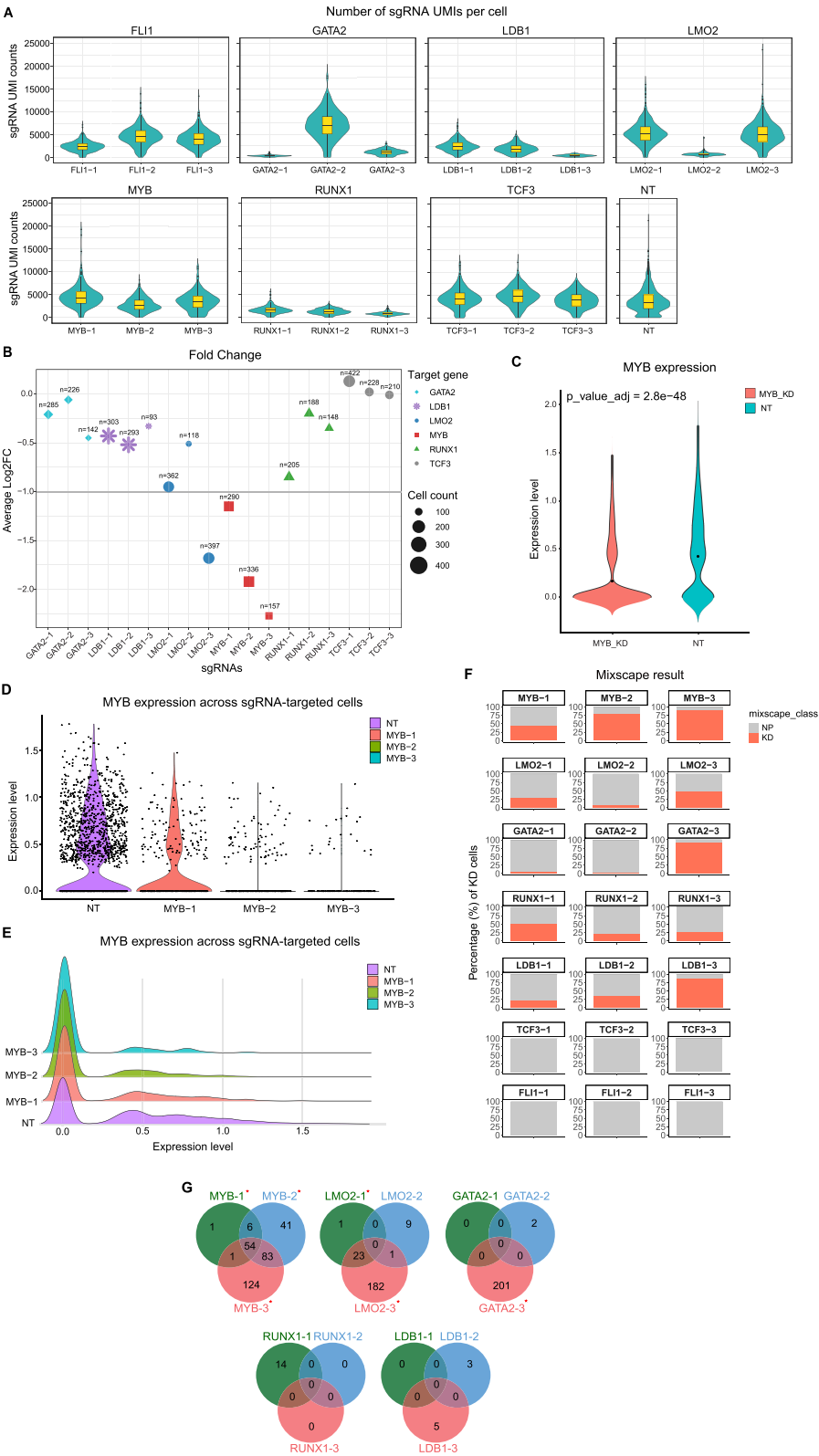
After confirming the ability of the modified CRISPRi-CROP-seq to successfully deliver one sgRNA per cell, in majority of cells, and capture the transcriptome profile in single-cells together with the sgRNA, we next evaluated the efficiency of gene inactivation. To start the analysis, we first determined the expression level of the gRNA

in the cells where it was identified - to ensure that the sgRNA is expressed in the cells where it was delivered and captured after single-cell sequencing. For all the captured sgRNA, we observed the expression of the sgRNAs with a median UMI count of 3535 (Figure 2A). In this analysis, we also noticed that even for the same gene, the number of captured sgRNAs per cell varied among the three different sgRNAs, and similar variability was also evident among sgRNAs targeting different genes (Figure 2A).

To understand the KD (knockdown) efficiency on the sgRNA level, we calculated their FC (Fold Change) and adjusted *P* value (*P*_val_adj) separately. The result (Figure 2B, Supplementary Table 1) showed that only 4 sgRNAs (MYB-1, MYB-2, MYB-3, LMO2-3) achieve efficient KD (with $\log_2\text{FC} < -1$ & *P*_val_adj < 0.05). To further visualize these findings, we generated violin plots at both the sgRNA and gene levels (Figure 2C-E, Supplementary Figure 2). The results revealed considerable variability among sgRNAs targeting the same gene.

To gain a more comprehensive view we then used the Mixscape computational framework that enhances signal-to-noise ratio in single-cell perturbation screens, and enable the separation of a cell mixture into perturbed and non-perturbed cells [26]. We performed Mixscape analysis at sgRNA level, to show the impact of each sgRNA on target gene expression (cells with each sgRNA) and found that 15 out of the 21 sgRNA perturbed (KD) the target gene in varying number of cells (Figure 2F). For some sgRNA that target the same gene, we observed a large variability. For example, for the sgRNAs targeting GATA2, the variability of cells, where the sgRNA targeting GATA2 were captured and the target gene was perturbed, varied between 5% to 90% (Figure 2F). For other genes, such as MYB, this variability is limited (Figure 2F). For other genes, such as TCF3, despite the sgRNAs is expressed (Figure 2A) and also captured, we could not identify cells where the gene is perturbed. For FLI1, that we used as a negative control as it is not expressed in the K562 cells, we were able to capture cells where the sgRNA targeting the gene is present, but as expected did not show perturbation (Figure 2F). This raises a question about the efficiency of the system. Even for genes that have a significant perturbation, when the sgRNAs were pooled, there is a large variability between the efficiency of sgRNAs (Figure 2B, F).

To further understand the contribution of each sgRNA, we tested if each of the 3 separate sgRNAs targeting the same gene produce similar differentially expressed gene (DEG) patterns in cells. To test this, for each individual sgRNA, we performed differential expression analysis by comparing cells showing efficient KD with cells carrying a non-targeting (NT) sgRNA. In these analyses, we observed that different sgRNAs generate different sets of



DEGs (Figure 2G). The overlap of DEGs by these sgRNA varies per targeted gene and, in some cases, this overlap is dependent on the efficiency of the sgRNA. For sgRNAs targeting the MYB gene, we observed the highest DEGs overlap between the sgRNAs targeting it (Figure 2G), suggesting that all three MYB sgRNAs are effective. For LMO2, the sgRNA LMO2-3 yielded the highest number of DEGs (Figure 2G). In contrast, the sgRNA LMO2-1 and LMO2-2 have a relatively low number of DEGs. Notably, there is a substantial overlap in DEGs between LMO2-1 and LMO2-3, suggesting that LMO2-1 is relatively efficient. However, LMO2-2 shows minimal overlap with the other two sgRNAs, indicating low efficiency. In the case of GATA2, only one sgRNA (GATA2-3) generated 201 DEGs. The other sgRNAs targeting RUNX1 and LDB1 did not yield significant results, and we only obtained a very low number of DEGs. In general, based on this differential gene expression analysis, only 6 out of the 18 sgRNAs had a significant impact.

To further investigate the impact of these perturbations, we focused on the 3 genes (MYB, GATA2 and LMO2) that show significant perturbation. We extracted all the KD cells from these 6 efficient KD sgRNAs and combine them on the gene level. We first clustered the cell population and plotted them using Uniform Manifold Approximation and Projection (UMAP), together with the cells with control sgRNAs that does not target any gene (NT), and we observed a clear separation of the cell where we perturbed the genes forming distinct clusters (Figure 3A), indicating the impact of the perturbation. We then performed differential expression analysis and plotted these DEGs (cutoff: $|\log FC| > 0.25$ adjusted p -value < 0.05), and identified DEGs (914 for MYB, 1364 for LMO2, 236 for GATA2, see Supplementary Table 2). The DEGs were visualized using a heatmap and showed a clear clustering of the cells based on the expression of these DEGs (Figure 3B, supplementary Figure 3). Among them, we found MYADM, STAT5A, LMNB1, TMX2, GATA2, and MYC, all of which have been previously reported as MYB target genes [40, 41].

We next validated these observations using RT-qPCR to confirm the downregulation of the targeted gene by the sgRNA and also the DEGs identified in the single-cell experiment following similar CRISPRi experimental protocol we applied for the single-cell experiment. We perturbed (KD) targeted genes separately by transfecting the sgRNA plasmid into k562 cells carrying dCas9. After sorting cells using Fluorescence-activated Cell Sorting (FACS) on day 2 and RNA extraction on day 3, we performed RT-qPCR to quantify the sgRNA targeted gene expression and DEG expression for each sgRNA. In this validation, we observed the expected downregulation of the target gene as observed in the single cell expression analysis (Figure 3C, Supplementary Figure 4A-B).

To validate our results, for MYB gene, we selected 5 DEGs that were upregulated (HBE1, HBZ, ALAS2, GYPB, HBA2) and 5 DEGs were downregulated (NMU, RP11-301G19.1, PDE4DIP, KCNH2, FAM178B) in the single-cell experiment. MYB is a key TF in erythropoiesis and hematopoietic stem cell regulation and cancer [42]. In this validation experiment, similar to what we observed at the single-cell RNA level, we observed comparable efficiency for each sgRNA targeting MYB (Figure 3C). Additionally, we validated the expression changes in these MYB target DEGs using RT-qPCR, supporting the impact of MYB inactivation observed in the single-cell experiment (Figure 3D, 3E). We performed a similar analysis for GATA2 and LMO2, two key transcription factors involved in hematopoiesis, vascular development, and immune system function [43, 44]. For GATA2, we selected the 10 DEGs, with five upregulated genes (ID3, APOE, PYCARD, CD52, and SPI1) and five downregulated genes (PRDX1, SNAP29, PAFAH1B3, HBD, and KRT19) (Supplementary Figure 4C-D). Similarly, for LMO2, we validated the expression of 10 DEGs: five upregulated genes (ID3, CD24, VIM, APOE, and SNCG), and five downregulated genes were HBZ, HBG1, GAL, KLF1, and HMBS (Supplementary Figure 4 E-F). In all instances, we were able to validate the perturbation of the sgRNA target genes (MYB, GATA2 & LMO2) and the DEGs using RT-qPCR.

We next performed gene enrichment analysis using the DEGs, and observed that for genes that are downregulated upon MYB perturbation, we observed enriched terms related to ribonucleoprotein complex biogenesis, metabolism of RNA, translation and protein maturation. For upregulated genes the enriched terms included erythrocyte development, porphyrin-containing compound metabolic process and regulation of hemoglobin biosynthetic process (Supplementary Figure 5A). This is consistent with previous findings on the MYB gene, which show that high MYB expression can inhibit cell differentiation in erythroid cell lines [42] and is essential for malignant self-renewal [45]. To further validate our findings, we analyzed a MYB ChIP-seq dataset performed on K562 cells [40]. The analysis revealed clear MYB occupancy around transcription start sites (TSS, ± 5 kb) of downregulated DEGs, whereas no significant enrichment was observed around the TSS (± 5 kb) of up-regulated DEGs (Supplementary Figure 5B). This is also consistent with the established role of MYB as a transcriptional activator [46]. In a similar fashion, we performed enrichment (Supplementary Figure 6A) and ChIP-seq (Supplementary Figure 6B) analysis [47] for GATA2-regulated DEGs. After enrichment analysis, we found that the downregulated GATA2 DEGs were mainly enriched in ribonucleoprotein complex biogenesis, PID MYC activation pathway, chaperone-mediated protein folding and cell

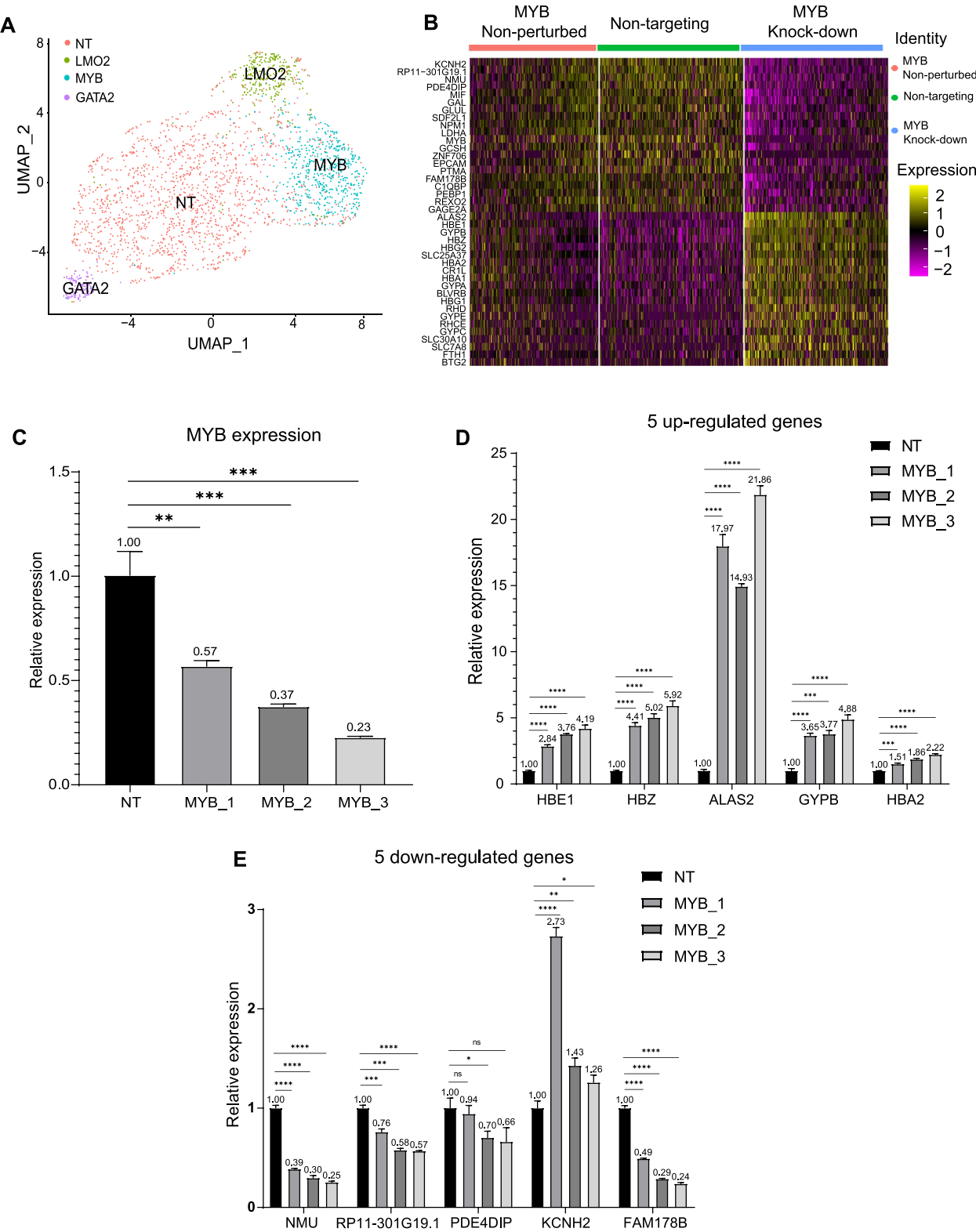


Fig. 3 Validation of CRISPRi-CROP-seq perturbation results. **A** UMAP plot showing effective gene knockdown cells across all six sgRNAs and non-targeting (NT) control cells. **B** Heatmap comparing MYB knock-down (KD) and non-perturbed (NP) cells (as estimated by Mixscape) to NT cells. **C** RT-qPCR analysis in K562 cells of MYB expression in MYB KD versus NT cells. **D** RT-qPCR validation of five upregulated DEGs identified in MYB KD cells compared to NT controls. **E** RT-qPCR validation of five downregulated DEGs following MYB knockdown. Data are presented as mean $\pm$ SD from three biological replicates ($n = 3$). Statistical significance was determined using two-tailed Student's t-test. Significance levels are indicated as follows: ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

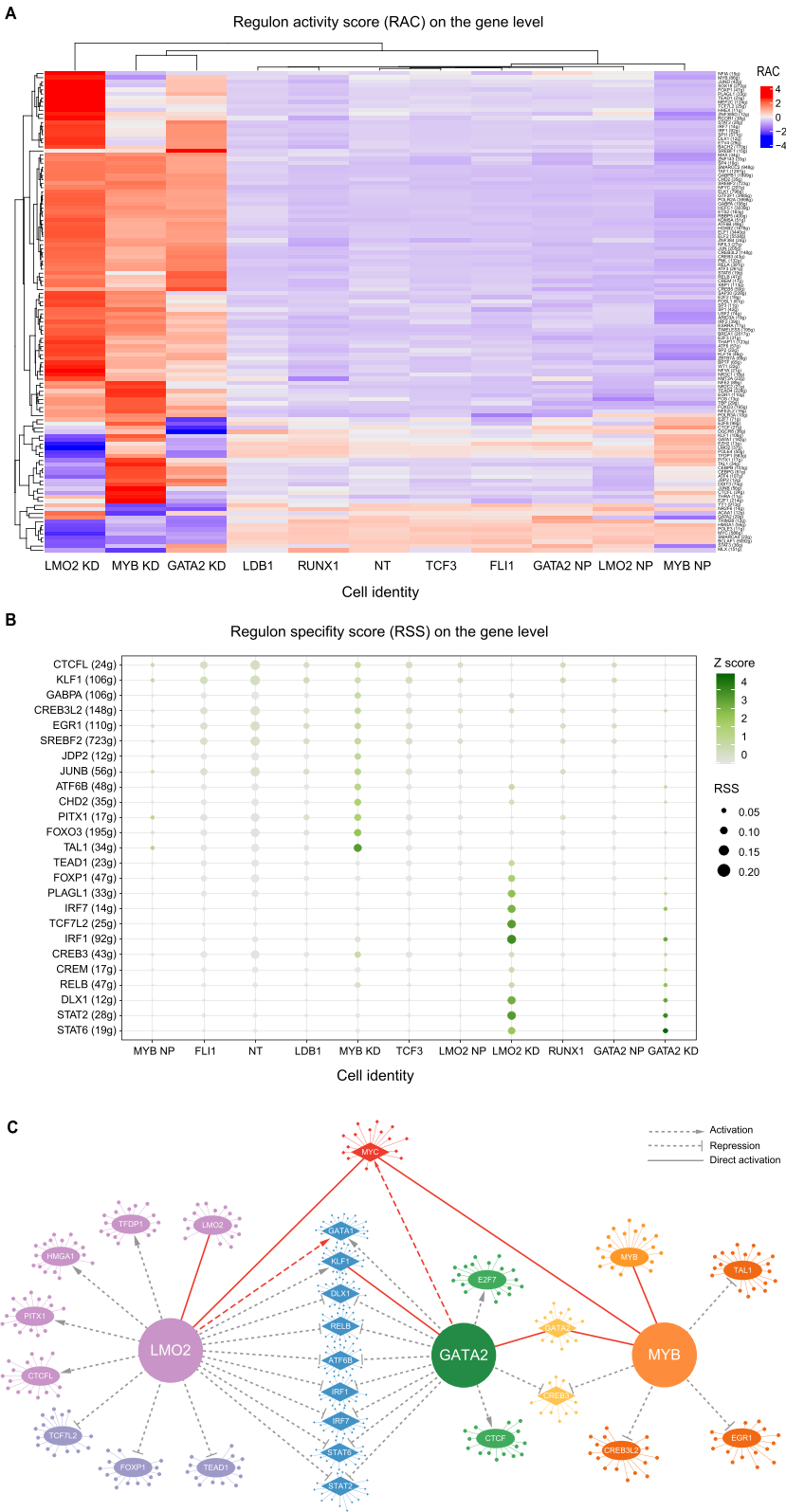


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Gene regulatory networks. **A** Regulon activity at perturbed gene (gene level summary, an average of 3 sgRNAs' KD cells). The X-axis represents cell identity (based on sgRNA that is identified), while the Y-axis represents regulons. "NT" refers to a combination of five non-targeting sgRNAs. **B** Regulon-specific score at the gene level. A higher Z-score indicates greater specificity of that regulon for the given cell group. **C** Gene Regulatory Networks based on this current sgRNA screen, and solely focused on LMO2, GATA2 and MYB. Grey dotted lines indicate putative regulatory connections (activation or repression). The red dotted line represents a previously reported indirect connection. Solid lines indicate direct regulatory interactions; among them, red solid lines correspond to interactions supported by published evidence. Except for the central genes LMO2, GATA2, and MYB, all other genes are regulons. For example, TAL1 is repressed by MYB (either directly or indirectly) and directly activates many genes (represented by the small dots connected to TAL1). Note that the number of small dots is symbolic and does not reflect the actual number of target genes

population proliferation (Supplementary Figure 6A). For the upregulated genes, they were enriched in cholesterol metabolism, maintenance of location, lysosome and neutrophil degranulation (Supplementary Figure 6A). When we analyzed GATA2 ChIP-seq data (Supplementary Figure 6B) that was generated on K562 cells, and mapped the occupancy around TSS of the up and down regulated genes, we observed clear enrichment for upregulated genes compared to downregulated genes.

In conclusion, we successfully perturbed (KD) three TFs—MYB, GATA2, and LMO2—out of the six TFs we had targeted in K562 cells, captured the impact of these perturbations on their target genes, and verified the results. However, despite this moderate success, these results indicate that, the efficiency of the CRISPRi-CROP-seq experiment depends on several aspects that need several considerations.

Impact of perturbations on gene regulatory networks (GRNs)

Perturbation of TFs is expected to alter GRNs, and since we observed significant changes in gene expression after knocking down MYB, GATA2, and LMO2, we investigated the impact of these perturbations on GRNs to better understand their effects. We applied SCENIC [48], a tool that infers GRNs from scRNA-seq data. In this analysis, we identified a total of 118 high-confidence regulons (a group of genes regulated by the same regulatory protein). We first analyzed regulon activity (regulon activity score, RAC) at both the sgRNA level (Supplementary Figure 7A) and gene level (Figure 4A). At the sgRNA level, most efficient sgRNAs exhibited distinct high or low regulon activity, except for MYB-1, which showed a similar but weaker activity pattern compared to MYB-2 and MYB-3. This result aligns with Fold Change and Mixscape analyses (Figure 2B and F), and confirming MYB-1 is less effective than MYB-2 and MYB-3. At the gene level (grouping sgRNA targeting the same gene), KD cell groups displayed distinct high- or low-activity regulons. As expected, MYB, LMO2, and GATA2 regulons showed reduced activity in their respective MYB, LMO2, and GATA2 KD groups, validating the specificity and efficiency of the knockdowns. To quantitatively assess how specifically each regulon is active in particular cell clusters, we performed regulon specificity (shown as regulon-specific score, RSS) analysis across all clusters

(Figure 4B). The results revealed that each KD cluster had distinct high-specificity and high-activity regulons, as indicated by high Z-scores and RSS. For example, the STAT6 regulon exhibited a significantly higher Z-score in the GATA2 KD cluster compared to other clusters, suggesting a strong activation following GATA2 KD, a pattern not observed in MYB KD or LMO2 KD.

Next, for each cluster (MYB KD, GATA2 KD, and LMO2 KD), we screened important regulons based on two criteria: 1) low-activity regulons ($RAC < -1$); (2) high-specificity and high-activity regulons ($RSS \text{ Z-score} > 0.5$). This initial filtering yielded 39 regulons for GATA2 KD, 40 for LMO2 KD, and 42 for MYB KD (Supplementary Table 3). To further refine the list of relevant regulons, we calculated the fold change (FC) of the core transcription factors (TFs) in each regulon (e.g., MYC in the MYC regulon), as these TFs are expected to show corresponding changes in expression. Regulons were retained only if their TF exhibited a significant expression change, defined as an absolute $\log_2$ fold change ($|\log_2 FC|$) greater than 0.5. This finally, resulted in 23 key regulons, and by leveraging the connections among these regulons, we constructed a gene regulatory network (GRN) (Figure 4C). In this network, the most strongest connection was between the LMO2 and GATA2, which is consistent with previous findings that identified LMO2 as interactor of GATA2 [49]. This network analysis not only confirmed several well-established direct gene interactions (e.g., MYB – MYC [40, 50], MYB – GATA2 [46], GATA2 – KLF1 [51], and LMO2 – MYC [52]), but also predicted indirect interactions, such as GATA2 – MYC [53] and LMO2 – GATA1 [54]. For the GATA2 – MYC interaction, one study demonstrated that pathway analysis of RNA-seq data following GATA2 deletion in AML leukemic stem cells (LSCs) revealed downregulation of MYC target genes [53]. This finding suggests that GATA2 may function as a positive regulator of the MYC pathway, rather than directly regulating MYC itself. Regarding the LMO2 – GATA1 interaction, it has been shown that LMO2 contributes to the expression of GATA1 target genes by modulating the assembly of the GATA1–SCL/TAL1 transcriptional complex at endogenous loci [54]. The presence of these known interactions further reinforces the accuracy and biological significance of our GRNs.

To further validate these inferred GRNs, we performed independent knockdown experiments for LMO2, and GATA2 in K562 cells. We then selected a subset of genes (based on the observed expression change in the GRN analysis) for RT-qPCR analysis to assess their expression changes. The results demonstrated that the majority of these genes exhibited the expected upregulation or downregulation (Supplementary Figure 7B-C). For example, LMO2 represses TFs such as IRF1, IRF7, and FOXP1, thus genes within these regulons are expected to be upregulated upon LMO2 KD. Consistent with this, we observed increased expression of ANK3 (IRF7 regulon), FN1 (IRF1), and IGF1 (FOXP1) in LMO2-repressed K562 cells. Similarly, GATA2 KD resulted in upregulation of AIF1 (STAT2 regulon), CD53 and PYCARD (IRF1), CORO1A (STAT6), and SPI1/TPM4 (CREB3), consistent with its repression role. In contrast, KRT19 (GATA1 regulon) was downregulated, as expected from GATA2's activation of GATA1. These results provide strong support for the reliability of our inferred GRNs.

In conclusion, we constructed GRNs for MYB, GATA2, and LMO2 using gene perturbation and identified key regulons and target genes regulated by these transcription factors. We were also able to validate these observations independently. Our results reveal both novel interactions and previously established ones, highlighting the power of this approach.

Pooled CRISPRi based gene perturbation and single-cell readout datasets have low efficiency

In recent years, several pooled CRISPRi single-cell screen experiments have been published and utilized to address various questions [55–57]. In the CRISPRi-CROP-seq experiment that we performed to assess the efficiency of this approach and unravel the regulatory network of the genes we targeted, the low efficiency we observed in our settings prompted us to question the reliability of other pooled CRISPRi single-cell screen experiments that have been recently published. To investigate and compare our results, we acquired two publicly available, published datasets [34, 58] that conducted pooled CRISPRi screen with single-cell sequencing experiments with direct capture sequencing, which is similar to our approach. The first dataset [34] (Dataset one, Wu et al. 2022) conducted a pooled CRISPRi single-cell screen experiment in induced pluripotent stem cells (iPSCs) using 480 single-guide RNAs (sgRNAs) to target and perturb 120 coding genes and 120 long non-coding RNAs (lncRNAs). Since our experiment focused primarily on coding genes, we retained the cells with sgRNAs targeting only coding genes and excluded those targeting lncRNAs. The second dataset [58] (Dataset two, by Joseph M.R et al. 2022), performed pooled CRISPRi single-cell screen in K562

cells using 384 sgRNAs targeting 128 genes, aiming to improve the CRISPRi efficiency by dual-sgRNA design.

We analyzed these datasets using a similar pipeline and approach used to analyze our CRISPRi-CROP-seq experiments data. First, we quantified the number of sgRNA that were used in the pool and what was obtained after the single cell sequencing. For dataset one, all the 240 sgRNA were captured, that is similar to what was published by the authors (Figure 5A). For dataset two, 384 sgRNAs, 377 were captured, also as initially reported by the authors (Figure 5A). Next, we analyzed the distribution of sgRNA per cell. For dataset one (Figure 5B), we obtained 78929 cells (69.1%) with one sgRNA, 27063 cells (23.7%) with 2 sgRNA and 8239 cells (7.2%) with 3 sgRNA. For dataset two (Figure 5C), we obtained 57989 cells (80.1%) with one sgRNA, 11799 cells (16.3%) with 2 sgRNA and 2571 cells (3.6%) with 3 sgRNA. This distribution of sgRNA is also in agreement with what the authors reported. Additionally, we also counted the median UMI for each sgRNA (Figure 5D-E), ranges from 1 to 600 in Dataset One and from 1 to 1200 in Dataset Two. This variation shows the variability between sgRNAs.

To estimate the efficiency of gene inactivation by sgRNA, we first calculated FC and P_{val_adj} for target gene of sgRNA ($\log_2FC < -1$ and $P_{val_adj} < 0.05$). Only sgRNA that passes the threshold can be identified as KD-like. For dataset one (Figure 5F), in iPSC, from the 240 sgRNA, only 133 show efficient gene inactivation, and for dataset two (Figure 5G), in K562 cells, only 185 out of the 377 sgRNA show significant gene inactivation. This indicates, in both datasets, only around half of the sgRNAs used are able to inactivate the target gene. Next, we used the Mixscape framework to determine the target gene perturbation efficiency at sgRNA level. For dataset one (Figure 5H, Supplementary Figure 8A), 162 sgRNAs were found to have efficient gene perturbation, and 104 sgRNAs for the dataset two (Figure 5H, Supplementary Figure 8B). When we plotted the impact of these sgRNA using UMAP (Supplementary Figure 8C-D), to visualize their impact and we observed that only few genes having a significant impact and differ from the sgRNA in the pool. Although this visualization is not quantitative, it demonstrates that only a few genes have a significant impact, which further confirms the low efficiency of the experiments. A large proportion of sgRNAs do not have a significant impact on gene perturbation.

In the next step, we conducted a more stringent analysis by integrating the fold change-based analysis of the target gene with the results from the Mixscape analysis. Only sgRNAs identified by both methods were classified as KD sgRNAs (Supplementary Figure 8E-F). For dataset one, in the iPSC dataset, only 97 out of 240 sgRNAs KD efficiently. For the K562 dataset, only 56 out of 377 sgRNAs KD efficiently. This again points to the

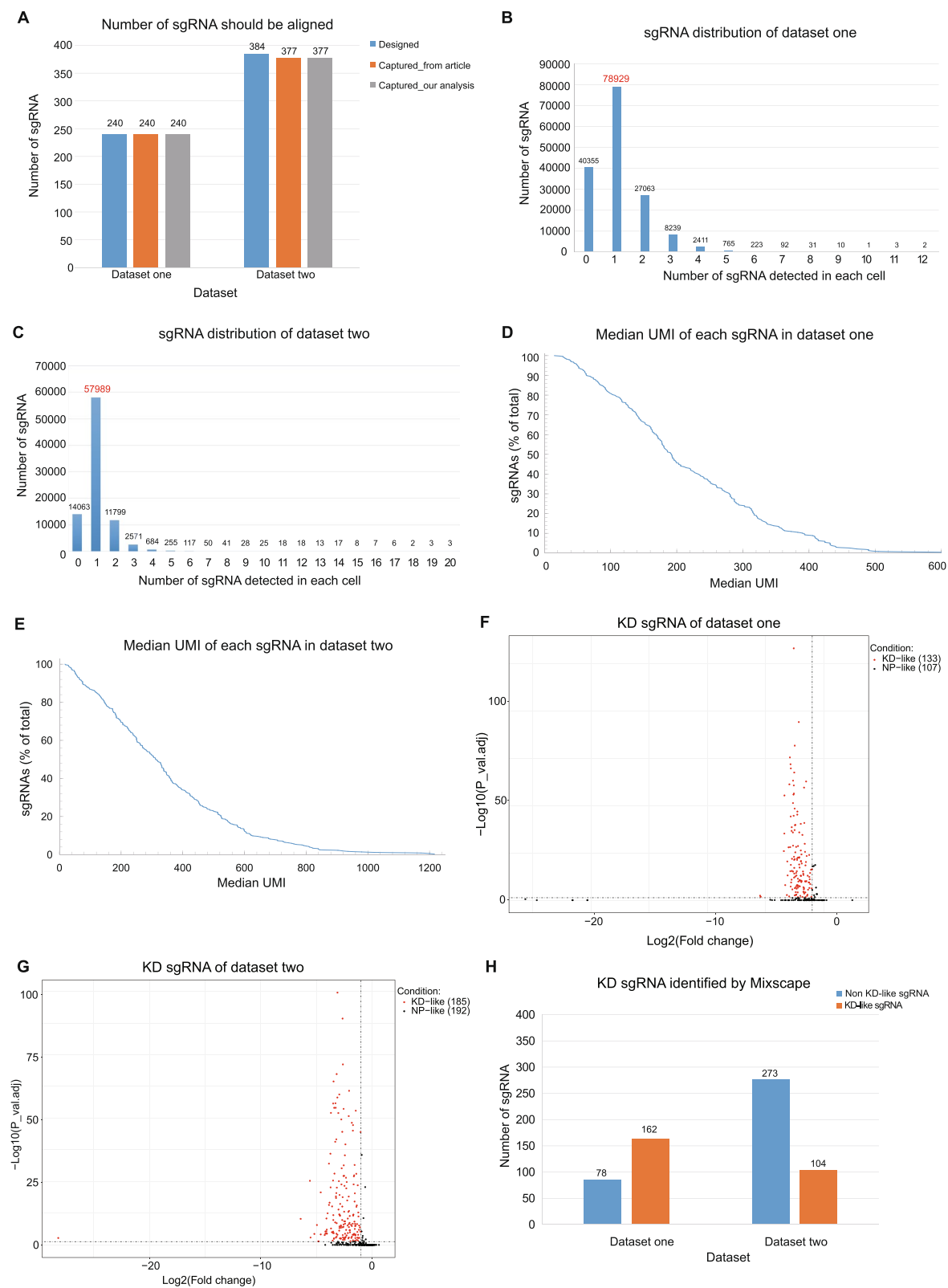


Fig. 5 Analysis of two published Perturb-seq CRISPRi datasets. **A** Number of sgRNA captured in the different analysis. **B, C** Distribution of the number of sgRNA detected in each cell for the two datasets. **D, E** Median UMI of each sgRNA for the two datasets. **F, G** Volcano plots which represent the number of sgRNA surpassing the Fold Change and p_{val_adj} thresholds in the two datasets. **H** Bar plots which represent the number of efficient KD sgRNA identified by the Mixscape program

fact that the limited efficiency of CRISPRi-CROP-seq experiments.

To conclude, in this analysis we show that, in these published datasets a significant proportion of the sgRNA does not produce significant perturbation to inactivate the targeted gene. This is in line with efficiency problems we observed in CRISPRi based single cell pooled screens.

Discussion

Massively parallel gene perturbation techniques have become indispensable for elucidating gene function, mapping gene regulatory networks (GRNs), and uncovering intricate molecular pathways. In our study, we employed CRISPRi-CROP-seq to interrogate the roles of six transcription factors (TFs), yielding detailed insights on three of them. By individually inactivating each TF, we identified their downstream target genes, reconstructed GRNs, and delineated the interrelationships among these regulators. Furthermore, we evaluated the efficiency of this powerful, but yet to be standardized, approach. Our results demonstrate that while CRISPRi-CROP-seq holds considerable promise, significant optimization and rigorous benchmarking, including a reevaluation of existing studies, are essential to fully harness its potential.

In our evaluation, we investigated the function and regulatory relationships of six TFs—MYB, GATA2, LMO2, RUNX1, LDB1, and TCF3—which were selected based on the high expression of these genes in AML patients. These factors play crucial roles in erythropoiesis, hematopoietic stem cell regulation, and cancer [52, 59–63]. While previous studies have examined these TFs in isolation, our study represents the first attempt to concurrently dissect their functions and interrelationships within a single experimental framework. This unified approach minimizes confounding variables by allowing us to assess multiple gene perturbations under identical conditions. As a result, we gained insights into the roles of three TFs—MYB, GATA2, and LMO2—demonstrating that their downstream target genes are integrally involved. For instance, MYB-associated differentially expressed genes (DEGs) were linked to biological processes consistent with its known function, and these findings were independently validated using RT-qPCR. Moreover, integrating MYB ChIP-seq data revealed significant enrichment of MYB binding at the transcription start sites of genes down-regulated in the MYB knock-down condition in our CRISPRi perturbation, further supporting its regulatory impact. In addition to identifying DEGs, we reconstructed the gene regulatory networks (GRNs) for MYB, GATA2, and LMO2 using CRISPRi perturbation data, and validated these networks by RT-qPCR. This network analysis not only confirmed known regulons but also uncovered novel regulatory circuits that illustrate the intricate interactions among these

TFs (Figure 4). One example is the predicted interaction between GATA2 and IRF1 [64]. Although no studies have directly reported this connection, one study, in acute promyelocytic leukemia (APL), showed MYC represses IRF1 expression through recruitment of the PML-RAR α to the IRF1 promoter. Given that GATA2 may act as a positive regulator of the MYC pathway, it is reasonable to propose that GATA2 could also contribute to the repression of IRF1 expression, potentially through an indirect mechanism involving MYC. However, we also observed one potentially spurious association within the reconstructed GRNs, suggesting that GATA2 may promote the expression of GATA1 [65]. This finding appears inconsistent with the well-known "GATA switch" during hematopoietic differentiation. In erythroid lineage, increased GATA1 expression would suppress GATA2, thereby facilitating differentiation [66]. Also, sustained high expression of GATA2 is known to promote leukemia cell self-renew and inhibit differentiation, suggesting that GATA2 should repress, rather than activate, GATA1 expression [67]. The observed positive regulation may therefore be a false positive, possibly resulting from the high sensitivity of the SCENIC algorithm, which can sometimes identify indirect or spurious regulatory links. This highlights the need for cautious interpretation of computationally inferred networks and further experimental validation. Despite this, the other findings of our study—confirming known interactions and identifying novel ones—lay the foundation for future hypothesis generation and investigation.

Although our investigation yielded key insights for three of the six targeted transcription factors, several limitations warrant discussion. First, inherent to single-cell approaches is the incomplete capture of the transcriptome. For instance, our median detection of 4,623 genes per cell is typical for the 10X Genomics platform; however, it raises important questions about whether this value accurately reflects the average gene expression per cell or if low-abundance transcripts are systematically underrepresented. A second, and more significant, limitation concerns the efficiency of the pooled CRISPRi perturbation method, as well as the broader challenges associated with massively parallel gene perturbation. Out of the six genes perturbed, meaningful insights were obtained for only three—an approximate 50% success rate. The low perturbation efficiency observed in our study echoes findings from other published works using similar setups. In our reanalysis of two of these datasets, we noted significantly lower efficiency rates (efficient KD was observed in 97 of 240 sgRNAs in dataset 1 and 56 of 377 sgRNAs in dataset 2) than originally reported, which prompts a careful reevaluation of these approaches. In our experiments, the result is particularly notable given that the experimental design was optimized for a high

cell count per sgRNA and deep sequencing coverage. In our work flow, we enriched for cells expressing dCas9 and sgRNA via FACS, employed a low multiplicity of infection to deliver a single sgRNA per cell, and effectively captured the sgRNAs through a 5' capture system (see Figure 1). Despite these measures, visible perturbation was limited to half of the targeted genes. Since every procedural step—from sgRNA delivery to capture—performed as expected, we suspect this limited efficacy is tied to factors intrinsic to sgRNA design. These factors include the sgRNA's ability to bind target promoters (which may be influenced by the local chromatin state) and its capacity to recruit the necessary repressive complexes. The sgRNAs used in our study were obtained from an existing library that might not fully account for these parameters. Future iterations of CRISPRi-CROP-seq could benefit from, careful design of sgRNA taking local chromatin state in to context, incorporating strategies that allow delivery of multiple sgRNAs per cell (such as delivering two or more sgRNAs in a single construct). Recently, an article titled "CROPseq-multi" [68] reported a redesigned CROP-seq vector that enables the delivery of two sgRNAs per cell. In this design, a single U6 promoter drives the expression of two adjacent but distinct sgRNAs, which are separated by a tRNA sequence. Their experiments showed that this strategy significantly improves perturbation efficiency. This study suggests a new and practical approach: designing vectors that express multiple sgRNAs may be an easy and effective way to enhance perturbation efficiency. Based on our observations, we advocate for a comprehensive benchmarking of all massively parallel gene perturbation methods to validate their reproducibility and enhance overall effectiveness.

In conclusion, our study employed massively parallel gene perturbation via CRISPRi-CROP-seq to investigate six transcription factors in K562 cells, and generated valuable insights into three key TFs—MYB, GATA2, and LMO2—and their interconnected GRNs. By integrating perturbation data with independent validations such as RT-qPCR and ChIP-seq, we confirmed the roles of these TFs. At the same time, our work highlights critical limitations of the current approach. Specifically, challenges with incomplete transcript capture at the single-cell level and the reduced efficiency of sgRNA-mediated perturbation underscore the need for careful optimization in sgRNA design and experimental strategy. We recommend that future studies consider incorporating multiple sgRNAs per cell and implementing rigorous benchmarking protocols to enhance data reproducibility and reliability. Ultimately, our findings not only advance our understanding of TF function in blood cell development but also pave the way for refining massively parallel gene perturbation methods for broader genomic applications.

Material and methods

Cell line production and maintenance

K562 cells were maintained in RPMI160 medium with GlutaMAX (Thermo scientific, catalog no. 61870044) supplemented with 10% fetal bovine serum (FBS), 100 units/mL penicillin and 100 µg/mL streptomycin. HEK293T cells were grown in Dulbecco's modified eagle medium (DMEM) in 10% FBS, 100 units/mL penicillin and 100 µg/mL streptomycin.

To construct the K562 cell line that can stably express dCas9-KRAB, K562 cells were transduced with lentivirus containing pHR-UCOE-SFFV-dCas9-mCherry-ZIM3-KRAB (Addgene, plasmid #154473) at a multiplicity of infection (MOI) approximately 0.5. Three days post-transduction, mCherry-positive polyclonal cells were isolated through fluorescence-activated cell sorting (FACS, BD FACS AriaII, Supplementary Figure 9).

sgRNA library cloning and quality control

For each target gene, three sgRNAs from the set A pool in the Human CRISPR Inhibition Pooled Library (Dolcetto) [39] were chosen. To enhance efficiency, a "G" was introduced at the 5' of each sgRNA. Additionally, five non-targeting sgRNAs sourced from literature were incorporated into our experiment. All these sgRNAs were synthesized by IDT as 73/72 base single-stranded oligonucleotides structured as follows: GGAGAACCA CCTTGTG (5' overhang to the mU6 promoter) + G + sgRNA + GTTTAAGAGCTAAGCTGGAAACAGCA TAGCAAGTT (3' overhang to the gRNA scaffold), and ordered as a pool (Supplementary Table 4).

The vector backbone was prepared by digesting the pBA950 plasmid (Addgene, plasmid #122239) with BspI (NEB, catalog no. R0585S) and BstXI (NEB, catalog no. R0113S) in NEBuffer r2.1 at 37°C for 1 hour. After subjecting the digested vector to gel electrophoresis, it was excised and purified using the gel clean-up kit (Takara, catalog no. 740609.250). The sgRNA library was then constructed following the CROP-seq protocol [3]. Library coverage was assessed by enumerating bacterial colonies on the 1:1,000 dilution plate, revealing that the sgRNA library was cloned at 770x coverage.

The sgRNA sequence was amplified from either the plasmid library or the genomic DNA (gDNA) extracted from cells transduced with the sgRNA library lentivirus. The sequencing libraries were prepared using the CRISPR library preparation method [3]. The libraries were sequenced on an Illumina NextSeq 2000 platform, generating 50-base paired-end reads: Read 1 encompassed the end of mU6 promoter (22 bases), the gRNA spacer (20/21 bases), and the beginning of scaffold (8 bases), while Read 2 only covered the scaffold. sgRNA sequences were aligned to the Read 1 files with a perfect match (no mismatch) to investigate their distribution.

Lentivirus production

One day before transfection, 7 million HEK293T cells were seeded onto a 10 cm dish in 12 ml Opti-MEM I (Gibco, catalog no. 51985-034) supplemented with 5% FCS and 200 μ M sodium pyruvate. Following overnight incubation to attain 90% confluency, HEK 293T cells were transfected using Lipofectamine 3000 reagent (Invitrogen, catalog no. L3000015). To make transfection mixture A, 10.2 μ g of the modified pBA950 (containing either one sgRNA or the entire library) or dCas9-KRAB (Addgene, plasmid #154473) was combined with 5.4 μ g each of the three packaging plasmids: pMDLg/pRRE (Addgene, plasmid #12251), pRSV-Rev (Addgene, plasmid #12253), and pMD2.G (Addgene, plasmid #12259). This mixture was added to 1.5 ml Opti-MEM I, supplemented with 36 μ l P3000 enhancer reagent. Simultaneously, transfection mixture B was prepared by blending 42 μ l Lipofectamine 3000 reagent with 1.5 ml Opti-MEM I. Lipid-DNA complexes were generated by combining transfection mixes A and B, followed by a 20 minutes' incubation at room temperature. Subsequently, 6 ml medium was removed from the HEK293T cells, and 3 ml lipid-DNA complexes was added. After a 6-hour incubation, the medium was replaced with 12 ml fresh medium. Lentivirus supernatants were harvested 48 hours post-transfection and filtered (0.45 μ m). Immediately after collection, lentivirus were concentrated using the Lenti-X-Concentrator reagent (Takara Bio, catalog no. 631232) following the manufacturer's instructions and resuspended in OPTI-MEM at 10% of the original culture volume. Lentiviral particles were subsequently aliquoted and frozen at -80°C .

Lentivirus titration

K562 cells were seeded onto 24-well plate at 50k cells per well with 250ul fresh medium, supplemented with 16ug/ml polybrene. Lentiviral particles were added to the treatment wells using a series of twofold dilutions, ranging from undiluted virus (1:1) to a 1:16 dilution, while an un-transduced control well received only fresh medium. Each dilution condition was prepared in duplicate. The final concentration of polybrene in all wells was adjusted to 8 μ g/ml. After culturing for 24 hours at 37°C , exchange the medium for 500ul fresh medium without polybrene. 72 hours post-transduction, flow cytometry (BD LSRFortessa) was employed to determine the ratio of transduced cells (TagBFP: sgRNA or mCherry: dCas9).

Single cell sequencing

Wild type (WT) K562 cells were first transduced with desired amount (estimated based on the titration result) of the dCas9-KRAB lentivirus with polybrene (8 ug/ml) to achieve a high MOI (~ 0.7 , day -2). Two days later (day 0), they were infected with pooled sgRNA lentivirus

(~ 0.2 MOI) and polybrene. On day 3 post-transduction, cells were sorted using FACS for the mCherry (dCas9-KRAB) and TagBFP (sgRNA) double-positive population. Sorted cells were cultured with fresh medium maintaining at a density of 100k cells/ml. Seven days after pooled sgRNA transduction (day 7), cells were sorted again for viability and double positive fluorescence, then harvested for sequencing immediately. Single cells were collected for the construction of 5' single-cell gene expression (GEX) libraries and CRISPR guide RNA (gRNA) capture libraries using the Chromium X platform (10x Genomics, PN-1000287, PN-1000451, PN-1000263, PN-1000215). The libraries were prepared with a target recovery of approximately 10,000 cells. Both GEX and CRISPR libraries were sequenced on an Illumina NovaSeq 6000 system using a 100-cycle kit. The sequencing configuration included read 1 (28 bp), read 2 (10 bp), read 3 (10 bp), and read 4 (90 bp).

Individual sgRNA cloning experiment

Eleven sgRNAs targeting 3 genes with 3 sgRNAs each and 2 non-targeting controls were individually cloned into pBA950 vector. After heat shock transformation, plasmids were extracted using the Miniprep kit (QIAGEN, catalog no. 27104). All plasmids were verified by XmnI (NEB, catalog no. R0194S) digest and Sanger sequencing with the primer 5'-CCTCGGCCTCTGCA TAAATA-3'. dCas9-KRAB K562 cells were transfected with the eleven sgRNA plasmids in parallel. In each reaction, 100ul 1X Nucleofection buffer was prepared with 4.5ug plasmids. Following count, spin down 1.5 million cells and resuspend the pellet in the Nucleofection buffer with plasmids. Cells were then transferred to an electroporation cuvette and shocked using program T-016 (Amaxa Nucleofector I). When shock is over, quickly add 1ml media to the cuvette and plate the cells. 48 hours post-transfection, sort the TagBFP and mCherry double-positive cells population using FACS. 72 hours post-transfection, cells were harvested and stored at -80°C , awaiting subsequent RNA extraction.

RT-qPCR

Total RNA was isolated using ReliaPrep[™] RNA Cell Miniprep System (Promega, catalog no. Z6012) and converted to cDNA using SuperScript IV Reverse Transcriptase kit (Invitrogen, catalog no. 18090200) following manufacturers' instructions. RT-qPCR reactions were prepared following Platinum[™] Taq DNA Polymerase kit (Invitrogen, catalog no. 10966034) supplemented with 1ul 1/2000 SYBR Green (Sigma, catalog no. S9430) in 25ul reactions. Target genes expression were determined with the primer pairs listed in Supplementary Table 5 and measured on a C1000 Touch Thermal Cycler (Bio-Rad). Relative expression of mRNA for different samples was determined by

the $\Delta\Delta\text{CT}$ method using GAPDH as the internal reference, and visualized by Prism v8.0.2. All RT-qPCR experiments were performed with three biological replicates ($n = 3$), with independent cell cultures and RNA extraction for each replicate.

Processing of single-cell sequencing data

Gene expression and sgRNA libraries were processed using the 10X Cell Ranger version 7.0.1 following default settings. Gene expression reads were aligned to the reference genome “refdata-gex-GRCh38-2020-A”. sgRNA reads were aligned to the reverse complemented sgRNA library using the pattern TTCCAGCTTAGCTCTTA AAC(BC). The “protospacer_calls_per_cell.csv” output file, generated through the application of the Gaussian Mixture Model within Cell Ranger, was used to assign sgRNAs to cells. For cells not listed in this file, they were either no guide molecules founded or lack of a confident call (didn’t pass the threshold). Cells with more than 1 feature (sgRNA) were filtered out.

Gene expression matrix were imported and analyzed using the Seurat package (version 4) in R. Cells were filtered for mitochondrial RNA percentage < 10%, unique genes detected > 400 and < 7000, and UMI counts < 60000. Finally, 6898 high-quality cells harboring single perturbations were used for all the further analysis. On-target knockdown effects (fold change and adjusted p -values) were quantified using the FindMarkers function in Seurat by comparing perturbed cells (cells containing sgRNAs directed against each target gene) with non-targeting control cells. Analyses were performed with pseudocount.use = 1e-9. Fold-change values and associated significance were computed using the Wilcoxon rank-sum test, and the resulting p -values were adjusted via the Bonferroni correction. Differentially expressed genes (DEGs) for each sgRNA were subsequently identified using FindMarkers applying the DESeq2 test for differential-expression assessment relative to cells harboring non-targeting sgRNAs, and the resulting p -values were adjusted via the Bonferroni correction.

Mixscape analysis

We analyzed the aforementioned processed count matrix following the Mixscape pipeline (https://satijalab.org/seurat/articles/mixscape_vignette). To investigate the effect at the sgRNA level, we ran the RunMixscape (Mixscape command) with find.mode = TRUE and prtb.type = “KD”. The results of this Mixscape analysis were visualized using bar plots and further explored through a heatmap of differentially expressed genes (DEGs) across KD, NP, and NT cell subsets. These DEGs were identified by comparing KD cells to NT cells. To visualize perturbation-specific clusters, Linear Discriminant Analysis (LDA) was

used as the dimensionality reduction method on NT cells and KD cells derived from efficient sgRNAs.

Gene regulatory networks inference and validation

Gene regulatory networks inference was started by first running GENIE3 [69] (SCENIC command) to identify co-expression modules. To enhance the accuracy of our results, we included all 6,898 cells with at least one perturbation, as incorporating multiple cell conditions improves module detection. In this step, only positive correlation were kept and used for the further analysis. Next, we used RcisTarget [48] (SCENIC command) to identify regulons, consisting of transcription factors (TFs) and their direct-binding targets. This analysis initially identified 307 regulons. However, 189 of these were labeled as “extended,” indicating low-confidence motif detection. To ensure robustness, we excluded these low-confidence regulons, retaining 118 high-confidence regulons for further analysis.

Next, we calculated regulon activity in each cell using AUCell [48] (SCENIC command) and generated heatmaps at both the sgRNA level and gene level to visualize activity patterns. At the gene level, the MYB KD group included all KD cells from three sgRNAs, the LMO2 KD group consisted of KD cells from LMO2-1 and LMO2-3, and the GATA2 KD group contained KD cells from GATA2-3. All other cells were categorized as NP and labeled as MYB NP, LMO2 NP, and GATA2 NP, respectively. Then we calculated the regulon-specific score (RSS) to filter out regulons that exhibited high activity across all clusters, as these were not of interest. Instead, we focused on regulons that displayed high activity in specific clusters, allowing us to identify key regulatory differences among the KD groups. To incorporate low-activity regulons, we applied a filter of $\text{RAC} < -1$.

After combining low-activity regulons ($\text{RAC} < -1$) and high-specificity regulons ($\text{RSS Z-score} > 0.5$) for each cluster (MYB KD, GATA2 KD, and LMO2 KD), we observed that some regulons were shared between two clusters, suggesting potential regulatory links between the corresponding KD genes. However, one limitation of this approach is its reliance on regulon activity measurements, which rank all genes in a regulon within each cell. This can lead to false positive regulons, where high regulon activity is driven by a few highly expressed genes that do not include the transcription factor (TF) itself. Since all genes in a regulon should be direct TF targets, the TF’s expression level is expected to change as well. To address this issue, we calculated Fold Change (FC) for the TF in each selected regulon and only considered regulons with $\text{TF } |\log_2 \text{FC}| > 0.5$ as potentially relevant regulons.

To validate the GRNs, we selected some genes from regulons with $|\log_2 \text{FC}| > 0.5$ and $\text{P_val_adj} < 0.05$. We then performed knockdown (KD) experiments like

“Individual sgRNA cloning experiment”. To ensure robust comparisons, we pooled cDNA from LMO2-1 and LMO2-3 as the LMO2 KD group, and GATA2-3 as the GATA2 KD group. RT-qPCR analysis was conducted to assess the expression changes of the selected validation genes.

Published dataset analysis

Raw data from two published papers, Wu et al. [34] and Joseph M.R et al. (2022)[58], were retrieved from the Gene Expression Omnibus (GEO) under the accession codes GSE150062 and GSM6210117, respectively. They were analyzed following the same approach we applied for the aforementioned TFs described with headers: “[Processing of single-cell sequencing data](#)” and “[Mixscape analysis](#)”.

Abbreviations

AML	Acute myeloid leukemia
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
CRISPRi-CROP-seq	pooled CRISPR-based gene inactivation
CRISPRi	CRISPR interference
CROP-seq	CRISPR droplet sequencing
dCas9	Dead Cas9
DEG	Differentially expressed gene
FACS	Fluorescence-activated Cell Sorting
FC	Fold Change
GEO	Gene Expression Omnibus
GRNs	Gene regulatory networks
iPSCs	Induced pluripotent stem cells
KD cell	Knock down cell
LDA	Linear Discriminant Analysis
LncRNA	Long non-coding RNA
LSCs	Leukemic stem cells
NT cell	Non targeted cell
RAC	Regulon activity score
RSS	Regulon-specific score
RT-qPCR	Reverse Transcription-quantitative Polymerase Chain Reaction
scRNA-seq	Single-cell RNA sequencing
TF	Transcription factor
UMAP	Uniform Manifold Approximation and Projection
WT	Wild type

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-026-12667-1>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We thank the China Scholarship Council for financially supporting P.Z., and H.Z. We would like to acknowledge Lucas Kuijpers for helpful discussions. We thank Tsung Wai Kan for assistance with FACS sorting of fluorescent cells. Wilfred van Ijcken and Zakia Azmani are gratefully acknowledged for designing and performing the next-generation sequencing used for library quality control.

Code availability

The analysis code supporting the findings of this study is publicly available at https://github.com/hongzhang0208-ux/CROP-seq_K562. The analysis builds

upon publicly available pipelines, including Mixscape (https://satijalab.org/seurat/articles/mixscape_vignette) and SCENIC (<https://github.com/aertslab/SCENIC>).

Authors' contribution

HZ and EM conceived the study. HZ performed the experiments and analyzed the data under the supervision of EM. EB performed the single-cell experiments. PZ assisted with experimental work and data analysis. HZ and EM wrote the manuscript. All authors read and approved the final version.

Funding

This research was supported by the Erasmus University Medical Center in Rotterdam, Netherlands.

Data availability

Raw and processed data have been deposited in NCBI-GEO under accession number GSE308682, and are available upon publication.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 October 2025 / Accepted: 13 February 2026

Published online: 13 March 2026

Reference

- Schraivogel D, Steinmetz LM, Parts L. Pooled genome-scale CRISPR screens in single cells. *Annu Rev Genet*. 2023;57:223–44. <https://doi.org/10.1146/annurev-genet-072920-013842>.
- Dixit A, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*. 2016;167:1853–1866.e1817. <https://doi.org/10.1016/j.cell.2016.11.038>.
- Datlinger P, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*. 2017;14:297–301. <https://doi.org/10.1038/nmeth.4177>.
- Agrotis A, Ketteler R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front Genet*. 2015;6:300. <https://doi.org/10.3389/fgenet.2015.00300>.
- Shalem O, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014;343:84–7. <https://doi.org/10.1126/science.1247005>.
- Adamson B, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*. 2016;167:1867–1882.e1821. <https://doi.org/10.1016/j.cell.2016.11.048>.
- Jaitin DA, et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq. *Cell*. 2016;167:1883–1896.e1815. <https://doi.org/10.1016/j.cell.2016.11.039>.
- So RWL, et al. Application of CRISPR genetic screens to investigate neurological diseases. *Mol Neurodegener*. 2019;14:41. <https://doi.org/10.1186/s13024-019-0343-3>.
- Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*. 2014;343:80–4. <https://doi.org/10.1126/science.1246981>.
- Shifrut E, et al. Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell*. 2018;175:1958–1971.e1915. <https://doi.org/10.1016/j.cell.2018.10.024>.
- Henriksson J, et al. Genome-wide CRISPR screens in T helper cells reveal pervasive crosstalk between activation and differentiation. *Cell*. 2019;176:882–896.e818. <https://doi.org/10.1016/j.cell.2018.11.044>.
- Guerriero ML, et al. Delivering robust candidates to the drug pipeline through computational analysis of arrayed CRISPR screens. *SLAS Discovery*. 2020;25:646–54. <https://doi.org/10.1177/2472555220921132>.
- Kim HS, et al. Arrayed CRISPR screen with image-based assay reliably uncovers host genes required for coxsackievirus infection. *Genome Res*. 2018;28:859–68. <https://doi.org/10.1101/gr.230250.117>.

14. Ross-Thriepand D, et al. Arrayed CRISPR screening identifies novel targets that enhance the productive delivery of mRNA by MC3-based lipid nanoparticles. *SLAS Discov.* 2020;25:605–17. <https://doi.org/10.1177/2472555220925770>.
15. Blomen VA, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science.* 2015;350:1092–6. <https://doi.org/10.1126/science.aac7557>.
16. Marceau CD, et al. Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature.* 2016;535:159–63. <https://doi.org/10.1038/nature18631>.
17. Porichis F, et al. High-throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. *Nat Commun.* 2014;5:5641. <https://doi.org/10.1038/ncomms6641>.
18. Haney MS, et al. Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat Genet.* 2018;50:1716–27. <https://doi.org/10.1038/s41588-018-0254-1>.
19. Fulco CP, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet.* 2019;51:1664–9. <https://doi.org/10.1038/s41588-019-0538-0>.
20. Song Q, et al. Direct-seq: programmed gRNA scaffold for streamlined scRNA-seq in CRISPR screen. *Genome Biol.* 2020;21:136. <https://doi.org/10.1186/s13059-020-02044-w>.
21. McFarland JM, et al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat Commun.* 2020;11:4296. <https://doi.org/10.1038/s41467-020-17440-w>.
22. Yan, X. et al. High-content imaging-based pooled CRISPR screens in mammalian cells. *J Cell Biol.* 2021;220. <https://doi.org/10.1083/jcb.202008158>.
23. Schraivogel D, et al. High-speed fluorescence image-enabled cell sorting. *Science.* 2022;375:315–20. <https://doi.org/10.1126/science.abj3013>.
24. Li Y, et al. UDA-seq: universal droplet microfluidics-based combinatorial indexing for massive-scale multimodal single-cell sequencing. *Nat Methods.* 2025;22:1199–212. <https://doi.org/10.1038/s41592-024-02586-y>.
25. Metzner E, Southard KM, Norman TM. Multiome Perturb-seq unlocks scalable discovery of integrated perturbation effects on the transcriptome and epigenome. *Cell Syst.* 2025;16:101161. <https://doi.org/10.1016/j.cels.2024.12.002>.
26. Papalexi E, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat Genet.* 2021;53:322–31. <https://doi.org/10.1038/s41588-021-00778-2>.
27. Yan RE, et al. Pooled CRISPR screens with joint single-nucleus chromatin accessibility and transcriptome profiling. *Nat Biotechnol.* 2024. <https://doi.org/10.1038/s41587-024-02475-x>.
28. Replogle JM, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol.* 2020;38:954–61. <https://doi.org/10.1038/s41587-020-0470-y>.
29. Schmidt R, et al. CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science.* 2022;375:eabj4008. <https://doi.org/10.1126/science.abj4008>.
30. Legut M, et al. A genome-scale screen for synthetic drivers of T cell proliferation. *Nature.* 2022;603:728–35. <https://doi.org/10.1038/s41586-022-04494-7>.
31. Horlbeck MA, et al. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife.* 2016;5. <https://doi.org/10.7554/eLife.19760>.
32. Boettcher M, McManus MT. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol Cell.* 2015;58:575–85. <https://doi.org/10.1016/j.molcel.2015.04.028>.
33. Tian R, et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron.* 2019;104:239–255.e212. <https://doi.org/10.1016/j.neuron.2019.07.014>.
34. Wu, D. et al. Dual genome-wide coding and lncRNA screens in neural induction of induced pluripotent stem cells. *Cell Genom.* 2022;2. <https://doi.org/10.1016/j.xgen.2022.100177>.
35. Truchi M, et al. Detecting subtle transcriptomic perturbations induced by lncRNAs knock-down in single-cell CRISPRi screening using a new sparse supervised autoencoder neural network. *Front Bioinform.* 2024;4:1340339. <https://doi.org/10.3389/fbinf.2024.1340339>.
36. Chen B, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell.* 2013;155:1479–91. <https://doi.org/10.1016/j.cell.2013.12.001>.
37. Tyner JW, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature.* 2018;562:526–31. <https://doi.org/10.1038/s41586-018-0623-z>.
38. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008;455:1061–1068. <https://doi.org/10.1038/nature07385>.
39. Sanson KR, et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun.* 2018;9:5416. <https://doi.org/10.1038/s41467-018-07901-8>.
40. Lemma RB, et al. Chromatin occupancy and target genes of the haematopoietic master transcription factor MYB. *Sci Rep.* 2021;11:9008. <https://doi.org/10.1038/s41598-021-88516-w>.
41. Fuglerud BM, et al. A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function. *Nucleic Acids Res.* 2017;45:7681–96. <https://doi.org/10.1093/nar/gkx364>.
42. Wang X, Angelis N, Thein SL. MYB - a regulatory factor in hematopoiesis. *Gene.* 2018;665:6–17. <https://doi.org/10.1016/j.gene.2018.04.065>.
43. Peters IJA, de Pater E, Zhang W. The role of GATA2 in adult hematopoiesis and cell fate determination. *Front Cell Dev Biol.* 2023;11:1250827. <https://doi.org/10.3389/fcell.2023.1250827>.
44. Zhu H, et al. Regulation of the lmo2 promoter during hematopoietic and vascular development in zebrafish. *Dev Biol.* 2005;281:256–69. <https://doi.org/10.1016/j.ydbio.2005.01.034>.
45. Zuber J, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature.* 2011;478:524–8. <https://doi.org/10.1038/nature10334>.
46. Lorenzo PI, et al. Identification of c-Myb target genes in K562 cells reveals a role for c-Myb as a master regulator. *Genes Cancer.* 2011;2:805–17. <https://doi.org/10.1177/1947601911428224>.
47. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489:57–74. <https://doi.org/10.1038/nature11247>.
48. Aibar S, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods.* 2017;14:1083–6.
49. Fagnan A, Drissen R, Di Genua C, Meng Y, Nerlov C. Myelo-Erythroid lineage segregation is regulated by the GATA-2 interactome. *Blood.* 2023;142:2687. <https://doi.org/10.1182/blood-2023-190285>.
50. Schmidt M, Nazarov V, Stevens L, Watson R, Wolff L. Regulation of the resident chromosomal copy of c-myc by c-Myb is involved in myeloid leukemogenesis. *Mol Cell Biol.* 2000;20:1970–81. <https://doi.org/10.1128/mcb.20.6.1970-1981.2000>.
51. Lohmann F, Bieker JJ. Activation of Ekf expression during hematopoiesis by Gata2 and Smad5 prior to erythroid commitment. *Development.* 2008;135:2071–82. <https://doi.org/10.1242/dev.018200>.
52. Lu L, et al. LMO2 promotes the development of AML through interaction with transcription co-regulator LDB1. *Cell Death Dis.* 2023;14:518. <https://doi.org/10.1038/s41419-023-06039-w>.
53. Menendez-Gonzalez JB, et al. Gata2 as a crucial regulator of stem cells in adult hematopoiesis and acute myeloid leukemia. *Stem Cell Reports.* 2019;13:291–306. <https://doi.org/10.1016/j.stemcr.2019.07.005>.
54. Inoue A, et al. Elucidation of the role of LMO2 in human erythroid cells. *Exp Hematol.* 2013;41:1062–1076.e1061. <https://doi.org/10.1016/j.exphem.2013.09.003>.
55. Pan Y, Tian R, Lee C, Bao G, Gibson G. Fine-mapping within eQTL credible intervals by expression CROP-seq. *Biol Methods Protoc.* 2020;5:bpaa008. <https://doi.org/10.1093/biomethods/bpaa008>.
56. Greatbatch CJ, et al. High throughput functional profiling of genes at intraocular pressure loci reveals distinct networks for glaucoma. *Hum Mol Genet.* 2024;33:739–51. <https://doi.org/10.1093/hmg/ddae003>.
57. Zhao X, Li J, Liu Z, Powers S. Combinatorial CRISPR/Cas9 screening reveals epistatic networks of interacting tumor suppressor genes and therapeutic targets in human breast cancer. *Cancer Res.* 2021;81:6090–105. <https://doi.org/10.1158/0008-5472.Can-21-2555>.
58. Replogle, J. M. et al. Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *Elife.* 2022;11. <https://doi.org/10.7554/eLife.81856>.
59. Dean A. PU.1 chromosomal dynamics are linked to LDB1. *Blood.* 2018;132:2615–6. <https://doi.org/10.1182/blood-2018-10-880781>.
60. Goyama S, et al. Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J Clin Invest.* 2013;123:3876–88. <https://doi.org/10.1172/jci68557>.
61. Kotmayer L, et al. GATA2 deficiency and MDS/AML: experimental strategies for disease modelling and future therapeutic prospects. *Br J Haematol.* 2022;199:482–95. <https://doi.org/10.1111/bjh.18330>.
62. Lin XC, et al. Molecular dysfunctions in acute myeloid leukemia revealed by integrated analysis of microRNA and transcription factor. *Int J Oncol.* 2016;48:2367–80. <https://doi.org/10.3892/ijo.2016.3489>.

63. Takao, S. et al. Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *Elife*. 2021;10. <https://doi.org/10.7554/eLife.65905>.
64. Lee LM, et al. A novel network pharmacology approach for leukaemia differentiation therapy using Mogrify(*). *Oncogene*. 2022;41:5160–75. <https://doi.org/10.1038/s41388-022-02505-5>.
65. Ikonomi P, et al. Levels of GATA-1/GATA-2 transcription factors modulate expression of embryonic and fetal hemoglobins. *Gene*. 2000;261:277–87. [https://doi.org/10.1016/s0378-1119\(00\)00510-2](https://doi.org/10.1016/s0378-1119(00)00510-2).
66. Suzuki M, et al. GATA factor switching from GATA 2 to GATA 1 contributes to erythroid differentiation. *Genes Cells*. 2013;18:921–33.
67. Nandakumar SK, et al. Low-level GATA2 overexpression promotes myeloid progenitor self-renewal and blocks lymphoid differentiation in mice. *Exp Hematol*. 2015;43:565–577 e510.
68. Walton R T, Qin Y, Kirby B, et al. CROPseq-multi: a universal solution for multiplexed perturbation in high-content pooled CRISPR screens. *bioRxiv*, 2025. 2024;03(17):585235. <https://doi.org/10.1101/2024.03.17.585235>
69. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. 2010;5:e12776.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.