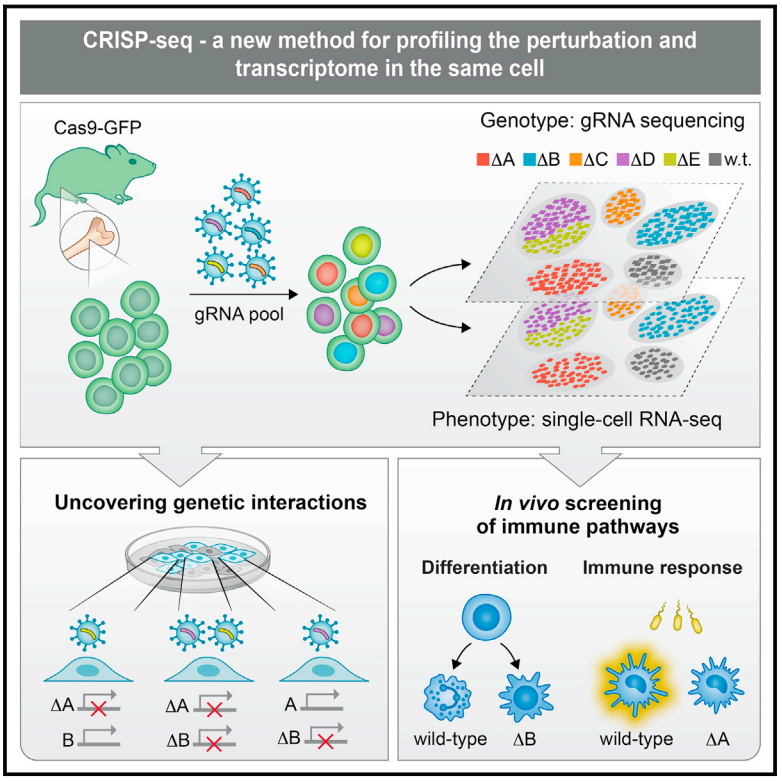
Resource

Cell

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Graphical Abstract



Authors

Diego Adhemar Jaitin, Assaf Weiner, Ido Yofe, ..., Amos Tanay, Alexander van Oudenaarden, Ido Amit

Correspondence

ido.amit@weizmann.ac.il

In Brief

Integrating massively parallel single-cell RNA-seq and CRISPR-pooled screens allows reconstruction of the regulatory networks controlling myeloid cell differentiation and their response to pathogens.

Highlights

- A new method for profiling the perturbation and transcriptome in the same cell
- An integrated approach for CRISPR-pooled screens with single-cell transcriptomics
- Immune development and signaling-dependent circuits revealed by CRISP-seq
- CRISP-seq uncovers gene modules regulated by Stat1 and Rela in myeloid cells

Data Resources

GSE90488
GSE90486
GSE90487

Jaitin et al., 2016, Cell *167*, 1883–1896
December 15, 2016
http://dx.doi.org/10.1016/j.cell.2016.11.039

CellPress

Resource

Cell

Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq

Diego Adhemar Jaitin,[1,5] Assaf Weiner,[1,2,5] Ido Yofe,[1,5] David Lara-Astiaso,[1] Hadas Keren-Shaul,[1] Eyal David,[1] Tomer Meir Salame,[3] Amos Tanay,[4] Alexander van Oudenaarden,[2] and Ido Amit[1,6,*]

[1]Department of Immunology, Weizmann Institute, Rehovot 76100, Israel
[2]Hubrecht Institute-KNAW, Royal Netherlands Academy of Arts and Sciences, 3584 CT Utrecht, the Netherlands
[3]Flow Cytometry Unit, Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel
[4]Department of Computer Science and Applied Mathematics, Department of Biological Regulation, Weizmann Institute, Rehovot 76100, Israel
[5]Co-first author
[6]Lead Contact
*Correspondence: ido.amit@weizmann.ac.il
http://dx.doi.org/10.1016/j.cell.2016.11.039

SUMMARY

In multicellular organisms, dedicated regulatory circuits control cell type diversity and responses. The crosstalk and redundancies within these circuits and substantial cellular heterogeneity pose a major research challenge. Here, we present CRISP-seq, an integrated method for massively parallel single-cell RNA sequencing (RNA-seq) and clustered regularly interspaced short palindromic repeats (CRISPR)-pooled screens. We show that profiling the genomic perturbation and transcriptome in the same cell enables us to simultaneously elucidate the function of multiple factors and their interactions. We applied CRISP-seq to probe regulatory circuits of innate immunity. By sampling tens of thousands of perturbed cells in vitro and in mice, we identified interactions and redundancies between developmental and signaling-dependent factors. These include opposing effects of Cebpb and Irf8 in regulating the monocyte/macrophage versus dendritic cell lineages and differential functions for Rela and Stat1/2 in monocyte versus dendritic cell responses to pathogens. This study establishes CRISP-seq as a broadly applicable, comprehensive, and unbiased approach for elucidating mammalian regulatory circuits.

INTRODUCTION

Functional diversity of immune cells is critical for the generation of the different regulatory and effector cell types required to safeguard the host not only against a broad range of threats, such as pathogens and cancer, but also from attacking its own healthy cells and tissues. Cellular diversity and response are controlled by regulatory networks, which function as decision-making circuits of the cell (Amit et al., 2011; Beyer et al., 2007; Carter et al., 2013). Two major confounding factors limit accurate characterization of cellular circuitry. The first is the complexity of biological circuits, which include redundancies and nonlinear crosstalk between pathways, and the second is cellular plasticity and heterogeneity within both in vivo and in vitro models (Junker and van Oudenaarden, 2014; Trapnell, 2015). Although synthetic lethality screens have greatly enhanced our understanding of the function of a large number of genes in various pathways (Amit et al., 2011; Blomen et al., 2015; Scholl et al., 2009; Tong et al., 2001), genome-wide methods to identify interactions between genetic elements at single-cell resolution are still lacking.

Myeloid cells are immune cells of hematopoietic origin that provide crucial innate defense (Geissmann et al., 2010; Hashimoto et al., 2011; Lavin et al., 2015). In addition, they have important tissue-specific functions that range from clearance of surfactant from the lungs to neuronal pruning and establishment of gut homeostasis (Lavin et al., 2015). Differentiation and tissue-specific activation of myeloid cells require precise regulation of gene expression (Glass and Natoli, 2016; Gosselin et al., 2014; Lavin et al., 2014; Lawrence and Natoli, 2011; Okabe and Medzhitov, 2014). Understanding the precise regulation of various immune modulatory pathways will enable improved diagnostics and cell engineering toward desired responses for better adjuvants, immunotherapy, and resolution of various neurodegeneration and autoimmune diseases. However, the plasticity, niche specificity, and complexity of immune regulatory networks are difficult to resolve with current methods applicable to bulk cell populations (Chattopadhyay et al., 2014; Moignard et al., 2013; Semrau and van Oudenaarden, 2015).

Single-cell genomic technologies enable unprecedented resolution in characterizing new cell types and states, transitions from normal to disease, and response to therapies (Junker and van Oudenaarden, 2014), including in cells of the brain, gut, liver, pancreas, bone marrow, as well as immune cell types in various tissues and disease states (Grün et al., 2015; Gury-BenAri et al., 2016; Jaitin et al., 2014; Junker and van Oudenaarden, 2014; La Manno et al., 2016; Paul et al., 2015; Segerstolpe et al., 2016; Shalek et al., 2014; Stegle et al., 2015; Treutlein et al., 2014; Zeisel et al., 2015). Single-cell technologies further our ability to identify potential regulators of the various cell states, promoting testable hypotheses to elucidate molecular mechanisms of regulation (Paul et al., 2015; Stegle et al., 2015; Yosef and Regev,

2016). Yet, current single-cell technologies are descriptive by nature, lacking the ability to elucidate causality unless they are combined with gene loss-of-function or knockout (KO) models (Junker and van Oudenaarden, 2014; Moignard and Göttgens, 2014; Stegle et al., 2015).

Clustered regularly interspaced short palindromic repeats (CRISPR/Cas)-based technologies have propelled the capacity to elucidate genetic function (Cong et al., 2013). Current CRISPR/Cas methods are used either for individual gene KOs or for pooled screens (Doudna and Charpentier, 2014; Hsu et al., 2014). While individual gene KOs are effective, they lack in scalability. Pooled genetic screens are powerful tools for the discovery and functional annotation of genetic elements, yet they lack the resolution needed to elucidate complex phenotypes. These screens are usually used to identify crude cellular phenotypes, such as survival, cell shape, or single gene markers (Chen et al., 2015; Hsu et al., 2014; Shalem et al., 2015). This may result in false-negative and false-positive results, as perturbations may change cell compositions and states that are difficult to decompose without more comprehensive phenotyping (Blomen et al., 2015; Scholl et al., 2009; Tong et al., 2001; Wang et al., 2013).

Together, these hurdles highlight the need for developing pooled multiplexed technologies for high-resolution characterization of mammalian pathways at the single-cell level. We developed CRISP-seq, an integrated method that combines the resolution of massively parallel single-cell RNA sequencing (RNA-seq) with the genome editing scale of pooled CRISPR screens. Here we demonstrate the potential of the method to uncover the function of multiple factors and their combinations in a single experiment. To identify the guide RNA (gRNA) combination that infected each single cell, we designed a scalable lentiviral vector backbone that contains, in addition to the gRNA module, a transcribed poly-adenylated unique guide index (UGI) and a fluorescent selection marker. We observe a high level of phenotype consistency between gRNAs targeting the same gene in both in vivo and in vitro models as well as in multiplex or individual gRNA experiments. We applied CRISP-seq to characterize the regulatory network controlling the differentiation of myeloid cells and their response to pathogen components by measuring tens of thousands of single cells perturbed for critical developmental and immune response regulators. We identified the role of Cebpb and Irf8 as controlling the commitment toward various myeloid lineages. In addition, we analyzed the rewiring of signaling-dependent transcription factors (TFs) in the inflammatory and antiviral pathways, highlighting different functions in monocyte versus dendritic cells (DCs). Our findings demonstrate CRISP-seq as a versatile approach to probe and infer the wiring of mammalian circuits, fundamental to the future engineering of immune cells toward desired responses, including immunotherapy.

RESULTS

CRISP-Seq: An Integrated Method for Single-Cell RNA-Seq and CRISPR-Pooled Screens

To elucidate the function of multiple regulatory factors at single-cell and genome-wide resolution, we developed CRISP-seq, an integrated method for pooled CRISPR/Cas genome editing followed by massively parallel single-cell RNA-seq. For this protocol, we engineered a scalable lentiviral backbone (CRISP-seq vector) that includes, in addition to a gRNA expression cassette, a transcribed UGI that allows the identification of the gRNA from single-cell RNA-seq data (Figure 1A; STAR Methods; Key Resources Table). Importantly, we also engineered the lentivirus to include a fluorescent selection marker that enabled us to study perturbed cells from specific niches in animal models. The combination of a gRNA index with the single-cell transcriptome data allowed us to generate deep and comprehensive phenotype profiling of multiplexed gene KOs and to study their function and interactions in a single experiment (Figure 1A).

The CRISP-seq protocol is highly reproducible for identifying the transcriptome in combination with the gRNA (Figures 1B–1D). To benchmark CRISP-seq, we cloned into our backbone a gRNA targeting the Itgam gene, which encodes for the CD11b integrin, alongside a blue fluorescent protein (BFP) marker and an mCherry fluorescent marker together with a gRNA targeting the Cebpb gene (STAR Methods; Key Resources Table). Bone marrow cells from C57BL/6 female mice harboring a GFP-labeled Cas9 knockin (Platt et al., 2014) were infected with a mix of lentiviruses expressing gRNA(CD11b)-BFP-UGI and gRNA(Cebpb)-mCherry-UGI. Then 5 days later, infected $GFP^+BFP^+/mCherry^+CD11c^+$ myeloid-derived cells were sorted for massively parallel single-cell RNA-seq analysis (STAR Methods; Key Resources Table). Information on BFP, mCherry, and CD11b intensities was recorded for each cell by index sorting (Paul et al., 2015). Comparison of CD11b protein expression levels and BFP intensities showed that, in 81% of the cells with high levels of BFP signal (Itgam-$gRNA^+$), CD11b protein expression decreased substantially (Figure 1B). Comparing the UGI read counts with CD11b and BFP intensities in each cell showed a high concordance among BFP-positive cells, CD11b perturbation, and CD11b-UGI expression (Figures 1B and S1A). We detected the CD11b-UGI with 84% precision, computed as the sum of true-positive and true-negative events relative to the BFP fluorescence-activated cell sorting (FACS) signal, and false-positive and false-negative events of 4% and 12%, respectively (Figures 1B and S1E).

To evaluate the potential of applying CRISP-seq for multiplexed genome editing, we assessed the accuracy of detecting individual gRNAs and their combinations. Successful Cas9 editing cleaves the gRNA complementary seed sequence in the DNA, creating mutations and small insertions or deletions (indels), but does not necessarily impact RNA expression directly. Because TFs are often regulated through auto-regulatory loops, their mRNA expression can potentially serve as a proxy for gRNA activity. The comparison of Cebpb mRNA expression versus mCherry intensities (Cebpb-$gRNA^+$) in single cells showed significant anti-correlation between the mCherry signal and Cebpb expression ($r = -0.39$; Figures 1C, S1B, and S1C). These experiments also confirmed that the BFP marker was more effective in labeling infected cells and displayed a lower false-negative rate, with better correlation among fluorescence intensity, UGI detection, and gene editing (Figures 1D–1F and S1B–S1E). To evaluate the genome editing efficiencies within UGI-positive cells targeting CD11b, Cebpb, and multiplexed CD11b and Cebpb, we sorted BFP−/mCherry−, BFP+/mCherry−, BFP−/mCherry+, and BFP+/mCherry+ cells (Figure 1E). We then amplified and sequenced the gRNA-targeted loci, which confirmed editing in

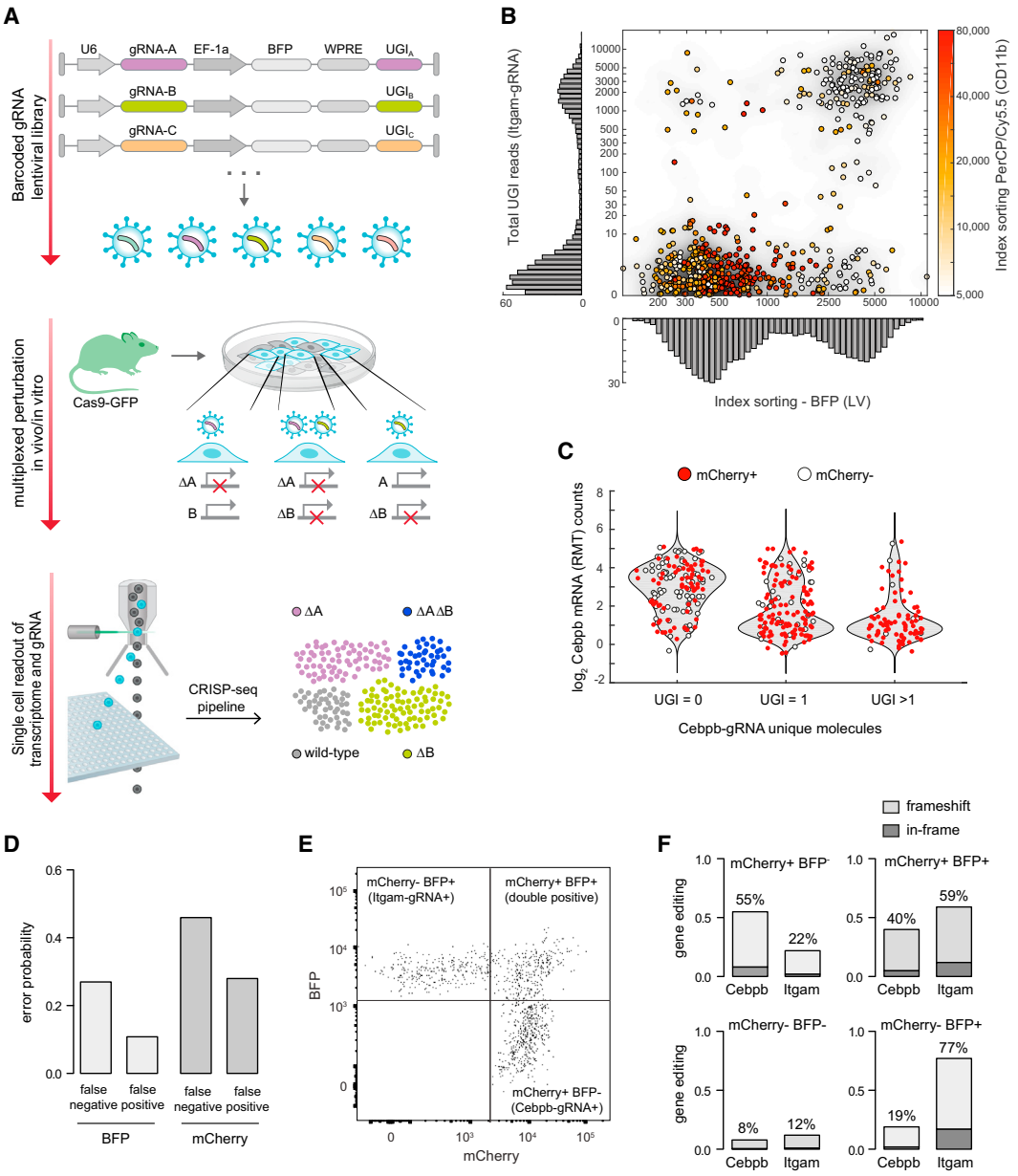


Figure 1. CRISP-Seq: An Integrated Method for Single-Cell RNA-Seq and CRISPR-Pooled Screens

(A) Schematics of the CRISP-seq experimental pipeline. A pooled gRNA library promotes gene KOs in Cas9-expressing cells through lentiviral infection, resulting in single or multiple gene KOs. Each gRNA in the pooled library is detected together with the cell transcriptome through the expressed unique guide index (UGI) during single-cell RNA-seq. A fluorescent marker enables selection of relevant cells, and downstream analysis elucidates genotype-to-phenotype relations in single cells. U6, small nuclear ribonucleoprotein promoter; EF-1a, elongation factor 1-alpha promoter; BFP, blue fluorescent protein; WPRE, Woodchuck hepatitis virus posttranscriptional regulatory element for enhanced BFP expression.

(B) Scatterplot showing Itgam/CD11b protein expression as recorded by FACS index sorting in each cell sequenced with CRISP-seq. BFP-recorded intensity and total UGI read count for Itgam are shown on the x and y axes, respectively. LV, lentivirus. Index sorting PerCP/Cy5.5, FACS-recorded intensity levels of the PerCP/Cy5.5 fluorophore conjugated to the CD11b antibody, is indicative of the CD11b protein expression levels.

(C) Violin plots of Cebpb mRNA expression is shown in single cells labeled for mCherry (Cebpb-gRNA) at different UGI detection cutoffs in each cell. RMT, random molecular tags.

(D) Bar graph shows measured error probabilities of false-positive and false-negative events for cells classified with UGI-seq versus BFP or mCherry fluorescent marker classification.

(E) FACS plot of myeloid cells expressing BFP (Itgam-gRNA$^+$) and mCherry (Cebpb-gRNA$^+$). Shown are cells sorted for indel-seq analysis, namely BFP+, mCherry+, or double-positive cells.

(F) Bar plots showing the percentage of genome editing around the targeted sites of each gRNA, for each quadrant indicated in (E). BFP/mCherry double-negative cells (negative control) were sorted from a gate defined outside both the mCherry- and the BFP-positive gates (Figure S1D).

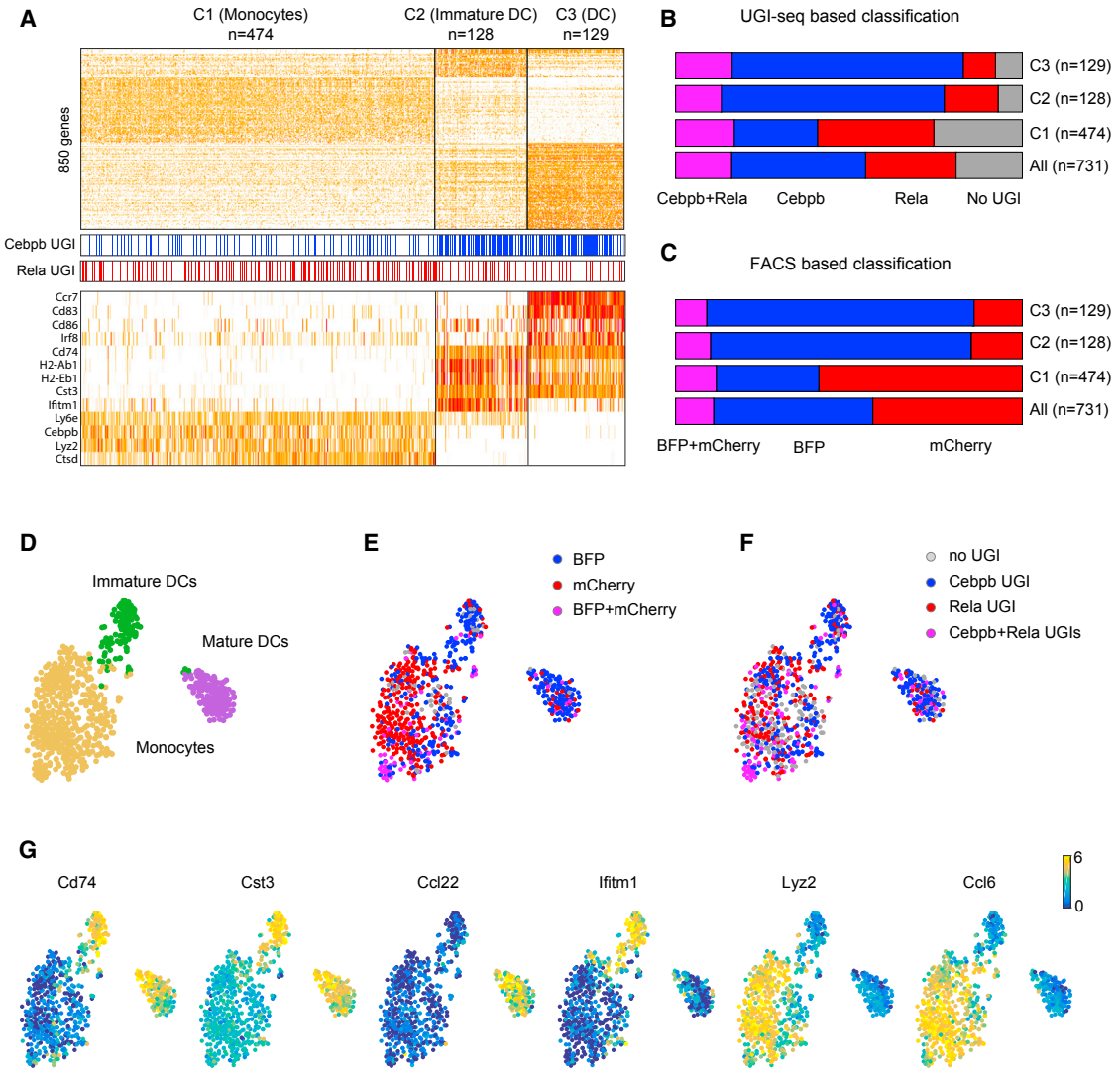


Figure 2. CRISP-Seq Analysis Identifies a Role for Cebpb in Monocyte Development

(A) Heatmap of 731 single myeloid cells. Top panel: expression of the 830 most differential genes across the clusters is shown. Middle bar: UGI detection of Cebpb-gRNA (blue) and Rela-gRNA (red) is shown. Bottom panel: expression of marker genes is shown.

(B and C) Bar plots show the ratio between the different gRNAs and their combinations in the different clusters, using either the UGI (B) or fluorescent marker (C).

(D) The t-distributed stochastic neighbor embedding (t-SNE) plot of the 731 single myeloid cells depicts the separation into monocytes (orange), immature DCs (green), and mature DCs (purple).

(E and F) Projection of the fluorescent marker (E) or UGI labeling (F) onto the t-SNE plots is shown.

(G) Projection of key marker genes onto the t-SNE plot is shown.

the expected quadrants (Figures 1F and S1F). Together, these results demonstrate the value of combining massively parallel single-cell RNA-seq and a UGI strategy for accurate identification of gRNA or combinations of gRNAs in single cells.

CRISP-Seq Analysis Identifies a Major Role for Cebpb in Monocyte Development

Next, we assessed the effectiveness of CRISP-seq in deciphering the function of genetic elements in a multiplexed experiment. The myeloid compartment is composed of environmental plastic cells with functional diversity in both cell state and response (Ginhoux and Jung, 2014; Glass and Natoli, 2016; Gosselin et al., 2014; Lavin et al., 2014, 2015). To better understand the pathways regulating this complexity, we infected bone marrow cells with gRNA targeting a combination of response (mCherry/Rela-gRNA) and developmental (BFP/Cebpb-gRNA) regulators and sorted CD11c^{+} myeloid cells, indexed for BFP and mCherry intensities, for CRISP-seq analysis.

Unsupervised graph-based clustering (PhenoGraph) (Levine et al., 2015) identified three major myeloid cell types in the culture (Figure 2A; Table S1; STAR Methods; Key Resources Table): a monocyte population expressing Lyz2, CD11b, Ly6c2, Cebpb, and lysosomal peptidases (Ctsb, Ctsd, and Ctss) and two DC populations expressing high levels of the MHC-II pathway

components (CD74, H2-Aa, etc.), Cst3, as well as checkpoint and co-stimulatory molecules (e.g., PD-L2 and CD86) (Helft et al., 2015) (Figure 2A). The two DC types could be subdivided into mature migratory DCs expressing CCR7, CCL22, CD83, and Irf8 and an immature subpopulation expressing Csf1r, Ifitm1, Fcgr3, and Lgals3 (Schlitzer et al., 2015). Using either the guide-specific fluorescent marker or the UGI indexes revealed that the two DC subtypes were significantly enriched for Cebpb KO cells, whereas the monocytes were enriched for wild-type (no UGI) and Rela KO cells (hyper-geometric p value $< 1 \times 10^{-4}$; Figures 2A–2F and S2A). We confirmed these results by infecting separate cultures with a single gRNA targeting Cebpb or a control gRNA and obtained comparable cell type phenotypes and distributions to the CRISP-seq pool (Figures S2B–S2E).

To further characterize these populations and their response to pathogens, we infected bone marrow cells with the same combination of Cebpb and Rela gRNAs and stimulated the myeloid culture with the Toll-like receptor 4 (TLR4) agonist lipopolysaccharide (LPS) (Akira et al., 2006; Amit et al., 2009), a purified component from gram-negative bacteria, for 4 hr prior to sorting. Clustering analysis identified the same three cell types (i.e., monocytes and immature and mature DCs), which exhibited highly diverse responses to LPS (Figures 3A and S3A–S3C; Table S2). The monocytes elicited a robust inflammatory response exemplified by the induction of Il1a/b, Il12b, and Cxcl2, as well as an antiviral response (e.g., Cxcl10, Oasl1, and Ifit2) (Figures 3A and 3B). In contrast, the DC subtypes activated the antiviral pathway and induced many co-stimulatory and checkpoint molecules, with minor induction of the inflammatory genes (Figures 3A, 3B, S3B, and S3C). Similar to the unstimulated culture, the two DC populations were significantly enriched for cells expressing the Cebpb-gRNA (hyper-geometric p value $< 1 \times 10^{-5}$; Figures 3A–3D). We observed a diminished inflammatory response in cluster II of the monocytes. This cluster was enriched for the UGI sequence matching the Rela-gRNA (hyper-geometric p value $< 1 \times 10^{-5}$). The response in this cluster was perturbed for dozens of inflammatory genes (Cxcl2, Il1b, Il12b, and Tnf), but not for antiviral response genes (Figures 3A–3E, S3D, and S3E).

In summary, CRISP-seq analysis identified the known role of Rela in regulating the inflammatory response in monocytic cells. Yet, it unexpectedly uncovered Cebpb as an important factor regulating the balance between DC and monocyte development (Feng et al., 2008; Heinz et al., 2010). When Cebpb was perturbed, cells were pushed toward the DC lineage expressing high levels of Irf8. Furthermore, our analysis showed that these two myeloid types boosted a dramatically different response to LPS. The plasticity in differentiation and rewiring of response pathways of myeloid cells would have made these results difficult to interpret without single-cell analysis coupled to perturbations (Paul et al., 2015).

Decoupling of Antiviral and Inflammatory Pathways by Multiplexed Perturbations

To better characterize the genotype-to-phenotype relation in single cells by CRISP-seq and to identify multiplexed perturbations, we developed an algorithm that would most accurately detect perturbed single cells with distinct phenotypes. Our framework relies on the assumption that cells with similar genotypes will be in close proximity in the phenotypic space; hence, a cell with a true loss-of-function hit will generate a similar phenotype that is different from in-frame mutations or non-targeted cells. Using this assumption, we sought to overcome two sources of potential outliers in our data, namely false-positive and false-negative cells. Regarding the former, targeting of Cas9 to a specific gene locus generates loss-of-function mutation/indels in up to 80% of the loci (Sternberg and Doudna, 2015). This implies that for any single cell for which a UGI was detected, there is at least a 20% chance that the targeted gene is fully or partially active. Conversely, with the current CRISP-seq/UGI strategy, up to 20% of the cells remain undetected but can potentially express the gRNA. To overcome the noisy and missing genotype labeling, we developed a label refinement algorithm based on *k*-Nearest Neighbors (kNN) graph (Blondel et al., 2008; Girvan and Newman, 2002; Levine et al., 2015) to correct the genotype labeling based on the genotype of neighboring cells (Figures 4A and 4B; Table S3; STAR Methods; Key Resources Table). In the first step after graph generation, cells that are connected to other cells with the same genotype more than expected by chance (bootstrap p value < 0.05) maintained their UGI label, whereas cells that are in disagreement with their neighbors lost their label. In the second step, we propagated the genotype labels to cells with missing labels based on the genotype of their neighbors.

To evaluate the effect of monocytic cells perturbed for multiplexed inflammatory and antiviral pathways, we infected bone marrow cultures with a pool of gRNAs targeting Rela and Irf9, known regulators of the two pathways, respectively. Then, we stimulated the culture with LPS for 4 hr and sorted cells (GFP^+CD11c^+) for CRISP-seq analysis (Figures 4B–4D). PhenoGraph clustering identified 691 monocytes and 81 DCs (Figure S4A). Because the DC populations are minor in this culture without Cebpb perturbation, we analytically removed all DCs from further analysis. Projecting the kNN graph after label refinement revealed four distinct genotype compartments corresponding to wild-type cells (no UGI), cells enriched for Rela KO, cells enriched for Irf9 KO, and cells enriched for Rela and Irf9 double KO (Figures 4B and S4A–S4C). The cells enriched for Irf9 KO displayed a diminished antiviral gene module response (e.g., Ifit2 and Cxcl10) (Figures 4C, 4D, 4F, and S4D). In contrast, the cells enriched for Rela KO displayed a diminished inflammatory response, including Il12b and Cxcl2 (Figures 4C–4E and S4D). We confirmed these results by infecting separate cultures with single gRNA targeting Rela or Irf9, and we obtained comparable phenotypes and distributions to the CRISP-seq pool (Figures S4E and S4F). The combination of Rela and Irf9 KO had, in most cases, an additive effect (A + B = C; Figures 4D, 4G, 4H, and S4D), indicating that, in monocytes, the inflammatory and antiviral pathways largely regulate different gene modules (Medzhitov, 2007; Napolitani et al., 2005). Among the few exceptions were Ptx3 and Cxcl3, a chemokine that controls monocyte migration. Cxcl3 displayed interesting combinatorics, repressed by IRF9 and activated by Rela, suggesting that this chemokine may have opposing effects depending on the immune stimuli. Together, our results demonstrate the ability of CRISP-seq to dissect multiplexed perturbations and non-overlapping regulation of inflammatory and antiviral responses in monocytes.

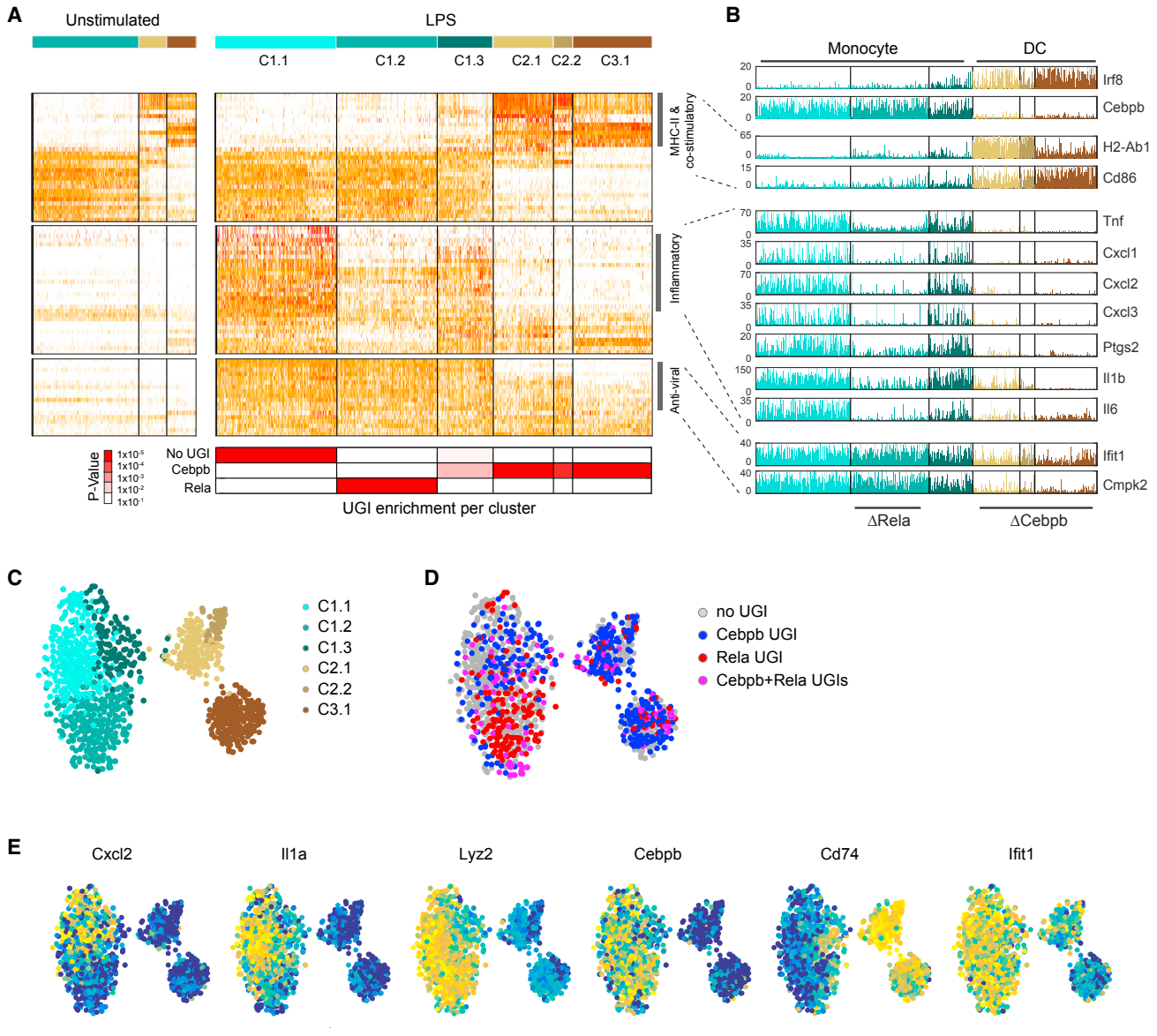


Figure 3. CRISP-Seq Analysis of Myeloid Cells Activated with LPS

(A) Heatmap of 1,186 single myeloid cells stimulated with LPS for 4 hr. Middle panel: expression of the 222 most differential genes across the clusters is shown. Left panel: expression of same genes from 731 unstimulated myeloid cells is shown. Bottom panel: enrichment of the different gRNAs across the clusters is shown.

(B) Expression level (RMT counts) of selected marker genes from different biological processes across the clusters following LPS stimulation. Perturbation of key inflammatory genes is observed within the cluster enriched for Rela KO.

(C) The t-SNE plot of the 1,186 single myeloid cells stimulated with LPS depicts the separation into the six different clusters shown in (A).

(D) Projection of the UGI onto the t-SNE plots is shown.

(E) Projection of key marker genes onto the t-SNE plot is shown.

Perturbations of Developmental and Signaling-Dependent TFs Reveal the Rewiring of Regulatory Circuits in Myeloid Cells

In order to extend our analysis to a larger group of TFs regulating the inflammatory and antiviral circuits, we screened for candidate TFs that are transcriptionally upregulated following LPS stimulation and may regulate these processes (Amit et al., 2009). We infected bone marrow cells with a pool of 57 gRNAs covering 22 genes and performed CRISP-seq on 6,144 cells stimulated with LPS for 4 hr (Figures 5A and S5A). In order to evaluate gRNAs effects on the antiviral or inflammatory programs, we calculated a significance score per gRNA for each program (Figure S5B). Significant perturbation ($p < 0.01$) was observed for 17 gRNAs covering 8 genes, and in all cases, the observed perturbations were specific to one of the programs (Figure S5B). Following this screen, we proceeded with a focused pool of TFs that robustly regulate the antiviral, inflammatory, or developmental processes, including Irf9, Irf4, Stat1,

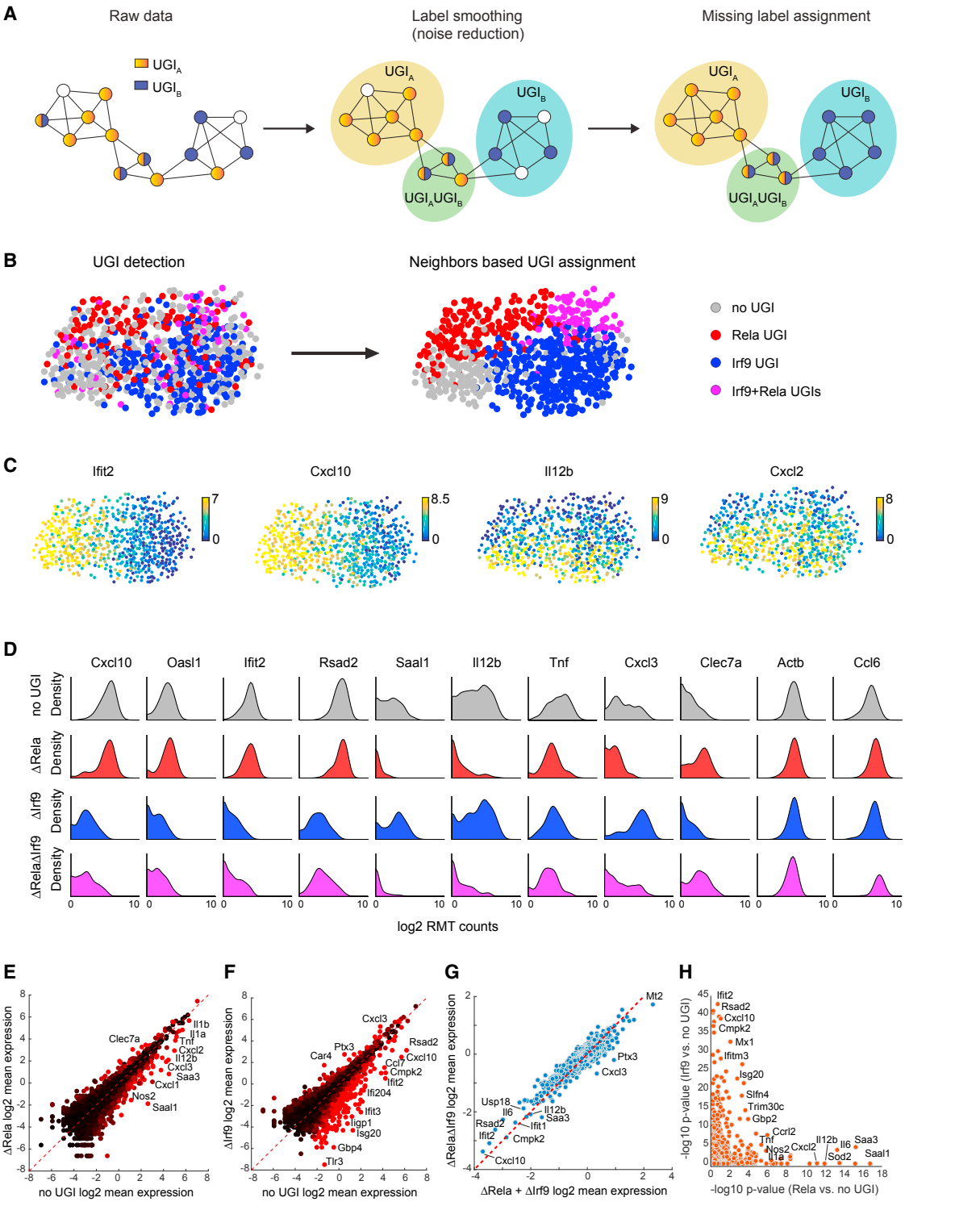


Figure 4. Decoupling of Antiviral and Inflammatory Pathways by Multiplexed Perturbations

(A) Cartoon showing the different processes and stages of UGI label refinement applied in our algorithm to most optimally model single cells targeted by Cas9. White circles indicate unlabeled cells, undetected UGI (no UGI).

(B) Projection of the UGI onto the kNN graph of 691 monocytes stimulated with LPS for 4 hr, before (left) and after (right) UGI assignment correction, is shown.

(C) Projection of key inflammatory and antiviral response genes onto the kNN graph is shown.

(D) Density histograms depict the expression of key inflammatory and antiviral response genes in the different KO combinations.

(legend continued on next page)

Stat2, Rela, Nfkb1, Cebpb, and Irf8 and performed CRISP-seq on 6,749 myeloid cells. Clustering analysis identified similar cell states as in previous perturbations, including two DC states enriched for Cebpb, and monocyte cells that were perturbed in the antiviral response module (Stat1, Stat2, Irf8, and Irf9) and in the inflammatory module (Rela and Nfkb1) (Figures 5B–5D). The monocytic cells perturbed for the antiviral regulators displayed a diminished antiviral response (e.g., Ifit2 and Cxcl10) and regulatory factors (Irf7 and Stat2), with almost no effect on the inflammatory gene module (Figures 5B–5D). In contrast, the cells perturbed for inflammatory regulators displayed a diminished inflammatory response (Cxcl2 and Il12b). To determine if these effects are direct or indirect, we compared the DNA-binding pattern of Stat1, Stat2, and Rela in monocytes to the transcriptional change upon perturbation of each factor (Figures 5E and 5F) (Garber et al., 2012). We found that most genes that were downregulated in Stat1/2 KOs also were bound directly by these factors (Pearson correlation $r = -0.52$), specifically within enhancer regions, suggesting that many of the transcriptional effects for these factors are direct.

We next addressed the rewiring of the same inflammatory and antiviral circuits in other myeloid cell types. We analyzed only factors perturbed in more than 30 cells, namely Rela and Stat2. KO of Stat2 in DC mimicked to a large degree the effects that we observed in monocytes, namely perturbation of a large set of antiviral genes (Figure S5C), of DC-specific response genes such as the migratory chemokine Ccr7, as well as of co-stimulatory molecules (CD83 and CD86). Since the inflammatory genes are not upregulated in DCs, we did not expect a specific effect of Rela KO in these cells. Surprisingly, Rela was found to affect a large number of activated DC genes (Ouaaz et al., 2002). These included co-stimulatory molecules, chemokines, and antiviral response genes (Figure S5C). Together, our analyses suggest that inflammatory and antiviral response circuits are rewired in different myeloid cells and that specific regulators control different gene modules in a cell-type-specific manner.

In Vivo CRISP-Seq Analysis Uncovers the Complexity of Myeloid Regulatory Circuits in Immune Niches

In vitro models identify many aspects of gene regulation and cellular function, but they do not recapitulate the full complexity of physiological interactions of diverse cell types within specific tissues (Chen et al., 2015). Immune niches within the spleen, lymph node, brain, or tumor represent a highly complex and dynamic network of interactions of various immune and non-immune cell types. Understanding the precise function of different regulatory circuits in these niches is important for both basic and clinical research. To study the regulatory function of developmental and signaling-dependent factors in immune niches, we sorted Lin^- $Sca1^+$ $c\text{-}kit^+$ (LSK) hematopoietic progenitors from GFP-labeled Cas9 knockin mice, and we infected them with a pool of Cebpb, Irf8, Rela, Stat1, Stat2, and two control gRNAs (Figure 6A; Table S4). The infected Cas9 (GFP^+) donor populations were mixed with unlabeled wild-type supportive bone marrow cells and injected into lethally irradiated recipient mice (Figure 6A; STAR Methods; Key Resources Table). Then, 7 days following transplantation, successful engraftment was observed and mice were injected with LPS (Figure S6A).

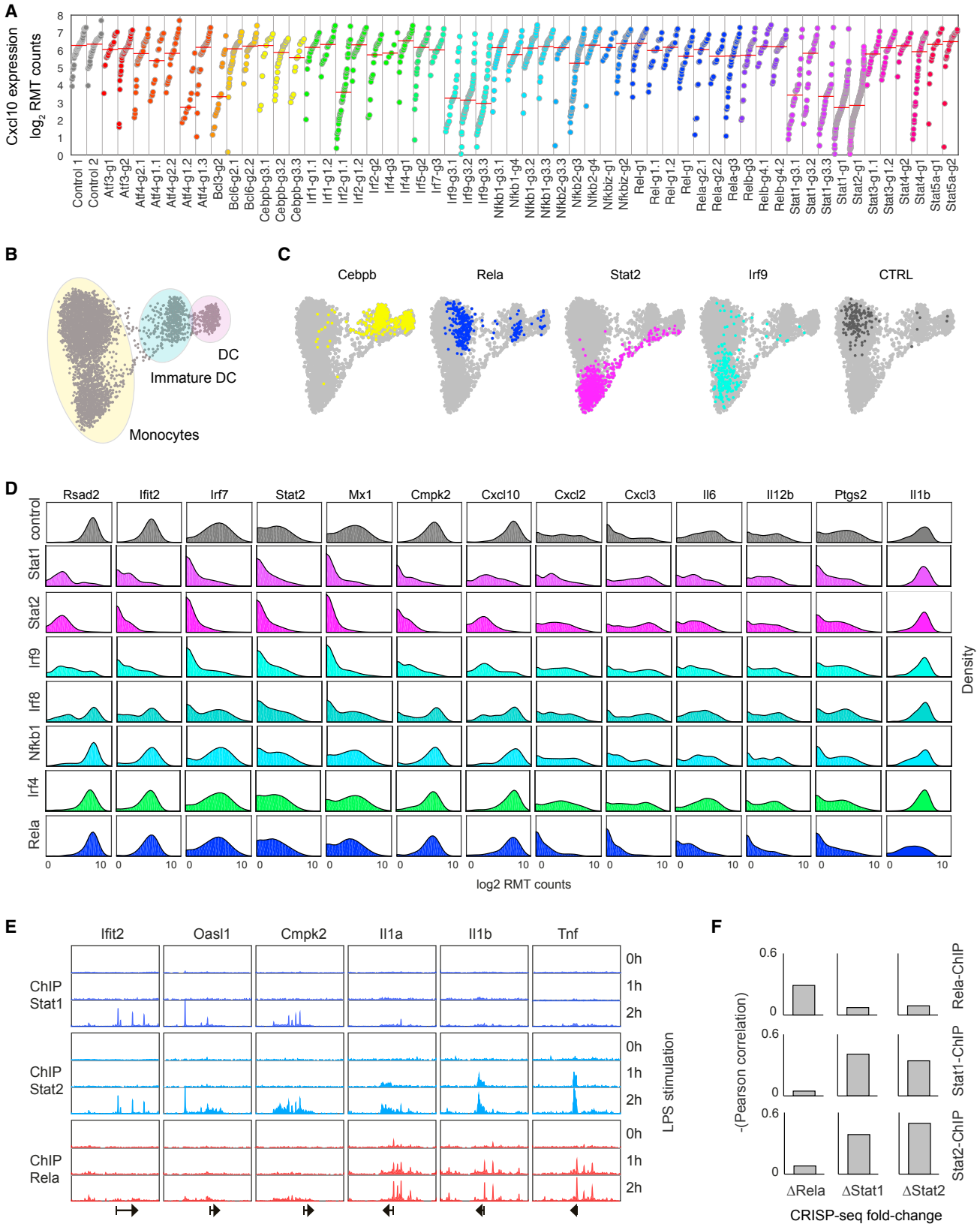
To focus on the regulation of myeloid cell response to pathogens in the splenic niche, 4 hr following LPS stimulation, 2,768 splenic myeloid cells ($CD11b^+$ or $CD11c^+$) positive for Cas9 (GFP^+) and gRNA (BFP^+) were sorted for CRISP-seq analysis. Unsupervised analysis of the single myeloid cells identified nine myeloid cell types and states (Figures 6B and S6B), including granulocytes, monocytes, conventional DCs (cDCs), and plasmacytoid DCs (pDCs). Similar to the in vitro model, monocytes were associated with high expression levels of antibacterial enzymes (Ctsc and Lyz2) and of the Cebpb TF (Figures 6B, 6C, and S6B). The cDCs were associated with high levels of MHC-II pathway genes and cysteine protease inhibitors (H2-Eb1 and Cst3), whereas pDCs were associated with the expression of classical markers of this type, such as Ly6D and Siglech, and relevant TFs (Irf8 and Tcf4). Granulocytes were associated with two clusters: an immature state expressing enzymes associated with neutrophil granule formation, namely MPO and Elane, and a more mature state expressing high levels of the antimicrobial peptide CAMP (Figures 6B, 6C, and S6B). Projecting the perturbation indexes (genotype) on this graph showed that perturbation of Cebpb was linked to the generation of cDCs ($p < 10^{-8}$), similar to the in vitro model, whereas perturbation of Irf8 was associated with granulopoiesis ($p < 10^{-12}$) (Figures 6D and S6C–S6E). To confirm the link between Irf8 KO and increased granulopoiesis, we infected bone marrow cells in vitro with mixtures of Cebpb, Irf8, and control gRNA, and we sorted both $CD11c^+$ and $CD11c^-$ cells for CRISP-seq analysis. Similar to the in vivo model, we observed that Cebpb was essential for the monocyte state, whereas Irf8 was essential for the DC and monocyte states (Figures S6D and S6E). These results are in agreement with KO experiments showing that Irf8 is required for the development of pDCs, cDCs, monocytes, and macrophages while inhibiting the generation of neutrophils (Becker et al., 2012; Kurotaki and Tamura, 2016).

Focusing on the perturbations of Stat1 or Stat2 resulted in largely overlapping phenotypes enriched for different activation states of monocytes, pDCs, and cDCs (Figures 6D–6F and S6C). Comparison of Stat1/2-perturbed versus wild-type cells within and across cell types revealed Stat1/2-dependent antiviral genes that were either common to various myeloid cell types (e.g., Irf7 and Isg15) or associated with specific cell types (Cxcl10, Ifi204, and Ifi27l2b) (Figures 6E and S6C). Together, these data show that the CRISP-seq technology is a powerful tool to elucidate the function of genes and pathways within various cell types in specific immune niches. In the future, CRISP-seq analysis of a larger set of developmental and signaling-dependent factors, in combination with different environmental and small molecule

(E and F) Scatterplots show the differentially expressed genes in control, Rela KO (E), and Irf9 KO (F) cells. Genes differentially expressed with significant p values are colored red.

(G) Scatterplot shows gene fold change for Rela/Irf9 double-KO cells over unperturbed cells compared to the linear combination of each individual KO effect.

(H) Scatterplot compares $-\log_{10}$ p values (Mann-Whitney U test) for differentially expressed genes in Irf9 KO population versus the unperturbed population and Rela KO population versus the unperturbed population.



(legend on next page)

conditions, will enable engineering of immune cells toward desired responses within specific niches, including for improved immunotherapy.

DISCUSSION

Gene regulatory networks function as decision-making circuits of the cell. Functional characterization of the regulatory pathways controlling cell fate and response is critical for the development of the next generation of targeted and combination therapies. Specifically, the plasticity and diversity in immune cell types and their responses have limited progress toward linking specific regulators with immune outcomes. CRISPR-based technologies have dramatically accelerated these efforts; however, they are used either for individual perturbations, assuming homogeneity in the population, or to measure interactions in particular loci (Parnas et al., 2015). Despite these important efforts, a robust technology that would systematically decipher the function of genetic elements at single-cell and genome-wide resolutions is still lacking. Here we describe a new and versatile method, CRISP-seq, which identifies in the same cell the specific perturbation and cell state. By generating a scalable lentiviral backbone that contains, in addition to the gRNA module, a fluorescent marker and a transcribed UGI, we show that CRISP-seq enables one to uncover in a single experiment the function of dozens of factors and their combinations. We successfully applied CRISP-seq to uncover regulators of cell state and response of myeloid cells, and we showed that our unique design is broadly applicable for both in vitro models and, importantly, for studying the regulation of immune cells in specific niches within animal models.

Biological circuits buffer, amplify, or temporally regulate signals from the environment through dedicated circuit designs. Genetic interactions can reveal functional relationships between genes and pathways in different cellular contexts, i.e., cell types and environmental conditions (Alon, 2006). Synthetic lethality screens in yeast and mammalian cells have greatly enhanced our understanding of the function of a large number of genes in various pathways, such as cytoskeletal organization in yeast (Tong et al., 2001), the KRAS pathway in cancer (Scholl et al., 2009), and the secretory pathway in viral budding (Blomen et al., 2015). We successfully applied CRISP-seq for multiplexed perturbations of an antiviral and inflammatory regulator and showed the potentially broad utility of CRISP-seq to screen for many-by-many interactions, limited only by the number of single cells that can be sequenced. Potentially, by using different fluorescent reporters, it is possible to sort via FACS the interacting markers of one or a few factors against a large pooled library of another fluorescent marker (few-by-many interactions). Such screens would effectively identify the genetic interactions of a smaller set of factors against a large genetic pool.

Our experiments also highlight several remaining challenges. First, more accurate identification of the UGI in every cell will improve the efficiency of CRISP-seq. This will be achieved by the expected improvements in the sensitivity of single-cell methods and by the second generation of CRISP-seq vectors with higher UGI expression levels. Second, although we have optimized the BFP and mCherry fluorescent markers used in CRISP-seq, additional fluorescent markers will enable multiplexing several in vivo experiments in one animal and focusing on specific interactions. Third, an additional gRNA expression cassette (multiplex gRNA vector) can be intuitively integrated into the CRISP-seq toolkit and further expand multiplexing of perturbations in single cells as well as testing for desired combinations (Sakuma et al., 2014; Vidigal and Ventura, 2015). Fourth, the analytical challenge of identifying false positives/negatives as well as of correctly modeling pathway hierarchies will require the development of new algorithms that optimally deal with the sparse and noisy nature of the data, such as a variant of the kNN graphs used here. Fifth, while our study focused on a small circuit of dozens of factors, the number of genetic elements that can be profiled in one experiment can be augmented significantly. As a rule of thumb, for most genes we found that ~50–100 cells are sufficient to accurately identify the perturbation phenotype. Sixth, while we developed CRISP-seq on the massively parallel RNA single-cell sequencing (MARS-seq) platform, it can be intuitively adopted for drop-based methods. Despite these limitations, our proof-of-principle study shows that CRISP-seq can now be applied to investigate the function of a large number of genetic elements and their interactions.

When profiling bulk cell populations following perturbations, a homogeneous population is tacitly and, in most cases, wrongly assumed. Hence, conclusions on gene function may be the misleading result of population heterogeneity. Our application of CRISP-seq to characterize the myeloid compartment following stimulation identified such heterogeneity in immune differentiation and response, highlighting the potential of using CRISP-seq for accurate dissection and characterization of immune circuits. The approach is not limited to coding genes or gene knockdown, but it can be used to perturb other genetic elements, such as non-coding RNA, promoters, and enhancers, as well as to induce specific gene expression, using a nuclease-dead Cas9 protein fused to a transcriptional activation domain together with a pool of indexed gRNAs targeting specific gene promoters. CRISP-seq also can be intuitively scaled in terms of the function of different circuit components under different environmental conditions. We have tested here two conditions, namely unstimulated and LPS-stimulated cells. However, the

Figure 5. Perturbations of Developmental and Signaling-Dependent TFs Reveal the Rewiring of Regulatory Circuits in Myeloid Cells

(A) Cxcl10 expression per single cell is shown for a pool of 57 gRNAs targeting signaling-dependent TFs in 2,600 UGI-confident myeloid cells stimulated with LPS for 4 hr. Cells are grouped by detected gRNA/UGI and ordered by expression level in each group. Red line, median expression per gRNA/UGI.

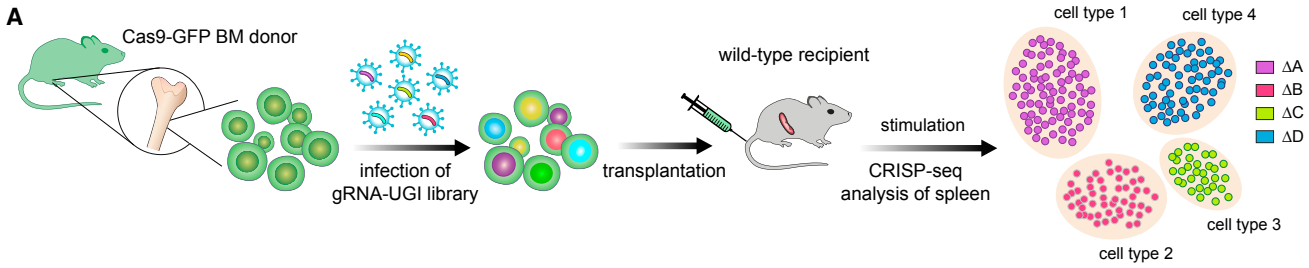
(B) PhenoGraph clustering of 6,749 myeloid cells stimulated with LPS for 4 hr is shown.

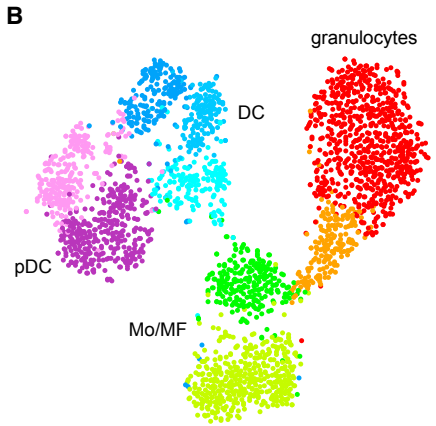
(C) Projection of the gRNA enrichments after UGI label refinement onto the kNN graph.

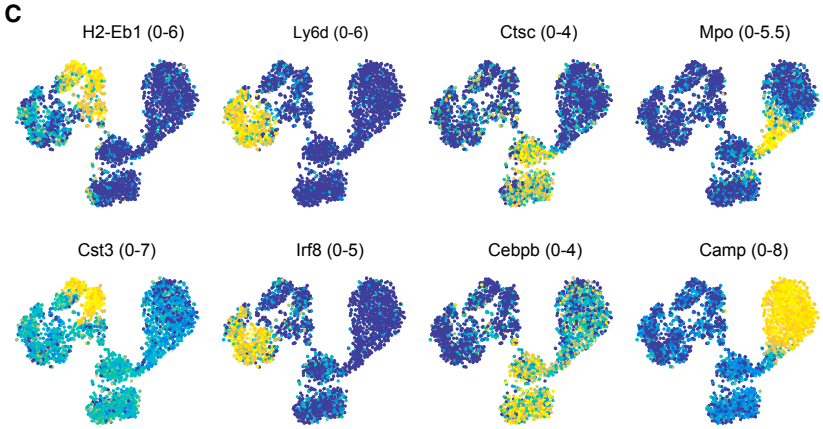
(D) Histograms of 5,674 monocyte cells depict the expression of key inflammatory and antiviral response genes in the different KOs.

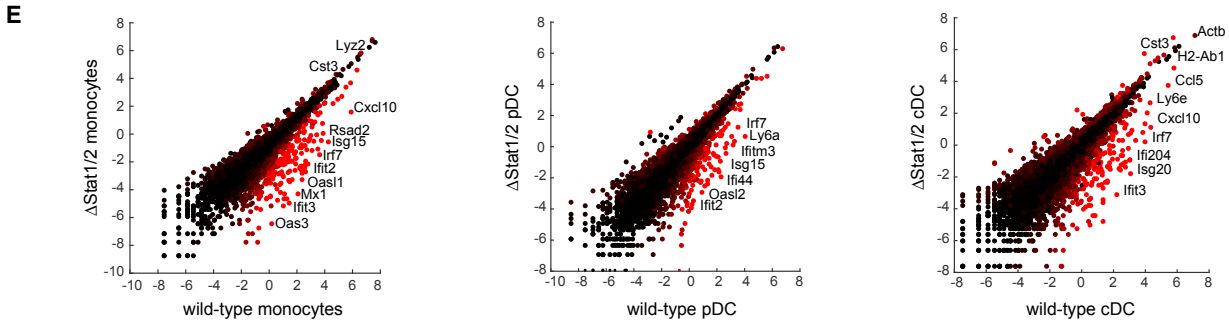
(E) ChIP-seq genome browser view of Stat1/2- and Rela-binding pattern in monocytes of key inflammatory and antiviral genes.

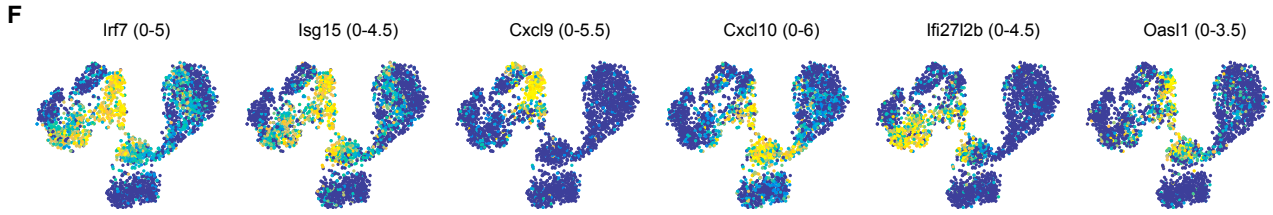
(F) Correlation of ChIP-seq binding (max peak height) compared to the transcriptional fold change upon perturbation of Stat1, Stat2, or Rela.











(legend on next page)

pool of gRNA-perturbed cells can be stimulated by different conditions or treated with various small molecules, with considerably greater flexibility and scalability than other approaches. Considering the high plasticity of myeloid and other immune cells, our results demonstrate the potential of using CRISP-seq to engineer immune cells for desired outcomes, immunotherapy, and other applications.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Cloning of guide RNA and Unique gRNA Index (UGI)
  - Lentivirus production
  - Isolation and culture of bone marrow-derived myeloid cells
  - Flow cytometry and single-cell capture
  - In Vivo CRISP-seq Assay
  - CRISP-seq library preparation
  - Sequencing and low-level processing
  - UGI-seq low-level processing
  - Graph-based clustering analysis
  - Graph based label refinement algorithm
  - Perturbation fold change analysis
  - CRISPR/Cas9 editing assessment (Indel-seq analysis)
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.11.039.

AUTHOR CONTRIBUTIONS

D.A.J. developed experimental protocols; designed, performed, and analyzed experiments; and wrote the manuscript. A.W. designed and analyzed experiments, developed computational methods, performed bioinformatic analysis, and wrote the manuscript. I.Y. designed and performed experiments and wrote the manuscript. D.L.-A. developed experimental protocols. H.K.-S., T.M.S., and E.D. contributed to the experiments. A.T. and A.v.O. contributed to project design and the analytical framework. I.A. directed the project, designed experiments, interpreted results, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Dana Peer for discussions on CRISP-seq analysis and Feng Zhang and Ophir Shalem for critical advice and reagents. We also thank Genia Brodsky for help with the artwork. A.v.O. is supported by European Research Council Advanced Grant ERC-AdG 294325-GeneNoiseControl. I.A. is supported by the European Research Council Consolidator Grant (ERC-COG) 724471-HemTree2.0; the Israel Science Foundation; the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine; the Helen and Martin Kimmel award for innovative investigation; a Minerva Stiftung research grant; the Israeli Ministry of Science, Technology, and Space; the David and Fela Shapell Family Foundation; and the Abramson Family Center for Young Scientists. I.A. is the incumbent of the Alan and Laraine Fischer Career Development Chair. A patent application has been filed related to this work, and the authors have made the CRISP-seq vectors widely available to the academic community through Addgene.

Received: October 27, 2016
Revised: November 14, 2016
Accepted: November 19, 2016
Published: December 15, 2016

REFERENCES

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783–801.

Alon, U. (2006). An introduction to systems biology: design principles of biological circuits (Boca Raton, Fla.; London: Chapman & Hall/CRC).

Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. Science *326*, 257–263.

Amit, I., Regev, A., and Hacohen, N. (2011). Strategies to discover regulatory circuits of the mammalian immune system. Nat. Rev. Immunol. *11*, 873–880.

Becker, A.M., Michael, D.G., Satpathy, A.T., Sciammas, R., Singh, H., and Bhattacharya, D. (2012). IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors. Blood *119*, 2003–2012.

Beyer, A., Bandyopadhyay, S., and Ideker, T. (2007). Integrating physical and genetic maps: from genomes to interaction networks. Nat. Rev. Genet. *8*, 699–710.

Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092–1096.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. *2008*, P10008.

Figure 6. In Vivo CRISP-Seq Analyses Uncover the Complexity of Myeloid Regulatory Circuits in Immune Niches

(A) Schematics of the in vivo CRISP-seq experimental pipeline. A pool of gRNA targeting genes of interest is infected into hematopoietic stem cells from Cas9-GFP+ donor mice and transplanted into recipient mice. Following LPS stimulation, immune cells are sorted from specific immune niches, and fluorescent markers are used for the selection of relevant cells. CRISP-seq analysis elucidates the genotype-to-phenotype relations within the specific niche.

(B) The kNN graph of 2,768 myeloid cells sorted from mice spleen following stimulation with LPS for 4 hr. Colors depict different cell types and states. Two pDC states (purple), three cDC states (blue), two monocyte states (green), and two neutrophil states (red/orange) are shown.

(C) Projection of cell type marker genes onto the kNN graph is shown. The expression range (RMT counts) is indicated next to the gene name.

(D) Projection of the gRNA enrichments after UGI label refinement onto the kNN graph. Stat1/2 KO cells are overlapping a pDC state (light purple), a cDC state (dark blue), and a monocyte state (light green).

(E) Scatterplots showing the differentially expressed genes (log2 mean expression) in control as compared to Stat1/2 KO in monocytes (left), pDCs (middle), and cDCs (right). Genes differentially expressed with significant p values are colored red.

(F) Projection of key antiviral response genes onto the kNN graph. Stat1/2 KO cells show a perturbed antiviral response in the different cell types.

Carter, H., Hofree, M., and Ideker, T. (2013). Genotype to phenotype via network analysis. Curr. Opin. Genet. Dev. *23*, 611–621.

Chattopadhyay, P.K., Gierahn, T.M., Roederer, M., and Love, J.C. (2014). Single-cell technologies for monitoring immune systems. Nat. Immunol. *15*, 128–135.

Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246–1260.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Feng, R., Desbordes, S.C., Xie, H., Tillo, E.S., Pixley, F., Stanley, E.R., and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc. Natl. Acad. Sci. USA *105*, 6057–6062.

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol. Cell *47*, 810–822.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. Science *327*, 656–661.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nat. Rev. Immunol. *14*, 392–404.

Girvan, M., and Newman, M.E. (2002). Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA *99*, 7821–7826.

Glass, C.K., and Natoli, G. (2016). Molecular control of activation and priming in macrophages. Nat. Immunol. *17*, 26–33.

Gosselin, D., Link, V.M., Romanoski, C.E., Fonseca, G.J., Eichenfield, D.Z., Spann, N.J., Stender, J.D., Chun, H.B., Garner, H., Geissmann, F., and Glass, C.K. (2014). Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell *159*, 1327–1340.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Gury-BenAri, M., Thaiss, C.A., Serafini, N., Winter, D.R., Giladi, A., Lara-Astiaso, D., Levy, M., Salame, T.M., Weiner, A., David, E., et al. (2016). The spectrum and regulatory landscape of intestinal innate lymphoid cells are shaped by the microbiome. Cell *166*, 1231–1246.

Hashimoto, D., Miller, J., and Merad, M. (2011). Dendritic cell and macrophage heterogeneity in vivo. Immunity *35*, 323–335.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Helft, J., Böttcher, J., Chakravarty, P., Zelenay, S., Huotari, J., Schraml, B.U., Goubau, D., and Reis e Sousa, C. (2015). GM-CSF mouse bone marrow cultures comprise a heterogeneous population of CD11c(+)MHCII(+) macrophages and dendritic cells. Immunity *42*, 1197–1211.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Junker, J.P., and van Oudenaarden, A. (2014). Every cell is special: genome-wide studies add a new dimension to single-cell biology. Cell *157*, 8–11.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357–360.

Klages, N., Zufferey, R., and Trono, D. (2000). A stable system for the high-titer production of multiply attenuated lentiviral vectors. Mol. Ther. *2*, 170–176.

Kurotaki, D., and Tamura, T. (2016). Transcriptional and epigenetic regulation of innate immune cell development by the transcription factor, interferon regulatory factor-8. J. Interferon Cytokine Res. *36*, 433–441.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R., Toledo, E.M., Villaescusa, J.C., et al. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. Cell *167*, 566–580.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell *159*, 1312–1326.

Lavin, Y., Mortha, A., Rahman, A., and Merad, M. (2015). Regulation of macrophage development and function in peripheral tissues. Nat. Rev. Immunol. *15*, 731–744.

Lawrence, T., and Natoli, G. (2011). Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nat. Rev. Immunol. *11*, 750–761.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell *162*, 184–197.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. Nature *449*, 819–826.

Moignard, V., and Göttgens, B. (2014). Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. BioEssays *36*, 419–426.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. *15*, 363–372.

Napolitani, G., Rinaldi, A., Bertoni, F., Sallusto, F., and Lanzavecchia, A. (2005). Selected Toll-like receptor agonist combinations synergistically trigger a T helper type 1-polarizing program in dendritic cells. Nat. Immunol. *6*, 769–776.

Okabe, Y., and Medzhitov, R. (2014). Tissue-specific signals control reversible program of localization and functional polarization of macrophages. Cell *157*, 832–844.

Ouaaz, F., Arron, J., Zheng, Y., Choi, Y., and Beg, A.A. (2002). Dendritic cell development and survival require distinct NF-kappaB subunits. Immunity *16*, 257–270.

Parnas, O., Jovanovic, M., Eisenhaure, T.M., Herbst, R.H., Dixit, A., Ye, C.J., Przybylski, D., Platt, R.J., Tirosh, I., Sanjana, N.E., et al. (2015). A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. Cell *162*, 675–686.

Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell *163*, 1663–1677.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Sakuma, T., Nishikawa, A., Kume, S., Chayama, K., and Yamamoto, T. (2014). Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. Sci. Rep. *4*, 5400.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Schlitzer, A., Sivakamasundari, V., Chen, J., Sumatoh, H.R., Schreuder, J., Lum, J., Malleret, B., Zhang, S., Larbi, A., Zolezzi, F., et al. (2015). Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. Nat. Immunol. *16*, 718–728.

Scholl, C., Fröhling, S., Dunn, I.F., Schinzel, A.C., Barbie, D.A., Kim, S.Y., Silver, S.J., Tamayo, P., Wadlow, R.C., Ramaswamy, S., et al. (2009). Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. Cell *137*, 821–834.

Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.M., Andréasson, A.C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. *24*, 593–607.

Semrau, S., and van Oudenaarden, A. (2015). Studying lineage decision-making in vitro: emerging concepts and novel tools. Annu. Rev. Cell Dev. Biol. *31*, 317–345.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Sternberg, S.H., and Doudna, J.A. (2015). Expanding the biologist's toolkit with CRISPR-Cas9. Mol. Cell *58*, 568–574.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat. Commun. *6*, 8083.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Yosef, N., and Regev, A. (2016). Writ large: genomic dissection of the effect of cellular environment on immune response. Science *354*, 64–68.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zufferey, R., Donello, J.E., Trono, D., and Hope, T.J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J. Virol. *73*, 2886–2892.

STAR★METHODS

KEY RESOURCES TABLE

REAGENTS or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse CD11c APC (clone N418)	eBioscience	17-0114-81; RRID: AB_469345
Anti-mouse CD11b PerCP/Cy5.5 (clone M1/70)	Biolegend	101228; RRID: AB_893232
Anti-mouse CD117 (c-kit) APC (clone 2B8)	eBioscience	17-1171-81; RRID: AB_469429
anti-mouse Sca-1 (Ly-6A/E) BV605 (clone D7)	Biolegend	108133; RRID: AB_2562275
Anti-mouse B220 Pacific Blue (clone RA3-6B2)	Biolegend	103227; RRID: AB_492876
Anti-mouse CD11b Pacific Blue (clone M1/70)	Biolegend	101224; RRID: AB_755986
Anti-mouse CD19 Pacific Blue (clone 6D5)	Biolegend	115523; RRID: AB_439718
Anti-mouse CD3 eFluor 450 (clone 17A2)	eBioscience	48-0032-82; RRID: AB_1272193
Anti-mouse CD4 eFluor 450 (clone GK1.5)	eBioscience	48-0041-82; RRID: AB_10718983
Anti-mouse CD8a eFluor 450 (clone 53-6.7)	eBioscience	48-0081-82; RRID: AB_1272198
Anti-Mouse Ly-6G (Gr-1) eFluor 450 (clone RB6-8C5)	eBioscience	48-5931-82; RRID: AB_1548788
Anti-mouse Ter-119 eFluor 450 (clone TER-119)	eBioscience	48-5921-82; RRID: AB_1518808
Anti-mouse Ly-6G APC/Cy7 (clone 1A8)	Biolegend	127624; RRID: AB_10640819
Anti-mouse CD3 PE (clone 17A2)	Biolegend	100205
Anti-mouse CD19 PE (clone 6D5)	Biolegend	115508; RRID: AB_313643
Anti-mouse TCR-beta PE (clone H57-597)	Biolegend	109208; RRID: AB_313431
Anti-mouse NK-1.1 PE (clone PK136)	Biogems	83712-60-25
CD117 MicroBeads, mouse	Miltenyi Biotec	130-091-224
Chemicals, Peptides, and Recombinant Proteins		
LPS-EK Ultrapure (used for in vitro cell stimulation)	Invivogen	tlrl-peklps
LPS *E. coli* 055:B5 (used for in vivo experiments)	Sigma-Aldrich	L2880
Ampicillin	Sigma-Aldrich	A1593
jetPEI, DNA in vitro transfection reagent	Polyplus Transfection	101-10N
Red blood cell (RBC) lysis solution	Sigma	R7757
GM-CSF	PeproTech	315-03-20
Murine TPO	PeproTech	315-14-10
Murine IL-3	PeproTech	213-13-10
Murine SCF (stem cell factor)	PeproTech	250-03-50
Commercial Kits		
QIAprep Spin Miniprep Kit	QIAGEN	27104
ZymoPURE Plasmid Maxiprep Kit	Zymo research	D4203
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	740609
Cloning Reagents		
BamHI	NEB	R0136S
Gibson Assembly Master Mix - 50 rxns	NEB	E2611L
FastDigest BsmBI (Esp3I)	Thermo Scientific	FD0454
FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific	EF0651
T4 Polynucleotide Kinase	NEB	M0201L
Quick ligase	NEB	M2200S
Stable Competent E. Coli (High Efficiency)	NEB	C3040H
GoTaq Green Master Mix	Promega	M7122
CRISP-Seq Library Reagents		
Superscript III	Invitrogen	18080-085
Exonuclease I	NEB	M0293

(*Continued on next page*)

Continued		
REAGENTS or RESOURCE	SOURCE	IDENTIFIER
Second Strand Synthesis module	NEB	E6111
T7 High Yield RNA Synthesis Kit	NEB	E2040
Turbo DNase I	Ambion	AM2239
RNA fragmentation reagents	Ambion	AM8740
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	M0204
AffinityScript Multi-Temp RT	Agilent	600109
Kapa HiFi HotStart PCR ReadyMix	Kapa Biosystems	KK2601
Sequence-Based Reagents		
All DNA oligonucleotides and primers were purchased from Integrated DNA Technologies (IDT), and appear below in dedicated tables.		
Deposited Data		
LentiGuide-Puro	Addgene	Addgene: 52963
CRISP-seq BFP backbone	Addgene	Addgene: 85707
CRISP-seq mCherry backbone	Addgene	Addgene: 85708
Raw data files for single-cell CRISP-seq	NCBI GEO	GEO: GSE90486
Raw data files for Indel-seq	NCBI GEO	GEO: GSE90487
Raw data files for CHIP-seq	NCBI GEO	GEO: GSE36104
Experimental Models: Organisms/Strains		
Cas9-GFP transgenic mice	The Jackson Laboratories	024858
C57BL/6JOlaHsd wild-type mice	Harlan	705
Software and Algorithms		
MATLAB R2016a software	MathWorks	http://mathworks.com/
Hisat	(Kim et al., 2015)	https://github.com/infphilo/hisat
PhenoGraph	(Levine et al., 2015)	https://www.c2b2.columbia.edu/danapeerlab/html/cyt-download.html
R version 3.2.3 (2015-12-10)	The R Project	https://www.r-project.org/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by Lead Contact Ido Amit (ido.amit@weizmann.ac.il).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Cas9-GFP transgenic mice were previously described (Platt et al., 2014). A founding breeding pair was purchased from The Jackson Laboratory. These mice were bred in the Weizmann Institute animal facility and backcrossed with wild-type black (C57BL/6); their progeny was crossed to produce Cas9-GFP homozygotes on a cleaner C57BL/6 background. In all experiments, wild-type black or Cas9-GFP young adult (7-11 weeks old) females were used. Mice were provided with food and water ad libitum and housed under a strict 12 hr light-dark cycle. All experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC).

METHOD DETAILS

Cloning of guide RNA and Unique gRNA Index (UGI)

For targeted loss-of-function screening using cell cytometry, we used the lentiviral vector lentiGuide-Puro (Platt et al., 2014) (plasmid #52963, Addgene) and replaced the puromycin resistance marker coding sequence (CDS) with either fluorophore EBFP or mCherry CDS. The gRNA cannot be identified during single-cell gene expression library construction, due to its short size and lack of a poly-adenylation tail. Therefore, to detect the gRNA in single cells in experiments where a mix of lentiGuide vectors with different gRNAs are used, we designed a unique gRNA identifier (UGI) barcode to be expressed at the 3′ end of the fluorophore transcript, immediately downstream to the Woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) (Zufferey et al., 1999) and upstream to

the polyadenylation signal in the lentiviral construct (Figure 1A). A library of UGIs, located 372 base-pairs (bp) upstream to the SV40 poly(A) signal, is processed in parallel to single-cell gene expression profiling (Jaitin et al., 2014), where the cell identifier links the cell to the gRNA through the UGI.

To incorporate UGIs in the lentivectors, we first introduced a BamHI restriction site immediately downstream to the WPRE, creating the CRISP-seq (BFP or mCherry) backbone vectors. Next, the random (N) 8 bp-long UGI sequences were synthesized in single stranded DNA (ssDNA) oligonucleotides with flanking sequences of homology with to a vector insertion site ("UGI coding oligo" in the table below). The insertion of the barcode sequences was performed by Gibson assembly cloning (Gibson et al., 2009). The CRISP-seq backbone vectors were digested with the restriction enzyme BamHI (NEB) and dephosphorylated using FastAP (Thermo Fisher): 6 μl 10X NEBuffer 3.1, 3 μl FastAP, 3 μl BamHI, 3 μg plasmid, water up to 60 μl. Incubate at 37°C for 1 hr. The digested plasmid was purified using a Nucleospin column (Macherey-Nagel), and Gibson assembly reactions were performed: 150 ng cut vector, 20 ng UGI oligo, 7.5 μl 2X Gibson master mix, water up to 15 μl. Incubate at 50°C for 1 hr. 10 μl were transformed into competent bacteria, which were plated on five 100 mg/L ampicillin LB agar plates. Plates were incubated overnight at 30°C and > 5,000 colonies were collected by scraping the plate surface and suspension in sterile media. A plasmid pool of UGIs was then purified from the suspended colonies using a plasmid maxiprep kit (Zymo research).

Guide RNA oligos were cloned into the CRISP-seq vectors as previously described (Sanjana et al., 2014). Briefly, gRNAs synthesized in pairs of oligonucleotides (IDT) with BsmBI-compatible overhangs were phosphorylated with T4 polynucleotide kinase (NEB) and annealed. The fragments were then ligated with purified, BsmBI-digested CRISP-seq plasmids. Ligated constructs were transformed into competent bacteria and single clones were picked and propagated. Finally, plasmids containing the fluorophore, a gRNA and a UGI were purified, and each gRNA was paired with its identifying UGI barcode by Sanger sequencing (see table below).

Primer name	Sequence
UGI coding oligo	ctccctttgggccgcctcCCCGCGTCGACGGATCCNNNNNNNNG acttacaaggcagctgtaga, N = random base
UGI Sanger sequencing	AATCCAGCGGACCTTCCT
gRNA Sanger sequencing	ATAATGATAGTAGGAGGCTTGG
fluorophore Sanger sequencing	TGGAGTACGTCGTCTTTAGG

A list of gRNA sequences used in this work appear in Table S5.

Lentivirus production

CRISP-seq-UGI lentiviral particles were produced by transfecting 293T cells growing in 15-cm. tissue culture dishes together with packaging plasmids, using the jetPEI transfection reagent (Polyplus-transfection) according to the manufacturer's instructions and following the standard lentivirus production protocol (Klages et al., 2000). Transfection efficiency was assessed by microscopic inspection of cell fluorescence one day later. Media was replaced with RPMI medium without additives 18 hr post transfection, and media containing virus particles were collected 48 and 72 hr post transfection. Virus particles from one or two plates (26mL per 293T plate) were concentrated using Amicon 100 KDa 15mL columns (Millipore) in a cold centrifuge at 1000 xg to a final concentration of 200-250 μl per virus, aliquoted and stored at −80°C until use.

Isolation and culture of bone marrow-derived myeloid cells

Mice were sacrificed by cervical dislocation. To isolate the bone marrow, femora and tibiae from one leg were removed, cleaned from flesh, and flushed with C10 culture medium (RPMI supplemented with 15% serum, 1% x100 non-essential amino acids, 10mM HEPES buffer, 1mM sodium pyruvate, 2mM L-glutamine, 1% L-glutaine and 50μM b-mercaptoethanol) using a G21 needle syringe. The bone marrow was filtered through a 70-μm cell strainer and spun down in a cold centrifuge at 300xg for 5 min. Cells were resuspended in 250 μl RBC lysis solution (Sigma) per leg and incubated for 5 min at room temperature, washed, and resuspended in C10 medium. Cultures were set by plating $6x10^5$ cells in 1mL C10 supplemented with 15 ng/ml GM-CSF in a 6-well non-tissue culture plate, and incubated under standard culture conditions (37°C, 5% CO_2). Cells were infected on culture day 2 by adding lentivirus and 8 μg/ml polybrene, and plates were centrifuged 1000 xg at 37°C for 45 min to enhance infection. At the end, 1mL C10 supplemented with GM-CSF was added. Cells were fed with 200 μl C10 supplemented with 30 ng/ml GM-CSF every second day.

Flow cytometry and single-cell capture

On day 7, cells were either treated with 100 ng/ml lipopolysaccharide (LPS) for 4 hr or left untreated as control. To obtain cell suspension, cells were scrapped from the well, washed and resuspended in cold FACS buffer (0.5% BSA and 2 mM EDTA in phosphate-buffered saline), stained with fluorophore-conjugated anti-mouse CD11c (and CD11b where indicated) antibody, and filtered through a 40-μm strainer. Cell sorting was performed using a BD FACSAria Fusion flow cytometer (BD Biosciences), gating for GFP (Cas9), and relatively high BFP or mCherry fluorescence (Figure S1C; CRISP-seq lentivirus-infected cells). This gate was chosen according to the maximal CD11b marker mean signal reduction in cells infected with Itgam-gRNA (the gene coding for CD11b), used to

calibrate the system. Single cells were sorted into 384-well capture plates containing 2 μl of lysis solution and barcoded poly(T) reverse-transcription (RT) primers for single-cell RNA-seq as described previously (Jaitin et al., 2014). To record marker levels of each single cell, the FACS Diva 7 "index sorting" function was activated during single-cell sorting. Results were exported into an Excel file containing the information about flow parameters of each cell (each well). Immediately after sorting, plates were spun down to ensure cell immersion into the lysis solution, snap frozen on dry ice and stored at −80°C until further processing.

In Vivo CRISP-seq Assay

Hematopoietic stem cells (HSCs) and multiple pluripotent progenitors (MPPs) were isolated from the bone marrow of Cas9-GFP donor mice, infected with a pool of CRISP-seq lentivirus containing the BFP fluorophore gene and different gRNAs, and injected into wild-type recipient mice (Figure 6A). Seven days post-transplantation, spleens were removed and single cell sorted for CRISP-seq analysis. On day 1, bone marrow from Cas9-GFP mice was isolated from mouse tibiae femora and ilia leg bones, filtered through a 70 mm cell strainer, and the cell suspension enriched for c-kit (CD117) expressing cells using magnetic cell separation (Miltenyi Biotec, Germany) according to manufacturer's instructions. Cells were then stained and FACS-sorted using a sterilized FACSAria Fusion cell sorter (BD Biosciences) into a tube already containing 500 μl of StemSpan supplemented with standard tissue culture penicillin/streptavidin (pen/strep) antibiotics. HSCs and MPPs were defined as GFP^+ Lin^- (lineage negative) Sca-1^+ c-kit^+ (LSK). Lineage markers included: anti-mouse Ter-119, Gr-1, CD11b, B220, CD19, CD3, CD4, and CD8, as previously described (Paul et al., 2015). About 90,000 donor cells were infected with a CRISP-seq lentivirus pool in tissue-culture 96-well plates for about 18 hr. The infection was carried out in 200 μl StemSpan Serum-Free Expansion medium (SFEM; StemCell technologies) supplemented with the cytokines Flt3, Il3, Tpo and SCF (stem cell factor), each at 10 ng/ml, and pen/strep antibiotics. On day 2, prior to bone marrow transplantation, wild-type recipient 8 weeks-old female mice were lethally irradiated with two subsequent X-ray doses of 550 cGy and 500 cGy that were administered 3 hr apart. After irradiation, drinking water was supplemented with 200 mg ciprofloxacin/ml. Four to six hours post irradiation, the donor cells were collected, washed twice with PBS and injected intravenously (tail injection) together with 200,000 recipient isogenic (wild-type) flushed whole bone marrow carrier cells for myeloprotection, 200 μl per mouse. On day 8, recipient mice were IP injected with either LPS (Sigma), 1 ng per 5 gr mouse, or PBS and 4 hr later, mice were sacrificed, their spleen extracted and dissociated into single splenocytes with a gentleMACS Dissociator (Miltenyi Biotec, Germany), filtered through a 70-μm strainer and incubated 5 min. in RBC lysis solution (Sigma). After washing, cells were stained and single-cell sorted as as described above. The myeloid niche coming from the donor was defined as GFP+ BFP+ lymphocyte-negatives (as CD19- TCRb- CD3- NK1.1-), and CD11c+ (DCs), CD11b+ or Ly-6G+ (granulocytes).

CRISP-seq library preparation

Libraries of single-cell gene expression (MARS-seq) and single-cell gRNA detection (UGI-seq) together with CRISP-seq, were prepared in parallel. For automated library production, Bravo robot station was used in combination with Nanodrop Express (BioNex, San Jose, CA). MARS-seq libraries were prepared as previously described (Jaitin et al., 2014). Briefly, mRNA from sorted cells was simultaneously barcoded, converted into cDNA and pooled using an automated pipeline. The pooled samples were then linearly amplified by T7 in vitro transcription (IVT). After DNase I treatment, the samples were cleaned up with 1.2x SPRI beads and the amplified RNA (aRNA). Half of the aRNA was fragmented and converted into a sequencing-ready library by tagging the samples with pool barcodes and Illumina sequences during ligation, RT, and PCR. For the corresponding gRNA information in each cell, a UGI-seq library was obtained from 10%–12% of the aRNA material, and processed in parallel as follows: Fragmentation was skipped and ligation was done together with MARS-seq samples, using the UGI ligation primer (see table below; pool barcode was added at a later step). Ligation cleanup and the subsequent reverse transcription (RT) reaction were the same as for MARS-seq samples, except for the use of a different RT primer (see table below). Then, an intermediate 10-cycle PCR step was done to amplify and add pool barcodes, using a barcoded forward primer and the reverse primer used in MARS-seq final step (see table below); PCR conditions were the same as in MARS-seq. Finally, another PCR reaction, as in MARS-seq, was done to complete and enrich the UGI-seq library. The resulting CRISP-seq product is a MARS-seq library and a corresponding UGI-seq library. Library quality assessment and concentration measurements were performed as previously described (Jaitin et al., 2014). The table below shows the sequence of primers used for UGI-seq library construction.

Primer name	Sequence
UGI ligation adaptor	ATGATCAAGCGACCACCGAG, modified with a phosphate group at the 5′ end, and a C3 spacer (blocker) at the 3′ end
Second RT UGI primer	CTCGGTGGTCGCTTGATCAT
Barcoded PCR forward	CTACACGACGCTCTTCCGATCTNNNNNXXXXTCCCCGCGTCGACGGATC, N = random base and XXXX = 4-bases plate barcode
P7_Rd2 PCR reverse	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Sequencing and low-level processing

CRISP-seq libraries, pooled at equimolar concentrations, were sequenced using an Illumina NextSeq 500 sequencer, at a sequencing depth of 60K-80K reads per cell for MARS-seq and about 4K reads per cell for UGI-seq. Reads are condensed into original molecules by counting same random molecular tags (RMT, a.k.a. unique molecular identifier or UMI). We used statistics on empty-well spurious RMT detection to ensure that the batches we used for analysis showed a low level of cross-single-cell contamination (less than 1%).

CRISP-seq reads were processed as previously described (Gury-BenAri et al., 2016). Mapping of reads was done using HISAT (version 0.1.6); reads with multiple mapping positions were excluded. Reads were associated with genes if they were mapped to an exon, using the UCSC genome browser for reference. Exons of different genes that shared genomic position on the same strand were considered a single gene with a concatenated gene symbol. Cells with less than 1500 UMIs were discarded from the analysis. Genes with mean expression smaller than 0.005 RMTs/cell or with above average expression and low coefficient of variance (< 1.2) were also discarded.

UGI-seq low-level processing

Sequenced reads containing the UGI-seq 5′ primer (TCCCCGCGTCGACGGATCC) up to 2 bp mismatches were extracted for further UGI-seq processing. We first extract plate barcode, cell-specific barcode (7bp), Random Molecular Tags (RMT - 8bp) and Unique Guide Identifier (UGI – 8bp) for each read. Reads with low quality (Phred < 27) or without a valid UGI sequence (up 1 bp mismatch), cell barcode (up 1 bp mismatch), or plate barcode (exact match) were discarded. Sequencing errors within a RMT may undermine the UGI counts by creating spuriously identified molecules from real molecules; this number is expected to increase linearly with sequencing depth. As UGI molecules were over-sequenced, these 'satellite' reads were easily detectable, and real molecule reads (in log scale) were normally distributed with an average of 2^{10} duplicated triplets (cell barcode, RMT, UGI) and a standard deviation of 2. Triplets with less than 30 reads were discarded as errors ($p < 0.01$, see Figure S1D) and each cell received a vector of UGI molecule counts. In some analyses, we consider the total number of reads with the same cell barcode and UGI (e.g., y axis in Figure 1B). To assign a binary label per cell, we consider UGI molecules > 1 as positive cells (See Figure S1D).

Graph-based clustering analysis

In order to assess the heterogeneity of cells in the samples, we used the PhenoGraph clustering algorithm (Levine et al., 2015). Briefly, low-level processing of CRISP-seq reads results in a matrix U with n rows and m columns, where rows represent genes and columns represent cells. Entry Uij contains the number of unique molecular identifiers (UMIs) from gene i that were found in cell j. The first step of the algorithm is to build a graph structure from this expression matrix. PhenoGraph first builds a *k*-Nearest Neighbors (kNN) graph using the Euclidean distance (we chose k = 30 and tested k = 15, 20, 25, 30, 40, 50, and got very similar results, not shown) and then refines this graph with the Jaccard similarity coefficient, where the edge weight between each two nodes is the number of neighbors they share divided by the total number of neighbors they have (Levine et al., 2015). To partition the graph into modules/communities PhenoGraph uses the Louvain Method (Blondel et al., 2008).

The graph is constructed and partitioned into modules based on the expression profile of the cells. We can now overlay the genotype information obtained from UGI-seq to calculate the enrichment of gRNA within clusters. We calculate the UGI enrichment p value within each cluster using the hyper geometric distribution, where N is the total number of cells, K is the number cells with UGI_A, n is the size of cluster c_i and k is the number of cells with UGI_A in cluster c_i. The probability of drawing k or more cells with UGI_A is:

$$p = F(k \mid N, K, n) = \sum_{i=k}^{n} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}.$$

Graph based label refinement algorithm

UGI-seq provides information on the expression of the reporter gene introduced in our lentivirus construct. This information is translated to a specific gRNA which was integrated together with the reporter gene. This gRNA will target Cas9 to a specific gene locus, but only in 70%–80% will generate true loss-of-function of the targeted gene (Sternberg and Doudna, 2015). In other cases, Cas9 may generate a non-harmful mutation (such as in-frame deletion) or no mutation at all. This implies that in 20%–30% of the cells with a detected UGI, the gene can be active or partially active and show a wild-type phenotype (false positive). On the other hand, as single-cell data is sparse by nature, cells with true edited gene loss-of-function can remain undetected by UGI-seq, becoming false negative events. We quantify the single-cell RNA detection error as 20% by comparing UGI-seq to FACS-based detection of the BFP fluorophore. In order to overcome the noisy and missing genotype label problem we developed a label refinement algorithm that can modify the labels themselves. Our algorithm is based on the assumption that the labels (= genotype) of the cells are consistent with their nearest neighborhoods, i.e., that cells sharing the same knockout mutation will have similar phenotype and this phenotype is distinct from the wild-type phenotype. The input data dataset $S = (X,Y)$. The expression matrix is denoted as X,

where $X = \{x_1, x_2, \ldots, x_N\}$ each cell expression $x \in \mathcal{R}^M$ is an M-dimensional vector. Their corresponding UGI labels are $Y = \{y_1, y_2, \ldots, y_N\}$, where $y \in \{0,1\}^K$ is a binary vector representing the UGIs detected in each cell. Our algorithm refines each UGI label separately. Based on the expression matrix we first build a Jaccard graph, similar to a PhenoGraph construction of the graph. An initial kNN graph in constructed based on the Euclidean distance between cells and the Jaccard index is calculated for every pair of nodes. The weight between nodes i and j, is given by:

$$W_{ij} = \frac{|v(i) \cap v(j)|}{|v(i) \cup v(j)|}$$

where $v(i)$ is the k-neighborhood of cell i. Our two-step algorithm first remove labels which disagree with their neighborhood and then assign labels to cells with significant neighbor's enrichment. For each cell, we define a neighborhood score for each UGI u as the sum of the Jaccard coefficients with all other labeled nodes in the graph:

$$s(i) = \sum_{j \in V, y_j^u = 1} W_{ij}$$

(most coefficients will be zero as most cells do not share common neighbors in the kNN graph). To calculate the p value of observing this score at random, we used bootstrapping, shuffling the labels randomly 100K times and counting the number of times $s(i)$ is bigger than the score obtained in each shuffled graph. Labels were removed from cells with p value > 0.05. In the next step, we repeated this process with the new filtered labels and added labels for cells with p value < 0.001. Changing the bounds within a reasonable range (0.01-0.2 for filtering out labels and 0.01-1e-5 for adding labels) modified the total number of labeled cells, but they still remained in the same neighborhood.

Perturbation fold change analysis

We calculated the perturbation effect for each gene knockout by comparing the average profile of the perturbed cells with the corresponding control cells. Two different strategies were used to select the group of cells, either by selecting cells by clusters (For Figures 2, 3, 5, and 6), or by selecting cells based on their labels after applying the refinement algorithm (Analysis done in Figure 4). When using the first strategy, clusters were assigned to genotype based on the UGI enrichment score. Scatterplots show $\log_2$ mean RMT counts in each group. p values were calculated using the Mann-Whitney U test (MATLAB 2016a *ranksum* function).

CRISPR/Cas9 editing assessment (Indel-seq analysis)

Cell sorting

About 4,000 cells per sample were sorted into a microfuge tube already containing 500 μl of cold FACS buffer. Tubes were gently vortexed and cells were pelleted in a cold centrifuge, at 1,500xg for 15 min at 4°C, to aspirate most of the supernatant, leaving about 50 μl, and stored at −80°C until further processing.

Genomic DNA extraction

Cells were lysed by three cycles of freeze/thaw by 37°C and dry ice incubation of 3 min each. Then SDS was added to a final concentration of 0.5% and the samples were incubated for 5 min at room temperature. Then, samples were incubated in RNase, DNase-free (Roche), 0.5 μL per 50 μL sample, for 30 min at 37°C. Next, two units of proteinase K (NEB) and 5 nM EDTA were added and samples were incubated at 37°C for 2 hr, followed by incubation at 65°C overnight. Alternatively, samples were incubated at 37°C for 30 min and then at 95°C for 10 min. Genomic DNA was cleaned up using 2.5 volumes of SPRI beads, and mass concentration was measured in a Qubit fluorometer with high-sensitivity DNA reagents (ThermoFisher Scientific).

Indel-seq library construction

Libraries were constructed around each exon-specific region in two PCR reactions, using target-specific primers with Illumina partial tags as overhangs for PCR1, and a second PCR to amplify and add the missing parts for Illumina sequencing (see table below). PCR1 protocol: To 5 ng of genomic DNA add 2 μl primer mix at 10 μM each primer, 25 μl 2x KAPA high-fidelity PCR mix (KAPA Biosystems, Roche), 50 μl reaction volume, 28 cycles. PCR program: 2 min at 98°C, 2 min, 28 x [20 s. at 98°C, 30 s. at 60°C, 40 s. at 72°C], 5 min at 72°C, 4°C end. Clean up the PCR1 product with 40 μl of SPRI beads (0.8 volumes). Measure concentration and assess expected size in a TapeStation instrument using high-sensitivity DNA reagents (Agilent Technologies) before PCR2. PCR2 protocol: To 5 ng of PCR1 product, add 1 μl of 10 μM P5_Rd1 primer, 1 μl of 10 μM indexed reverse primer, choosing specific barcodes for each sample, 10 μl 2x KAPA high-fidelity PCR mix, 20 μl reaction volume, 5 cycles. PCR program: 2 min at 98°C, 2 min, 2 cycles x [20 s. at 98°C, 30 s. at 58°C, 45 s. at 72°C], 3 cycles x [20 s. at 98°C, 30 s. at 65°C, 45 s. at 72°C], 5 min at 72°C, 4°C end. Clean up the PCR2 product with one volume of SPRI beads. Measure molar concentration with Qubit and TapeStation. Indel-seq libraries were sequenced using a Miseq Illumina sequencer. The table below shows the sequence of the primers used for Indel-seq library construction.

Primer name	Sequence
Cebpb Indel-seq partial Rd1	ACACGACGCTCTTCCGATCTCCTGGTAGCCCAGGTAGGC
Cebpb Indel-seq partial Rd2	CTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCCGACCTCTTCGCCG
Itgam Indel-seq partial Rd1	ACACGACGCTCTTCCGATCTTGTCTGGTTAACAGCCTTTG
Itgam Indel-seq partial Rd2	CTGGAGTTCAGACGTGTGCTCTTCCGATCTCCATTTCCCATCCTAACTTC
P5-Rd1 forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
P7-i7-partial Rd2 reverse	CAAGCAGAAGACGGCATACGAGATXXXXXXXGTGACTGGAGTTCAGACGTGTGCT, XXXXXXX = 7 bases index

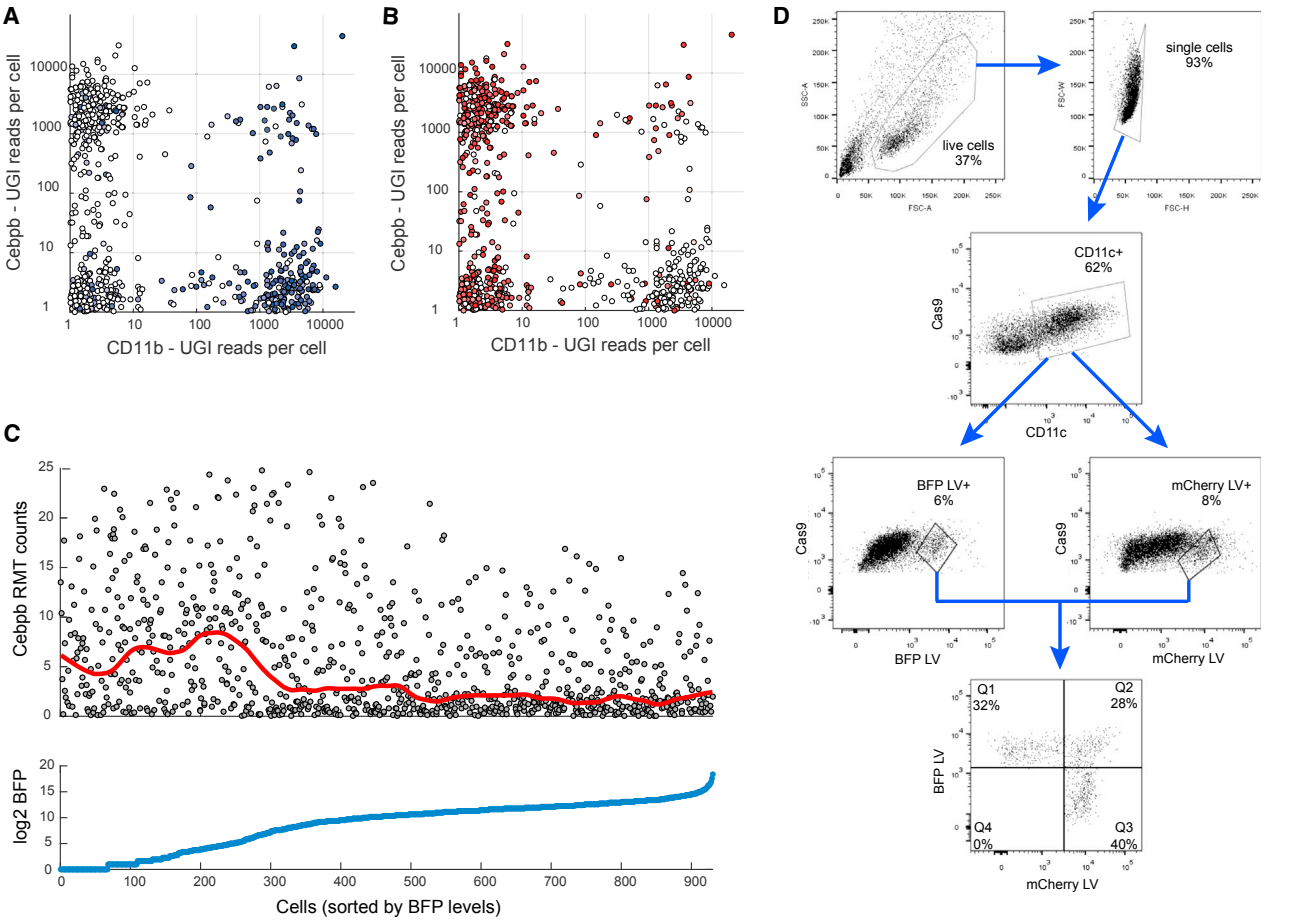
QUANTIFICATION AND STATISTICAL ANALYSIS

Significance of UGI enrichments in unsupervised clusters was calculated using the hypergeometric p value. A p value $< 10^{-3}$ was considered significant. Significance in gene expression differential was estimated using the non-parametric Mann-Whitney U test; all p values were corrected using FDR < 0.01. For in vivo experiments, at least two mice were analyzed as biological replicates.

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession number for the sequencing data reported in this paper is NCBI GEO: GSE90488. This parent directory includes the following datasets: GEO: GSE90486 (single-cell CRISP-seq) and GEO: GSE90487 (Indel-seq).



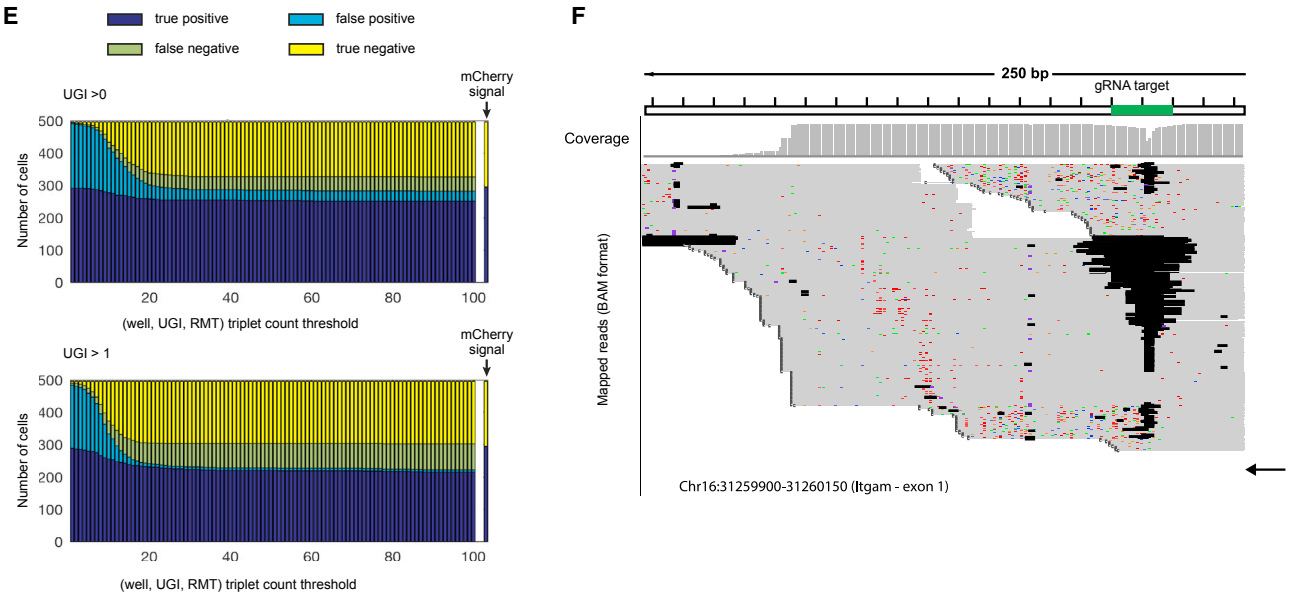


Figure S1. A UGI Strategy Is Effective in Detecting gRNA in Single Cells, Related to Figure 1

(A) Scatterplot showing BFP fluorescent reporter expression levels as recorded by FACS index sorting in each cell sequenced with CRISP-seq. Unique guide index (UGI) count for Itgam (CD11b) and for Cebpb are shown on the X and Y axes, respectively.

(legend continued on next page)

(B) Same as in (A), showing mCherry fluorescent reporter expression levels in the same experiment.
(C) Single cells sorted by BFP levels (bottom) show downregulation of Cebpb mRNA as BFP increases. Red line, Cebpb expression trend.
(D) FACS plot showing the gating strategy to identify CD11c+ myeloid cells. Single cells were sorted from either LV+ gate, as the union between the BFP and mCherry gates.
(E) Calculation of UGI-based detection sensitivity and specificity compared to the index sorting-based gRNA detection (mCherry signal, Cebpb-gRNA). Bar plot showing detection accuracy as a function of (well,UGI,RMT) triplet count threshold. Top panel consider UGI count > 0 as positive cells and bottom panel consider UGI count > 1 as positive. Rightmost bar shows the FACS positive/negative cells ratio.
(F) Interactive genome viewer (IGV) of mutations (indels, insertions, SNPs) in the genomic loci around the Itgam-gRNA target.

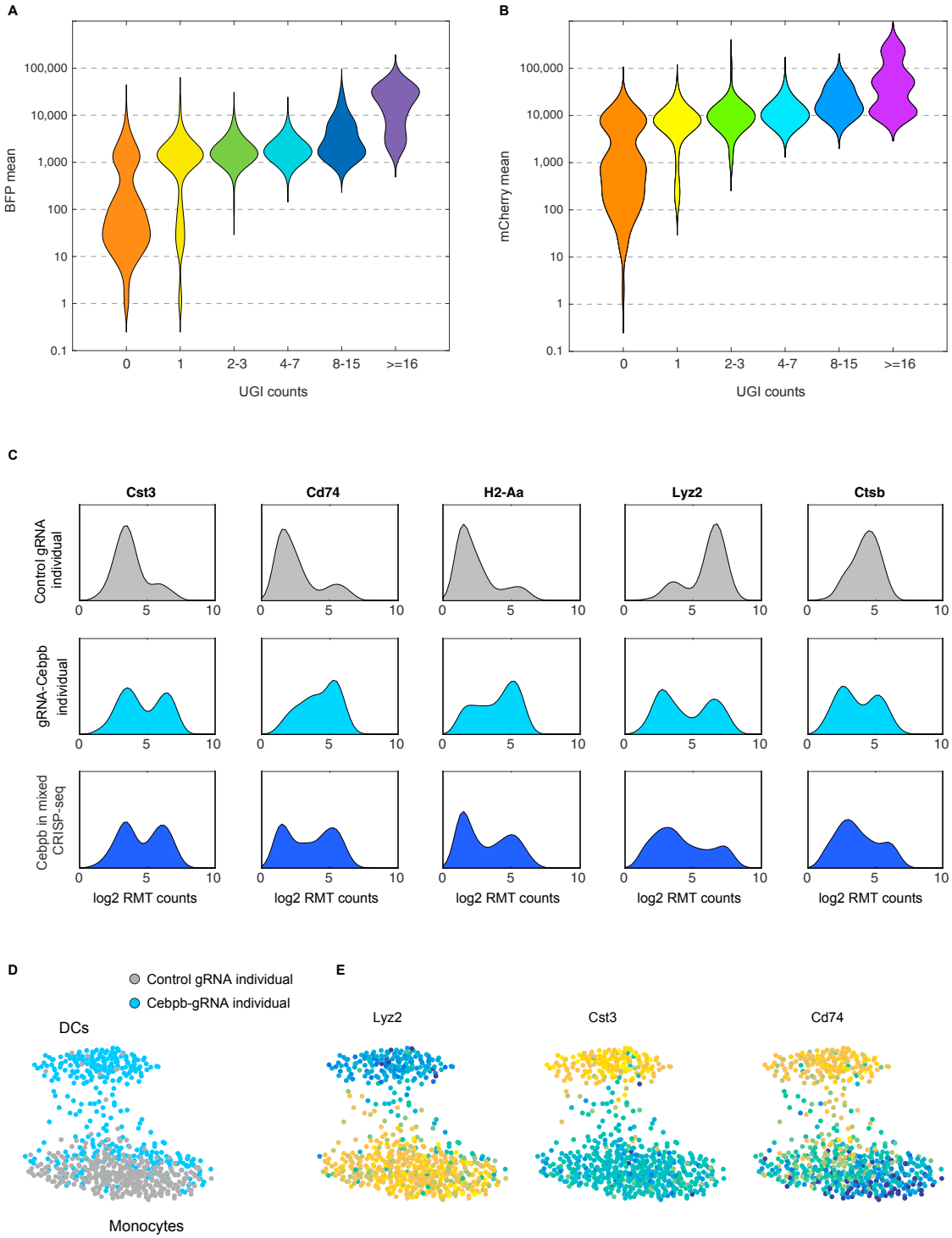


Figure S2. Validation of gRNA Detection from Single Cells Infected with a Pool of gRNAs, Related to Figure 2

(A and B) Violin plots of lentivirus marker expression in single cells labeled for BFP (Rela-gRNA) (A) or for mCherry (Cebpb-gRNA) (B) in binned UGI counts.

(C) Density histograms of representative genes in cells infected with a control gRNA (upper panels), in cells infected only with Cebpb-gRNA (individual; middle panels), and Cebpb UGI positive cells from CRISP-seq mix from Figure 2.

(D) tSNE plot of 349 cells infected separately with control gRNA (gray dots) combined with 359 cells infected with only Cebpb-gRNA.

(E) Projection of representative genes of each cell type onto the tSNE plot.

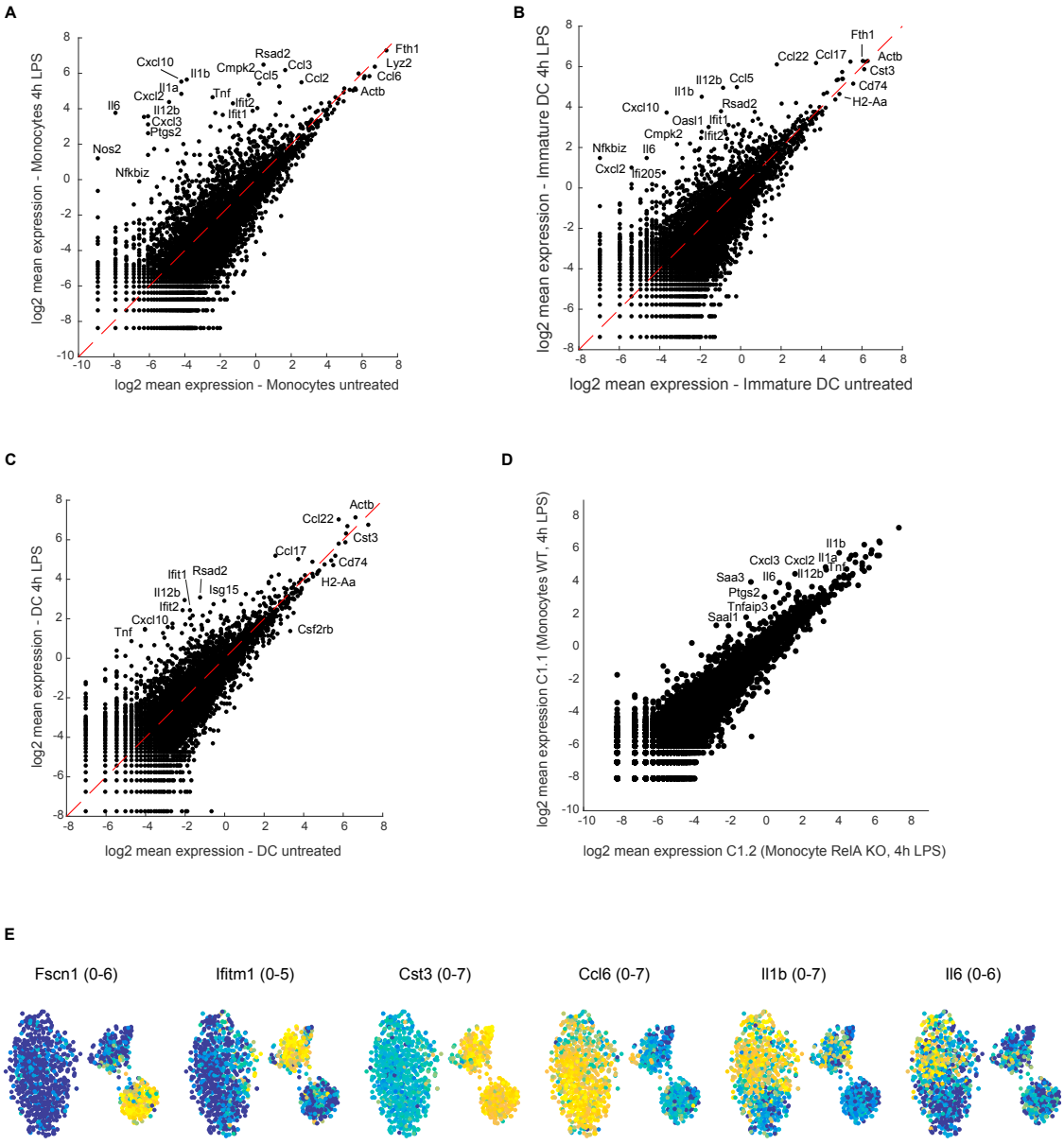


Figure S3. The Cebpb TF Is Essential for the Monocyte State, Related to Figure 3

(A–C) Scatterplots showing differentially expressed genes in 4h LPS stimulated cells compared to untreated cells in monocytes (A), immature DCs (B), and mature DCs (C).

(D) Scatterplots of mean expression in cluster 1.1 (unperturbed monocytes) versus cluster 1.2 (RelA KO monocytes).

(E) Projection of additional key marker genes onto the t-SNE plot (Figure 3E).

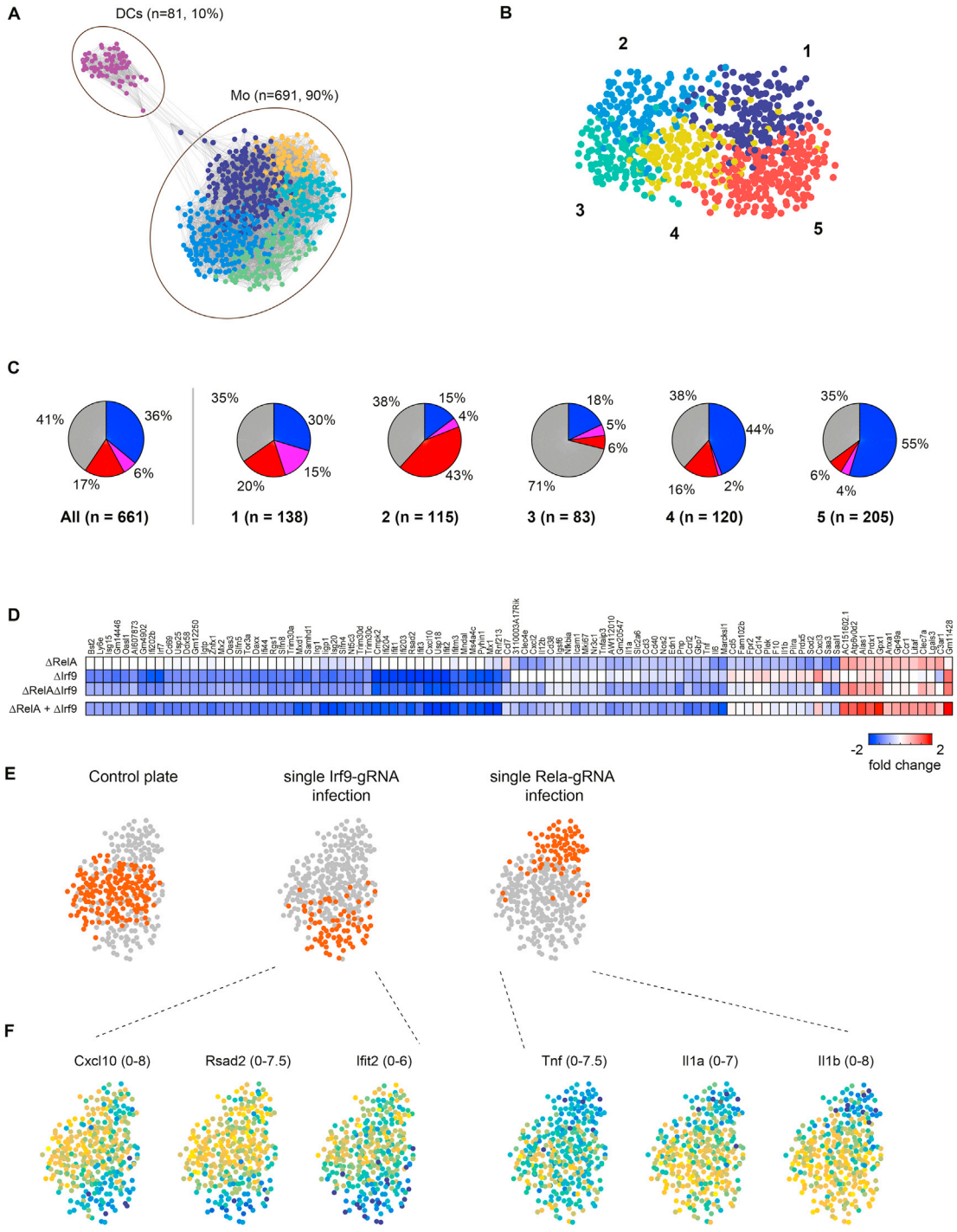


Figure S4. Multiplexed Perturbations of Irf9 and Rela in Single Cells, Related to Figure 4

(A) PhenoGraph clustering of 772 myeloid cells.

(B) Clustering analysis of 691 monocytes cells identified five distinct clusters, defined by their different UGI composition.

(C) Pie charts of the distribution of UGIs per cluster in (B).

(D) Heatmap showing fold change from unperturbed cells for RelA KO, Irf9 KO and RelA/Irf9 double KO. Bottom row correspond to levels computed as the sum of the individual KO.

(E) Combined tSNE plot of cells infected with individual gRNAs (three separate cultures): left, projection of cells infected with a control gRNA (red dots, 180 cells); middle, for Irf9-gRNA only (81 cells); right, for Rela-gRNA only (80 cells).

(F) Projection of representative inflammatory or antiviral gene expression levels onto the tSNE plot of individual plates.

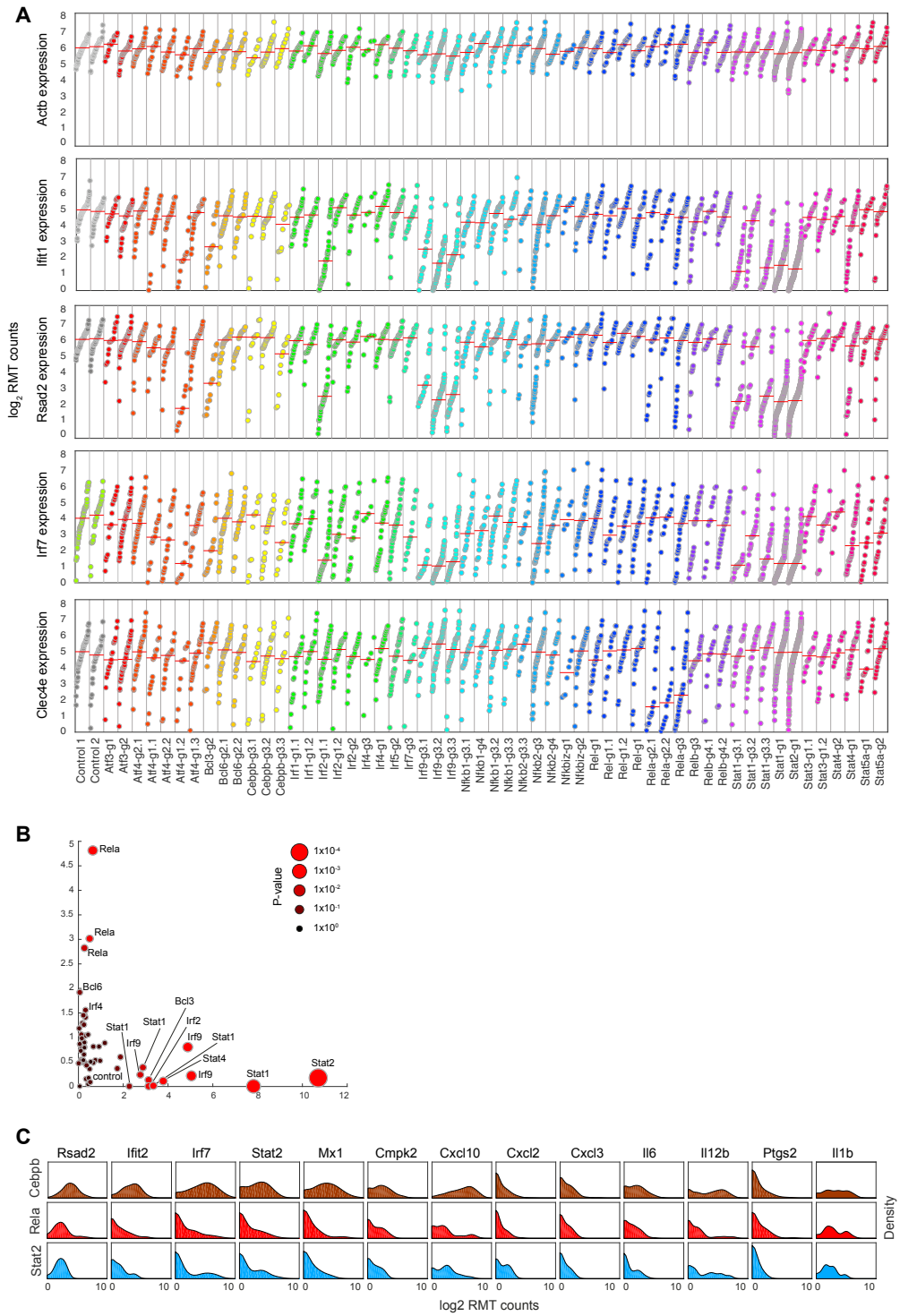


Figure S5. Perturbations of Developmental and Signaling-Dependent TFs, Related to Figure 5

(A) Expression of Actb (top), Ifit1 (middle), and Rsad2 (bottom) per single cell is shown for a pool of 57 gRNAs targeting signaling-dependent TFs in 2,600 UGI-confident myeloid cells stimulated with LPS for 4 hr. Cells are grouped by detected gRNA/UGI and ordered by expression level in each group. Red line, median expression per gRNA/UGI.

(B) Effect of each gRNA on the inflammatory (y axis) and the antiviral (x axis) pathways, indicated by their Z statistic. The signatures are defined by the expression 10 most induced inflammatory or antiviral genes. The p values (Mann-Whitney U-test) are indicated both by color and size of label.

(C) Histograms of 1,075 dendritic cells depict the expression of key inflammatory and antiviral response genes in the different KOs.

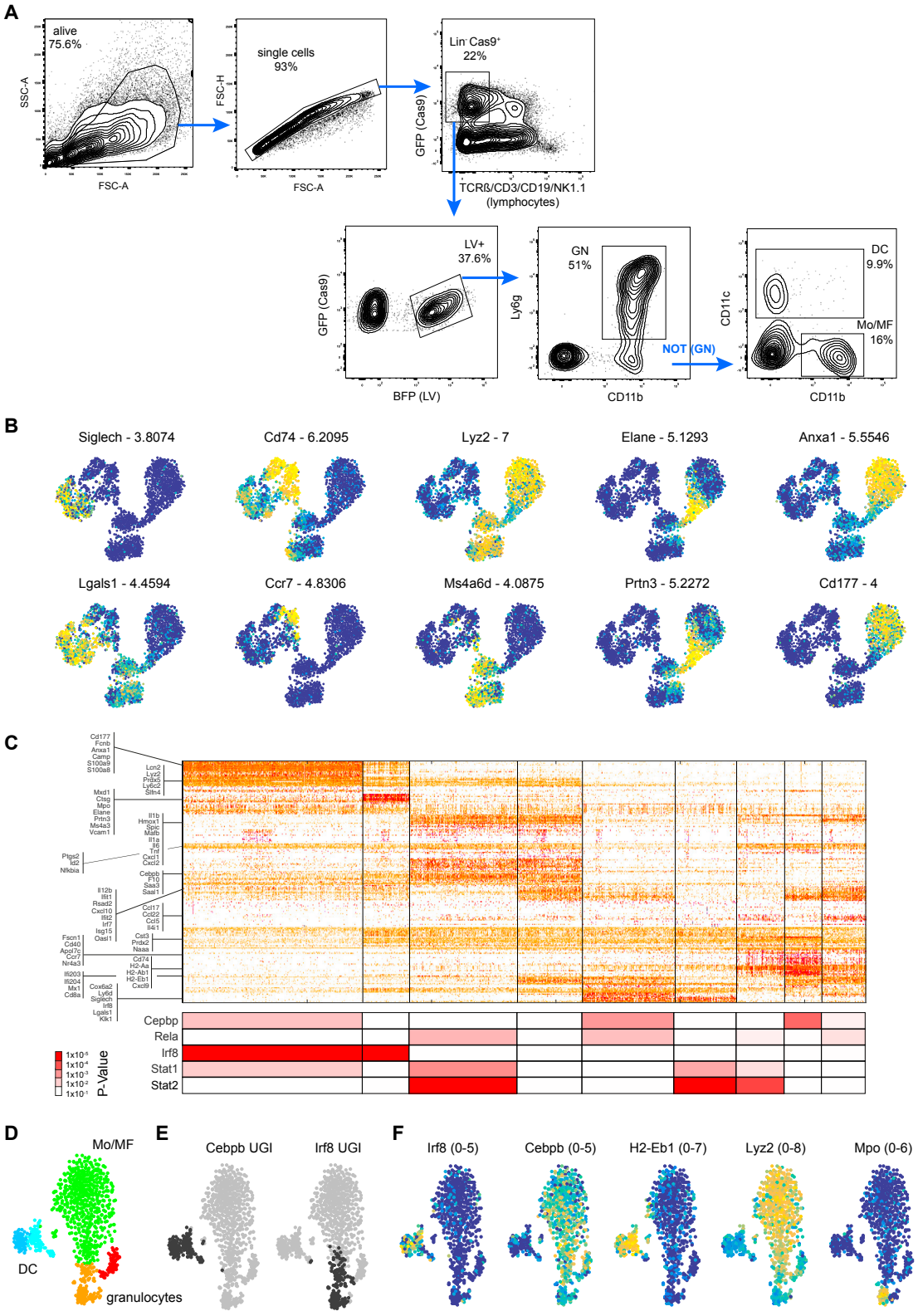


Figure S6. In Vivo CRISP-Seq Analysis Identifies Non-overlapping Targets of Stat1/2 in Different Myeloid Cells, Related to Figure 6

(A) FACS plot showing the gating strategy to identify the donor myeloid cells from the spleen. Cas9-GFP+ BFP+ (infection positive) cells we single-cell sorted after excluding recipient lymphocytes and cells with no markers (negative for both CD11b and CD11c).

(legend continued on next page)

(B) Projection of the expression levels of additional developmental genes onto the t-SNE plot (Figure 6C).
(C) Heatmap of 2768 myeloid cells from in vivo experiment, clusters correspond to Figure 6B. Bottom panel show UGI enrichment of different gRNAs within clusters. (D–F) Ex vivo validation of Irf8 and Cebpb role in myeloid development.
(D) tSNE analysis of 925 myeloid cells infected with Cebpb/Rela/Control gRNAs depicting their separation into monocytes/macrophages (green), DCs (blue) and granulocytes at different developmental stages (red and orange).
(E) Left, refined UGI labels of Cebpb overlapping the DC cluster; right, refined Irf8 UGI labels overlapping the granulocytes clusters.
(F) Projection of the expression levels of Irf8, Cebpb and three representative genes onto the tSNE.