Single-cell analysis of mixed-lineage states leading to a binary cell fate choice

Andre Olsson1, Meenakshi Venkatasubramanian2, Viren K. Chaudhri1, Bruce J. Aronow2, Nathan Salomonis3, Harinder Singh1§ & H. Leighton Grimes1,3§

Delineating hierarchical cellular states, including rare intermediates and the networks of regulatory genes that orchestrate cell-type specification, are continuing challenges for developmental biology. Single-cell RNA sequencing is greatly accelerating such research, given its power to provide comprehensive descriptions of genomic states and their presumptive regulators1-5. Haematopoietic multipotential progenitor cells, as well as bipotential intermediates, manifest mixed-lineage patterns of gene expression at a single-cell level6-7. Such mixed-lineage states may reflect the molecular priming of different developmental potentials by co-expressed alternative-lineage determinants, namely transcription factors. Although a bistable gene regulatory network has been proposed to regulate the specification of either neutrophils or macrophages7,8, the nature of the transition states manifested in vivo, and the underlying dynamics of the cell-fate determinants, have remained elusive. Here we use single-cell RNA sequencing coupled with a new analytic tool, iterative clustering and guide-gene selection, and clonogenic assays to delineate hierarchical genomic and regulatory states that culminate in neutrophil or macrophage specification in mice. We show that this analysis captured prevalent mixed-lineage intermediates that manifested concurrent expression of haematopoietic stem cell/progenitor and myeloid progenitor cell genes. It also revealed rare metastable intermediates that had collapsed the haematopoietic stem cell/progenitor gene expression programme, instead expressing low levels of the myeloid determinants, Irf8 and Gfi1 (refs 9–13). Genetic perturbations and chromatin immunoprecipitation followed by sequencing revealed Irf8 and Gfi1 as key components of counteracting myeloid-gene-regulatory networks. Combined loss of these two determinants ‘trapped’ the metastable intermediate. We propose that mixed-lineage states are obligatory during cell-fate specification, manifest differing frequencies because of their dynamic instability and are dictated by counteracting gene-regulatory networks.

To analyse discrete genomic states and the transitional intermediates that span myelopoiesis, we performed single-cell RNA sequencing (scRNA-seq) on stem/multipotent progenitor cells (LSK; lin-), common myeloid progenitor (CMP) cells, granulocyte monocyte progenitor (GMP) cells14, and LKCD34+ cells (lin- c-Kit+CD34+)15 that included granulocytic precursors. Analysis of the data using six independent computational approaches13,14,16,17 resulted in the varied delineation of cellular states and intermediates (Extended Data Figs 1–5 and Supplementary Information). Therefore, we developed a method, iterative clustering and guide-gene selection (ICGS, http://www.AltAnalyze.org), which utilizes the pairwise correlation of dynamically expressed genes and iterative clustering with pattern-specific guide genes to delineate coherent gene-expression patterns (Fig. 1a, Supplementary Information). Exclusion of cell-cycle genes improved predictions of developmental states (Extended Data Fig. 6a–c and Supplementary Information). ICGS resolved nine hierarchically ordered cellular states (Fig. 1b) that encompassed all those described above. GO-Elite pathway enrichment (http://www.AltAnalyze.org) assigned cellular identities to the following states: HSCP-1 (haematopoietic stem cell progenitor), HSCP-2, megakaryocytic, erythroid, Multi-Lin* (multi-lineage primed), MDP (monocyte-dendritic cell precursor), monocyte, granulocytic and myelocyte (myelocytes and metamyelocytes). Gene expression patterns of Csrlr (encoding CD115), Fli3 and Cx3cr1 suggested that both CMP and GMP cell populations contain macrophage and dendritic cell precursors (MDP; CX3CR1+CD115+CD135+)18, which we confirmed by flow cytometry (Extended Data Fig. 6d-f). Notably, the unbiased ICGS analysis inferred a developmental order that agreed with the experimentally determined haematopoietic sequence19 (Fig. 1b, bottom). Similarly, clustering of LKCD34+ cells recreated the entire developmental ordering, with granulocytic precursors at one end of the continuum (Extended Data Fig. 6b). Thus, ICGS generated a refined order of discrete myeloid cell states independent of, but consistent with, prior knowledge.

Next, we displayed the incidence and amplitude of expression of key genes within the predicted ICGS haematopoietic hierarchy (Fig. 1c). Notably, the Multi-Lin* population co-expressed the transcription factors Gata2, Meis1, PU.1 (encoded by Sip1) and C/EBPα—the latter two are key regulators of myelopoiesis20,21. They also manifested infrequent and variable-amplitude expression of megakaryocytic, erythroid, granulocytic and monocytic genes (Fig. 1c). Thus, during steady-state myelopoiesis a prevalent mixed-lineage state is encountered that expresses haematopoietic stem cell/progenitor (HSCP) and myeloid progenitor genes (Ctsg, Mpo, and Elane), while displaying molecular priming of erythroidic, megakaryocytic, granulocytic and monocytic potentials. Each ICGS-delineated cellular state is expected to have an underlying regulatory state characterized by distinct combinations of transcription factors. Clustering of Pearson-correlation coefficients for ICGS-delineated transcription factor–gene pairs (Fig. 1d, e and Extended Data Fig. 6g–j) revealed three distinct regulatory states within GMP cells (Fig. 1e). Two were demarcated by expression of transcription factors involved in granulocyte (for example, Cebpe and Gfi1) or monocyte (for example, Irf8 and Klf4) specification22. The third state was indicated by the expression of HSCP transcription factors Gata2 and Meis1, along with the signal-induced transcription factors Jun, Fos and Egr1. The combined analysis of myelopoiesis suggests a multipotential ground state associated with a large set of transcription factors, including Gata2, Meis1, PU.1 and C/EBPα, that is acted on by signal-induced transcription factors such as Fos and Egr1 to generate myeloid progenitors that then undergo a strong bifurcation into demarcated monocytic and granulocytic genomic states.
To infer regulatory interactions among transcription factors that are reflective of granulocytic and monocytic specification, their pairwise expression was correlated with cellular genomic states (Fig. 1f–i and Extended Data Fig. 7a). This both confirms established regulatory relationships (for example, Irf8–Klf4 (ref. 23)) and suggested new regulatory interactions (Irf8–Zeb2 and Gfi1–Per3). Notably, Gfi1 and Irf8, which are required for normal granulopoiesis and monopoiesis, respectively (9, 11, 12, 24, 25), displayed strong partitioning within granulocyte–specified versus monocyte–specified cells (Fig. 1f). Given their reciprocal expression, we analysed the consequences of Gfi1 or Irf8 loss on genes strongly correlated with their expression within wild-type GMP cells (Fig. 2a). Loss of either transcription factor reduced the heterogeneity of genomic states manifested at the single-cell level (Fig. 2a). Furthermore, loss of Irf8 or Gfi1 reciprocally perturbed the expression of transcription factors that were associated with the mononuclear (Klf4, Zeb2 and Irf5) and granulocytic (Per3 and Ets1) regulatory states, respectively (Fig. 2a and Extended Data Fig. 7b, c). To explore the underlying molecular mechanisms, we performed chromatin immunoprecipitation with sequencing (ChIP–seq) analyses in GMP cell populations (Fig. 2b). Notably, Irf8 peaks were enriched for EICE motifs, which are co-bound by PU.1 (ref. 26). Intersection of the Gfi1 and Irf8 peaks revealed the presence of shared regions that were deemed accessible in GMP cells based on assay for transposable-accessible chromatin with sequencing (ATAC–seq) data (27) (Fig. 2b, c). Gfi1 recruits Lsd1 (ref. 28), a histone demethylase acting on H3K4me2. The shared genomic regions displayed increased H3K4me2 levels upon Gfi1 loss (Fig. 2c) or Lsd1 inhibition, correlating with enhanced monocytic potential (Extended Data Fig. 7d, e). Genes located near the shared genomic regions were associated with monocyte–dendritic–precursor cells or abnormal mononuclear cell morphology, and were reciprocally dysregulated in Irf8−/− or Gfi1−/− GMP cells (Extended Data Fig. 7f). Thus Gfi1 antagonizes the specification of the monocyte–dendritic programme in GMP cells by repressing enhancers activated by PU.1–Irf8. Pertinently, similar binding patterns for Gfi1 and Irf8 were seen on the Irf8, Klf4 and Zeb2 genes (Extended Data Fig. 8a). It is therefore likely that Gfi1 represses the Irf8, Klf4 and Zeb2 genes (Extended Data Fig. 8a). It is therefore likely that Gfi1 represses the Irf8, Klf4 and Zeb2 genes by interrupting positive regulation by PU.1 and Irf8, a mechanism similar to its antagonism of PU.1 on the Spi1 gene (29). To test regulatory interactions further, we varied levels of Gfi1 within GMP cells using an inducible Gfi1 allele (Extended Data Fig. 8b). Gfi1 induction in GMP cells increased granulocytic potential, while diminishing monocytic potential (Extended Data Fig. 8c, d). In CD115+ GMP cells, inducing Gfi1 repressed monocytic gene expression (including Irf8), and induced neutrophil gene expression in a dose-dependent manner (Fig. 2d). In agreement with regulatory-state (Fig. 1e) and loss–of–function (Fig. 2a) analyses, key transcription factors were reciprocally altered by increased expression of Gfi1; Klf4, Zeb2 and Irf5 were repressed whereas Ets1 and
ATAC-seq analysis in GMP cells (GSE60103) and H3K4me2 ChIP-seq overlapping ChIP-seq peaks. Genotypes and cell clusters of Irf8-high (blue) and Gfi1-high (green) or neither (purple) are indicated, along with genes shared with ICGS.

Per3 were induced. The perturbation and ChIP-seq data were used to assemble a gene regulatory network underlying myeloid cell fate specification (Fig. 2e).

Given that Irf8 and Gfi1 function as antagonistic determinants, we investigated how their dynamic expression shapes the genomic state and developmental potential of a GMP cell. Analysis of GMP cell populations using an Irf8-GFP reporter (Fig. 3a) and CD115 revealed two major Irf8-expressing GMP subpopulations (IG1 and IG3) and a minor intermediate (IG2) (Fig. 3b). Expression of CD115 protein and transcripts was strongly correlated with Irf8 expression (Fig. 3b, c). Conversely Gfi1 transcripts were anti-correlated with those of Irf8 and Csf1r (Fig. 3c).

Figure 3 | Detection of rare transition state poised to undergo myeloid cell-fate determination. a, Schematic representation of the Irf8-GFP (IG) reporter allele. b, Flow cytometric analysis of CD115 and GFP expression in GMP cell populations. Gating strategy for IG1, IG2 and IG3 cells is indicated. c, TaqMan analysis of indicated transcripts in IG cells. d, CFU assays with indicated IG cells. Percentage distribution of colonies containing granulocytes (G), macrophages (M), or both (GM) is displayed on the y axis (n = 3). e, Schematic representation Gfi1-GFP (GG) reporter allele. f, Flow cytometry analysis of CD115 and GFP expression in GMP cell populations. Gating strategy for GG1, GG2 and GG3 cells is indicated. g, TaqMan of indicated transcripts in GG cells as in c. h, CFU assays with indicated GG cells as in d (n = 3). Representative plots from one of three independent experiments with each reporter are shown. c and g display two technical replicates.

Select genes shared with ICGS are indicated. b, ChIP-seq analysis of Gfi1 and Ir8 in GMP cells. Statistically enriched Gfi1 (P = 6.63 × 10^-6) and Ets-Irf composite element (EICE) motifs (P = 1.08 × 10^-6) are displayed. Venn diagram illustrates Gfi1, Irf8 and -correlated genes.
which expressed low levels of Irf8 and Gfi1 (Fig. 3b, c), appeared to represent cells poised to undergo specification as they gave rise to equal proportions of monocytic and granulocytic colonies. Next, we examined GMP cell populations from Gfi1−/− reporter mice (Fig. 3e), using CD115 as a surrogate for Irf8. Flow cytometry analysis revealed two major Gfi1-expressing GMP cell population intermediates (GG2, GG3) and one minor population (GG1) (Fig. 3f). GG2 cells expressed the highest levels of Gfi1 and represented specified granulocytic progenitors, while GG3 cells, which expressed the highest levels of Irf8, were oppositely specified as monocytic progenitors (Fig. 3f–h). The GG1 cell population was rare and expressed intermediate levels of both transcription factors (Fig. 3g). The Gfi1−/− reporter expresses stable GFP, which can overestimate Gfi1 expression, probably accounting for higher GFP expression in GG3 cells (Fig. 3f) despite the very low level of Gfi1 transcripts (Fig. 3g). Pertinently, the GG1 population was enriched for bipotential cells (CFU-GM) as well as those undergoing lineage specification (CFU-G and CFU-M; Fig. 3h, Extended Data Fig. 9a). Thus, by using reporters for reciprocally expressed transcription factors we were able to distinguish between bipotential cells, their lineage-committed progeny and rare intermediates poised to undergo binary cell-fate choice.

We next performed scRNA-seq on GG1 and IG2 cell populations (Fig. 4a). Four clusters of cells could be delineated within the GG1 population (Extended Data Fig. 9b–d and Supplementary Information). One group was enriched for HSCP genes, including Gata1, Gata2, Egr1, Fosb and Jun (Fig. 4a). These cells were not contaminants as they expressed CD16/32 and CD34 (Extended Data Fig. 9c, e, f) and corresponded to the bipotential cells (CFU-GM) within the GG1 population (which we explore shortly). The second cluster downregulated HSC genes with the exception of Gata1 and expressed Gfi1, Il5ra, Prg2 and Epx. These were eosinophilic progenitor cells based on data from CFU assays, cytopsins and flow cytometry (Extended Data Fig. 10a–h). The remaining two groups of cells expressed low levels of Gfi1 and Irf8 along with the myeloid genes Etv6, Mpo, Elane and Hoxa1, although a subset of these expressed higher levels of Irf8 along with Cybb and Ly6a (Fig. 4a). The genomic states of these latter groups suggest they represent mixed-lineage intermediates that are poised for binary cell fate choice. To test this, we analysed the IG2 subpopulation (Fig. 4a and Extended Data Fig. 10i), which lacks bipotential progenitors (CFU-GM) and are highly enriched for cells undergoing lineage specification, resulting in CFU-G and CFU-M. HSC gene expression waned in the IG2 subpopulation and Gfi1 and Irf8 were co-expressed. In contrast, the GG1 and IG1 subpopulations, which both contain the bipotential progenitors (CFU-GM), were enriched for cells expressing key HSC genes (Extended Data Fig. 10j, k), linking the HSC gene expression module with CFU-GM developmental output. We were thus able to assign genomic states at a single-cell level to the well-known myeloid intermediates CFU-M, CFU-G and CFU-GM.

We note that the induction of Irf8 at low levels is associated with loss of the multipotential programme but is not accompanied by the specification for monocytic differentiation. Higher-amplitude Irf8 expression appears necessary for the latter. Similarly, an intermediate level of Gfi1 expression is associated with loss of the multipotential programme, but a further increase in its expression coincides with neutrophil-differentiation specification. Thus, the haematopoietic intermediates, which express a multipotential programme (HSCP1 and HSCP2), span the LSK, CMP and GMP flow cytometric gates. Those rare cells within the GMP gate that are undergoing monocytic versus neutrophil specification have collapsed the multipotential programme, resulting in CFU-G and CFU-M. HSCP gene populations based on ICGS-delineated genes (Fig. 1b), with indicated myeloid cellular states (right) (see global analysis in Extended Data Fig. 10l), a finding underscored by the quantitative indexing of monocytic and granulocytic signature genes (Fig. 4c). Notably, the Irf8−/− Gfi1−/− GMP cells were more tightly correlated as a group than the IG2 subpopulation. Accordingly, we propose that IG2 cells manifest dynamic instability because of the counteracting functions of Irf8 and Gfi1, and that this metastable state is trapped by the elimination of both developmental determinants.

We were able to identify both prevalent and rare mixed-lineage genomic states that are encountered during myeloipisis (Fig. 5). Multi-Lin* intermediates expressing HSCP genes induce both robust myeloid progenitor cell gene expression and transcripts for alternative lineage genes. Notably, myeloid priming occurs in cells which express the transcription factors PU.1 and/or C/EBPα (ref. 31). A remarkable feature of this mixed-lineage state is its prevalence and apparent stability, despite the mixing of alternative lineage determinants. Expression of the HSCP gene module in GMP cells is associated with CFU-GM potential. In some rare cell subpopulations, HSCP gene expression wanes with the simultaneous acquisition of CFU-G and CFU-M potentials. Based upon the frequency at which it is observed, this state is inferred to be metastable, but could in fact be trapped due to the elimination of counteracting determinants. The concept of trapping rare developmental intermediates by genetic perturbation is analogous to the trapping of unstable transition states in chemical reactions using physicochemical strategies32. We propose that the simultaneous
expression of counteracting regulatory gene network components manifests as dynamic instability. This may generate oscillations in the regulatory states of multi- or bi-potential intermediates, resulting in bursts of alternative lineage gene expression. This oscillatory behaviour may be a reflection of the partial assembly of counteracting regulatory states or a lack of regulatory-state stabilization.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 December 2015; accepted 9 August 2016.
Published online 31 August 2016.

1. Grün, D. et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525, 251–255 (2015).
2. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell 153, 1653–1677 (2013).
3. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).
4. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature 524, 288–293 (2015).
5. Yan, L. et al. Single-cell mRNA-Seq profiling of human preimplantation embryos reveals hidden subpopulations of cells. Nature 533, 331–336 (2016).
6. Morrow, K. R. et al. Neuron specification is driven by cell type-specific long-range CRX transcription factor binding. Nat. Neurosci. 20, 155–160 (2017).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. C. Morse for supplying Irb8-GFP mice. We acknowledge the assistance of the Cincinnati Children’s Hospital Medical Center (CHCCH) Research Flow Cytometry Core (supported in part by NIH grants AR-47363, NIH DK78392 and NIH DK90971) and DNA Sequencing Center (CCHMC) Research Flow Cytometry Core (supported in part by NIH grants AR-47363, NIH DK78392 and NIH DK90971) and DNA Sequencing.

Author Contributions A.O., H.S. and H.L.G. designed and interpreted experiments. A.O. performed the experiments. N.S. conceived and developed the software with considerable input from B.J.A., H.S. and H.L.G. Bioinformatics data were analysed by M.V., V.K.C., B.J.A., N.S., H.S. and H.L.G. The paper was written by A.O., N.S. and H.L.G. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.L.G. (Lee.grimes@cchmc.org), N.S. (nathan.salomonis@cchmc.org) or H.S. (harinder.singh@cchmc.org).

Reviewer Information Nature thanks L. Zon and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mice. Gfi1 expression was measured in the Gfi1-IRES-Venus (G3GV) knock-in mice, we first modified the pBS31 vector49 to contain 7 tetracycline-responsive elements with revised sequence and spacing, termed G3 (ref. 36) (pBS31-G3). To generate the Gfi1-IRES-Venus sequence, the internal ribosomal entry site (IRES) from the encephalomyocarditis virus was cloned 5′ of a rapidly maturing YFP variant (Venus)5. The mouse Gfi1 open reading frame was then cloned 5′ of IRES–Venus. Finally, the Gfi1-IRES-Venus fragment was cloned into the pBS31 G3 plasmid (pBS31-G3V). Inducible Gfi1–IRES-Venus knock-in mice were generated by the electroporation of KH2 embryonic stem (ES) cells with both the pBS31-G3GV vector and a FLP recombinase expression vector (pCAGS-FLPe-Puro). FLPe recombinase is expected to recombine the pBS31 plasmid into the Col5a1 locus of KH2 mouse ES cells, and repair a defective hygromycin resistance gene. KH2 cells were maintained on DR4 feeders (Mirimix) as previously described. After electroporation, the KH2 cells were selected in hygromycin and the first eight hygromycin-resistant clones were expanded. Since KH2 cells also contain a ROSA allele encoding rtTA-M2, a split of each recombinant clone was treated with doxycycline in vitro and then analysed by immunoblot using anti-Gfi1 (AF3450, R&D Systems) or anti-GFP (632593, Clontech) antibodies. Four independent Gfi1–IRES-Venus ES cell clones were injected into 8-cell embryos, resulting in an average of 50% chimaerism. Progeny were backcrossed to C57Bl/6 ROSA26rtTA122 mice (JAX stock number: 006965).

Flow cytometry and cell sorting. Mice were euthanized with carbon dioxide and by cervical dislocation. Femurs, tibiae and the iliac crest were harvested immediately after euthanasia and put into cold PBS with 2% FBS. Bones were crushed with a pestle and mortar, filtered and washed in cold PBS with 2% FBS, then enriched using CD117 Microbeads on a Automacs Pro separator (Miltenyi). CD117+ cells were stained with lineage: CD3-biotin (clone 145-2C11, BioLegend), CD4-biotin (clone RM4-5, eBioscience), CD117–APC (clone 2B8, Becton, Dickinson and Company), CD11b-biotin (clone M1/70, Becton, Dickinson and Company), CD19-biotin (clone 6D5, Biotest), Gr1-biotin (clone RB6-8C5, Bioclegen), Ter119-biotin (clone Ter-119, Biolegend) and CD45R-biotin (clone RA3-68B2, Biotest). To isolate LSK, CMP and GMP cells, lineage-stained cells were stained with streptavidin APC-Cy7 (Becton, Dickinson and Company), CD16/32-PerCP-eFl70 (clone 93, e Bioscience), CD117–APC (clone 2B8, Becton, Dickinson and Company), Sca–1-PerCy5 (clone D7, Becton, Dickinson and Company) and CD34–BV421 (clone RAM34, Becton, Dickinson and Company). GMP and CMP gates were set using CD34-PE. To isolate Irf8–GFP and Gfi1–GFP GMP cell subpopulations, the LSK, CMP, GMP panel was supplemented with CD115–BV605 (clone TR15-12F1, 2.2, BioLegend). MDP were analysed by adding CD115–BV605 (clone TR15–12F1 2.2, BioLegend) and CD135–PE (A2F10.1, eBioscience) to the LSK, CMP, GMP panel. CD115–BV605 and CD135–PE were co-stained with CD117–APC (clone 2B8, Becton, Dickinson and Company), Sca–1–APC (clone A20, Becton, Dickinson and Company), Gr1–APC (clone RB6-8C5, Becton, Dickinson and Company) and CD34–BV421 (clone RAM34, Becton, Dickinson and Company). Eosinophil differentiation was assayed by staining washed CFU cells with lineage: CD3-biotin (clone 145-2C11, BioLegend), CD4-biotin (clone RM4-5, eBioscience), CD117–APC (clone 2B8, Becton, Dickinson and Company), Sca–1–APC (clone A20, Becton, Dickinson and Company), Gr1–APC (clone RB6-8C5, Becton, Dickinson and Company) and CD34–BV421 (clone RAM34, Becton, Dickinson and Company). The resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantification of the total chromatin yield. An aliquot of chromatin (30 μg) was pre-cleared with protein A (for Gfi1) or protein G (for Irf8) agarose beads (Life Technologies). Genomic DNA regions of interest were isolated using μg of antibody against Gfi1 (ref. 50) or Irf8 (sc-6058, Santa Cruz). Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65 °C, and ChIP DNA was purified by phenol–chloroform extraction and ethanol precipitation. Illumina sequencing libraries were prepared from the ChIP and input DNAs by the standard consecutive enzymatic steps of end-polishing, da-addition, and adapter ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and 50 nucleotide single-end reads were sequenced on Illumina HiSeq 2500 (Gfi1) or NexSeq 500 (Irf8).

Alternatively, lineage-negative bone marrow cells were lysed in cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP40). Chromatin from nuclei, lysed in Nucleus lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS), was diluted in immunoprecipitation buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS) and sheared using a Bioruptor (Diagenode). Chromatin immunoprecipitation was performed with anti-H3K4me2 (Ab2-035-050, Diagenode), then isolated with Protein A/G Magnetic Beads (Pierce). After unshearing, libraries were prepared (Illumina Inc.) and sequenced on Genome Analyzer II (Illumina Inc.). CEBP-Pro ChIP–seq fastq were downloaded from the GEO, accession number GSE43007. ATAC-seq in GMP cells was downloaded from the GEO, accession number GSE59992.

RNA-seq and ChIP-seq data processing. RNA-seq and ChIP–seq reads were aligned to the mouse genome using Bowtie2 (ref. 39). Single-cell and bulk sorted RNA-seq were analysed using RSEM to estimate TPM for all genes. Genomic aligned sequences were visualized with IGV41. Outlier cells were excluded based on cell library read-depth and overall percentage sequence alignment (Supplementary Information). Evaluation with different RNA-seq alignment and expression quantification algorithms was used to verify the overall accuracy of our results (Supplementary Information). The ICGS automated workflow has been implemented in the easy-to-use, open-source transcriptome analysis toolkit AltAnalyze53. AltAnalyze facilitates both pairwise correlation of dynamically expressed genes and iterative clustering with pattern-specific guide genes to delineate highly coherent gene expression patterns (Supplementary Information). Cells displaying these expression patterns are ordered in a second step using the HOPACH clustering algorithm. An evaluation

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
of the ability of ICGS to identify known and novel populations was performed for three previously described single-cell RNA-Seq datasets (Extended Data Fig. 5a–d and Supplementary Information). In order to identify subpopulations of 3 cells, present at 10%, a minimum of 30 cells was required. For primary discovery analyses n = 90 was required. Differentially expressed genes were identified using AltAnalyze with a discovery-rate-adjusted empirical Bayes–moderated t-test P < 0.05. Hierarchical clustering and heat map visualization was produced using AltAnalyze and R. All AltAnalyze heat maps are scaled to a contrast factor of 2.5 and median-centred normalized. Details on the ICGS analysis pipeline and GGI1/G2 associated population identification are detailed in Supplementary Information.

ChIP–seq peaks were called using Homer software46 using options “-style factor, -size 500 minDist 1000”. ChIP–seq heat plot was generated in R using the heatmap utility from the bioconductor package ‘mage’47,48. For visualization, RNA and ChIP–seq were processed and aligned to mm10 using Bioawardrobe46 (which requires mm10). Tracks were displayed using the UCSC Genome Browser47.

The data sets are reposited in GEO as a SuperSeries under accession number GSE70245. ICGS-ordered genes and cell expression profiles can be queried and visualized for selected gene and gene-sets of interest at http://www.altanalyzr.org/hematopoietic.html.

RT–PCR. High-capacity cDNA reverse transcription kit (Applied Biosystems) was used to generate cDNA. Quantitative PCR was performed using Taqman universal master mix (Applied Biosystems) and the following gene expression assays (all Applied Biosystems): Csf1r (Mm00432689_m1), Egr1 (Mm00656724_m1), Ela2 (Mm00469310_m1), Eps (Mm00487425_m1), Gata1 (Mm01352636_m1), Gata2 (Mm00492301_m1), Gapdh (Mm99999915_g1), Gfi1 (Mm00515855_m1), Il6ra (Mm00432848_m1), Irf8 (Mm00492567_m1), JunB (Mm00423546_s1), Meis1 (Mm00487664_m1), Pbx1 (Mm004207617_m1) and Prog2 (Mm0036479_m1).

Methylcellulose assays and liquid culture. For methyl cellulose assays, 750 μg ml−1 doxycycline (SIGMA D9891) was added to either liquid or methyl cellulose. SCF (25 ng ml−1). To test Lsd1 dependency, CD117+ cells were treated with 0.5 μM LSD1-C76 (Xcessbio), for either 24 h in liquid culture or in methyl cellulose for 7 days. For IL-5 driven eosinophil colony assays, Gfi1-GFPtrans, CD115+ GMP cells were plated in M3231 (StemCell Technologies) supplemented with IL-3 (20 ng ml−1), IL-5 (50 ng ml−1), GM–CSF (10 ng ml−1), SCF (25 ng ml−1) or IL-5 (50 ng ml−1) and SCF (25 ng ml−1) only. Cytospins were prepared by washing the cells twice in PBS. 10,000 cells were loaded onto VistaVision HistoBond (VWR) slides using a Cytospin 4 Cytocentrifuge (Thermo Fisher Scientific). Slides were dried overnight and then stained with Camco Stain Pak (Cambridge Diagnostic Products, Inc.).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 1 | Experimental design, optimization, quality control, and validation of scRNA-seq data. a, Schematic illustrating haematopoietic cell intermediates that constitute the myeloid developmental pathway in the bone marrow. Select cell-surface markers expressed by the various intermediates are indicated (top). The mouse bone-marrow populations used for scRNA-seq are colour coded and indicated at bottom of schematic. b, Flow cytometry strategy for isolation of LSK, CMP and GMP haematopoietic cell subsets. c, Sorting strategy for lineage-negative CD117+CD34+ (LK CD34+) myeloid progenitor-precursor populations. d, Optimization of size of scRNA-seq library fragments by varying the amount of input cDNA (amount shown on top of each electropherogram). e, An scRNA-seq library fragment distribution before and after optimization reveals an increase in reads that can be mapped to the genome (reads mapping to a single genomic/transcriptomic location using BowTie2/TopHat2; 61.8% before and 87.3% after optimization). f, scRNA-seq summary statistics. g, Distribution of number of aligned reads for scRNA-seq libraries. h, Scatter plot of the fraction of aligned reads (y axis: RSEM-transcriptome-aligned reads relative to the total number of sequenced reads) versus the total number of genes with a TPM > 1 (x axis) for each cell. RNA-seq libraries in red were considered outliers and eliminated from downstream analyses. Of the 96 LSK, 96 CMP, 136 GMP and 66 LKCD34+ cell libraries that were sequenced, 3 LSK, 2 CMP, 4 GMP and 3 LKCD34+ samples failed the quality control analysis. i, Histogram showing the inverse-cumulative count of genes for each sequenced cell greater than the TPM cut-off in each bin (200 bins, red samples represent outliers from Extended Data Fig. 1h). j, Correlation between bulk RNA-seq (TRUseq Stranded mRNA HT kit) for each sorted population compared to average single-cell-gene-expression values from the same sort (for example, LSK, CMP or GMP cells).
Extended Data Figure 2 | Analysis of scRNA-seq data with Monocle, SCUBA, RaceID and principle component analysis. a, Pseudotemporal ordering of all 382 haematopoietic progenitors along the 9 identified Monocle cell states. These states were identified from the original four annotated flow cytometric gates, based on 5,110 Monocle-identified genes ($P < 0.01$). b, Hierarchical clustering of the 1,000 most significant genes for the 9 Monocle states (correlation and ward hierarchical clustering for genes). c, SCUBA pseudotemporal ordered cell states. d, Hierarchically clustering of the 1,000 most variable SCUBA genes. e, RaceID identified $k$-means cell clusters. f, RaceID t-SNE visualization of these cells and $k$-means cell states. g, Heat map of the t-SNE based cell-state ordering and the top 150 most significant RaceID genes (identified using RaceID R code, clustdiffgenes function p-value) associated with each of the 14 $k$-means clusters. Gene-set predictions are assigned on the left of the heat map following ImmGen (https://www.immgen.org/) gene-set enrichment analysis in AltAnalyze. h, Hierarchical clustering of 584 principle component analysis (PCA)-identified genes using the workflow outlined in ref. 48.
Extended Data Figure 3 | See next page for caption.
Extended Data Figure 3 | Analysis of scRNA-seq data with Seurat, scLVM and ICGS. 

a, Seurat-significant gene-weighted PCA, coloured by the sorted cell population annotations. 
b, Seurat t-SNE displayed output for the 9 predicted cell states. 
c, Seurat hierarchical relationships between the predicted cell states, based on 161 differentially expressed genes. 
d, Expanded 766 significant genes displayed along the Seurat ordered cell states. 
e, Uncorrected PCA using the top 2,361 scLVM variance (https://www.synapse.org/#!Synapse:syn4975059) genes (left). PCA of the scLVM-corrected normalized expression matrix (right). 
f, Hierarchically clustered heat map using the top 2,361 genes from the corrected scLVM-normalized expression matrix. 
g, ICGS analysis and integration of cell-type prediction analyses in AltAnalyze. ICGS-produced expression heat map for the haematopoietic progenitor scRNA-seq data. On the right of the ICGS heat map are the default predicted GO-Elite BioMarker enrichment predictions (60 top-genes for each of the 300 cells/tissue microarray data sets evaluated) for each HOPACH cluster. On the left is a similar set of gene-set enrichments derived in AltAnalyze for all mouse ImmGen profiles (enrichment analysis and visualization available through the AltAnalyze heat map viewer). Fisher-exact enrichment P values are displayed with each term, along with the associated HOPACH cluster number. Terms used to derive the final predicted cell-types are manually highlighted in red.
Extended Data Figure 4 | Comparing ICGS with Monocle, SCUBA, RaceID, PCA, Seurat and scLVM for the analysis of scRNA-seq data.

a, Table comparing ICGS-derived cell population gene-set results using different ICGS parameters. These parameters include minimum cells differing between the highest and lowest values for a gene for an indicated minimum fold difference and minimum Pearson correlation threshold. The results indicated under ICGS Steps refer to the gene outputs from each differing between the highest and lowest values for a gene for an indicated different ICGS parameters. These parameters include minimum cells a
RaceID, PCA, Seurat and scLVM for the analysis of scRNA-seq data.

b, PCA visualization of the first two principal components of all expressed genes (ICGS step 1), following Z-score normalization of all TPM values. Cells are coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different coloured according to the flow cytometric gate (top left), according to their ICGS clusters (top right) or for cell populations indicated by the different
a) Dataset 1) Pre-implantation embryos and hESCs (human)
   - single cells: n=124
   - depth: 39.3 million

   Zygote → 2-cell → 4-cell → 8-cell → Morula → Late-blastocyst

   Dollyt

   p0 HESCs → p10 HESCs

b) Dataset 2) Myoblast differentiation (human)
   - single cells: n=271
   - depth: 4 million

   Myoblast → Intermediate myocytes → Mature myotubes

   Putative contaminating interstitial mesenchymal cells

Dataset 3) Intestinal Organoid (mouse)
   - single cells: n=288
   - depth: 300 thousand

   Intestinal Stem Cells → Transit amplifying cells → Secretory Cells

   Enteroctye precursor → Enteroctyes

Extended Data Figure 5 | Application of ICGS to diverse scRNA-seq data sets. a, Schematic overview of published embryonic, myoblast and intestinal organoid scRNA-seq data sets. b, HOPACH-generated clusters derived using ICGS for human scRNA-seq data for pre-implantation embryos and ES cells. c, HOPACH-generated clusters using ICGS for human myoblast differentiation, without the inclusion of cell-cycle-associated genes. d, HOPACH-generated clusters using ICGS for mouse intestinal organoids for the discovery of rare cell populations. Novel rare population markers reported by the original study authors are highlighted in red. To the left of the heat maps are predicted cell-types and tissues using AltAnalyze enrichment analysis (GO-Elite algorithm), using genes from the embedded MarkerFinder database. Enriched terms are ordered based on significance from the bottom to top in each indicated HOPACH cluster. Genes to the right of the heat map are guide genes delineated by ICGS. The colour bars above the heat maps indicate either HOPACH clusters or input cell identities.
Extended Data Figure 6 | See next page for caption.
Extended Data Figure 6 | Cell cycle, monocyte–dendritic precursor, and transcription factor–gene correlation analyses. a–c, Activation of a mitotic gene expression programme in developmentally distinct cell populations. a, Heat map of single-cell ICGS-gene-expression clusters generated (in AltAnalyze using the HOPACH algorithm) with the allowed inclusion of cell-cycle regulators as guide genes. Each column represents a single cell library; each row represents a different gene. ICGS-identified guide genes are indicated to the right of each plot. ICGS-identified HOPACH clusters are indicated at the top. b, ICGS from a reordered by the gates used for flow cytometric isolation (indicated at the top). Cell types (to the left) were predicted using GO-Elite (AltAnalyze) and ToppGene enrichment analysis, in addition to prior literature knowledge. c, PCA visualization of the first two principal components of all expressed genes (ICGS step 1), following Z-score normalization of all TPM values. Cells shaded to signify the mean expression of cell-cycle-associated genes (GO: 0022402). d–f, MDP and nascent dendritic cells within myeloid progenitor cell gates. d, Column plots displaying the incidence and amplitude of expression of select genes (in Fig. 1b ICGS-clustered order is shown as ‘Clusters’ at the top). The origin (flow-cytometric-gate) of each cell is indicated. Expression of Flt3, Csf1r and Cx3cr1 identifies MDP, while expression of Batf3 and Ili205 suggests dendritic cell differentiation. e, Flow cytometric analysis of lineage-negative Cx3CR1–GFP+ mouse bone marrow cells confirms the presence of phenotypic CD135+ (Flt3), CD115+ (Csf1r) MDP in CMP and GMP gates. f, Bar graph representing the relative abundance of MDP cells within each gate. Mean ± s.e.m. for three biological replicates. g–j, Transcription factor-to-gene correlation analysis. g, ICGS clustering of LSK cells (n = 93) with cell-cycle genes excluded. h, ICGS clustering of CMP cells (n = 94) with cell-cycle genes excluded. ICGS-selected guide genes are displayed on the right of each heat map. i, Heat map displays clustering of Pearson correlation coefficients among genes and transcription factors using HOPACH, with corresponding ICGS clusters from LSK in g. j, Heat map displays clustering of Pearson correlation coefficients among genes and transcription factors using HOPACH, with corresponding ICGS clusters from CMP cell population in h. Columns represent genes and rows transcription factors that are captured by ICGS analysis of CMP cells.
Extended Data Figure 7 | See next page for caption.
Extended Data Figure 7 | Transcription factor to transcription factor correlations and transcription factor loss-of-function analyses.

a, Scatter plots reveal the single-cell structure underlying correlations between transcription factors. Scatter plots generated in R (using the pairs function) show TPM of select transcription factor pairs in individual GMP cell populations (colours corresponding to ICGS groups in Fig. 1d, top). Expression is given as TPM. Pearson correlation coefficients are indicated opposite to each plot. b, Plots displaying the incidence and amplitude of expression of select genes in Fig. 2a. Expression clusters of Irf8-high (blue) and Gfi1-high (green) or neither (Multi-Lin*; purple) are shown. Significant changes in the expression of key genes between Irf8−/− versus Irf8-high wild-type GMP, or Gfi1−/− versus Gfi1-high wild-type GMP cells are noted (*P < 0.05, **P < 0.01, ***P < 0.001; Benjamini–Hochberg adjusted). Note that Irf8−/− and Gfi1−/− GMP cells continue to express non-productive transcripts emanating from the mutant Gfi1 and Irf8 alleles. c, Gfi1−/− GMP cells show a significant increase in cell-cycle-related gene expression compared to wild-type or Irf8−/− GMP cells. HOPACH clustering of Gfi1−/− and Irf8−/− GMP cells using haematopoietic guide genes from Fig. 2a. All cells were first clustered by HOPACH and then grouped according to sorting gates. In agreement with our previous report that Gfi1 controls two genetically separable programmes; granulopoiesis and Hox-based myeloid progenitor proliferation25, Gfi1−/− GMP cells demonstrate significantly increased HSC and cell-cycle-associated gene expression. Cell-cycle-associated genes were enriched (Z > 1.96) in Gfi1−/− and depleted (Z < −1.96) in Irf8−/− GMP cells. d, e, Lsd1 inhibition results in monocytic colony formation and increased Irf8 expression. d, CFU assays performed with CD117+ bone marrow cells with and without Lsd1-inhibitor (GSK C-76) treatment. The y axis displays percentage distribution of colony types. Mean CFU number of three technical replicates shown. e, TaqMan analysis of Irf8 expression in CD117+ bone marrow cells with and without treatment with C-76 (16 h). Mean of 3 technical replicates with similar results from 3 biological replicates. Representative plot from one of the 3 independent experiments performed is displayed for both d and e. f, Heat map showing the expression of a subset of genes (214) associated with Gfi1- and Irf8-shared ChIP–seq peaks. All displayed genes are significantly differentially expressed (P < 0.05, Benjamini–Hochberg adjusted) among at least one of the four comparisons (Irf8−/− versus wild type; Irf8−/− versus Irf8-high wild-type; Gfi1−/− versus wild-type; Gfi1−/− versus Gfi1-high wild-type). Marked genes (−) are associated with ImmGen monocyte-dendritic-precursor genes sets, and named genes are associated with abnormal mononuclear cell morphology (Mouse Phenotype Ontology; GO-Elite).
Extended Data Figure 8 | See next page for caption.
Extended Data Figure 8 | Counteracting functions of Irf8 and Gfi1 in myeloid cell fate choice. a, Gfi1, Irf8 and CEBPα ChIP–seq and RNA-seq tracks illustrating co-regulation at select loci. Gfi1 and Irf8 ChIP–seq was carried using crosslinked wild-type GMP cells, whereas RNA-seq was performed using non-crosslinked wild-type GMP cells. Cebpα ChIP–seq data was obtained from GEO record, accession number GSE43007. Significant peaks called by MACS are represented as bars under each ChIP–seq track. Regions that have called peaks overlapping for Gfi1, Irf8 and Cebpα are highlighted by a box. Strand-specific RNA-seq data displayed as black and grey peaks, respectively. Refseq gene structure presented at bottom for Irf8, Gfi1, Klf4, Per3, Zeb2 and Ets1. b–d, G3-tetracycline-inducible promoter-driven Gfi1 allele G3GV results in granulocytic differentiation. b, Schematic representation of the Col1a1 locus of KH2 ES cells engineered using FLP recombinase to harbour a G3-tetracycline-inducible promoter-driven Gfi1 allele. KH2 ES cells also contain a ROSA-allele which expresses the rtTA-M2 protein. Immunoblot of Gfi1 and Venus eYFP expression in ES cells. G3GV KH2 ES cells were treated with 1 μg ml⁻¹ doxycycline for 48 h, then analysed for Gfi1 and Venus expression by immunoblotting. For gel source data, see Supplementary Fig. 1. c, TaqMan analysis of gene expression in Csf1r− and Csf1r+ GMP cells, with or without doxycycline induction of G3GV using one allele encoding rtTA-M2. Mean of two technical replicates represented. d, CFU assays using lineage-negative bone marrow cells from wild type B6 or G3GV knock-in mice. Cells were cultured with or without 1 μg ml⁻¹ doxycycline, in methyl-cellulose media. The percentage distribution of colony types is displayed on the y axis. Mean CFU number (bottom) (n = 3 wells per condition). Representative plot from one of three independent experiments performed is displayed in c and d.
Extended Data Figure 9 | Bipotential GG1 cells comprise transcriptionally distinct progenitor populations. a, Colony appearance of CFU-G, CFU-M and CFU-GM respectively. Photos were taken with a 10× objective lens. b, ICGS analysis of GG1 cells with cells spanning the entire myeloid developmental spectrum (Fig. 1b). Cells were separated according to flow cytometric sort gates. c, Hierarchical clustering using genes in panel b that are expressed in GG1 cells (TPM >1) identifies four distinct sub clusters. d, Finding GG1-like cells in the existing scRNA-seq data set. HOPACH clustering of the same genes and cells from Extended Data Fig. 9b with arrows indicating the GG1 and 16 GG1-like cells identified in the other sort gates. GG1-like cells were identified by comparing centroids from c to those from Fig. 1b. HOPACH clusters, using the LineageProfiler classification option in AltAnalyze (n = 16) (Supplementary Methods). Arrows at the top of the heat map denote GG1 and GG1-like cells. e, f, Back-gating of sorted Ir8–GFP GMP cell subpopulations (e; IG1, IG2 and IG3) or Gfi1–GFP GMP cell subpopulations (f; GG1, GG2 and GG3) showing that all populations are phenotypically GMP cells (CD16/32high CD34high).
Extended Data Figure 10 | See next page for caption.
Extended Data Figure 10 | Clustering intermediates and Irf8−/− Gfi1−/− double-knockout GMP cells. a–h, GMP cell subpopulations enriched for CFU-GM also contain eosinophil–granulocyte progenitors. a, Plots displaying the incidence and amplitude of expression of select genes (from Fig. 4a). b, TaqMan analysis of eosinophil gene expression (Il5ra, Epx and Prg2) in the GMP cell subpopulations from Gfi1−GFP heterozygous mice. c, CFU assays with GMP cell subsets in media containing IL-3, GM-CSF, IL-5, SCF and TPO. d, CFU assays with GMP cell subsets with media containing IL-5 and SCF (which supports eosinophil–granulocyte colonies). e, TaqMan analysis of eosinophil gene expression in colonies from GG1 cells. Mean CFU number of 2 technical replicates with similar results from 2 biological replicates. f, Cytospin analysis of eosinophils in GG1-derived CFU from i. g, Flow cytometry analysis for eosinophil–granulocyte markers CCR3 and SiglecF on colonies from GG1 cells. Nearly all the GG1-derived IL-5 and SCF CFUs are positive for eosinophil markers. Representative FACS plot shown. h, TaqMan analysis of eosinophil genes (Il5ra, Epx, Prg2) in the GMP cell subpopulations from Irf8−GFP heterozygous mice. i, ICGS of GG1 and IG2 cells with cells spanning the entire myeloid developmental spectrum (Fig. 1b). Cells were separated according to flow cytometry sort gates. j–k, GG1 and IG1 cells that are enriched for CFU-GM also preferentially express HSCP1– and HSCP2-cluster genes. j, TaqMan analysis of HSCP1–HSCP2 genes in the GMP GG subpopulations. k, TaqMan analysis of HSCP1–HSCP2 genes in sorted IG subpopulations. l, Clustering Irf8−/− Gfi1−/− double-knockout single-cell libraries. HOPACH hierarchical clustering of all cells from Fig. 1b, as well as IG2 and Irf8−/− Gfi1−/− double-knockout, single-cell libraries. Only genes from Fig. 1b and in the previously clustered results were included. Genes and cells outlined in the dotted box were re-clustered with HOPACH to delineate relationships between monocytic and granulocytic cell programming among the different indicated cell populations (Fig. 4c). Representative plot of the mean of 2 technical replicates from 1 of 3 independent experiments performed displayed in b–e, h, j and k.
Author Correction: Single-cell analysis of mixed-lineage states leading to a binary cell fate choice

Andre Olsson, Meenakshi Venkatasubramanian, Virendra K. Chaudhri, Bruce J. Aronow, Nathan Salomonis, Harinder Singh & H. Leighton Grimes

Correction to: Nature https://doi.org/10.1038/nature19348, published online 31 August 2016.

In this Letter, the first name of author Virendra K. Chaudhri was incorrectly spelled 'Viren'. In addition, author Meenakshi Venkatasubramanian should have also been associated with the affiliation 'Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, Ohio 45221, USA' and authors Bruce J. Aronow, Nathan Salomonis, Harinder Singh and H. Leighton Grimes should all have also been associated with the affiliation 'Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio 45229, USA'. The original Letter has not been corrected online.