



Published in final edited form as:

Nature. 2013 June 13; 498(7453): 236–240. doi:10.1038/nature12172.

Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells

Alex K. Shalek^{1,*}, Rahul Satija^{2,*}, Xian Adiconis², Rona S. Gertner¹, Jellert T. Gaublomme¹, Raktima Raychowdhury², Schragi Schwartz², Nir Yosef², Christine Malboeuf², Diana Lu², John T. Trombetta², Dave Gennert², Andreas Gnirke², Alon Goren^{2,3}, Nir Hacohen^{2,4}, Joshua Z. Levin², Hongkun Park^{1,2,†}, and Aviv Regev^{2,5,†}

¹Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

²Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

³Department of Pathology & Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129, USA

⁴Center for Immunology and Inflammatory Diseases & Department of Medicine, Massachusetts General Hospital, Charlestown, MA 02129, USA

⁵Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02140, USA

Abstract

Recent molecular studies have revealed that, even when derived from a seemingly homogenous population, individual cells can exhibit substantial differences in gene expression, protein levels, and phenotypic output^{1–5}, with important functional consequences^{4,5}. Existing studies of cellular heterogeneity, however, have typically measured only a few pre-selected RNAs^{1,2} or proteins^{5,6} simultaneously because genomic profiling methods³ could not be applied to single cells until very recently^{7–10}. Here, we use single-cell RNA-Seq to investigate heterogeneity in the response of bone marrow derived dendritic cells (BMDCs) to lipopolysaccharide (LPS). We find extensive, and previously unobserved, bimodal variation in mRNA abundance and splicing patterns, which we validate by RNA-fluorescence *in situ* hybridization (RNA-FISH) for select transcripts. In particular, hundreds of key immune genes are bimodally expressed across cells, surprisingly even for genes that are very highly expressed at the population average. Moreover, splicing patterns demonstrate previously unobserved levels of heterogeneity between cells. Some of the observed bimodality can be attributed to closely related, yet distinct, known maturity states of BMDCs; other portions reflect differences in the usage of key regulatory circuits. For example, we identify

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[†]To whom correspondence should be addressed: Hongkun_Park@harvard.edu (HP), aregev@broadinstitute.org (AR).

*These authors contributed equally to this work.

Author Contributions

AR, HP, JZL, NH, AKS, RS, AIG, & AG conceived and designed the study. AKS, XA, RSG, JTG, RR, CM, DL, JTT, DG & JG performed experiments. RS, AKS, SS, & NY performed computational analyses. RS, AKS, AIG, NH, JZL, HP, & AR wrote the manuscript, with extensive input from all authors. The authors declare no competing financial interests.

a module of 137 highly variable, yet co-regulated, antiviral response genes. Using cells from knockout mice, we show that variability in this module may be propagated through an interferon feedback circuit involving the transcriptional regulators Stat2 and Irf7. Our study demonstrates the power and promise of single-cell genomics in uncovering functional diversity between cells and in deciphering cell states and circuits.

To characterize the extent of expression variability on a genomic scale and decipher its functional implications, we used single-cell RNA-Seq to profile a temporal snapshot of the BMDC response to LPS. This is an attractive model system for single-cell analyses for several reasons. First, LPS, a component of gram-negative bacteria and a ligand of Toll-like receptor 4, strongly synchronizes cellular responses and mitigates temporal phasing¹¹. Second, LPS activation evokes a robust transcriptional program that has been extensively investigated at the population level¹². Third, LPS stimulation should increase the correlation between mRNA and protein levels for induced genes, thus reducing a potentially confounding factor¹³. Lastly, differentiated BMDCs are post-mitotic, largely removing cell cycle-dependent transcriptional variation³.

We stimulated BMDCs with LPS and harvested single cells after four hours¹² (Supplementary Information (SI)). Using SMART-Seq⁹, we constructed cDNA libraries from 18 single BMDCs (S1–S18), three replicate populations of 10,000 cells, and two negative controls (empty wells), and sequenced each to an average depth of 27-million read-pairs. Negative control libraries failed to align (<0.25%) to the mouse genome, and were discarded from all further analyses. Library quality metrics, such as genomic alignment rates, rRNA contamination, and 3' or 5' coverage bias, were similar across all libraries (Supplementary Table 1). We estimated expression levels for all UCSC-annotated genes using RSEM¹⁴ (Supplementary Table 2, SI) and discarded genes that were not appreciably expressed (transcripts per million (TPM) > 1) in at least three individual cells, retaining 6,313 genes for further analysis.

While the gene expression levels of population replicates were tightly correlated with one another (Pearson $r > 0.98$, log-scale, Fig. 1a), there were substantial differences in expression between individual cells ($0.29 < r < 0.62$, mean: 0.48, Fig. 1b, Supplementary Fig. 1). Despite this extensive cell-to-cell variation, expression levels for an “average” single cell correlated well with the population samples ($0.79 < r < 0.81$, Fig. 1c, Supplementary Fig. 1).

We used RNA-FISH, an amplification-free imaging technique², to verify that heterogeneity in our single-cell expression data reflected true biological differences, rather than technical noise associated with the amplification of small amounts of cellular RNA. For 25 genes, selected to cover a wide range of expression levels, the variation in gene expression detected by RNA-FISH closely mirrored the heterogeneity observed in our sequencing data (Fig. 1d–g, Supplementary Fig. 2). For example, expression of housekeeping genes (*e.g.*, Beta-Actin (Actb), Beta-2-microglobulin (B2m)) matched a log-normal distribution in both single-cell RNA-Seq and RNA-FISH measurements, consistent with previous studies¹. In contrast, many genes involved in the LPS response, although highly expressed on average, showed

significantly greater levels of heterogeneity, with expression levels deviating ~1,000 fold between individual cells in extreme cases (Fig. 1e–g).

More generally, we observed that single cell variability existed across a wide range of population expression levels (Fig. 2a). Of the 522 most highly expressed genes (single-cell average TPM > 250, Fig. 2a: unshaded region, Supplementary Table 3), 281 had low cell-to-cell variability (coefficient of variation (CV) < 0.25, SI) and were well described by log-normal distributions (RNA-Seq: Fig. 2b,c top, RNA-FISH (Actb, B2m): Supplementary Fig. 2). These 281 genes were enriched for housekeeping genes, encoding ribosomal and other structural proteins (Supplementary Table 2 & 3, Bonferroni-corrected $p=1.5\times 10^{-6}$), consistent with previous findings in yeast¹⁵ and mammalian cells¹.

Surprisingly, however, 185 of the remaining 241 (coefficient of variation (CV) > 0.25, SI) highly expressed genes had bimodal expression patterns (Fig. 2b,c bottom): mRNA levels for these genes were high in many of the cells, but were at least an order of magnitude lower (often very low or undetectable) than the single-cell average in three or more cells. We independently verified this disparity by RNA-FISH (*e.g.*, Cxcl1, Cxcl10, Ifit1, and others: Fig. 1f,g & Supplementary Fig. 2), confirming that it was not a result of technical noise. This variable set included both antiviral and inflammatory response genes, and was highly enriched for genes whose expression was increased by at least two-fold upon LPS stimulation at the population level¹⁶ ($p = 2.7\times 10^{-7}$; hypergeometric test; Supplementary Table 2). Still, bimodal expression was not a universal feature of immune response transcripts; some key chemokines and chemokine receptors (Ccl3, Ccl4, Ccr12), cytokines (Cxcl2), and signaling molecules (Tank) were highly expressed in every cell (Supplementary Fig. 3), indicating that all cells were indeed activated by LPS.

This degree of variation in expression for highly expressed (on average) transcripts has not been observed in previous reports^{7–10}. For example, examination of published single-cell RNA-Seq datasets of human embryonic stem cells⁹ (Fig. 2a), mouse embryonic stem cells, and terminally differentiated fibroblasts¹⁰ (Supplementary Fig. 4) revealed far less heterogeneity in expression for highly abundant (population average) genes. Similarly, studies of protein expression in mid-log yeast cells and dividing human cell lines^{15,17} did not find such bimodality in (on average) highly expressed genes. We thus hypothesized that widespread variability in single-cell gene expression may reflect functionally important differences in the stimulated BMDC population.

Furthermore, we found that splicing patterns also showed previously unobserved levels of heterogeneity across single cells. Specifically, for genes that have multiple splice isoforms at the population level, individual cells predominantly expressed one particular isoform. We calculated the frequency (percent spliced in, PSI) of previously annotated splicing events in each of our samples using MISO¹⁸, a Bayesian framework for calculating isoform ratios (Supplementary Table 4). Although the population-derived estimates were highly reproducible, single cells exhibited significant variability in their exon-inclusion frequencies (Fig. 3a,b).

We considered the possibility that PCR amplification (intrinsic to the library preparation process) could potentially produce an overestimation of isoform regulation variability, particularly for weakly expressed transcripts¹⁹. However, even when we limited our analysis to 89 alternatively spliced exons ($0.2 < \text{population PSI} < 0.8$) that were very highly expressed within a single cell (single cell TPM > 250 , SI), we still observed the same variability in splicing patterns amongst individual cells, with highly skewed expression towards a single splice variant (Fig. 3b). We obtained similar results when we generated three additional single-cell cDNA libraries using a slightly modified SMART-Seq protocol (SI) in which a four nucleotide barcode was introduced onto each RNA molecule during reverse transcription¹⁹, enabling us to estimate the number of unique RNA transcripts that existed prior to PCR (Supplementary Fig. 5 & 6 and SI).

To the best of our knowledge, single-cell variation in splicing patterns has rarely been studied for individual genes, and never been analyzed on a genomic scale. One recent report²⁰ used RNA-FISH to study variation in alternative isoforms in two genes, and observed lower levels of isoform variability across single cells (the levels of heterogeneity differed in different cell types). Another study that used fluorescent reporters to quantify single-cell exon inclusion levels for one gene discovered highly variable and bimodal splicing patterns²¹.

To independently verify the existence of extensive differences in isoform ratios between cells, we designed RNA-FISH probes targeting constitutive and isoform-specific exons in two genes (Irf7 and Acpp, Fig. 3c and Supplementary Fig. 7 & 8)²⁰. We found substantial expression variability in overall Irf7 levels between individual cells (as reflected by the 'constitutive' probes, Fig. 3c, bottom and top panels), mirroring our single-cell sequencing results (and further explored below). Additionally, within each Irf7-expressing cell, we observed a bias towards either the inclusion or exclusion of the cassette exon (Fig. 3c, Supplementary Fig. 7, middle panel, *e.g.*, compare 'high' and 'low' marked cells). We obtained comparable results for Acpp using two probes designed to detect mutually exclusive alternative final exons (Supplementary Fig. 8).

We next explored the sources and functional implications of expression variability. Bimodality amongst highly expressed immune response genes may reflect either the presence of distinct cellular subtypes or stochastic differences in the activation of signaling circuits¹¹. We performed a principal components analysis (PCA, Fig. 4a) on our single-cell expression profiles, focusing on the 632 genes that were induced at least twofold in the population-wide response to LPS¹⁶ (Supplementary Table 5). We found two distinct subpopulations, clearly distinguishable by the first principal component (PC1, 15% of the total variation, Fig. 4a). One group of fifteen cells expressed a core set of antiviral and inflammatory defense cytokines (including: Tnf, Il1a, Il1b, and Cxcl10) at extremely high levels (TPM $> 1,000$), while the remaining three cells expressed them at far weaker levels (TPM < 50). Some cell surface proteins (Ccr7, Cd83) and chemokines (Ccl22), which are known markers of BMDC maturation, showed the opposite expression pattern (Fig. 4b, Supplementary Fig. 9).

During maturation, BMDCs switch from antigen-capturing to antigen-presenting cells that prime the adaptive immune system²². Maturation can occur either in response to pathogen-derived ligands (pathogen-dependent maturation), such as LPS, or when clusters of BMDCs are disrupted in culture²² (pathogen-independent maturation). Both processes lead to induction of maturation markers, but only pathogen-dependent maturation results in co-expression of defense cytokines.

Examining the expression of maturation markers and defense cytokines (Supplementary Fig. 9) suggested that our 18 cells represent two distinct maturity states: **(1)** fifteen cells that were in the early stages of pathogen-dependent maturation (Fig. 4a, 'maturing', triangles; grey triangles, the two cells furthest along in this process); and, **(2)** three cells that likely matured during the culturing process (Fig. 4a, 'mature', squares; pathogen-independent). We further verified the existence of these sub-populations via RNA-FISH (Supplementary Fig. 10), single-cell quantitative reverse transcription polymerase chain reaction (qRT-PCR; Supplementary Fig. 11, SI, Supplementary Table 6), and cell sorting based on surface markers identified from the RNA-Seq data (Supplementary Fig. 12, SI). These results highlight that single-cell RNA-Seq can sensitively distinguish between closely related, yet distinct, developmental states, even within the same cell type.

Since differences in cell state explain only a small portion of the observed heterogeneity, we next examined the variation that might arise from the differential activity of regulatory circuits. We reasoned that co-variation across single cells between the mRNA levels of a transcription factor and its targets would represent a potential regulatory interaction, and, furthermore, would suggest that heterogeneity in the regulator's expression may underlie the variability of its targets. Such a correlative approach has successfully identified regulatory connections from population-level transcription profiles measured in different conditions^{12,23}. Here, we attempted to apply it to multiple single cells in the same condition.

To this end, we calculated the correlation in expression profiles between every pair of induced genes across all single cells, and identified a cluster of 137 genes that varied in a correlated way and were strongly discriminated by the second principal component (PC2, 8% of the variation, Fig. 4a,b). The cluster's genes included the known antiviral master regulators *Irf7* and *Stat2*, and were highly enriched for members of the antiviral response¹² (60 of 137 genes, $p = 2.5 \times 10^{-3}$, hypergeometric test, Supplementary Table 5), as well as *STAT2* targets¹⁶ (73/137 genes, $p = 4.5 \times 10^{-5}$, hypergeometric test). Most (100/137) of the cluster's genes were bimodally expressed across single cells (Fig. 2c, bottom) despite being strongly expressed at the population level (13 genes TPM > 250; 53 genes TPM > 50). We independently validated a subset of these correlations using single-cell qRT-PCR and RNA-FISH (Fig. 4c,d). Moreover, single-cell qRT-PCR analysis of additional time points demonstrated that these correlations persisted at 6h as well (Supplementary Discussion in SI, Supplementary Fig. 13).

We hypothesized that bimodal variation in the expression of the cluster's genes may be related to differences in the levels and activities of *Stat2* and *Irf7*. To test this hypothesis, we measured expression of a set of antiviral genes by single-cell qRT-PCR in LPS-stimulated BMDCs from *Irf7* knockout (*Irf7* $-/-$) mice (SI). As expected, this perturbation ablated

expression of most of the variable antiviral transcripts in our signature, while leaving non-variable antiviral transcripts relatively unaffected (Fig. 4e). However, Stat2 expression and variability levels were unaffected by the Irf7 knockout, implying that Stat2 may act either upstream or in parallel to Irf7 during the response²⁴ (Supplementary Fig. 14). As both Stat2 and Irf7 are targets of the interferon-signaling pathway, we stimulated and profiled BMDCs from interferon receptor knockout (Ifnr $-/-$) mice. In these cells, we found drastically reduced expression for both Stat2 and Irf7, as well as all other measured cluster genes (Fig. 4f).

Our analysis provides a proof-of-concept demonstrating how co-variation between transcripts across seemingly homogeneous single cells can help to identify and assemble regulatory circuits. Specifically, in our variable circuit (Supplementary Fig. 14) interferon signaling is required for induction of Stat2 and Irf7, which, in turn, act to induce our variable antiviral cluster genes. Our experiments do not definitively determine, however, which component of the circuit causes the observed heterogeneity *per se*. One compelling possibility is that upstream noise is propagated from the interferon-signaling pathway first to Stat2 and Irf7 and then to the target genes^{25,26}. This hypothesis is supported by the variation we observed in STAT protein levels and nuclear localization (Supplementary Discussion in SI, Supplementary Fig. 15 & 16). However, since temporal snapshots of RNA and protein are not always directly comparable (Supplementary Discussion in SI, Supplementary Fig. 15 & 16), new strategies for tracing the spatiotemporal dynamics of both proteins and RNA in single living cells are needed to fully test this hypothesis¹¹.

A similar approach could potentially be used to explore the consequences of bimodality in splicing. Even looking at just 18 cells, we witnessed interesting examples of bimodal splicing patterns for genes whose isoforms have distinct functional consequences. For example, the splicing regulators Srsf3 and Srsf7 are each known to contain a “poison cassette exon” that, when included, targets the RNA for degradation via nonsense-mediated decay²⁷ (Supplementary Fig. 17). Meanwhile, splicing differences in other regulatory genes may further enhance expression diversity: for example, proteins encoded by different isoforms of Irf7 (Fig. 3c) differentially activate interferon-responsive genes *in vitro*²⁴. These examples suggest that heterogeneity in splicing may represent another layer of response encoding.

In conclusion, our study reveals extensive bimodality in the transcriptional response of BMDCs to LPS, reflected in gene expression, alternative splicing, and regulatory circuit activity. While some variation in expression reflects differences in developmental state, other bimodal patterns reflect the differential activity of an antiviral regulatory circuit in this temporal snapshot. These phenomena allowed us to treat each cell as a “perturbation system” for reconstructing cell circuits²⁸, even with relatively few cells.

Moreover, our results demonstrate how co-variation across single cells can help dissect and refine gene modules that may be indistinguishable in population-scale measurements. For instance, in a recent population-scale study¹⁶, we identified a large cluster of 808 “late-induced” LPS genes that was enriched for both maturation genes and STAT-regulated antiviral genes. These two subsets could not be separated by population-level expression

profiles alone¹⁶, but our single-cell data from a single timepoint clearly distinguishes them. Similarly, the unexpected and prevalent skewing we discovered in alternative splicing between single cells revises our molecular view of this process.

Finally, although many of our analyses focused on highly expressed genes to reduce the potential influence of amplification noise, our data also revealed substantial bimodality amongst more moderately expressed transcripts, such as large non-coding RNAs (lincRNAs, Supplementary Fig. 18). This suggests that the low population-level expression of these transcripts²⁹ may sometimes reflect high expression in a small subset of cells as opposed to uniform levels of low expression. While further technical improvements will be necessary to disentangle these two hypotheses (Supplementary Fig. 5), single-cell measurements should help facilitate the discovery and annotation of lincRNAs.

Comparing our results to other single-cell RNA-Seq data sets (*e.g.*, Fig. 2a, Supplementary Fig. 4) indicates that the source of the analyzed tissue (*in vitro* vs. *ex vivo*), the biological condition of the individual cells (steady state vs. dynamically responding), and the cellular microenvironment all likely influence the extent of single-cell heterogeneity within a system. When applied to complex tissues – such as unsorted bone marrow, developing embryos, tumors, and other rare clinical samples – the variability seen through single-cell genomics may help determine new cell classification schemes, identify transitional states, discover previously unrecognized biological distinctions, and map markers that differentiate them. Fulfilling this potential would require novel strategies to address the high levels of noise inherent in single-cell genomics – both technical, due to minute amounts of input material, and biological, *e.g.*, due to short bursts of RNA transcription³⁰. Future studies that couple technological advances in experimental preparation with novel computational approaches would enable analyses, based on hundreds or thousands of single cells, to reconstruct intracellular circuits, enumerate and redefine cell states and types, and transform our understanding of cellular decision-making on a genomic scale.

Methods Summary

BMDCs, prepared as previously described¹², were stimulated with LPS for 4h and then sorted as single cells or populations (10,000 cells) directly into TCL lysis buffer (Qiagen) supplemented with 1% v/v 2-mercaptoethanol. After performing a 2.2x clean up with Agencourt RNAClean XP Beads (Beckman Coulter), whole transcriptome-amplified cDNA products were generated using the SMARTer Ultra-low RNA Kit (Clontech), and conventional Illumina libraries were made and sequenced to an average depth of 27 million read pairs (HiSeq 2000, Illumina). Expression levels and splicing ratios were quantified using RSEM¹⁴ and MISO¹⁸, respectively. Additional experiments were performed using RNA-FISH (Panomics), Immunofluorescence, FACS, and single-cell qRT-PCR (Single Cell-to-CT (Invitrogen) and BioMark (Fluidigm)). **Full Methods** and any associated references are provided in SI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank N. Chevrier, C. Villani, M. Jovanovic, M. Bray and J. Shuga for scientific discussions; N. Friedman and E. Lander for comments on the manuscript; B. Tilton, T. Rogers, and M. Tam for assistance with cell sorting; J. Bochicchio, E. Shefler, and C. Guiducci for project management; the Broad Genomics Platform for all sequencing work; K. Fitzgerald for the *Irf7* $-/-$ bone marrow; and L. Gaffney for help with artwork. Work was supported by an NIH Postdoctoral Fellowship (1F32HD075541-01, RS), an NIH grant (U54 AI057159, NH), an NIH New Innovator Award (DP2 OD002230, NH), an NIH CEGS Award (1P50HG006193-01, HP, AR and NH), NIH Pioneer Awards (5DP1OD003893-03 to HP, DP1OD003958-01 to AR), the Broad Institute (HP and AR), HHMI (AR), and the Klarman Cell Observatory at the Broad Institute (AR).

References

1. Bengtsson M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Research*. 2005; 15:1388–1392. doi:papers2://publication/doi/10.1101/gr.3820805. [PubMed: 16204192]
2. Raj A, Van Oudenaarden A. Single-Molecule Approaches to Stochastic Gene Expression. *Annual Review of Biophysics*. 2009; 38:255–270.10.1146/annurev.biophys.37.032807.125928
3. Kalisky T, Blainey P, Quake SR. Genomic Analysis at the Single-Cell Level. *Annual review of genetics*. 2011; 45:431–445. doi:papers2://publication/doi/10.1146/annurev-genet-102209-163607.
4. Feinerman O, et al. Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Molecular Systems Biology*. 2010; 6:1–16. doi:papers2://publication/doi/10.1038/msb.2010.90.
5. Cohen AA, et al. Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science*. 2008; 322:1511–1516.10.1126/science.1160165 [PubMed: 19023046]
6. Bendall SC, Nolan GP. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science (New York, NY)*. 2011; 332:677–678.10.1126/science.1206351
7. Islam S, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*. 2011 doi:papers2://publication/doi/10.1101/gr.110882.110.
8. Tang F, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*. 2009; 6:377–382.10.1038/nmeth.1315 [PubMed: 19349980]
9. Ramskold D, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*. 2012; 30:777–782. doi:papers2://publication/doi/10.1038/nbt.2282.
10. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports*. 10.1016/j.celrep.2012.08.003
11. Tay S, et al. Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature*. 2010; 466:267–271. doi:papers2://publication/doi/10.1038/nature09145. [PubMed: 20581820]
12. Amit I, et al. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science*. 2009; 326:257–263.10.1126/science.1179050 [PubMed: 19729616]
13. Li GW, Xie XS. Central dogma at the single-molecule level in living cells. *Nature*. 2011; 475:308–315. [PubMed: 21776076]
14. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011; 12:323–323. [PubMed: 21816040]
15. Bar-Even A, et al. Noise in protein expression scales with natural protein abundance. *Nature Genetics*. 2006; 38:636–643. [PubMed: 16715097]
16. Garber M, et al. A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Molecular Cell*. 2012; 47:810–822.10.1016/j.molcel.2012.07.030 [PubMed: 22940246]
17. Sigal A, et al. Variability and memory of protein levels in human cells. *Nature*. 2006; 444:643–646.10.1038/nature05316 [PubMed: 17122776]
18. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*. 2010; 7:1009–1015. [PubMed: 21057496]

19. Kivioja T, et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*. 2011; 9:72–74. doi:10.1038/nmeth.1778 [PubMed: 22101854]
20. Waks Z, Klein AM, Silver PA. Cell-to-cell variability of alternative RNA splicing. *Molecular Systems Biology*. 2011; 7:1–12. doi:papers2://publication/doi/10.1038/msb.2011.32.
21. Gurskaya NG, et al. Analysis of alternative splicing of cassette exons at single-cell level using two fluorescent proteins. *Nucleic Acids Research*. 2012;40.10.1093/nar/gkr1314
22. Jiang A, et al. Disruption of E-Cadherin-Mediated Adhesion Induces a Functionally Distinct Pathway of Dendritic Cell Maturation. *Immunity*. 2007; 27:610–624. doi:papers2://publication/doi/10.1016/j.immuni.2007.08.015. [PubMed: 17936032]
23. Friedman N. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science (New York, NY)*. 2004; 303:799–805. doi:10.1126/science.1094068
24. Ning S, Huye LE, Pagano JS. Regulation of the Transcriptional Activity of the IRF7 Promoter by a Pathway Independent of Interferon Signaling. *Journal of Biological Chemistry*. 2005; 280:12262–12270. [PubMed: 15664995]
25. Zhao M, Zhang J, Phatnani H, Scheu S, Maniatis T. Stochastic Expression of the Interferon- β Gene. *PLoS biology*. 2012; 10:e1001249. [PubMed: 22291574]
26. Rand U, et al. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular Systems Biology*. 2012; 8:10.1038/msb.2012.17
27. Änkö M-L, et al. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biology*. 2012; 13:10.1186/gb-2012-13-3-r17
28. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science (New York, NY)*. 2005
29. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*. 2011; 25:1915–1927. [PubMed: 21890647]
30. Cai L, Dalal CK, Elowitz MB. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*. 2008; 455:485–490. doi:papers2://publication/doi/10.1038/nature07292. [PubMed: 18818649]

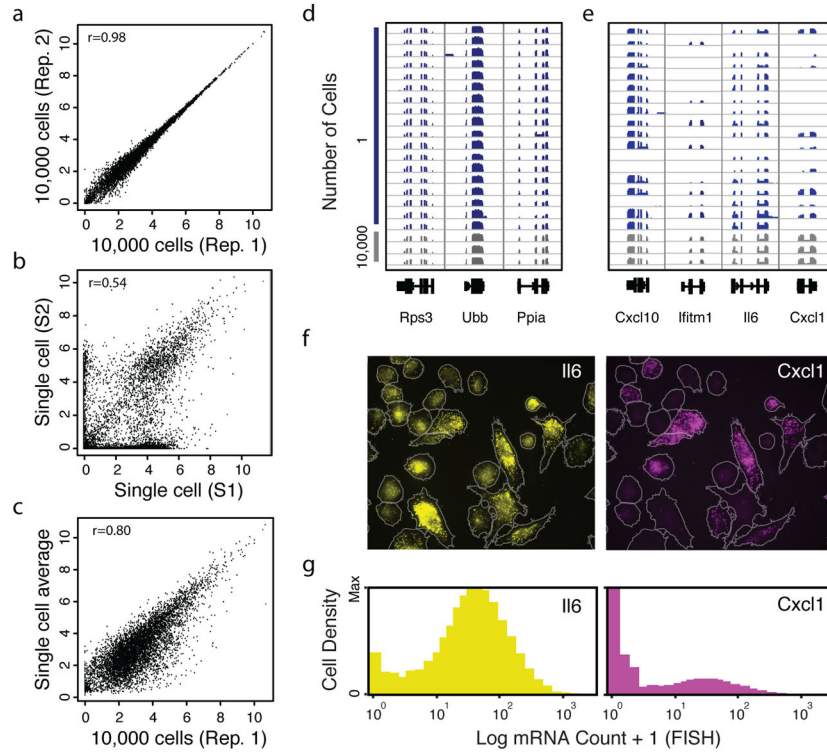


Figure 1. Single-cell RNA-Seq of LPS-stimulated BMDCs reveals extensive transcriptome heterogeneity

a–c, Correlations of transcript expression levels (x & y-axes: log-scale TPM+1) between two 10,000-cell population replicates (**a**), two single cells (**b**), and the ‘average’ single cell and a population (**c**). **d,e**, RNA-Seq read densities in single cells (blue) and population replicates (grey) for three non-variable genes (**d**) and four variable ones (**e**). **f–g**, RNA-FISH of representative transcripts. Optical micrographs (cell boundaries; grey outlines) and maximum-normalized distributions of expression levels from a RNA-FISH co-staining (n = 3,193 cells) for Il6 (yellow) and Cxcl1 (magenta).

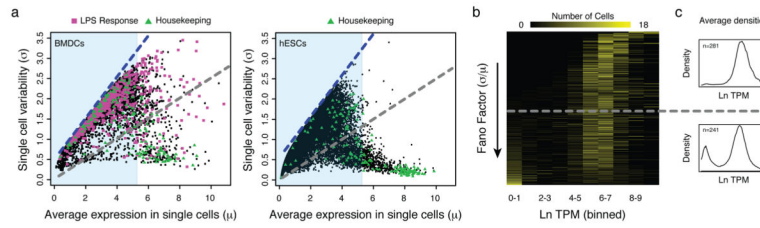


Figure 2. Bimodal variation in expression levels across single cells

a, Relationship between average expression level in single cells (μ , X axis) and standard deviation (σ , Y axis) for 6,313 genes (Supplementary Table 2). Blue dashed line: maximum theoretical σ for an average expression level (SI); Grey dashed line: constant Fano factor ($\sigma/\mu = 0.25$). Magenta: immune response genes; Green: housekeeping genes; light blue shaded region: single-cell average TPM < 250. **b**, Cellular heterogeneity for the 522 most highly expressed genes (single cell average; Supplementary Table 3). Each row represents a discretized histogram for a single gene (sorted by the Fano factor from low to high (top to bottom)). Color represents the number of cells (yellow: 18 cells; black: 0) that express the gene at the noted level. Grey dashed line denotes the constant Fano factor (0.25) highlighted in (a). **c**, Averaged expression density distributions for the 281 low-variability genes (top) and the 241 highly variable genes (bottom).

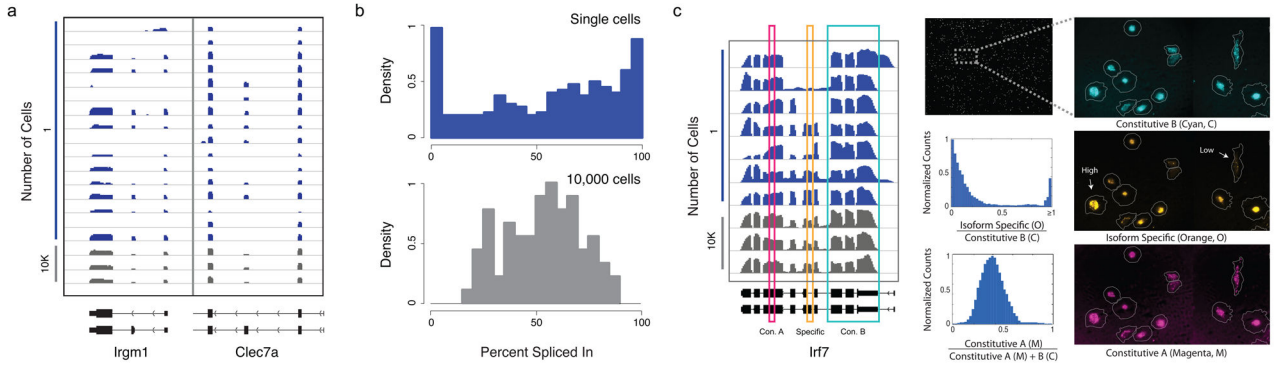


Figure 3. Variation in isoform usage between single cells
a. RNA-Seq read densities in single cells (blue) and population replicates (grey) for two illustrative loci, each with two different isoforms (bottom). **b.** Distributions of exon inclusion (PSI scores, X axis) for alternatively spliced exons of highly expressed genes (single-cell TPM > 250) in individual cells (blue histogram, top) and populations (grey histogram, bottom). **c.** Left: RNA-Seq read densities for Irf7 (only cells where the transcript is expressed are shown). Colored boxes mark exons analyzed by RNA-FISH. Right: RNA-FISH images from simultaneous hybridization with probes for two constitutive ('Con') regions of the transcript (A: cyan (C); B: magenta (M)) and one alternatively spliced exon ('Specific': orange (O)). White arrows (middle panel) highlight two cells with high levels of Irf7, but opposite preferences for the alternatively spliced exon. Histograms showing global abundance ratios for isoform-specific and constitutive probes (cells with less than 5 constitutive counts have been excluded; n = 490 cells; bottom histogram deviates from 0.5 due to probe design, see SI).

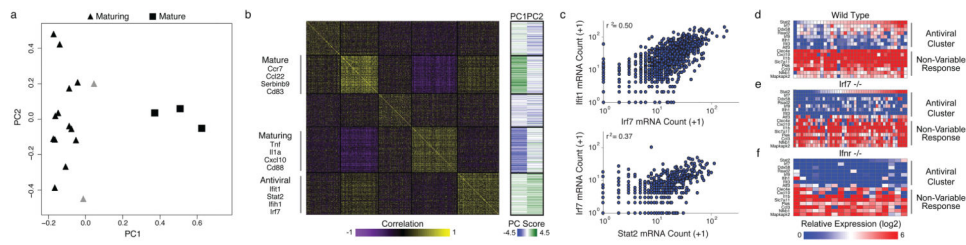


Figure 4. Analysis of co-variation in single-cell mRNA expression levels reveals distinct maturity states and an antiviral cell circuit

a, PCA of 632 LPS-induced genes. Contributions of each cell (points) to the first two principal components. b, Clustered correlation matrix of induced genes. Left: the Pearson correlation coefficients (r) between single-cell expression profiles of every pair of 632 LPS-induced genes (rows, columns). Right: the projection score (green: high; blue: low) for each gene (row) onto PC1 (left) and PC2 (right). c, Confirmation of correlations for Irf7-Stat2 ($n = 655$ cells) and Irf7-Ifit1 ($n = 934$ cells) by RNA-FISH. d-f, Expression levels for 16 genes in single BMDCs (columns), measured using single-cell qRT-PCR, in wild type (WT) ($n = 36$) (d), Irf7 $-/-$ ($n = 47$) (e), and Ifnr $-/-$ ($n = 18$) (f) at 4h after LPS stimulation (SI).