LETTER

doi:10.1038/nature12172

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

Alex K. Shalek[1]*, Rahul Satija[2]*, Xian Adiconis[2], Rona S. Gertner[1], Jellert T. Gaublomme[1], Raktima Raychowdhury[2], Schraga Schwartz[2], Nir Yosef[2], Christine Malboeuf[2], Diana Lu[2], John J. Trombetta[2], Dave Gennert[2], Andreas Gnirke[2], Alon Goren[2,3], Nir Hacohen[2,4], Joshua Z. Levin[2], Hongkun Park[1,2] & Aviv Regev[2,5]

**Recent molecular studies have shown 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 methods[3] could not be applied to single cells until very recently[7–10]. Here we use single-cell RNA sequencing to investigate heterogeneity in the response of mouse bone-marrow-derived dendritic cells (BMDCs) to lipopolysaccharide. We find extensive, and previously unobserved, bimodal variation in messenger RNA abundance and splicing patterns, which we validate by RNA-fluorescence *in situ* hybridization 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 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 sequencing (RNA-Seq) to profile a temporal snapshot of the BMDC response to lipopolysaccharide (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 phasing[11]. Second, LPS activation evokes a robust transcriptional program that has been extensively investigated at the population level[12]. Third, LPS stimulation should increase the correlation between mRNA and protein levels for induced genes, thus reducing a potentially confounding factor[13]. Finally, differentiated BMDCs are post-mitotic, largely removing cell cycle-dependent transcriptional variation[3].

We stimulated BMDCs with LPS and collected single cells after four hours[12] (Supplementary Information). Using SMART-Seq[9], we constructed complementary DNA 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, ribosomal RNA contamination, and 3′ or 5′ coverage bias, were similar across all libraries (Supplementary Table 1). We estimated expression levels for all University of California Santa Cruz (UCSC)-annotated genes using RSEM (RNA-Seq by expectation maximization)[14] (Supplementary Table 2) 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.

Although 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 and Supplementary

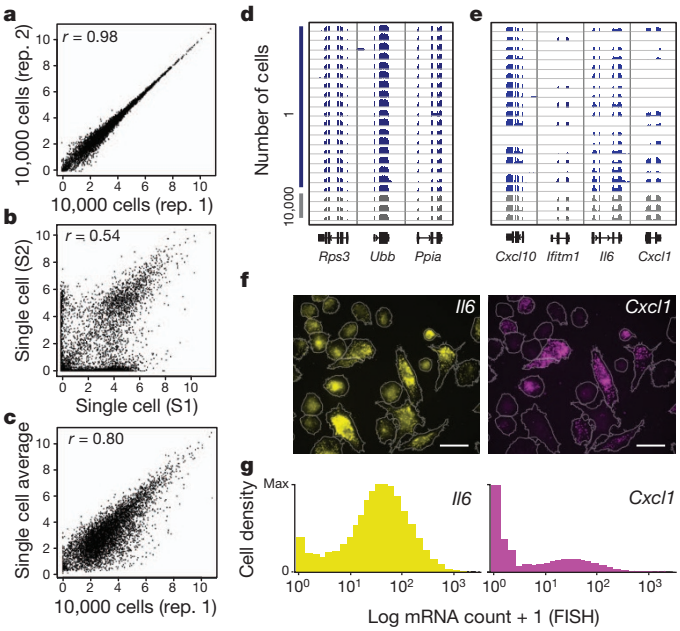


**Figure 1 | Single-cell RNA-Seq of LPS-stimulated BMDCs reveals extensive transcriptome heterogeneity.** **a**–**c**, Correlations of transcript expression levels (*x* and *y*-axes: log-scale TPM + 1) between two 10,000-cell population replicates (rep.) (**a**), two single cells (S1 and S2) (**b**), 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). Scale bars, 25 μm.

[1]Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. [2]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachuestts 02142, USA. [3]Department of Pathology & Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [4]Center for Immunology and Inflammatory Diseases & Department of Medicine, Massachusetts General Hospital, Charlestown, Massachuestts 02129, USA. [5]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02140, USA.
*These authors contributed equally to this work.

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 and Supplementary Fig. 1).

We used RNA-fluorescence *in situ* hybridization (RNA-FISH), an amplification-free imaging technique[2], 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 and Supplementary Fig. 2). For example, expression of housekeeping genes (such as β-actin (*Actb*) and β2-microglobulin (*B2m*)) matched a log-normal distribution in both single-cell RNA-Seq and RNA-FISH measurements, consistent with previous studies[1]. By 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, and Supplementary Table 3), 281 had low cell-to-cell variability (coefficient of variation (CV, $\sigma/\mu$) < 0.25; Supplementary Information) 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 Tables 2 and 3; Bonferroni-corrected $P = 1.5 \times 10^{-6}$), consistent with previous findings in yeast[15] and mammalian cells[1].

Notably, however, 185 of the remaining 241 (CV > 0.25; Supplementary Information) 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 (for example, *Cxcl1*, *Cxcl10* and *Ifit1*; Fig. 1f, g and 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 in which expression was increased by at least twofold after LPS stimulation at the population level[16] ($P = 2.7 \times 10^{-7}$; 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* and *Ccrl2*), cytokines (*Cxcl2*), and signalling 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 data sets of human embryonic stem cells[9] (Fig. 2a), mouse embryonic stem cells, and terminally differentiated fibroblasts[10] (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 proposed 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 (percentage spliced in (PSI)) of previously annotated splicing events in each of our samples using MISO[18], 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 transcripts[19]. However, even when we limited our analysis to 89 alternatively spliced exons (0.2 < population PSI < 0.8) that were very highly expressed within a single cell (single cell TPM > 250; Supplementary Information), we still observed the same variability in splicing patterns among 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 (Supplementary Information) in which a four-nucleotide barcode was introduced onto each RNA molecule during reverse transcription[19], enabling us to estimate the number of unique RNA transcripts that existed before PCR (Supplementary Figs 5 and 6 and Supplementary Information).

To the best of our knowledge, single-cell variation in splicing patterns has rarely been studied for individual genes, and never been analysed on a genomic scale. One recent report[20] used RNA-FISH to

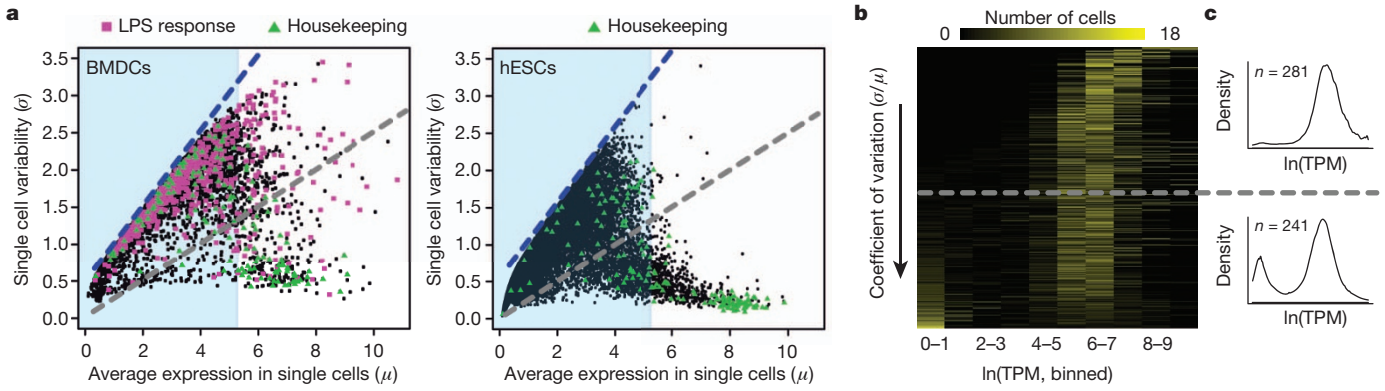


**Figure 2 | Bimodal variation in expression levels across single cells.** **a**, Relationship between average expression level in single cells ($\mu$, $x$ axis) and standard deviation ($\sigma$, $y$ axis) for 6,313 genes (Supplementary Table 2). Blue dashed line denotes maximum theoretical $\sigma$ for an average expression level (Supplementary Information). Grey dashed line denotes the constant coefficient of variation (CV, $\sigma/\mu = 0.25$). Magenta represents immune response genes; green denotes housekeeping genes; light blue shaded region represents 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 CV from low to high (top to bottom)). Colour represents the number of cells (yellow: 18 cells; black: 0) that express the gene at the noted level. Grey dashed line denotes the constant CV (0.25) highlighted in (**a**). **c**, Averaged expression density distributions for the 281 low-variability genes (top) and the 241 highly variable genes (bottom).

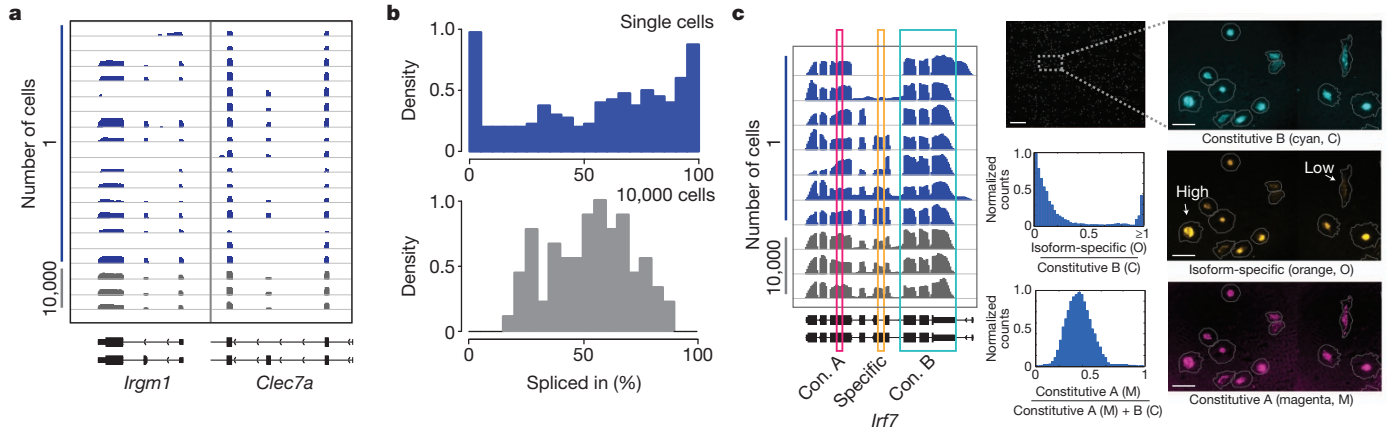


**Figure 3 | Variation in isoform usage between single cells.** **a**, RNA-Seq read densities in single cells (blue) and population replicates (grey) for two illustrative loci, each with two different isoforms (bottom). **b**, Distributions of exon inclusion (PSI scores, *x* axis) for alternatively spliced 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 in which the transcript is expressed are shown). Coloured boxes mark exons analysed 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) highlight two cells with high levels of *Irf7*, but opposite preferences for the alternatively spliced exon. Histograms show global abundance ratios for isoform-specific and constitutive probes (cells with less than five constitutive counts have been excluded; $n = 490$ cells; bottom histogram deviates from 0.5 owing to probe design; see Supplementary Information). Scale bars, 250 μm (left); 25 μm (right).

study variation in alternative isoform usage 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 patterns[21].

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[20] (*Irf7* and *Acpp*; Fig. 3c and Supplementary Figs 7 and 8). We found substantial expression variability in overall *Irf7* levels between individual cells (as reflected by the 'constitutive' probes; Fig. 3c, top and bottom), mirroring our single-cell sequencing results (and further explored below). Furthermore, within each Irf7-expressing cell, we observed a bias towards either the inclusion or exclusion of the cassette exon (Fig. 3c and Supplementary Fig. 7, middle; for example, 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 among highly expressed immune response genes may reflect the presence of distinct cellular subtypes or stochastic differences in the activation of regulatory circuits[11]. We performed a principal components analysis (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 LPS[16] (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 defence cytokines (including *Tnf*, *Il1a*, *Il1b* and *Cxcl10*) at extremely high levels (TPM > 1,000), whereas the remaining three cells expressed them at far weaker levels (TPM < 50). Some cell surface proteins (*Ccr7* and *Cd83*) and chemokines (*Ccl22*), which are known markers of BMDC maturation, showed the opposite expression pattern (Fig. 4b and Supplementary Fig. 9).

During maturation, BMDCs switch from antigen-capturing to antigen-presenting cells that prime the adaptive immune system[22]. Maturation can occur either in response to pathogen-derived ligands (pathogen-dependent maturation), such as LPS, or when clusters of BMDCs are disrupted in culture[22] (pathogen-independent maturation).

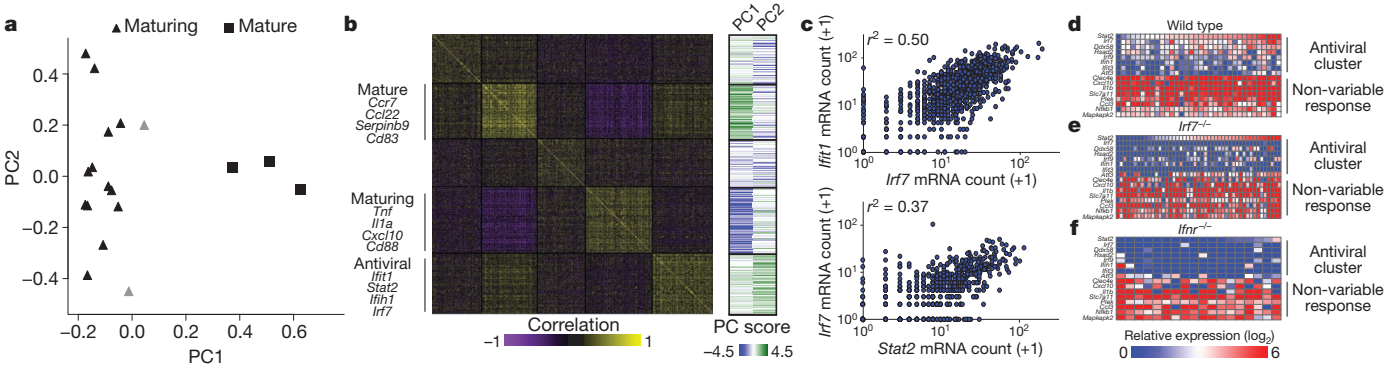


**Figure 4 | Analysis of co-variation in single-cell mRNA expression levels reveals distinct maturity states and an antiviral cell circuit.** **a**, Principal components analysis of 632 LPS-induced genes. Contributions of each cell (points) to the first two principal components (PC1 and PC2). **b**, 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 ($n = 36$) (**d**), $Irf7^{-/-}$ ($n = 47$) (**e**) and $Ifnr^{-/-}$ ($n = 18$) (**f**), at 4 h after LPS stimulation (Supplementary Information).

Both processes lead to induction of maturation markers, but only pathogen-dependent maturation results in co-expression of defence cytokines.

Examining the expression of maturation markers and defence cytokines (Supplementary Fig. 9) suggested that our 18 cells represent two distinct maturity states: (1) 15 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 probably 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 PCR (qRT–PCR; Supplementary Fig. 11, Supplementary Information and Supplementary Table 6), and cell sorting based on surface markers identified from the RNA-Seq data (Supplementary Fig. 12 and Supplementary Information). These results highlight that single-cell RNA-Seq can sensitively distinguish between closely related, yet distinct, developmental states, even within the same cell type.

Because 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 several 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 genes of this cluster included the known antiviral master regulators *Irf7* and *Stat2*, and were highly enriched for members of the antiviral response[12] (60 out of 137 genes, $P = 2.5 \times 10^{-3}$, hypergeometric test; Supplementary Table 5), as well as STAT2 targets[16] (73 out of 137 genes, $P = 4.5 \times 10^{-5}$, hypergeometric test). Most (100 out of 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 6 h as well (Supplementary Discussion and 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 (Supplementary Information). 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, indicating that Stat2 may act either upstream or in parallel to Irf7 during the response[24] (Supplementary Fig. 14). As both Stat2 and Irf7 are targets of the interferon-signalling pathway, we stimulated and profiled BMDCs from interferon receptor knockout ($Ifnr^{-/-}$) mice. In these cells, we found markedly 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 signalling is required for the 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-signalling pathway first to *Stat2* and *Irf7* and then to the target genes[25,26]. This hypothesis is supported by the variation we observed in Stat1 and Stat2 protein levels and nuclear localization (Supplementary Discussion and Supplementary Figs 15 and 16). However, because temporal snapshots of RNA and protein are not always directly comparable (Supplementary Discussion and Supplementary Figs 15 and 16), new strategies for tracing the spatiotemporal dynamics of both proteins and RNA in single living cells are needed to fully test this hypothesis[11].

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 decay[27] (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. Although 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 circuits[28], 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 study[16], 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 alone[16], but our single-cell data from a single time point clearly distinguishes them. Similarly, the unexpected and prevalent skewing we discovered in alternative splicing between single cells revises our molecular view of this process. Furthermore, although many of our analyses focused on highly expressed genes to reduce the potential influence of amplification noise, our data also revealed substantial bimodality among more moderately expressed transcripts, such as large non-coding RNAs (lincRNAs; Supplementary Fig. 18). This suggests that the low population-level expression of these transcripts[29] may sometimes reflect high expression in a small subset of cells as opposed to uniform levels of low expression. Although further technical improvements will be necessary to disentangle these two hypotheses (Supplementary Fig. 5), single-cell measurements should help to facilitate the discovery and annotation of lincRNAs.

Comparing our results to other single-cell RNA-Seq data sets (for example, Fig. 2a and Supplementary Fig. 4) indicates that the source of the analysed tissue (*in vitro* versus *ex vivo*), the biological condition of the individual cells (steady state versus dynamically responding), and the cellular microenvironment all probably influence the extent of single-cell heterogeneity within a system. When applied to complex tissues—such as unsorted bone marrow, developing embryos, tumours and other rare clinical samples—the variability seen through single-cell genomics may help to determine new cell classification schemes, identify transitional states, discover previously unrecognized biological distinctions, and map markers that differentiate them. Fulfilling this potential would require new strategies to address the high levels of noise inherent in single-cell genomics—both technical, owing

to minute amounts of input material, and biological, for example, owing to short bursts of RNA transcription[30]. Future studies that couple technological advances in experimental preparation with new 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 described[12], were stimulated with LPS for 4 h 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 2.2× 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 RSEM[14] and MISO[18], respectively. Additional experiments were performed using RNA-FISH (Panomics), immunofluorescence, FACS and single-cell qRT–PCR (Single Cell-to-CT (Invitrogen) and BioMark (Fludigm)). Full Methods and any associated references are provided in the Supplementary Information.

**Received 2 November 2012; accepted 5 April 2013.**

**Published online 19 May; corrected online 12 June 2013 (see full-text HTML version for details).**

1. Bengtsson, M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.* **15,** 1388–1392 (2005).
2. Raj, A. & Van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Ann. Rev. Biophys.* **38,** 255–270 (2009).
3. Kalisky, T., Blainey, P. & Quake, S. R. Genomic analysis at the single-cell level. *Ann. Rev. Gen.* **45,** 431–445 (2011).
4. Feinerman, O. *et al.* Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol. Sys. Biol.* **6,** 1–16 (2010).
5. Cohen, A. A. *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322,** 1511–1516 (2008).
6. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332,** 687–696 (2011).
7. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21,** 1160–1167 (2011).
8. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6,** 377–382 (2009).
9. Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotech.* **30,** 777–782 (2012).
10. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2,** 666–673 (2012).
11. Tay, S. *et al.* Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature* **466,** 267–271 (2010).
12. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326,** 257–263 (2009).
13. Li, G.-W. & Xie, X. S. Central dogma at the single-molecule level in living cells. *Nature* **475,** 308–315 (2011).
14. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
15. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38,** 636–643 (2006).
16. Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* **47,** 810–822 (2012).
17. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
18. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7,** 1009–1015 (2010).
19. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* **9,** 72–74 (2012).
20. Waks, Z., Klein, A. M. & Silver, P. A. Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol.* **7,** 1–12 (2011).
21. Gurskaya, N. G. *et al.* Analysis of alternative splicing of cassette exons at single-cell level using two fluorescent proteins. *Nucleic Acids Res.* **40,** e57 (2012).
22. Jiang, A. *et al.* Disruption of E-cadherin-mediated adhesion induces a functionally distinct pathway of dendritic cell maturation. *Immunity* **27,** 610–624 (2007).
23. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303,** 799–805 (2004).
24. Ning, S., Huye, L. E. & Pagano, J. S. Regulation of the transcriptional activity of the IRF7 promoter by a pathway independent of interferon signaling. *J. Biol. Chem.* **280,** 12262–12270 (2005).
25. Zhao, M., Zhang, J., Phatnani, H., Scheu, S. & Maniatis, T. Stochastic expression of the interferon-β gene. *PLoS Biol.* **10,** e1001249 (2012).
26. Rand, U. *et al.* Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Mol. Syst. Biol.* **8,** 584 (2012).
27. Änkö, M.-L. *et al.* The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* **13,** R17 (2012).
28. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308,** 523–529 (2005).
29. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
30. Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455,** 485–490 (2008).

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

**Acknowledgements** 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 a National Institutes of Health (NIH) Postdoctoral Fellowship (1F32HD075541-01, to R.S.), a Charles H. Hood Foundation Postdoctoral Fellowship (to A. Goren), an NIH grant (U54 AI057159, to N.H.), an NIH New Innovator Award (DP2 OD002230, to N.H.), an NIH CEGS Award (1P50HG006193-01, to H.P., A.R. and N.H.), NIH Pioneer Awards (5DP1OD003893-03 to H.P., DP1OD003958-01 to A.R.), the Broad Institute (to H.P. and A.R.), HHMI (to A.R.), and the Klarman Cell Observatory at the Broad Institute (to A.R.).

**Author Contributions** A.R., H.P., J.Z.L., N.H., A.K.S., R.S., A. Goren and A. Gnirke conceived and designed the study. A.K.S., X.A., R.S.G., J.T.G., R.R., C.M., D.L., J.J.T., D.G. and J.T.G. performed experiments. R.S., A.K.S., S.S. and N.Y. performed computational analyses. R.S., A.K.S., A. Goren, N.H., J.Z.L., H.P. and A.R. wrote the manuscript, with extensive input from all authors.

**Author Information** Data have been deposited in GEO under accession number GSE41265. Reprints and permissions information is available at www.nature.com/reprints. 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.P. (Hongkun_Park@harvard.edu) or A.R. (aregev@broad.mit.edu).