

## Supplemental Discussion

### Temporal heterogeneity in the antiviral response to LPS

To characterize how the variation in the antiviral circuit may change during the response, we performed single-cell quantitative reverse transcription-polymerase chain reaction (qRT-PCR) expression profiling for a signature of 13 genes (nine antiviral cluster genes, two uniformly induced genes, and two housekeeping controls) in unstimulated BMDCs and at 2h, 4h, and 6h post-LPS stimulation (**Supplementary Fig. 13**). We found that the percentage of cells expressing the antiviral cluster genes increased with time (**Supplementary Fig. 13**), and was mirrored by changes in the fraction of cells that exhibit high mRNA levels for antiviral master regulators. In contrast, the uniformly induced genes (*Cxcl10*, *Clec4e*) were robustly induced after two hours in all cells. Importantly, the quantitative correlations between the expression levels of the transcripts that encode master regulators and the downstream targets existed at both the 4h and 6h time points.

### Role of Stat proteins in driving antiviral variability

One possibility is that earlier variation in *Stat2* levels underlies the extensive variation in the anti-viral cluster at 4h, including in the *Stat2* transcript itself (via autoregulation<sup>1</sup>). For example, while the majority of immune response genes (e.g., *Ifit1*) were not expressed in unstimulated cells, the *Stat2* transcript was variably expressed even prior to LPS stimulation (**Supplementary Fig. 13**). Cells with high levels of *Stat2* prior to stimulation may be the ones that express our antiviral cluster at the 4h time point.

To further examine this link, we co-stained cells for Ifit1, Stat1, and Stat2 mRNAs as well as STAT1, pSTAT1, and STAT2 proteins (**Supplementary Methods, Supplementary Fig. 15 and 16**), and quantified these mRNA/protein levels and protein localization in BMDCs simulated with LPS for 0, 2, and 4h. While overall protein levels increased in all cases throughout the time course, we found substantial heterogeneity in the induction of STAT1, pSTAT1, and STAT2 (**Supplementary Fig. 15**). At 2h, all three proteins showed heterogeneity in both their expression and nuclear translocation. By 4h, protein levels were more homogeneous, and nuclear translocation was less pronounced. Ifit1 mRNA distributions displayed highly similar patterns, exhibiting more bimodal expression at early time points and becoming more uniform by 4h. However, STAT protein and Ifit1 mRNA levels within individual cells were not correlated early ( $0.00 < r^2 < 0.12$ ), and only very weakly correlated at 4h ( $0.00 < r^2 < 0.28$ ). This may be due to the fact that a target's mRNA accumulation reflects the integrated temporal activity of a transcriptional regulator, which may not be well represented by a single temporal snapshot<sup>2</sup>. Thus, in cells with high Ifit1 mRNA levels, STAT proteins may already have left the nucleus. Validating such a hypothesis requires real-time tracing of protein and multiple transcripts simultaneously<sup>3</sup>, a task significantly complicated by difficulties of adding endogenous fluorescent tags in primary immune cells<sup>4</sup>, and to the STAT proteins specifically<sup>5</sup>. Conversely, even at 4h, Ifit1 mRNA levels correlated better with Stat1 and Stat2 mRNA levels than their protein levels (**Supplementary Fig. 16**). Since STAT proteins autoregulate their own gene expression<sup>1</sup>, this is consistent with our hypothesis of an earlier regulatory event.

## Materials and Methods

### Mice

C57BL/6 wild-type (wt) were obtained from Jackson Laboratory (Bar Harbor, ME). *Irf7*<sup>-/-</sup> bone marrow was provided by Kate Fitzgerald from University of Massachusetts Medical School. *Ifnr*<sup>-/-6,7</sup> bone marrow was provided by Nir Hacohen from Massachusetts General Hospital (MGH).

All animals were housed and maintained in a conventional pathogen-free facility at the MIT in Cambridge, MA (IUCAC protocol: 0609-058015 (AR)) or at MGH in Boston, MA (IUCAC protocol: 2003N000284 (NH)). All experiments were performed in accordance to the guidelines outlined by the MIT Committee on Animal Care (Cambridge, MA).

### Cell culture, sorting, and lysis

We prepared cultures of bone marrow derived dendritic cells (BMDCs from 6-8 week old female B6 mice as previously described<sup>8,9</sup>. At 9 days of *in vitro* culture, we stimulated the cells with lipopolysaccharide (LPS, Invivogen) as previously described<sup>8,9</sup> for 4h, transferred the cells to a 15 mL conical tube on ice, added 5  $\mu$ M calcein AM and 5  $\mu$ M ethidium homodimer (EthD-1, Invitrogen), and then sorted single calcein-positive, EthD-1-negative cells into individual wells of a 96-well plate, each containing 5  $\mu$ l TCL buffer supplemented with 1% 2-mercaptoethanol (Qiagen, Valencia, CA). After centrifuging, we froze the plates immediately at -80°C. The total time elapsed between removal from

the incubator and lysis was less than 15 minutes. Right before cDNA synthesis, we thawed the cells on ice and purified them with 2.2x RNAClean SPRI beads (Beckman Coulter Genomics, Danvers, MA) without final elution. The beads with captured RNA were air-dried and processed immediately for cDNA synthesis. We also prepared wells with no cells as negative controls and extracted total RNA from ensembles of 10,000 cells as population samples (see below).

### **cDNA synthesis and amplification**

We used the SMARTer Ultra Low RNA Kit (Clontech, Mountain View, CA) to prepare amplified cDNA. We added 1  $\mu$ l of 12  $\mu$ M 3' SMART primer (5'-AAGCAGTGGTATCAACGCAGAGTACT<sub>(30)</sub>N-1N (N = A, C, G, or T; N-1 = A, G, or C)), 1  $\mu$ l of H<sub>2</sub>O, and 2.5  $\mu$ l of Reaction Buffer onto the RNA-capture beads. We mixed them well by pipetting, heated the mixture at 72°C for 3 minutes and placed it on ice. First-strand cDNA was synthesized from this RNA primer mix by adding 2  $\mu$ l of 5x first-strand buffer, 0.25  $\mu$ l of 100mM DTT, 1  $\mu$ l of 10 mM dNTPs, 1  $\mu$ l of 12  $\mu$ M SMARTer II A Oligo (5'-AAGCAGTGGTATCAACGCAGAGTACXXXXX (X = undisclosed base in the proprietary SMARTer oligo sequence)), 100 U SMARTScribe RT, and 10 U RNase Inhibitor in a total volume of 10  $\mu$ l and incubating at 42°C for 90 minutes followed by 10 minutes at 70°C. We purified the first strand cDNA by adding 25  $\mu$ l of room temperature AMPure XP SPRI beads (Beckman Coulter Genomics, Danvers, MA), mixing well by pipetting, incubating at room temperature for 8 minutes. We removed the supernatant from the beads after a good separation was established. We carried out all of the above steps in a PCR product-free clean room. We amplified the cDNA by adding 5

μl of 10x Advantage 2 PCR Buffer, 2 μl of 10 mM dNTPs, 2 μl of 12 μM IS PCR primer (5'- AAGCAGTGGTATCAACGCAGAGT), 2 μl of 50x Advantage 2 Polymerase Mix, and 39 μl H<sub>2</sub>O in a total volume of 50 μl. We performed the PCR at 95°C for 1 minute, followed by 21 cycles of 15 seconds at 95°C, 30 seconds at 65°C and 6 minutes at 68°C, followed by another 10 minutes at 72°C for final extension. We purified the amplified cDNA by adding 90 μl of AMPure XP SPRI beads and washing with 80% ethanol.

For molecule-counting experiments<sup>10</sup> (outlined in **Supplementary Fig. 5 and 6**), the SMARTer II A Oligo was replaced with a custom RNA oligonucleotide containing four random bases (Molecular barcoded SMARTer II A Oligo sequence: 5'- AAGCAGTGGTATCAACGCAGAGTNNNNrGrGrG-3'). Library construction was performed as described above; however, the SPRI purification between reverse transcription and LD-PCR amplification (see above) was omitted for these libraries in order to minimize material loss.

### **cDNA shearing and library construction**

We added the purification buffer (Clontech) to the amplified cDNA to make a total volume of 76 μl. We sheared the cDNA in a 100 μl tube with 10% Duty Cycle, 5% Intensity and 200 Cycles/Burst for 5 minutes in the frequency sweeping mode (Covaris S2 machine, Woburn, MA). We purified the sheared cDNA with 2.2 volumes AMPure XP SPRI beads.

We prepared indexed paired-end libraries for Illumina sequencing as described<sup>11</sup>, with the following modifications. First, we used a different indexing adaptor (containing an 8-base barcode) for each library. Second, we size-selected the ligation product by using two rounds of 0.7 volume of AMPure XP SPRI bead cleanup with the first round starting volume at 100 µl. Third, we performed PCR with Phusion High-Fidelity DNA polymerase with GC buffer and 2 M betaine. Fourth, we used 55°C as the annealing temperature in PCR with the universal indexing primers (forward primer 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC, reverse primer 5'-CAAGCAGAAGACGGCATACGAGAT). Fifth, we performed 12 cycles of PCR. Sixth, we removed PCR primers using two rounds of 1.0 volume of AMPure beads.

### **Population controls and negative controls**

For positive (population) controls, we isolated 13.8 ng of total RNA, as measured by BioAnalyzer (Agilent, Santa Clara, CA), from 10,000 cells using PrepEase RNA Spin Kit (Affymetrix, Santa Clara, CA). We used 1 ng of total RNA in the above processes except that only 12 cycles were used in the cDNA amplification step. For negative controls, we carried out all of the above processes starting with zero sorted cells in TCL-buffer-containing wells. We used 18 cycles in the final PCR of Illumina library construction.

### **Read trimming and mapping**

During reverse transcription, the SMART polymerase adds short (SMARTer II A Oligo) and long (SMART primer oligo) adapters to the beginning of the second read for fragments originating from the 5' and 3' ends of the transcript, respectively. Before

mapping reads, we removed these adapter sequences using Btrim64 with command line arguments -l 1 -e 100 -v 1 -b 28 -a -100. We trimmed adapter sequences from approximately a third of the second reads. Trimmed reads were mapped to the mm9 version of the mouse genome using Tophat v1.4.1<sup>12</sup> with default parameters. Genome mappings were used to visualize data in the Integrative Genome Viewer<sup>13</sup> and to compute a set of library quality metrics, as described below.

Reads where the short adapter (5' end) was trimmed mapped at approximately equal rates to untrimmed reads. However, read pairs where the long adapter (3' end) was trimmed often contained polyA stretches even after trimming, and mapped at extremely low rates (<1%). Since these reads should originate from the 3' end of the transcript, this low mapping percentage results in a depletion of reads from the 3' end of the transcript. This depletion may cancel out the 3' coverage bias that is a byproduct of the SMART protocol (see below)<sup>14</sup>.

### **Quantifying unique mRNA molecules**

When processing the three single-cell libraries where the SMARTer oligo was modified to include a four-nucleotide random barcode sequence, reads containing the SMARTer II A Oligo were isolated and trimmed as described above. Four additional bases (corresponding to the barcode) were then trimmed and maintained for later processing. Trimmed reads were mapped to the mouse mm9 genome as described above. For each gene, we then identified the subset of these reads that mapped to the first 1,500 bases of exonic sequence on the correct strand, and retrieved their original four-nucleotide

barcodes. We counted the unique number of barcodes for each gene and used this count as an alternative quantification of single-cell gene expression. Both unique molecular barcode counts and TPM estimates are provided for all three cells in **Supplementary Table 7**.

### Library quality metrics

Library quality metrics, including genomic mapping rates, coefficients of variation of coverage of each transcript, the fraction of ribosomal RNA in each library, and positional coverage biases, were calculated using PicardTools version 1.42 (picard.sourceforge.net). We observed less 3' bias in our data than previously reported<sup>14</sup>, likely due to the differences in library construction noted above (**Supplementary Fig. 19**).

### Expression level calculation

We created a Bowtie<sup>15</sup> index based on the UCSC knownGene<sup>16</sup> transcriptome, and aligned paired-end reads directly to this index using Bowtie v 0.12.7 with command line options -q --phred33-quals -n 2 -e 99999999 -l 25 -I 1 -X 1000 -a -m 200. Next, we ran RSEM v1.11<sup>17</sup> with default parameters on these alignments to estimate expression levels. RSEM's gene level expression estimates (tau) were multiplied by 1,000,000 to obtain transcript per million (TPM) estimates for each gene. To transform expression levels to log-space, we took the  $\ln(\text{TPM}+1)$ . When calculating the “average” single-cell expression level, we first averaged TPM levels from each of the 18 single cells, and then transformed this average estimate into log space.



We applied identical procedures to a previously published dataset<sup>1</sup>, consisting of an RNA-Seq time course after LPS stimulation of BMDCs. We used this dataset to identify a set of 632 genes that were induced at least two-fold in the population at 4h following LPS stimulation as compared to pre-stimulation. These genes were analyzed in **Fig. 2a**, **Fig. 2d**, and **Fig. 4b**.

### RNA fluorescence *in situ* hybridization (FISH)

We measured the expression levels for 25 different mRNA transcripts *in situ* using RNA-FISH probes (Panomics). Briefly, BMDCs were sorted on Cd11c (Miltenyi Biotech) at 8 days *in vitro* and plated on poly-*l*-lysine coated glass coverslips. The following morning, some cells were stimulated with LPS as previously described<sup>8,9</sup>. Ten minutes prior to fixation, cell culture media was replaced with a 1:500 dilution of Alexa-350 wheat germ agglutinin (WGA, Invitrogen) in HBSS. Subsequently, cells were fixed and stained according to the manufacturer's recommendations. After curing overnight, Slowfade (Invitrogen) mounted coverslips were raster scanned at 60x magnification (1.42 NA, oil immersion) in x, y, and z using an epifluorescence microscope (Olympus) outfitted with Metamorph software. On average, 100 individual 3-dimensional stacks were taken for each sample. For all samples, four-color imaging was performed to obtain the following information: excitation (ex) 405nm – WGA & DAPI stains; ex 488nm, ex 546nm, ex 647nm – Probes 1, 2, and 3, respectively.

The obtained images were processed in two phases. First, CellProfiler<sup>18</sup> was used to determine cell numbers and locations for each stack of images taken using the UV filter

set (ex 405nm). Brightly stained nuclear regions (DAPI) were used to identify individual nuclei and were then used as seeds for determining the extents of each cell from the duller membrane outlines (WGA). The locations and extents of individual cells were then extracted for each imaging position using the software. Next, for each color channel, individual mRNAs were identified and counted in Matlab using a previously described analysis package<sup>19</sup>. Identified mRNAs were then allotted to individual cells using the output of CellProfiler. Final analysis and plotting was also performed using Matlab. The displayed RNA-FISH images were false-colored and overlaid using Adobe Photoshop.

For all RNA FISH histograms, counts were binned ( $n = 50$ ) and smoothed with a window of 5 bins in Matlab. As controls, BMDCs that were not stimulated with LPS were also analyzed to ensure the specificity of our induced-gene RNA-FISH probes (**Supplementary Fig. 20**).

For splicing analyses, custom RNA fish probes (Panomics) were designed to either Irf7 or Acpp as follows:

#	Accession	Target	Start	Stop	Length	Approx. bDNAs	Name and color in Fig. 3c and Supplementary Fig. 7 and 8
1	NM_019807.2	Acpp	1199	2667		20	Exon A (Orange, O)
2	NM_207668.2	Acpp	1199	4488		20	Exon B (Magenta, M)
3	NM_016850	Irf7	891	992	101	3	Isoform Specific (Orange, O)
4	NM_016850	Irf7	-	-		20	Constitutive B (Cyan, C)
5	NM_016850	Irf7	1461			3	Constitutive A (Magenta, M)

The difference in the number of bDNAs between the two constitutive Irf7 probes led to slightly better binding and thus higher counts for the constitutive probe B. As a result, the

metric, probe A counts/(probe A counts + probe B counts) (used in the histogram in **Fig. 3c**), is normally distributed with a mean of ~0.45 (instead of ~0.5). Plotting the isoform-specific probe over constitutive probe B gave a similar curve (compare **Fig. 3c** with **Supplementary Fig. 7**). A cell was only included if the number of counted mRNAs for the constitutive probe (Irf7) was at least 5 or if the sum of alternative exon counts was at least 5. For Acpp (**Supplementary Fig. 8**),  $n = 615$  cells; for Irf7,  $n = 490$  cells.

### Immunofluorescence (IF) measurements

IF co-staining was performed as previously described<sup>4,9</sup> directly after RNA-FISH staining. STAT1, pSTAT1, and STAT2 antibodies, all used at 1:200, were obtained from Santa Cruz Biotechnology. Average and total fluorescence levels, as well as the percentage of the fluorescence localized to the nucleus, were quantified from epifluorescence images using locations and extents of individual cells and their nuclei, as above (**Supplementary Fig. 15 and 16**). For all protein histograms, counts were binned ( $n = 100$ ) and smoothed with a window of 5 bins in Matlab. Single-plane and 3-dimensional scans yielded similar results (data not shown).

### Single-cell qRT-PCR

Single BMDCs were prepared for qRT-PCR using the Single-Cell-to-Ct kit (Ambion) with minor modifications. Namely, individual BMDCs were sorted into one-fourth of the recommended lysis buffer volume and all subsequent steps were scaled to match. After specific target amplification, an exonuclease I digestion (NEB) was performed by adding 0.5  $\mu$ L Exonuclease I, 0.25  $\mu$ L Exonuclease I Reaction Buffer, and 1.75  $\mu$ L water to each

sample, vortexing, centrifuging, and heating to 37°C for 30 minutes. After an 80°C heat inactivation for 15 minute, samples were diluted 1:5 in Buffer TE. Single cells, negative controls, and population controls (prepared equivalently using extracted total RNA) were analyzed using 96x96 gene expression chips (Fluidigm Biomark)<sup>20</sup>.

### Coefficient of variation calculation

We calculated the coefficient of variation (CV, normalized standard deviation) as the ratio of the standard deviation of gene expression values (log space) across single cells and the average single cell expression level (log space, see above). The dashed grey lines in **Fig. 2a,b** represent a constant CV of 0.25, and broadly separate highly expressed genes into two groups of variable and non-variable genes (**Supplementary Table 3**). Functional enrichment analysis of these two gene sets (see below) was highly robust to small changes in the CV threshold (between 0.2 and 0.3) that we used.

The dashed blue line in **Fig. 2a** represents the maximum theoretical standard deviation for the 18 single cells given their single cell average. This theoretical maximum occurs when the cells are perfectly bimodally distributed about a value of  $(\mu + \log(2))/2$  and is represented by the relationship:  $\sigma_{\max} = \sqrt{18/17} * (\mu + \log(2))/2$ .

### Functional enrichment of variable/non-variable gene sets

Functional enrichment (GO annotation) of non-variable highly expressed gene sets was performed using DAVID v6.7<sup>21,22</sup>. The full list of 522 highly expressed genes was used as the background set. For **Fig. 2a** and **Supplementary Table 3** we combined two lists to

form a set of housekeeping genes. The first list is a set of ribosomal subunit proteins defined in GO annotations (GO term 0003735: structural constituent of ribosome)<sup>21,22</sup> and the second list is taken from a table of commonly used murine housekeeping genes that were downloaded directly from Qiagen: (<http://b2b.qiagen.com/products/pcr/quantitect/housekeepinggenes.aspx>). All genes included in this list are provided in **Supplementary Table 2**.

### **Correlation matrix and principal component analysis (PCA)**

PCA for 632 induced genes was performed in R using the `prcomp` function. We transformed the expression values of each gene to have zero mean and unit variance across single cells in order to appropriately compare variability patterns across genes with different overall abundance in the population.

We calculated a correlation matrix based on the log-scale (but non-transformed) gene-expression estimates, and clustered the matrix using k-means. We chose a parameter of five clusters based on the “elbow method”<sup>23</sup> (data not shown), but our identification of a strongly enriched antiviral cluster (and its high degree of overlap with PC2) was highly robust to the parameter choice or stochasticity of k-means.

In our set of 632 genes, we annotated a set of antiviral genes targets from previous work<sup>8</sup>. We annotated STAT2 targets from a previously defined set of “promoter ChIP peaks”<sup>1</sup> on a set of identically stimulated (at 4h) BMDCs. Cluster-specific enrichment analyses were

performed using a hypergeometric test in R, using the full set of 632 induced genes as a background set.

### **Population fluorescence-activated cell-sorting (FACS) analysis and qRT-PCR**

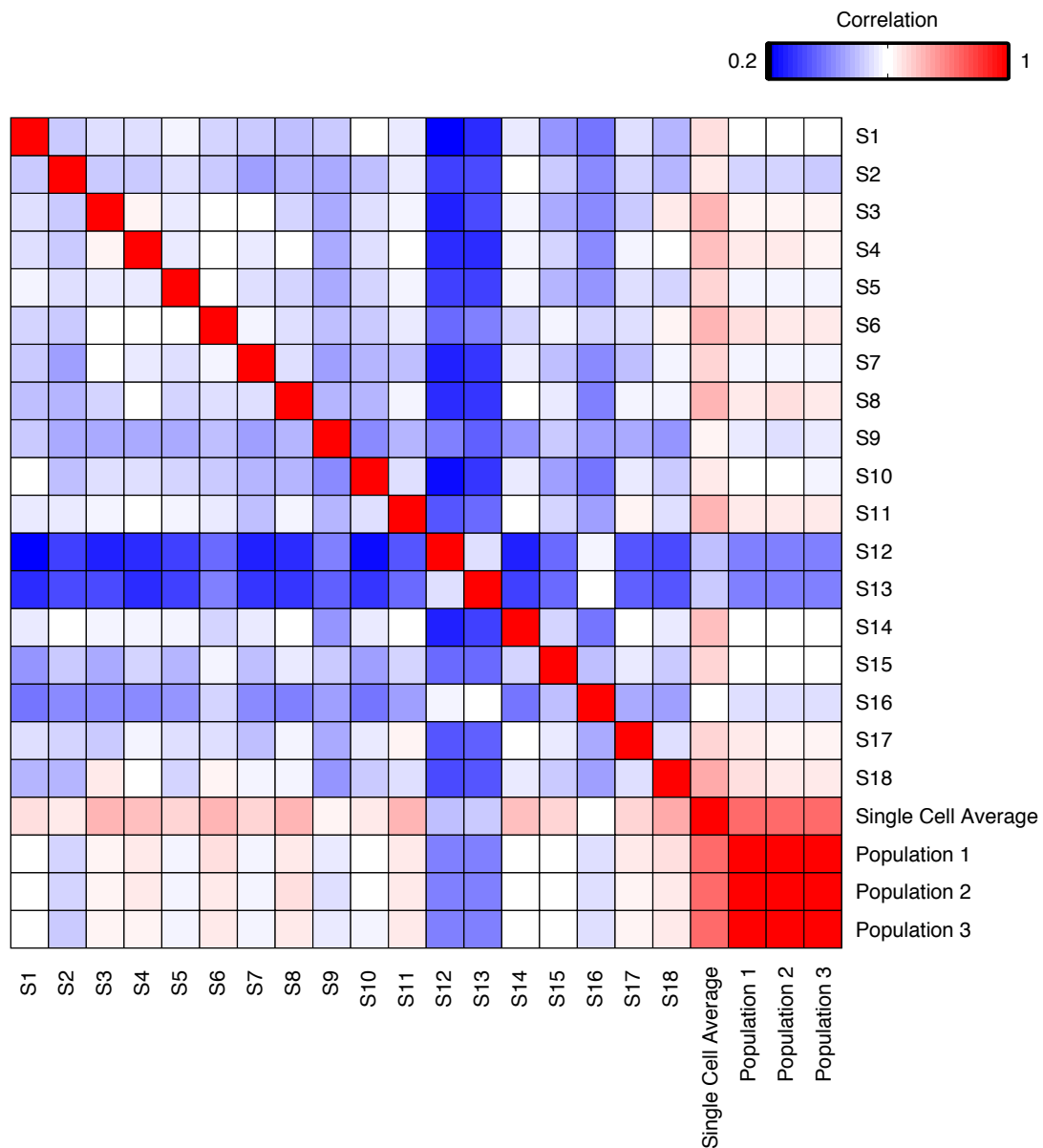
BMDCs were stimulated with LPS for 4h. Fifteen minutes prior to sorting, cells were stained with each of 11 antibodies from Biolegend that defined our mature (1) or maturing (2) cells: Cd83 (1), Cd273 (1), Ccr7 (1), Cd40 (1), Cd201 (1), Cd137 (1), Cd68 (2), Cd120b (2), Cd53 (2), Cd88 (2), and Cd16/32 (2). Three groups of 1,000 cells either positive or negative for each of the tested surface markers were sorted in 100  $\mu$ L of buffer TCL supplemented with 1% 2-mercaptoethanol. Total RNA was then extracted from each of the 20 samples using an RNeasy Mini Kit (Qiagen) and cDNA was prepared using Sensiscript RT (Qiagen) as previously described<sup>24</sup>. Population-wide expression levels for different transcripts were then analyzed relative to GAPDH using qRT-PCR, as previously described<sup>24</sup> (**Supplementary Fig. 12**). Primers for qRT-PCR are presented in **Supplementary Table 6**.

### **Splicing analysis**

We downloaded a set of ~67,000 previously annotated alternatively spliced events (skipped exons, mutually exclusive splice events)<sup>25</sup>. We ran MISO<sup>26</sup> with default parameters to estimate the percent spliced in (PSI) for every event in each of our single cells and population replicates. The vast majority of events were not expressed at sufficient depth in any of our samples to be analyzed by MISO. For the remaining 4,338 events, we noted that PSI estimates derived from 10,000 cell replicates were tightly

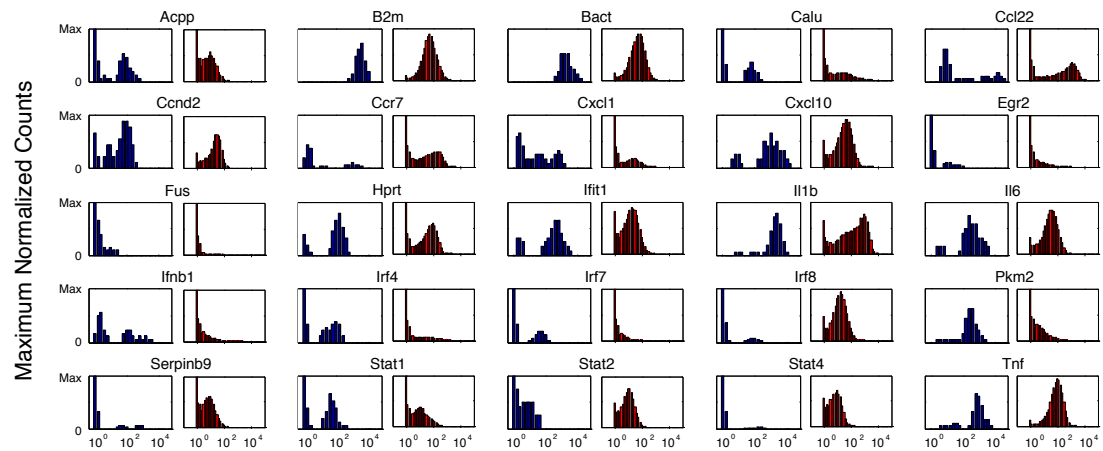
correlated (mean  $r = 0.91$ ). We averaged the PSI values for the three population replicates and focused the remainder of our analyses on the 352 “alternatively spliced” events ( $20\% < \text{population PSI average} < 80\%$ ) in 322 genes (28 genes had at least two alternative splicing events).

We then examined the PSI distribution of these 352 alternative splicing events across single cells (**Fig. 3b**). To ensure that we only examined reliable splicing events from highly expressed transcripts, we only considered PSI estimates for single cell/splice event pairs where the alternatively spliced gene was expressed at high levels (single-cell TPM  $> 250$ ) within that single cell. This resulted in 89 unique alternative splice events from 79 genes. After applying this filter, we plotted a histogram of PSI estimates across single cells (**Fig. 3b**, top). **Fig. 3b** (bottom) shows a histogram of PSI estimates from the first 10,000-cell replicate for the same 89 splice events from **Fig. 3b** (top).



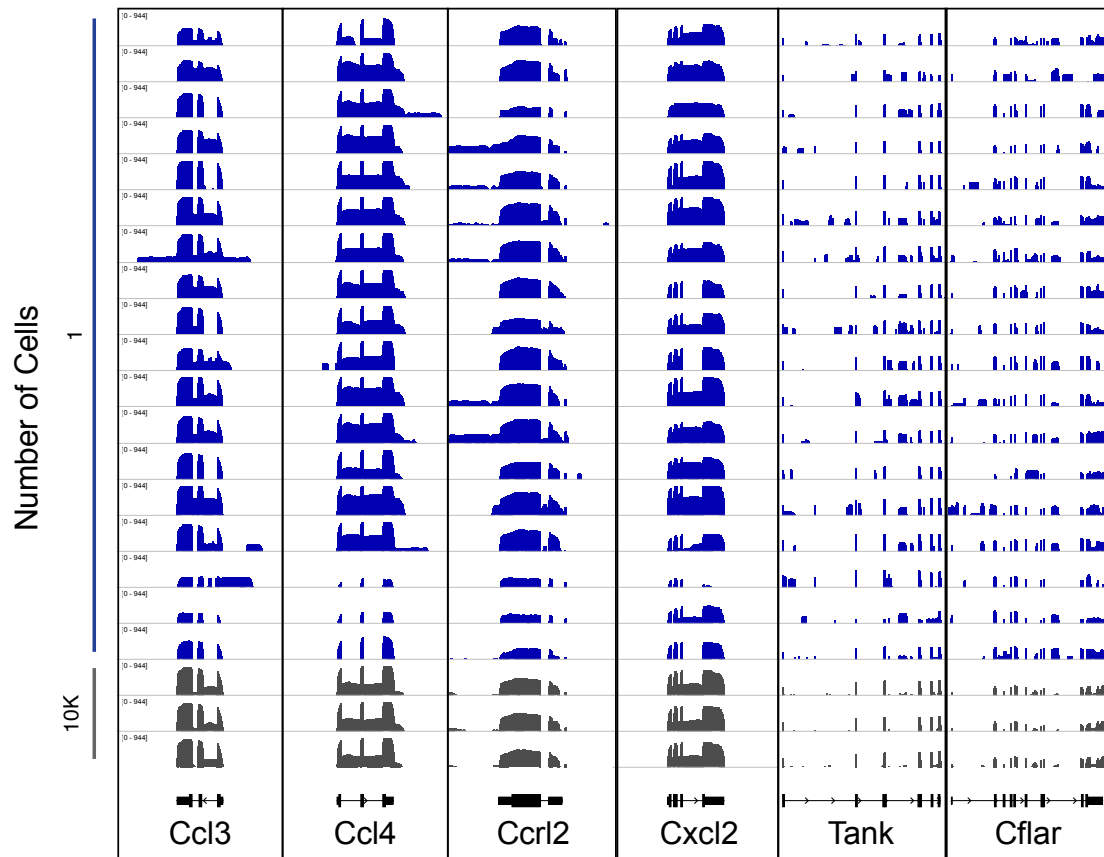
**Supplementary Fig. 1 | Global correlations in mRNA expression between single LPS stimulated BMDCs.** Shown are the Pearson correlation coefficients between global expression profiles of each of 18 individual cells, the single cell average, and three populations of 10,000 cells each (rows, columns). All correlations were computed on log-scale expression profiles. Single cells (S) 12, 13, and 16 were mature prior to stimulation, while single cells 9 and 15 are the most mature of the cells which are undergoing pathogen-dependent maturation, correspond to light grey triangles in **Fig. 4a**.



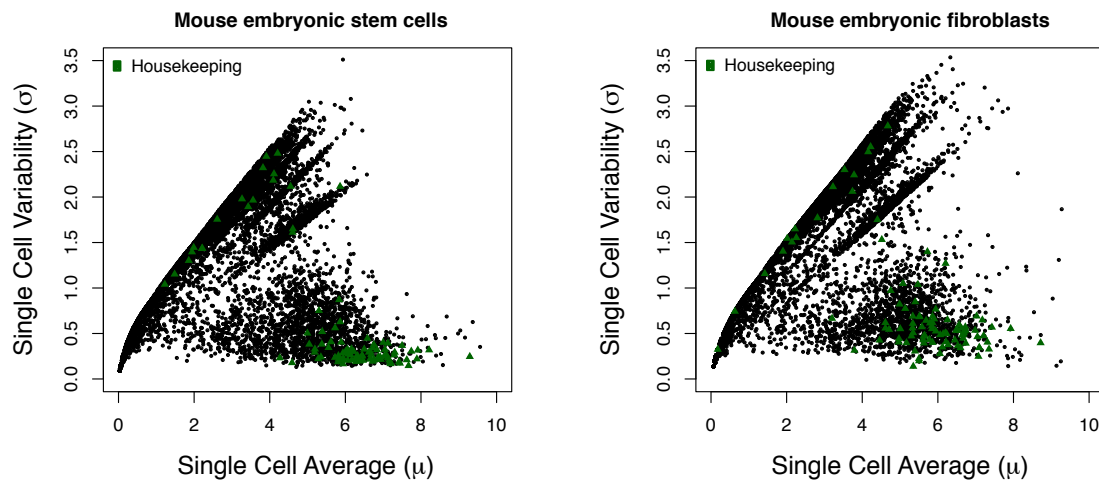


Single Cell RNA-Seq (Blue, x-axis: Log(TPM)) & Single Cell RNA-FISH (Red, x-axis: log(mRNA Count))

**Supplementary Fig. 2 | Agreement between single-cell RNA-Seq and RNA-FISH for 25 different transcripts.** Shown are the distributions of gene expression levels for each of 25 transcripts in single-cell RNA-Seq of 18 cells (left, blue) and in single-cell RNA-FISH of, on average, 1600 cells (right, red). All histograms are normalized to their maximums.



**Supplementary Fig. 3 | Robust LPS response across all cells.** Shown are tracks of RNA-Seq reads from the Integrative Genomics Viewer (IGV) for the levels of key response genes (columns, gene name at bottom) in each single cell (blue) and the population average (grey). The genes include key chemokines and chemokine receptors (Ccl3, Ccl4, Ccl2), cytokines (Cxcl2), and other important components of the LPS response (Tank, Cflar).

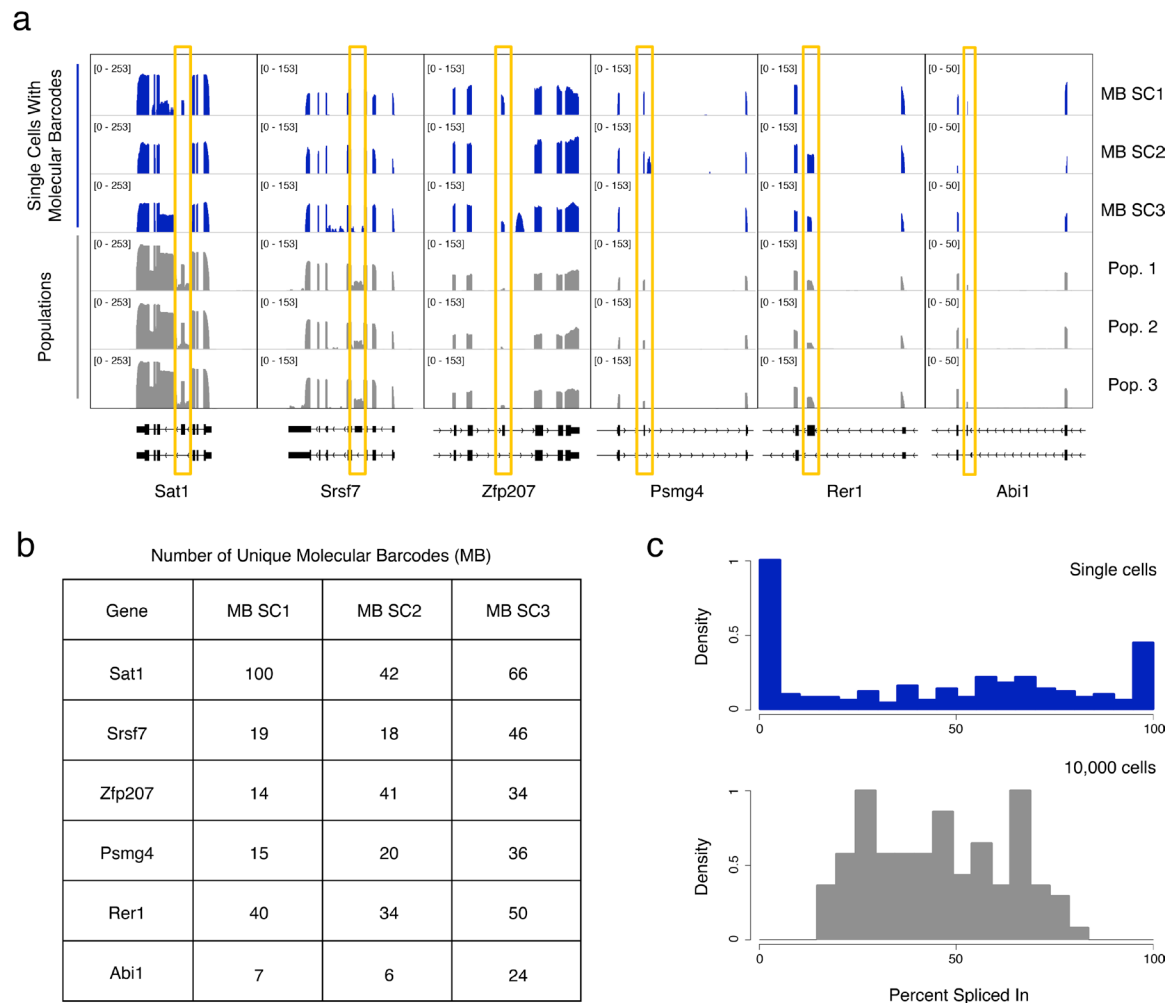


**Supplementary Fig. 4 | Variation in gene expression from single-cell RNA-Seq in other cell types.** Shown is the relationship between the single-cell expression average ( $\mu$ , X axis) and single-cell variability (standard deviation,  $\sigma$ , Y axis) in mouse embryonic stem cells (left) and mouse embryonic fibroblasts (right). These figures show a re-analysis of previously published single-cell RNA-Seq data<sup>23</sup>. Housekeeping genes are green. In both cases we found substantially less variability in single-cell gene expression compared to LPS-stimulated BMDCs (**Fig. 2a**).

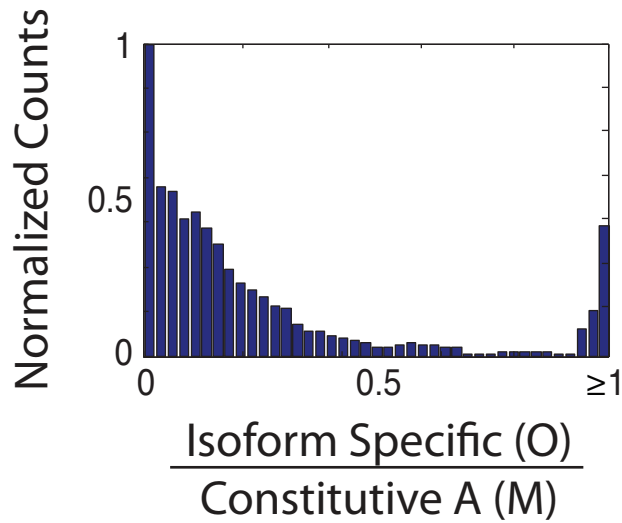


**Supplementary Fig. 5 | Quantification of unique mRNA molecules in three single cells. a**, Modified protocol. We modified the SMARTer II A oligo, introducing a random four nucleotide barcode onto each mRNA molecule during reverse transcription. Shown

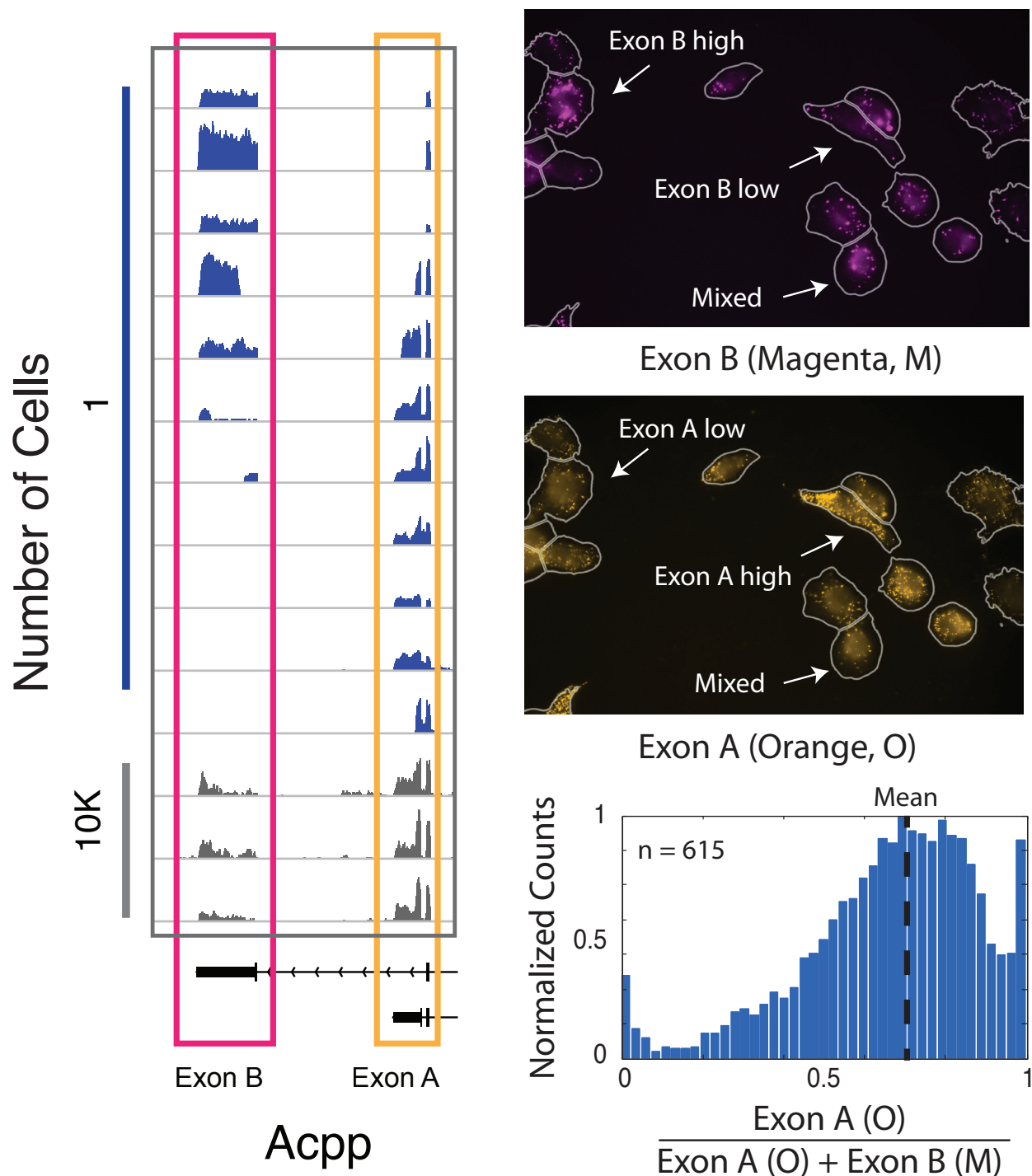
is the structure of modified oligo (barcode is represented by NNNN). This barcode is retained through PCR amplification and library preparation. **b**, IGV screenshot showing read densities at one locus for the three barcoded single-cell cDNA libraries (blue) and the three 10,000 cell replicate experiments (grey). Two single cells express exclusively one of two isoforms. **c**, Detailed examination of reads mapping to 5' end of transcript. The 242 reads represent 38 unique barcodes, affirming that the observed splicing result is not simply due to stochastic amplification of just one or a few molecules. **d**, Relationship between single-cell TPM (X axis, log scale) and uniquely identified barcodes (Y axis, log scale) for the three barcoded single-cell libraries. Only genes represented by at least one unique barcode are plotted. Light blue shaded area represents single-cell TPM < 250, the threshold used throughout our study. The two alternate quantifications of single-cell gene expression are well correlated overall ( $0.84 < r < 0.88$ ) with a tight linear relationship for highly expressed genes (TPM > 250). Extremely highly expressed genes (TPM > 3,000) likely represented more unique molecules than the number of available unique barcodes (256), resulting in saturation.



**Supplementary Fig. 6 | Variation in isoform expression between single cells based on the 3 barcoded single-cell libraries. a**, IGV screenshots showing read densities for 6 alternatively spliced genes. For each gene, the alternatively spliced exon is boxed in orange. **b**, Table showing the number of unique molecular barcodes (rows) counted for each transcript shown in (a) in each molecular barcoded single cell (MB SC, columns). **c**, Shown are the distributions of exon inclusion (PSI scores, X axis) for alternatively spliced exons in genes represented by at least 15 barcodes in single cells (blue histogram, top) and in the populations (grey histogram, bottom). Results are highly similar to our splicing analysis of highly expressed genes across the 18 cells (single-cell TPM > 250; Fig. 3). Single cells exhibit a strong skew towards one isoform or the other.

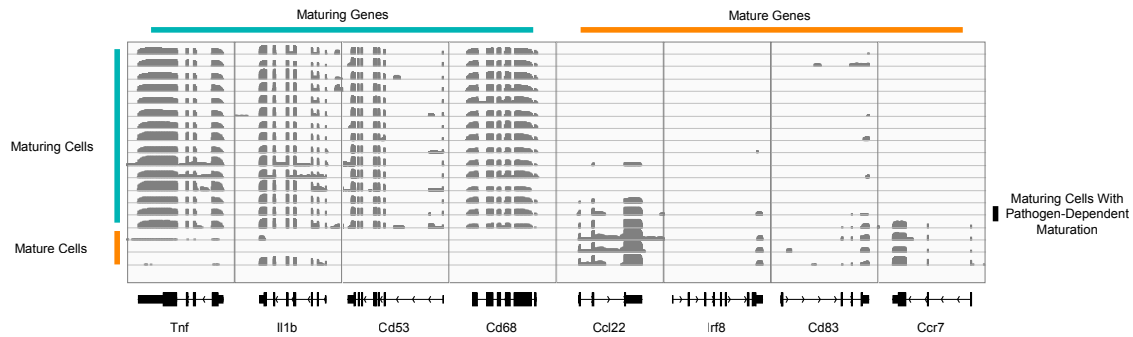


**Supplementary Fig. 7 | RNA-FISH validation of splicing variation in Irf7 in single cells.** Shown is the distribution across cells of the ratio of Irf7 transcripts displaying the isoform-specific Irf7 probe (Orange, **Fig. 3c**) relative to the shorter constitutive probe (Magenta, **Fig. 3c**). The distribution is similarly bimodal to that obtained when calculating the ratio of the specific probe to the longer constitutive probe (**Fig. 3c**).

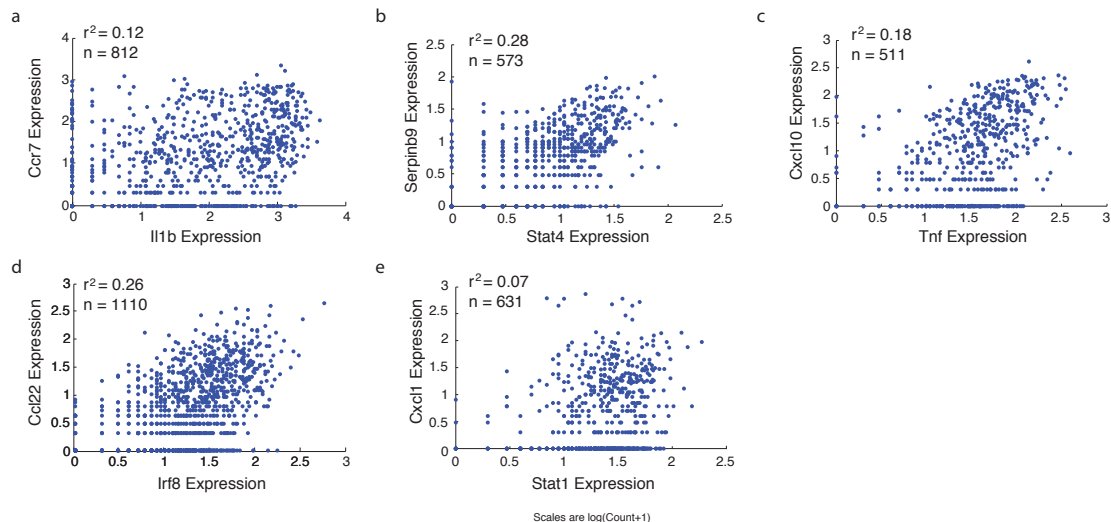


**Supplementary Fig. 8 | Confirmation of variation in Acpp isoform expression between single cells using RNA-FISH.** Left: RNA-Seq read densities for Acpp (only cells where the transcript is expressed are shown). Color boxes mark mutually-exclusive final exons analyzed by RNA-FISH. Right: RNA-FISH images from simultaneous hybridization with probes for the two exons denoted on the left. The distribution (bottom right) only includes the 615 cells for which the total number of Acpp transcripts (sum of exon A and exon B) was greater than 5.





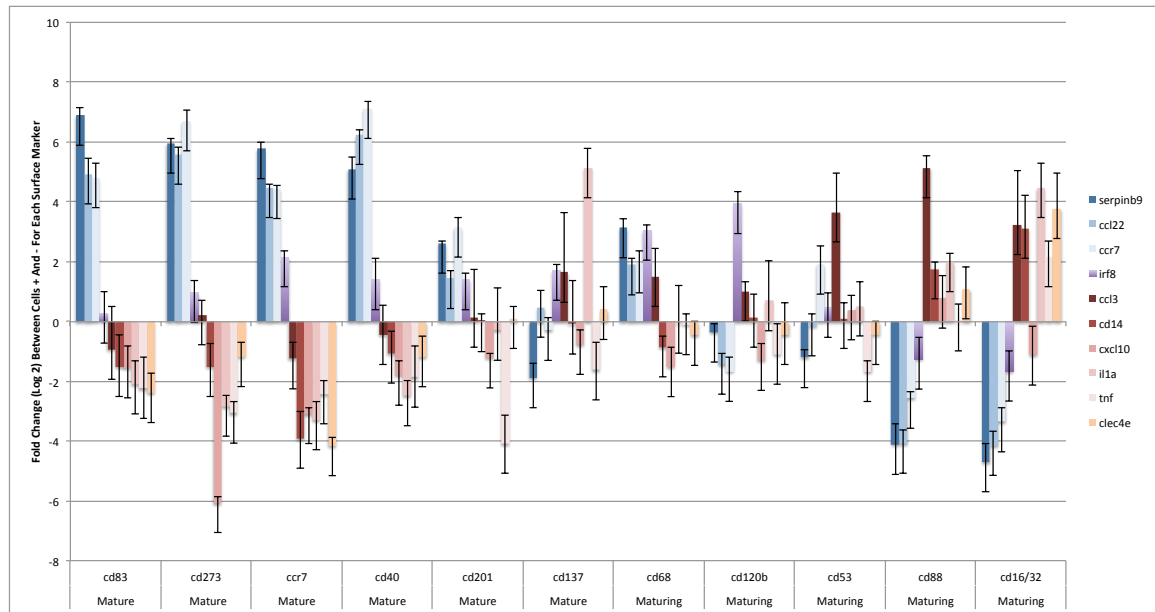
**Supplementary Fig. 9 | IGV screenshots exhibiting the separation between maturing and mature cells.** These genes have either very high (positive) or low (negative) projection scores for PC1. A black vertical bar on the right highlights two cells that express both mature and maturing markers, suggesting that they have begun to undergo pathogen-dependent maturation.



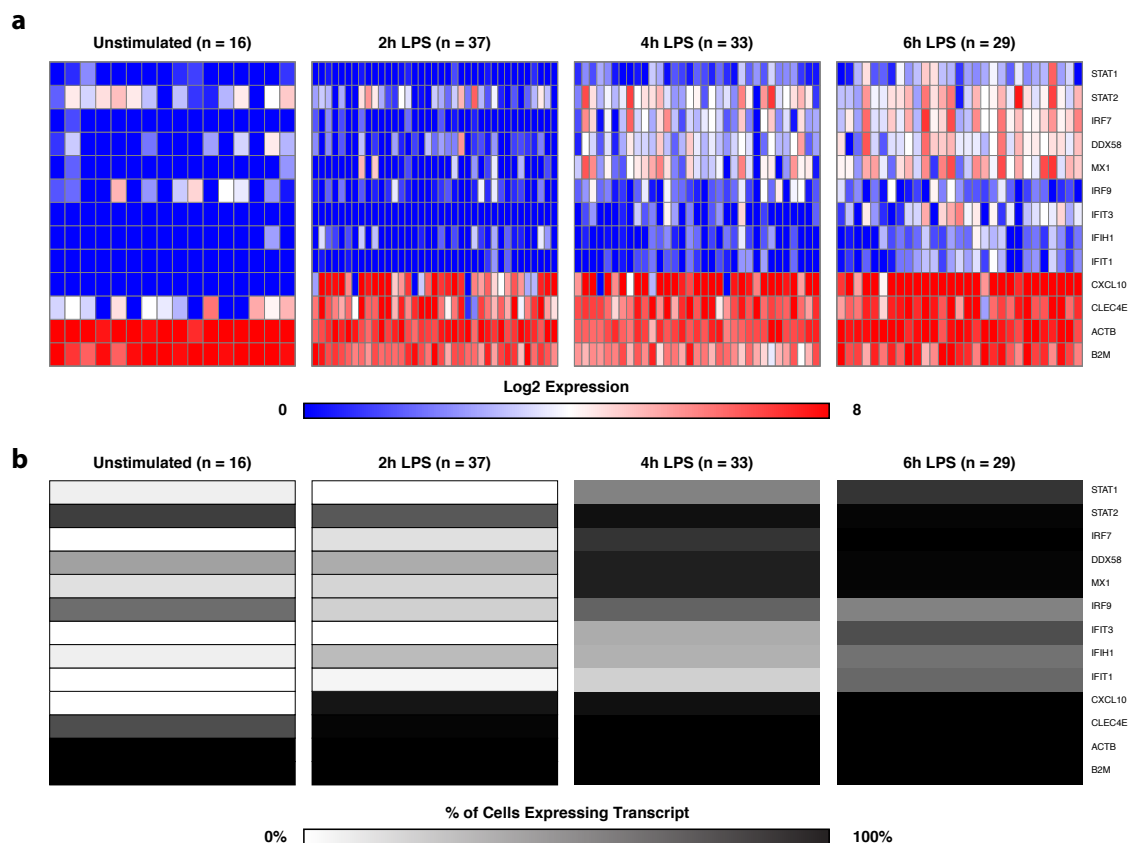
**Supplementary Fig. 10 | Confirmation of co-variation patterns by RNA-FISH.** Shown are the relationships in expression levels (log (Count+1)) for pairs of transcripts simultaneously measured by RNA-FISH. **a**, Expression levels for Ccr7 (expressed more in mature cells) and Il1b (expressed more in maturing cells) do not correlate strongly (Pearson  $r^2 = 0.12$ ,  $n = 812$ ). **b**, Expression levels for Stat4 (expressed more in mature cells) and Serpinb9 (expressed more in mature cells) correlate more strongly (Pearson  $r^2 = 0.28$ ,  $n = 573$ ). **c**, Expression levels for Cxcl10 and Tnf (both expressed more in maturing cells) correlate mildly (Pearson  $r^2 = 0.18$ ,  $n = 511$ ). **d**, Ccl22 and Irf8 (both expressed in mature cells) show moderate correlation (Pearson  $r^2 = 0.26$ ,  $n = 1110$ ). **e**, Stat1 (antiviral, specific to neither) and Cxcl1 (inflammatory, specific to neither) correlate very weakly (Pearson  $r^2 = 0.07$ ,  $n = 631$ ).



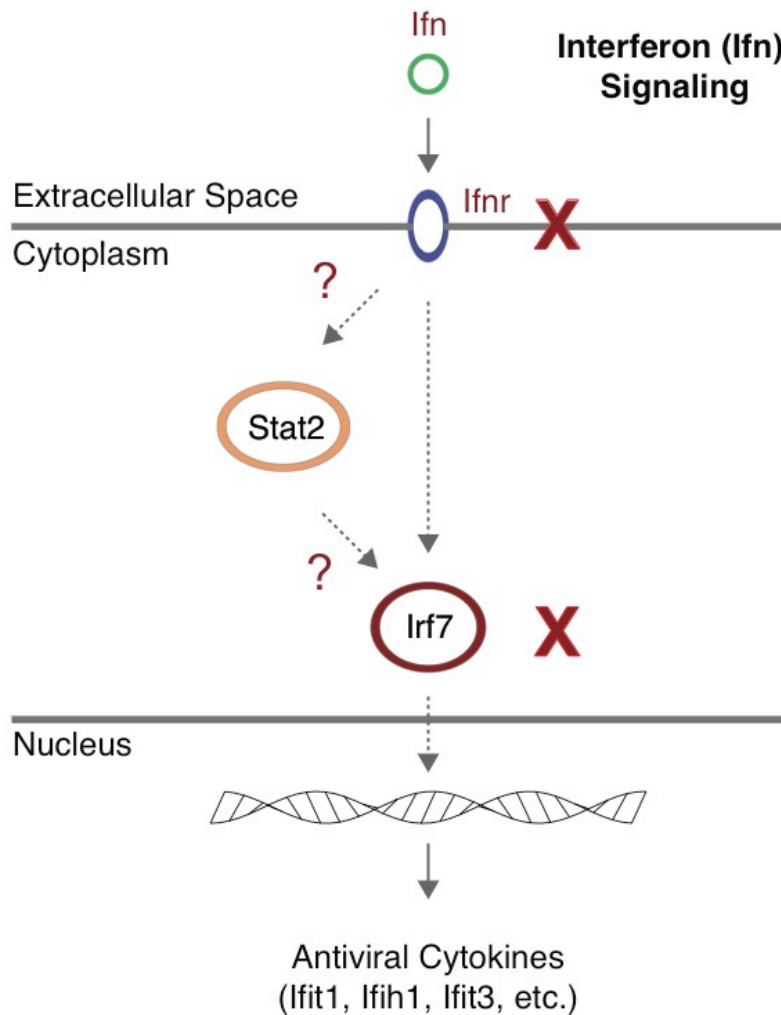
**Supplementary Fig. 11 | Individual LPS-stimulated BMDCs cluster into two distinct populations by single-cell qRT-PCR.** Shown are the normalized expression levels (red: high; blue: low, scale on top) from single-cell qRT-PCR (Fluidigm) for 50 genes (rows) in each of 46 individual cells (columns). The cells are hierarchically clustered based on their expression profiles (dendrogram, top) and form two main clusters (mature and maturing, bottom).



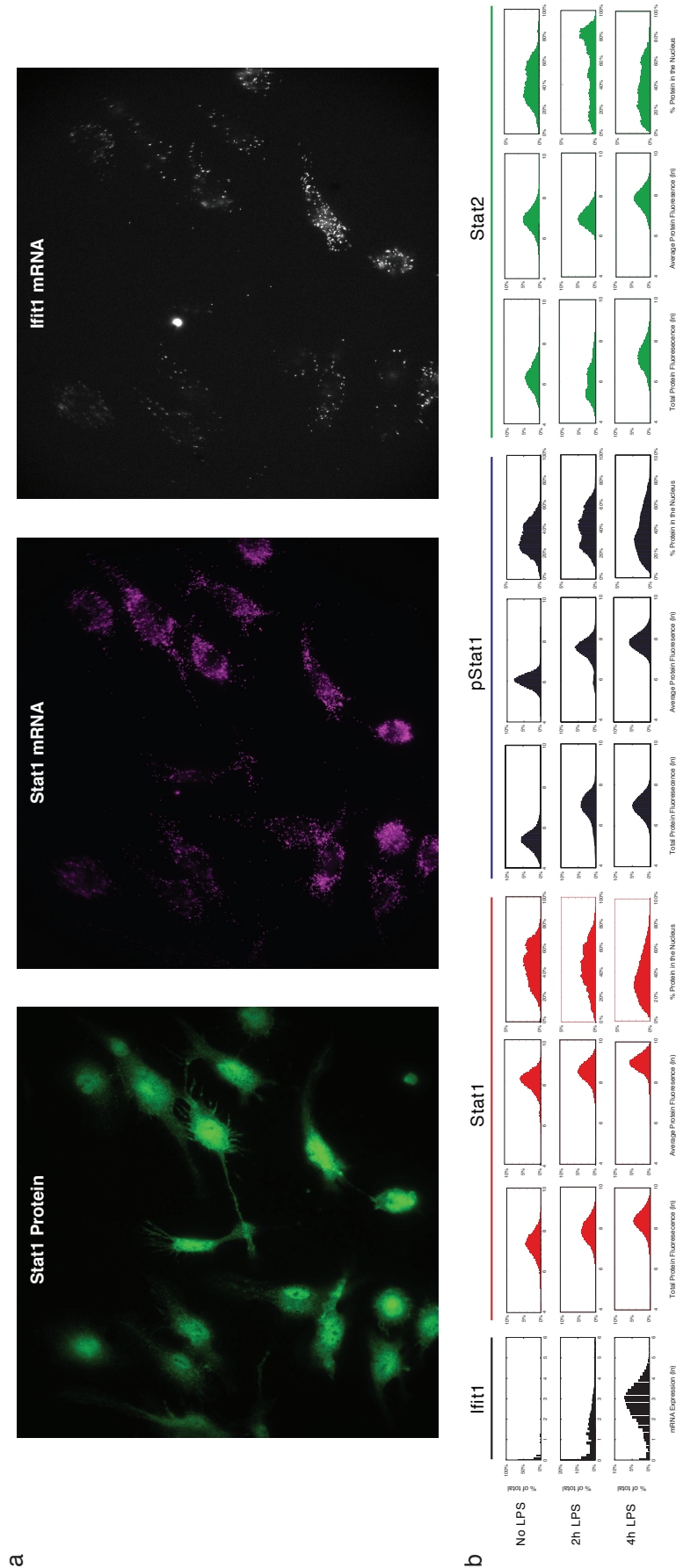
**Supplementary Fig. 12 | RNA-Seq identifies genes which differentiate BMDC maturity states.** We used our PCA analysis (PC1 scores) to select eleven genes, coding for cell surface markers, whose RNA levels best discriminated between our mature and maturing subpopulations. For each marker, we flow sorted cells into positive and negative subpopulations, and profiled a gene signature (blue bars: genes highly expressed in ‘mature’ cells; red: genes highly expressed in ‘maturing’ cells) in each with qRT-PCR. Shown are the differential expression levels (Y axis) of each of 10 signature genes (bars, color legend, right) for each cell surface marker (X axis).



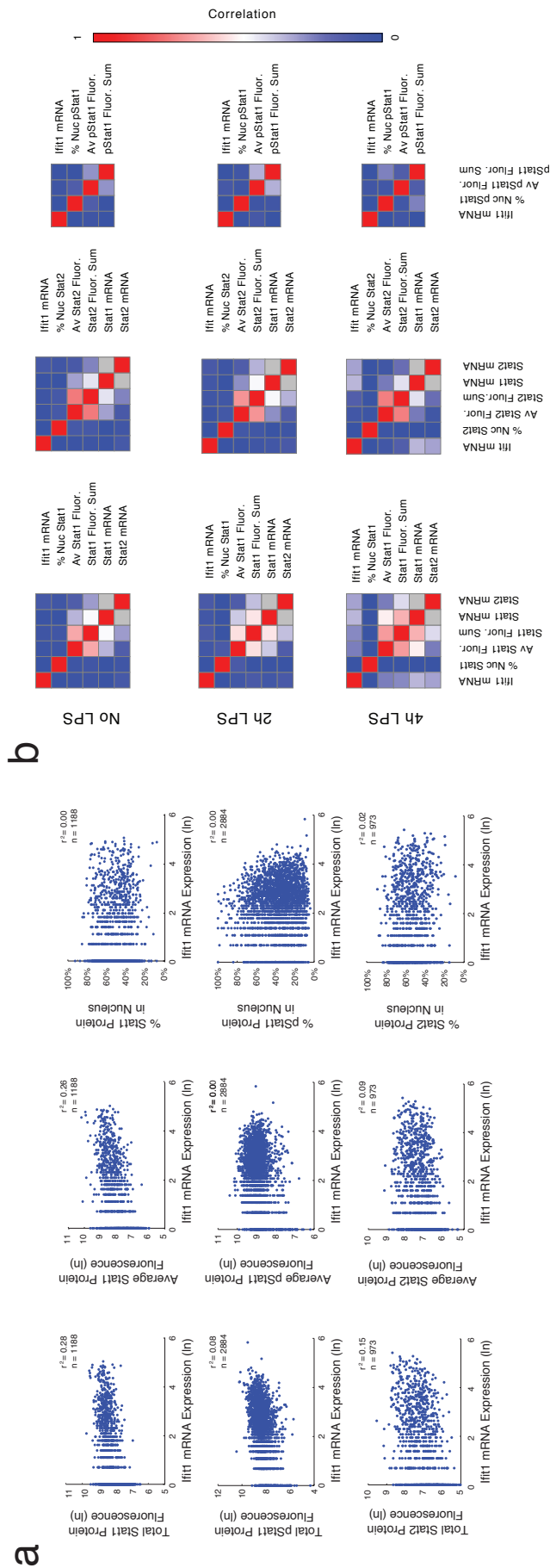
**Supplementary Fig. 13 | Single-cell qPCR expression profiling for a signature of 13 genes along an LPS response time course.** **a**, Shown are the expression levels of each gene (row) in each cell (column) in unstimulated BMDCs and at 2h, 4h, and 6h post-LPS stimulation. The gene signature consists of nine antiviral cluster genes, two uniformly induced genes, and two housekeeping controls. **b**, The percentage of cells that express each gene (rows) at each time point (column). A cell was scored as positive for a gene if the gene's expression was higher than a Ct of 23 on the Fluidigm BioMark. While some immune response genes, *Cxcl10* and *Clec4e*, were uniformly induced in all cells and persisted across time points, the percentage of cells expressing the antiviral cluster genes increased in a time-dependent manner.



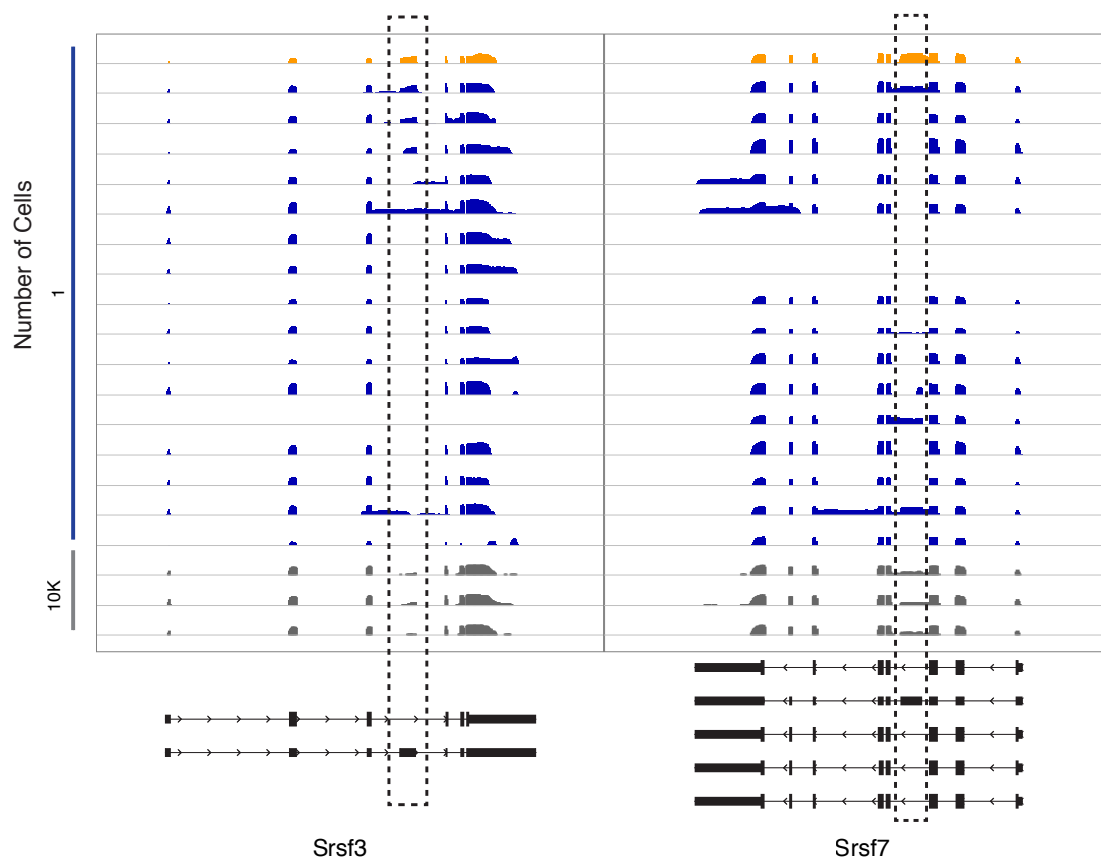
**Supplementary Fig. 14 | A simple model for the identified antiviral circuit.** X's represent points of perturbation. Interferon (Ifn) feedback drives expression of Irf7 and Stat2. Variability in the expression of Irf7 propagates to variability in the expression of antiviral genes, such as Ifit1. Stat2 is implicated as well, though its relation to Irf7 cannot be established by the current experiments.



**Supplementary Fig. 15 | RNA (RNA-FISH) and protein (immunofluorescence) co-staining. a**, An example of a co-staining image for Stat1 protein (green), Stat1 mRNA (magenta), and Ifit1 mRNA (white). **b**, Distributions of the levels of Ifit1 mRNA (black) and STAT1 (red), pSTAT1 (grey), and STAT2 (green) proteins (total fluorescence level, left histogram; average fluorescence level, middle; and percent nuclear localization, right) after exposure to LPS for 0 (top), 2 (middle) or 4 (bottom) hours. While overall protein levels increased in all cases throughout the time course, there is substantial variation in the induction of STAT1, pSTAT1, and STAT2. STAT1 levels rose gradually while pSTAT1's shifts were most pronounced early. STAT2, meanwhile, showed strong nuclear localization by 2h, followed by strong induction from 2 to 4h. By 4hr, protein levels were more homogeneous and nuclear translocation was less pronounced.

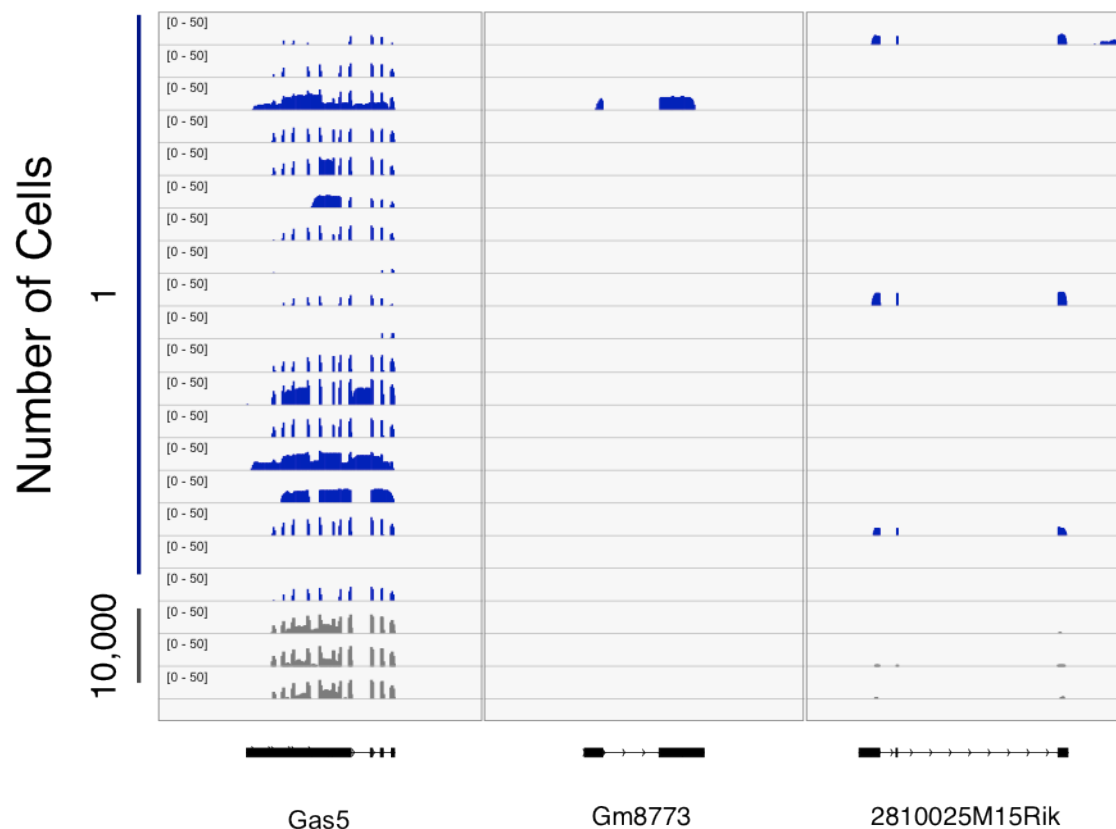


**Supplementary Fig. 16 | Correlation between STAT protein and Ifit1 mRNA expression.** **a**, Representative scatter plots showing the correlation between STAT proteins (Y axis) and Ifit mRNA levels (X axis) after a 4h LPS stimulation. Top row: STAT1, middle row: pSTAT1; bottom row: STAT2. Left column: total protein fluorescence; middle column: average protein fluorescence; right column: percent of nuclear protein. **b**, Heatmaps showing the correlation ( $r^2$ ; blue = 0; red = 1) between different measured parameters after exposure to LPS for 0 (top), 2 (middle), or 4 hours (bottom).

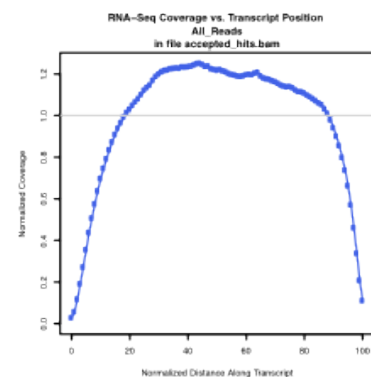
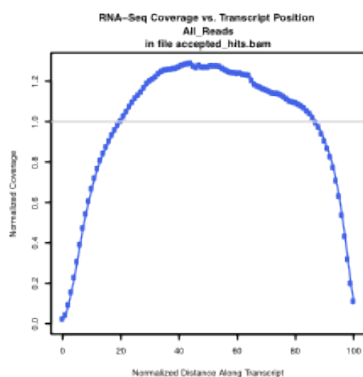
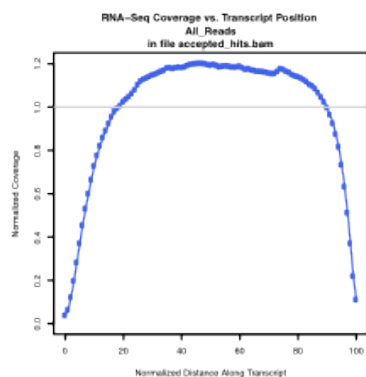
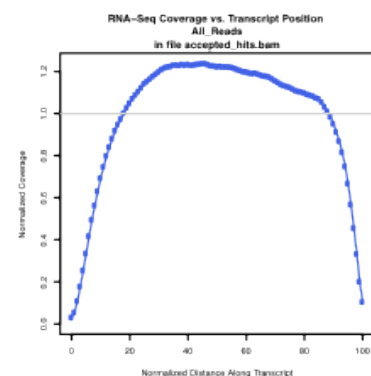
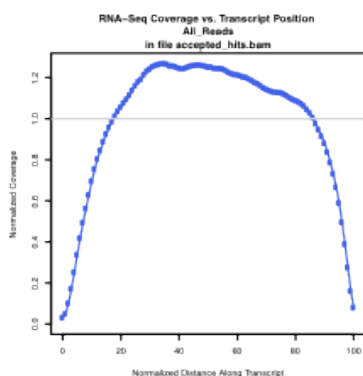
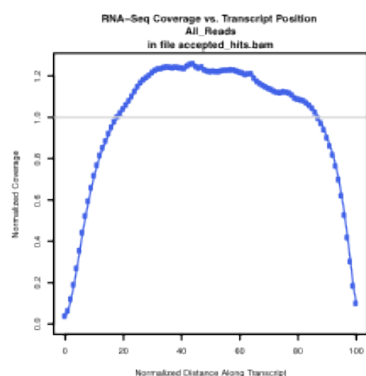
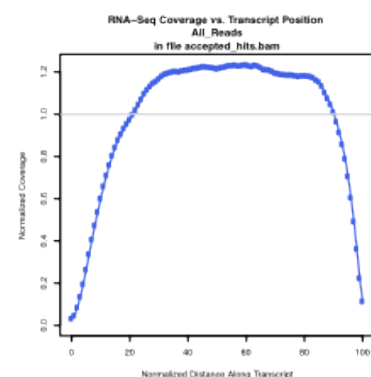
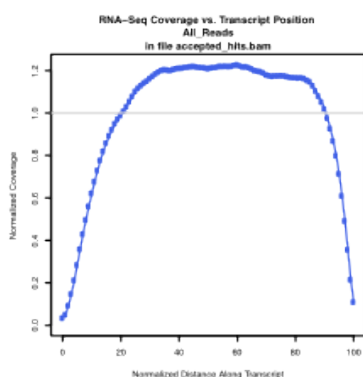
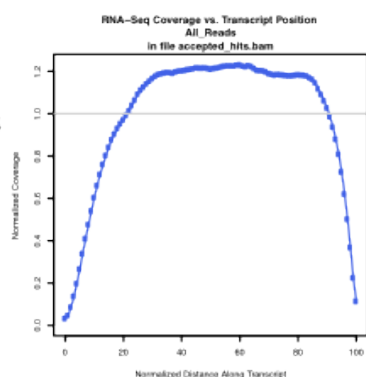


**Supplementary Fig. 17 | Splicing patterns for ‘poison’ cassette exons of the splicing factors Srsf3 and Srsf7.** Shown are the RNA-Seq read densities in each individual cell (‘1’, blue) and the population average (‘10,000’, grey) for two genes encoding the splicing factors Srsf3 and Srsf7, each of which is known to have an alternatively spliced poison cassette exon (dashed box). The known annotated isoforms for each gene is shown at the bottom. One cell, S13, highlighted in orange at the top, expressed only the Srsf3 and Srsf7 isoforms that contain the ‘poisonous’ exons. For each gene, 11 cells exclusively expressed the alternative isoform.

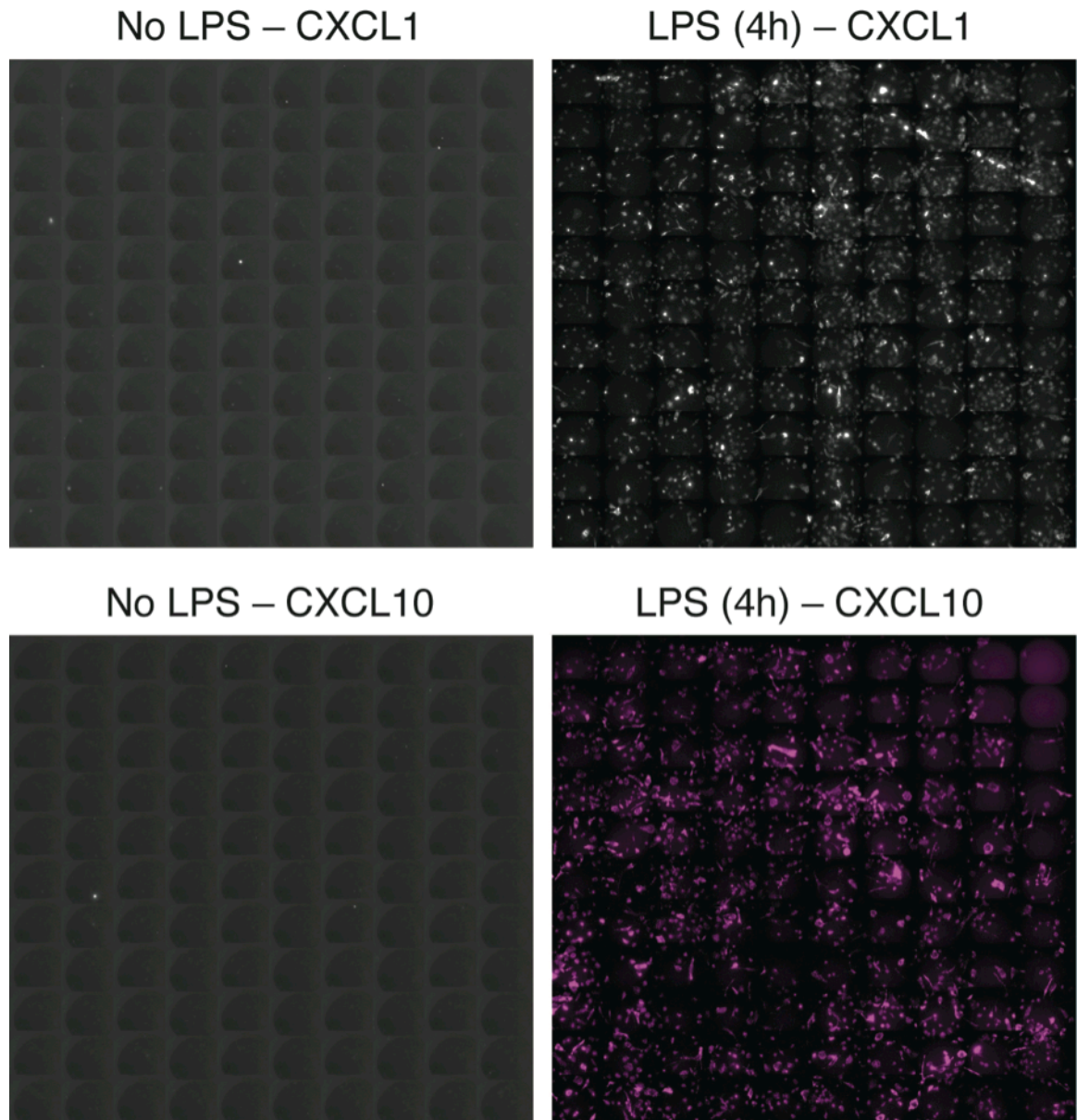




**Supplementary Fig. 18 | Expression variation in long intergenic non-coding (linc) RNAs.** Shown are the RNA-Seq read densities in each individual cell ('1', blue) and the population average ('10,000', grey) for three previously annotated lincRNA genes. A lincRNA relatively highly expressed at the population level (Gas5, left), is bimodally expressed at the single-cell level. Two lincRNAs lowly expressed or undetectable at the population (Gm8773, 2810025M15Rik) can be detected in some individual cells.

Single Cells,  
1-3Single Cells  
4-610,000 replicates  
1-3

**Supplementary Fig. 19 | Quality control for 3' bias.** Shown are plots of normalized RNA-Seq coverage at each normalized transcript position from 5' (left) to 3' (right) for 6 single cells (top two rows) and all three 10,000 populations (bottom row). Both the single cells and the populations show little 3' bias.



**Supplementary Fig. 20 | RNA-FISH of the immune-response genes *Cxcl1* (Top) and *Cxcl10* (Bottom) in the absence of LPS stimulation (left) and after 4h of an LPS stimulus (right).** *Cxcl10* and *Cxcl1*, although expressed at negligible levels prior to stimulation, are strongly induced by LPS.

## References

- 1 Garber, M. *et al.* A High-Throughput Chromatin Immunoprecipitation Approach Reveals Principles of Dynamic Gene Regulation in Mammals. *Molecular Cell* **47**, 810-822, doi:10.1016/j.molcel.2012.07.030 (2012).
- 2 Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485-490, doi:papers2://publication/doi/10.1038/nature07292 (2008).
- 3 Cohen, A. A. *et al.* Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science* **322**, 1511-1516, doi:10.1126/science.1160165 (2008).
- 4 Shalek, A. K. *et al.* Nanowire-Mediated Delivery Enables Functional Interrogation of Primary Immune Cells: Application to the Analysis of Chronic Lymphocytic Leukemia. *Nano Lett.* **12**, 6498-6504, doi:10.1021/nl3042917 (2012).
- 5 Meyer, T., Begitt, A. & Vinkemeier, U. Green fluorescent protein-tagging reduces the nucleocytoplasmic shuttling specifically of unphosphorylated STAT1. *The FEBS journal* **274**, 815-826, doi:10.1111/j.1742-4658.2006.05626.x (2007).
- 6 Darnell, J. E., Jr., Kerr, I. M. & Stark, G. R. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science (New York, N.Y.)* **264**, 1415-1421 (1994).
- 7 Gough, D. J. *et al.* Functional crosstalk between type I and II interferon through the regulated expression of STAT1. *PLoS biology* **8**, e1000361-e1000361 (2010).
- 8 Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257-263, doi:10.1126/science.1179050 (2009).
- 9 Chevrier, N. *et al.* Systematic Discovery of TLR Signaling Components Delineates Viral-Sensing Circuits. *Cell* **147**, 853-867, doi:10.1016/j.cell.2011.10.022 (2011).
- 10 Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* **9**, 72-74, doi:10.1038/nmeth.1778 (2011).
- 11 Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods* **7**, 709-715 (2010).
- 12 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
- 13 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
- 14 Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**, 777-782, doi:papers2://publication/doi/10.1038/nbt.2282 (2012).
- 15 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, doi:10.1186/gb-2009-10-3-r25 (2009).
- 16 Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Research*, doi:10.1093/nar/gkq963 (2010).
- 17 Li, B. & Dewey, C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, doi:10.1186/1471-2105-12-323 (2011).
- 18 Carpenter, A. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* **7**, doi:10.1186/gb-2006-7-10-r100 (2006).
- 19 Raj, A., Van Den Bogaard, P., Rifkin, S. A., Van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**, 877-879, doi:10.1038/nmeth.1253 (2008).
- 20 Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature Biotechnology* **29**, 1120-1127, doi:10.1038/nbt.2038 (2011).

- 21 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
- 22 Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
- 23 Diday, E. *New approaches in classification and data analysis*. (Springer-Verlag, 1994).
- 24 Shalek, A. K. *et al.* Vertical silicon nanowires as a universal platform for delivering biomolecules into living cells. *Proceedings of the National Academy of Sciences* **107**, 1870-1875, doi:10.1073/pnas.0909350107 (2010).
- 25 Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476, doi:10.1038/nature07509 (2008).
- 26 Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**, 1009-1015, doi:10.1038/nmeth.1528 (2010).