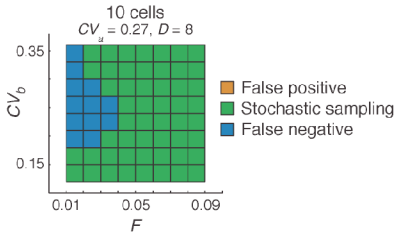
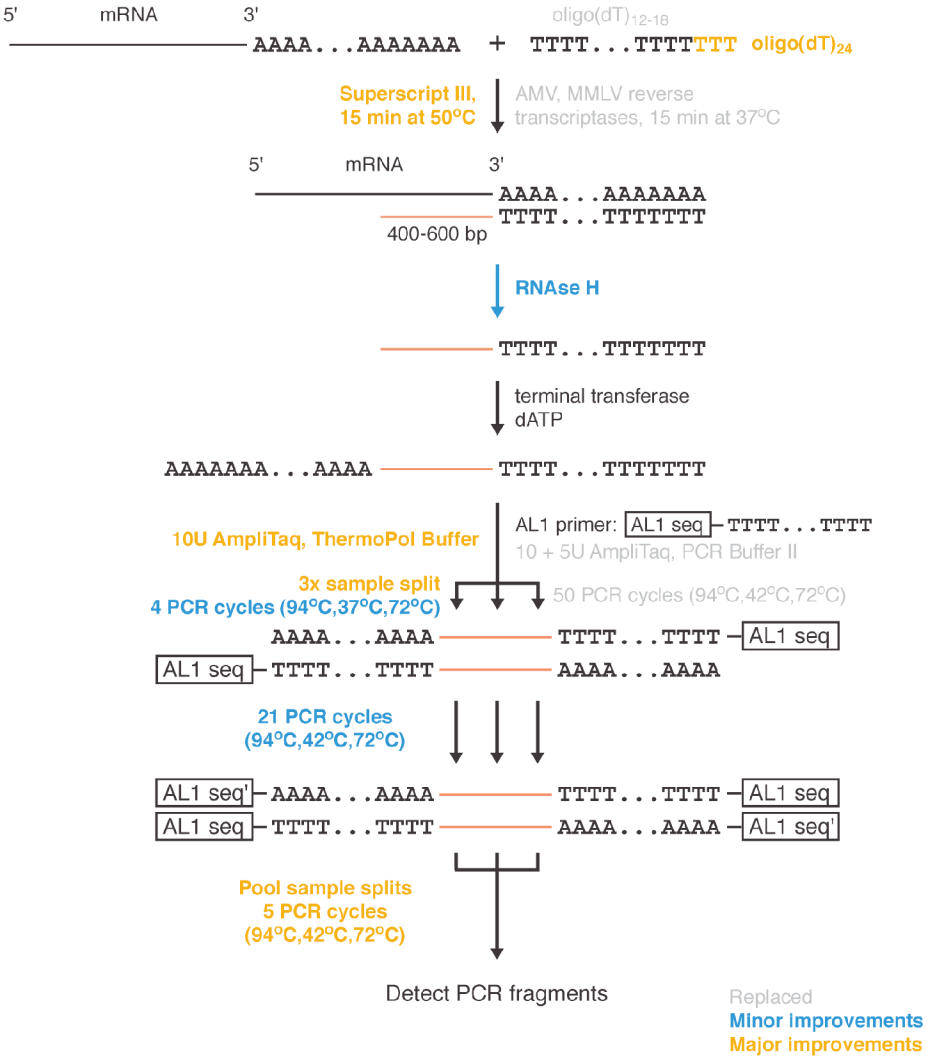
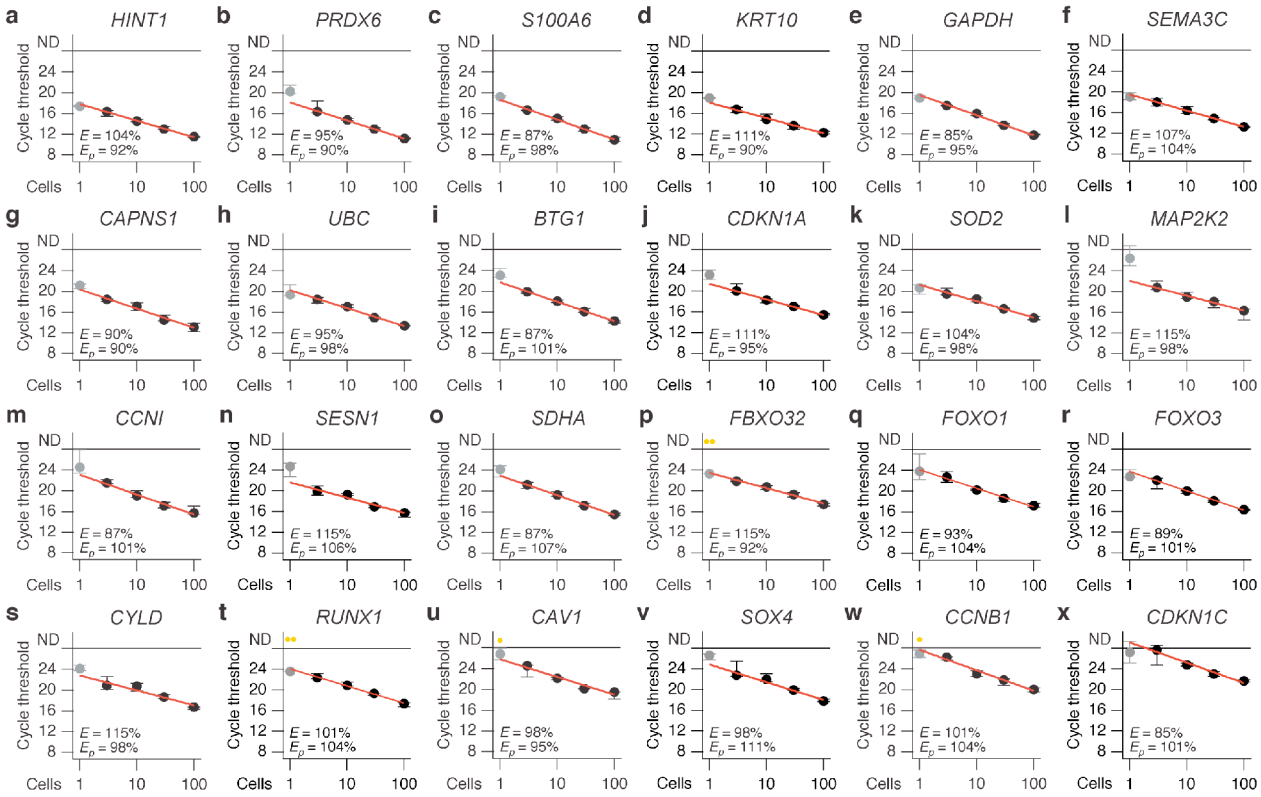


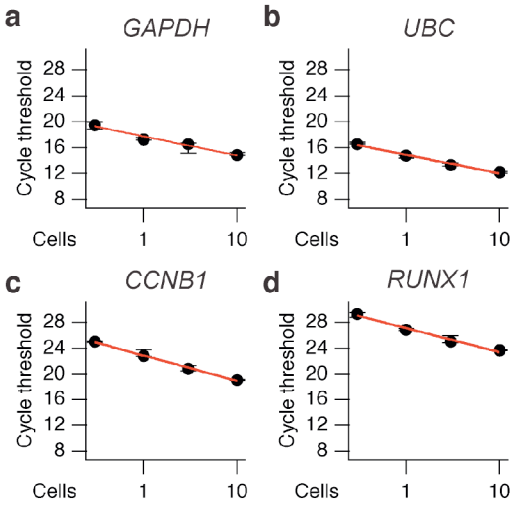
Supplementary Figure 1. | Stochastic profiling is prone to false negatives when heterogeneous subpopulations are rare.
Simulations were performed as described in **Figure 1g** for 10 cells averaged with CV_a = 27%, CV_b = 12–36%, D = 8, and F = 0.01–0.09. Note the increase in false negatives when F < 0.05, corresponding to heterogeneous expression in one out of 20 cells.



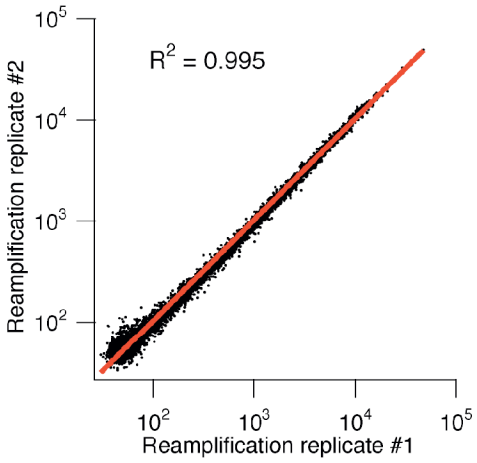
Supplementary Figure 2. | Small-cell poly(A) PCR.
Polyadenylated mRNA extracted from cells is primed with oligo(dT)$_{24}$ and an abbreviated reverse transcription is performed using a reverse transcriptase with high thermal stability. RNA bound to first-strand cDNA is degraded with RNAse H, and cDNAs are tailed with poly(A). Poly(A)-tailed cDNA is amplified by PCR with a T_{24}-containing universal primer (AL1, Ref. [1]) and the indicated cycling temperatures. For the first 25 cycles, each amplification is split into three aliquots and then pooled to improve reproducibility (see the Experimental Procedures). Modifications of the original protocol[1] that caused major (orange) and minor (blue) improvements in precision or accuracy are highlighted.



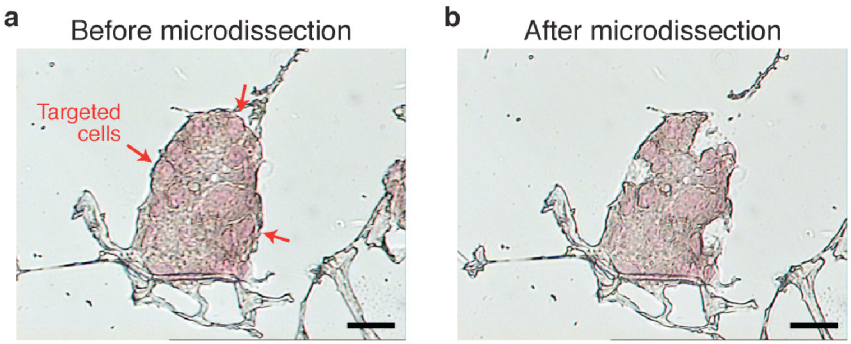
Supplementary Figure 3. | Quantitative small-cell amplification of high- to low-abundance transcripts from 3–100 cells.
(a-x) 100-cell samples were serially diluted, amplified by small-cell poly(A) PCR (**Supplementary Fig. 2**), and quantified by RT-qPCR as described in the Online Methods. The RT-qPCR cycle threshold for each gene is plotted as a function of starting cellular material and is shown as the median ± range of three replicate small-cell amplifications. Amplification efficiencies (E) based on a log-linear fit of the 3–100-cell dilutions (red line) are listed along with primer efficiencies (E_p) calculated by serially diluting the template before RT-qPCR. Genes are ordered **a** through **x** in the order of increasing median cycle threshold from the 10-cell replicates, which was used as an approximation of relative abundance (lower cycle thresholds suggest increased relative abundance). Note that the one-cell amplifications (gray) of higher-abundance transcripts (**b,i,j,l,n**) often deviate from the log-linear fit, and the one-cell amplification of lower-abundance transcripts (**p,t,u,w**) are frequently not detectable (yellow, ND).



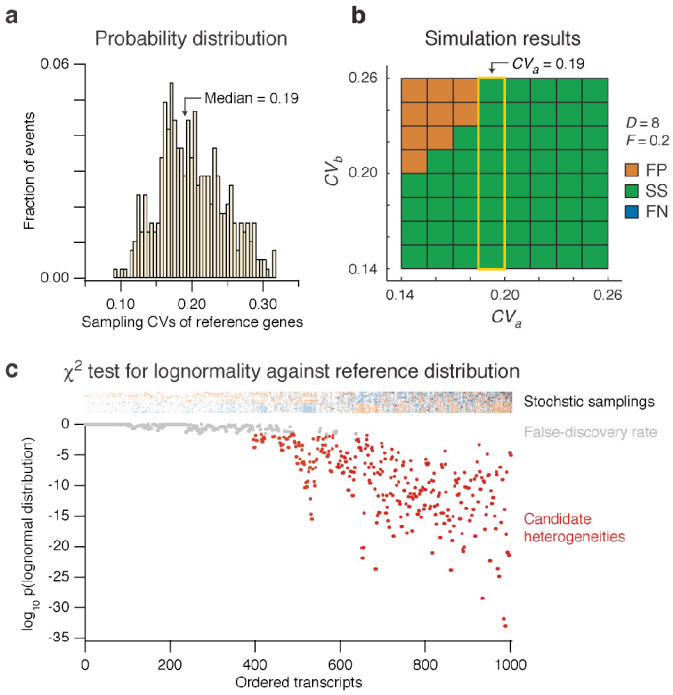
Supplementary Figure 4. | Limited reamplification does not affect the quantitative accuracy of small-cell PCR.
(a-d) A 10-cell equivalent of amplified cDNA from the procedure described in **Supplemental Figure 2** was serially diluted to 3, 1, and 0.3 cell equivalents and reamplified as described in the **Online Methods**. Reamplified material was then used as the template for RT-qPCR for the high-abundance genes **(a)** *GAPDH* and **(b)** *UBC* and the low-abundance genes **(c)** *CCNB1* and **(d)** *RUNX1*.



Supplementary Figure 5. | Reproducible reamplification, labeling, and microarray hybridization following small-cell PCR. Two 10-cell equivalents of the same starting small-cell amplification were reamplified, labeled, and hybridized to Illumina HumanRef-8 microarrays as described in the **Online Methods**.

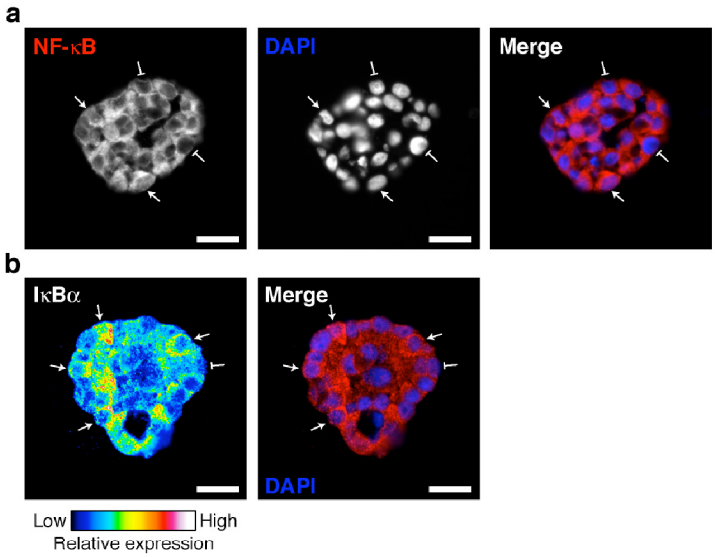


Supplementary Figure 6. | Accurate laser-capture microdissection of single matrix-attached cells in cryosections of MCF10A 3D structures.
(**a,b**) 8 μm thick frozen sections of MCF10A-5E acini were stained with nuclear fast red and prepared for laser-capture microdissection as described in the **Online Methods**. Phase-contrast image of a single MCF10A-5E structure is shown (**a**) before and (**b**) after microdissection of three matrix-attached cells (arrows). Note that the microdissection is effective and that there is minimal collateral pickup of material from adjacent cells. Scale bar is 20 μm.

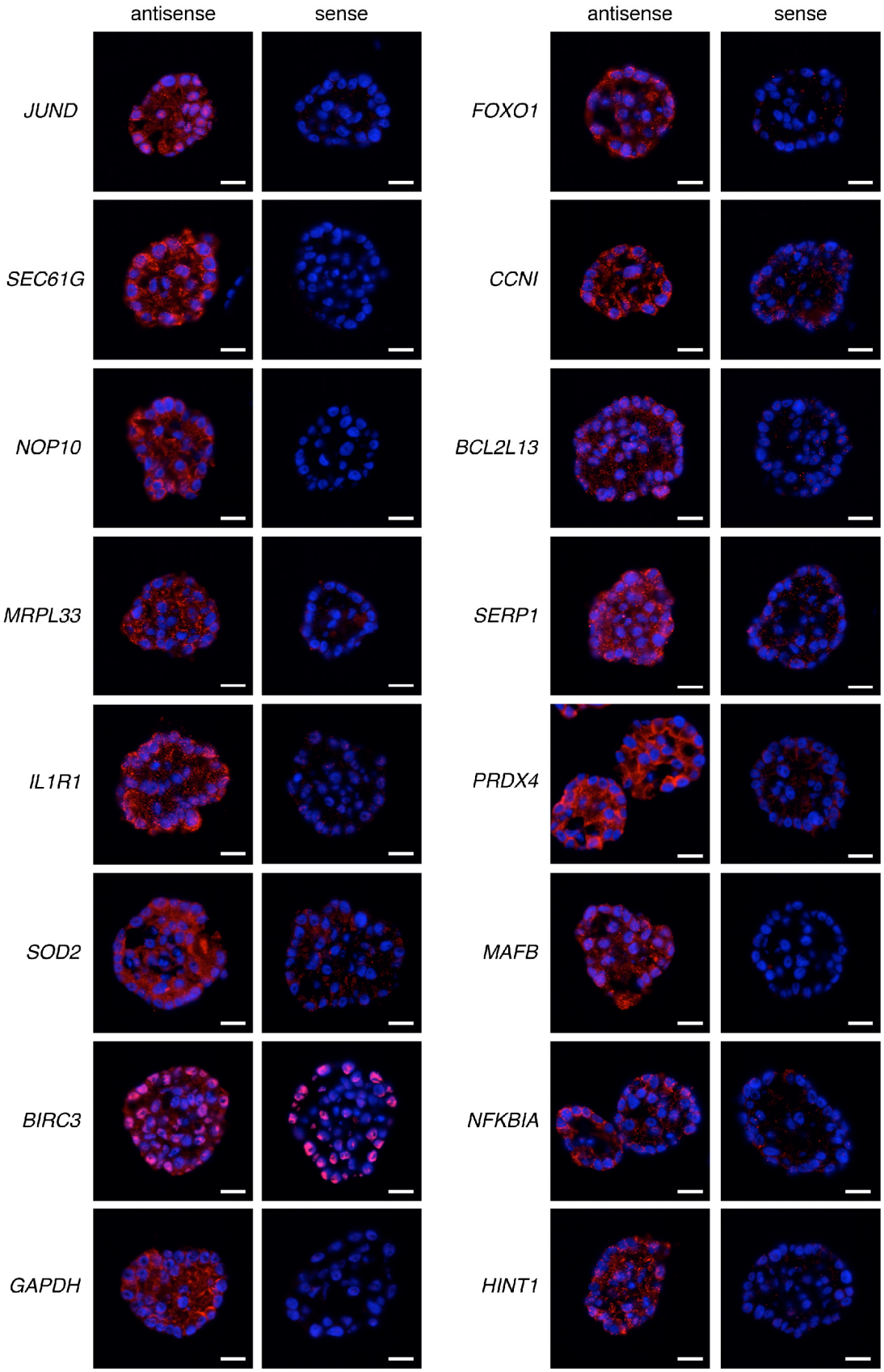


Supplementary Figure 7. | Applied stochastic profiling for discovering candidate single-cell heterogeneities.

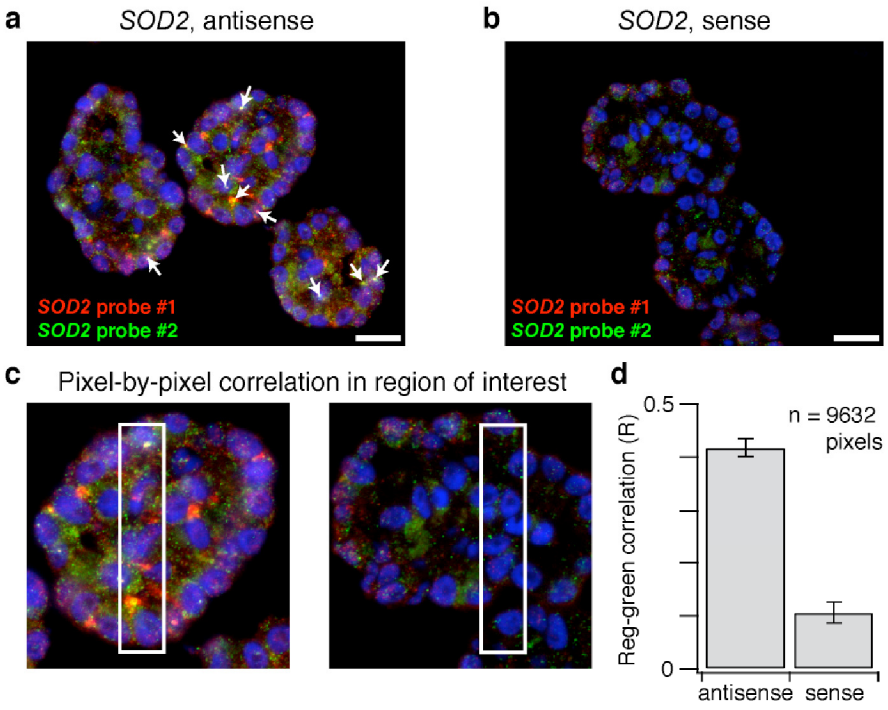
(**a**) Derivation of an empirical reference distribution for stochastic profiling. CVs for the reference genes highlighted in **Figure 3a** were calculated and analyzed as a histogram, yielding a median CV of 19%. (**b**) Stochastic profiling should be effective given an empirically derived reference distribution. Simulations were performed as described in **Figure 1g** for 10 cells averaged with CV_a = 19%, CV_b = 14–26% (the interquartile range of the distribution shown in **a**), D = 8, and F = 0.2. Note that stochastic sampling is much more robust when using the empirically derived reference distribution compared with the literature-derived reference distribution shown in **Figure 1g**. (**c**) Local neighborhoods of three genes were grouped and tested against a log-normal reference distribution with CV_a = 19%. This procedure was iterated through all three possible combinations of three local genes, and the median p value from the χ^2 test for lognormality is shown. Genes whose median p value is not below false discovery are shown in gray, and genes with significant deviations from a log-normal distribution are the candidate heterogeneities shown in red. The clustering results from **Figure 3a** are shown for reference above the plot.



Supplementary Figure 8. | Heterogeneous activation of the NF-κB pathway in matrix-attached cells of MCF10A 3D structures.
(**a,b**) 5 μm thick frozen sections of MCF10A-5E acini were stained with (**a**) the p65 subunit of NF-κB or (**b**) IκBα as described in the **Online Methods**. For **a**, the image is shown in grayscale to emphasize p65 localization. Single cells showing strong nuclear localization of p65 (high NF-κB activation) are highlighted with arrows, whereas those showing strong cytoplasmic localization (low NF-κB activation) are highlighted with flat markers. For **b**, the image is pseudocolored to highlight quantitative differences in fluorescence intensity. Single cells showing strong IκBα expression (low NF-κB activation) are highlighted with arrows, whereas those showing weak expression (high NF-κB activation) are highlighted with flat markers. For **a** and **b**, grayscale or pseudocolored images are shown (left) together with merged images including a DAPI counterstain to highlight nuclei (center, right). Scale bar is 25 μm.

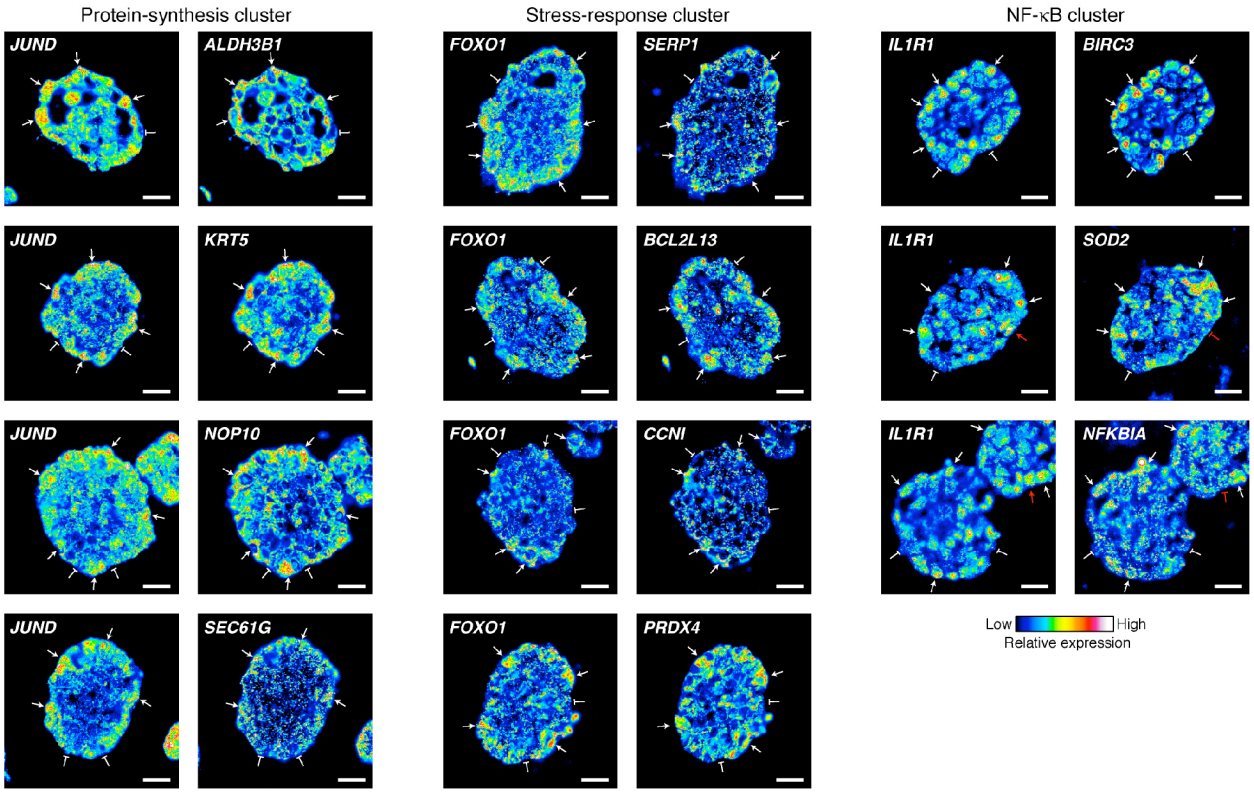

antisense
sense
antisense
sense
JUND
FOXO1
SEC61G
CCNI
NOP10
BCL2L13
MRPL33
SERP1
IL1R1
PRDX4
SOD2
MAFB
BIRC3
NFKBIA
GAPDH
HINT1

Supplementary Figure 9. | Fluorescence intensities from RNA FISH are specific to endogenous transcripts.
Haptenylated riboprobes for the indicated genes were hybridized to day 10 frozen sections of MCF10A-5E acini and imaged by widefield immunofluorescence (see **Online Methods**). For the indicated gene, the antisense and sense control are shown from identically treated specimens. The exposure time and scaling of images for each gene is the same to allow comparison of the specific (antisense) and background (sense) signals. Discussion of the *BIRC3* probe is included in **Supplementary Note 3**. Scale bar is 20 μm.



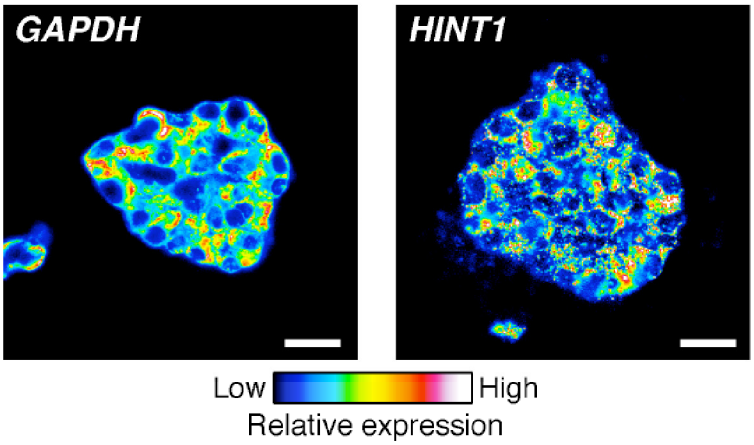
Supplementary Figure 10. | Fluorescence intensities from RNA FISH are specific to the intended transcript.

(**a**,**b**) Antisense and sense probes against two different regions of *SOD2* were labeled with different haptens and red or green fluorophores where indicated and stained by RNA FISH (see **Online Methods**). Note that the antisense probes are brighter and extensively colocalized in areas of active transcription (arrows). (**c**) Regions of interest from **a** and **b** used for quantitative analysis. (**d**) Strong correlation between red and green fluorescence intensities per pixel in the antisense image shown in **a**. Correlations are shown ± 90% Fisher Z-transformed confidence intervals for the indicated number of pixels. The sense control is essentially uncorrelated.



Supplementary Figure 11. | Validation of coregulated single-cell heterogeneities predicted by stochastic profiling.

Two-color RNA FISH images were collected at day 10 of MCF10A morphogenesis for the indicated gene pairings. Images are pseudocolored to emphasize quantitative differences in fluorescence intensity, and single cells showing strong coexpression are highlighted with arrows (high expression) or flat markers (low expression). Single-positive cells (red) for the *IL1R1–SOD2* and *IL1R1–NFKBIA* pairings are discussed in **Supplementary Note 2**. Scale bar is 20 μm.



Supplementary Figure 12. | Reduced cell-to-cell expression variation for genes not predicted to be heterogeneously expressed by stochastic profiling.
Two-color RNA FISH images were collected at day 10 of MCF10A morphogenesis for *GAPDH* and *HINT1*, two genes with sampling fluctuations not significantly different than the reference distribution. Images are pseudocolored to highlight quantitative differences in fluorescence intensity. Note that matrix-attached cells show considerably less variation in *GAPDH* and *HINT1* expression compared to the genes examined in **Figures 4** and **5** and **Supplementary Figure 11**. Scale bar is 20 μm.

Supplementary Table 1. | Primers used in this study.

Gene	Forward primer	Reverse primer	Primer amount (pmol)	Amplification efficiency (%)
BTG1	CCTACAGAATTGGAGAGGATGG	AGTTTTTGGAAGGGCTCGTT	5	107
CAPNS1	GATCAGGGACCATTTGCAGT	GATTTGAAGGCACGGAACAT	10	90
CAV1	CATCTGGGGCATTTACTTCG	GATGCGGACATTGCTGAATA	20	95
CCNB1	ATTGCAGCAGGAGCTTTTTG	CATGCTTCGATGTGGCATAC	10	85
CCNI	AAGGACAACAGCAAGCCAGA	GAGCCGTTTGATTCCATCAT	10	101
CDKN1A	GGTGGCAGTAGAGGCTATGG	GCCGAGAGAAAACAGTCCAG	10	95
CDKNIC	GACCGTTCATGTAGCAGCAA	AGCTTTACACTTTGGGACCAG	20	101
CYLD	TCCCTTGCCAGAATATGGAG	ATGGGGTGACTTGAGGAATG	5	93
FBXO32	GCAGCTGGATTGGAAGAAGA	GTCCTGGGGTGAAAGTGAAA	10	92
FOXO1	TAACCCCAGCCCCAACTTA	AATGCCAGGTTGGTCTGTTC	10	104
FOXO3	TTGATTTGAAGCACCTCATCC	GAAAGGGGAAGATTCCCAAA	10	101
GAPDH	AACGTGTCAGTGGTGGACCT	TCGCTGTTGAAGTCAGAGGA	10	95
HINT1	GCCTTGCTTTCCATGACATT	CCTTATTCAGGCCCAGATCA	10	92
KRT10	GCCCGACGGTAGAGTTCTTT	CCATAGACCATCAAGACAGAAGTG	10	90
MAP2K2	GTGAACGAGCCACCTCCTAA	GGCAAAATCCACTTCTTCCA	10	98
PRDX6	CGTGTGGTGTTTGTTTTTGG	CTTCTTCAGGGATGGTTGGA	10	90
RUNX1	ATCCCAGAACCCTCTCTCCA	CTCTGGTGGCTCCTGAACA	10	105
S100A6	ACACCCTGAGCAAGAAGGAG	CCCCAGGAAGGTGACATACT	10	98
SDHA	ACACAGACCTGGTGGAGACC	CAAAGGGCTTCTTCTGTTGC	10	107
SEMA3C	TGGTGGTTGATTGGTTTTATTTT	CACAGATTGAACATTCACCATT	20	104
SESN1	AGGCTGTGCAGCTCCTTAAA	AAAGCAATTCACGCTTCCAG	20	106
SOD2	CTGGACAAACCTCAGCCCTA	TTTGTAAGTGTCCCCGTTCC	10	98
SOX4	TCAAGGCCACAAAAACAATG	GGGTTCAGTTCCAGCTGATT	20	111
UBC	CAGAGGTGGGATGCAAATCT	TTTCCCAGCAAAGATCAACC	10	98

Supplementary Note 1. | Detection of cell-to-cell heterogeneities by RNA FISH.

Nearly all genes were clearly observed as heterogeneously expressed by RNA FISH (**Fig. 4a-c** and **Supplementary Fig. 11**). However, two genes (*MRPL33* and *MAFB*) showed expression patterns that were too punctate to determine whether cell-to-cell heterogeneities existed (**Supplementary Fig. 9**). The probes for *MRPL33* and *MAFB* were specific when compared to the sense control, indicating that the punctate patterns arose because of subcellular RNA localization[2]. A recent large-scale study of RNA localization in *Drosophila* indicated that many transcripts are routed to the site of protein function[3]. The L33 subunit encoded by *MRPL33* comprises part of the mitoribosome[4], suggesting that the speckled *MRPL33* expression indicates localization to mitochondria.

The pronounced subcellular expression pattern of *MAFB* is surprising, given that MafB is a transcription factor[5]. Such RNA foci could arise if *MAFB* were located in stress granules or P-bodies[6]. Attempts to quantify either speckle intensity or speckle number were inconclusive as a measure of transcript abundance. Thus, we were unable to validate the expression heterogeneities predicted by stochastic profiling for these two genes.

Supplementary Note 2. | Coregulated expression patterns identified by RNA FISH.

We identified multiple examples of coregulated expression for the clusters shown in **Figure 3b** (**Fig. 4a-c** and **Supplementary Fig. 11**). Nevertheless, we were unable to confirm coexpressed heterogeneities between any transcripts and *MRPL33* or *MAFB* because of their punctate expression pattern (**Supplementary Note 1**). In addition, we identified two genes in the NF-κB cluster (*SOD2* and *NFKBIA*) where the single-cell fluorescence patterns by RNA FISH did not exactly match that predicted by stochastic profiling. We believe that this discrepancy is because of the difference in sample processing that was required for RNA FISH compared to laser-capture microdissection (see **Online Methods**). For random sampling by laser-capture microdissection, the 3D structures were embedded immediately as fresh-frozen tissue. For RNA FISH, samples were washed and then fixed with paraformaldehyde for 15 min before proceeding to the cryopreservation and embedding steps. *SOD2* and *NFKBIA* are rapidly induced by NF-κB[7, 8], but both transcripts are highly unstable[9, 10]. Thus, we identified single cells that expressed *IL1R1* but were negative for *SOD2* or *NFKBIA* (**Supplementary Fig. 11**), suggesting that the *SOD2* or *NFKBIA* transcripts in these cells had degraded during sample processing for RNA FISH.

Supplementary Note 3. | Validation of the *BIRC3* probe for RNA FISH.

We were severely constrained in our ability to design a 175-225 bp riboprobe that was specific for *BIRC3* and representative of the microarray data, because ~1.9 kb of its 3' end is 78% identical to *BIRC2*. We therefore cloned a probe against the one 3' region of *BIRC3* that is not conserved in *BIRC2*, which is the probe shown in **Supplementary Figure 9**. Unfortunately, the corresponding sense probe for this region causes very-high nuclear staining. Lower probe concentrations or higher-stringency washes did not improve the background staining with the sense probe. However, the antisense probe stains the nucleus heterogeneously (as opposed to the homogeneous staining observed with the sense probe) and exhibits a much higher cytoplasmic staining. Moreover, we found that *BIRC3* expression as visualized with this probe tightly matches *IL1R1* expression (**Fig. 4c,d**). Thus, we believe the antisense *BIRC3* probe is largely specific for the endogenous transcript.

REFERENCES

1. Brady, G. & Iscove, N.N. Construction of cDNA libraries from single cells. *Methods Enzymol* **225**, 611-623 (1993).
2. Martin, K.C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719-730 (2009).
3. Lecuyer, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174-187 (2007).
4. Suzuki, T. et al. Proteomic analysis of the mammalian mitochondrial ribosome. Identification of protein components in the 28 S small subunit. *J Biol Chem* **276**, 33181-33195 (2001).
5. Kataoka, K. Multiple mechanisms and functions of maf transcription factors in the regulation of tissue-specific genes. *J Biochem* **141**, 775-781 (2007).
6. Balagopal, V. & Parker, R. Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr Opin Cell Biol* **21**, 403-408 (2009).
7. Pahl, H.L. Activators and target genes of Rel/NF-kappaB transcription factors. *Oncogene* **18**, 6853-6866 (1999).
8. Wong, G.H. & Goeddel, D.V. Induction of manganous superoxide dismutase by tumor necrosis factor: possible protective mechanism. *Science* **242**, 941-944 (1988).
9. Blattner, C. et al. UV-Induced stabilization of c-fos and other short-lived mRNAs. *Mol Cell Biol* **20**, 3616-3625 (2000).
10. Akashi, M. et al. Irradiation increases manganese superoxide dismutase mRNA levels in human fibroblasts. Possible mechanisms for its accumulation. *J Biol Chem* **270**, 15864-15869 (1995).