Published in final edited form as:
Nat Protoc. 2013 February ; 8(2): . doi:10.1038/nprot.2012.158.

Generalized stochastic profiling of transcriptional regulatory heterogeneities in tissues, tumors, and cultured cells

Lixin Wang[1] and **Kevin A. Janes**[1,*]
[1]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908

Abstract

Single-cell variations in gene and protein expression are important during development and disease. Cell-to-cell heterogeneities can be directly inspected one cell at a time, but global methods are usually not sensitive enough to work with such a small amount of starting material. Here, we provide a detailed protocol for stochastic profiling, a method that infers single-cell regulatory heterogeneities by repeatedly sampling small collections of cells at random. Repeated stochastic sampling is performed by laser capture microdissection or limiting dilution, followed by careful exponential cDNA amplification, hybridization to microarrays, and statistical analysis. Stochastic profiling surveys the transcriptome for programs that are heterogeneously regulated among cellular subpopulations in their native tissue context. The protocol is readily optimized for specific biological applications and takes about one week to complete.

Keywords

heterogeneity; single-cell; stochastic; systems biology; gene expression; cancer; development

INTRODUCTION

Even within a clonal population, no two cells are truly equal[1-4]. Nonuniformities in the cellular microenvironment[5-7] combine with random fluctuations caused by transcription[5, 8, 9], translation[10, 11], and cell division[12] to yield cell-to-cell heterogeneities that can be profound. Biological mechanisms exist to suppress variation[13], but they are energetically costly[14]. Thus, isolating "pure" subpopulations by lineage or surface markers is largely artificial, because these cells will eventually drift back to a steady-state heterogeneity[15-18]. Instead, a better strategy for understanding cell-to-cell differences may be to exploit population variability and consider each cell as its own self-contained experiment[19-21].

This approach has now become possible with the development of global techniques for analyzing single cells[22]. Genomic[23-26] and proteomic[27-30] methods are actively advancing, but the first to appear was single-cell transcriptomics[31-33]. The details of mRNA expression profiling in single cells can vary widely depending on the method[31-43]. Algorithmically, however, the different protocols all involve roughly the same steps: 1) extract cellular RNA

*Correspondence should be addressed to K.A.J. (kjanes@virginia.edu).

AUTHOR CONTRIBUTIONS

L.W. designed the current implementation of stochastic profiling and the optimization protocol for different cellular contexts. K.A.J. conceived of the method, supervised the development of the current implementation, coded all computer simulations, and wrote the manuscript with contributions from L.W.

COMPETING FINANCIAL INTERESTS

The authors declare that they have no competing financial interests.

by chemical, thermal, or enzymatic methods; 2) perform an oligo(dT)-based capture or an abbreviated oligo(dT)-primed reverse transcription (RT) to prepare a cDNA library of roughly uniform length; 3) tail the library with homopolymer; 4) exponentially pre-amplify the tailed cDNA with a universal homopolymer-containing primer; and 5) detect the amplification products by quantitative PCR (qPCR), oligonucleotide microarrays, or RNA-seq. Steps 1–5 are iterated across dozens or hundreds of single cells in an effort to reconstruct the population-level distribution and identify recurrent expression states.

Studies using the above workflow have uncovered many qualitative heterogeneities in the areas of neuroscience[32, 44, 45] and tissue development[33, 41, 46, 47]. Interestingly, when similar approaches were applied to cells from a common lineage—where regulatory heterogeneities are possibly more quantitative rather than qualitative—the findings were much more generic[48-50]. This suggested that existing transcriptomic methods did not clearly separate biological variability from measurement variability when using one cell worth of starting material[42, 43]. Indeed, certain steps essential to the procedure, such as the RT step, are known to add substantial measurement variation when starting with minute amounts of input RNA[51, 52]. A second confounding factor was that virtually all protocols required tissue dissociation to isolate single cells by micropipette aspiration or FACS[34-40]. This is a major drawback for studying adherent populations such as epithelia, where cell detachment alters signaling and gene expression within minutes[53-55]. Cell-to-cell variation caused by the dissociation procedure could distort the true heterogeneities in the resident tissue. To study single-cell biology through transcriptomics, it would be crucial to reduce measurement and handling artifacts substantially.

The collective challenges with existing methods prompted us to develop stochastic profiling[56], an alternative approach for gaining single-cell information quantitatively, efficiently, and in situ. Stochastic profiling does not focus on intrinsic biological noise[5] but rather on cell-to-cell heterogeneities in the regulation of gene expression. The method is based on the premise that heterogeneities in single-cell regulation can be inferred without measuring single cells explicitly (Fig. 1a). Instead, small collections of ~10 cells are repeatedly sampled at random by laser-capture microdissection[57], and mRNA expression for each of these “stochastic samplings” is quantitatively profiled with a customized small-sample cDNA amplification procedure[56]. Highly accurate and precise expression profiles for the stochastic samplings are achievable because of the 10-fold increase in starting material compared to single cells. After building gene-by-gene histograms from 15–20 stochastic samplings (Fig. 1b), statistical hypothesis testing is then used to identify transcripts whose distribution is significantly different from the lognormal distribution, a common null model for ordinary biological variability[58, 59]. Transcripts subject to dichotomous single-cell regulation are identified at this step because of binomial fluctuations in the proportion of high- and low-expressing cells collected during each stochastic sampling. Finally, the filtered transcripts are clustered to identify groups of genes with correlated sampling fluctuations that suggest putative single-cell expression programs[56].

Comparison with other methods

Our procedure holds two main advantages over existing methods, both of which are related to the 10-fold increase in starting material. First, stochastic-profiling measurements are much more sensitive to low-abundance transcripts and less prone to experimental noise compared to single-cell amplifications[56]. For this reason, stochastic profiling is readily compatible with laser-capture microdissection, which is superior to dissociation-based methods for retaining adherent cells in their native context[57, 60]. Tissues are snap-frozen within seconds and can be ready for microdissection with minimal processing in aqueous solutions where RNA can be damaged or degraded. However, because the RNA must be

proteolytically released from the microdissected specimen, the overall yield is lower than would be obtained by chemical lysis of suspension cells. 10-cell sampling offsets the decrease in RNA yield per cell and enables quantitatively accurate expression profiling[56].

Second, by measuring 10 cells at a time, stochastic profiling surveys the overall population heterogeneity more efficiently. For strict single-cell methods, large numbers of samples must be processed to ensure that the major subpopulations have been captured[34, 40, 41, 46]. These sample numbers eventually become prohibitive because of the most-expensive step: gene detection (Step 5 above). By contrast, stochastic profiling gathers information about 10 cells per sample, which hones in on recurrent heterogeneities with fewer samples. Through computer simulations, we determined that stochastic profiling should be able to detect high-low transcriptional heterogeneities that occur with a frequency of 5–50% (Fig. 1c). It would be difficult to identify infrequent heterogeneities (~5–20%) confidently with 15–20 single-cell analyses, but such patterns are readily uncovered by stochastic profiling[56, 61].

Applications of the method

In general, stochastic-profiling clusters are extremely informative[56], because spurious correlations among 15–20 random samplings are unlikely, even when surveying the transcriptome. For example, a Pearson correlation of $R = 0.7$ among 18 samplings has a 0.06% probability of being observed by chance, implying only six false correlations when 10,000 genes are surveyed. The simplest explanation for a correlated transcriptional cluster is that the constituent genes are jointly controlled by a common upstream regulatory factor, which is heterogeneously activated.

We have recently used this reasoning to study the single-cell regulation of FOXO transcription factors during 3D organotypic culture of breast epithelial cells[61-63]. Stochastic profiling identified a clear split in the sampling fluctuations of FOXO-regulated genes, which we independently validated in single cells by multicolor RNA FISH[61]. Over 90% of gene pairs within a single FOXO cluster were strongly correlated among single cells ($R >$ 0.6), whereas over 60% of gene pairs across clusters were weakly correlated or uncorrelated ($R < 0.4$). Bioinformatic analysis[64, 65] of promoters together with chromatin immunoprecipitation revealed that one cluster of FOXO target genes was coregulated by another transcription factor, RUNX1 (ref. [61, 62]). FOXO–RUNX1 crosstalk was unanticipated and became apparent only when examining the heterogeneous expression state of single cells via stochastic profiling. Very recently, *RUNX1* was found to be recurrently mutated in breast cancer[66, 67], independently validating our earlier predictions of its tumor-suppressive role[61, 62].

Looking forward, we anticipate that stochastic profiling will be useful as a tool for studying heterogeneous cell-to-cell regulation. For example, it was shown that co-fluctuating proteins in yeast act as “noise regulons” that coordinate important biological processes[68]. Remarkably, the functions of several regulons (e.g., stress response and protein biosynthesis) were identical to the expression programs identified previously by stochastic profiling of 3D breast-epithelial cultures[56]. This suggests that there may be some inherent circuits linked to heterogeneous regulation that are widely conserved[20]. Another future direction for stochastic profiling is to examine the mechanisms of partially penetrant phenotypes[6, 69, 70]. Conceivably, such phenotypes are driven by upstream molecular heterogeneities before the phenotype is obvious. Stochastic profiling could be used to search for these heterogeneities in an unbiased way.

Last, we emphasize that the principle of stochastic profiling is completely general. Although implemented for transcriptomics, the concept of random sampling could be applied to other high-sensitivity methods that analyze small numbers of cells[29, 71-73]. For protein analysis,

the 10-cell threshold of stochastic profiling may be much easier to reach than a one-cell threshold because of the inability to amplify the starting material. Particularly exciting would be small-sample stochastic profiling of chromatin modifications at a genome-wide level[74, 75]. The analysis pipeline described at the end of the protocol here could be immediately adapted to such alternative implementations of our method.

In the Supplementary Software, we provide a script (`StochProfParameters.m`) that simulates stochastic profiling with six user-defined parameters: 1) the number of cells per sampling; 2) the number of samplings used to build the distribution; 3) the coefficient of variation (CV) of the log-normal reference distribution (CV_{ref}) that specifies the null model for hypothesis testing; 4) the underlying CV of the log-normal test distribution (CV_{test}), which is used to diagnose false positives; 5) the fold difference in expression (D) between high and low subpopulations; and 6) the expression fraction (F) of cells in the high subpopulation. By running this script, users can survey up to two parameter ranges at a time to assess the performance of the method for different biological applications (Fig. 1).

Experimental design

Cryopreservation and frozen sectioning—When working with tissues and tissue-like material, proper cryosectioning is an important first step for stochastic profiling. Ideally, fresh samples are embedded and frozen simultaneously in a dry ice-isopentane bath. However, the procedure also works with tissues snap-frozen in liquid nitrogen and then embedded afterwards. Fixed-frozen specimens are incompatible because RNA is crosslinked within the tissue and cannot be released enzymatically.

We prefer to use cryostats with disposable microtome blades that can be replaced after each set of sections is collected. During sectioning, the goal is to keep specimens at the lowest temperature possible. The downstream histology procedure is meant to preserve RNA integrity, not morphology. Thus, we routinely cut sections at temperatures that cause some chattering of the blade and flaking of the tissue. After wicking each section, the slide is placed immediately into a slide box within the refrigerated cryostat to refreeze the section as quickly as possible. The sample should never thaw thereafter. Because of the atypical sectioning requirements, we prefer to cut sections ourselves rather than submitting samples to a core histology facility.

Rapid histology and laser-capture microdissection—Various stains have been reported to be compatible with laser-capture microdissection[76]. However, the stochastic profiling protocol uses nuclear fast red because of its superior ability to maintain RNA integrity[77]. Since RNA is most susceptible to hydrolysis during aqueous processing steps[78], a broad-spectrum RNAse inhibitor is spiked into the staining solution immediately before use. The staining protocol described here is versatile and can be applied to various tissue types as well as cultured adherent cells plated on coverslips (Fig. 2).

After washing briefly, samples are dehydrated with an ethanol series and cleared with xylenes. We recommend purchasing ethanol in small quantities, because ethanol is hygroscopic and opened containers will draw moisture from the air. Likewise, the xylene step should be precisely controlled for effective microdissection. Excessive clearing can lead to overdrying and collateral pickup of cells adjacent to individual laser shots. Conversely, insufficient clearing will dry the section too slowly, causing ambient moisture to enter the section and making microdissection impossible (see TROUBLESHOOTING).

For stochastic profiling, it is critical to maintain the RNA integrity of each cell that is microdissected. Ultraviolet-based microdissection platforms cut tissues very cleanly, but RNA near the ultraviolet laser is severely degraded. Thus, our protocol uses an infrared-

based microdissection instrument for gentle, mechanical dissociation of a single cell from its neighbors. Collateral pickup of surrounding tissue is easily removed by gently pressing the microdissection cap against a weak adhesive note. As a positive control, a larger sampling of 100 cells is included to assess the overall amplification efficiency. To control for amplification variability, a large pool of microdissected cells is split into multiple identical 10-cell aliquots after elution from the microdissection cap. Usually, the 15–20 repeated samplings are split across two separate days so that sampling-to-sampling variation and day-to-day variation can be compared. If the observed expression heterogeneities are strongly associated with specific amplification groups, it indicates problems with day-to-day reproducibility of the procedure.

The biological strategy for random sampling must be clearly defined up front, as it critically influences the types of regulatory heterogeneities that will be uncovered by the method. Stochastic profiling starts with a hypothesis about where such heterogeneities might lie hidden. Experimental conditions (time point, treatment condition, etc.) should be optimized beforehand to focus on the sought-after heterogeneities as cleanly as possible. To avoid complications from obvious cell-to-cell variation, such as differences in the cell cycle or cellular microenvironment, samplings should be collected as uniformly as possible. Key parameters to control for include cell size, distance from blood vessels, and contact with ECM or neighboring cell types. Hidden variations arising from clonal cell subpopulations can be averaged out within each sampling by microdissecting cells across different regions of the tissue. Alternatively, by collecting the sampled cells locally, clone-to-clone variations can be enriched if desired.

RNA elution and small-sample cDNA amplification—Small-sample cDNA amplification[56] for stochastic profiling involves: 1) cellular proteolysis to release RNA from the specimen, 2) an abbreviated oligo(dT)-primed RT to yield a cDNA pool of uniform length, 3) poly(A) tailing with terminal transferase, and 4) poly(A) PCR with a universal oligo(dT)-containing primer (AL1, ref. [79]; Fig. 3). Care must be taken to avoid contaminating samples with poly(A) PCR amplicons from previous experiments, and lack of contamination should be confirmed with a blank control that has been run through the entire procedure. Genomic DNA contamination is not ordinarily a problem because of the small amount of starting material, but this should be evaluated with a no-RT control. Because there is no DNAse step and the abbreviated RT often does not span an intron, we sometimes observe slight amplification in the no-RT sample for transcripts with many pseudogenes. We consider this acceptable as long as the no-RT levels are negligible compared to the transcript levels observed in the samples.

When working with samplings obtained by microdissection, the initial cellular proteolysis is critically important for high-sensitivity amplification. Tissue sections are solvent fixed and bound to the polymer on the microdissection cap (see above). Thus, RNA must be freed from precipitated ribonucleotide binding proteins in a way that is compatible with the downstream amplification steps. We use proteinase K as a broad specificity protease because of its high activity at elevated temperatures. To avoid digestion of the later enzymes in the procedure, proteinase K is irreversibly inhibited with saturating concentrations of phenylmethanesulfonyl fluoride (PMSF). Excess PMSF is then rapidly hydrolyzed in the alkaline pH of the first-strand synthesis buffer without noticeable inhibition of the RT step itself. More-stable protease inhibitors are not as effective, presumably because of interference later in the procedure.

Although our protocol was originally designed for microdissected cells[56], the amplification is also compatible with suspension cells obtained by FACS or limiting dilution. The digestion buffer components are separated into two parts for washing-storage and lysis-

digestion, with saponin added to the lysis component as a gentle permeabilizing agent (see PROCEDURE). Small quantities of suspension cells can be stored frozen in buffer before starting the lysis-digestion without loss of amplification efficiency (Supplementary Fig. 1). This provides a convenient means for archiving primary or flow-sorted samples before starting the amplification.

We have recently discovered that the performance of the PCR amplification depends critically on the cell type and sample format. This is likely because of differences between the overall mRNA content of different cells and the efficiency of RNA extraction during cellular proteolysis. Thus, we recommend a sample-specific optimization protocol that should be followed when adapting stochastic profiling to new biological contexts. Among all parameters, we have found that the amount of AL1 primer and the number of poly(A) PCR cycles are the most crucial for sample-specific optimization (Fig. 3). The original primer concentration (5 μg AL1 per 100 μl PCR reaction[56]) is the minimum required for high-sensitivity detection. Amplification of some samples continues to improve with up to tenfold higher amounts of AL1, and thus our optimization protocol recommends testing 5–50 μg in pilot experiments.

When performing the optimization, fractions of the PCR amplification should be collected from 25–40 cycles for monitoring by qPCR. The goal is to identify the maximum number of cycles where high- and low-abundance transcripts (defined by qPCR cycle threshold) are still amplifying efficiently with a 100-cell sample. We use housekeeping genes[80] as abundant mRNA species and then screen various surface receptors and transcription or splicing factors that can act as low-copy readouts of the amplification. By surveying 6–8 genes within this range, the optimal AL1 concentration and PCR cycle number is readily identified for a cell and sample type (Fig. 4). Next, the amplification is repeated under the optimized AL1 and cycling conditions with serial dilutions of starting material from 100 cells to one cell. We consider the amplification valid when all transcripts tested show a reproducible log-linear increase in qPCR cycle threshold with decreasing starting material down to three cells (see ANTICIPATED RESULTS). At the 10-cell input level used for stochastic profiling, there should be no need to exclude “unsuccessful” amplifications[38].

Reamplification and aminoallyl labeling—The cDNA prepared by small-sample amplification is immediately suitable as a template for qPCR, but samples must be labeled before global profiling by microarrays. Amplified cDNA is diluted and reamplified in the presence of aminoallyl-dUTP, which provides a strong nucleophile for conjugation to fluorescent succinimidyl esters. Reamplification involves design criteria that are different from small-sample amplification. During the 10-cell amplification, processivity and sensitivity of the PCR reaction are paramount. For reamplification, sensitivity is less of an issue and a higher priority is placed on achieving a high degree of labeling. We screened several polymerase blends for their ability to efficiently incorporate high levels of aminoallyl-dUTP and had the greatest success with Roche High Fidelity polymerase. Our protocol replaces 80% of thymidine bases with aminoallyl-uracil to maximize dye coupling. The aminoallyl moiety is located at the 5-position of the pyrimidine ring of uracil, which is not adjacent to the 2- and 3-positions involved in base pairing. Consequently, unreacted aminoallyl groups are not expected to interfere with microarray hybridization.

As with small-sample amplification, the number of PCR cycles during reamplification must be optimized empirically. To obtain accurate cycle-by-cycle estimates for the extent of amplification, a pilot reaction is performed in the presence of SYBR Green and monitored by real-time qPCR[81]. Great care must be taken to avoid saturating the reaction and ruining quantitative accuracy. Thus, the maximum number of reamplification cycles for all samples must fall near the mid-exponential phase of the first sample that amplifies detectably (Fig.

5a). Varying the number of PCR cycles on a sample-by-sample basis is not recommended, because the SYBR Green estimates by qPCR are a mixture of amplified material and primer dimer. Instead, samples with lower amounts of starting material can be reamplified in several parallel reactions that are pooled and concentrated during the purification step. This conservative strategy avoids overamplifying some of the samples inadvertently, and it retains the quantitative accuracy and reproducibility of the starting material[56].

Before dye coupling, it is important to remove primer dimers from the reamplification. Primer dimers will cause an overestimation of cDNA yield, and aminoallyl-labeled primer dimers will compete for the dye label. We sought to avoid the need for a gel-purification step[82], because the DNA yields after gel extraction and isolation are typically poor. Instead, we use PureLink spin columns with a modified protocol that achieves near-stoichiometric isolation and recovery of cDNA from the anion-exchange resin (see PROCEDURE). Aminoallyl-labeled cDNA is coupled to Alexa Fluor 555, which is interchangeable with Cy3 but shows superior performance for microarray applications[83]. Additionally, Alexa Fluor 555 decapacks are available as single-use aliquots, which save cost. After dye conjugation and secondary purification, the degree of labeling is determined by spectrophotometry with the Invitrogen Dye:Base Ratio Calculator: http://probes.invitrogen.com/resources/calc/basedyeratio.html. Our protocol enables the conjugation of 7–10 dye molecules per ~500 bp amplicon (see ANTICIPATED RESULTS) (Fig. 5b).

Microarray hybridization and data analysis—Alexa Fluor 555-labeled cDNA should be compatible with any commercial microarray platform. However, we have performed stochastic profiling exclusively with Expression BeadChips from Illumina because of their reduced cost, higher throughput, and equivalent performance compared to more-expensive alternatives[84]. The hybridization protocol is followed as recommended by the manufacturer with the following modifications: 1) 1 □g of cDNA is added to each well rather than 750 ng cRNA to account for the fact that only the complementary strand of the cDNA sample will hybridize, and 2) the samples are denatured briefly at 95 °C and then added to the slide prewarmed at the 58 °C hybridization temperature. The second modification attempts to minimize re-annealing of the labeled cDNA with its complementary strand before hybridization. From this point, slides are incubated, washed, and scanned exactly as recommended.

A stochastic profiling experiment typically involves 16–20 random 10-cell samplings and the same number of amplification controls (a larger pool of 160–200 cells split into 10-cell aliquots before small-sample amplification). Each microarray is normalized to its mean fluorescence intensity, and then genes are filtered according to two criteria for the amplification controls. First, the gene must be reproducibly amplified. Irreproducible transcripts are readily flagged, because an unsuccessful amplification causes dramatic fluctuation artifacts in the controls. We apply a loose filter to the amplification controls that removes genes with control fluctuations >5-fold. Second, each gene must be reproducibly detected. We retain genes with a median detection $P < 0.1$ across the amplification controls as determined by the microarray manufacturer. After filtering, the data are re-normalized by median fluorescence intensity to adjust for residual post-filtering differences in overall signal. The re-normalized 10-cell samplings comprise the final preprocessed dataset for analysis.

The first step in the analysis pipeline is to extract the genes with sampling fluctuations significantly greater than measurement fluctuations. Because eukaryotic gene-expression variability is often log-normally distributed[58, 85], we logarithmically transform the data for analysis. To standardize the log-transformed data, each transcript is then scaled by its geometric mean taken across all samplings, and each sampling is scaled by its geometric

mean taken across all transcripts. Next, we must isolate the transcripts with sampling variations that are significantly greater than the measurement variation intrinsic to the amplification. This allows us to estimate a reference distribution with which to compare the fluctuations of the 10-cell sampling measurements. In our original work[56], we compared the CV of the sampling fluctuations with the CV of the amplification controls by using McKay's approximation[86]. However, we now prefer to avoid approximations and instead directly examine the ratio of variances with respect to the *F* distribution[87]. Genes with significantly higher sampling variances relative to controls (at a user-defined false-discovery rate, FDR_{var}) are extracted and then sorted based on their CV for subsequent distribution testing.

There are a variety of methods for comparing empirical data with a (log)-normal distribution. Our earlier work used the χ^2 goodness-of-fit test[56, 61], but we now favor the K-S test because it is conservative and can be accurately applied on a gene-by-gene basis. Other alternatives could be considered when seeking greater statistical power[88]. To define a reference distribution for the K-S test, we inspect the cumulative distribution function of CVs from genes with measurable sampling variations (Fig. 6a). Using this empirical plot, we week to identify a CV_{ref} that encompasses the baseline biological variation but is distinct from variance caused by cell-to-cell heterogeneities in transcriptional regulation. Simulations indicate that there is a wide tolerance for detecting heterogeneities as long as CV_{ref} is below 35% and the underlying variability of the test distribution is within three-fold of CV_{ref} (Fig. 1d). Usually, a reasonable CV_{ref} can be identified around the first inflection point of the cumulative distribution function (Fig. 6a, red). The method assumes that this inflection point indicates the median baseline biological variation (low CV), which can be used as the base condition to test for heterogeneous regulation (high CV). If the earlier variance filtering is too stringent, then fewer low-CV transcripts will enter into the cumulative distribution function, making it harder to identify CV_{ref} (Fig. 6a, darker gray). Ideally, the function would appear as the superposition to two staggered sigmoid curves, indicating a clear separation of the baseline variation (reference distribution) and the heterogeneous cell-to-cell regulation comprising the variation of the test distributions. Last, we threshold the *P* value of each gene according to a second user-defined false-discovery rate (FDR_{het}), which is generally less stringent than FDR_{var} because of the conservative nature of the K-S test. The genes falling below this threshold are predicted by stochastic profiling to be heterogeneously expressed (Fig. 6b, green).

To facilitate the filtering and analysis of stochastic-profiling data, we provide here a pair of MATLAB functions that execute the necessary calculations (Supplementary Software). `StochProfMicroarrayFilt.m` takes tab-delimited ASCII files of gene names, relative microarray fluorescence intensities, and detection *P* values and outputs the filtered, median-scaled array data. This output can be saved as a MATLAB workspace so that the time-consuming filtering step only needs to be done once. `StochProfAnalysis.m` takes the filtered output as input, performing the variance and distribution tests to arrive at the final gene set, which can be standardized by Z score and clustered hierarchically.

As representative microarray data, we include two ASCII files containing 16 stochastic 10-cell samplings and 16 amplification controls (**Supplementary Data 1** and **2**). When executing the analysis pipeline, there are three user-defined inputs to consider: 1) FDR_{var}, the false-discovery rate for testing significant biological variation above measurement variation; 2) CV_{ref}, the reference CV estimating background biological variation (Fig. 6a); and 3) FDR_{het}, the false-discovery rate for testing significant cell-to-cell heterogeneity above background biological variation (Fig. 6b). All three parameters will influence the total number of genes predicted to be heterogeneously expressed. However, our analysis of an earlier dataset[56] suggests that the fundamental clusters of single-cell gene expression are less

sensitive to the exact parameter values (Fig. 6c, white boxes). We recommend that the user iterate through `StochProfAnalysis.m` several times with different combinations of FDR_{var}, CV_{ref}, and FDR_{het} to identify the salient clusters of interest.

Validation and follow-up studies—Stochastic profiling provides a global means for identifying candidate genes that may be subject to heterogeneous transcriptional regulation. However, this is just a starting point for more-specific observations and perturbations of the phenomenon. To validate predicted heterogeneities, we use RNA FISH because gene-specific reagents are readily synthesized, which can be multiplexed in different fluorescence channels. When verifying a heterogeneous transcriptional cluster, multiple gene pairs should be tested in different combinations to examine the extent of coregulation. Overall, we have observed extremely good concordance between stochastic-profiling predictions and RNA FISH experiments with single genes or gene pairs[56, 61].

Validated transcripts can be pursued further to test for functions of the cell-to-cell heterogeneity. We usually start by following up RNA FISH observations with immunofluorescence to confirm that regulatory heterogeneities propagate to the protein level. (Here, it is not uncommon to see some dampening in the cell-to-cell variation due to the extra steps of translation and protein turnover.) Direct functional testing can be challenging, because one needs to separate the role of the heterogeneity from the general role of the gene-protein itself. We initially seek to homogenize the cell-to-cell expression pattern by eliminating the minority expression state observed by RNA FISH. For example, if a high-expression state is observed in 15% of the overall population, we will target the endogenous gene by RNAi with the goal of eliminating the high population. Conversely, if a high-expression state is observed in 85% of the overall population, we will constitutively express the gene to eliminate the low population. The difficulty is that either of these perturbations will also change the overall levels of expression. Ultimately, assigning function to a heterogeneity requires addback approaches, where the endogenous gene is knocked down by RNAi and then an RNAi-resistant version is expressed constitutively at near-endogenous levels. Unlike specificity tests for RNAi targeting sequences[89], here the expectation is that addback will not revert the phenotype caused by knockdown but instead may yield another phenotype caused by disruption of the cell-to-cell heterogeneity.

Limitations

The biggest drawback of stochastic profiling is that the method does not provide a direct single-cell readout, which can be problematic for some applications. If gene-expression clusters are partially correlated, for example, stochastic profiling cannot distinguish whether single cells have a partial coexpression or whether the sampling pattern is caused by an admixture of cells with uncorrelated expression. We are actively working to develop analytical approaches for extracting accurate single-cell information from stochastic-profiling data (manuscript in preparation).

A related limitation is that stochastic profiling cannot diagnose all forms of heterogeneity. Analytically, the method assumes that baseline biological variation is log-normally distributed, which is not true for transcripts with low transcriptional burst frequencies relative to their mRNA degradation rate[90]. This could create problems with false positives, where regulatory heterogeneities would be predicted for genes that simply have an intrinsically noisy expression pattern. Such transcripts would need to be distinguished at the validation and follow-up phase (see above). Problems will also arise with extremely low-abundance transcripts, where some cells will have exactly zero copies, because the log-normal distribution is only defined for values greater than zero[59]. It is unclear whether such

transcripts would be amplified with enough technical reproducibility to reach the distribution-testing phase (see above).

As with nearly all single-cell transcriptomic methods, stochastic profiling focuses on polyadenylated mRNAs and therefore cannot monitor other RNA species (miRNAs, ncRNAs, etc.). Consequently, the current method focuses on oligonucleotide microarrays for detection and not RNA-seq[36, 47]. A final limitation is that stochastic sampling thus far has only been performed based on cell morphology or tissue geography together with simple histological stains. In principle, fluorescent reporters or rapid immunofluorescence[91] could be used in the future to achieve stochastic profiling within molecularly defined cellular subtypes.

MATERIALS

REAGENTS

- Ethanol (Ultrapure #200CSPTP)
- Nuclear fast red (Vector Laboratories #H-3403)
- RNAsin Plus (Promega #N2611, 40 U μl^{-1})
- Xylenes (Fisher Scientific #HC700-1GAL)

  ! CAUTION Xylenes are toxic and should only be used in a chemical fume hood.
- RNAse Away (Molecular Bioproducts #7003)
- Isopentane (Sigma #320404-1L)

  ! CAUTION Isopentane is highly flammable and should not be disposed of down the sink.
- Neg-50 embedding medium (Richard-Allan Scientific #6502)
- Nuclear fast red (Vector #H3403)
- SuperScript III (Invitrogen #18080-044, 200 U μl^{-1})

  CRITICAL This reverse transcriptase retains its activity at elevated temperatures and is compatible with the downstream steps in the protocol. Other reverse transcriptases cannot be substituted.
- 5× first-strand buffer (included with Invitrogen #18080-044)
- Nuclease-free H_2O (Promega #P119C)
- Proteinase K (Sigma #P2308)
- Saponin (TCI America #S0019)
- Anti-RNase (Ambion #AM2690, 20 U μl^{-1})

  CRITICAL RNAse inhibitors must be compatible with all of the downstream steps in the protocol and cannot be substituted with other inhibitors.
- SUPERase-In (Ambion #AM2694, 20 U μl^{-1})

  CRITICAL RNAse inhibitors must be compatible with all of the downstream steps in the protocol and cannot be substituted with other inhibitors.
- PMSF (Sigma #P7626)

CRITICAL PMSF comes in various purities depending on the vendor, and the procedure has been optimized with this supplier.

- 100 mM dATP, dCTP, dGTP, dTTP (dNTP set: Roche #11277049001)
- Oligo(dT)$_{24}$ (25 nmol synthesis from Invitrogen)
- RNAse H (New England Biolabs #M0297L, 5 U μl^{-1})
- 25 mM $MgCl_2$ (Applied Biosystems #H05143)
- Terminal transferase buffer (Invitrogen #16314-015)

  ! CAUTION The terminal transferase buffer contains Co^{2+}, which is toxic if ingested or inhaled. Use appropriate precautions.
- Terminal transferase (Roche #03333574001, 400 U μl^{-1})
- 10× ThermoPol buffer (New England Biolabs #B9005S)
- 100 mM $MgSO_4$ (New England Biolabs #B1003S)
- BSA (Roche #10711454001)
- Roche Taq polymerase (Roche #04728858001, 5 U μl^{-1})
- AL1 primer (200 nmol synthesis from IDT or MWG):
  ATTGGATCCAGGCCGCTCTGGACAAAATATGAATTCTTTTTTTTTTTTTT TTTTTTT TTT

  CRITICAL Long desalted primers will have varying purities depending on the manufacturer.
- 100× SYBR Green I (Invitrogen #S7563 diluted 100-fold in DMSO to 100×)
- 50 mM aminoallyl-dUTP (Ambion #1103015)
- High-Fidelity polymerase (Roche #11732650001, 3.5 U μl^{-1})
- 10× High-Fidelity PCR buffer without Mg^{2+} (included with Roche #11759175001)
- PureLink PCR Purification Kit (Invitrogen #46-6056)
- NaOAc (Calbiochem #567418)
- 20 mg/ml glycogen (Invitrogen #10814-010)
- $NaHCO_3$ (Acros #21712500)
- Alexa Fluor 555 reactive dye decapack (Invitrogen #A32756)
- GEX hybridization buffer (included with Illumina #BD-103-0204 or equivalent)

EQUIPMENT

- Thermocycler (Bio-Rad MyCycler or equivalent)
- Cryostat (Leica CM1950)
- Microtome blades (Thermo Scientific #MX36 premier+)
- Pixcell II laser capture microdissection instrument (Arcturus)
- Inverted microscope (Olympus CKX41 or equivalent)
- Cell counter (Beckman Coulter #AM35308)
- Hemacytometer (Hausser Scientific #437757)

- SuperFrost Plus glass microscope slides (VWR #48311-703)
- 24 × 50 mm No. 1.5 glass coverslips (Fisher Scientific #12-544D)
- Coplin staining jars (Fisher #S90130)
- Cryomold (Sakura #4566)
- PrepStrip (Arcturus #LCM0207)
- Capsure HS LCM caps and ExtracSure adaptor (Arcturus #LCM0214)
- 0.5 ml thin-walled PCR tube (Applied Biosystems #N8010611)
- 0.2 ml thin-walled PCR tube (Applied Biosystems # N8010612)

  CRITICAL The amplification procedure has only been validated with these PCR tubes.
- PCR adaptors (Fisher #11-715-125D)
- qPCR instrument (Bio-Rad CFX 96 Real-Time System or equivalent)
- 96-well qPCR plates and optically clear film
- NanoDrop (NanoDrop Technologies #ND-1000)
- Expression BeadChip (Illumina #BD-103-0204 or equivalent)
- BeadArray reader (Illumina)
- MATLAB software (MathWorks)

REAGENT SETUP

20 mg ml^{-1} proteinase K—Prepare a 20 mg ml^{-1} solution in nuclease-free H_2O and store in 20 μl aliquots at −20 °C for up to six months. After thawing, keep at 4 °C for up to one month.

100 mM PMSF—Prepare a 17.42 mg ml^{-1} solution in 100% (vol/vol) ethanol shortly before use.

25× stock primer mix—Prepare a solution containing 15 μl nuclease-free H_2O, 5 μl 100 mM dATP, 5 μl 100 mM dCTP, 5 μl 100 mM dGTP, 5 μl 100 mM dTTP, and 5 μl 80 OD ml^{-1} oligo(dT)$_{24}$. Store in 5 μl aliquots at −20 °C for up to six months.

RNAse H–Mg^{2+} mix—Prepare a solution containing 5 μl RNAse H (5 U μl^{-1}) and 5 μl 25 mM $MgCl_2$. Mix on ice and use immediately.

2.6× tailing buffer—Prepare a solution containing 363 μl nuclease-free H_2O, 400 μl 5× Invitrogen terminal transferase buffer, and 15 μl 100 mM dATP. Store in 100 μl aliquots at −20 °C for up to one year. Add 0.2 μl of terminal transferase (400 U μl^{-1}) per 3.5 μl 2.6× tailing buffer immediately before use.

CRITICAL Do not use the Roche 5× TdT reaction buffer that come with the terminal transferase. This buffer lacks the Co^{2+} cofactor that is important for transferase activity.

! CAUTION Co^{2+} is toxic if ingested or inhaled. Use appropriate precautions.

Saponin-proteinase K solution—Prepare a 25 mg ml^{-1} proteinase K and 1% (wt/vol) saponin solution in nuclease-free H_2O immediately before use.

15 μg μl^{-1} AL1 primer—Prepare a 15 μg μl^{-1} solution in nuclease-free H_2O and store in 10 μl aliquots at −20 °C for up to one year.

1 M $NaHCO_3$—Dissolve 12.6 g of $NaHCO_3$ in 100 mL of H_2O. Adjust the volume to 150 mL to yield a final concentration of 1 M. Filter-sterilize the solution and store at room temperature (22 °C) for up to one month.

3 M NaOAc (pH 5.2)—Dissolve 408.3 g of sodium acetate-3 H_2O in 800 mL of H_2O. Adjust the pH to 5.2 with glacial acetic acid. Adjust the volume to 1 L with H_2O. Dispense into aliquots and sterilize by autoclaving. Store at room temperature for one year or more.

PROCEDURE

Embedding and cryosectioning ●TIMING 1 d

1 | This step can be performed using option A, option B, or option C depending on whether tissue specimens, cultured adherent cells, or suspension cells are used.

A. *Tissue specimens.* Equilibrate a dry ice-isopentane bath in a plastic beaker.

! CAUTION Isopentane will bubble violently when first added to the dry ice. Wear safety goggles and gloves to avoid frostbite.

B. *Cultured adherent cells.* Skip to Step 10.

C. *Suspension cells.* Skip to Step 26.

2 | Place fresh or snap-frozen tissue into a small cryomold and cover with Neg-50 embedding medium

CRITICAL STEP Move quickly to minimize changes in RNA expression or integrity during the freezing process.

3 | Pick up the cryomold with large forceps and freeze the specimen on top of the dry ice-isopentane bath. Try not to submerge the cryomold so that the progress of the embedding can be monitored from above.

PAUSE POINT After the specimen is completely frozen, the sample can be stored on dry ice while additional samples are embedded. For long-term storage from months to years, wrap cryomolds in tinfoil and store at −80 °C.

! CAUTION Isopentane can be reused indefinitely and should not be disposed of down the sink.

4 | Transport embedded samples on dry ice to the cryostat, place samples and a slide rack in the cryostat box, and equilibrate the box temperature to −24 °C.

5 | Replace the microtome blade and carefully wipe the blade, cryostat platform, and anti-roll bar with a Kimwipe moistened with both ethanol and RNAse Away.

! CAUTION Be sure to wipe away from the direction of the microtome blade.

6 | Remove the sample from the cryomold and mount with Neg-50 on a cryostat chuck.

7 | Trim the sample and cut 8 μm sections using either the anti-roll bar or a small paintbrush.

8 | Wick sections onto slides and move slides immediately to the slide rack inside the cryostat box. Up to two sections can be wicked per slide.

CRITICAL STEP Each section must be frozen a quickly as possible after wicking to avoid RNA degradation. The second section must be cut quickly so that the slide is still warm enough to wick the second section.

9 | Move the slide box containing the frozen sections to dry ice and dispose of the remaining embedded block.

PAUSE POINT Frozen sections can be stored for months at −80 °C.

Staining and laser-capture microdissection ●TIMING 2 hr

10 | This step can be performed using option A or option B depending on whether frozen sections or cultured adherent cells are used.

A. *Frozen sections.* Remove four slides from −80 °C and place immediately in 75% (vol/vol) ethanol for 30 s.

CRITICAL STEP Bring a staining jar containing 75% (vol/vol) ethanol to the −80 °C freezer and immerse the slides before they thaw or accumulate excessive frost.

B. *Cultured adherent cells.* Plate cells on 24 × 50 mm coverslips as desired and then place immediately in 75% (vol/vol) ethanol for 30 s.

11 | Transfer slides to distilled water for 30 s.

CRITICAL STEP All aqueous staining steps should be followed precisely to maintain consistent RNA integrity. Reserve a set of new Coplin staining jars exclusively for laser-capture microdissection.

12 | Stain slides with a few drops of nuclear fast red containing 1 U ml^{-1} RNAsin Plus for 30 s. 100 μl nuclear fast red + 2.5 μl RNAsin Plus is sufficient for four coverslips or slides containing two sections per slide.

13 | Transfer slides to distilled water for 15 s (dip slide, remove slide, dip slide again).

14 | Repeat Step 13 with a second staining jar of distilled water.

15 | Dehydrate slides in 70% (vol/vol) ethanol for 30 s, 95% (vol/vol) ethanol for 30 s, and finally 100% (vol/vol) ethanol for 30 s.

16 | Clear slides in xylenes for 2 min.

! CAUTION Xylenes are toxic and should only be used in a chemical fume hood.

17 | Air dry slides for 5–10 min in a chemical fume hood. For cells cultured on coverslips, mount dried coverslips face up on glass microscope slides by using clear nail polish.

18 | Transport slides in a desiccator to the microdissector.

19 | Turn on the instrument and spray hands with RNAse Away

20 | Clear away loosely adherent tissue from the slide by gently pressing down a PrepStrip on the surface of the slide.

21 | Load Capsure HS LCM caps onto the instrument.

22 | Detach an LCM cap, focus the laser, and begin dissecting with the following laser settings: 0.175 V, 50–65 mW, 750 μs laser power. If the sample has been appropriately dehydrated, this laser power should allow good capture and resolution (1–2 cells per laser shot). Multiple shots are often required to cause polymer wetting at this laser power.

? TROUBLESHOOTING

23 | If there is extensive collateral pickup from adjacent nondissected cells, press the LCM cap lightly on an adhesive note.

CRITICAL STEP The weakest possible adhesive note should be used to avoid removing material from the microdissected cells.

24 | Load the LCM cap onto the ExtracSure adaptor included with the LCM caps and store upside down at room temperature.

PAUSE POINT The LCM cap can be stored for 1–2 h as Steps 21–24 are repeated with additional samples or random samplings.

25 | After completing all microdissections, proceed immediately to small-sample cDNA amplification.

Sample-specific cDNA amplification ●TIMING 11 hr

26 | This step can be performed using option A or option B depending on whether microdissected cells or suspension cells are used.

A. *Microdissected cells.* Add 4 μl of digestion buffer to the ExtracSure adaptor. Cover the ExtracSure adaptor with a 0.5 ml thin-walled PCR tube. Incubate both the LCM cap and the remaining digestion buffer at 42 °C for 1 h.

Prepare the digestion buffer as follows:

Reagent	Volume
5× first-strand buffer	20 μl
1× stock primer mix	2 μl
20 mg ml^{-1} proteinase K	1 μl
Nuclease-free H_2O	57 μl
Total volume	80 μl

CRITICAL STEP Make sure to cover the adaptors tightly with PCR tubes, but do not crimp the PCR tube or dislodge the ExtracSure adaptor.

B. *Suspension cells.* Resuspend cells in pre-digestion buffer lacking proteinase K.

Prepare the pre-digestion buffer as follows:

Reagent	Volume
5× first-strand buffer	22 μl
1× stock primer mix	2.2 μl
Nuclease-free H_2O	55.8 μl
Total volume	80 μl

Then spike in 1/10th volume saponin-proteinase K solution to make 1× digestion buffer containing 0.1% (wt/vol) saponin for cell digestion and RNA extraction. Incubate both the cell suspension and the remaining digestion buffer at 42 °C for 1 hr.

PAUSE POINT The cells suspended in pre-digestion buffer can be stored in −80 °C without loss of amplification efficiency.

CRITICAL STEP The remaining digestion buffer must be treated identically as the buffer in contact with cells so that the proteinase K partially inactivates itself and the buffer can be used for dilution of more-concentrated samples after RNA elution.

27 | Spin tubes for 2 min at 2500 × g on a benchtop centrifuge at room temperature.

28 | Prepare the digestion stop buffer as follows:

Reagent	Volume
20 U μl^{-1} Anti-RNase	1 μl
20 U μl^{-1} SUPERase-In	1 μl
100 mM PMSF	1 μl
Nuclease-free H_2O	17 μl
Total volume	20 μl

CRITICAL STEP Be sure that the PMSF is added right before Step 29, or it will precipitate over time in the digestion stop buffer.

29 | Immediately add 1 μl of digestion stop buffer to each sample and mix by pipetting. Vortex and centrifuge briefly.

30 | If performing a serial or replicate dilution of a more-concentrated sample, dilute the sample with digestion buffer + digestion stop buffer, mixed at a 4:1 ratio immediately beforehand.

CRITICAL STEP The digestion buffer must be incubated 1 hr at 42 °C as described in Step 26.

31 | Transfer 4.5 μl of each sample to a 0.2 ml thin-walled PCR tube and place the tubes on ice.

32 | Prepare a blank tube for the amplification by adding 4 μl of digestion buffer + digestion stop buffer and 0.5 μl of nuclease-free water to a 0.2 ml thin-walled PCR tube. Heat-denature the blank sample at 65 °C for 1 min and allow to cool at room temperature for 90 s. Spin for 2 min at 12,000 × g on a benchtop centrifuge at 4 °C to collect condensation within the tube.

CRITICAL STEP Perform all heating and incubation steps as protocols on a PCR thermocycler to ensure temperature accuracy and stability.

33 | Add 0.5 □ of SuperScript III to each sample, vortex briefly, and incubate at 50 °C for 15 min.

CRITICAL STEP Make sure that the SuperScript III is well mixed within the sample, but do not allow the solution to flick up from the base of the tube.

34 | Heat-inactivate the enzyme by incubating at 70 °C for 15 min.

35 | Place samples on ice and spin for 2 min at 12,000 × g on a benchtop centrifuge at 4 °C to collect condensation within the tube.

36 | Add 1 □ of RNAse H–Mg^{2+} mix and incubate at 37 °C for 15 min.

37 | Place samples on ice and spin for 2 min at 12,000 × g on a benchtop centrifuge at 4 °C to collect condensation within the tube.

38 | Add 3.5 □ of 2.6× tailing buffer containing 0.2 □ of 400 U □$^{-1}$ terminal transferase per sample and incubate at 37 °C for 15 min.

39 | Heat-inactivate the enzyme by incubating at 65 °C for 10 min.

40 | Place samples on ice and spin for 2 min at 12,000 × g on a benchtop centrifuge at 4 °C to collect condensation within the tube.

41 | Prepare the ThermoPol PCR buffer as follows:

Reagent	Volume
10× ThermoPol buffer	10 □
100 mM $MgSO_4$	2.5 □
20 mg ml^{-1} BSA	0.5 □
100 mM dNTP	1 □
15 U □$^{-1}$ RocheTaq polymerase	2 □
15 □g □-1AL1	0.3–3 □
Nuclease-free H_2O	68–70.7 □
Total volume	90 □

Add 90 □ of ThermoPol PCR buffer and split sample into three 33- □ aliquots: add 33 □ to two new 0.2 ml thin-walled PCR tubes and use the remainder as the third aliquot.

CRITICAL STEP If optimizing for a new biological format, the amount of primer in the ThermoPol PCR buffer should be varied in five replicate 100-cell samples that include 5, 10, 15, 25, or 50 □g AL1.

42 | On a thermocycler containing a heated lid, run the following PCR amplification protocol:

Cycle	Denaturation	Annealing	Extension
1–4	94 °C for 1 min	32 °C for 2 min	72 °C for 6 min with 10 sec increase at each cycle

Cycle	Denaturation	Annealing	Extension
5–25	94 °C for 1 min	42 °C for 2 min	72 °C for 6 min 40 s with 10 sec increase at each cycle

CRITICAL STEP If optimizing for a new biological format, collect a 10 μl fraction of the PCR reaction after cycle 25.

43 | Cool to 4 °C and move the samples to ice. Pool the three splits into the original 0.2 ml PCR tube, vortex and centrifuge briefly.

44 | On a thermocycler containing a heated lid, run the following PCR amplification protocol:

Cycle	Denaturation	Annealing	Extension
26–30	94 °C for 1 min	42 °C for 2 min	72 °C for 6 min
31–35	94 °C for 1 min	42 °C for 2 min	72 °C for 6 min
35–40	94 °C for 1 min	42 °C for 2 min	72 °C for 6 min

CRITICAL STEP If optimizing for a new biological format, collect a 10 μl fraction of the PCR reaction after cycles 30, 35, and 40. If the protocol has already been optimized, simply stop the entire reaction at the optimal number of PCR cycles.

45 | Cool to 4 °C and move the samples to ice.

PAUSE POINT Samples can be frozen at −20 °C for months to years and can be thawed several times without noticeable degradation of the amplification products.

46 | Dilute 1 μl of each amplified cDNA sample 450–500-fold in water and measure genes of interest by qPCR.

CRITICAL STEP qPCR amplicons must target within ~500 bp of the 3′ end of the transcript to be detected reliably because of the abbreviated RT in Step 33.

? TROUBLESHOOTING

cDNA reamplification and labeling ●TIMING 2–3 d

47 | Prepare the following master mix (for 10 reactions) to perform a pilot reamplification:

Reagent	Volume
10× High-Fidelity PCR buffer without Mg^{2+}	20 μl
25 mM $MgCl_2$	28 μl
100 mM dATP	0.4 μl
100 mM dCTP	0.4 μl
100 mM dGTP	0.4 μl
50 mM aminoallyl-dUTP	0.64 μl
10 mM dTTP	0.8 μl

Reagent	Volume
20 mg ml^{-1} BSA	1 μl
15 μg μl^{-1} AL1 primer	0.6 μl
3.5 U μl^{-1} High-Fidelity polymerase	2 μl
100× SYBR green	0.5 μl
PCR-grade water	135.3 μl
Total volume	190 μl

CRITICAL STEP Note that the dTTP stock is at a 10-fold lower concentration than the other dNTPs.

48 | Add 0.5 μl of each amplified cDNA sample to 2 μl of PCR-grade water, add 1 μl of this diluted cDNA to 19 μl of the reamplification master mix, and monitor the amplification products by qPCR with the following amplification protocol:

Cycle	Denaturation	Annealing	Extension
1–40	94 °C for 1 min	42 °C for 2 min	72 °C for 3 min, then measure fluorescence

CRITICAL STEP The goal of the pilot amplification is to identify the amplification cycle where the first sample hits the middle of its exponential phase. Running the amplification later into the exponential phase will overamplify the sample and cause loss of quantitative accuracy.**?**

TROUBLESHOOTING

49 | Prepare the aminoallyl-cDNA (aa-cDNA) reamplification reaction as follows:

Reagent	Volume
10× High-Fidelity PCR buffer without Mg^{2+}	10 μl
25 mM $MgCl_2$	14 μl
100 mM dATP	0.2 μl
100 mM dCTP	0.2 μl
100 mM dGTP	0.2 μl
50 mM aminoallyl dUTP	0.32 μl
10 mM dTTP	0.4 μl
20 mg ml^{-1} BSA	0.5 μl
15 μg μl^{-1} AL1 primer	0.3 μl
Amplified cDNA	1 μl
3.5 U μl^{-1} High-Fidelity polymerase	1 μl
PCR-grade water	71.9 μl
Total volume	100 μl

50 | On a thermocycler containing a heated lid, run the following PCR amplification protocol:

Cycle	Denaturation	Annealing	Extension
1–OPT	94 °C for 1 min	42 °C for 2 min	72 °C for 3 min

CRITICAL STEP The aa-cDNA reamplification is performed with the optimum number of PCR cycles (OPT) identified in Step 48.

51 | To each aa-cDNA sample, add 400 μl PureLink Binding Buffer.

CRITICAL STEP Do not use the High-Cutoff Binding Buffer included with the PureLink columns. This will cause the aa-cDNA to flow through the column.

52 | Apply the entire solution to a PureLink column and spin at 10,000 × g for 1 min.

53 | Discard flow through, wash column with 650 μl Wash Buffer, and spin at 10,000 × g for 1 min.

54 | Discard flow through, and spin again at 10,000 × g for 1 min.

55 | Transfer the column to a clean elution tube and add 50 μl elution buffer.

56 | Seal the cap on the column and incubate at 65 °C for 10 min.

CRITICAL STEP The high-temperature elution maximizes the yield off of the PureLink columns.

57 | Spin at 10,000 × g for 1 min.

58 | Add another 50 μl elution buffer, reseal the cap on the column, and incubate at 65 °C for 10 min.

59 | Spin at 10,000 × g for 1 min.

60 | Add 10 μl 3 M NaOAc (pH 5.2) and 1 μl 20 mg ml^{-1} glycogen. Vortex.

61 | Add 250 μl ice-cold 100% (vol/vol) ethanol, vortex, and incubate the samples at −20 °C for at least 30 min.

PAUSE POINT Samples can be frozen at −20 °C overnight if needed.

62 | Spin for 10 min at max speed on a benchtop centrifuge.

63 | Carefully aspirate the supernatant and wash the pellet with 500 μl 70% (vol/vol) ethanol at room temperature.

64 | Spin for 1 min at max speed on a benchtop centrifuge.

65 | Carefully aspirate the supernatant and wash the pellet with 500 μl 70% (vol/vol) ethanol at room temperature.

CRITICAL STEP The second wash is important to remove trace amines from the precipitated aa-cDNA for maximal labeling efficiency afterwards.

66 | Spin for 1 min at max speed on a benchtop centrifuge.

67 | Carefully aspirate the supernatant and remove residual ethanol by hand with a pipette tip.

68 | Air dry the aa-cDNA pellets for 5–10 min at room temperature.

69 | Resuspend the pellet in 5 μl nuclease-free water and incubate for 15 min at 37 °C to redissolve.

70 | Determine the aa-cDNA concentration by A_{260} spectrophotometry on a NanoDrop.

PAUSE POINT Samples can be frozen at −20 °C for months to years and can be thawed several times without noticeable degradation of the aa-cDNA.

71 | Mix 1 μg aa-cDNA and 3 μl of 1 M $NaHCO_3$ in a total volume of 8 μl.

72 | For each labeling reaction, dissolve one vial of Alexa Fluor 555 succinimidyl ester dye from the decapack in 2 μl DMSO.

CRITICAL STEP Add DMSO to the side of each tube, then spin down all the tubes together to minimize the time the dye is sitting in DMSO outside of the reaction.

73 | Add 2 μl of the resuspended Alexa Fluor 555 dye to the mixture and vortex a max speed for 15 s.

CRITICAL STEP The lengthy vortexing time is critical to ensure high coupling efficiencies.

74 | Spin down the labeling reactions briefly and incubate for 1 h at room temperature.

75 | Add 10 μl 3 M NaOAc (pH 5.2) and 80 μl water to each labeling reaction.

CRITICAL STEP Neutralizing the pH allows for more efficient binding to the PureLink column in the subsequent purification.

76 | Repeat Steps 51–64 with the Alexa Fluor 555-labeled cDNA (555-cDNA).

77 | Carefully aspirate the supernatant and remove residual ethanol by hand with a pipette tip.

78 | Air dry the 555-cDNA pellets for 5–10 min at room temperature.

79 | Resuspend the pellet in 5 μl nuclease-free water and incubate for 15 min at 37 °C to redissolve.

80 | Determine the 555-cDNA concentration and degree of labeling by A_{260} and A_{555} spectrophotometry on a NanoDrop. The concentration and degree of labeling can be determined using the base:dye ratio calculator on the Invitrogen web site: http://probes.invitrogen.com/resources/calc/basedyeratio.html

PAUSE POINT Samples can be frozen at −20 °C for months without noticeable degradation of the 555-cDNA.

? TROUBLESHOOTING

Microarray hybridization and data analysis ●TIMING 2–3 d

81 | Mix 555-cDNA sample with 10 μl GEX hybridization buffer and denature at 94 °C for 4 min.

82 | Add samples directly to each lane of an Expression BeadChip prewarmed to 58 °C and incubate at 58 °C for 20 hr.

83 | Wash and dry slides according to the manufacturer's recommendation and scan on a BeadArray reader.

PAUSE POINT After scanning the Expression BeadChip, the analysis can be performed at any time thereafter.

84 | Export gene-probe name, fluorescence intensity, and detection *P* value for each lane as a tab-delimited ASCII text file.

85 | Download `StochProfMicroarrayFilt.m` and `StochProfAnalysis.m` into a directory recognized by the MATLAB path.

86 | On the MATLAB Command Window, type:

```
[Genes, Samples, StochSamplings, ControlSamplings] =
StochProfMicroarrayFilt;
```

? TROUBLESHOOTING

87 | Follow the prompts to select the ASCII text file of random samplings (first prompt) and the ASCII text file of amplification controls (second prompt). Set the median detection *P* value to 0.1 and the maximum fold-change threshold to 5.

88 | Save the workspace containing the filtered microarray data by typing the following on the MATLAB Command Window:

```
save('FilteredMicroarrays')
```

The filtered microarray data can now be recovered to the workspace at any time by typing:

```
load FilteredMicroarrays
```

89 | On the MATLAB Command Window, type:

```
[HetGenes, HetData, HetGenesPval]=StochProfAnalysis(Genes, Samples,
StochSamplings, ControlSamplings);
```

90 | Follow the prompts to select FDR_{var}, CV_{ref}, and FDR_{het}. Display heterogeneous expression patterns as a clustergram if desired. The predicted heterogeneous transcript names are now stored as `HetGenes`, the corresponding sampling data are stored as `HetData`, and the exact *P* values for the K-S test of these data against a log-normal distribution are stored as `HetGenesPval`.

91 | To test additional values of FDR_{var}, CV_{ref}, and FDR_{het}, type:

```
clear all

close all

load FilteredMicroarrays
```

and return to Step 89.

●TIMING

Steps 1–9, embedding and cryosectioning, 1 d

Steps 10–25, staining and laser-capture microdissection, 2 hr

Steps 26–46, sample-specific cDNA amplification, 11 hr

Steps 47–80, cDNA reamplification and labeling, 2–3 d

Steps 81–91, microarray hybridization and data analysis, 2–3 d

ANTICIPATED RESULTS

The rapid histology protocol should give faint pink nuclear staining in cells and tissue sections, which is easily identified during microdissection (Fig. 2). For small-sample cDNA amplification, a reasonably clear optimum should exist for AL1 primer amount and cycle number (Fig. 4). By following the two-step optimization procedure, we have identified conditions for microdissected primary melanoma cells (Fig. 7a and Supplementary Fig. 2), HT-29 colon adenocarcinoma cells microdissected off of coverslips (Fig. 7b and Supplementary Fig. 3), and SKW 6.4 lymphoblastoid suspension cells isolated by limiting dilution (Fig. 7c and Supplementary Fig. 4). With microdissected samples, quantitative accuracy and reproducibility are usually lost with one-cell equivalents of starting material[56] (Fig. 7a,b). This reemphasizes the importance of the random 10-cell sampling approach for microdissected tissue. Interestingly, one-cell measurements are possible with suspension cells (Fig. 7c and Supplementary Fig. 1), which is consistent with earlier results from single cells obtained by micropipette aspiration or FACS[34-40]. Based on simulations, the overall reproducibility of 10-cell amplification replicates must be within 35%, because background biological variation will only amplify this error and ultimately can give rise to false negatives (Fig. 1d).

During reamplification and labeling, it is not uncommon to see some spread in the total cDNA levels on a sampling-to-sampling basis (Fig. 5a, yellow). This reflects global differences in the extent of mRNA extraction from the microdissection cap. For qPCR, the differences can be accounted for with a panel of loading-control genes[92]. For microarrays, it is better to perform replicate reamplifications of low-abundance samples and pool before labeling (see TROUBLESHOOTING). When labeling aa-cDNA, we typically observe a degree of labeling of ~1.5 Alexa Fluor 555 dye molecules per 100 bases (Fig. 5b). The yield of 555-cDNA should be very close to 100% relative to the input aa-cDNA provided that the modified elution protocol is used with the PureLink columns.

Each microarray sample should detect a comparable number of genes to that obtained by conventional methods (typically 7000–10,000 genes depending on the platform). After running the `StochProfMicroarrayFilt.m` algorithm, at least half of the detected transcripts on the array should be measured with sufficient reproducibility for analysis[56]. The extent of cell-to-cell heterogeneities identified by `StochProfAnalysis.m` can vary widely depending upon the biological context and the exact analysis parameters (Fig. 6c). When a clonal cell line was globally profiled in 3D culture[56], we found that 10–20% of transcripts were predicted to be heterogeneously expressed. Conceivably, this percentage could be substantially higher when considering a population of cells that is actively proliferating (Fig. 7b,c) or genomically unstable (Fig. 7a,b). Regardless of the exact numbers, stochastic profiling provides a general method for uncovering cell-to-cell heterogeneities in a variety of biological settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Sameer Bajikar for critically reading this manuscript, Craig Slingluff for providing the primary melanoma sample, Byong Kang for help with screening low-abundance genes, and Cheryl Borgman for help with frozen sectioning. This work was supported by the National Institutes of Health Director's New Innovator Award

Program (1-DP2-OD006464), the American Cancer Society (120668-RSG-11-047-01-DMC), the Pew Scholars Program in the Biomedical Sciences, and the David and Lucile Packard Foundation.

References

1. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135:216–226. [PubMed: 18957198]
2. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? Cell. 2010; 141:559–563. [PubMed: 20478246]
3. Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol. 2011; 12:119–125. [PubMed: 21224886]
4. Losick R, Desplan C. Stochasticity and cell fate. Science. 2008; 320:65–68. [PubMed: 18388284]
5. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002; 297:1183–1186. [PubMed: 12183631]
6. Snijder B, et al. Population context determines cell-to-cell variability in endocytosis and virus infection. Nature. 2009; 461:520–523. [PubMed: 19710653]
7. Snijder B, et al. Single-cell analysis of population context advances RNAi screening at multiple levels. Mol Syst Biol. 2012; 8:579. [PubMed: 22531119]
8. Yu J, Xiao J, Ren X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. Science. 2006; 311:1600–1603. [PubMed: 16543458]
9. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005; 123:1025–1036. [PubMed: 16360033]
10. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. Nat Genet. 2002; 31:69–73. [PubMed: 11967532]
11. Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006; 440:358–362. [PubMed: 16541077]
12. Huh D, Paulsson J. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat Genet. 2011; 43:95–100. [PubMed: 21186354]
13. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression. PLoS Biol. 2004; 2:e137. [PubMed: 15124029]
14. Lestas I, Vinnicombe G, Paulsson J. Fundamental limits on the suppression of molecular fluctuations. Nature. 2010; 467:174–178. [PubMed: 20829788]
15. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009; 459:428–432. [PubMed: 19363473]
16. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008; 453:544–547. [PubMed: 18497826]
17. Sharma SV, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010; 141:69–80. [PubMed: 20371346]
18. Gupta PB, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011; 146:633–644. [PubMed: 21854987]
19. Xie XS, Yu J, Yang WY. Living cells as test tubes. Science. 2006; 312:228–230. [PubMed: 16614211]
20. Pelkmans L. Cell Biology. Using cell-to-cell variability--a new era in molecular biology. Science. 2012; 336:425–426. [PubMed: 22539709]
21. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. Science. 2012; 336:183–187. [PubMed: 22499939]
22. Wang D, Bodovitz S. Single cell analysis: the new frontier in ‘omics’. Trends Biotechnol. 2010; 28:281–290. [PubMed: 20434785]
23. Navin N, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011; 472:90–94. [PubMed: 21399628]
24. Gerlinger M, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012; 366:883–892. [PubMed: 22397650]

25. Hou Y, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell. 2012; 148:873–885. [PubMed: 22385957]
26. Xu X, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell. 2012; 148:886–895. [PubMed: 22385958]
27. Uhlen M, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010; 28:1248–1250. [PubMed: 21139605]
28. Bendall SC, et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science. 2011; 332:687–696. [PubMed: 21551058]
29. Shi Q, et al. Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. Proc Natl Acad Sci U S A. 2012; 109:419–424. [PubMed: 22203961]
30. Cohen AA, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science. 2008; 322:1511–1516. [PubMed: 19023046]
31. Iscove NN, et al. Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. Nat Biotechnol. 2002; 20:940–943. [PubMed: 12172558]
32. Tietjen I, et al. Single-cell transcriptional analysis of neuronal progenitors. Neuron. 2003; 38:161–175. [PubMed: 12718852]
33. Chiang MK, Melton DA. Single-cell transcript analysis of pancreas development. Dev Cell. 2003; 4:383–393. [PubMed: 12636919]
34. Sanchez-Freire V, Ebert AD, Kalisky T, Quake SR, Wu JC. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. Nat Protoc. 2012; 7:829–838. [PubMed: 22481529]
35. Citri A, Pang ZP, Sudhof TC, Wernig M, Malenka RC. Comprehensive qPCR profiling of gene expression in single neuronal cells. Nat Protoc. 2012; 7:118–127. [PubMed: 22193304]
36. Tang F, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc. 2010; 5:516–535. [PubMed: 20203668]
37. Esumi S, Kaneko R, Kawamura Y, Yagi T. Split single-cell RT-PCR analysis of Purkinje cells. Nat Protoc. 2006; 1:2143–2151. [PubMed: 17487206]
38. Kurimoto K, Yabuta Y, Ohinata Y, Saitou M. Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. Nat Protoc. 2007; 2:739–752. [PubMed: 17406636]
39. Tang F, et al. 220-plex microRNA expression profile of a single cell. Nat Protoc. 2006; 1:1154–1159. [PubMed: 17406397]
40. Hartmann CH, Klein CA. Gene expression profiling of single cells on large-scale oligonucleotide arrays. Nucleic Acids Res. 2006; 34:e143. [PubMed: 17071717]
41. Dalerba P, et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 2011; 29:1120–1127. [PubMed: 22081019]
42. Klein CA, et al. Combined transcriptome and genome analysis of single micrometastatic cells. Nat Biotechnol. 2002; 20:387–392. [PubMed: 11923846]
43. Klein CA, Zohlnhofer D, Petat-Dutter K, Wendler N. Gene expression analysis of a single or few cells. Curr Protoc Hum Genet. 2005:18. **Chapter 11**, Unit 11. [PubMed: 18428383]
44. Trimarchi JM, Stadler MB, Cepko CL. Individual retinal progenitor cells display extensive heterogeneity of gene expression. PLoS One. 2008; 3:e1588. [PubMed: 18270576]
45. Trimarchi JM, et al. Molecular heterogeneity of developing retinal ganglion and amacrine cells revealed through single cell gene expression profiling. J Comp Neurol. 2007; 502:1047–1065. [PubMed: 17444492]
46. Kurimoto K, et al. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008; 22:1617–1635. [PubMed: 18559478]
47. Tang F, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009; 6:377–382. [PubMed: 19349980]
48. Ramos CA, et al. Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. PLoS Genet. 2006; 2:e159. [PubMed: 17009876]

49. Bahar R, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature. 2006; 441:1011–1014. [PubMed: 16791200]
50. Taniguchi K, Kajiyama T, Kambara H. Quantitative analysis of gene expression in a single cell by qPCR. Nat Methods. 2009; 6:503–506. [PubMed: 19525960]
51. Reiter M, et al. Quantification noise in single cell experiments. Nucleic Acids Res. 2011; 39:e124. [PubMed: 21745823]
52. Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M. Properties of the reverse transcription reaction in mRNA quantification. Clin Chem. 2004; 50:509–515. [PubMed: 14726469]
53. Zhao B, et al. Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. Genes Dev. 2012; 26:54–68. [PubMed: 22215811]
54. Geller SF, Lewis GP, Fisher SK. FGFR1, signaling, and AP-1 expression after retinal detachment: reactive Muller and RPE cells. Invest Ophthalmol Vis Sci. 2001; 42:1363–1369. [PubMed: 11328752]
55. Shibata Y, Nakamura H, Kato S, Tomoike H. Cellular detachment and deformation induce IL-8 gene expression in human bronchial epithelial cells. J Immunol. 1996; 156:772–777. [PubMed: 8543832]
56. Janes KA, Wang CC, Holmberg KJ, Cabral K, Brugge JS. Identifying single-cell molecular programs by stochastic profiling. Nat Methods. 2010; 7:311–317. [PubMed: 20228812]
57. Emmert-Buck MR, et al. Laser capture microdissection. Science. 1996; 274:998–1001. [PubMed: 8875945]
58. Bengtsson M, Stahlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res. 2005; 15:1388–1392. [PubMed: 16204192]
59. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: Keys and clues. Bioscience. 2001; 51:341–352.
60. Kamme F, et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. J Neurosci. 2003; 23:3607–3615. [PubMed: 12736331]
61. Wang L, Brugge JS, Janes KA. Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. Proc Natl Acad Sci U S A. 2011; 108:E803–812. [PubMed: 21873240]
62. Janes KA. RUNX1 and its understudied role in breast cancer. Cell Cycle. 2011; 10:3461–3465. [PubMed: 22024923]
63. Wang CC, Jamal L, Janes KA. Normal morphogenesis of epithelial tissues and progression of epithelial tumors. Wiley Interdiscip Rev Syst Biol Med. 2012; 4:51–78. [PubMed: 21898857]
64. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994; 2:28–36. [PubMed: 7584402]
65. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007; 8:R24. [PubMed: 17324271]
66. Banerji S, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–409. [PubMed: 22722202]
67. Ellis MJ, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature. 2012; 486:353–360. [PubMed: 22722193]
68. Stewart-Ornstein J, Weissman JS, El-Samad H. Cellular noise regulons underlie fluctuations in Saccharomyces cerevisiae. Mol Cell. 2012; 45:483–493. [PubMed: 22365828]
69. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. Nature. 2010; 463:913–918. [PubMed: 20164922]
70. Eldar A, et al. Partial penetrance facilitates developmental evolution in bacteria. Nature. 2009; 460:510–514. [PubMed: 19578359]
71. O'Neill RA, et al. Isoelectric focusing technology quantifies protein signaling in 25 cells. Proc Natl Acad Sci U S A. 2006; 103:16153–16158. [PubMed: 17053065]
72. Jain A, et al. Probing cellular protein complexes using single-molecule pull-down. Nature. 2011; 473:484–488. [PubMed: 21614075]

73. Rapkiewicz A, et al. The needle in the haystack: application of breast fine-needle aspirate samples to quantitative protein microarray technology. Cancer. 2007; 111:173–184. [PubMed: 17487852]
74. Adli M, Bernstein BE. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. Nat Protoc. 2011; 6:1656–1668. [PubMed: 21959244]
75. Wu AR, et al. Automated microfluidic chromatin immunoprecipitation from 2,000 cells. Lab Chip. 2009; 9:1365–1370. [PubMed: 19417902]
76. Espina V, et al. Laser-capture microdissection. Nat Protoc. 2006; 1:586–603. [PubMed: 17406286]
77. Burgemeister R, Gangnus R, Haar B, Schutze K, Sauer U. High quality RNA retrieved from samples obtained by using LMPC (laser microdissection and pressure catapulting) technology. Pathol Res Pract. 2003; 199:431–436. [PubMed: 12924446]
78. Wang H, et al. Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. BMC Genomics. 2006; 7:97. [PubMed: 16643667]
79. Brady G, Iscove NN. Construction of cDNA libraries from single cells. Methods Enzymol. 1993; 225:611–623. [PubMed: 8231874]
80. Miller-Jensen K, Janes KA, Brugge JS, Lauffenburger DA. Common effector processing mediates cell-specific responses to stimuli. Nature. 2007; 448:604–608. [PubMed: 17637676]
81. Nagy ZB, et al. Real-time polymerase chain reaction-based exponential sample amplification for microarray gene expression profiling. Anal Biochem. 2005; 337:76–83. [PubMed: 15649378]
82. Kurimoto K, et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. 2006; 34:e42. [PubMed: 16547197]
83. Cox WG, Beaudet MP, Agnew JY, Ruth JL. Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. Anal Biochem. 2004; 331:243–254. [PubMed: 15265729]
84. Shi L, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol. 2006; 24:1151–1161. [PubMed: 16964229]
85. Warren L, Bryder D, Weissman IL, Quake SR. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc Natl Acad Sci U S A. 2006; 103:17807–17812. [PubMed: 17098862]
86. McKay AT. Distribution of the Coefficient of Variation and the Extended '*t*' Distribution. J Roy Stat Soc. 1932; 95:695–698.
87. Sokal, RR.; Rohlf, FJ. Biometry. Edn. 4th.. W.H. Freeman and Company; New York: 2012.
88. Sheskin, DJ. Handbook of Parametric and Nonparametric Statistical Procedures. Edn. 4th.. Chapman & Hall; New York: 2007.
89. Kaelin WG Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. Science. 2012; 337:421–422. [PubMed: 22837515]
90. Raj A, van Oudenaarden A. Single-molecule approaches to stochastic gene expression. Annu Rev Biophys. 2009; 38:255–270. [PubMed: 19416069]
91. Fend F, et al. Immuno-LCM: laser capture microdissection of immunostained frozen sections for mRNA analysis. Am J Pathol. 1999; 154:61–66. [PubMed: 9916919]
92. Vandesompele J, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002; 3 RESEARCH0034.

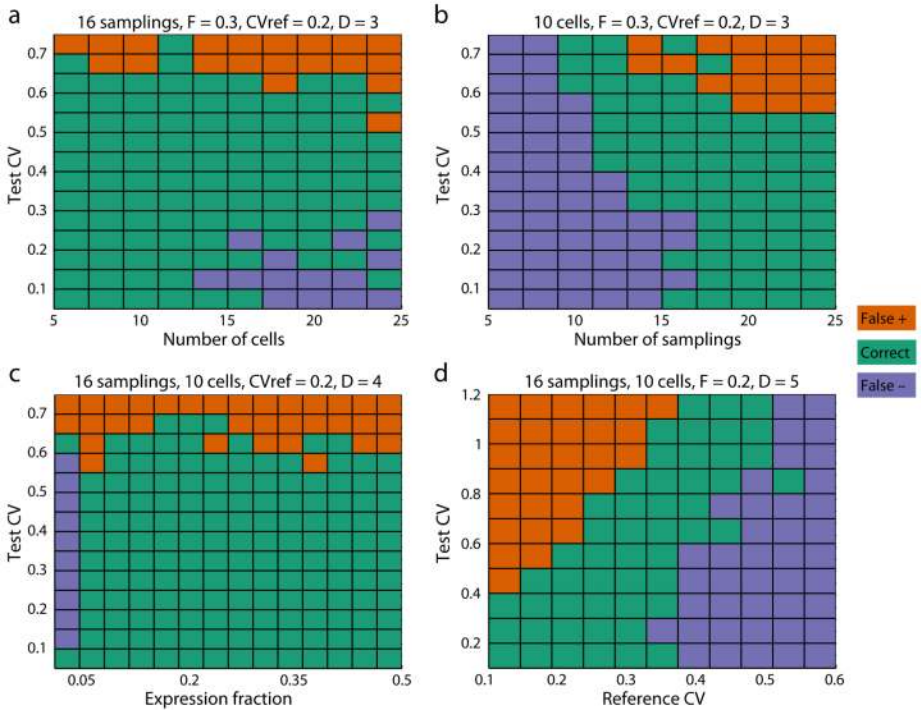

Figure 1.

Theoretical simulations of stochastic profiling under different parameter sets. **(a)** Stochastic profiling reliably uncovers a 20% expression dichotomy when samplings are comprised of fewer than 12 cells. **(b)** Stochastic profiling requires at least 15-20 samplings to work effectively. **(c)** Stochastic profiling works effectively for a wide range of expression fractions. **(d)** Stochastic profiling correctly distinguishes a 20% expression dichotomy when the reference CV (CV_{ref}) is less than 35% and the test CV (CV_{test}) is not more than ~threefold greater than CV_{ref}. Simulations were performed with the indicated parameter sets by using `StochProfParameters.m` in the Supplementary Software. False positives (defined as predicting a dichotomy in the test distribution when one does not exist) are shown in red, false negatives (defined as predicting no dichotomy in the test distribution when one exists) are shown in blue, and correct assignments are shown in green.

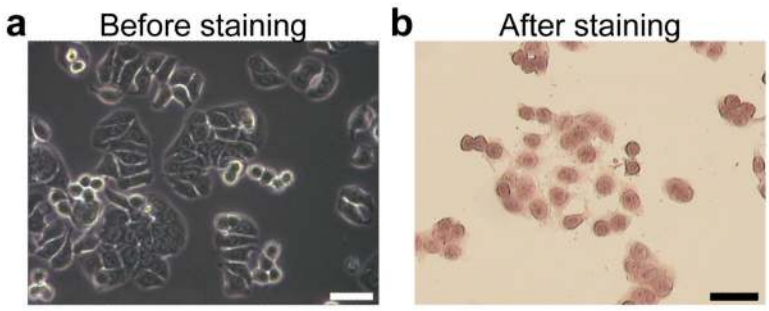

Figure 2.

Rapid nuclear fast red staining of HT-29 colon adenocarcinoma cells for laser capture microdissection. **(a)** HT-29 cells were plated directly on glass coverslips and imaged by phase-contrast microscopy. **(b)** Coverslips were fixed and stained with nuclear fast red as described in the PROCEDURE and imaged by brightfield microscopy. Scale bar = 20 μm.

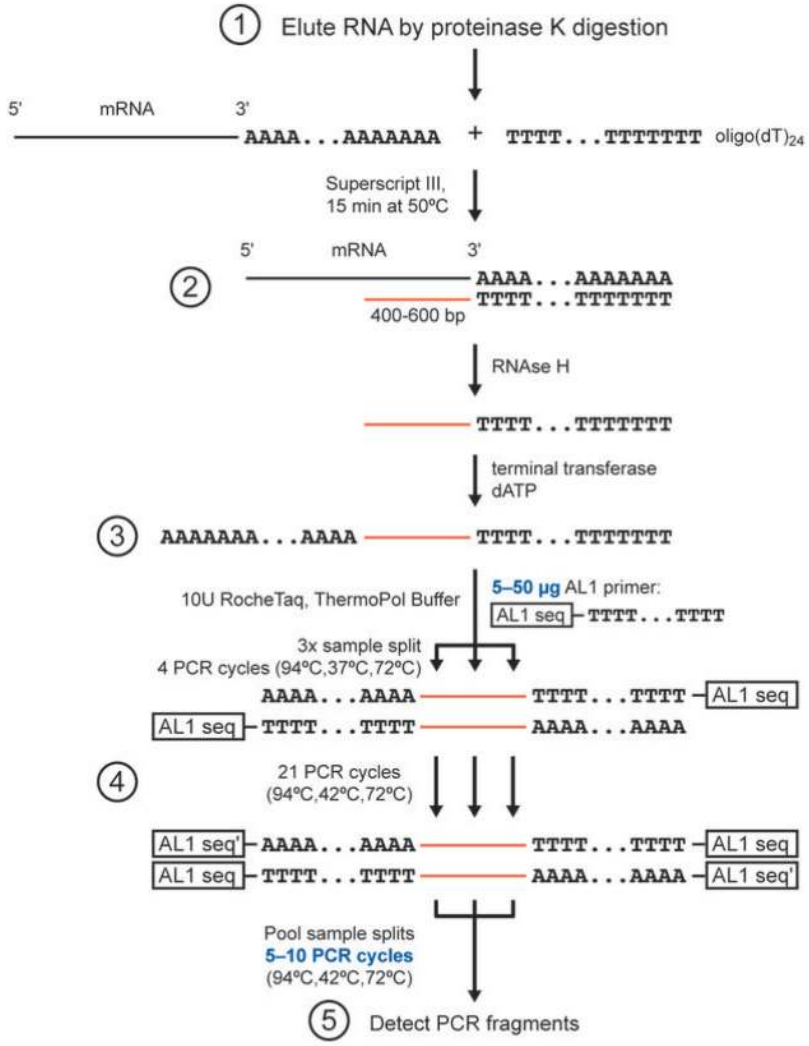

Figure 3.
Workflow for small-sample cDNA amplification. Adapted from [56] to include the optimization steps (blue) described in the text.

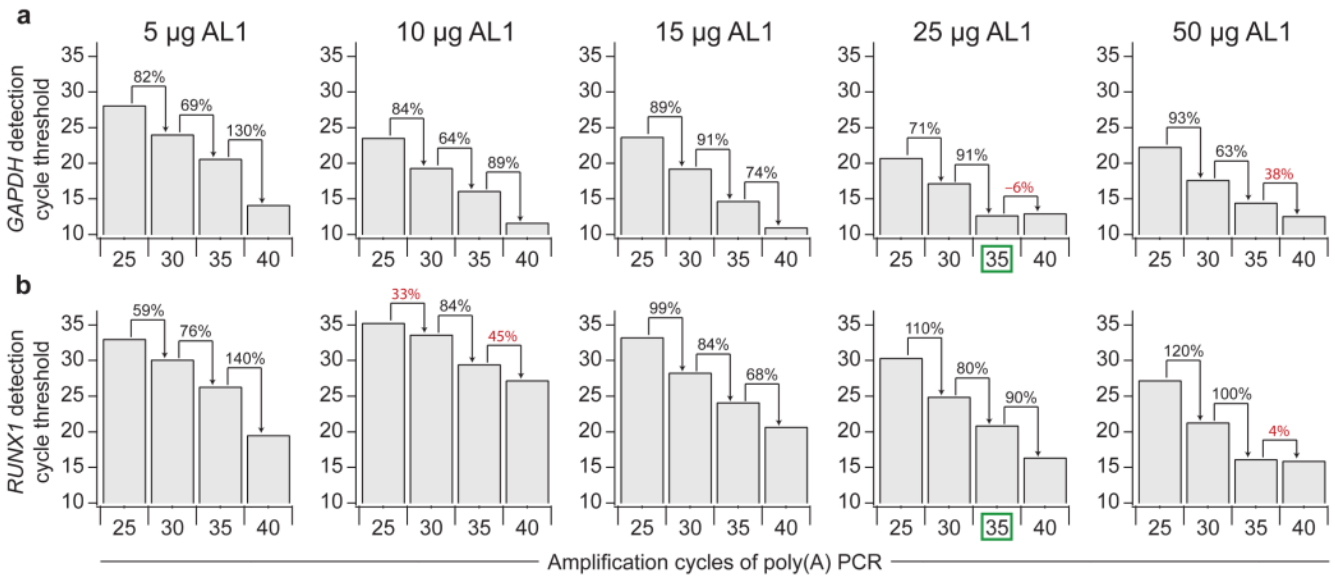

Figure 4.

Optimizing AL1 primer amount and poly(A) PCR cycle number for small-sample cDNA amplification. **(a,b)** Representative poly(A) PCR optimization for microdissected melanoma cells. 100-cell aliquots were amplified with the indicated AL1 primer amounts and total amplification cycles and (a) high-abundance genes such as *GAPDH* and (b) low-abundance genes such as *RUNX1* were profiled by qPCR. Efficiencies of gene amplification are shown between five-cycle amplification steps, and amplification efficiencies below 50% (red) were flagged as indications of suboptimal conditions or saturation of the amplicon. The condition leading to the highest concentration (lowest cycle number) of *GAPDH* and *RUNX1* with the fewest total amplification cycles was 25 □g AL1 for 35 cycles (green box).

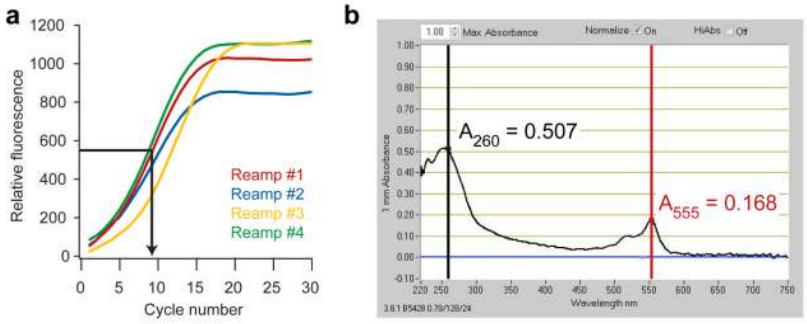

Figure 5.

Reamplification and labeling of 10-cell samples. **(a)** Representative reamplification of four 10-cell samplings monitored by real-time qPCR with SYBR Green fluorescence. The black arrow indicates the maximum number of reamplification cycles (nine). Note that Sample #3 (yellow) will give lower yield than the other reamplifications and will likely need to be reamplified in duplicate or triplicate before purification and labeling. **(b)** Representative 555-cDNA spectrum with the A_{260} and A_{555} peaks highlighted to calculate the extent of coupling. For this sample, the degree of labeling was 1.5 Alexa Fluor 555 molecules per 100 bases.

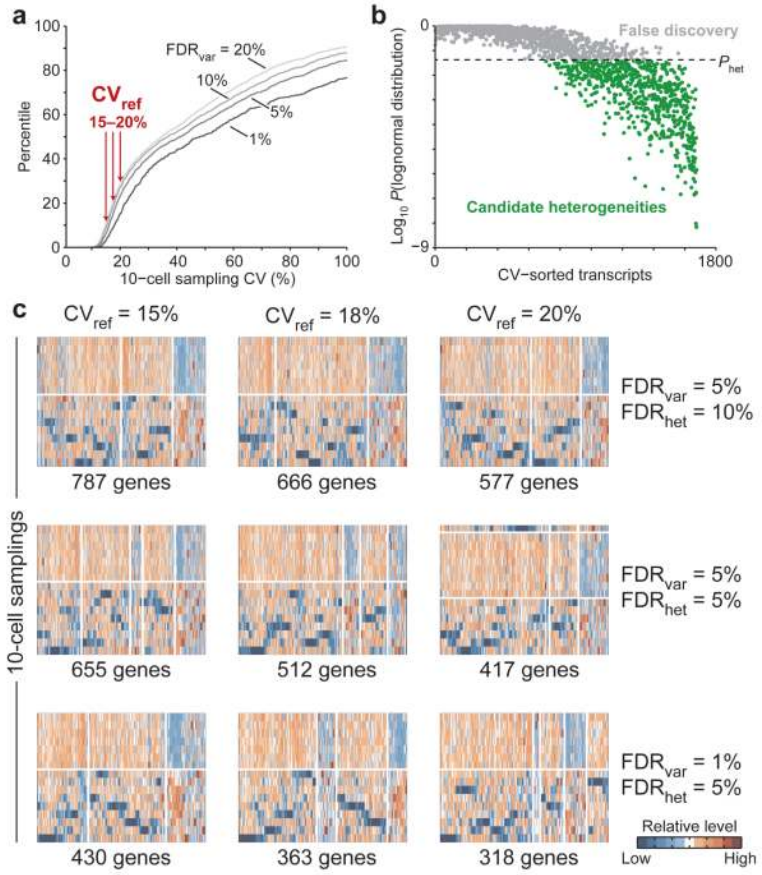

Figure 6.

Statistical and bioinformatic analysis of stochastic-profiling data. **(a)** Empirical cumulative distribution function of 10-cell sampling CVs for genes with significant biological variation at FDR_{var} = 20, 10, 5, and 1% (increasing gray). Candidate reference CVs (CV_{ref}) are highlighted in red. Note that the cumulative distribution function at FDR_{var} = 1% is substantially reduced compared to the others, suggesting overly stringent filtering at this step. **(b)** *P* values for the χ^2 goodness-of-fit test with the following parameters: FDR_{var} = 5%, CV_{ref} = 18%, FDR_{het} = 10%. Candidate heterogeneities (green) fall below the *P*-value threshold (P_{het}) adjusted based on FDR_{het}. **(c)** The number of predicted expression heterogeneities decreases with decreasing FDR_{var}, increasing CV_{ref}, and decreasing FDR_{het}, but the fundamental expression clustering does not change. The stochastic-profiling data from [56] was analyzed with the indicated parameters, and the resulting gene sets were hierarchically clustered with Ward's linkage. The major expression clusters are marked in white.

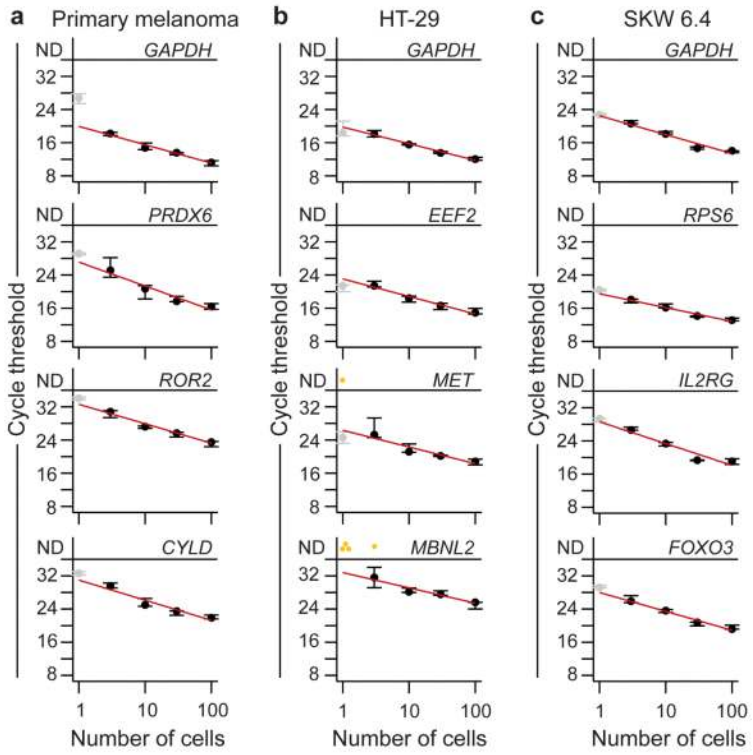

Figure 7.

Optimized small-sample cDNA amplification in three distinct biological contexts. **(a-c)** 100-cell samples were serially diluted and amplified by poly(A) PCR under optimal conditions for **(a)** microdissected primary melanoma cells (25 ng AL1, 35 cycles), **(b)** microdissected HT-29 colon adenocarcinoma cells (10 ng AL1, 35 cycles), and **(c)** SKW 6.4 lymphoblastoid suspension cells (50 ng AL1, 30 cycles). High- and low-abundance genes were monitored by qPCR, and data are shown as the median ± range of three replicate small-sample amplifications. Red lines show the log-linear fit of the 3–100-cell dilutions. Note that the one-cell amplifications (gray) often deviate from the log-linear fit or are frequently not detectable (yellow, ND).

TROUBLESHOOTING TABLE

Step	Problem	Possible reason	Solution
22	The polymer does not wet after many laser shots	Laser power is not high enough	Increase the laser power in 5 mW increments
	The tissue does not detach from the section after polymer wetting	Tissue is insufficiently dehydrated	Replace ethanol dehydration solutions. Increase the time of xylene clearing in Step 16
	Too much collateral tissue pickup during microdissection	Tissue is overly dehydrated	Reduce the time of xylene clearing in Step 16
46	qPCR cycle thresholds are all very low	cDNA is overamplified	Reduce AL1 primer amount or PCR cycle numbers
	qPCR cycle thresholds are all very high	RNA in tissue is degraded or amplification is defective	Perform an amplification with ~100 pg purified RNA
		cDNA is underamplified	Increase AL1 primer amount or PCR cycle numbers
48	There is insufficient cDNA material in some samples for labeling	Lower overall global cDNA amplification	Run multiple reamplifications of the same low-concentration cDNA template in parallel and pool them during purification
80	Degree of labeling is less than 1.5 dyes per 100 bases	Contamination with trace amines	Add an additional ethanol wash after Step 65
		Too much aa-cDNA added to the reaction	Obtain a stable Nanodrop reading of A_{260} before labeling
	Low 555-cDNA yield	Followed column purification protocol in PureLink manual	Carefully follow Steps 51–64
86	Error returned by MATLAB during filtering	ASCII text files are not exactly formatted as specified	Compare with example data in **Supplementary Data 1** and **2**