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

Lixin Wang & Kevin A Janes

Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, USA. Correspondence should be addressed to K.A.J. (kjanes@virginia.edu).

Published online 10 January 2013; doi:10.1038/nprot.2012.158

Single-cell variations in gene and protein expression are important during development and disease. Such cell-to-cell heterogeneities can be directly inspected one cell at a time, but global methods are usually not sensitive enough to work with the starting material of a single cell. Here we provide a detailed protocol for stochastic profiling, a method that infers single-cell regulatory heterogeneities by repeatedly sampling small collections of cells selected 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 1 week to complete.

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 often an artificial undertaking, because these cells will eventually drift back to a steady-state heterogeneity[15–18]. Instead, a better strategy for studying 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 examine heterogeneity globally was single-cell transcriptomics[31–33]. The details of mRNA expression profiling in single cells can vary widely depending on the method used[31–43]. Algorithmically, however, the different protocols all involve roughly the same steps: (i) extracting cellular RNA by chemical, thermal or enzymatic methods; (ii) performing an oligo(dT)-based capture or an abbreviated oligo(dT)-primed reverse transcription (RT) to prepare a cDNA library of roughly uniform length; (iii) tailing the library with a homopolymer; (iv) exponentially preamplifying the tailed cDNA with a universal homopolymer-containing primer; and (v) detecting the amplification products by quantitative PCR (qPCR), oligonucleotide microarrays or RNA-seq. These five steps 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]. Notably, when similar approaches were applied to cells from a common lineage—in which regulatory heterogeneities are possibly more quantitative than qualitative in nature—the findings were limited to general descriptions of variability[48–50]. These results suggested that existing transcriptomic methods did not clearly separate biological variability from measurement variability when using starting material from a single cell[42,43]. Indeed, certain steps essential to the procedure, such as RT, are known to add substantial measurement variation when minute amounts of input RNA are used[51,52]. A second confounding factor in the earlier studies was that virtually all protocols required tissue dissociation to isolate single cells by micropipette aspiration or FACS[34–40]. The tissue-dissociation step 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 in gene expression 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 substantially reduce measurement and handling artifacts.

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 is not meant to examine transcriptional noise (a type of stochasticity) that is intrinsic to the process of RNA polymerase binding–unbinding and elongation[5,57]. Rather, 'stochastic' refers to a key facet of the method, which focuses on cell-to-cell heterogeneities in the regulation of gene expression. Stochastic profiling is based on the premise that heterogeneities in single-cell regulation can be inferred without measuring them explicitly in single cells (**Fig. 1a**). Instead, randomly selected collections of approximately ten cells are sequentially sampled by laser-capture microdissection[58], and then mRNA expression for each of these 'stochastic samplings' is quantitatively profiled via a customized small-sample cDNA amplification procedure[56]. Highly accurate and precise expression profiles for the stochastic samplings are achievable because of the tenfold increase in starting material compared with the use of 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 log-normal distribution, a common null model for ordinary biological variability[59,60]. Transcripts subject to dichotomous single-cell regulation are identified at this step because of binomial fluctuations in the proportion of high-expressing and low-expressing cells collected during

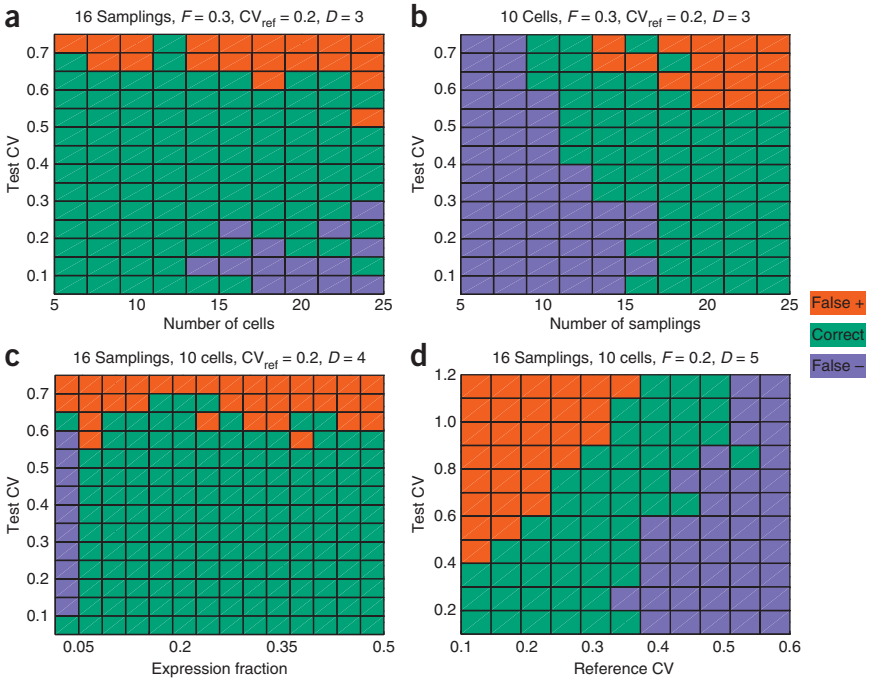

Figure 1 | Theoretical simulations of stochastic profiling for various expression dichotomies defined by different parameter sets. (**a**) Stochastic profiling reliably uncovers a 30% expression dichotomy (*F*) when samplings are composed of fewer than 12 cells. (**b**) Stochastic profiling requires 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 <35% and the test CV (CV_{test}) is not more than approximately threefold greater than CV_{ref}. Simulations were performed with the indicated parameter sets by using `StochProfParameters.m` in **Supplementary Method 1**. False positives (defined as predicting a dichotomy in the test distribution when one does not exist) are shown in orange, false negatives (defined as predicting no dichotomy in the test distribution when one exists) are shown in purple and correct assignments are shown in green.

each stochastic sampling. Finally, the dichotomously fluctuating transcripts are clustered to identify groups of genes with correlated sampling fluctuations, the presence of which suggests expression programs that are coordinately regulated in single cells[56].

Comparison with other methods

Our procedure has two main advantages with respect to existing methods, both of which are related to the tenfold increase in the amount of starting material. First, stochastic profiling measurements are much more sensitive to low-abundance transcripts and less influenced by experimental noise compared with single-cell amplifications[56]. For this reason, stochastic profiling is readily compatible with laser-capture microdissection, which is superior to dissociation-based methods for maintaining adherent cells in their native context[58,61]. Tissues are snap-frozen within seconds and can be ready for microdissection with minimal processing in aqueous solutions that can damage or degrade RNA. However, because the RNA must be proteolytically released from the microdissected specimen, the overall yield in this approach is lower than that obtained by chemical lysis of suspension cells. Ten-cell sampling offsets the decrease in RNA yield per cell and enables one to obtain quantitatively accurate expression profiles[56].

Second, by examining ten 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 identified[34,40,41,46]. These sample numbers eventually become prohibitive for the most expensive step: gene detection. By contrast, stochastic profiling gathers information about ten cells per sample, which hones in on recurrent heterogeneities that will be detected in 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**)[56]. Identifying infrequent heterogeneities (~5–20%) confidently would be difficult with 15–20 single-cell analyses, but such patterns are readily uncovered by stochastic profiling[56,62].

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 one is 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 used this reasoning to study the single-cell regulation of the FOXO transcription factors during 3D organotypic culture of breast epithelial cells[62–64]. Stochastic profiling identified a clear separation in the sampling fluctuations of FOXO-regulated genes, which we independently validated in single cells by multicolor RNA fluorescence *in situ* hybridization (FISH)[62]. 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[65,66] of promoters together with chromatin immunoprecipitation revealed that one cluster of FOXO target genes was co-regulated by another transcription factor, RUNX1 (refs. 62,63). FOXO-RUNX1 cross talk was unanticipated and became apparent only when the heterogeneous expression state of single cells was examined via stochastic profiling. One year later, *RUNX1* was found to be recurrently mutated in breast cancer[67,68], independently validating our earlier predictions of its tumor-suppressive role[62,63].

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 proteins with co-fluctuating expression levels in yeast work together to coordinate important

biological processes[69]. Remarkably, the functions that co-fluctuated in single cells (e.g., stress response and protein biosynthesis) were identical to those functions identified previously by stochastic profiling of 3D breast-epithelial cultures[56]. This evidence suggests that there may be some inherent circuits linked to cell-to-cell heterogeneous regulation that are widely conserved[20]. Another future direction for stochastic profiling is to examine the mechanisms of incompletely penetrant phenotypes[6,70,71]. Conceivably, the emergence of such phenotypes is driven by upstream molecular heterogeneities that are active before the phenotype becomes obvious. Stochastic profiling could be used to search for these heterogeneities in an unbiased way.

Finally, it is important to 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[29,72–74]. In protein analysis, the ten-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. Small-sample stochastic profiling of chromatin modifications at a genome-wide level would be particularly desirable[75,76]. The analysis pipeline described at the end of the protocol here could be immediately adapted to such alternative implementations of our method.

In **Supplementary Method 1**, we provide a script (`StochProfParameters.m`) that simulates stochastic profiling with six user-defined parameters: (i) the number of cells per sampling; (ii) the number of samplings used to build the distribution; (iii) the coefficient of variation (CV) of the log-normal reference distribution (CV_{ref}) that specifies the null model for hypothesis testing; (iv) the underlying CV of the log-normal test distribution (CV_{test}), which is used to diagnose false positives; (v) the fold difference in expression (D) between high and low subpopulations; and (vi) 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 an equilibrated dry ice–isopentane bath. However, the procedure also works with tissues that have been snap-frozen in liquid nitrogen and subsequently embedded. Fixed-frozen specimens are incompatible with our approach because RNA is cross-linked within the tissue in these specimens and cannot be released enzymatically.

We prefer using 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 molecular integrity, not its 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 cutting the sections ourselves rather than submitting samples to a core histology facility.

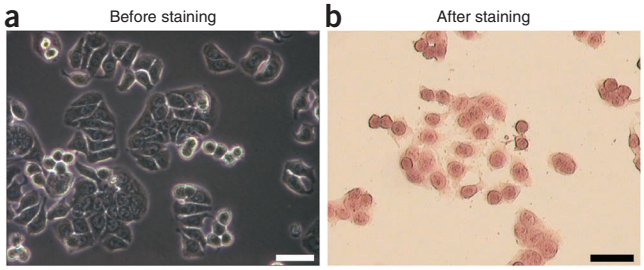

Figure 2 | Rapid nuclear fast red staining of HT-29 colon adenocarcinoma cells for laser-capture microdissection. (**a**) HT-29 cells plated directly on glass coverslips and imaged by phase-contrast microscopy. (**b**) Coverslips fixed and stained with nuclear fast red as described in the PROCEDURE and imaged by bright-field microscopy. Scale bars, 20 μm.

Rapid histology and laser-capture microdissection. Various stains have been reported to be compatible with laser-capture microdissection[77]. However, this protocol uses nuclear fast red because of its superior ability to maintain the molecular integrity of RNA[78]. As RNA is most susceptible to hydrolysis during aqueous processing steps[79], 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 and cultured adherent cells plated on coverslips (**Fig. 2**).

After washing briefly, samples are dehydrated using a series of increasingly more-concentrated ethanol solutions and then cleared with xylenes. We recommend purchasing ethanol in small (~500 ml) quantities because ethanol is hygroscopic and opened containers will draw moisture from the air. Similarly, 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 crucial to maintain the molecular integrity of the RNA in each cell that is microdissected. Ultraviolet-based microdissection platforms cut tissues very cleanly, but RNA strands near the area dissected by the ultraviolet laser are severely degraded. Thus, our protocol uses an infrared-based microdissection instrument for gentle mechanical dissociation of a single cell from its neighbors. Surrounding tissue collaterally picked up is easily removed from the microdissection cap by gently pressing the cap against a weak adhesive (e.g., Post-It) note before RNA elution. As a positive control, a sampling of 100 cells is carried out to assess the overall amplification efficiency. To control for amplification variability, a large pool of microdissected cells is split into multiple identical ten-cell aliquots after elution from the microdissection cap. Usually, the complete set of 15–20 stochastic samplings is performed across two different days, so that sampling-to-sampling variation and day-to-day variation can be compared. A strong association between expression heterogeneities and specific amplification groups indicates problems with day-to-day reproducibility of the procedure.

The biological strategy for random sampling must be clearly defined upfront, 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

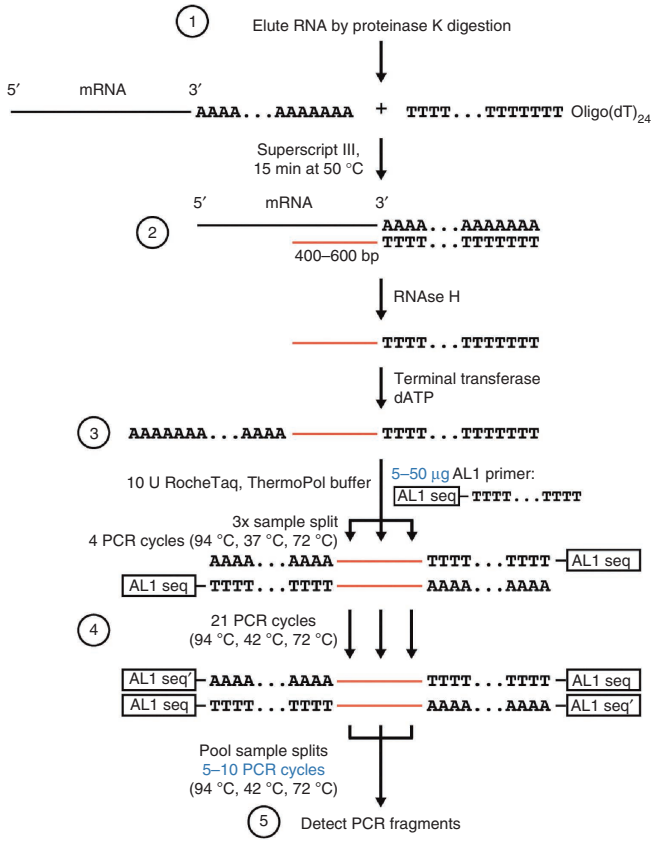

Figure 3 | Workflow for small-sample cDNA amplification. The numbered steps correspond to (1) extraction of cellular RNA, (2) abbreviated reverse transcription, (3) poly(A) tailing, (4) exponential poly(A) PCR with AL1 primer and (5) downstream detection of PCR products as described in the text. Adapted from Janes *et al.*[56] to include the optimization steps (blue) described in the text.

might lie. Experimental conditions (time point, treatment condition and so on) should be optimized beforehand to focus on the sought-after heterogeneities as exclusively as possible. To avoid complications from obvious cell-to-cell variation, such as differences in the cell cycle or cellular microenvironment, the characteristics of the collected samples should be as uniform as possible. Key parameters to control for this purpose include cell size, distance from blood vessels and contact with the extracellular matrix or neighboring cell types. Hidden variations arising from clonal cell subpopulations can be averaged out within each sample by microdissecting cells across different regions of the tissue. Alternatively, by collecting the 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 the following: (i) cellular proteolysis to release RNA from the specimen; (ii) an abbreviated oligo(dT)-primed RT to yield a cDNA pool of uniform length; (iii) poly(A) tailing of the cDNA pool with terminal transferase; and (iv) poly(A) PCR with a universal oligo(dT)-containing primer (AL1 (ref. 80), **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 subjected to 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. As no DNase step is performed in this protocol, and the cDNA synthesized by abbreviated RT often does not contain a splice junction, we sometimes observe slight amplification in the no-RT sample for transcripts with many pseudogenes. We consider this artifact acceptable as long as relative levels in the no-RT control are negligible compared with those observed in the samples.

When working with samples 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. 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 enzymes used later in the procedure, proteinase K is irreversibly inhibited with saturating concentrations of 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. Protease inhibitors that are more stable than PMSF are not as effective, presumably because they interfere with subsequent steps in the procedure.

Although our protocol was originally designed for microdissected cells[56], the amplification process we describe 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 lysis–digestion without loss of amplification efficiency (**Supplementary Fig. 1**). This approach 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 crucially on the cell type and microdissection-suspension format (L.W. and K.A.J., unpublished observations). This effect probably stems from 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 of AL1 for a 100-µl PCR reaction mixture[56]) is the minimum required for high-sensitivity detection. Amplification of some samples continues to improve as AL1 concentration increases up to tenfold, and thus our optimization protocol recommends testing 5–50 µg in pilot experiments with a 100-cell equivalent of starting material.

When performing the optimization of this protocol, fractions of the PCR amplification should be collected from 25–40 cycles for monitoring by qPCR. The goal of collecting fractions at this stage is to identify the maximum number of cycles where high-abundance and low-abundance transcripts (defined by qPCR cycle threshold) are still amplifying efficiently with a 100-cell sample. We use housekeeping genes[81] as abundant mRNA species and then screen various surface receptors and transcription or splicing

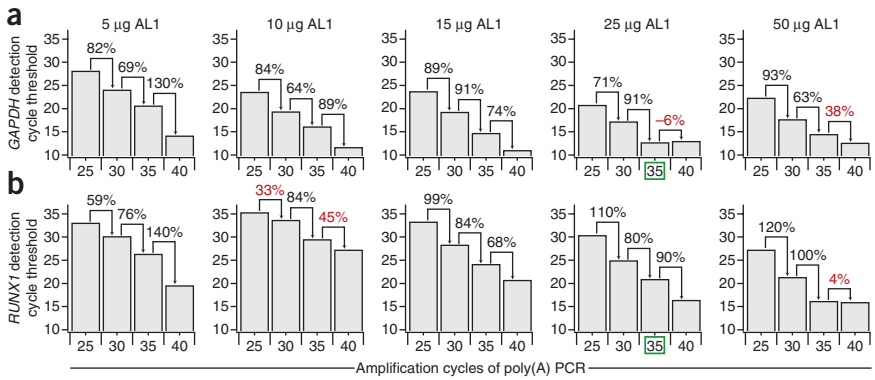

Figure 4 | Representative optimization of AL1 primer amount and poly(A) PCR cycle number for small-sample cDNA amplification implemented with microdissected melanoma cells. (**a**,**b**) Aliquots of 100 cells were amplified with the indicated AL1 primer amounts and total amplification cycles (shown in the *x* axes) and profiled by qPCR for the high-abundance gene *GAPDH* (**a**) and the low-abundance gene *RUNX1* (**b**). 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 µg of AL1 for 35 cycles (green box).

factors that can act as low-copy readouts of the amplification. By surveying six to eight genes within this range, the optimal AL1 concentration and PCR cycle number is readily identified for a specific cell and sample type (**Fig. 4**). Next, the amplification is repeated under the optimized AL1 concentration and cycling conditions with serial dilutions of starting material from 100 cells to 1 cell. We consider the amplification successfully optimized 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 ten-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. Design criteria for the reamplification step are different from those of small-sample amplification. During the ten-cell amplification, processivity and sensitivity of the PCR reaction are paramount. In reamplification, sensitivity is less of an issue, and achieving a high degree of labeling instead becomes the main priority. 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-position and 3-position that are 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 of the extent of amplification, a pilot reaction is performed in the presence of SYBR Green and monitored by real-time qPCR[82]. Great care must be taken to avoid saturating the PCR reamplification 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 of amplicon abundance during qPCR are derived from a mixture of amplified material and primer dimer. Instead, samples containing small amounts of starting cDNA 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, enables one to retain accurate quantitative information about the amount of starting material and ensures reproducibility of the procedure[56].

Before dye coupling, primer dimers should be removed from the reamplification mixture. The presence of 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[83] 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. Aminoallyl-labeled cDNA is coupled to Alexa Fluor 555, which is spectrally interchangeable with Cy3 but shows superior performance for microarray applications[84]. In addition, Alexa Fluor 555 decapacks are available as single-use aliquots, which can mitigate costs. After dye conjugation and secondary purification, the degree of labeling is determined by spectrophotometry with the

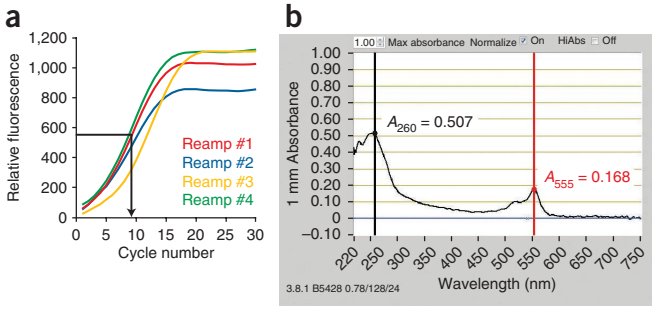

Figure 5 | Reamplification and labeling of ten-cell samples. (**a**) Representative reamplification of four ten-cell samplings monitored by real-time qPCR with SYBR Green fluorescence. The black arrow indicates the maximum number of reamplification cycles (nine). Note that sample no. 3 (yellow profile) will give a lower yield than the other reamplifications and will probably 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.

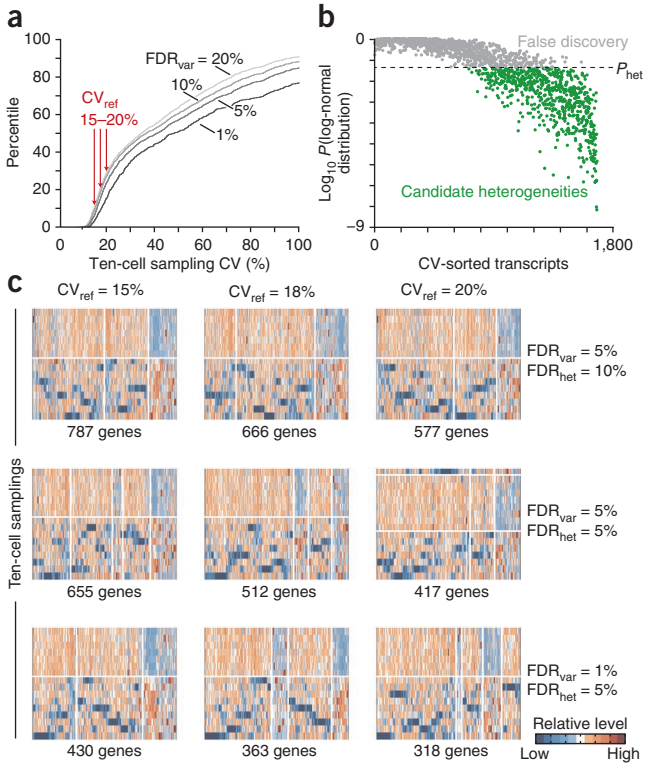

Figure 6 | Statistical and bioinformatic analysis of stochastic-profiling data. (**a**) Empirical cumulative distribution function is shown as the aggregate percentile (shown in the *y* axis) of ten-cell sampling CVs for genes with significant biological variation at FDR_{var} = 20, 10, 5 and 1% (profiles of increasingly dark gray color). Candidate reference CVs (CV_{ref}) are highlighted in red. Note that the cumulative distribution function at FDR_{var} = 1% is substantially reduced compared with the others, suggesting overly stringent filtering at this step. (**b**) *P* values for the χ^2 goodness-of-fit test as a function of CV-ordered transcripts (shown in the *x* axis) 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 on the basis of 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 set from Janes *et al.*[56] was analyzed with the indicated parameters, and the resulting gene sets were hierarchically clustered with Ward's linkage, wherein rows indicate the stochastic ten-cell samplings, columns indicate the genes retained by the analysis and the color bar indicates the z-score–standardized fluctuations of each gene about its mean. The number of genes retained by the indicated parameter set is shown underneath the clustered data set. The major expression clusters separated by hierarchical clustering are marked in white.

Invitrogen dye:base ratio calculator: http://probes.invitrogen.com/resources/calc/basedyeratio.html. Our protocol enables the conjugation of seven to ten dye molecules per ~500-bp amplicon (see ANTICIPATED RESULTS and **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 they have a lower cost and higher throughput than alternative systems, and they perform equivalently[85]. The hybridization protocol is performed in accordance with the recommendations of the manufacturer, except for the following modifications: (i) 1 µg of cDNA is added to each well in place of the 750 ng of cRNA recommended by the manufacturer; this increased amount of cDNA is chosen to account for the fact that only the complementary strand of the cDNA sample will hybridize. (ii) The samples are denatured briefly at 95 °C and then added to a slide that has been prewarmed at the 58 °C hybridization temperature. This second modification is introduced to minimize re-annealing of the labeled cDNA with its complementary strand before hybridization. From this point on, slides are incubated, washed and scanned exactly as recommended.

A stochastic profiling experiment typically involves 16–20 random ten-cell samples and 16–20 amplification controls (a larger pool of 160–200 cells split into ten-cell aliquots before small-sample amplification). Each microarray is normalized to have the same overall 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 marked fluctuation artifacts in the amplification controls, which should ordinarily be very precise. By using the amplification controls, we apply a loose filter that removes genes from the data set with control fluctuations greater than fivefold. 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 renormalized by median fluorescence intensity to adjust for residual postfiltering differences in the overall signal. The renormalized ten-cell samplings comprise the final preprocessed data set for analysis.

The first step in the analysis pipeline is the identification of genes whose expression levels among ten-cell samples show significantly larger fluctuations than the amplification replicates. Because eukaryotic gene-expression variability is often log-normally distributed[59,86], we logarithmically transform the data for analysis. To standardize the log-transformed data, the level of each transcript is then scaled by its geometric mean calculated across all ten-cell samples, and each ten-cell microarray is normalized to its overall geometric mean. Next, we must identify those transcripts whose expression levels have significantly larger variations between independent ten-cell samples than between amplification replicates. Separating biological variation from measurement variation enables one to estimate a reference distribution with which to compare the fluctuations measured in the ten-cell samples. In our original work[56], we compared the CV of the sample-to-sample fluctuations with the CV of the amplification controls by using McKay's approximation[87]. However, we now prefer to avoid approximations and instead directly examine the ratio of variances with respect to the *F* distribution[88]. Genes with significantly higher sample-to-sample variances than controls (at a user-defined false-discovery rate, FDR_{var}) are identified and then sorted on the basis of their CV for subsequent distribution testing.

Many methods exist for comparing empirical data with a (log)-normal distribution[89]. Our early work with stochastic profiling used the χ^2 goodness-of-fit test[56,62], but we now favor the K-S test because it is conservative and can be accurately applied on a gene-by-gene basis. To define a reference distribution for the K-S test, we inspect the cumulative distribution function of CVs from genes with measurable sample-to-sample variations (**Fig. 6a**).

By using this empirical plot, we seek to identify a CV_{ref} that accounts for 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 the CV_{ref} is below 35% and the underlying variability of the test distribution is not more than three times larger than the 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). This approach assumes that the 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 initial variance filter is too stringent, then few low-CV transcripts will enter into the cumulative distribution function, making it harder to identify CV_{ref} (**Fig. 6a**, curves of increasingly dark gray color). 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 underlying the variation of the test distributions. By using the selected CV_{ref}, we perform the K-S test on a gene-by-gene basis, imposing a threshold for the resulting *P* value of each gene according to a second user-defined false-discovery rate (FDR_{het}). FDR_{het} is generally less stringent than FDR_{var} because of the conservative nature of the K-S test. The genes whose sample-to-sample fluctuations yield *P* values below this threshold are predicted by stochastic profiling to be heterogeneously expressed (**Fig. 6b**, green spots).

To facilitate the filtering and analysis of stochastic-profiling data, we provide here a pair of MATLAB functions that perform the necessary calculations (**Supplementary Methods 2** and **3**). `StochProfMicroarrayFilt.m` (**Supplementary Method 2**) 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 performed once. `StochProfAnalysis.m` (**Supplementary Method 3**) 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 ten-cell samplings and 16 amplification controls for matrix-attached breast epithelial cells in 3D organotypic culture[56] (**Supplementary Data 1** and **2**). When executing the analysis pipeline, there are three user-defined inputs to consider: (i) FDR_{var}, the false-discovery rate for testing significant biological variation above measurement variation; (ii) CV_{ref}, the reference CV estimating background biological variation (**Fig. 6a**); and (iii) 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 early data set[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, it is just a starting point for more specific observations and perturbations of the candidate genes and their single-cell expression patterns. To validate predicted heterogeneities, we use RNA FISH because gene-specific reagents are readily synthesized and 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[56,62].

Validated transcripts can be pursued further to test for functions of the cell-to-cell regulatory 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 involved in translation and protein turnover.) Direct functional testing can be challenging because the role of the heterogeneity and the general role of the gene or protein itself need to be separated. 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 RNA interference (RNAi) with the goal of eliminating the high-expressing 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-expressing population. The difficulty is that either of these perturbations will also change the overall levels of expression. Ultimately, assigning function to a heterogeneity requires add-back 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[90], here the expectation is that add-back 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.

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 rates[57]. Deviations from a log-normal baseline 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. Problems will also arise with extremely low-abundance transcripts, in which some cells will have exactly zero copies, because the log-normal distribution is only defined for values

greater than zero[60]. 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, noncoding RNAs and so on). 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 on the basis of 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

- Tissue specimens ▲ **CRITICAL** If you are using tissue specimens as starting material, implement the entire procedure.
- Cultured adherent cells ▲ **CRITICAL** If you are using cultured adherent cells as starting material, begin the procedure at Step 10.
- Suspension cells ▲ **CRITICAL** If you are using suspension cells as starting material, begin the procedure at Step 26.
- Ethanol (Ultrapure, cat. no. 200CSPTP)
- Nuclear fast red (Vector Laboratories, cat. no. H-3403)
- RNasin Plus (Promega, cat. no. N2611, 40 U μl^{-1})
- Xylenes (Fisher Scientific, cat. no. HC700-1GAL) ! **CAUTION** Xylenes are toxic and should only be used in a chemical fume hood.
- RNase Away (Molecular Bioproducts, cat. no. 7003)
- Isopentane (Sigma-Aldrich, cat. no. 320404-1L) ! **CAUTION** Isopentane is highly flammable and should not be disposed of down the sink.
- Neg-50 embedding medium (Richard-Allan Scientific, cat. no. 6502)
- Clear nail polish (Fisher, cat. no. S30697)
- SuperScript III (Invitrogen, cat. no. 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 used.
- First-strand buffer, 5× (included with Invitrogen, cat. no. 18080-044)
- Nuclease-free H_2O (Promega, cat. no. P119C)
- Proteinase K (Sigma-Aldrich, cat. no. P2308)
- Saponin (TCI America, cat. no. S0019)
- Anti-RNase (Ambion, cat. no. 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, cat. no. 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-Aldrich, cat. no. P7626) ▲ **CRITICAL** PMSF comes in various purities depending on the vendor, and the procedure has been optimized for the chemical provided by this supplier.
- dATP, dCTP, dGTP, dTTP, 100 mM (dNTP set; Roche, cat. no. 11277049001)
- $Oligo(dT)_{24}$ (25 nmol synthesis from Invitrogen)
- RNase H (New England Biolabs, cat. no. M0297L, 5 U μl^{-1})
- $MgCl_2$, 25 mM (Applied Biosystems, cat. no. H05143)
- Terminal transferase buffer (Invitrogen, cat. no. 16314-015) ! **CAUTION** The terminal transferase buffer contains Co^{2+}, which is toxic if ingested or inhaled. Use appropriate precautions.
- Terminal transferase (Roche, cat. no. 03333574001, 400 U μl^{-1})
- ThermoPol buffer, 10× (New England Biolabs, cat. no. B9005S)
- $MgSO_4$, 100 mM (New England Biolabs, cat. no. B1003S)
- BSA (Roche, cat. no. 10711454001)
- Roche Taq polymerase (Roche, cat. no. 04728858001, 5 U μl^{-1})
- AL1 primer (200 nmol synthesis from IDT or MWG): 5′-ATTGGATCCAGGCCGCTCTGGACAAAATATGAAT-TCTTTTTTTTTTTTTTTTTTTTTTTTT-3′ ▲ **CRITICAL** Long desalted primers will have varying purities, depending on the manufacturer.
- SYBR Green I, 100× (Invitrogen, cat. no. S7563; diluted 100-fold in DMSO to 100×)
- Aminoallyl-dUTP, 50 mM (Ambion, cat. no. 1103015)
- High-fidelity polymerase (Roche, cat. no. 11732650001, 3.5 U μl^{-1})
- High-fidelity PCR buffer without Mg^{2+}, 10× (included with Roche, cat. no. 11759175001)
- PureLink PCR purification kit (Invitrogen, cat. no. 46-6056)
- Sodium acetate (Calbiochem, cat. no. 567418)
- Glycogen, 20 mg ml^{-1} (Invitrogen, cat. no. 10814-010)
- $NaHCO_3$ (Acros, cat. no. 21712500)
- Alexa Fluor 555 reactive dye decapack (Invitrogen, cat. no. A32756)
- DMSO (Mediatech, cat. no. MT-25-950-CQC)
- GEX hybridization buffer (included with Illumina, cat. no. BD-103-0204 or equivalent)
- PCR-grade water

EQUIPMENT

- Thermocycler with a heated lid (Bio-Rad MyCycler or equivalent)
- Cryostat (Leica, cat. no. CM1950)
- Microtome blades (Thermo Scientific, cat. no. MX36 premier+)
- Pixcell II laser-capture microdissection instrument (Arcturus)
- Inverted microscope (Olympus, cat. no. CKX41 or equivalent)
- Cell counter (Beckman Coulter, cat. no. AM35308)
- Hemocytometer (Hausser Scientific, cat. no. 437757)
- SuperFrost Plus glass microscope slides (VWR, cat. no. 48311-703)
- Glass coverslips, 24 mm × 50 mm, no. 1.5 (Fisher Scientific, cat. no. 12-544D)
- Coplin staining jars (Fisher, cat. no. S90130)
- Cryomold (Sakura, cat. no. 4566)
- PrepStrip (Arcturus, cat. no. LCM0207)
- Capsure HS LCM caps and ExtracSure adaptor (Arcturus, cat. no. LCM0214)
- Thin-walled PCR tubes, 0.5 ml (Applied Biosystems, cat. no. N8010611)
- Thin-walled PCR tubes, 0.2 ml (Applied Biosystems, cat. no. N8010612) ▲ **CRITICAL** The amplification procedure has only been validated with these PCR tubes.
- PCR adaptors (Fisher, cat. no. 11-715-125D)
- qPCR instrument (Bio-Rad, CFX 96 Real-Time system or equivalent)
- 96-well qPCR plates and optically clear film
- NanoDrop spectrophotometer (NanoDrop Technologies, cat. no. ND-1000)
- Expression BeadChip (Illumina, cat. no. BD-103-0204 or equivalent)
- BeadArray reader (Illumina)
- MATLAB software (MathWorks)
- Kimwipes
- Desiccator

REAGENT SETUP

Proteinase K, 20 mg ml^{-1} Prepare a 20 mg ml^{-1} solution in nuclease-free H_2O and store it in 20-µl aliquots at −20 °C for up to 6 months. After thawing, keep the solution at 4 °C for up to 1 month.

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

Stock primer mix, 25× Prepare a solution containing 15 µl of nuclease-free H_2O, 5 µl of 100 mM dATP, 5 µl of 100 mM dCTP, 5 µl of 100 mM dGTP, 5 µl of 100 mM dTTP and 5 µl of 80 OD per ml $oligo(dT)_{24}$. Store the primer mix in 5-µl aliquots at −20 °C for up to 6 months.

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

Tailing buffer, 2.6× Prepare a solution containing 363 µl of nuclease-free H_2O, 400 µl of 5× Invitrogen terminal transferase buffer and 15 µl of 100 mM dATP. Store the buffer in 100-µl aliquots at −20 °C for up to 1 year. Add 0.2 µl of terminal transferase (400 U μl^{-1}) per 3.5 µl of 2.6× tailing buffer immediately before use. ! **CAUTION** Co^{2+} is toxic if ingested or inhaled. Use the buffer with

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

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

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

$NaHCO_3$, 1 M 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 it at room temperature (22 °C) for up to 1 month.

Sodium acetate, 3 M (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 liter with H_2O. Dispense the solution into aliquots and sterilize it by autoclaving. Store it at room temperature for 1 year or more.

PROCEDURE

Embedding and cryosectioning of tissue specimens ● TIMING 1 d

▲ CRITICAL Steps 1–9 are only implemented in the processing of tissue specimens; if you are using cultured adherent cells as starting material, go directly to Step 10. If you are using suspension cells, go instead directly to Step 26.

1| Equilibrate a dry ice–isopentane bath in a plastic beaker.

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

2| Place fresh or snap-frozen (in liquid nitrogen) tissue into a small cryomold and cover it with Neg-50 embedding medium.

▲ CRITICAL STEP Proceed 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. After the specimen is completely frozen, the sample can be stored on dry ice while additional samples are embedded.

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

■ PAUSE POINT For long-term storage from months to years, wrap cryomolds in tinfoil and store them at −80 °C.

4| Transport the embedded samples on dry ice to the cryostat. Place the 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 it with Neg-50 embedding medium 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 the sections onto slides and move the 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 as quickly as possible after wicking in order 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 h

10| Fix the samples with 75% (vol/vol) ethanol implementing option A or option B, depending on whether frozen sections or cultured adherent cells are used, respectively.

(A) Fixation of frozen sections

(i) Remove four slides from dry ice or from a −80 °C freezer and place them 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 or the dry ice container and immerse the slides before they thaw or accumulate excessive frost.

(B) Fixation of cultured adherent cells

(i) Plate the cells on 24 mm × 50 mm coverslips as desired, and then place them immediately in 75% (vol/vol) ethanol for 30 s.

11| Transfer the slides to distilled water at room temperature and wait for 30 s.
▲ **CRITICAL STEP** All aqueous staining steps should be followed precisely in order to maintain consistent RNA integrity. Reserve a set of new Coplin staining jars exclusively for laser-capture microdissection.

12| Place the slides face up on top of a paper towel and then add a few drops of nuclear fast red containing 1 U ml^{-1} RNasin Plus to their surfaces. Let the staining proceed for 30 s. A volume of 100 µl of nuclear fast red + 2.5 µl of RNasin Plus is sufficient to stain four coverslips or slides containing two sections per slide.

13| Move the slides to a new Coplin staining jar containing distilled water. Remove each slide individually from the Coplin staining jar and dip it back in the jar again for a total rinse time of 15 s in distilled water.

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

15| Transfer the slides to a new Coplin staining jar containing 70% (vol/vol) ethanol and let the dehydration proceed for 30 s. Transfer the slides to another Coplin staining jar containing 95% (vol/vol) ethanol and leave them to dehydrate for 30 s. Transfer the slides finally to another Coplin staining jar containing 100% (vol/vol) ethanol and let the dehydration proceed for 30 s.

16| Place the slides in a new Coplin staining jar containing xylenes and allow the ethanol to be cleared for 2 min.
! **CAUTION** Xylenes are toxic and should only be used in a chemical fume hood.

17| Air-dry the slides face up for 5–10 min in a chemical fume hood. For cells cultured on coverslips, place dried coverslips face up on glass microscope slides and use clear nail polish to ensure that the edges of the coverslips adhere to the slides.

18| Place the slides in a desiccator and transport them to the microdissector.

19| Turn on the instrument and spray your 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 (one or two cells per laser shot). Multiple shots are often required to cause polymer wetting at this laser power.
? **TROUBLESHOOTING**

23| (Optional) If there is extensive collateral pickup from adjacent nondissected cells, press the LCM cap lightly on an adhesive (e.g., Post-It) note.
▲ **CRITICAL STEP** The weakest possible adhesive note should be used in order to avoid removing the material from the microdissected cells.

24| Load the LCM cap onto the ExtracSure adaptor included with the LCM caps and store it 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 h**
26| Prepare digestion mixtures according to option A or option B, depending on whether microdissected cells or suspension cells are used, respectively.

(A) Digestion of microdissected cells

(i) Prepare the digestion buffer at room temperature in a microcentrifuge tube as follows:

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

(ii) Add 4 μl of digestion buffer to the ExtracSure adaptor containing the microdissected cells. Cover the ExtracSure adaptor with a 0.5-ml thin-walled PCR tube.
▲ **CRITICAL STEP** Make sure to cover the adaptors tightly with PCR tubes, but do not crimp the PCR tubes or dislodge the ExtracSure adaptor.

(iii) Incubate both the LCM cap and the remaining digestion buffer at 42 °C for 1 h.
▲ **CRITICAL STEP** The remaining digestion buffer must be treated in an identical manner as the buffer in contact with cells, so that the proteinase K partially inactivates itself and the buffer can be used to dilute concentrated samples after RNA elution.

(B) Digestion of suspension cells

(i) Prepare the predigestion buffer at room temperature in a microcentrifuge tube as follows:

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

(ii) Resuspend the cells at the desired concentration (number of cells per 3.6 μl) in predigestion buffer lacking proteinase K.
■ **PAUSE POINT** The cells suspended in predigestion buffer can be stored at −80 °C for months without loss of amplification efficiency.

(iii) Add a one-tenth volume of saponin-proteinase K solution to the predigestion mixture to make 1× digestion buffer containing 0.1% (wt/vol) saponin for cell digestion and RNA extraction.

(iv) Incubate both the cell suspension and the remaining digestion buffer at 42 °C for 1 h.
▲ **CRITICAL STEP** The remaining digestion buffer must be treated in an identical manner as the buffer in contact with cells, so that the proteinase K partially inactivates itself and the buffer can be used to dilute concentrated samples after RNA elution.

27| Centrifuge tubes containing the digested cells from Step 26 for 2 min at 2,500*g* in a benchtop centrifuge at room temperature.

28| Prepare the digestion-stop buffer at room temperature in a microcentrifuge tube as follows:

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

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

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

30| (Optional) If you are performing a serial or replicate dilution of a more-concentrated sample, dilute the sample from Step 29 with digestion buffer + digestion-stop buffer, mixed at a 4:1 ratio shortly beforehand.
▲ **CRITICAL STEP** The digestion buffer used for dilution at this step must be incubated for 1 h at 42 °C as described in Step 26 before it is mixed with digestion-stop buffer.

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 control sample 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 it to cool at room temperature for 90 s. Spin the mixture for 2 min at 12,000*g* on a benchtop centrifuge at 4 °C to collect condensation within the tube.
▲ **CRITICAL STEP** From this step onward, perform all heating and incubation steps in a PCR thermocycler to ensure temperature accuracy and stability.

33| Add 0.5 µl of SuperScript III to each sample, vortex it briefly and incubate it at 50 °C for 15 min.
▲ **CRITICAL STEP** Ensure 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 SuperScript III reverse transcriptase by incubating it at 70 °C for 15 min.

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

36| Add 1 µl of RNase H–Mg^{2+} to each sample, mix it and incubate it at 37 °C for 15 min.

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

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

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

40| Place the samples on ice and spin them 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 on ice in a microcentrifuge tube as follows:

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

▲ **CRITICAL STEP** If you are 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 µg of AL1.

42| Add 90 µl of ThermoPol PCR buffer to the 0.2-ml, thin-walled PCR tube containing the tailed and heat-inactivated samples.

43| Transfer two 33-µl aliquots of the mixture from Step 42 to two 0.2-ml, thin-walled PCR tubes and leave the third remaining 33-µl aliquot in the original tube.

44| In a thermocycler with 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 s increase at each cycle
5–25	94 °C for 1 min	42 °C for 2 min	72 °C for 6 min 40 s with 10 s increase at each cycle

45| (Optional) If you are optimizing for a new biological format, collect a 10-µl fraction of the PCR reaction after cycle 25.

46| Cool the PCR mixtures to 4 °C and place them in ice. Pool the three aliquots into the original 0.2-ml PCR tube, vortex the tube and centrifuge the mixture briefly.

47| In a thermocycler with 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

48| If the protocol has already been optimized, stop the entire reaction at the optimal number of PCR cycles; otherwise, proceed until the advised number of cycles has been implemented. Collect a 10-µl fraction of the PCR reaction after cycles 30, 35 and 40.

49| Cool the samples to 4 °C in the thermocycler and then place them on ice.
■ **PAUSE POINT** Samples can be frozen and stored at −20 °C for months to years and can be thawed several times without noticeable degradation of the amplification products.

50| Dilute 1 µl of each amplified cDNA sample 450- to 500-fold in water and quantify the genes of interest by qPCR as described in ref. 81, or by using an equivalent qPCR procedure.
▲ **CRITICAL STEP** Because of the abbreviated RT in Step 33, qPCR primers must be designed to anneal within ~400 bp from the 3′ end of the transcript for the transcript to be detected reliably.
? TROUBLESHOOTING

cDNA reamplification and labeling ● TIMING 2–3 d

51| Prepare the following master mix (sufficient for ten reactions) to perform a pilot reamplification:

Reagent	Volume (µl)
10× high-fidelity PCR buffer without Mg^{2+}	20
25 mM $MgCl_2$	28
100 mM dATP	0.4
100 mM dCTP	0.4
100 mM dGTP	0.4
50 mM aminoallyl-dUTP	0.64

10 mM dTTP	0.8
20 mg ml^{-1} BSA	1
15 μg μl^{-1} AL1 primer	0.6
3.5 U μl^{-1} high-fidelity polymerase	2
100× SYBR Green	0.5
Nuclease-free H_2O	135.3
Total volume	190

▲ **CRITICAL STEP** Note that the concentration of the dTTP stock is tenfold lower than that of the other dNTPs, so as to increase the extent of aminoallyl-dUTP incorporation.

52| Add 0.5 μl of each amplified cDNA sample from Step 49 to 2 μl of PCR-grade water; next, add 1 μl of this diluted cDNA to 19 μl of the reamplification master mix, and monitor the amplification products by qPCR via 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 the fluorescence

▲ **CRITICAL STEP** The goal of the pilot reamplification is to identify the amplification cycle at which the first sample hits the middle of its exponential phase. Carrying on the amplification into the late exponential phase will overamplify the cDNA strands and cause loss of quantitative accuracy.

? TROUBLESHOOTING

53| Prepare the following aminoallyl-cDNA (aa-cDNA) reamplification reaction mixture:

Reagent	Volume (μl)
10× high-fidelity PCR buffer without Mg^{2+}	10
25 mM $MgCl_2$	14
100 mM dATP	0.2
100 mM dCTP	0.2
100 mM dGTP	0.2
50 mM aminoallyl-dUTP	0.32
10 mM dTTP	0.4
20 mg ml^{-1} BSA	0.5
15 μg μl^{-1} AL1 primer	0.3
Amplified cDNA from Step 49	1
3.5 U μl^{-1} high-fidelity polymerase	1
PCR-grade water	71.9
Total volume	100

54| Run the following PCR amplification protocol in a thermocycler with a heated lid:

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 52.

55| To each aa-cDNA sample, add 400 µl of PureLink binding buffer included in the PureLink PCR purification kit.
▲ **CRITICAL STEP** Do not use the high-cutoff binding buffer included with the PureLink columns, as using this buffer will cause the aa-cDNA to flow through the column.

56| Apply the entire solution to a PureLink column and centrifuge it at 10,000*g* for 1 min at room temperature.

57| Discard the flow-through, wash the column with 650 µl of wash buffer included in the PureLink PCR purification kit and centrifuge the mixture at 10,000*g* for 1 min.

58| Discard the flow-through again, and centrifuge it once again at 10,000*g* for 1 min at room temperature.

59| Transfer the column to a clean elution tube and add 50 µl of the elution buffer included in the PureLink PCR purification kit.

60| Seal the cap on the column and incubate the column at 65 °C for 10 min.
▲ **CRITICAL STEP** The high-temperature elution maximizes the yield of the aa-cDNA dissolved off the PureLink columns.

61| Centrifuge the column at 10,000*g* for 1 min at room temperature and retain the eluate at the bottom of the tube.

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

63| Centrifuge the column at 10,000*g* for 1 min at room temperature, discard the column and save the eluate.

64| Add 10 µl of 3 M sodium acetate (pH 5.2) and 1 µl of 20 mg ml^{-1} glycogen to the eluate and vortex the resulting mixture.

65| Add 250 µl of ice-cold 100% (vol/vol) ethanol, vortex and incubate the samples at −20 °C for at least 30 min.
PAUSE POINT The samples can be stored at −20 °C overnight if needed.

66| Spin the samples from Step 65 for 10 min at 4 °C at maximum speed on a benchtop centrifuge.

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

68| Spin the samples for 1 min at room temperature at maximum speed on a benchtop centrifuge.

69| Carefully aspirate the supernatant and wash the pellets with 500 µl of 70% (vol/vol) ethanol at room temperature.
▲ **CRITICAL STEP** The second wash is important to remove amine traces from the precipitated aa-cDNA and thus achieve maximal labeling efficiency afterward.

70| Spin the samples for 1 min at maximum speed at room temperature on a benchtop centrifuge.

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

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

73| Resuspend each pellet in 5 µl of nuclease-free water and incubate the resulting mixture for 15 min at 37 °C to re-dissolve the aa-cDNA pellet.

74| Determine the aa-cDNA concentration by measuring the absorbance at 260 nm (A_{260}) on a NanoDrop spectrophotometer.
■ **PAUSE POINT** The samples can be frozen and stored at −20 °C for months to years and can be thawed several times without noticeable degradation of the aa-cDNA.

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

76| For each labeling reaction, dissolve one vial of Alexa Fluor 555 succinimidyl ester dye from the decapack in 2 µl of DMSO.
▲ **CRITICAL STEP** Add DMSO to the side of each tube, and then spin down all the tubes together to minimize the time during which the dye is sitting in DMSO outside of the reaction.

77| Add 2 µl of the resuspended Alexa Fluor 555 dye to the mixture and vortex it at maximum speed for 15 s.
▲ **CRITICAL STEP** The lengthy vortexing is crucial to ensure high coupling efficiencies.

78| Spin down the labeling reaction mixtures briefly and incubate them for 1 h at room temperature.

79| Add 10 µl of 3 M sodium acetate (pH 5.2) and 80 µl of water to each labeling reaction mixture.
▲ **CRITICAL STEP** Neutralizing the pH with 3 M sodium acetate (pH 5.2) increases the efficiency of the binding to the PureLink column in the subsequent purification.

80| Repeat Steps 55–68 with the Alexa Fluor 555–labeled cDNA (555-cDNA).

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

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

83| Resuspend each pellet in 5 µl of nuclease-free water and incubate the resulting mixture for 15 min at 37 °C to re-dissolve the 555-cDNA pellet.

84| Determine the 555-cDNA concentration and the 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).
? TROUBLESHOOTING
■ **PAUSE POINT** The samples can be frozen and stored at −20 °C for months without noticeable degradation of the 555-cDNA.

Microarray hybridization and data analysis ● TIMING 2–3 d

85| Mix each 555-cDNA sample with 10 µl of GEX hybridization buffer and place the resulting mixture at 94 °C for 4 min to achieve DNA denaturation.

86| Add individual samples directly to each lane of an Expression BeadChip prewarmed at 58 °C, and incubate the Expression BeadChip at 58 °C for 20 h.

87| Wash and dry the slides according to the manufacturer's recommendation, and scan them on a BeadArray reader.
■ **PAUSE POINT** After scanning the Expression BeadChip, data analysis can be performed at any time.

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

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

90| On the MATLAB command window, type

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

? TROUBLESHOOTING

91| 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.

92| 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
```

93| On the MATLAB command window, type

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

94| 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`.

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

```
clear all

close all

load FilteredMicroarrays
```

and return to Step 93.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 1**.

TABLE 1 | 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
50	qPCR cycle thresholds are all very low (<15)	cDNA is overamplified	Reduce AL1 primer amount or PCR cycle numbers
	qPCR cycle thresholds are all very high (>25)	RNA in tissue is degraded or amplification is defective	Perform an amplification with ~100 pg of purified RNA
		cDNA is underamplified	Increase AL1 primer amount or PCR cycle numbers
52	There is insufficient cDNA material in some samples for labeling	Low overall global cDNA amplification	Run multiple reamplifications of the same low-concentration cDNA template in parallel and pool them together during purification

(continued)

TABLE 1 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
84	Degree of labeling is less than 1.5 dye molecules per 100 bases	Contamination with trace amounts of amines	Add an additional ethanol wash after Step 69
		Too much aa-cDNA added to the reaction	Obtain a stable NanoDrop reading of A_{260} before labeling
	Low 555-cDNA yield	Column purification protocol described in PureLink manual was mistakenly followed	Carefully follow Steps 55–63
90	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**

● TIMING

Steps 1–9, embedding and cryosectioning, 1 d
Steps 10–25, staining and laser-capture microdissection, 2 h
Steps 26–50, sample-specific cDNA amplification, 11 h
Steps 51–84, cDNA reamplification and labeling, 2–3 d
Steps 85–95, microarray hybridization and data analysis, 2–3 d

ANTICIPATED RESULTS

The rapid histology protocol should yield a 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 observation emphasizes further the importance of the random ten-cell sampling approach for microdissected tissue. Interestingly, one-cell measurements are possible with suspension cells (**Fig. 7c** and **Supplementary Fig. 4**), which is consistent with earlier results from single cells obtained by micropipette aspiration or FACS[34–40]. On the basis of simulations, the overall reproducibility of ten-cell amplification replicates must be within 35% because background biological variation will only amplify this error and can ultimately 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 sample-to-sample basis (**Fig. 5a**, yellow). This observation 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 them before labeling

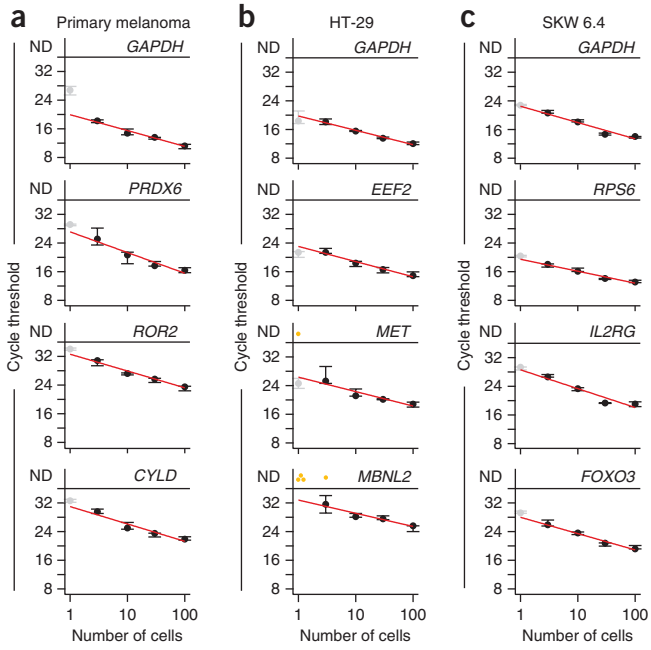

Figure 7 | Optimized small-sample cDNA amplifications in three distinct biological contexts. (**a**–**c**) 100-cell samples were serially diluted and amplified by poly(A) PCR under optimal conditions for microdissected primary melanoma cells (25 µg of AL1, 35 cycles) (**a**), microdissected HT-29 colon adenocarcinoma cells (10 µg of AL1, 35 cycles) (**b**) and SKW 6.4 lymphoblastoid suspension cells (50 µg of AL1, 30 cycles) (**c**). High-abundance 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- to 100-cell dilutions. Note that the one-cell amplifications (gray) often deviate from the log-linear fit or are frequently not detectable (ND, yellow).

(see TROUBLESHOOTING). When labeling aa-cDNA, we typically observe ~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 7,000–10,000 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.

ACKNOWLEDGMENTS We thank S. Bajikar for critically reading this manuscript, C. Slingluff for providing the primary melanoma sample, B. Kang for help with screening low-abundance genes and C. Borgman for help with frozen sectioning. This work was supported by the US 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.

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 no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nprot.2012.158. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
2. Altschuler, S.J. & Wu, L.F. Cellular heterogeneity: do differences make a difference? *Cell* **141**, 559–563 (2010).
3. Snijder, B. & Pelkmans, L. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.* **12**, 119–125 (2011).
4. Losick, R. & Desplan, C. Stochasticity and cell fate. *Science* **320**, 65–68 (2008).
5. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
6. Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**, 520–523 (2009).
7. Snijder, B. *et al.* Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol. Syst. Biol.* **8**, 579 (2012).
8. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X.S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).
9. Golding, I., Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
10. Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).
11. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single-molecule level. *Nature* **440**, 358–362 (2006).
12. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* **43**, 95–100 (2011).
13. Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J. & Eisen, M.B. Noise minimization in eukaryotic gene expression. *PLoS Biol.* **2**, e137 (2004).
14. Lestas, I., Vinnicombe, G. & Paulsson, J. Fundamental limits on the suppression of molecular fluctuations. *Nature* **467**, 174–178 (2010).
15. Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M. & Sorger, P.K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432 (2009).
16. Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008).
17. Sharma, S.V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
18. Gupta, P.B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
19. Xie, X.S., Yu, J. & Yang, W.Y. Living cells as test tubes. *Science* **312**, 228–230 (2006).
20. Pelkmans, L. Cell Biology. Using cell-to-cell variability—a new era in molecular biology. *Science* **336**, 425–426 (2012).
21. Munsky, B., Neuert, G. & van Oudenaarden, A. Using gene expression noise to understand gene regulation. *Science* **336**, 183–187 (2012).
22. Wang, D. & Bodovitz, S. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol.* **28**, 281–290 (2010).
23. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
24. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
25. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012).
26. Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886–895 (2012).
27. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
28. Bendall, S.C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
29. Shi, Q. *et al.* Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. *Proc. Natl. Acad. Sci. USA* **109**, 419–424 (2012).
30. Cohen, A.A. *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511–1516 (2008).
31. Iscove, N.N. *et al.* Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat. Biotechnol.* **20**, 940–943 (2002).
32. Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
33. Chiang, M.K. & Melton, D.A. Single-cell transcript analysis of pancreas development. *Dev. Cell* **4**, 383–393 (2003).
34. Sanchez-Freire, V., Ebert, A.D., Kalisky, T., Quake, S.R. & Wu, J.C. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nat. Protoc.* **7**, 829–838 (2012).
35. Citri, A., Pang, Z.P., Sudhof, T.C., Wernig, M. & Malenka, R.C. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat. Protoc.* **7**, 118–127 (2012).
36. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
37. Esumi, S., Kaneko, R., Kawamura, Y. & Yagi, T. Split single-cell RT-PCR analysis of Purkinje cells. *Nat. Protoc.* **1**, 2143–2151 (2006).
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.* **2**, 739–752 (2007).

39. Tang, F. *et al.* 220-plex microRNA expression profile of a single cell. *Nat. Protoc.* **1**, 1154–1159 (2006).
40. Hartmann, C.H. & Klein, C.A. Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Res.* **34**, e143 (2006).
41. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).
42. Klein, C.A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
43. Klein, C.A., Zohlnhöfer, D., Petat-Dutter, K. & Wendler, N. Gene expression analysis of a single or few cells. *Curr. Protoc. Hum. Genet.* 11.8.1–11.8.18 (2005).
44. Trimarchi, J.M., Stadler, M.B. & Cepko, C.L. Individual retinal progenitor cells display extensive heterogeneity of gene expression. *PLoS ONE* **3**, e1588 (2008).
45. Trimarchi, J.M. *et al.* Molecular heterogeneity of developing retinal ganglion and amacrine cells revealed through single cell gene expression profiling. *J. Comp. Neurol.* **502**, 1047–1065 (2007).
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.* **22**, 1617–1635 (2008).
47. Tang, F. *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
48. Ramos, C.A. *et al.* Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. *PLoS Genet.* **2**, e159 (2006).
49. Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011–1014 (2006).
50. Taniguchi, K., Kajiyama, T. & Kambara, H. Quantitative analysis of gene expression in a single cell by qPCR. *Nat. Methods* **6**, 503–506 (2009).
51. Reiter, M. *et al.* Quantification noise in single cell experiments. *Nucleic Acids Res.* **39**, e124 (2011).
52. Stahlberg, A., Hakansson, J., Xian, X., Semb, H. & Kubista, M. Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.* **50**, 509–515 (2004).
53. Zhao, B. *et al.* Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. *Genes Dev.* **26**, 54–68 (2012).
54. Geller, S.F., Lewis, G.P. & Fisher, S.K. FGFR1, signaling, and AP-1 expression after retinal detachment: reactive Muller and RPE cells. *Invest. Ophthalmol. Vis. Sci.* **42**, 1363–1369 (2001).
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.* **156**, 772–777 (1996).
56. Janes, K.A., Wang, C.C., Holmberg, K.J., Cabral, K. & Brugge, J.S. Identifying single-cell molecular programs by stochastic profiling. *Nat. Methods* **7**, 311–317 (2010).
57. Raj, A. & van Oudenaarden, A. Single-molecule approaches to stochastic gene expression. *Annu. Rev. Biophys.* **38**, 255–270 (2009).
58. Emmert-Buck, M.R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
59. 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.* **15**, 1388–1392 (2005).
60. Limpert, E., Stahel, W.A. & Abbt, M. Log-normal distributions across the sciences: keys and clues. *Bioscience* **51**, 341–352 (2001).
61. Kamme, F. *et al.* Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J. Neurosci.* **23**, 3607–3615 (2003).
62. Wang, L., Brugge, J.S. & Janes, K.A. Intersection of FOXO- and RUNX1-mediated gene expression programs in single breast epithelial cells during morphogenesis and tumor progression. *Proc. Natl. Acad. Sci. USA* **108**, E803–812 (2011).
63. Janes, K.A. RUNX1 and its understudied role in breast cancer. *Cell Cycle* **10**, 3461–3465 (2011).
64. Wang, C.C., Jamal, L. & Janes, K.A. Normal morphogenesis of epithelial tissues and progression of epithelial tumors. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4**, 51–78 (2012).
65. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
66. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. & Noble, W.S. Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).
67. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409 (2012).
68. Ellis, M.J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
69. Stewart-Ornstein, J., Weissman, J.S. & El-Samad, H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol. Cell* **45**, 483–493 (2012).
70. Raj, A., Rifkin, S.A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918 (2010).
71. Eldar, A. *et al.* Partial penetrance facilitates developmental evolution in bacteria. *Nature* **460**, 510–514 (2009).
72. O'Neill, R.A. *et al.* Isoelectric focusing technology quantifies protein signaling in 25 cells. *Proc. Natl. Acad. Sci. USA* **103**, 16153–16158 (2006).
73. Jain, A. *et al.* Probing cellular protein complexes using single-molecule pull-down. *Nature* **473**, 484–488 (2011).
74. Rapkiewicz, A. *et al.* The needle in the haystack: application of breast fine-needle aspirate samples to quantitative protein microarray technology. *Cancer* **111**, 173–184 (2007).
75. Adli, M. & Bernstein, B.E. Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq. *Nat. Protoc.* **6**, 1656–1668 (2011).
76. Wu, A.R. *et al.* Automated microfluidic chromatin immunoprecipitation from 2,000 cells. *Lab Chip* **9**, 1365–1370 (2009).
77. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603 (2006).
78. 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.* **199**, 431–436 (2003).
79. Wang, H. *et al.* Histological staining methods preparatory to laser capture microdissection significantly affect the integrity of the cellular RNA. *BMC Genomics* **7**, 97 (2006).
80. Brady, G. & Iscove, N.N. Construction of cDNA libraries from single cells. *Methods Enzymol.* **225**, 611–623 (1993).
81. Miller-Jensen, K., Janes, K.A., Brugge, J.S. & Lauffenburger, D.A. Common effector processing mediates cell-specific responses to stimuli. *Nature* **448**, 604–608 (2007).
82. Nagy, Z.B. *et al.* Real-time polymerase chain reaction-based exponential sample amplification for microarray gene expression profiling. *Anal. Biochem.* **337**, 76–83 (2005).
83. Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **34**, e42 (2006).
84. Cox, W.G., Beaudet, M.P., Agnew, J.Y. & Ruth, J.L. Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. *Anal. Biochem.* **331**, 243–254 (2004).
85. Shi, L. *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* **24**, 1151–1161 (2006).
86. Warren, L., Bryder, D., Weissman, I.L. & Quake, S.R. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proc. Natl. Acad. Sci. USA* **103**, 17807–17812 (2006).
87. McKay, A.T. Distribution of the coefficient of variation and the extended 't' distribution. *J. Roy. Stat. Soc.* **95**, 695–698 (1932).
88. Sokal, R.R. & Rohlf, F.J. *Biometry* 4th edn. (W.H. Freeman and Company, 2012).
89. Sheskin, D.J. *Handbook of Parametric and Nonparametric Statistical Procedures* 4th edn. (Chapman & Hall, 2007).
90. Kaelin, W.G. Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science* **337**, 421–422 (2012).
91. Fend, F. *et al.* Immuno-LCM: laser capture microdissection of immunostained frozen sections for mRNA analysis. *Am. J. Pathol.* **154**, 61–66 (1999).
92. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, RESEARCH0034 (2002).