SCIENTIFIC REPORTS

OPEN

In situ 10-cell RNA sequencing in tissue and tumor biopsy samples

Shambhavi Singh[1], Lixin Wang[1], Dylan L. Schaff[1], Matthew D. Sutcliffe[1], Alex F. Koeppel[2], Jungeun Kim[3], Suna Onengut-Gumuscu[4], Kwon-Sik Park[3], Hui Zong[3] & Kevin A. Janes[1,4,5]

Received: 23 November 2018
Accepted: 4 March 2019
Published online: 18 March 2019

Single-cell transcriptomic methods classify new and existing cell types very effectively, but alternative approaches are needed to quantify the individual regulatory states of cells in their native tissue context. We combined the tissue preservation and single-cell resolution of laser capture with an improved preamplification procedure enabling RNA sequencing of 10 microdissected cells. This *in situ* 10-cell RNA sequencing (10cRNA-seq) can exploit fluorescent reporters of cell type in genetically engineered mice and is compatible with freshly cryoembedded clinical biopsies from patients. Through recombinant RNA spike-ins, we estimate dropout-free technical reliability as low as ~250 copies and a 50% detection sensitivity of ~45 copies per 10-cell reaction. By using small pools of microdissected cells, 10cRNA-seq improves technical per-cell reliability and sensitivity beyond existing approaches for single-cell RNA sequencing (scRNA-seq). Detection of low-abundance transcripts by 10cRNA-seq is comparable to random 10-cell groups of scRNA-seq data, suggesting no loss of gene recovery when cells are isolated *in situ*. Combined with existing approaches to deconvolve small pools of cells, 10cRNA-seq offers a reliable, unbiased, and sensitive way to measure cell-state heterogeneity in tissues and tumors.

Tumors are complex mixtures of cells that are heterogeneous in their genetics, lineage, and microenvironment[1,2]. Whole-tumor profiles of genes and transcript abundances yield inter-tumor differences that are clinically important for patient prognosis[3–5], but these cellular profiles are population averages[6]. The tumor microenvironment contains several different cell types that vary among cases[7–12]. At the single-cell level, cancer cells are heterogeneous and genetic subclones evolve as the disease progresses[13,14]. Tumor cells also display non-genetic heterogeneity and can switch between regulatory states in a reversible and context-dependent manner[15–17]. Together, these variations dictate phenotypic differences such as proliferative index, metastatic potential, and response to therapy[16,18–22].

Assessing intra-tumor heterogeneity of gene regulation requires precise transcriptomic measurements of a very small number of cells isolated from within the tumor context. The current methods for single-cell RNA sequencing (scRNA-seq) are powerful in their ability to profile thousands of individual cells and identify differences in genotype or lineage in a mixed population. However, the first step of most large-scale scRNA-seq methods is some form of tissue dissociation and single-cell isolation, which can alter transcriptional profiles and confound downstream analyses[23,24]. Approaches such as laser-capture microdissection (LCM) can obtain samples for RNA-seq[25–28], but they usually require so many cells for reliable measurement that single-cell variation is obscured (Supplementary Fig. S1). Dissociation-based scRNA-seq methods also struggle with technical variability, including "dropout" of medium-to-low abundance transcripts that yield zero aligned reads[29–32]. The 3–20% conversion efficiency[29,30,33–35] of RNA to amplifiable cDNA is problematic given estimates that 90% of the transcriptome is expressed at 50 copies or fewer per cell[36]. While valid for the most consistently expressed genes and markers within a sample, scRNA-seq data miss a large proportion of the transcriptome[36,37]. Measuring single-cell expression profiles *in situ* is even more challenging because of losses incurred during biomolecule extraction as well as non-mRNA contaminants, which can be considerable in stroma-rich specimens. Collectively, these hurdles make it difficult to measure tumor-cell regulatory heterogeneities reliably and evaluate their functional consequences.

[1]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, USA. [2]Bioinformatics Core, University of Virginia, Charlottesville, VA, 22908, USA. [3]Department of Microbiology, Immunology & Cancer Biology, University of Virginia, Charlottesville, VA, 22908, USA. [4]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA. [5]Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, 22908, USA. Shambhavi Singh and Lixin Wang contributed equally. Correspondence and requests for materials should be addressed to K.A.J. (email: kjanes@virginia.edu)

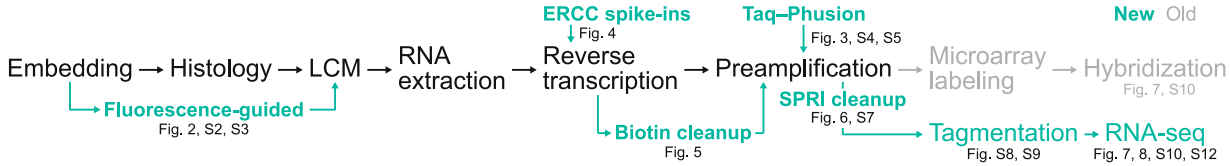

Figure 1. A revised transcriptomic pipeline for *in situ* 10-cell RNA sequencing. Substantive changes are indicated in green and gray.

Multiple studies have reported a pronounced improvement in gene detection and technical reproducibility when using 10–30 cells of starting material rather than one cell[28,35,38–42]. The increased cellular RNA offsets losses incurred during reverse transcription, enabling more reliable downstream amplification. The gains are irrespective of amplification strategy and detection platform, and they are more dramatic than when increasing the starting material another tenfold to 100 cells. Previously, we combined the technical advantages of 10-cell pooling with the *in-situ* fidelity of LCM to devise a random-sampling method called "stochastic profiling"[41,42]. The method identifies single-cell regulatory heterogeneities by analyzing the statistical fluctuations of transcriptomes measured repeatedly as 10-cell pools microdissected from a cell lineage[41,43]. Pooling increases gene detection and technical reproducibility; repeated sampling is used to extract the single-cell information that is retained in pools of 15 cells or smaller (Supplementary Fig. S1). Genes with bimodal regulatory states[44] create skewed deviations from a null model of biological and technical noise, which parameterize the underlying population-level distribution more accurately than single-cell measurements[39,45]. By applying stochastic profiling to spatially organized breast-epithelial spheroids and gene panels measured by quantitative PCR or microarray, we uncovered multiple regulatory states relevant to 3D organization and stress responses[18,46,47]. However, this early work did not stringently evaluate the importance of sample integrity for primary tissues from animals or patients, nor did it involve probe-free measures of 10-cell data like RNA sequencing.

Here, we report improvements in sample handling, amplification, and detection that enable RNA sequencing of 10-cell pools isolated from tissue and tumor biopsies by LCM and its extensions. We find that cryoembedding of freshly isolated tissue pieces is crucial to preserve the localization of genetically encoded fluorophores in engineered mice used for fluorescence-guided LCM. By incorporating ERCC spike-ins at non-disruptive input amounts in the amplification, we calibrate sensitivity and provide a standard reference to compare with other scRNA-seq methods[48]. Sample tagging and fragmentation (tagmentation) is accomplished by Tn5 transposase[49], which is compatible with the revised procedure as well as with past 10-cell amplifications. We sequence archival samples that had previously been measured by BeadChip microarray to provide a side-by-side comparison of transcriptomic platforms with limiting material[41,50]. Applying 10-cell RNA sequencing (10cRNA-seq) to various mouse and human cell types isolated by LCM, we obtain substantially better exonic alignments, with increases in gene coverage that are consistent with the single-cell sensitivity of prevailing scRNA-seq methods. The realization of 10cRNA-seq by LCM creates new opportunities for stochastic profiling[45] and other unmixing approaches[39] to deconvolve single-cell regulatory states *in situ*.

Results

Methods for profiling small quantities of cellular RNA have evolved considerably over the past decade, but they all involve the same fundamental steps: (1) cell isolation, (2) RNA extraction, (3) reverse transcription, (4) preamplification, and (5) detection[51]. The original protocol for *in situ* 10-cell profiling combines LCM for cell isolation followed by proteinase K digestion for RNA extraction[42]. The extracted material undergoes an abbreviated high-temperature reverse transcription with oligo$(dT)_{24}$, and cDNA is carefully preamplified by poly(A) PCR[52] that generates sufficient 3′ ends (~500 bp in size) for microarray labeling and hybridization[42] (Fig. 1).

Unsurprisingly, the earliest steps in the procedure are the most critical for achieving the maximum amount of amplifiable starting material. To avoid losses, steps 1–4 (cell isolation through preamplification) are normally performed without intermediate purification. Therefore, buffers and reagents must be carefully tested and titrated to be mutually compatible throughout the "one-pot" protocol. Since description of the procedure[41,42], multiple commercial providers merged or were acquired, leading to the discontinuation of multiple RNAse inhibitors, the Taq polymerase, and the BeadChip microarrays. The collective disruptions in sourcing prompted a modernization of 10-cell profiling toward RNA-seq of primary material at a biopsy scale, including how tissue–tumor samples were handled before the start of the procedure (Fig. 1).

Protein localization for LCM requires fresh cryoembedding. To minimize extra handling steps that could degrade RNA, *in situ* profiling of clinical samples is ordinarily performed with rapid histological stains[41,51,53,54] (Fig. 1). LCM can also be guided by fluorescence in place of histology when using cells or animals engineered to encode genetic labels[55,56]. However, new challenges arise when seeking to preserve localization and brightness of encoded fluorophores during single-cell isolation and RNA extraction. Compared to polysome-bound mRNAs, fluorescent proteins diffuse much more readily, and chromophores may be damaged by the fixation and dehydration steps needed to preserve RNA integrity. Fluorescent-protein structure is preserved by chemical fixatives, but covalent crosslinking of biomolecules is unsuitable for extracting RNA from tissue. Fluorescence-guided profiling therefore entails a competing set of tradeoffs that must be balanced for optimal performance.

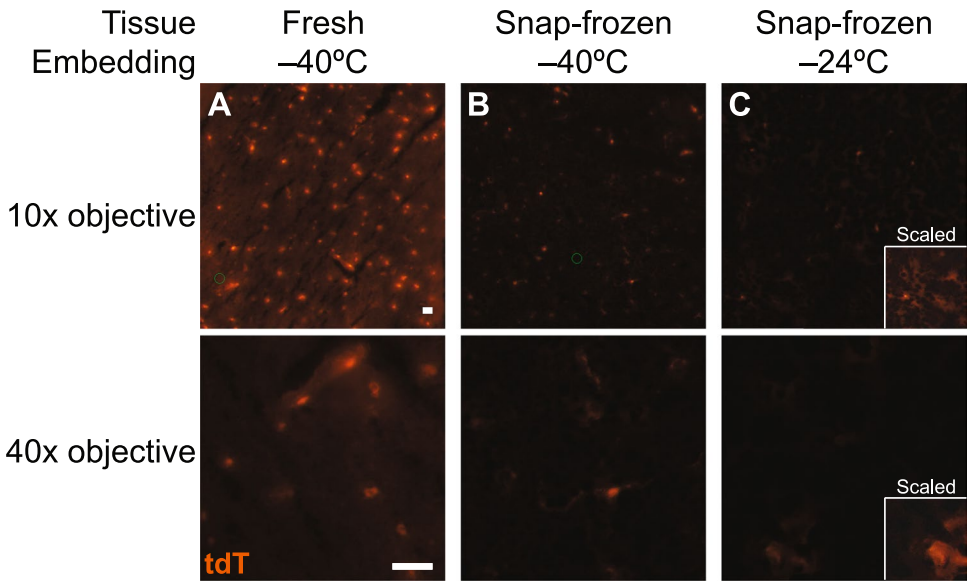

Figure 2. Fresh cryoembedding preserves tandem-dimer Tomato (tdT) fluorescence and localization better than snap-frozen alternatives. Brain samples from *Cspg4-CreER;Trp53$^{F/F}$;Nf1$^{F/F}$;Rosa26-LSL-tdT* animals were (**A**) freshly cryoembedded in Neg-50 medium with dry ice-isopentane (−40 °C), (**B**) snap-frozen in dry ice-isopentane and then cryoembedded, or (**C**) snap-frozen and slowly cryoembedded in a cryostat (−24 °C). Low- and high-magnification images were captured with the factory-installed color camera on the Arcturus XT LCM instrument. Images were exposure matched and are displayed with a gamma compression of 0.67. Insets have been rescaled to emphasize tdT diffusion away from the cell body. Scale bar is 25 µm. Brightfield images from the same sections are shown in Supplementary Fig. S3.

We reasoned that the greatest flexibility would be afforded by reporter mice expressing tandem-dimer Tomato (tdT)—a bright, high molecular-weight derivative of DsRed[57]. Key handling parameters were evaluated using *Cspg4-CreER;Trp53$^{F/F}$;Nf1$^{F/F}$;Rosa26-LSL-tdT* mice, a model of malignant glioma[58]. In these animals, administration of tamoxifen elicits sparse labeling of oligodendrocyte precursor cells (OPCs) in the brain, enabling fluorescence retention to be assessed in single cells. Extensive optimization of cryosectioning and wicking conditions was required to preclude fluorophore diffusion while ensuring reliable LCM pickup (see Methods). We found that an accelerated 70–95–100% ethanol series[41,42] maintained tdT fluorescence and localization of labeled cells through xylene clearing and dehydration (Fig. 2A). Separately, using freshly embedded tissue from a "mosaic analysis of double markers" (MADM) animal that labels various brain lineages with enhanced green fluorescent protein (EGFP), tdT, or both[59,60], we confirmed that EGFP fluorescence was also acceptably retained with the 70–95–100% ethanol series (Supplementary Fig. S2). Although EGFP diffusion was noticeably greater compared to tdT owing to its smaller size (~28 kDa vs. ~54 kDa), we could nonetheless reliably identify the cell bodies of single EGFP-positive cells for LCM. Surprisingly, we found that fresh-tissue embedding was critically important for preserving single-cell localization and brightness. Snap-freezing before cryoembedding caused considerable loss and delocalization of tdT fluorescence, even when prefrozen material was rapidly embedded in dry ice-isopentane (−40 °C) (Fig. 2B,C). Brightfield images of these cryosections also showed considerable tissue damage compared to freshly embedded material (Supplementary Fig. S3). For mechanically challenging tissues in which embedding support is important for cryosectioning, we conclude that fresh-tissue embedding is essential for maximum biomolecular retention and integrity.

Improving poly(A) preamplification for modern RNA-seq. Previously, *in situ* 10-cell profiling was optimized for quantification by BeadChip microarray[41,42], but microarrays have been supplanted by RNA-seq for unbiased measures of the transcriptome[61] (Fig. 1). An advantage of RNA-seq is that nucleic acids are detected regardless of origin, enabling use of exogenous RNA standards to calibrate sensitivity and quantitative accuracy when spiked into a biological sample[62–64]. The versatility of RNA-seq is also a caveat, because all nucleic acids in a sample will be sequenced, including unwanted preamplification byproducts and contaminating DNA from mitochondria or the nucleus[65–67]. In the original scRNA-seq report that used a variant of poly(A) PCR, only 37 ± 9% of sequenced reads aligned to RefSeq transcripts[68], and exonic alignment rates below 50% remain common[69]. Therefore, we focused improvements to poly(A) preamplification towards ensuring that most sequencing reads aligned to the 3′ ends of cellular mRNAs.

In poly(A) PCR, cDNA is 3′ adenylated and then preamplified with a universal T_{24}-containing primer called AL1[52]. We previously found that the amount of AL1 strongly influenced overall sensitivity of gene detection, with improvements noted at concentrations as high as 25 µM[42]. Excess AL1 also drives nonspecific amplification of low molecular-weight primer concatemers[70], which do not influence gene measurements by quantitative PCR or microarray but create overwhelming contamination for RNA-seq. To improve poly(A) PCR, we screened a range of commercial Taq and proofreading polymerases along with empirical blends of those that maximized the

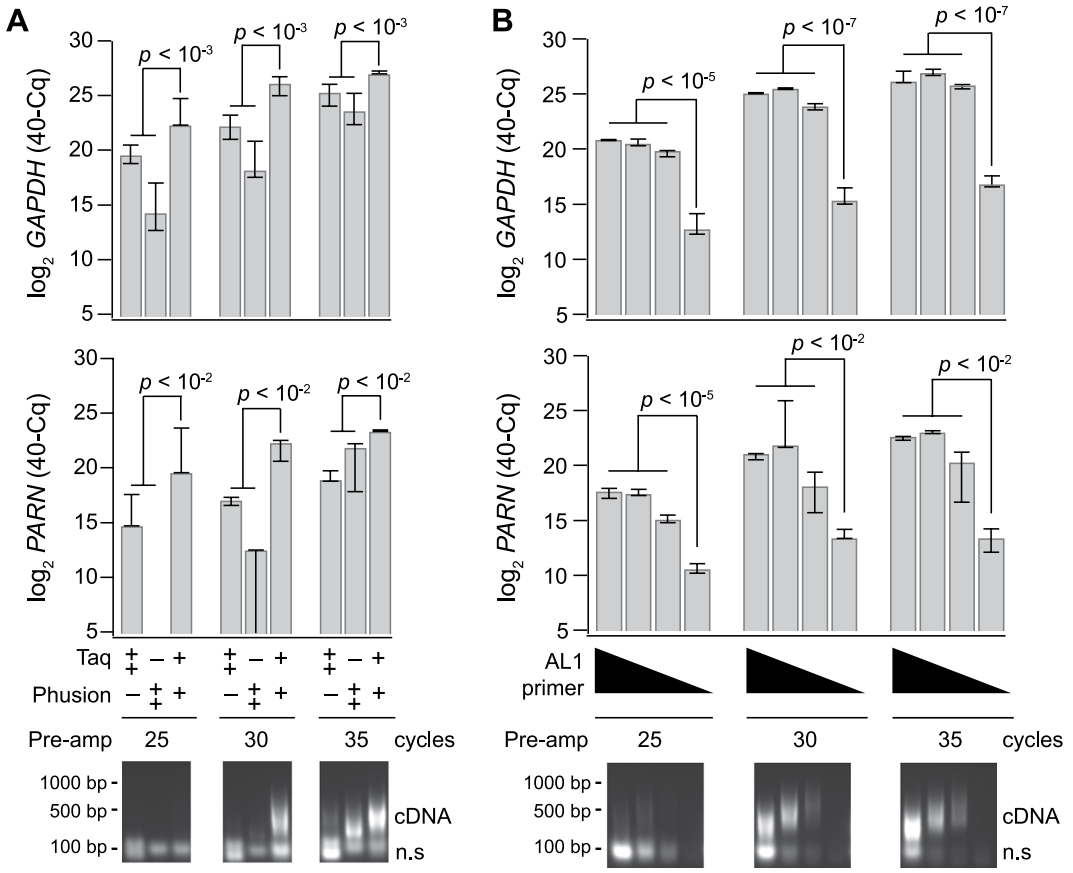

Figure 3. A blend of Taq–Phusion polymerases improves selective poly(A) amplification of cDNA and reduces AL1 primer requirements. Cells were obtained by LCM from a human breast biopsy and split into 10-cell equivalent amplification replicates. (**A**) Poly(A) PCR was performed with 15 µg of AL1 primer with Taq alone (10 units), Phusion alone (4 units) or Taq/Phusion combination (3.75 units/1.5 units). (**B**) Poly(A) PCR was performed with either 25, 5, 2.5, or 0.5 µg of AL1 primer and the Taq–Phusion blend from (A). Above—Relative abundance for the indicated genes and preamplification conditions was measured by quantitative PCR (qPCR). Data are shown as the median inverse quantification cycle (40–Cq) ± range from $n = 3$ amplification replicates and were analysed by two-way (A) or one-way (B) ANOVA with replication. Below—Preamplifications were analysed by agarose gel electrophoresis to separate poly(A)-amplified cDNA from nonspecific, low molecular-weight concatemer (n.s.). Qualitatively similar results were obtained separately three times. Lanes were cropped by poly(A) PCR cycles for display but were electrophoresed on the same agarose gel and processed identically. The uncropped image is shown in Supplementary Fig. S13A.

intended ~500 bp cDNA products relative to nonspecific concatemer. We obtained a better-than-additive pre-amplification by combining Taq and Phusion polymerases (see Methods). An equal mixture of the two enzymes dramatically increased the yield of ~500 bp preamplification products relative to nonspecific concatemer (Fig. 3A, lower). The empirical blend also significantly improved the preamplification of both high-abundance (*GAPDH*) and low-abundance (*PARN*) targets as measured by quantitative PCR (Fig. 3A, upper). The two-enzyme blend further enabled a 10-fold decrease in AL1 primer concentration without detectable loss in preamplification efficiency (Fig. 3B). The Taq-Phusion combination was superior for a primary breast-cancer biopsy (Fig. 3) as well as two murine tissue sources: a murine small-cell lung cancer line derived from $Trp53^{\Delta/\Delta}Rb^{\Delta/\Delta}$ lung epithelium[71] and tdT-labeled OPCs (Supplementary Figs S4 and S5), illustrating its generality. The enzyme modification created a viable starting point for combining poly(A) PCR preamplification with RNA-seq.

Sensitivity, accuracy, and precision of the updated poly(A) PCR approach were assessed using recombinant RNA spike-ins as internal positive controls[64]. A dilution of ERCC spike-ins was defined that did not detectably perturb the measured abundance of endogenous transcripts in RNA equivalents from 10 microdissected cells (Fig. 4A). After poly(A) PCR of the spike-in dilution plus 100 pg RNA (~10 cells), we measured the relative abundance of individual spike-ins, using quantitative PCR (qPCR) to eliminate RNA-seq read depth as a complicating factor. Purified qPCR end products served as an absolute reference of each spike-in for cross-comparison (see Methods). We observed good linearity across 22 spike-ins spanning an abundance of ~10^4 (Fig. 4B). Deviations, technical noise, and dropouts all increased considerably for spike-ins below ~250 copies per reaction, consistent with previous reports[29]. This collective measurement uncertainty restricts interpretation of single-cell data to highly expressed transcripts, but 10-cell pooling reduces the threshold to ~25 copies on average per cell. With

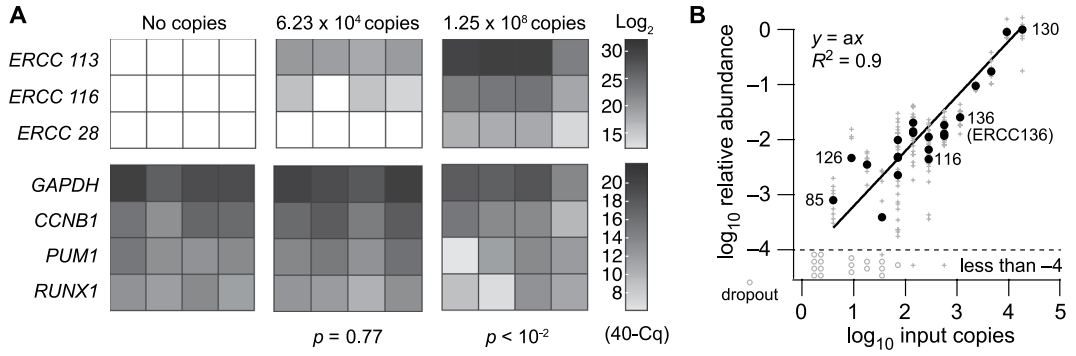

Figure 4. Optimized ERCC spike-in dilutions assess poly(A) PCR sensitivity and dynamic range without suppressing cDNA amplification of endogenous transcripts. (**A**) 100 pg RNA was supplemented with ERCC Mix 1 at the indicated dilutions and amplified via optimized poly(A) PCR. ERCC and endogenous gene abundances were measured by qPCR, and data are shown in grayscale as the inverse quantification cycle (40–Cq) from $n = 4$ amplification replicates. Negative effects of the ERCC spike-ins on endogenous genes (lower) were assess by two-way ANOVA with replication. (**B**) ERCC Mix 1 (6.23×10^4 copies) was spiked into 100 pg RNA and amplified via optimized poly (A) PCR. Proportional abundance of ERCC standards was estimated with a seven-log dilution series from purified qPCR end products. Data are shown as the median 40–Cq (black) for 22 ERCC spike-in standards from $n = 8$ amplification replicates (gray) with undetected "dropouts" reported below (circles).

poly(A) PCR, we did not observe qualitative dropout in more than 50% of technical replicates for spike-ins as dilute as four copies per reaction (ERCC85; Fig. 4B), indicating good sensitivity. RNA spike-ins do not mimic the characteristics of endogenous transcripts extracted from cells, but they can provide a common reference to benchmark preamplification methods for RNA-seq[48]. These experiments indicated that the improved poly(A) preamplification was sufficiently reliable for unbiased profiling of 10-cell transcriptomes.

For RNA extraction from the LCM cap, an optimized digestion buffer is used containing proteinase K to release mRNAs from precipitated ribosomes[41]. Proteinase K also digests nucleosomes, which may cause elution of contaminating genomic DNA. In past and current analyses of human LCM samples preamplified ± reverse transcription, we never found genomic copies of genes amplified within ~0.4% of measured mRNA transcripts ($\Delta Cq \geq 8$ for 16 genes measured in four human cell types, Supplementary Fig. S6). For mouse tissues, however, genomic copies were more prevalent and variable, with some genes measured as abundantly without reverse transcription as with it (Figs 5A and S6). Gel electrophoresis showed weak-but-detectable bands above the desired ~500 bp product in preamplifications without reverse transcription, implying nonspecific amplification (Fig. 5A, lower). Concerned that the murine genome could compete with the amplification of cDNA, we appended an intermediate purification following reverse transcription with 5′-biotin-modified oligo$(dT)_{24}$. Biotinylated cDNA was purified on streptavidin-conjugated magnetic beads, which could be separated from contaminants in the LCM extract and used as a starting template for poly(A) preamplification. Addition of the biotin cleanup step mildly improved the amplification of cDNAs and, importantly, eliminated the confounding abundance of murine genomic DNA (Fig. 5B). We recommend biotinylated oligo$(dT)_{24}$ and bead purification for mouse samples considering the recurrent challenges with genomic DNA (Supplementary Fig. S6 and see Discussion).

Poly(A) PCR samples are kept dilute to avoid saturating the preamplification, but aliquots can be carefully reamplified up to microgram scale for microarray hybridization[41,42]. In preparing libraries for sequencing, we pursued tagmentation using Tn5 transposase because addition of sequencing adapters is sterically impossible within the ~40 bp distal ends of a PCR amplicon[72]. The steric restrictions of Tn5 were advantageous for pruning away the long, A-repetitive universal primer from poly(A) amplicons that would otherwise be wastefully sequenced. Commercial Tn5 tagmentation kits (Nextera XT) require 1000-fold less material than past microarray hybridizations, prompting reevaluation of how the 10-cell libraries were prepared. We retained the mid-logarithmic reamplification approach described previously[41] but substituted paramagnetic Solid Phase Reversible Immobilization (SPRI) beads for library purification[73]. Two rounds of purification with 70% (vol/vol) SPRI beads eliminated ~99% of primer dimers and concatemers in 10-cell reamplifications from various sources (Figs 6 and S7). Reamplified samples yielding at least 200 ng of purified product (Supplementary Fig. S8) were tagmented at 1-ng scale according to the Nextera XT protocol. Although poly(A) amplicon sizes are centered at ~500 bp (Fig. 6A), we found that the higher SPRI bead ratio recommended for 300–500 bp inputs (180% [vol/vol] beads) was essential for purification of tagmented libraries (Supplementary Fig. S9). Under these conditions, both new and archival poly(A) PCR preamplifications are compatible with RNA sequencing.

Paired comparison of 10-cell transcriptomics by BeadChip microarray and RNA-seq. Poly(A) PCR provides an abundant source of material for transcript quantification, creating an opportunity to revisit 10-cell samples profiled earlier on BeadChip microarrays. In the original application of stochastic profiling, 10-cell samples were locally microdissected from 3D spheroids of a clonal human breast-epithelial cell line[41]. We sequenced 18 biological replicates from this study (6.6 ± 2.3 million reads) along with three 10-cell pool-and-split controls that assessed technical variability[33,41]. Technical correlation was as high within pool-and-split replicates

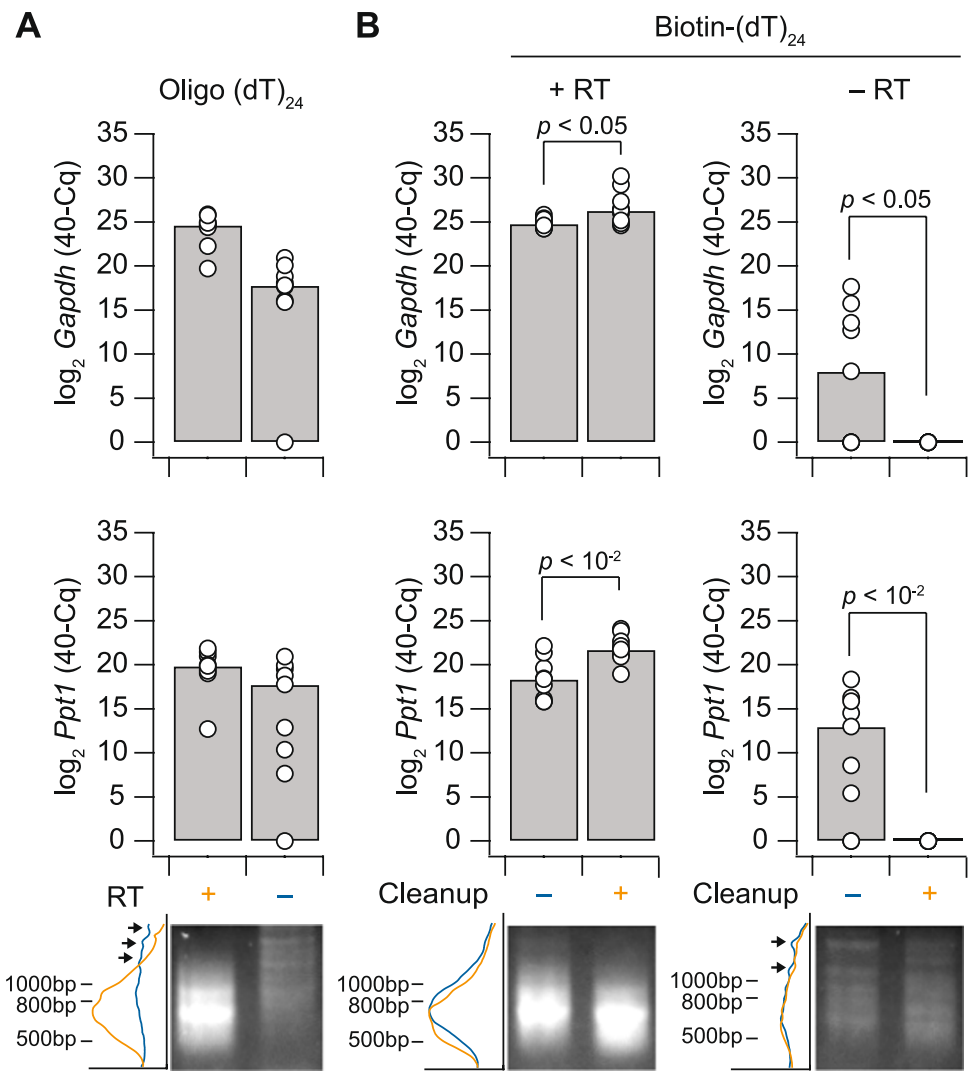

Figure 5. Poly(A) amplification of murine sequences without reverse transcription is eliminated with 5′-biotin-modified oligo(dT)$_{24}$ and streptavidin bead cleanup. (**A**) Reverse transcription-free preamplification of genomic DNA confounds accurate quantification of some mRNAs. (**B**) Bead cleanup eliminates nonspecific preamplification of genomic DNA. Above—Data are shown as the median inverse quantification cycle (40–Cq, gray) of $n = 3$ independent experiments (three amplification replicates per experiment). Differences with and without bead cleanup were assessed by Wilcoxon rank sum test. Below—Preamplifications were analysed by agarose gel electrophoresis to separate poly(A)-amplified cDNA and genomic amplification. Electrophoretic traces were analysed by densitometry to the left of the image, with genomic amplicons highlighted (arrows). Lanes were cropped by the indicated conditions for display but were electrophoresed on the same agarose gel and processed identically. The uncropped image is shown in Supplementary Fig. S13D.

measured by RNA-seq as when the same replicates were measured by microarray ($R \sim 0.9$; Fig. 7B,C,D,F–H). For both platforms, undetectable genes in one technical replicate were quantified up to ~$10^2 = 100$ transcripts per million (TPM) or ~$10^{3.3} = 2000$ BeadChip fluorescence intensity in another replicate. Among detected genes with at-least one technical replicate yielding zero measured TPM, we found that RNA-seq correlated with BeadChip intensity across replicates ($R \sim 0.4$, $p \sim 0$; Supplementary Fig. S10A). The concordance between the two platforms strongly argues that transcript losses are authentic dropout events[74], not artifacts of RNA-seq read depth or BeadChip detection sensitivity. Combining the reliable detection limits of 100 TPM (Fig. 7B,C,F) and ~250 ERCC copies/reaction (Fig. 4B), we predict (250 copies/reaction)/(10 cells/reaction x 100 TPM) = 250,000 mRNA copies per cell, consistent with published estimates[38].

When 10-cell transcript representation was compared, we found that RNA-seq TPM and BeadChip microarray intensities were correlated ($R \sim 0.6$; Fig. 7A,E,I), albeit not as strongly as reported elsewhere[50,75]. Some genes yielded background fluorescence on microarrays but moderate-to-high TPM, likely due to BeadChip probe sequences absent from the amplicons generated by poly(A) PCR. Among genes with a median TPM > 1000 by RNA-seq, we identified 27 BeadChip probes exhibiting a median fluorescence less than $10^{2.5}$. The median distance of the 27 probes from the 3′ end of the corresponding gene was 845 bases (interquartile range: 492–1392 bases),

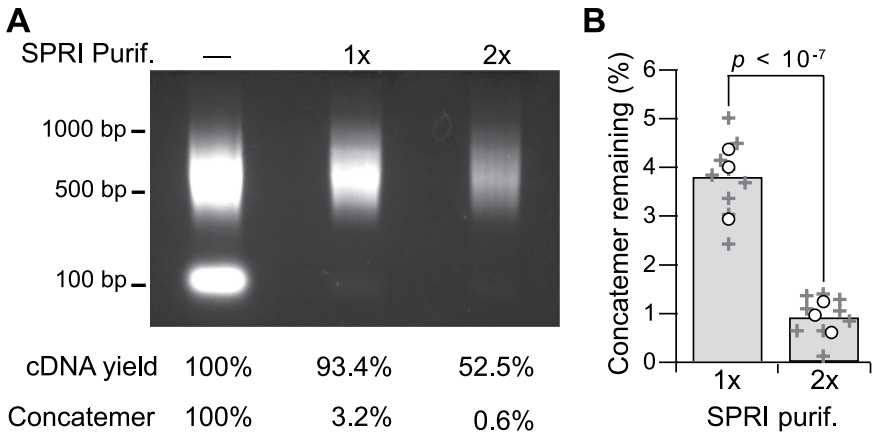

Figure 6. Iterative SPRI bead purification eliminates low molecular-weight contaminants before tagmentation. (**A**) Poly(A) PCR reamplifications[41] of 10-cell human breast cancer samples were analysed by gel electrophoresis without purification or after one (1×) or two (2×) rounds of purification with 70% (vol/vol) SPRI beads. The uncropped image is shown in Supplementary Fig. S13E. (**B**) Contaminating low molecular-weight concatemers are significantly reduced after two rounds of SPRI bead purification. Data are shown as the mean (gray) of $n = 3$ independent reamplifications (circles) each purified three times (+). Differences were assessed by two-way ANOVA with replication. The uncropped gel image used for concatemer densitometry is shown in Supplementary Fig. S13F (upper).

upstream of the distal ~500 bp 3′ ends amplified by poly(A) PCR. The probe-independent nature of RNA-seq reinforces one of its critical advantages for 10-cell transcriptomics.

We also evaluated quantitative concordance of the 18 10-cell samples measured both by BeadChip microarray and RNA-seq. The variance of 7713 genes was twice their mean value measured on each platform, suggesting significant biological variation across the 18 samples ($p < 0.01$). For biologically variable genes, the median sample-by-sample Pearson correlation between BeadChip microarray and RNA-seq was 0.42 (interquartile range: 0.16–0.63), with 599 transcripts showing $R \geq 0.8$ (Supplementary Fig. S10B). Considering a median TPM of 17 (interquartile range: 4–49) for the 10-cell data analysed, these cross-platform correlations fall within the range reported for TCGA microarrays and RNA-seq (R ~ 0.4–0.9)[75]. Our retrospective analysis indicates that 10cRNA-seq data corroborate BeadChip microarrays and provide broader access to 3′ mRNA ends not represented on oligonucleotide probe sets.

Advantages of 10cRNA-seq for diverse mouse and human cell types. Last, we aggregated the intermediate revisions to 10-cell transcriptomic profiling (Fig. 1) and asked whether there were more-overarching benefits to sequencing small pools versus single cells. Different methods for scRNA-seq have already been rigorously compared by multiple groups[48,69]. Since 10-cell sampling could be adopted by many of these approaches, we focused instead on the data quality from published scRNA-seq datasets of various types relative to similar cells profiled by our 10cRNA-seq approach, including biological replicates and pool-and-split controls. We identified two scRNA-seq datasets for murine OPCs[76,77], two for murine lung neuroendocrine cells[78], two for human breast cancer[79,80], and one for MCF-10A cells[81] (Supplementary Table S1). All raw data were identically processed and aligned to the transcriptome with RSEM[82]. Using transcriptome references stringently emphasized exonic read alignments, and the RSEM model for expectation maximization enabled the degeneracy of 3′-end sequences to contribute to transcript quantification. Data quality was gauged by the percentage of reads aligned, and sensitivity was assessed by the number of Ensembl genes with an estimated TPM greater than one.

For the mouse cell types, we observed significant increases in gene detection between 10cRNA-seq and certain scRNA-seq datasets (Fig. 8A). OPCs isolated by fluorescence-guided LCM showed increased gene detection with 10cRNA-seq compared to scRNA-seq of OPCs purified by fluorescence-activated cell sorting (GSE75330)[77]. Gene detection in the sorted OPCs was poorer than when OPCs were collected randomly in a cell atlas of the mouse cortex (GSE60361)[76], emphasizing the stresses caused by non-LCM methods of enrichment. We were unable to detect a significant increase in gene detection between small-cell lung cancer cells profiled by 10cRNA-seq and single neuroendocrine cells randomly dissociated from the mouse airway and profiled by plate-based scRNA-seq[78]. However, neuroendocrine cells are so rare in this tissue that plate-based scRNA-seq was very underpowered ($n = 5$ cells). When droplet-based scRNA-seq was used to increase statistical power to $n = 92$ cells, there was a significant reduction in gene counts compared to 10cRNA-seq profiling the equivalent of 120 cells ($n = 12$ 10-cell replicates). Results were similar but even more striking for human cell types (Fig. 8B). 10cRNA-seq of MCF-10A cells and primary breast cancer cells showed high alignment rates and routinely detected more than 10,000 Ensembl genes, the upper limit for any single cell profiled by three different scRNA-seq methods[79–81]. In cases where gene sensitivities were comparable, we noted dramatically improved alignment rates for 10cRNA-seq (Fig. 8C,D), reinforcing the efficiency of data collection by adopting a 10-cell approach.

The increased detection of transcripts in 10cRNA-seq data could arise from the accumulation of sporadic gene-expression events among single cells in the 10-cell pool. 10cRNA-seq collects 10-cell pools that are

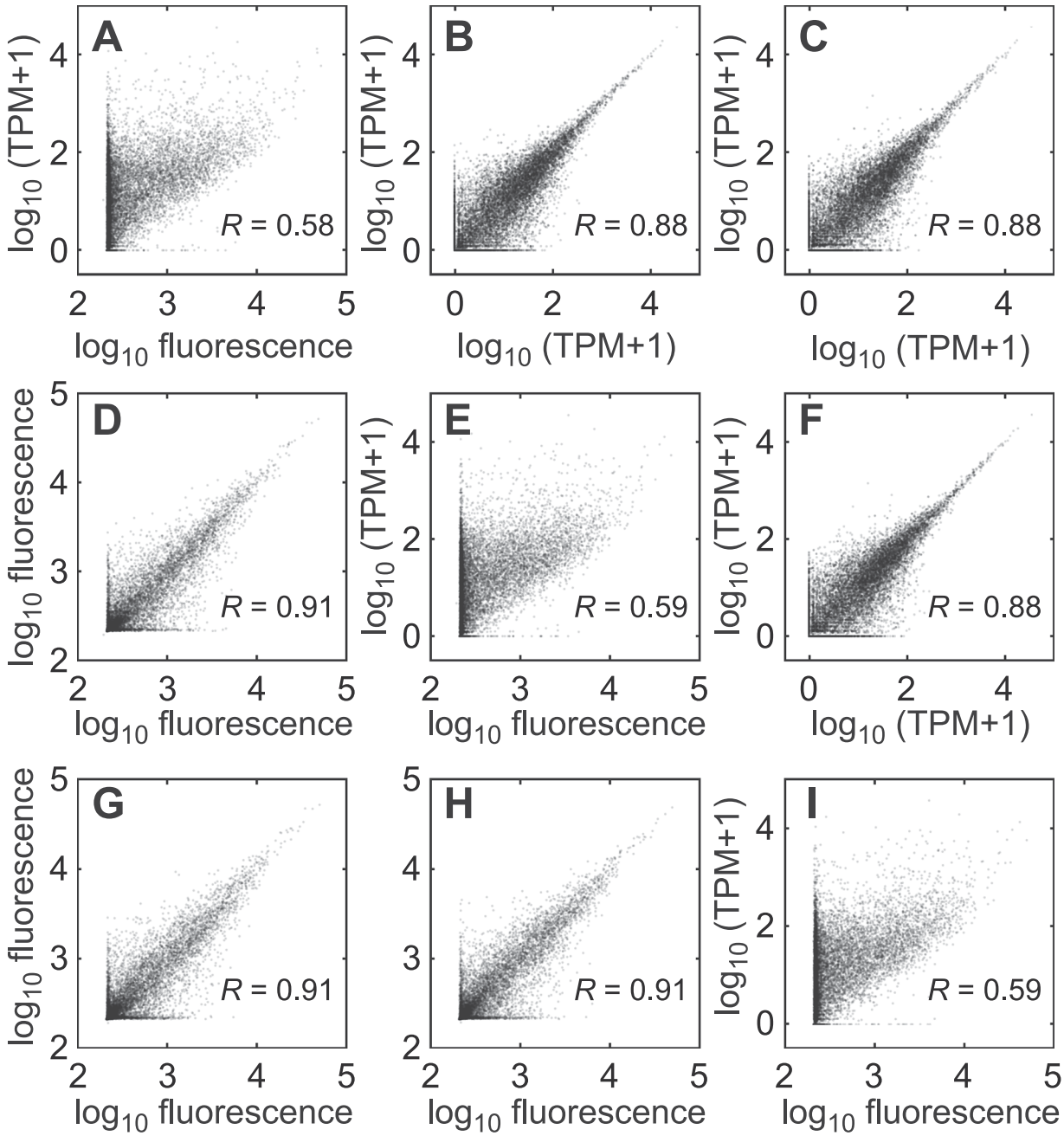

Figure 7. Paired comparison of 10-cell transcriptomes profiled by BeadChip microarray and 10cRNA-seq. **(A–I)** Three pool-and-split 10-cell replicates from before[41] were reamplified, purified, and tagmented for RNA-seq. Inter-replicate correlations among BeadChip microarray triplicates (**D**,**G**,**H**) and 10cRNA-seq triplicates (**B**,**C**,**F**) as well as intra-replicate correlations between platforms (**A**,**E**,**I**) are shown together with the log-scaled Pearson correlation (*R*).

histologically indistinguishable by LCM, but it does not control for noisy transcriptional bursting or differences in cell-cycle phase. To evaluate whether the 10cRNA-seq detection statistics were consistent with those from scRNA-seq data, we randomly combined similar single-cell transcriptomes into 10-cell groups, modeling drop-outs as a binomial probability for RNA-to-cDNA conversion (see Methods). We aggregated 48 random 10-cell assemblies within each of the six scRNA-seq datasets[76–81] and noted a significant increase in gene counts that was comparable to 10cRNA-seq data (Supplementary Fig. S11). On a per-cell basis, 10cRNA-seq matches the gene-recovery sensitivity of scRNA-seq and may be preferable when isolating single cells *in situ* is critical.

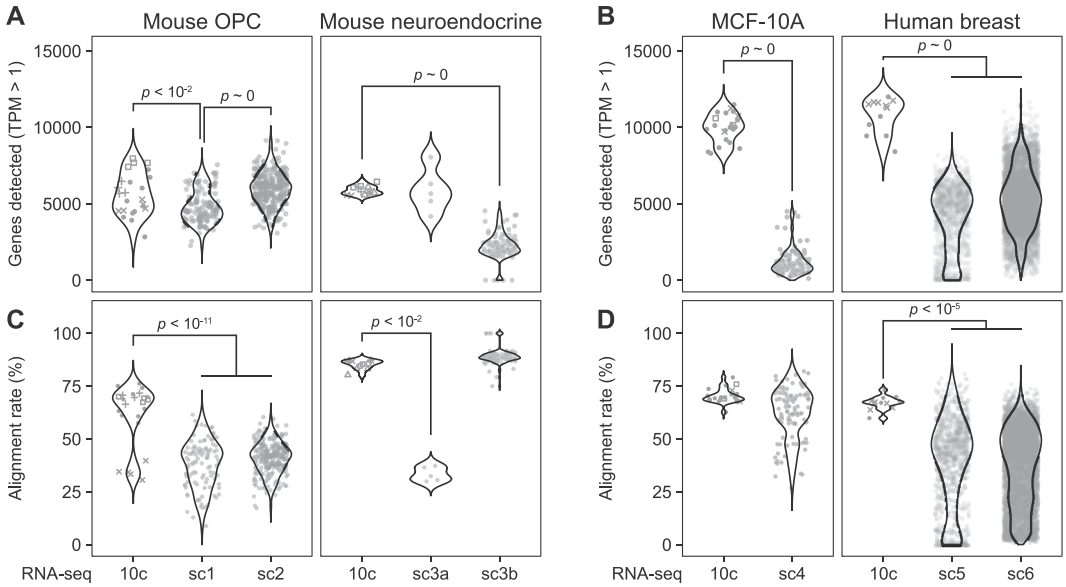

Figure 8. Increased gene detection and improved exonic alignment rates for 10cRNA-seq compared to scRNA-seq. (**A**) Detection of murine Ensembl genes for mouse oligodendrocyte precursor cells (OPCs) and lung neuroendocrine-derived cells. (**B**) Detection of human Ensembl genes for MCF-10A cells and human breast cells. (**C**) Exonic alignment rate comparison for OPCs and lung neuroendocrine-derived cells. (**D**) Exonic alignment rate comparison for MCF-10A cells and human breast cells. Public scRNA-seq data were obtained from the indicated accession numbers: sc1 = GSE75330, sc2 = GSE60361, sc3a = GSE103354 (plate-based), sc3b = GSE103354 (droplet-based), sc4 = GSE66357, sc5 = GSE113197, sc6 = PRJNA396019. 10cRNA-seq data were aggregated from independent 10-cell samples (circles) and 10-cell equivalents from pool-and-split controls. Pool-and-split controls from the same day are indicated with non-circular markers corresponding to the shared day. Pairwise differences between 10-cell and single-cell methods were assessed by permutation test.

Discussion

Single-cell transcriptomics has expanded or rewritten the catalog of cell types in tissues, organs, and organisms[78,83–88]. Yet, scRNA-seq does not obviate the need for complementary approaches, which accurately profile regulatory-state changes within a given cell lineage[43]. The technical advances reported here demonstrate the immediate feasibility of 10cRNA-seq for mouse and human samples obtained *in situ* by LCM. We combined straightforward extensions of ERCC spike-ins and tagmentation with new approaches for fluorescence-guided LCM and cDNA purification that may prove beneficial for other applications (Fig. 1). Although small-sample RNA-seq is never fully dissociated from tissue acquisition or cell handling, our data illustrate a workflow that can be paused and restarted when LCM is used as an intermediate step.

Previous descriptions of fluorescence-guided LCM relied upon exogenous fluorophores added by lectins, antibodies, or viruses[27,55,56,89]. Through careful optimization of cryoembedding and LCM, we identified conditions that preserved the most-common fluorescent proteins used to engineer the mouse germ line. Compatibility with genomically encoded labels creates new opportunities for combining 10cRNA-seq with lineage tracing[90] to examine early regulatory-state changes in development and disease. Compared to fluorophore localization, RNA integrity was not as exquisitely sensitive to sample preparation and handling. Nevertheless, we recommend fresh cryoembedding of all samples in case other protein-guided approaches, such as immuno-LCM[91], might be pursued. The breast core biopsies profiled here were prospectively obtained and cryoembedded during an outpatient procedure. However, a nearly identical protocol has been deployed intra-operatively for surgical pathology[92], implying that fresh cryoembedding is not prohibitive for biobanked clinical samples.

A startling result from the revised protocol was the extent of poly(A) amplification observed in murine samples when reverse transcription was omitted. Nonspecific amplification was not as prominent in human samples obtained by LCM, pointing to specific differences in genome composition and the susceptibility to priming with AL1. A plausible explanation lies in transposable elements—specifically, the distinct classes of short interspersed nuclear elements (SINEs) in rodents and humans[93]. Human-specific Alu SINEs and rodent-specific B-type SINEs both contain stretches of 10–20 As that could partially anneal to the T homopolymer sequence on the 3′ end of AL1[94]. However, to amplify during poly(A) PCR, an antisense SINE must be sufficiently nearby. The mouse genome is ~20% smaller than humans, and B-type SINEs are ~25% more numerous in mice compared to Alu SINEs in humans[93]. The differences reduce the expected spacing of sense-antisense SINEs from ~6 kb in humans to ~4 kb in mice, consistent with a prior analysis of sense-antisense SINEs around transcription start sites[95]. The shorter average spacing may be close enough for genomic fragments to compete with the ~500 bp cDNA amplicons generated during reverse transcription (Figs 3, 5A). Such nonspecific products were prevented from coamplifying with cDNA by using biotinylated oligo$(dT)_{24}$ and streptavidin beads, akin to the bead capture and primer

extension of droplet-based approaches[34,96]. This strategy may prove useful in other non-murine settings, such as suspension cells, where genomic contamination will be more extensive than with LCM[42].

ERCC spike-ins provide a standard to compare 10cRNA-seq against single-cell methods for transcriptomic profiling. Using the metrics of Svennson *et al.*[48], we estimate a 50% detection sensitivity of 45 copies per reaction (90% nonparametric CI: [15–485]) and a Pearson product-moment correlation coefficient of $R = 0.86$ (90% nonparametric CI: [0.71–0.91] from $n = 72$ samples). The R accuracy is somewhat lower than prevailing techniques, but that may be overly pessimistic because 10cRNA-seq uses such a dilute mix of spike-ins (4 million-fold dilution of the ERCC stock). Detection sensitivity is comparable to that reported for the most popular plate-based scRNA-seq methods, including SMART-seq2[97] and CEL-seq[98]. The strength of 10cRNA-seq lies in the use of 10-cell pooling to improve the per-cell technical sensitivity beyond the best microfluidic- and droplet-based approaches for scRNA-seq[48]. LCM minimizes disruptive tissue handling and provides histologic cues for microdissecting pools of cells within the same lineage. Adopting a 10-cell approach may prove similarly beneficial for other microdissection-based approaches, such as GEO-seq[26] and the recent pairing of SMART-seq2 with LCM[28].

When 10cRNA-seq was compared to scRNA-seq, we often observed significant improvements in exonic alignment. Methods for scRNA-seq typically yield exonic alignment rates below 50%[81], with the remainder of aligned reads splitting equally between intronic and intergenic sequences[97]. 10cRNA-seq achieves exonic alignments of 70% or higher despite using oligo(dT)-primed reverse transcription with the same potential to prime internal A homopolymer sequence as with scRNA-seq[99,100]. Interestingly, in one instance of similarly high exonic alignment (GSE66357, Fig. 8B), the RNA-printing approach to scRNA-seq incorporated a DNase treatment absent from all other methods[81]. This study also yielded a significantly reduced gene-detection sensitivity compared to 10cRNA-seq. Commingling genomic DNA may dilute exonic alignment percentages and inflate the number of genes detected due to chance sequencing of genomic DNA from exonic loci. Multiple scRNA-seq approaches incorporate unique molecular identifiers appended to oligo(dT)[48,81,101]. The identifiers avoid redundantly counting the same product of reverse transcription, and they also retrospectively exclude sequenced reads that do not come from cDNA. The biotin cleanup approach we devised for mouse cells (Fig. 5) achieves cDNA selection prospectively in situations where genomic contamination may be problematic.

Our work illustrates that 10-cell profiling can extend beyond microarrays[45] and quantitative PCR[39,40] to compete favorably with scRNA-seq. Although ill-suited for lineage mapping of highly mixed cell populations[43], 10cRNA-seq exploits the precision of LCM to target specific cell types *in situ* and define their regulatory heterogeneities. LCM is also advantageous for sequencing cells that are delicate or difficult to dissociate rapidly[28]. We anticipate immediate applications of 10cRNA-seq to cancer biology, where the initiation, progression, and diversification of tumors could be tackled in modern animal models as well as in patients.

Materials and Methods

Cell and tissue sources. The MCF10A-5E breast epithelial cell samples were described previously[41]. KP1 small-cell lung cancer cells[71] were grown as spheroids in RPMI Medium 1640 with 10% FBS, 1% penicillin-streptomycin, and 1% glutamine. KP1 spheroids were pelleted and mixed in Neg-50 (Richard-Allan Scientific) before cryoembedding. Animal housing and experimental procedures were carried out in compliance with regulations and protocols approved by the IACUC at the University of Virginia. *Cspg4-CreER;Trp53$^{F/F}$;Nf1$^{F/F}$;Rosa26-LSL-tdT* mice[58] were housed in accordance with IACUC Protocol #3955 at the University of Virginia. As per the approved protocol, animals were administered 200 mg/kg tamoxifen by oral gavage for five days, and brains were harvested at 12 days or 183 days after the last administration. A labelled glioma arising the olfactory bulb at 165 days after the last tamoxifen administration was also used. Human sample acquisition and experimental procedures were carried out in compliance with regulations and protocols approved by the IRB-HSR at the University of Virginia. In accordance with IRB Protocol #19272, breast cancer samples were collected as ultrasound-guided core needle biopsies during diagnostic visits from participants who provided informed consent. Each core biopsy was divided into multiple pieces before cryoembedding. Unless otherwise indicated, all samples were freshly cryoembedded in a dry ice-isopentane bath and stored at −80 °C wrapped in aluminium foil.

Cryosectioning. Samples were equilibrated to −24 °C in a cryostat before sectioning. 8 µm sections were cut and wicked onto Superfrost Plus slides. To preserve fluorescence localization of tdT and EGFP, slides were precooled on the cutting platform for 15–30 sec before wicking, and the section was carefully placed atop the cooled slide with forceps equilibrated at −24 °C. Then, the slide was gently warmed from underneath by tapping with a finger until the section was minimally wicked onto the slide. All wicked slides were stored in the cryostat before transfer to −80 °C storage on dry ice. Frost build-up was minimized by storing cryosections in five-slide mailers.

Staining, dehydration, and laser-capture microdissection. For cryosections lacking fluorophores, slides were stained and dehydrated as described previously[41,42]. Briefly, slides were fixed immediately in 75% ethanol for 30–60 sec, rehydrated quickly with water, stained with nuclear fast red (Vector Labs) containing 1 U/ml RNAsin-Plus (Promega) for 15 sec, and rinsed two more times with water before dehydrating with 70% ethanol for 30 sec, 95% ethanol for 30 sec, and 100% ethanol for 1 min and clearing with xylene for 2 min. tdT- and EGFP-labelled cryosections were not stained and instead began with the 70% ethanol dehydration step that also provided solvent fixation. After air drying, slides were microdissected immediately on an Arcturus XT LCM instrument (Applied Biosystems) using Capsure HS caps (Arcturus). The smallest spot size was used, and typical instrument settings of ~50 mW power and ~2 msec duration yielded ~25 µm spot diameters capturing 1–3 cells per laser shot.

RNA extraction and first-strand synthesis. RNA extraction and first-strand synthesis were similar to earlier protocols[41,42] with some minor modifications. Capsure HS caps were eluted for 1 hr at 42 °C with 4 µl

of digestion buffer containing 1.25x First-strand buffer (Invitrogen), 100 µM dNTPs (Roche), 0.08 OD/ml oligo$(dT)_{24}$ with or without 5′-biotin modification (IDT), and 250 µg/ml proteinase K (Sigma). Samples containing ERCC spike-ins included a four-million-fold dilution of ERCC spike-in mixture 1 (Ambion). Eluted samples were centrifuged into 0.5 ml PCR tubes at 560 rcf for 2 min, the digestion buffer was quenched with 1 µl of digestion stop buffer containing 2 U/µl SuperAse-in (Invitrogen) and 5 mM freshly prepared PMSF (Sigma). 4.5 µl of the quenched extract was transferred to a 0.2 ml PCR tube, and reverse transcription was performed with 0.5 µl of SuperScript III (Invitrogen) for 30 min at 50 °C followed by heat inactivation at 70 °C for 15 min. Samples were placed on ice and centrifuged for 2 min at 18,000 rcf on a benchtop microcentrifuge.

Streptavidin bead cleanup of biotinylated first-strand products. For 5′-biotin-containing samples, streptavidin magnetic beads (Pierce) were prepared in a 0.2 ml PCR tube on a 96 S Super Magnet Plate (Alpaqua). Beads (6 µl per sample) were magnetized, aspirated, and resuspended in binding buffer (5 µl per sample) containing 1x First-strand buffer (Invitrogen), 4 M NaCl, and 0.02% (vol/vol) Tween-20. 5 µl of resuspended beads were added after first-strand synthesis, and samples were incubated for 60 min at room temperature with mixing every 15 min. Beads were pelleted on the magnet plate, resuspended in 100 µl high-salt wash buffer (50 mM Tris [pH 8.3], 2 M NaCl, 75 mM KCl, 3 mM $MgCl_2$, 0.01% Tween-20). Beads were pelleted again on the magnet plate, and the pellet was washed once with 100 µl high-salt wash buffer. Next, beads were resuspended in 100 µl low-salt wash buffer (50 mM Tris [pH 8.3], 75 mM KCl, 3 mM $MgCl_2$) and transferred to a fresh 0.2 ml PCR tube. Beads were pelleted again on the magnet plate, and the pellet was washed once with 100 µl low-salt wash buffer. After the last wash, the beads were resuspended in 5 µl 1x First-strand buffer for RNAse H treatment and poly(A) tailing.

RNAse H treatment and poly(A) tailing. RNAse H digestion and poly(A) tailing were performed exactly as described previously[41,42]. Briefly, template mRNA strands were hydrolyzed for 15 min at 37 °C with 1 µl of RNAse H solution containing 2.5 U/ml RNAse H (USB Corporation) and 12.5 mM $MgCl_2$. After RNAse H treatment, cDNA templates were poly(A)-tailed with 3.5 µl of 2.6x tailing solution containing 80 U terminal transferase (Roche), 2.6x terminal transferase buffer (Invitrogen) and 1.9 mM dATP. The tailing reaction was incubated for 15 min at 37 °C and then heat-inactivated at 65 °C for 10 min. Samples were placed on ice and spun for 2 min at 18,000 rcf on a benchtop centrifuge.

Poly(A) PCR. Poly(A) PCR was carried out with several modifications to the earlier procedure[41,42]. To each tailed sample, 90 µl of poly(A) PCR buffer was added to a final concentration of 1x ThermoPol buffer (New England Biolabs), 2.5 mM $MgSO_4$, 1 mM dNTPs (Roche), 100 µg/ml BSA (Roche), 3.75 U Taq polymerase (NEB) and 1.5 U Phusion (NEB) and 2.5 µg AL1 primer (ATTGGATCCAGGCCGCTCTG GACAAAATATGAATTCTTTTTTTTTTTTTTTTTTTTTTT). Each reaction was split into three thin-walled 0.2 ml PCR tubes and amplified according to the following thermal cycling scheme: four cycles of 1 min at 94 °C (denaturation), 2 min at 32 °C (annealing) and 2 min plus 10 sec per cycle at 72 °C (extension); 21 cycles of 1 min at 94 °C (denaturation), 2 min at 42 °C (annealing) and 2 min 40 sec plus 10 sec per cycle at 72 °C (extension). The tubes were cooled, placed on ice, and the reactions from three tubes for each sample were pooled and amplified according to the following thermal cycling scheme: five cycles of 1 min at 94 °C (denaturation), 2 min at 42 °C (annealing) and 6 min at 72 °C (extension). Amplified samples were stored at −20 °C until further use.

Poly(A) PCR re-amplification. For sequencing, poly(A) PCR cDNA samples were reamplified as before[41,42] in a 100 µl PCR reaction containing 1x High-Fidelity buffer (Roche), 3.5 mM $MgCl_2$, 200 µM dNTPs (Roche), 100 µg/ml BSA (Roche), 5 µg AL1 primer, 1 µl Expand High Fidelity polymerase (Sigma), and 1 µl of poly(A) PCR sample. Each reaction was amplified according to the following thermal cycling scheme: 1 min at 94 °C (denaturation), 2 min at 42 °C (annealing) and 3 min at 72 °C (extension). The appropriate number of PCR cycles was determined by a pilot reamplification containing 20 µl of the PCR reaction above plus 0.25x SYBR Green monitored on a CFX96 real-time PCR instrument (Bio-Rad). The number of amplification cycles for each sample was selected to ensure that the reamplification remained in the exponential phase and there was sufficient cDNA for SPRI bead purification (typically 5–12 cycles).

qPCR. For detection of specific targets in poly(A) PCR samples, qPCR was performed on a CFX96 real-time PCR instrument (Bio-Rad) as previously described[102]. 0.1 µl or 0.01 µl of each preamplification was used with the qPCR primers listed in Supplementary Table S2. For relative quantification between ERCC spike-ins, qPCR amplicons were purified by gel electrophoresis, extracted, ethanol precipitated, and quantified by spectrophotometry. Purified amplicons were used to create a six-log standard curve based on ERCC amplicon copy number. All spike-ins were normalized to ERCC130 copy numbers to obtain relative abundance.

SPRI bead purification. Re-amplified samples were purified twice with 70% (vol/vol) Ampure Agencourt XP SPRI beads. SPRI beads were equilibrated to room temperature for 30 min, and 70 µl beads were added to the 100 µl reamplification product. After a 15-min incubation at room temperature, samples were magnetized for 5 min. The supernatant was removed with a gel-loading pipette tip, leaving ~5 µl volume in the well. Beads were gently washed twice on the magnet with 200 µl freshly prepared 80% (vol/vol) ethanol and aspirated with a gel-loading pipette tip. Residual ethanol was removed after the second wash, and beads were air-dried at room temperature for 10 min before resuspension in 10 µl elution buffer (10 mM Tris-HCl [pH 8.5]). Samples were magnetized at room temperature for 1 min, and the eluted supernatant was transferred to a new 0.2 ml PCR tube. The 10 µl elution was purified a second time with 7 µl beads and the same incubation, ethanol wash, and elution conditions as the first purification.

RNA sequencing and analysis. Bead-purified cDNA libraries were quantified with the Qubit dsDNA BR Assay Kit (Thermo Fisher) using a seven-point standard curve and a CFX96 real-time PCR instrument (Bio-Rad) for detection. Samples were diluted to 0.2 ng/μl before tagmentation with the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's earlier recommendation to purify libraries with 180% (vol/vol) SPRI beads (Supplementary Fig. S9). For each run, samples were multiplexed at an equimolar ratio, and 1.3 pM of the multiplexed pool was sequenced on a NextSeq 500 instrument with NextSeq 500/550 Mid/High Output v1/v2 kits (Illumina) to obtain 75-bp paired-end reads at an average depth of 4.2 million reads per sample (Supplementary Fig. S8) or 7.5 million reads per sample (all others). Simulated read depths of 10cRNA-seq data from MCF10A-5E cells confirmed saturation of gene detection above ~5 million reads per sample (Supplementary Fig. S12). Adapters were trimmed using fastq-mcf in the EAutils package (version ea-utils.1.1.2-537) with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, no nucleotide skew causing cycle removal). Quality checks were performed with FastQC (version 0.11.7) and multiqc (version 1.5). Datasets were aligned to either the human (GRCh38.84) or the mouse (GRCm38.82) transcriptome along with reference sequences for ERCC spike-ins, using RSEM (version 1.3.0) with the following options: --bowtie2 --single-cell-prior --paired-end (Bowtie2 transcriptome aligner, single-cell prior to account for dropouts, paired-end reads). RSEM read counts were converted to transcripts per million (TPM) by dividing each value by the total read count for each sample and multiplying by 10^6. Mitochondrial genes and ERCC spike-ins were not counted towards the total read count during TPM normalization. The number of genes with TPM > 1 for each sample was calculated relative to the number of unique Ensembl IDs for the organism excluding ERCC spike-ins.

Analysis of public scRNA-seq datasets. FASTQ files were downloaded from GSE75330, GSE60361, GSE103354 (plate-based), GSE66357, GSE113197, and PRJNA396019. FASTQ files were not available for the droplet-based dataset of GSE103354; therefore, BAM files were downloaded from SRR7621182 and converted to FASTQ format. Adapters were trimmed using fastq-mcf with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, no nucleotide skew causing cycle removal). To compare with the other datasets, seqtk (version 1.3) was used to clip 15 bp unique molecular identifiers from the beginning of sequences in GSE60361 and GSE75330. All RNA-seq datasets were aligned to either the human (GRCh38.84) or the mouse (GRCm38.82) transcriptome as well as reference sequences for ERCC spike-ins, using RSEM with the following options: --bowtie2 --single-cell-prior (Bowtie2 transcriptome aligner, single-cell prior to account for dropouts). GSE103354 (plate-based), GSE113197, and PRJNA396019 also used --paired-end (paired-end reads). TPM conversion and gene detection quantification were calculated as above. For post-hoc pooling (Supplementary Fig. S11), individual scRNA-seq profiles were selected at random ($n = 48$ per dataset) and grouped with the nine scRNA-seq profiles in the dataset that were nearest by Jaccard distance. To model dropouts, TPM values for each scRNA-seq profile were scaled to expected copies per cell assuming 250,000 mRNA copies per cell[38] and transmitted to the 10-cell pool as a binomial random variable (N = expected copies per cell, p = RNA-to-cDNA conversion efficiency = 10% for Supplementary Fig. S11). Post-hoc pooling results were similar up to a conversion efficiency of ~30%.

Paired analysis of BeadChip microarrays and 10cRNA-seq. Microarray data (GSE120030)[41] were batch processed with the lumi R package[103] using a detection threshold of 0.05 and simple scaling normalization to obtain $\log_2$-normalized values that were converted to $\log_{10}$-normalized values. Gene names from the BeadChip files were merged to the extent possible with Ensembl IDs from the RSEM alignments by using HUGO Gene Nomenclature synonym tables to match current and retired gene names.

Monte Carlo simulations. Simulations of stochastic-profiling experiments were performed in MATLAB using StochProfGUI[42]. Each parameter set was run 50 times to measure median p values and nonparametric confidence intervals. False positives were called when the median p value was less than 0.05 for a unimodal population (expression fraction = 0). False negatives were called when the median p value was greater than 0.05 for a bimodal population (expression fraction ≠ 0).

Data Availability

All 10cRNA-seq data are available through the NCBI Gene Expression Omnibus (GSE120261). Step-by-step protocols for 10cRNA-seq, including critical steps and troubleshooting, are available here as a Supplementary Note and will be maintained on the Janes Laboratory website (http://bme.virginia.edu/janes/protocols/).

References

1. Fidler, I. J. Tumor heterogeneity and the biology of cancer invasion and metastasis. *Cancer Res.* **38**, 2651–2660 (1978).
2. Allinen, M. *et al*. Molecular characterization of the tumor microenvironment in breast cancer. *Cancer Cell* **6**, 17–32, https://doi.org/10.1016/j.ccr.2004.06.010 (2004).
3. Sorlie, T. *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98**, 10869–10874 (2001).
4. Alizadeh, A. A. *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511, https://doi.org/10.1038/35000501 (2000).
5. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54, https://doi.org/10.1038/ng1060 (2003).
6. Levsky, J. M. & Singer, R. H. Gene expression and the myth of the average cell. *Trends Cell Biol.* **13**, 4–6 (2003).
7. Place, A. E., Jin Huh, S. & Polyak, K. The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Res.* **13**, 227, https://doi.org/10.1186/bcr2912 (2011).
8. Carmona-Fontaine, C. *et al*. Emergence of spatial structure in the tumor microenvironment due to the Warburg effect. *Proc. Natl. Acad. Sci. USA* **110**, 19402–19407, https://doi.org/10.1073/pnas.1311939110 (2013).

9. Cai, D. L. & Jin, L. P. Immune Cell Population in Ovarian Tumor Microenvironment. *J. Cancer* **8**, 2915–2923, https://doi.org/10.7150/jca.20314 (2017).
10. Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322, https://doi.org/10.1016/j.ccr.2012.02.022 (2012).
11. Kalluri, R. The biology and function of fibroblasts in cancer. *Nat. Rev. Cancer* **16**, 582–598, https://doi.org/10.1038/nrc.2016.73 (2016).
12. Yuan, Y. Spatial Heterogeneity in the Tumor Microenvironment. *Cold Spring Harb. Perspect. Med.* **6**, https://doi.org/10.1101/cshperspect.a026583 (2016).
13. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313, https://doi.org/10.1038/nature10762 (2012).
14. Navin, N. *et al*. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94, https://doi.org/10.1038/nature09807 (2011).
15. Gupta, P. B. *et al*. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644, https://doi.org/10.1016/j.cell.2011.07.026 (2011).
16. Shaffer, S. M. *et al*. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435, https://doi.org/10.1038/nature22794 (2017).
17. Sharma, S. V. *et al*. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80, https://doi.org/10.1016/j.cell.2010.02.027 (2010).
18. Wang, C. C., Bajikar, S. S., Jamal, L., Atkins, K. A. & Janes, K. A. A time- and matrix-dependent TGFBR3-JUND-KRT5 regulatory circuit in single breast epithelial cells and basal-like premalignancies. *Nat. Cell Biol.* **16**, 345–356, https://doi.org/10.1038/ncb2930 (2014).
19. Gerlinger, M. *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892, https://doi.org/10.1056/NEJMoa1113205 (2012).
20. Maley, C. C. *et al*. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* **38**, 468–473, https://doi.org/10.1038/ng1768 (2006).
21. Proia, T. A. *et al*. Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* **8**, 149–163, https://doi.org/10.1016/j.stem.2010.12.007 (2011).
22. Tirosh, I. *et al*. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196, https://doi.org/10.1126/science.aad0501 (2016).
23. van den Brink, S. C. *et al*. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936, https://doi.org/10.1038/nmeth.4437 (2017).
24. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* **144**, 3625–3632, https://doi.org/10.1242/dev.151142 (2017).
25. Peng, G. *et al*. Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. *Dev. Cell* **36**, 681–697, https://doi.org/10.1016/j.devcel.2016.02.020 (2016).
26. Chen, J. *et al*. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* **12**, 566–580, https://doi.org/10.1038/nprot.2017.003 (2017).
27. Pereira, M. *et al*. Direct Reprogramming of Resident NG2 Glia into Neurons with Properties of Fast-Spiking Parvalbumin-Containing Interneurons. *Stem Cell Reports* **9**, 742–751, https://doi.org/10.1016/j.stemcr.2017.07.023 (2017).
28. Nichterwitz, S. *et al*. Laser capture microscopy coupled with Smart-seq. 2 for precise spatial transcriptomic profiling. *Nat Commun* **7**, 12139, https://doi.org/10.1038/ncomms12139 (2016).
29. Brennecke, P. *et al*. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095, https://doi.org/10.1038/nmeth.2645 (2013).
30. Bhargava, V., Head, S. R., Ordoukhanian, P., Mercola, M. & Subramaniam, S. Technical variations in low-input RNA-seq methodologies. *Sci. Rep.* **4**, 3678, https://doi.org/10.1038/srep03678 (2014).
31. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145, https://doi.org/10.1038/nrg3833 (2015).
32. Pierson, E. & Yau, C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* **16**, 241, https://doi.org/10.1186/s13059-015-0805-z (2015).
33. Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640, https://doi.org/10.1038/nmeth.2930 (2014).
34. Klein, A. M. *et al*. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201, https://doi.org/10.1016/j.cell.2015.04.044 (2015).
35. Picelli, S. *et al*. Smart-seq. 2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098, https://doi.org/10.1038/nmeth.2639 (2013).
36. Eberwine, J., Sul, J. Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11**, 25–27 (2014).
37. Islam, S. *et al*. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167, https://doi.org/10.1101/gr.110882.110 (2011).
38. Marinov, G. K. *et al*. From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510, https://doi.org/10.1101/gr.161034.113 (2014).
39. Narayanan, M., Martins, A. J. & Tsang, J. S. Robust Inference of Cell-to-Cell Expression Variations from Single- and K-Cell Profiling. *PLoS Comput. Biol.* **12**, e1005016, https://doi.org/10.1371/journal.pcbi.1005016 (2016).
40. Martins, A. J. *et al*. Environment Tunes Propagation of Cell-to-Cell Variation in the Human Macrophage Gene Network. *Cell Syst* **4**, 379–392 e312, https://doi.org/10.1016/j.cels.2017.03.002 (2017).
41. 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, https://doi.org/10.1038/nmeth.1442 (2010).
42. Wang, L. & Janes, K. A. Stochastic profiling of transcriptional regulatory heterogeneities in tissues, tumors and cultured cells. *Nat. Protoc.* **8**, 282–301, https://doi.org/10.1038/nprot.2012.158 (2013).
43. Janes, K. A. Single-cell states versus single-cell atlases - two classes of heterogeneity that differ in meaning and method. *Curr. Opin. Biotechnol.* **39**, 120–125, https://doi.org/10.1016/j.copbio.2016.03.015 (2016).
44. Shalek, A. K. *et al*. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240, https://doi.org/10.1038/nature12172 (2013).
45. Bajikar, S. S., Fuchs, C., Roller, A., Theis, F. J. & Janes, K. A. Parameterizing cell-to-cell regulatory heterogeneities via stochastic transcriptional profiles. *Proc. Natl. Acad. Sci. USA* **111**, E626–635, https://doi.org/10.1073/pnas.1311647111 (2014).
46. 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, https://doi.org/10.1073/pnas.1103423108 (2011).
47. Bajikar, S. S. *et al*. Tumor-Suppressor Inactivation of GDF11 Occurs by Precursor Sequestration in Triple-Negative Breast Cancer. *Dev. Cell* **43**, 418–435 e413, https://doi.org/10.1016/j.devcel.2017.10.027 (2017).
48. Svensson, V. *et al*. Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387, https://doi.org/10.1038/nmeth.4220 (2017).

49. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040, https://doi.org/10.1101/gr.177881.114 (2014).
50. Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**, e78644, https://doi.org/10.1371/journal.pone.0078644 (2014).
51. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620, https://doi.org/10.1016/j.molcel.2015.04.005 (2015).
52. Brady, G. & Iscove, N. N. Construction of cDNA libraries from single cells. *Methods Enzymol.* **225**, 611–623 (1993).
53. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
54. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603, https://doi.org/10.1038/nprot.2006.85 (2006).
55. Kreklywich, C. N. *et al.* Fluorescence-based laser capture microscopy technology facilitates identification of critical *in vivo* cytomegalovirus transcriptional programs. *Methods Mol. Biol.* **1119**, 217–237, https://doi.org/10.1007/978-1-62703-788-4_13 (2014).
56. Murakami, H., Liotta, L. & Star, R. A. IF-LCM: laser capture microdissection of immunofluorescently defined cells for mRNA analysis rapid communication. *Kidney Int.* **58**, 1346–1353, https://doi.org/10.1046/j.1523-1755.2000.00295.x (2000).
57. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572, https://doi.org/10.1038/nbt1037 (2004).
58. Galvao, R. P. *et al.* Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. *Proc. Natl. Acad. Sci. USA* **111**, E4214–4223, https://doi.org/10.1073/pnas.1414389111 (2014).
59. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492, https://doi.org/10.1016/j.cell.2005.02.012 (2005).
60. Liu, C. *et al.* Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* **146**, 209–221, https://doi.org/10.1016/j.cell.2011.06.014 (2011).
61. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63, https://doi.org/10.1038/nrg2484 (2009).
62. Cronin, M. *et al.* Universal RNA reference materials for gene expression. *Clin. Chem.* **50**, 1464–1471, https://doi.org/10.1373/clinchem.2004.035675 (2004).
63. Baker, S. C. *et al.* The External RNA Controls Consortium: a progress report. *Nat. Methods* **2**, 731–734, https://doi.org/10.1038/nmeth1005-731 (2005).
64. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551, https://doi.org/10.1101/gr.121095.111 (2011).
65. Lusk, R. W. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One* **9**, e110808, https://doi.org/10.1371/journal.pone.0110808 (2014).
66. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29, https://doi.org/10.1186/s13059-016-0888-1 (2016).
67. Consortium, S. M.-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914, https://doi.org/10.1038/nbt.2957 (2014).
68. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382, https://doi.org/10.1038/nmeth.1315 (2009).
69. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631–643 e634, https://doi.org/10.1016/j.molcel.2017.01.023 (2017).
70. Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **34**, e42 (2006).
71. Schaffer, B. E. *et al.* Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res.* **70**, 3877–3883, https://doi.org/10.1158/0008-5472.CAN-09-4228 (2010).
72. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119, https://doi.org/10.1186/gb-2010-11-12-r119 (2010).
73. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
74. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742, https://doi.org/10.1038/nmeth.2967 (2014).
75. Chen, L. *et al.* Correlation between RNA-Seq and microarrays results using TCGA data. *Gene* **628**, 200–204, https://doi.org/10.1016/j.gene.2017.07.056 (2017).
76. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142, https://doi.org/10.1126/science.aaa1934 (2015).
77. Marques, S. *et al.* Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329, https://doi.org/10.1126/science.aaf6463 (2016).
78. Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324, https://doi.org/10.1038/s41586-018-0393-7 (2018).
79. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879–893 e813, https://doi.org/10.1016/j.cell.2018.03.041 (2018).
80. Nguyen, Q. H. *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* **9**, 2028, https://doi.org/10.1038/s41467-018-04334-1 (2018).
81. Bose, S. *et al.* Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* **16**, 120, https://doi.org/10.1186/s13059-015-0684-3 (2015).
82. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, https://doi.org/10.1186/1471-2105-12-323 (2011).
83. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, https://doi.org/10.7554/eLife.27041 (2017).
84. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667, https://doi.org/10.1126/science.aam8940 (2017).
85. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. *Nature* **542**, 352–356, https://doi.org/10.1038/nature21065 (2017).
86. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, https://doi.org/10.1038/nature14966 (2015).
87. Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M. & Reddien, P. W. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* **360**, https://doi.org/10.1126/science.aaq1736 (2018).
88. Plass, M. *et al.* Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**, https://doi.org/10.1126/science.aaq1723 (2018).
89. Hunter, F., Xie, J., Trimble, C., Bur, M. & Li, K. C. Rhodamine-RCA *in vivo* labeling guided laser capture microdissection of cancer functional angiogenic vessels in a murine squamous cell carcinoma mouse model. *Mol. Cancer* **5**, 5, https://doi.org/10.1186/1476-4598-5-5 (2006).
90. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148**, 33–45, https://doi.org/10.1016/j.cell.2012.01.002 (2012).

91. Fend, F. *et al*. Immuno-LCM: laser capture microdissection of immunostained frozen sections for mRNA analysis. *Am. J. Pathol.* **154**, 61–66, https://doi.org/10.1016/S0002-9440(10)65251-0 (1999).
92. Steu, S. *et al*. A procedure for tissue freezing and processing applicable to both intra-operative frozen section diagnosis and tissue banking in surgical pathology. *Virchows Arch.* **452**, 305–312, https://doi.org/10.1007/s00428-008-0584-y (2008).
93. Mouse Genome Sequencing, C. *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562, https://doi.org/10.1038/nature01262 (2002).
94. Shedlock, A. M. & Okada, N. SINE insertions: powerful tools for molecular systematics. *Bioessays* **22**, 148–160, doi:10.1002/(SICI)1521-1878(200002)22:2<148::AIDBIES6>3.0.CO;2-Z (2000).
95. Tsirigos, A. & Rigoutsos, I. Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput. Biol.* **5**, e1000610, https://doi.org/10.1371/journal.pcbi.1000610 (2009).
96. Macosko, E. Z. *et al*. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214, https://doi.org/10.1016/j.cell.2015.05.002 (2015).
97. Picelli, S. *et al*. Full-length RNA-seq from single cells using Smart-seq. 2. *Nat. Protoc.* **9**, 171–181, https://doi.org/10.1038/nprot.2014.006 (2014).
98. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* **2**, 666–673, https://doi.org/10.1016/j.celrep.2012.08.003 (2012).
99. Nam, D. K. *et al*. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl. Acad. Sci. USA* **99**, 6152–6156, https://doi.org/10.1073/pnas.092140899 (2002).
100. La Manno, G. *et al*. RNA velocity of single cells. *Nature* **560**, 494–498, https://doi.org/10.1038/s41586-018-0414-6 (2018).
101. Kivioja, T. *et al*. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74, https://doi.org/10.1038/nmeth.1778 (2012).
102. 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).
103. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548, https://doi.org/10.1093/bioinformatics/btn224 (2008).

Acknowledgements

We thank Kathy Repich and Jennifer Harvey for help with clinical sample acquisition, Emily Farber for technical assistance with RNA sequencing, and Stephen Turner from the UVA Bioinformatics Core for guidance on alignment approaches. This work was supported by the National Institutes of Health #R01-CA194470 and #U01-CA215794 (K.A.J.), the David & Lucile Packard Foundation #2009-34710 (K.A.J.), a Medical Scientist Training Program Fellowship #T32-GM007267, a UVA Cancer Center support grant #P30-CA044579, a Wagner Fellowship (S.S.), and a Harrison Undergraduate Research Award (D.L.S.).

Author Contributions

S.S., L.W. and K.A.J. conceived the experiments with assistance from J.K., S.O., K.P. and H.Z. involving mouse models and RNA-seq. S.S., L.W., D.L.S., M.D.S. and K.A.J. conducted the experiments. A.F.K. helped with RNA-seq data analysis. S.S. and K.A.J. drafted the manuscript and received feedback from all authors.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-41235-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.