Development and applications of single-cell transcriptome analysis

Fuchou Tang[1,2], Kaiqin Lao[3] & M Azim Surani[1]

Dissecting the relationship between genotype and phenotype is one of the central goals in developmental biology and medicine. Transcriptome analysis is a powerful strategy to connect genotype to phenotype of a cell. Here we review the history, progress, potential applications and future developments of single-cell transcriptome analysis. In combination with live cell imaging and lineage tracing, it will be possible to decipher the full gene expression network underlying physiological functions of individual cells in embryos and adults, and to study diseases.

© 2011 Nature America, Inc. All rights reserved.

Development is driven and controlled by temporal and spatial changes in gene transcription, followed by translation of the resulting mRNAs into proteins. The transcriptome is broadly defined as the entire RNA component of an individual cell, or it is narrowly and practically defined as the polyadenylated products of RNA polymerase II activity[1,2]. Recent advances in high-throughput sequencing technology make it possible to obtain information on single-cell transcriptomes at high resolution by RNA sequencing (RNA-seq) analysis, which can be instructive in regards to how individual cells respond to signals and other environmental cues at critical stages of cell-fate determination or when they acquire aberrant phenotypes. Essentially all cells in an individual organism have a virtually identical genotype, but the individual transcriptomes reflect expression of a subset of genes, which is determined by their epigenetic state. Diverse cell types have unique transcriptomes, which can be used to assess the gene regulation network underlying their physiological functions, behavior and phenotype during development in multicellular organisms[2].

Because cells have unique transcriptomes, such analysis should be carried out at single-cell resolution. The analysis should also encapsulate the exact sequence, quantity, localization, activity (for example, being actively translated or degraded) and modifications (such as base methylation) of all types of full-length RNAs at single-base resolution[1–3]. Indeed, mRNA, rRNA, tRNA and small nuclear RNA can have more than 100 structurally distinct post-transcriptional modifications at thousands of different sites[3]. Because of technical limitations, such as the sensitivity of detection, the majority of studies on transcriptomes have been carried out with hundreds of thousands or even millions of cells[4–7]. In some instances, however, it is not possible to collect large numbers of cells, such as from very early embryos, which makes their transcriptome analysis very difficult, if not impossible[8].

Recent studies have also shown that gene expression is invariably heterogeneous even in evidently similar cell types[9,10]. Such stochastic variations in the transcriptomes have important implications for cell-fate decisions[11]. For example, Sunney Xie and colleagues recently showed that a stochastic single-molecule event can trigger phenotype switching of a bacterial cell[12,13]. Differences in transcriptomes may also provide critical information on the composition of cell types in diseased tissues, including tumors that could contain a small number of cancer stem cells[9,14]. The heterogeneity of gene expression among similar cell types can be due to differences in the epigenetic status of the genome, circadian clock, cell cycle, microenvironment or niche as well as intrinsic transcriptional 'noise'[15–22]. In fact, gene expression is stochastic in essentially all model organisms from bacteria to humans[19–22]. This is partly because for the majority of genes in an individual cell, only one (for prokaryotes) or two

[1]Wellcome Trust/Cancer Research UK Gurdon Institute of Cancer and Developmental Biology, University of Cambridge, Cambridge, UK. [2]Biodynamics Optical Imaging Center, School of Life Sciences, Peking University, Beijing, China. [3]Genetic Systems, Applied Biosystems, part of Life Technologies, Foster City, California, USA. Correspondence should be addressed to K.L. (kaiqinlao@gmail.com) or M.A.S. (a.surani@gurdon.cam.ac.uk).

PUBLISHED ONLINE 30 MARCH 2011; DOI:10.1038/NMETH.1557

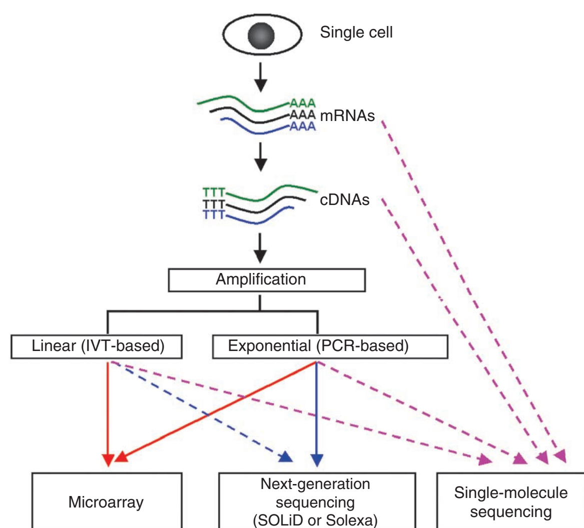

Figure 1 | Strategies for single-cell transcriptome analysis. Solid lines represent strategies that have been demonstrated experimentally; dotted lines represent proposed strategies that will probably be realized in the near future.

copies (for most of the eukaryotic organisms) of genomic DNA templates are available for transcription, and the molecular events that trigger their expression will intrinsically have stochastic characteristics[22]. To understand the basis and importance of heterogeneity and stochastic aspect of gene expression, it is essential to examine transcriptomes of individual cells.

History of single-cell transcriptome analysis

Transcriptome analysis at single-cell resolution was pioneered two decades ago by Norman Iscove, using exponential amplification of single cell cDNAs by PCR[23], and by James Eberwine using linear amplification of cDNAs by T7 RNA polymerase–based *in vitro* transcription (IVT)[24,25] (**Fig. 1**). These approaches have accelerated insights on the molecular mechanisms of development and function of the mammalian neural system, especially because these cells are probably the most heterogeneous group of cells. In this case, transcriptomes obtained at the cellular or even subcellular resolution in a long axon can be informative[26–29].

Later, the use of commercially available high-density DNA microarray chips led to the development of single-cell microarrays[30–39] (**Table 1**). Although this method is powerful and can be used to obtain whole-genome gene expression patterns[32,40], the cDNA fragments being amplified are in general short (several hundred base pairs) and cannot be used to detect transcripts generated through alternative splicing. Most importantly, the method can only be used to detect known genes.

Generating single-cell transcriptomes

To generate single-cell transcriptomes, individual intact cells are isolated and transferred into a test tube containing lysis buffer. The whole-cell lysate is then reverse-transcribed using oligo(dT) primers to convert mRNAs with poly(A) tails into first-strand cDNAs. The residual mRNA templates are degraded and a poly(A) tail is added to the 3′ ends of the first-strand cDNAs. These cDNAs are uniformly amplified with universal oligo(dT) primers. A key requirement for the procedure is that the buffers for earlier and later reaction steps are compatible[30]. In addition, enzymes used in the earlier steps need to be inactivated by heat treatment. This approach avoids additional isolation, precipitation and purification steps. The amplified single-cell cDNAs can be quality-checked and tested by quantitative (q)PCR, and only samples of interest and of high quality need to be analyzed further.

The method for the isolation of individual cells can vary. Picking single cells manually using a mouth pipette is the most straightforward option[30,41], although this can be time-consuming and technically challenging. Laser-assisted microdissection or fluorescence-activated cell sorting can also be used to isolate a specific subpopulation of cells based on cell-surface markers or fluorescent reporters[42–45]. Cells of higher plants with cell walls are difficult to dissociate using enzymes, but isolation of nuclei after homogenization of tissues might be effective[46].

It should be possible to use microfluidics systems in the future to isolate and track thousands of single cells in parallel in a nanoliter of solution[34,47]. This will greatly enhance the accuracy and efficiency of analyzing single cells from a variety of sources, including adult stem cells or cancer cells.

Analysis of all mRNAs in individual cells first requires their release from cells, which can be accomplished with detergents but these should not interfere with the subsequent reverse-transcription process[30,48]. Several types of detergents can be considered, such as guanidine thiocyanate and Nonidet P-40 (refs. 30,32,33,40,48). Both the type and amount of detergents need to be adjusted to obtain the best results for different cell types, which depends on the propensity of cells to undergo lysis. When working with a new cell type, several detergents should be tested in parallel at different concentrations to obtain the best conditions for the specific cell type. In most instances, the whole-cell lysate can be used directly for reverse transcription without the need to remove detergents[23,30]. It is also possible to isolate and purify mRNAs from single-cell lysate first, by using oligo(dT)-coated magnetic beads or oligo(dT) peptide nucleic acids that will capture the mRNA amid proteins, metabolites and cell debris[32,33,40]. This strategy has been reported to work in conjunction with single-cell cDNA microarray analysis[33]. This approach can be combined with the isolation of genomic DNA from the same cell for genotype analysis[32]. As the lysate buffer is washed away while specifically isolating mRNAs with poly(A) tails, this approach allows for the use of much stronger lysate conditions for a quick and efficient release of mRNAs[32].

Several types of reverse transcriptases are available for the preparation of cDNAs[49]. The SuperScript III (Invitrogen) is most widely used for this purpose, which can potentially generate full-length cDNAs of up to 10 kilobases (kb) (refs. 30,41,49). When using the whole-cell lysate, oligo(dT) primers are usually used for this step because the use of random primers would result in the amplification of the cDNAs of rRNA and tRNA, which are two orders of magnitude more abundant than mRNAs. To prevent this, in principle, mRNAs can be specifically isolated, purified and then combined with random primers for priming the reverse transcription of full-length mRNAs. dNTP concentrations need to be carefully adjusted to permit efficient reverse transcription and to avoid interference with a later poly(A)-tailing step. As mRNA 5′ ends for different genes and the corresponding 3′ ends of the first-strand cDNAs are different, use of terminal deoxynucleotidyl transferase to add poly(A) tails can enable unbiased amplification of all expressed genes' cDNAs.

npg © 2011 Nature America, Inc. All rights reserved.

Table 1 | Single-cell multiplex gene expression analysis strategies

Techniques	Target RNAs	Amplification	Throughput	References
Single-cell RNA-seq	mRNAs with poly(A) tail	Yes	High (whole transcriptome)	41,53
Single-cell cDNA microarray (PCR-based)	mRNAs with poly(A) tail	Yes	High (whole transcriptome with predesigned probes)	31–34,36
Single-cell cDNA microarray (IVT-based)	mRNAs with poly(A) tail	Yes	High (whole transcriptome with predesigned probes)	35
Single-cell cDNA microarray (combination of PCR and IVT)	mRNAs with poly(A) tail	Yes	High (whole transcriptome with predesigned probes)	15,30,64,84
Single-cell microRNA profiling (PCR-based)	Known microRNAs	Yes	High (up to several hundred microRNAs)	84–86
Single-cell multiplex qPCR	Known mRNAs	Yes	Low (up to several hundred genes)	15,44,45,53,87
Single-cell multiplex qPCR	Known mRNAs	No	Low (up to five genes)	49
Single-molecule RNA fluoresence *in situ* hybridization	Known mRNAs	No	Low (up to several genes)	88,89

After reverse transcription and tailing, the single-cell cDNAs can be amplified. One representative individual mammalian cell contains ~10 pg of total RNA and ~0.1 pg of mRNA, which usually needs to be amplified around ten million–fold to match the requirement for a standard microarray analysis. Either PCR or IVT can be used for amplification[30,50,51]. The advantage of the PCR strategy is the exponential amplification of cDNAs so that single-cell cDNAs can be amplified millions-fold in several hours[30,50]; its disadvantage is the accumulation of primer dimers and other nonspecific byproducts during amplification, especially during later cycles of PCR[23,30]. The merit of the IVT strategy is its stringent specificity while reducing accumulation of nonspecific byproducts[24]; its drawback is that cRNAs typically less than 1 kb are generated[50]. The IVT procedure is also more tedious and time-consuming, and every round of IVT can amplify the cDNAs only up to 1,000-fold[50,51]. In practice, single-cell cDNAs can be amplified sufficiently for microarray analysis in two rounds of PCR amplification[32–34], in three rounds of IVT amplification[35] or via a combination of PCR and IVT amplification[30].

We recently improved a widely used single-cell cDNA amplification protocol[30,52], which is highly quantitative but generates only about 0.85-kb fragments at the 3′ ends of the mRNAs. We combined it with next-generation sequencing to develop single-cell RNA-seq analysis[41]. We increased the efficiency of the protocol and can generate up to 3-kb fragments of cDNAs[2,41]. We also used amine-modified primers for the second round of PCR to remove the residual free primers and primer dimers from the sequencing library to improve throughput. Furthermore, a previous single-cell cDNA microarray analysis detected about 6,800 distinct gene transcripts in an individual embryonic stem cell[30], whereas we detected expression of about 10,800 genes, which means that our assay detected expression of nearly 60% more genes in an individual embryonic stem cell[53].

Because of the higher sensitivity of next-generation sequencing, amplification by additional IVT step is no longer needed. We showed that the amplification method works faithfully for the detection of the full transcriptome of individual early mouse blastomeres. Up to 60% of all the transcripts in the mouse genome are expressed at this stage in an individual cell[41]. We also found that up to 20% of genes with known splicing isoforms express multiple transcript variants in a single cell, highlighting the complexity of an individual cell's transcriptome[41]. And we found thousands of previously unknown exon-exon junctions in the transcriptome from an individual cell, indicating that our understanding of the mammalian cell transcriptome is far from complete[41]. Recently, we applied the technique to trace the process of the derivation of embryonic stem cells from the inner cell mass of blastocysts, which illustrated that the approach works faithfully for the analysis of relatively small-sized individual cells[53]. Thus, the technique is potentially applicable for the analysis of many cell types in developing embryos and in adult tissues, although the method has so far not been used on different cell types.

After amplification, single-cell cDNAs can be analyzed either on a microarray or by deep sequencing[30,41]. The latter provides more detailed and accurate information on transcriptomes with at least five-log dynamic range[1,41,54], but it is expensive and requires more computational power for data analysis[41]. In general, 20 to 40 million sequencing reads per cell are enough for most purposes, such as detection of new genes, splicing variants, polyadenylation sites and new exons of known genes[2]. To date, only the SOLiD system has been used for single cell RNA-seq, but the protocol is platform-independent. Paired-end reads will allow determination of splicing junctions from single-cell RNA-seq data more accurately. Microarray approaches are appropriate for obtaining general transcriptome information on the up- or downregulation of transcripts of known genes.

For bioinformatics analysis, both standard commercial software and free academic software are available[55]. Recently developed bioinformatics tools for RNA-seq data analysis, such as Cufflinks, Scripture, alternative expression analysis by sequencing (ALEXA-seq), mixture of isoforms (MISO) and Trans-ABySS (assembly by short sequences), could also be used for single cell RNA-seq[56–60]. Data normalization is crucial for comparison of transcriptomes of different samples in the same batch and of samples from different laboratories and platforms. For relative quantification, normalization can be either quantile, reads per kilobases of exon model per million mapped reads (RPKM) or reads per million mapped reads (RPM)[54]. The recently developed 'normalization by expected uniquely mappable area' method will probably improve the quantification of single-cell RNA-seq data[61]. We have shown that both quantile and RPM normalization work well for our single-cell RNA-seq data[41]. But RPKM quantification would be preferable when RNA-seq methods can detect full-length cDNAs.

The absolute quantification of cDNAs can be obtained by using spike-in RNA[30], which could be any predetermined quantity of poly(A)-tailed RNA that is not present in the transcriptome of the cell to be analyzed. For mammalian cell transcriptome analysis,

© 2011 Nature America, Inc. All rights reserved.

usually the RNAs encoding polylysine, poly-diaminopimelic acid, polyphenylalanine or polythreonine with poly(A) tail are used as spike-in RNAs[30,36,52]. When only hundreds of copies of spike-in RNAs are added into the single-cell sample, caution is necessary to ensure that there is no substantial degradation of the 'spike-in' RNAs. It is also important to note that the volume of different types of cells is highly variable and can differ by hundreds-fold[62]. Absolute quantification is only a measure of the absolute copy number of each gene's mRNA in a cell, without consideration of the cell volume. The concentration rather than the absolute copy number of mRNAs is important for determining its function in a cell. Theoretically, the absolute concentration of RNAs for every expressed gene in an individual cell can be determined from the amount of spike-in RNAs used, combined with determining the volume of individual cells[63].

Applications of single-cell transcriptome analysis

Single-cell transcriptome analysis can be used for determining gene regulatory networks at whole-genome scale and can be combined with overexpression, knockout or knockdown of a gene of interest to reveal how it regulates gene expression in target cells[41,64,65]. This is especially relevant for the analysis of stem cells and cells during early embryonic development because of the highly dynamic and heterogeneous nature of subpopulations of the cells[41,44]. Analysis of heterogeneity among cells is emerging as an important application of single-cell transcriptomes. Even highly similar cell types can have different gene expression patterns for a wide variety of reasons[9]. More importantly, gene expression is intrinsically stochastic owing to different microenvironments or because of the small number of molecules involved in transcription and translation[19–22]. It can be safely claimed that heterogeneity of gene expression is an intrinsic property of living cells, and there are no strictly identical cells in an organism[9]. Furthermore, stochastic characteristics of gene expression can profoundly affect the fate and phenotype of a cell[12,13,66]. Dissecting the heterogeneity of gene expression in a cell population will thus be an important application of single-cell analysis. Indeed, there is evidence for heterogeneity in subpopulations of embryonic stem cells based on the expression of Nanog, Rex1 or Stella[15,53,67,68]. Cell heterogeneity between cells in a tumor has been known for a long time[69]. Single-cell transcriptome analysis is a feasible strategy to identify the subpopulations in a tumor and to detect putative cancer stem cells. As only one individual cell needs to be isolated and lysed from a tissue, it is theoretically possible to analyze gene expression networks noninvasively to monitor the progress of human disease, or monitor a rare or precious biological sample, and to continuously trace gene expression dynamics of a tissue during physiological or pathological processes without disturbing or consuming the entire sample.

Another application is to determine the gene expression profiles of subcellular compartments. It is well known that there is active transport of mRNAs from cell body to the axons or dendrites in the neuron for local translation[26–29,36,37,70]. Single-cell transcriptome analysis can be used to detect mRNAs specifically localized in axons or dendrites, which is often of great importance for determining physiological functions of these neurons.

As next-generation sequencing techniques provide information at single-base resolution, it is also possible to analyze allele-specific gene expression in an individual cell, provided single-nucleotide polymorphisms are available to discriminate between the two alleles[71]. This will greatly improve our understanding of how the genetic and epigenetic elements influence allele-specific gene expression in an individual cell. Allelic imbalance can accurately describe small differences between cells, which can arise in a number of ways through changes in relative allele-specific expression by mutations such as point mutations or by RNA editing[3].

Current limitations

Although recent advances in RNA-seq analysis offer substantial opportunities to evaluate properties of cells, there are some drawbacks of current single-cell RNA-seq methods. First, the strandedness of mRNAs is lost in the library construction, which prevents discrimination between sense and antisense transcripts from the same locus[41]. Clearly, the strandedness should be preserved to permit accurate annotation of the sense and antisense RNA transcripts from the same gene locus. Combined with the existing strand-specific cDNA library preparation strategies, such as T7 RNA polymerase–based *in vitro* transcription and dUTP second-strand synthesis strategies, it will be possible to recover the strandedness information for single-cell transcriptomes in the near future[72].

Second, currently only the 3′ end and up to 3-kb fragments of mRNAs are obtained, which leaves nearly 36% of genes with mRNAs longer than 3 kb that cannot be fully examined. Methods that allow for the use of random primers for reverse transcription of purified mRNAs will make it possible to recover full-length cDNAs, including 5′ untranslated and transcription start regions for all expressed genes.

Third, because the method is based on reverse transcription with oligo(dT) primers, only mRNAs with poly(A) tail are detected, which will exclude some long noncoding RNAs and most of the small noncoding RNAs[41]. In the future, a more sophisticated protocol should be developed to deplete abundant rRNA and tRNA while preserving all the mRNA transcripts, primary transcripts before polyadenylation.

Fourth, the current single-cell RNA-seq method does not permit analysis of the transcriptome and genomic sequence of individual cells simultaneously. Improving the method based on the strategy developed by Christoph Klein and colleagues[32,33,40], it is possible to use next-generation sequencing to simultaneously obtain both full genome and transcriptome information from an individual cell. This will fulfill the central goal of biology and medicine, which is to connect the genotype and phenotype of individual cells under physiological or pathological conditions.

Perspectives

Single-cell transcriptome analysis will eventually permit connections between gene expression networks, cell lineage and phenotype of individual cells. Combined with live-cell imaging, this is potentially a powerful tool for tracing cell lineage during development or cell differentiation, especially in conjunction with florescent protein reporters[73]. Live-cell imaging together with single-cell RNA-seq will greatly improve our understanding of how cell differentiation is achieved and dynamically regulated by gene expression networks. This approach can also be used to analyze cellular reprogramming and transdifferentiation[74].

Currently, all available single-cell transcriptome analyses rely on cDNA amplification. Recently developed single-molecule

© 2011 Nature America, Inc. All rights reserved. npg

sequencing has the potential to sequence full-length mRNAs from a single cell directly without reverse transcription and amplification steps, which can be used to more accurately determine expression levels of different splicing isoforms[75–77]. Moreover, the full-length mRNA sequences will accurately determine allele-specific gene expression with defined phase information of each locus. However, the sequencing efficiency of current single-molecule sequencing techniques still requires a few hundred cells, and these methods only detect about 15–25% of expressed mRNAs, which needs to be improved to achieve single-cell RNA-seq, because only dozens of copies of mRNAs are produced from the majority of individual expressed genes in a cell[76]. Furthermore, the accuracy of the single-molecule sequencing technique is still relatively low and needs to be improved to acquire the exact sequences of every mRNA molecule at high accuracy comparable to current RNA-seq[76,77]. This will probably be achieved by improving the single-molecule sequencing technique to permit sequencing a single mRNA or cDNA molecule repeatedly without damaging it.

At the moment, only the static amount of mRNAs is measured by single-cell transcriptome analysis, which is the result of the balance between transcription and degradation of mRNAs. More detailed analysis will require accurate quantification of mRNAs being actively translated[78–83]. This single cell translating RNA-seq will directly reflect the translational activity and function of the genes at particular time points[78–83]. All the genetic and epigenetic information in the genome needs to be read and released through transcription into RNAs. We are now witnessing the opportunity to link gene expression network with the physiology, function and phenotype of every individual cell. It will be possible in the future to model the behavior and phenotype of an individual cell based on its environment and its transcriptome. Finally, we may also understand how a cell survives and functions properly in a complex and noisy environment, with a complex and noisy transcriptome.

© 2011 Nature America, Inc. All rights reserved.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Published online at http://www.nature.com/naturemethods/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
2. Tang, F. *et al.* RNA-seq analysis to capture the transcriptome landscape of a single cell. *Nat. Protoc.* **5**, 516–535 (2010).
3. He, C. Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **6**, 863–865 (2010).
4. Schulze, A. & Downward, J. Navigating gene expression using microarrays—a technology review. *Nat. Cell Biol.* **3**, E190–E195 (2001).
5. Cloonan, N. & Grimmond, S.M. Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol.* **9**, 234 (2008).
6. Wold, B. & Myers, R.M. Sequence census methods for functional genomics. *Nat. Methods* **5**, 19–21 (2008).
7. Schuster, S.C. Next-generation sequencing transforms today's biology. *Nat. Methods* **5**, 16–18 (2008).
8. Saitou, M., Barton, S.C. & Surani, M.A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300 (2002).
9. Huang, S. Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862 (2009).
10. Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545 (2010).
11. Eldar, A. & Elowitz, M.B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
12. Choi, P.J., Cai, L., Frieda, K. & Xie, X.S. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446 (2008).
13. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
14. Frank, N.Y., Schatton, T. & Frank, M.H. The therapeutic promise of the cancer stem cell concept. *J. Clin. Invest.* **120**, 41–50 (2010).
15. Hayashi, K., Lopes, S.M., Tang, F. & Surani, M.A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
16. Panda, S., Hogenesch, J.B. & Kay, S.A. Circadian rhythms from flies to human. *Nature* **417**, 329–335 (2002).
17. Maury, E., Ramsey, K.M. & Bass, J. Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. *Circ. Res.* **106**, 447–462 (2010).
18. Wittenberg, C. & Reed, S.I. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* **24**, 2746–2755 (2005).
19. Arias, A.M. & Hayward, P. Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.* **7**, 34–44 (2006).
20. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
21. Losick, R. & Desplan, C. Stochasticity and cell fate. *Science* **320**, 65–68 (2008).
22. Shahrezaei, V. & Swain, P.S. The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.* **19**, 369–374 (2008).
23. Brady, G., Barbara, M. & Iscove, N.N. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol. Cell. Biol.* **2**, 17–25 (1990).
    This first report on the preparation of single-cell cDNAs by exponential amplification based on PCR established the foundation for single-cell cDNA microarray and RNA-seq analysis.
24. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* **89**, 3010–3014 (1992).
    This study reported preparation of single cell cDNAs by linear amplification based on IVT.
25. Van Gelder, R.N. *et al.* Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667 (1990).
26. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
27. Tanabe, Y., William, C. & Jessell, T.M. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell* **95**, 67–80 (1998).
28. Yamagata, M., Weiner, J.A. & Sanes, J.R. Sidekicks: synaptic adhesion molecules that promote lamina-specific connectivity in the retina. *Cell* **110**, 649–660 (2002).
29. Shumyatsky, G.P. *et al.* Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory specifically related to learned fear. *Cell* **111**, 905–918 (2002).
30. Kurimoto, K. *et al.* An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.* **34**, e42 (2006).
    This was the first report of highly quantitative (R^2 = 0.89 for embryonic stem cells) single-cell cDNA microarray analysis.
31. Jensen, K.B. & Watt, F.M. Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. *Proc. Natl. Acad. Sci. USA* **103**, 11958–11963 (2006).
32. Klein, C.A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
    This report describes the isolation and amplification of cDNAs and genomic DNAs simultaneously from a single cell for transcriptome and comparative genomic hybridization analysis, which facilitated the analysis of evolution of tumor cells at single-cell resolution.
33. Hartmann, C.H. & Klein, C.A. Gene expression profiling of single cells on large-scale oligonucleotide arrays. *Nucleic Acids Res.* **34**, e143 (2006).
34. Bontoux, N. *et al.* Integrating whole transcriptome assays on a lab-on-a-chip for single cell gene profiling. *Lab Chip.* **8**, 443–450 (2008).
35. Sul, J.Y. *et al.* Transcriptome transfer produces a predictable cellular phenotype. *Proc. Natl. Acad. Sci. USA* **106**, 7624–7629 (2009).
36. Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
    This was the first report describing the use of high-density oligonucleotide arrays for single-cell cDNA microarray analysis.

37. Tietjen, I., Rihel, J. & Dulac, C.G. Single-cell transcriptional profiles and spatial patterning of the mammalian olfactory epithelium. *Int. J. Dev. Biol.* **49**, 201–207 (2005).
38. Sugino, K. *et al.* Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* **9**, 99–107 (2006).
39. Xie, D. *et al.* Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20**, 804–815 (2010).
40. Klein, C.A. Single cell amplification methods for the study of cancer and cellular ageing. *Mech. Ageing Dev.* **126**, 147–151 (2005).
41. Tang, F. *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
This first report of single-cell RNA-seq digital transcriptome analysis based on next-generation sequencing described complex features of single-cell transcriptomes at single-base resolution.
42. Schütze, K. & Lahr, G. Identification of expressed genes by laser-mediated manipulation of single cells. *Nat. Biotechnol.* **16**, 737–742 (1998).
43. Galbraith, D.W., Elumalai, R. & Gong, F.C. Integrative flow cytometric and microarray approaches for use in transcriptional profiling. *Methods Mol. Biol.* **263**, 259–280 (2004).
44. 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).
45. Warren, L.A. *et al.* Transcriptional instability is not a universal attribute of aging. *Aging Cell* **6**, 775–782 (2007).
46. Zhang, C., Barthelson, R.A., Lambert, G.M. & Galbraith, D.W. Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. *Plant Physiol.* **147**, 30–40 (2008).
47. Spiller, D.G., Wood, C.D., Rand, D.A. & White, M.R. Measurement of single-cell dynamics. *Nature* **465**, 736–745 (2010).
48. Ståhlberg, A. & Bengtsson, M. Single-cell gene expression profiling using reverse transcription quantitative real-time PCR. *Methods* **50**, 282–288 (2010).
49. Taniguchi, K., Kajiyama, T. & Kambara, H. Quantitative analysis of gene expression in a single cell by qPCR. *Nat. Methods* **6**, 503–506 (2009).
50. Livesey, F.J. Strategies for microarray analysis of limiting amounts of RNA. *Brief. Funct. Genomics Proteomics* **2**, 31–36 (2003).
51. Kawasaki, E.S. Microarrays and the gene expression profile of a single cell. *Ann. NY Acad. Sci.* **1020**, 92–100 (2004).
52. 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).
53. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. *Cell Stem Cell* **6**, 468–478 (2010).
54. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).
55. Pepke, S., Wold, B. & Mortazavi, A. Computation for ChIP-seq and RNA-seq studies. *Nat. Methods* **6**, S22–S32 (2009).
56. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
57. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28**, 503–510 (2010).
58. Griffith, M. *et al.* Alternative expression analysis by RNA sequencing. *Nat. Methods* **7**, 843–847 (2010).
59. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
60. Robertson, G. *et al.* De novo assembly and analysis of RNA-seq data. *Nat. Methods* **7**, 909–912 (2010).
61. Lee, S. *et al.* Accurate quantification of transcriptome from RNA-seq data by effective length normalization. *Nucleic Acids Res.* **39**, e9 (2011).
62. Baserga, R. Is cell size important? *Cell Cycle* **6**, 814–816 (2007).
63. Crissman, H.A. & Steinkamp, J.A. Rapid, simultaneous measurement of DNA, protein, and cell volume in single cells from large mammalian cell populations. *J. Cell Biol.* **59**, 766–771 (1973).
64. 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).
65. Kurimoto, K. & Saitou, M. Single-cell cDNA microarray profiling of complex biological processes of differentiation. *Curr. Opin. Genet. Dev.* **20**, 470–477 (2010).
66. 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).
67. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
68. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
69. Shackleton, M., Quintana, E., Fearon, E.R. & Morrison, S.J. Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell* **138**, 822–829 (2009).
70. Willis, D.E. & Twiss, J.L. Regulation of protein levels in subcellular domains through mRNA transport and localized translation. *Mol. Cell. Proteomics* **9**, 952–962 (2010).
71. Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat. Methods* **6**, 613–618 (2009).
72. Levin, J.Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7**, 709–715 (2010).
73. Schroeder, T. Imaging stem-cell-driven regeneration in mammals. *Nature* **453**, 345–351 (2008).
74. Smith, Z.D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat. Biotechnol* **28**, 521–526 (2010).
75. Ozsolak, F. *et al.* Direct RNA sequencing. *Nature* **461**, 814–818 (2009).
76. Ozsolak, F. *et al.* Amplification-free digital gene expression profiling from minute cell quantities. *Nat. Methods* **7**, 619–621 (2010).
77. Treffer, R. & Deckert, V. Recent advances in single-molecule sequencing. *Curr. Opin. Biotechnol.* **21**, 4–11 (2010).
78. Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
79. Doyle, J.P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
80. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
81. Mustroph, A. *et al.* Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **106**, 18843–18848 (2009).
82. Sanz, E. *et al.* Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. USA* **106**, 13939–13944 (2009).
83. Zanetti, M.E., Chang, I.-F., Gong, F.C., Galbraith, D.W. & Bailey-Serres, J. Immunopurification of polyribosomal complexes of *Arabidopsis* for global analysis of gene expression. *Plant Physiol.* **138**, 624–635 (2005).
84. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21**, 644–648 (2007).
85. Tang, F., Hajkova, P., Barton, S.C., Lao, K. & Surani, M.A. MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res.* **34**, e9 (2006).
86. Tang, F. *et al.* 220-plex microRNA expression profile of a single cell. *Nat. Protoc.* **1**, 1154–1159 (2006).
87. Guo, G. *et al.* Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685 (2010).
88. Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
89. Femino, A.M., Fay, F.S., Fogarty, K. & Singer, R.H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).

© 2011 Nature America, Inc. All rights reserved.

npg