Published in final edited form as:
Nat Methods. 2011 April ; 8(4 Suppl): S6–11. doi:10.1038/nmeth.1557.

Development and applications of single cell transcriptome analysis

Fuchou Tang[1,3], **Kaiqin Lao**[2], and **M. Azim Surani**[1,*]

[1]Wellcome Trust/Cancer Research UK Gurdon Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QN, UK

[2]Genetic Systems, Applied Biosystems, part of Life Technologies, 850 Lincoln Centre Drive, Foster City, CA 94404, USA

[3]BIOPIC, School of Life Sciences, Peking University, Beijing, 100871, China

Abstract

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. Combined 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 for the analysis of diseases.

Importance of single cell transcriptome analysis

Development is driven and controlled by temporal and spatial changes in gene transcription, followed by translation of the resulting messenger RNAs (mRNA) into proteins. The transcriptome is broadly defined as the entire RNA component of an individual cell, or narrowly defined as the polyadenylated products of RNA polymerase II[1,2]. Recent advances make it possible to obtain information on single cell transcriptomes at high resolution by RNA-Seq analysis, which can be instructive concerning how individual cells respond to signals and other environmental cues at critical stages of cell fate determination, or when they acquire aberrant phenotype. Essentially all cells within an individual organism share a virtually identical genotype, but the individual transcriptomes reflect expression of a subset of genes, which is determined by their epigenetic state, including DNA methylation and histone modifications. Individual diverse cell types exhibit a unique transcriptome, which can be used to assess gene regulation network underlying their physiological functions, behavior and phenotype during development, and for their role in multi-cellular organisms[2]. Although potentially every individual cell has a unique transcriptome, differences in mRNA abundance of some genes may not necessarily lead to a measurable functional consequence.

Ideally transcriptome analysis should be carried out at single cell resolution and should encapsulate the exact sequence, quantity, localization, nature (including modifications), and activity (for example, being actively translated or degraded) of all types of full length RNAs at single-base resolution[1-3]. Due to technical limitations, such as the efficiency of RNA purification and sensitivity of detection, majority of studies on transcriptomes have been carried out with hundreds of thousands or even millions of cells[4-7]. However, in some instances, it is not possible to collect large numbers of cells, such as from very early embryos where only a few precursor cells of specific lineages are available for analysis,

*Correspondence should be addressed to K.L (Kai.Lao@lifetech.com) or M.A.S. (a.surani@gurdon.cam.ac.uk) .

Europe PMC Funders Author Manuscripts

which makes their transcriptome analysis very difficult, if not impossible[8]. Indeed, there are also only limited numbers of some adult progenitor and stem cells.

Recent studies have also shown that gene expression is invariably heterogeneous even when comparing evidently similar cell types[9,10]. Such stochastic variations in the transcriptomes have important implications for cell fate decisions[11,]. For example, Xie's lab 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 are heterogeneous and could contain a small number of cancer stem cells[9,14]. The heterogeneity of gene expression amongst similar cell types can be due to the differences in the epigenetic status of the genome, the circadian clock, cell cycle, microenvironment or the niche, as well as the 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 majority genes within an individual cell, only one (for prokaryotes) or two copies (for majority 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 comprehend the basis and significance 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 the resolution of single cells was pioneered two decades ago by the groundbreaking work of Norman N. Iscove, using exponential amplification of cDNAs by PCR[23], and by James Eberwine using linear amplification of cDNAs by T7 RNA polymerase-based in vitro transcription[24,25] (T7-IVT) (Figure 1). This approach greatly facilitated and accelerated insights on the molecular mechanisms of development and function of mammalian neural system, especially because they are probably the most heterogeneous group of cells. In this case, transcriptomes at the cellular or even sub-cellular resolution within 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). While this method is powerful and able to provide gene expression pattern at whole genome scale[32,40], the cDNA fragments being amplified are in general short (several hundred base pairs), and unable to detect transcripts generated through alternative splicing. Most importantly, the method can only detect known genes.

Generating single cell transcriptomes

To generate single cell transcriptomes, individual intact cells are isolated and transferred into an Eppendorf tube containing the lysate buffer. This is followed by direct reverse transcription of the whole cell lysate using oligo-dT primers to convert mRNAs with poly(A) tail into first-strand cDNAs. The residual mRNA templates are degraded and a poly(A) tail is added to the 3′ end of the first-strand cDNAs. These cDNAs are uniformly amplified with oligo-dT primers. A key requirement for the procedure is that the buffers used for earlier steps are compatible with the reactions for later steps[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, all of which can potentially lead to a significant loss of the small amounts of mRNAs, and subsequently cDNAs, generated from individual cells. The amplified single cell cDNAs can be quality-checked and tested by qPCR, and only selected samples need to be analyzed further by cDNA microarray or deep-sequencing.

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 step can be time-

consuming and technically challenging. Laser-assisted microdissection (LMD) or fluorescent-activated cell sorting (FACS) can also be used to isolate specific sub-population of cells based on cell surface markers or fluorescent reporters[42-45], which can be used to achieve high accuracy and throughput. Cells of higher plants with cell walls are difficult to dissociate by enzymes, but isolation of nuclei after homogenization of tissues might be effective in this instance[46].

It should be possible to use microfluidics systems in the future to isolate and track thousands of single cells accurately and in parallel within a nanolitter of solution[34,47]. This approach 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 different types of detergents can be considered, such as guanidine thiocyanate and Nonidet P-40[30,32,33,40,48]. Clearly, both the type and amount of detergents used need to be adjusted to obtain the best results for different cell types. This depends on the propensity of cells to undergo lysis. When working on a new cell type, several detergents should be tested in parallel at different concentrations to obtain the best conditions for the specific cell type. An alternative strategy reported to work for single cell cDNA microarray analysis was to isolate and purify mRNAs from a single cell using oligo-dT coated magnetic beads that will remove proteins, metabolites, and the cell debris, from the mRNAs[32,33,40]. This approach can also be combined with the isolation of genomic DNA from the same cell for the analysis of the genotype. Using this method, the lysate buffer can also be washed away while specifically isolating mRNAs with poly(A) tails, allowing the use of much stronger lysate conditions for the quick and efficient release of mRNAs[32]. However, in most instances, the simpler option of the whole cell lysate can be directly used for reverse transcription without the necessity for the removal of the detergents[23,30].

Several types of reverse transcriptases are available for the preparation of cDNAs[49]. The Superscript III is most widely used for this purpose, which can potentially generate full length cDNAs of up to 10 kb[30,41,49]. When using the whole cell lysate, oligo-dT primer is usually used for this step since the rRNAs/tRNAs are two orders more abundant than mRNAs. The use of random primers would result in the amplification of 95% to 98% of the cDNAs later during the PCR step as cDNAs of rRNAs/tRNAs. To prevent this, mRNAs could be specifically isolated and purified, which can then be combined with the use of random primers as an alternative for priming the reverse transcription to obtain full length mRNAs. Here the concentrations of dNTPs need to be carefully adjusted to permit efficient reverse transcription and to avoid interference with a later poly(A) tailing step. The sequences of the 5′ ends of the mRNAs of different genes and the corresponding 3′ ends of the first-strand cDNAs are different. Thus, the use of terminal deoxynucleotidyl transferase will add a poly(A) tail to it, and provide unbiased amplification of all expressed genes' cDNAs.

After reverse transcription and tailing steps, the single cell cDNAs can be amplified. One representative individual mammalian cell contains about 10 pg of total RNAs, and about 0.1 pg of mRNAs, which usually needs to be amplified around ten million folds to match the requirement for a standard microarray analysis. Either PCR or IVT can be used for amplification[30,50,51]. The advantage of PCR strategy is the exponential amplification of the cDNAs so that single cell cDNAs can be amplified millions of folds within several hours[30,50]. However, its disadvantage is the accumulation of primer dimers and other non-specific byproducts during amplification, especially during later cycles of PCR[23,30]. The merit of IVT strategy is its stringent specificity while reducing accumulation of non-specific

byproducts[24]. Its drawback is that the protocol generates cRNAs and their lengths are typically less than 1 kb[50]. IVT procedure is also more tedious and time-consuming and every round of IVT can amplify the cDNAs for only up to one thousand folds[50,51]. In practice, single cell cDNAs can be amplified sufficiently for microarray analysis by two rounds of PCR amplification[32-34], three rounds of IVT amplification[35], or a combination of PCR and IVT amplification[30]. In general, all these strategies can be tailored for individual needs to work robustly.

We have recently improved a widely used single cell cDNA amplification protocol[30,52], which is highly quantitative, except that it generates only about 0.85kb fragments at the 3′ end of the mRNAs that is relatively short and lacks a large segment of the 5′ end sequence and quantity information. Moreover, we combined it with the next generation sequencing technique to develop single cell RNA-Seq analysis[30,41,52]. We have increased the efficiency of the protocol, and can generate up to 3kb fragments of cDNAs instead of 0.85kb fragments reported previously[2,41,52]. 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 its throughput. Furthermore, a previous single cell cDNA microarray analysis detected about 6,800 distinct gene transcripts in an individual embryonic stem (ES) cell, whereas we detected expression of about 10,800 genes, which means that our assay detected expression of nearly 60% more genes in an individual ES cell[53].

Due to the higher sensitivity of next generation sequencing, further amplification by additional IVT step is no longer needed. We showed that the method works faithfully for the detection of full transcriptome of individual early mouse blastomeres. Up to 65% 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 within a single cell, highlighting the complexity of individual cell's transcriptome[41]. Furthermore, thousands of previously unknown exon-exon junctions were found 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 as well as from adult tissues, although the method has not so far been used independently on different cell types.

After the amplification step, single cell cDNAs can be analysed either by microarray or by deep sequencing[30,41]. The deep sequencing approach is more accurate with at least 5 log dynamic range[1,41,54], but it is expensive and needs more computer power for the analysis of the data[41]. Microarray approach is appropriate to obtain general transcriptome information on the presence of transcripts, and on the up or down regulation of transcripts of the known genes. However, deep sequencing provides a more detailed and accurate information on transcriptomes. In general, 20 to 40 million sequencing reads per cell are enough for most purposes, such as detection of novel genes, splicing variants, polyadenylation sites, and for the detection of novel exons of known genes[2].

For bioinformatics analysis, both standard commercial softwares and free academic softwares are available[55]. Normalization of the data is crucial for comparison of transcriptomes of different samples within the same batch, and of those from different labs and platforms. For relative quantification, normalization can be either quantile, RPKM (Reads per kilobases per million reads), or RPM (Reads per million reads)[54]. We have shown that both quantile and RPM normalization work well for our single cell RNA-Seq data[41]. Since our method detected about 3kb cDNA fragments at their 3′ ends and not full

length cDNAs of all mRNAs, RPM works just as well as RPKM. The absolute quantification of cDNAs can be obtained through the use of 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, usually the poly-Lys, Dap, Phe, and Thr peptide RNAs with poly(A) tail are used as spike-in RNAs[30,36,52]. However, 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 significant degradation of the 'spike-in' RNAs during manipulations of the samples, such as dilution and transfer, because the stability of RNAs in such low quantity is often questionable. It is also important to note that the volume of different types of cells is highly variable, which can differ by up to hundreds of folds[56]. The absolute quantification only measures the absolute copy number of each gene's mRNAs within a cell, without consideration of the cell volume. The concentration rather than the absolute copy number of mRNAs is relatively 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[57].

Applications of single cell transcriptome analysis

Single cell transcriptome technique is potentially widely applicable for the analysis of gene expression for essentially any cell type. The single cell transcriptome analysis can be used for determining gene regulatory networks at whole genome scale in an objective and non-biased way. When combined with overexpression, knockout or knockdown of a gene of interest, single cell transcriptome analysis can reveal how it regulates the gene expression network in target cells[41,58,59]. This is especially relevant for the analysis of cells during early embryonic development and for stem cells, when there are limited numbers of cells, and because of the highly dynamic and heterogeneous nature of subpopulations of the cells[41,44]. Analysis of heterogeneity amongst 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]. For example, gene expression networks in a mammalian cell can change dramatically during the day based on the circadian clock[16,17]. More importantly, gene expression is intrinsically stochastic due to different microenvironments, or due to the small number of molecules involved in transcription and translation, which is a general feature of gene expression[19-22]. So 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[12,13,60]. Dissecting the heterogeneity of gene expression within a cell population will thus be an important application of single cell analysis. Indeed, there is evidence for heterogeneity in subpopulations of ES cells based on the expression of Nanog, Rex1, or Stella[15,53]. Cell heterogeneity amongst tumors has been known for a long time[61]. Single cell transcriptome analysis is a feasible strategy to identify the subpopulations within a tumor, and to detect putative cancer stem cells. Since only one individual cell needs to be isolated and lysed from a tissue, it is theoretically possible to analyze gene expression network in a non-invasive way to monitor the progress of human disease, or monitor 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 whole sample.

An additional application of the method described here is to determine gene expression profile of sub-cellular 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[26-29,36,37,62]. Single cell transcriptome analysis can be used to detect mRNAs specifically localized in axon or dendrite, which is often of great importance for determining

physiological functions of these neurons. If the cell body can be separated from axons, then either individual or pooled cell bodies and axons can be analyzed independently by single cell RNA-Seq technique to find all the differentially localized transcripts within cell bodies and axons.

Since next generation sequencing techniques provides information at single base resolution, it is also possible to analyze allele specific gene expression (ASE) within an individual cell, that is the separate mRNA quantities of the two copies of the genomic DNAs of a diploid cell, provided single nucleotide polymorphisms (SNPs) are available to discriminate between the two alleles[63], which will greatly improve our understanding of how the genetic and epigenetic elements influence allelic gene expression within an individual cell. Allelic imbalance (AI) can accurately measure small differences between individual cells, which can arise in a number of ways through changes in relative expression of alleles (ASE) by mutations such as point mutations, or by RNA editing[3]. While recent advances in RNA-Seq analysis offer significant opportunities to evaluate properties of cells, there are also some drawbacks of current single cell RNA-Seq methods. First, the strandness of mRNAs is lost in the library construction, which prevents discrimination between sense and antisense transcripts from the same locus[41]. In the future, it may be possible to use the T7 RNA polymerase-based in vitro transcription and dUTP second strand synthesis strategies to obtain information on sense and antisense transcripts[64].

Second, only the 3′ end and up to 3kb fragment of mRNAs is obtained, which leaves nearly 36% of genes with mRNAs longer than 3kb that cannot be fully elucidated, especially their 5′ UTR and TSS regions. This is primarily due to the use of oligo-dT primers instead of random primers for reverse transcription, and not due to the RNA-Seq library preparation[41]. 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′ UTR and TSS regions for all expressed genes.

Third, the method is based on reverse transcription with oligo-dT primers, so only the mRNAs with poly(A) tail are detected, which will exclude some long non-coding RNAs and most of the small non-coding RNAs[41]. In the future, a more sophisticated protocol should be used to deplete abundant rRNAs and tRNAs while preserving all the mRNA transcripts, including those without poly(A) tail or primary transcripts before polyadenylation. And the strandness 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, it will be possible to recover the strandness information for single cell transcriptomes in the near future[64].

Fourth, the current method does not permit analysis of the transcriptome and genomic sequence of individual cells simultaneously. Improving the method based on the strategy elegantly developed by Klein's lab[32,33,40], it is possible to use next generation sequencing to simultaneously get 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 reporter[65]. 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

Europe PMC Funders Author Manuscripts

approach can also be used for the analysis of cellular reprogramming and trans-differentiation[66].

Currently, all available single cell transcriptome analysis relies on amplification of their cDNAs. Recently developed single molecule sequencing (SMS) has the potential to sequence full length mRNAs from a single cell directly without reverse transcription and amplification steps, which will more accurately determine expression levels of different splicing isoforms[67-69]. Moreover, the full-length mRNA sequences will accurately determine allele specific gene expression (ASE) with defined phase information of each locus. However, the sequencing efficiency of current SMS techniques still requires a few hundred cells, and only detects about 15-25% of expressed mRNAs, which needs to be improved to achieve single cell RNA-Seq, since only dozens of copies of mRNAs are produced from the majority of individual expressed genes in a cell[68]. Furthermore, the accuracy of the SMS 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[68,69]. This will probably be achieved by improving the SMS technique to permit sequencing a single mRNA/cDNA molecule repeatedly without damaging it. At the moment, only the static amount of mRNAs is obtained by single cell transcriptome analysis, which is the result of the balance between transcription and degradation of mRNAs. More significant analysis will require accurate quantification of mRNAs being actively translated[70-75], which are the mRNAs with ribosomes on them, or single cell translating-RNA-Seq. This will directly reflect the translational activity and function of the genes at particular time points[70-75]. 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, and how it survives and behaves appropriately with a complex and noisy transcriptome.

References

1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 2009; 10:57–63. [PubMed: 19015660]
2. Tang F, et al. RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protoc. 2010; 5:516–535. [PubMed: 20203668]
3. He C. Grand Challenge Commentary: RNA epigenetics? Nat. Chem. Biol. 2010; 6:863–865. [PubMed: 21079590]
4. Schulze A, Downward J. Navigating gene expression using microarrays--a technology review. Nat. Cell Biol. 2001; 3:E190–195. [PubMed: 11483980]
5. Cloonan N, Grimmond SM. Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. 2008; 9:234. [PubMed: 18828881]
6. Wold B, Myers RM. Sequence census methods for functional genomics. Nat. Methods. 2008; 5:19–21. [PubMed: 18165803]
7. Schuster SC. Next-generation sequencing transforms today's biology. Nat. Methods. 2008; 5:16–18. [PubMed: 18165802]
8. Saitou M, Barton SC, Surani MA. A molecular programme for the specification of germ cell fate in mice. Nature. 2002; 418:293–300. [PubMed: 12124616]
9. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development. 2009; 136:3853–3862. [PubMed: 19906852]
10. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. Science. 2010; 327:542–545. [PubMed: 20110496]

11. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010; 467:167–173. [PubMed: 20829787]
12. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science. 2008; 322:442–446. [PubMed: 18927393]
13. Taniguchi Y, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. 2010; 329:533–538. [PubMed: 20671182]
14. Frank NY, Schatton T, Frank MH. The therapeutic promise of the cancer stem cell concept. J. Clin. Invest. 2010; 120:41–50. [PubMed: 20051635]
15. Hayashi K, Lopes SM, Tang F, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell. 2008; 3:391–401. [PubMed: 18940731]
16. Panda S, Hogenesch JB, Kay SA. Circadian rhythms from flies to human. Nature. 2002; 417:329–35. [PubMed: 12015613]
17. Maury E, Ramsey KM, Bass J. Circadian rhythms and metabolic syndrome: from experimental genetics to human disease. Circ. Res. 2010; 106:447–462. [PubMed: 20167942]
18. Wittenberg C, Reed SI. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. Oncogene. 2005; 24:2746–2755. [PubMed: 15838511]
19. Arias AM, Hayward P. Filtering transcriptional noise during development: concepts and mechanisms. Nat. Rev. Genet. 2006; 7:34–44. [PubMed: 16369570]
20. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135:216–226. [PubMed: 18957198]
21. Losick R, Desplan C. Stochasticity and cell fate. Science. 2008; 320:65–68. [PubMed: 18388284]
22. Shahrezaei V, Swain PS. The stochastic nature of biochemical networks. Curr. Opin. Biotechnol. 2008; 19:369–374. [PubMed: 18662776]
23. Brady G, Barbara M, Iscove NN. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. Methods Mol. Cell Biol. 1990; 2:17–25.
24. Eberwine J, et al. Analysis of gene expression in single live neurons. Proc. Natl. Acad. Sci. U S A. 1992; 89:3010–3014. [PubMed: 1557406]
25. Van Gelder RN, et al. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. U S A. 1990; 87:1663–1667. [PubMed: 1689846]
26. Dulac C, Axel R. A novel family of genes encoding putative pheromone receptors in mammals. Cell. 1995; 83:195–206. [PubMed: 7585937]
27. Tanabe Y, William C, Jessell TM. Specification of motor neuron identity by the MNR2 homeodomain protein. Cell. 1998; 95:67–80. [PubMed: 9778248]
28. Yamagata M, Weiner JA, Sanes JR. Sidekicks: synaptic adhesion molecules that promote lamina-specific connectivity in the retina. Cell. 2002; 110:649–660. [PubMed: 12230981]
29. Shumyatsky GP, et al. Identification of a signaling network in lateral nucleus of amygdala important for inhibiting memory specifically related to learned fear. Cell. 2002; 111:905–918. [PubMed: 12526815]
30. Kurimoto K, et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. 2006; 34:e42. [PubMed: 16547197]
31. Jensen KB, Watt FM. Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. Proc. Natl. Acad. Sci. U S A. 2006; 103:11958–11963. [PubMed: 16877544]
32. Klein CA, et al. Combined transcriptome and genome analysis of single micrometastatic cells. Nat. Biotechnol. 2002; 20:387–392. [PubMed: 11923846]
33. Hartmann CH, Klein CA. Gene expression profiling of single cells on large-scale oligonucleotide arrays. Nucleic Acids Res. 2006; 34:e143. [PubMed: 17071717]
34. Bontoux N, et al. Integrating whole transcriptome assays on a lab-on-a-chip for single cell gene profiling. Lab Chip. 2008; 8:443–450. [PubMed: 18305863]
35. Sul JY, et al. Transcriptome transfer produces a predictable cellular phenotype. Proc. Natl. Acad. Sci. U S A. 2009; 106:7624–7629. [PubMed: 19380745]

36. Tietjen I, et al. Single-cell transcriptional analysis of neuronal progenitors. Neuron. 2003; 38:161–175. [PubMed: 12718852]
37. Tietjen I, Rihel J, Dulac CG. Single-cell transcriptional profiles and spatial patterning of the mammalian olfactory epithelium. Int. J. Dev. Biol. 2005; 49:201–217. [PubMed: 15906233]
38. Sugino K, et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. 2006; 9:99–107. [PubMed: 16369481]
39. Xie D, et al. Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Res. 2010; 20:804–815. [PubMed: 20219939]
40. Klein CA. Single cell amplification methods for the study of cancer and cellular ageing. Mech. Ageing Dev. 2005; 126:147–151. [PubMed: 15610773]
41. Tang F, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods. 2009; 6:377–382. [PubMed: 19349980]
42. Schütze K, Lahr G. Identification of expressed genes by laser-mediated manipulation of single cells. Nat. Biotechnol. 1998; 16:737–742. [PubMed: 9702771]
43. Galbraith DW, Elumalai R, Gong FC. Integrative flow cytometric and microarray approaches for use in transcriptional profiling. Methods Mol. Biol. 2004; 263:259–280. [PubMed: 14976371]
44. Warren L, Bryder D, Weissman IL, Quake SR. Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc. Natl. Acad. Sci. U S A. 2006; 103:17807–17812. [PubMed: 17098862]
45. Warren LA, et al. Transcriptional instability is not a universal attribute of aging. Aging Cell. 2007; 6:775–782. [PubMed: 17925006]
46. Zhang C, Barthelson RA, Lambert GM, Galbraith DW. Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol. 2008; 147:30–40. [PubMed: 18354040]
47. Spiller DG, Wood CD, Rand DA, White MR. Measurement of single-cell dynamics. Nature. 2010; 465:736–745. [PubMed: 20535203]
48. Ståhlberg A, Bengtsson M. Single-cell gene expression profiling using reverse transcription quantitative real-time PCR. Methods. 2010; 50:282–288. [PubMed: 20064613]
49. Taniguchi K, Kajiyama T, Kambara H. Quantitative analysis of gene expression in a single cell by qPCR. Nat. Methods. 2009; 6:503–506. [PubMed: 19525960]
50. Livesey FJ. Strategies for microarray analysis of limiting amounts of RNA. Brief. Funct. Genomic Proteomic. 2003; 2:31–36. [PubMed: 15239941]
51. Kawasaki ES. Microarrays and the gene expression profile of a single cell. Ann. N Y Acad. Sci. 2004; 1020:92–100. [PubMed: 15208186]
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. 2007; 2:739–752. [PubMed: 17406636]
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. 2010; 6:468–478. [PubMed: 20452321]
54. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 2008; 5:621–628. [PubMed: 18516045]
55. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. Nat. Methods. 2009; 6:S22–S32. [PubMed: 19844228]
56. Baserga R. Is cell size important? Cell Cycle. 2007; 6:814–816. [PubMed: 17404503]
57. Crissman HA, Steinkamp JA. Rapid, simultaneous measurement of DNA, protein, and cell volume in single cells from large mammalian cell populations. J. Cell Biol. 1973; 59:766–771. [PubMed: 4128323]
58. Kurimoto K, et al. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008; 22:1617–1635. [PubMed: 18559478]
59. Kurimoto K, Saitou M. Single-cell cDNA microarray profiling of complex biological processes of differentiation. Curr. Opin. Genet. Dev. 2010; 20:470–477. [PubMed: 20619631]

60. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009; 459:428–432. [PubMed: 19363473]
61. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009; 138:822–829. [PubMed: 19737509]
62. Willis DE, Twiss JL. Regulation of protein levels in subcellular domains through mRNA transport and localized translation. Mol. Cell. Proteomics. 2010; 9:952–962. [PubMed: 20167945]
63. Zhang K, et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. Nat. Methods. 2009; 6:613–618. [PubMed: 19620972]
64. Levin JZ, et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods. 2010; 7:709–715. [PubMed: 20711195]
65. Schroeder T. Imaging stem-cell-driven regeneration in mammals. Nature. 2008; 453:345–351. [PubMed: 18480816]
66. Chan EM, et al. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nat. Biotechnol. 2009; 27:1033–1037. [PubMed: 19826408]
67. Ozsolak F, et al. Direct RNA sequencing. Nature. 2009; 461:814–818. [PubMed: 19776739]
68. Ozsolak F, et al. Amplification-free digital gene expression profiling from minute cell quantities. Nat. Methods. 2010; 7:619–621. [PubMed: 20639869]
69. Treffer R, Deckert V. Recent advances in single-molecule sequencing. Curr. Opin. Biotechnol. 2010; 21:4–11. [PubMed: 20202812]
70. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010; 466:835–840. [PubMed: 20703300]
71. Doyle JP, et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell. 2008; 135:749–762. [PubMed: 19013282]
72. Heiman M, et al. A translational profiling approach for the molecular characterization of CNS cell types. Cell. 2008; 135:738–748. [PubMed: 19013281]
73. Mustroph A, et al. Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. Proc. Natl. Acad. Sci. U S A. 2009; 106:18849–18854. [PubMed: 19846784]
74. Sanz E, et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc. Natl. Acad. Sci. U S A. 2009; 106:13939–13944. [PubMed: 19666516]
75. Zanetti ME, Chang I-F, Gong FC, Galbraith DW, Bailey-Serres J. Immunopurification of polyribosomal complexes of arabidopsis for global analysis of gene expression. Plant Physiol. 2005; 138:624–635. [PubMed: 15955926]
76. Tang F, et al. Maternal microRNAs are essential for mouse zygotic development. Genes Dev. 2007; 21:644–648. [PubMed: 17369397]
77. Tang F, Hajkova P, Barton SC, Lao K, Surani MA. MicroRNA expression profiling of single whole embryonic stem cells. Nucleic Acids Res. 2006; 34:e9. [PubMed: 16434699]
78. Tang F, et al. 220-plex microRNA expression profile of a single cell. Nat. Protoc. 2006; 1:1154–1159. [PubMed: 17406397]
79. Guo G, et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell. 2010; 18:675–685. [PubMed: 20412781]
80. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods. 2008; 5:877–879. [PubMed: 18806792]
81. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of single RNA transcripts in situ. Science. 1998; 280:585–590. [PubMed: 9554849]

Europe PMC Funders Author Manuscripts

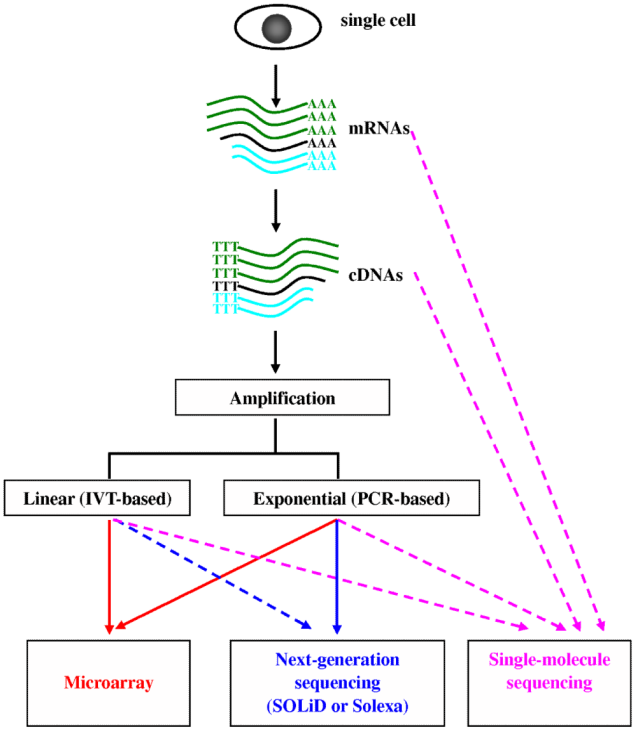

Figure 1.
Strategies for single cell transcriptome analysis. Solid lines represent strategies that have been demonstrated experimentally; dotted lines represent proposed strategies that will probably be realized in the near future.

Europe PMC Funders Author Manuscripts

Table 1

Single cell multiplex gene expression analysis strategies

Techniques	target RNAs	With Amplification	Throughput	References
single cell RNA-Seq	mRNAs with polyA tail	Yes	high (whole transcriptome)	41,53
single cell cDNA microarray (PCR-based)	mRNAs with polyA tail	Yes	high (whole transcriptome with pre-designed probes)	31,32,33,34,36
single cell cDNA microarray (IVT-based)	mRNAs with polyA tail	Yes	high (whole transcriptome with pre-designed probes)	35
single cell cDNA microarray (Combination of PCR & IVT)	mRNAs with polyA tail	Yes	high (whole transcriptome with pre-designed probes)	15,30,58,76
single cell microRNA profiling	known microRNAs	Yes	high (up to several hundred microRNAs)	76,77,78
single cell multiplex qPCR	known mRNAs	Yes	low (up to several hundred genes)	15,44,45,53,79
single cell multiplex qPCR	known mRNAs	No	low (up to five genes)	49
single molecule RNA FISH	known mRNAs	No	low (up to several genes)	80,81