REVIEW

Technologies of high-throughput tissue-specific gene expression profiling

Zhichao Xu, Xiao Liu

MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China

Correspondence: Xiao Liu
E-mail: xiaoliu@tsinghua.edu.cn
Received: December 11, 2017
Published: September 24, 2018

Gene expression profiling is an important strategy to study animal development, response to stimuli and diseases. RNAs measured in gene expression profiling experiments are frequently purified from mixture of multiple cell types. The resultant data have low resolution, incapable of distinguishing transcriptome of different cell types and likely biased towards up-regulated genes in dominant tissues. These problems can be solved by obtaining tissue-specific gene expression profile. For dozens of years, there have been several strategies developed to isolate specific tissues or purify RNAs from tissue of interest, and combined with high-throughput RNA assays to generate transcriptome of various specific tissues or cell types. This review will introduce basic principles of these methods and their application in large-scale transcriptome analysis, and discuss on their advantages and limitation.

Introduction

In multicellular organisms, vital processes including development, physiology and response to stimuli require precise regulation of genome on the molecular level, which is manifested as temporal and spatial expression patterns of genes. Gene expression profiling is a substantial measurement of gene expression patterns under various conditions that contribute to a global picture of gene transcription genome-wide. Tissue-specific transcriptome analysis on this basis provides us with comprehensive insights into the complicated nature of genome.

Fluorescence microscopy played a significant role among approaches of profiling tissue-specific gene expression. Fluorescence in situ hybridization (FISH) [1] uses fluorescent probes to bind specific nucleic acid targets, while immunofluorescence can use fluorophore-linked antibodies to target specific proteins [2]. Recombinant technology allows a fluorescent protein to ligate covalently to the product of target gene [3, 4]. Generally, gene profiling technologies based on fluorescence microscopy are not only intuitive but also precise enough to reach single molecular level accuracy [5]. However, these technologies are expensive and labor intensive in large scale gene expression profiling projects, let alone that image analysis is not automated enough to

improve the throughput.

Methods of large scale gene expression profiling have also been developed for decades. Microarrays are capable of producing large amounts of gene expression data based on hybridization of mRNA output to probes [6]. Meanwhile, approaches based on next-generation sequencing technology (SGS) such as serial analysis of gene expression (SAGE) [7], RNA-Seq [8], chromatin immunoprecipitation sequencing (ChIP-Seq) [9, 10], DNase I hypersensitive sites sequencing (DNase-Seq) [11, 12] and assay for transposase-accessible chromatin with high throughput sequencing (ATAC-Seq) [13] can profile gene expression in the form of sequence read, which is feasible to quantification and automated analysis. In this way, gene expression patterns and regulatory networks can be characterized systematically. However, despite the fact that these methods can be scaled up, the biological meanings and regulation mechanisms can often be ambiguous. One of key reasons is that sample preparation process can not reach tissue level sometime, especially in isolation or cultivation of a specific cell type in small organisms, e.g. *Caenorhabditis elegans* and *Drosophila melanogaster*.

Recent work focusing on functional genomics have enabled leaps in methodology of efficient profiling of tissue-specific transcriptome, including cell sorting, RNA immunoprecipitation, RNA tagging. New designs are constantly developed with improvement in not only throughput but also precision. The focus of this review is to outline basic principles of these methods and their application in large-scale transcriptome analysis, and discuss on their advantages and limitation.

Cell sorting

The idea to obtain tissue-specific transcriptome based on cell sorting has been developed for decades. Depending on tissue types, frequently used sample dissection techniques include direct isolation [14], enzymatic disaggregation [15], laser-microdissection [16] and direct single cell RNA sequencing [17]. For samples that are easily dissociated, cells can be labelled with fluorescence using fluorescent protein reporter assay or immunofluorescence targeting tissue-specific antigens on cell membranes before fluorescence-activated cell sorting (FACS) is used to collect fluorescence labelled cells for further gene expression profiling. A recent study has proven that gene expression profiling based on cell sorting can reach the level of single cells in mammalian embryos. After the mouse embryo was dissected and dissociated into a single-cell suspension, cells from the epiblast and nascent Flk1+ mesoderm of gastrulating mouse embryos were immunofluorescence labelled before FACS. This way, researchers were able to investigate early mesoderm diversification through single-cell RNA sequencing [18]. Cell sorting is feasible not only in mammalian tissue and cell culture, but also in other model organisms. For instance, in order to identify genes related to touch receptor neurons in *C. elegans*, researchers generated worm strains expressing green fluorescent protein (GFP) only in mechanosensory neurons and their precursor cells. After disaggregating worm embryo into single cells, cells were cultured *in vitro* until some of them expressed GFP and then these cells were isolated by FACS. Comparing microarray analysis result from cells from wild type and *mec-3* mutant, which fails to develop mechanosensory neurons, researchers were able to discover genes overexpressed or suppressed in the process of mechanosensory neuron development and unveil some regulators previously unknown [19]. Scientists also isolated cells from maize shoot apical meristem (SAM) using laser-microdissection. After next generation sequencing of the selected cells' cDNA library, new transcripts were discovered [20].

However, the efficiency of methods based on cell sorting drastically depends on sample types. Thus, cell sorting is frequently used in embryo, blood, bone marrow, etc., while in other cases, additional manipulations are usually needed. Taking *C. elegans* as an example, in order to circumvent the tough cuticle barriers that larvae and adults carry, neurons and muscle cells can be cultured *in vitro* after isolation from embryo [21]. This result in obscure understanding of biological processes *in vivo*. In another study, researchers applied the isolation of Nuclei Tagged in A specific Cell Type (INTACT) technique [22, 23]. In this method, nuclei were affinity-labeled through the transgenic co-expression of nuclear tagging fusion (NTF) carrying biotin ligase recognition peptide (BLRP) together with biotin ligase (BirA) in the cell type of interest. After total nuclei were isolated, biotin-labeled nuclei purified using streptavidin-coated magnetic beads were prepared for transcriptome and epigenetic analysis. INTACT has been successfully applied in many model organisms, e.g. *C. elegans*, *D. melanogaster* and *A. thaliana* [22-24].

RNA immunoprecipitation

RNA immunoprecipitation (RIP) is the method focused on interaction between RNA and protein *in vivo*. The basic principle of RIP is generating protein-RNA cross-links between molecules in close proximity in *vivo*, then RNA molecules cross-linked with a given protein are isolated by immunoprecipitation of the protein [25]. RIP assay targeting poly(A)-binding protein (PAB) driven by tissue-specific promoter can achieve the collection of messenger RNAs

from a certain group of cells. This method called PAB-RIP was first tested in *C. elegans* muscle [26]. PAB-RIP is very sensitive, even up to single-cell level. As proved in a recent study, two transgenetic worm strains separately expressing PAB in the amphid sensilla neuron ASEL and ASER was generated. Using PAB-RIP and genome-wide microarray, researchers compared the gene expression profiles of two functionally divergent neurons in the *C. elegans* gustatory neuron class ASE and discovered 188 genes with asymmetry expression in ASEL and ASER [27]. PAB-RIP also played an important role in *C. elegans* functional genomic studies. In the model organism ENcyclopedia of DNA Element (modENCODE) project, PAB-RIP was used to measure the transcriptome of 25 different tissues from distinct developmental stages. Tiling array profiles revealed the spatial and temporal expression pattern in 75% of genes while more than 200 long non-coding RNAs (lncRNA) were discovered, most of which also showed tissue-specific expression patterns [28]. Generally, PAB-RIP generates large scale tissue-specific transcriptome data, which establishes the foundation of high-throughput and precise gene expression profile.

The shortcoming of PAB-RIP lies in the insufficient signal-to-noise ratio when targeting a very small number of tagged cells among the whole tissue extracts in large organisms. RNA-IP also requires the non-covalent interaction between mRNA and PAB to be fixed by formaldehyde cross-linking, which is time consuming to optimize the protocol. In order to reduce the background noise, sample can be firstly dissected and selected to remove unnecessary input. In an enhanced technique named translating ribosome affinity purification (TRAP), researchers introduced fusions of ribosomal proteins with enhanced green fluorescent protein (EGFP) in several specific mouse neurons. Here EGFPs not only provided visualization of target cells, which are more conveniently isolated under a microscope, but are also easily purified by anti-EGFP-coated magnetic beads after the immunoprecipitation assay targeting ribosomal proteins. Using TRAP, researchers compared mRNA profiles from EGFP-bound ribosomes and those from unbound fractions to distinguish two morphologically identical, incorporated subclasses of medium spiny neurons at the transcriptome level [29, 30].

RNA tagging

Protozoan genomes encode a few nucleotide synthetases that are able to process nucleoside analogues into nucleotide analogues, which can be incorporated into RNA molecules. The principle of tissue-specific gene expression profiling using covalent RNA labeling is transforming one of those nucleotide synthetases driven by a tissue-specific promoter into cells of interest and label RNA molecules in these cells with nucleoside analogues. By isolating labeled RNAs, the cell type-specific transcriptome is obtained. For example, uracil phosphoribosyltransferase (UPRT) from *Toxoplasma gondii* can transform the thio-substituted compound 4-thiouracil (4tU) into 4tUMP [31]. Researchers used the GAL4/UAS system to drive UPRT expression specifically in glial cells of Drosophila larval brains. After total RNAs were extracted from transgenic larvae fed with 4tU, RNA molecules carrying 4tU were biotin labeled and then purified using streptavidin affinity chromatography. Results showed that glia-specific RNA were effectively isolate by TU-tagging without prior cell dissociation [32]. The bias that transcripts with more uracils had larger chance to be observed was removed by normalization using regression equation [32].

Gene expression profiling using the TU-tagging method depends on not only the specific expression of UPRT but also the 4TU treatment, which provides an opportunity for researchers to investigate temporal and spatial expression patterns through a chemical/genetic intersectional method. In this advanced design, the spatial specificity of gene expression is confined by cell-specifically expressed UPRT while the temporal specificity is confined by precise treatment of 4TU [33]. However, the TU-tagging method is not widely used currently because of its complexity.

Another RNA tagging method named trans-Splicing based RNA Tagging (SRT) was invented based on the mechanism of trans-splicing between the spliced leader (SL) RNA and pre-mRNA in *C. elegans* [34]. Through driving the expression of tagged SL1 transgenes by a tissue-specific promoter, cell type-specific transcriptome can be separated by tag. For example, the Illumina adaptor sequence was used to tag SL1 and the tagged SL1 was driven by the tissue-specific myo-3 promoter. Therefore researchers were able to profile the gene expression of body muscle and discovered some new transcripts in body muscle. However, trans-splicing between the spliced leader (SL) RNA and pre-mRNA does not exist in most organisms. Even in *C. elegans*, there are ~30% of genes that are not trans-spliced under the mechanism and thus can not be profiled.

Discussion

The main purpose of studying gene regulation is to investigate the precise time and place of gene expression. Tools of gene expression profiling have witnessed and proved the substantial progress in genomics. For example, SRT is based on full study in Splicing-Leader-guided trans-splicing, TU-tagging depends on the discovery of

UPRT and PAB-RIP relies on the discovery of PAB and development in the immunoprecipitation assay. Therefore, with new discoveries and progress in genomics in the future, we can anticipate new methods in high-throughput gene expression profiling and further development in existing techniques. Together with advances in imaging, statistical methods and computational approaches, study of gene expression profiling will guided us to a deeper understanding of gene regulation.

In conclusion, precise measurement of amount, time and place of gene expression is a foundation of novel discoveries across all areas of biology. Tools of high-throughput tissue-specific gene expression profiling provide us with means to test our models in gene regulation that have been built in functional genomic studies. With further development in gene expression profiling techniques, we will be able to understand gene regulation in a more fundamental perspective.

Conflicting interests

The authors have declared that no conflict of interests exist.

Abbreviation

FISH: Fluorescence in situ hybridization; SAGE: serial analysis of gene expression; ChIP-seq: chromatin immunoprecipitation sequencing; ATAC-seq: assay for transposase-accessible chromatin with high throughput sequencing; FACS: fluorescence-activated cell sorting; GFP: green fluorescent protein; SAM: shoot apical meristem; INTACT: Nuclei Tagged in A specific Cell Type; NTF: nuclear tagging fusion; BLRP: biotin ligase recognition peptide; RIP: RNA immunoprecipitation; PAB: poly(A)-binding protein; ASEL/ASER: amphid sensilla left/right; modENCODE: model organism ENcyclopedia Of DNA Element; lncRNA: long non-coding RNA; TRAP: translating ribosome affinity purification; EGFP: enhanced green fluorescent protein; UPRT: uracil phosphoribosyltransferase; 4tU: 4-thiouracil; SL: spliced leader.

References

1. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, *et al*. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res 2013; 41:D996-D1008.

2. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, *et al*. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014; 515:568-571.

3. Murray JI, Boyle TJ, Preston E, Vafeados D, Mericle B, Weisdepp P, *et al*. Multidimensional regulation of gene expression in the C. elegans embryo. Genome Res 2012; 22:1282-1294.

4. Liu X, Long F, Peng H, Aerni SJ, Jiang M, Sánchez-Blanco A, *et al*. Analysis of cell fate from single-cell gene expression profiles in C. elegans. Cell 2009; 139:623-633.

5. Wang C, Han B, Zhou R, Zhuang X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. Cell 2016; 165:990-1001.

6. Ishida Y, Takabatake T, Kakinuma S, Doi K, Yamauchi K, Kaminishi M, *et al*. Genomic and gene expression signatures of radiation in medulloblastomas after low-dose irradiation in Ptch1 heterozygous mice. Carcinogenesis 2010; 31:1694-1701.

7. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 2009; 457:1038-1042.

8. modENCODE Consortium, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, *et al*. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 2010; 330:1787-1797.

9. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57-74.

10. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, *et al*. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 2012; 22:1813-1831.

11. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, *et al*. The accessible chromatin landscape of the human genome. Nature 2012; 489:75-82.

12. Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell 2016; 165:1375-1388.

13. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol Ed Frederick M Ausubel Al 2015; 109:21.29.1-9.

14. Factor DC, Najm FJ, Tesar PJ. Generation and characterization of epiblast stem cells from blastocyst-stage mouse embryos. Methods Mol Biol Clifton NJ 2013; 1074:1-13.

15. Volovitz I, Shapira N, Ezer H, Gafni A, Lustgarten M, Alter T, *et al*. A non-aggressive, highly efficient, enzymatic method for dissociation of human brain-tumors and brain-tissues to viable single-cells. BMC Neurosci 2016; 17:30.

16. Sanna PP, Repunte-Canonigo V, Guidotti A. Gene profiling of laser-microdissected brain regions and individual cells in drug abuse and schizophrenia research. Methods Mol Biol Clifton NJ 2012; 829:541-550.

17. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet 2013; 14:618-630.

18. Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, Macaulay IC, *et al*. Resolving early mesoderm diversification through single-cell expression profiling. Nature 2016; 535:289-293.

19. Zhang Y, Ma C, Delohery T, Nasipak B, Foat BC, Bounoutas A, *et al*. Identification of genes expressed in C. elegans touch

receptor neurons. Nature 2002; 418:331-335.

20. Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 2007; 17:69-73.

21. Christensen M, Estevez A, Yin X, Fox R, Morrison R, McDonnell M, *et al*. A primary culture system for functional analysis of C. elegans neurons and muscle cells. Neuron 2002; 33:503-514.

22. Deal RB, Henikoff S. The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc 2011; 6:56-68.

23. Steiner FA, Talbert PB, Kasinathan S, Deal RB, Henikoff S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Res 2012; 22:766-777.

24. Henry GL, Davis FP, Picard S, Eddy SR. Cell type-specific genomics of Drosophila neurons. Nucleic Acids Res 2012; 40:9691-9704.

25. Gilbert C, Svejstrup JQ. RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol Ed Frederick M Ausubel Al 2006; Chapter 27:Unit 27.4.

26. Roy PJ, Stuart JM, Lund J, Kim SK. Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature 2002; 418:975-979.

27. Takayama J, Faumont S, Kunitomo H, Lockery SR, Iino Y. Single-cell transcriptional analysis of taste sensory neuron pair in Caenorhabditis elegans. Nucleic Acids Res 2010; 38:131-142.

28. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, *et al*. A spatial and temporal map of C. elegans gene expression. Genome Res 2011; 21:325-341.

29. Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, *et al*. A translational profiling approach for the molecular characterization of CNS cell types. Cell 2008; 135:738-748.

30. Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, *et al*. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 2008; 135:749-762.

31. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol 2005; 23:232-237.

32. Miller MR, Robinson KJ, Cleary MD, Doe CQ. TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat Methods 2009; 6:439-441.

33. Gay L, Miller MR, Ventura PB, Devasthali V, Vue Z, Thompson HL, *et al*. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. Genes Dev 2013; 27:98-115.

34. Ma X, Zhan G, Sleumer MC, Chen S, Liu W, Zhang MQ, *et al*. Analysis of C. elegans muscle transcriptome using trans-splicing-based RNA tagging (SRT). Nucleic Acids Res Published Online First: 23 August 2016. doi:10.1093/nar/gkw734