

EDITORIAL

Open Access

# Single-cell transcriptome in the identification of disease biomarkers: opportunities and challenges

Zhitu Zhu<sup>1\*</sup>, Diane C Wang<sup>2</sup>, Laurențiu M Popescu<sup>3</sup> and Xiangdong Wang<sup>4\*</sup>

## Abstract

Single cell transcriptome defined as the entire RNA or polyadenylated products of RNA polymerase II on a cell can describe the gene regulation networks responsible for physiological functions, behaviours, and phenotypes in response to signals and microenvironmental changes. Single cell transcriptome/sequencing has the special power to investigate small groups of differentiating cells, circulating tumour cells, or tissue stem cells. A large number of factors may influence the extent of single-cell heterogeneity within a system. It is the opportunity that the single-cell sequencing can be used for the identification of genetic changes in rare cells, e.g. cancer and tissue stem cells, in clinical samples. The methodologies of single-cell sequencing have been improved and developed with the increase of the understanding and attention. The clinical research and application of the single cell sequencing analysis are expected to identify and validate disease-specific biomarkers, network biomarkers, dynamic network biomarkers. The single cell research and value will be also dependent upon the understanding of genomic heterogeneity, planning and design of study protocol, representative of selected and targeted cells, and sensitivity and repeatability of the methodology. The single cell sequencing can be used to develop new diagnostics, monitor disease progresses, measure responses to therapies, and predict the prognosis of patients, although there are still a large number of challenges and difficulties to be faced. It would be more values and specificities of the single cell sequencing to integrate with the function of cells, organs, and systems of the body, the clinical phenotypes of patients, and the description of clinical bioinformatics.

**Keywords:** Single cell, Sequencing, Biomarkers, Heterogeneity, Disease

Single cell transcriptome defined as the entire RNA or polyadenylated products of RNA polymerase II on a cell can describe the gene regulation networks responsible for physiological functions, behaviours, and phenotypes in response to signals and microenvironmental changes. Methods of single cell transcriptome analyses have been developed and significantly improved with advantages and disadvantages [1,2]. Single cell transcriptome can demonstrate the heterogeneity of gene expressions, interactions and regulations of gene regulatory networks, subpopulations within a tumour, characteristics of putative cancer stem cells, gene expression profiles of intracellular compartments, mRNA locations, allele specific gene expression, or the information combined

with the existing strand-specific cDNA library preparation strategies. With next generation sequencing, the information on both full genome and transcriptome can be obtained from an individual cell. Single-cell genome analysis becomes more important in the understanding of the diversity in microbial ecology, cancer development, prenatal genetic diagnosis, or human genome structures.

Single cell transcriptome/sequencing has the special power to investigate small groups of differentiating cells, circulating tumour cells, or tissue stem cells. Tang et al. defined the dynamic molecular changes with cell-fate changes during the conversion from the inner cell mass cells of blastocysts to pluripotent embryonic stem cells using RNA-Seq transcriptome analysis at the resolution of single cells [3]. Molecular mechanisms by which genetic elements could be switched depend upon the pluripotency, changes in epigenetic regulators, the stability of the newly acquired epigenotype, or self-renewal. It should be an excellent example to further study the regulation and differentiation of small numbers of stem cells in adults,

\* Correspondence: zhuzhitu@163.com; xiangdong.wang@clintransmed.org

<sup>1</sup>Department of Oncology, Liaoning Medical University Hospital, Jinzhou, China

<sup>4</sup>Fudan University Center of Clinical Bioinformatics, Shanghai Respiratory Research Institute, Fudan University School of Medicine, Zhongshan Hospital, Shanghai, China

Full list of author information is available at the end of the article

genes for general metabolism, expressions of repressive epigenetic regulators, or changes in microRNAs, to identify targeted genes of pluripotency. Furthermore, single cell transcriptome was used to measure the frequency of large-scale genome instability and de novo mutation rates with distinct characteristics [4].

A large number of factors may influence the extent of single-cell heterogeneity within a system, e.g. the source of analyzed tissues, biological conditions or microenvironments of the individual cell. The single cell transcriptome was suggested to avoid, or at least reduce, the variability of clinical samples, in order to define new cell classifications, transitional states, biological distinctions, or biomarkers. Shalek et al. recently demonstrated that the inflammatory stimulus lipopolysaccharide could induce an extensive, previously unobserved, bimodal variation in mRNA abundance and splicing patterns in bone marrow-derived dendritic cells using single-cell RNA-Seq analysis [5]. It indicates that stimuli from autocrine-based or external sources may initiate the heterogeneity of the cell, different from the traditional understanding. It is possible that we can identify and discover more sensitive and specific elements responsible for the development of cell heterogeneity as biomarkers to monitor dynamic changes and interactions of gene regulation networks or develop new molecular or cellular therapies. The multiplexed primer design can be introduced for selected targeted genes to lower the pre-amplification primer concentration and primer dimer signals, increase the feasibility of measurements, or make the operation easier for classification of cell type, dissection of heterogeneity, mapping of cellular hierarchy, and computational construction of genetic networks [6].

It is the opportunity that the single-cell sequencing can be used for the identification of genetic changes in rare cells, e.g. cancer and tissue stem cells, in clinical samples which may contain hundred cells in body fluids and biopsy with fine-needle. It seems that one of clinical applications of single-cell sequencing should be to understand the genomic profiling of copy number or sequence mutations in rare cells. The challenge in the identification and validation of therapeutic targets is to define the cell heterogeneity, especially to differ between cancer and pre-cancer cells, activated and non-activated inflammatory cells, or functional and non-functional cells, between sampling micro-regions, between cells at various stages, or between responses to therapy. For example, our previous studies on genetic comparison of mouse lung telocytes with mesenchymal stem cells and fibroblasts demonstrated that about 10-20% of genes expressed differently between 5 and 10 day cell culture of lung telocytes per se [7]. It indicates that gene expressions, phenotypes, or functional networks can be varied in the differentiation process or developmental duration. Single cell sequencing has been widely considered as the

powerful and helpful tool to identify disease-specific biomarkers to diagnose and monitor the occurrence and progress of the disease, as well as the responses of patients to the treatment. It has been emphasized that disease-specific biomarkers, network biomarkers, and dynamic network biomarkers can monitor and predict the responses to therapies, prognoses of patients, and reoccurrences of the disease, by integrating genomic/proteomic profiles with clinical phenotypes [8-12]. It is a common challenge for those methodologies including single cell sequencing to define the existence of discovered genes before the occurrence of the disease, before clinical diagnosis, or before the application of therapies. It should be also considered whether the same type of the single cell can be collected for measuring the genomic heterogeneity and frequency of resistant clones in the primary disease before and after a single therapy or multiple manipulations, or during the dynamic follow-up in patients. It is highly expected to have the single cell sequencing study with a clear clinical design and definite disease-specific or associated groups, including similar diagnoses, severities, stages, genetic backgrounds, pathogeneses, or therapies. One of critical parts to apply single-cell sequencing for disease biomarkers should be to integrate clinical experience and knowledge with bioinformatics from the single cell sequencing and ensure the understanding of diseases.

The methodologies of single-cell sequencing have been improved and developed with the increase of the understanding and attention. The single cell sequencing is expected to detect the genomic heterogeneity in the primary disease and provide useful information for selecting therapies, monitoring the efficacy and efficiency of the therapy, predicting the potential outcomes of patients, if the disease would be monogenomic one which hardly exists in clinical practice. It is more important to consider the interactions between genes, proteins, cells, organs, as well as the critical role of the tissue micro-environment, when clinical studies are designed for the single cell sequencing analysis. Single-cell sequencing was evidenced to have the high sensitivity in the identification of rare chemoresistant clones in the primary cancer and to investigate questions of whether resistant clones were pre-existing in the primary cancer or appeared in response to therapies [13]. The challenge to translate single cell sequencing analysis from clinical research to clinical application is to standardize the study design and protocol, the study and operation process, data analysis and mining, and the application of clinical bioinformatics. The clinical application of single cell sequencing per se faces the difficulty of problem-solving or question-based study design, and defines the criteria of disease-specific diagnosis, monitoring, or prediction. On the other hand, the single cell sequencing can be used as the tool to screen and identify

disease-specific biomarkers which are further developed into a diagnostic panel. It is still difficult to develop the gene panel-based diagnostic kits after the identification, selection, validation, and optimization from the single cell sequencing, due to the unexpected barriers, e.g. for stability and repeatability of the expression, the representative of the targeted panel, or description of the measured results. It is also important to understand the dynamic existence of genomic variation, interaction between genes, and definition of gene scale, origin, rate, and nature.

Another potential or new hope is to develop new type cells for therapies by bioengineering or modifying targeted genes selected from the single cell sequencing. It believes that the construction of a phylogenetic tree of a human lifetime and the contribution of genetic heterogeneity to the organism can be defined and figured out by the single cell sequencing analysis of the different cells from endodermal, mesodermal, and ectodermal tissues from an individual. The sub-classification or new discovered populations of cell types can be found by gene expression profiles and epigenetic status, although the relationship between genome sequence, epigenetic status, and gene expression, and the functional capacity of the cell remain unclear [14]. The new methods have been suggested to expedite applications in basic genome research and provide novel approaches for clinical genetic diagnosis [15]. The paired-end sequence analysis of single-cell whole-genome amplification products was developed to explore the nature and pace of genome mutation, the genetic composition, or de novo mutations. Such exciting approach could detect DNA mutations, DNA copy number changes from allelic whole-genome amplification -amplification, and the break points and architecture of structural variants. We call the special attention and believe it is a time to further translate the concept of single-cell analysis, sequencing, or transcriptomics into clinical research and application for the identification and validation of disease-special biomarkers, on basis of recent emphases from a growing number of outstanding publications [16-19].

In conclusion, the clinical research and application of the single cell sequencing analysis are expected to identify and validate disease-specific biomarkers, network biomarkers, dynamic network biomarkers. The single cell research and value will be also dependent upon the understanding of genomic heterogeneity, planning and design of study protocol, representative of selected and targeted cells, and sensitivity and repeatability of the methodology. The single cell sequencing can be used develop new diagnostics, monitor disease progresses, measure responses to therapies, and predict the prognosis of patients, although there are still a large number of challenges and difficulties to be faced. It would be more values and specificities of the single cell sequencing to integrate with the function of cells, organs, and systems of the body,

the clinical phenotypes of patients, and the description of clinical bioinformatics.

#### Author details

<sup>1</sup>Department of Oncology, Liaoning Medical University Hospital, Jinzhou, China. <sup>2</sup>Department of Medicine, Royal Berkshire Hospital, Reading, UK. <sup>3</sup>Department of Morphological Sciences, Division of Cellular and Molecular Medicine, Faculty of Medicine, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. <sup>4</sup>Fudan University Center of Clinical Bioinformatics, Shanghai Respiratory Research Institute, Fudan University School of Medicine, Zhongshan Hospital, Shanghai, China.

Received: 17 June 2014 Accepted: 21 July 2014

Published: 12 August 2014

#### References

1. Tang F, Lao K, Surani MA: Development and applications of single-cell transcriptome analysis. *Nat Methods* 2011, **8**(4 Suppl):S6-S11.
2. Blainey PC, Quake SR: Dissecting genomic diversity, one cell at a time. *Nat Methods* 2014, **11**(1):19-21.
3. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, Lao K, Surani MA: Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 2010, **6**(5):468-478.
4. Wang J, Fan HC, Behr B, Quake SR: Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 2012, **150**(2):402-412.
5. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaubblomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H, Regev A: Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 2013, **498**(7453):236-240.
6. Guo G, Luc S, Marco E, Lin TW, Peng C, Kerenyi MA, Beyaz S, Kim W, Xu J, Das PP, Neff T, Zou K, Yuan GC, Orkin SH: Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* 2013, **13**(4):492-505.
7. Zheng Y, Zhang M, Qian M, Wang L, Cismasiu VB, Bai CX, Wang XD: Genetic comparison of mouse lung telocytes with mesenchymal stem cells and fibroblasts. *J Cell Mol Med* 2013, **4**:567-577.
8. Liu R, Wang XD, Aihara K, Chen L: Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers. *Med Res Rev* 2013, **17**. doi:10.1002/med.21293.
9. Chen H, Song ZJ, Qian MJ, Bai CX, Wang XD: Selection of disease-specific biomarkers by integrating inflammatory mediators with clinical informatics in AECOPD patients: a preliminary study. *J Cell Mol Med* 2012, **16**:1286-1297.
10. Wang XD: Role of clinical bioinformatics in the development of network-based Biomarkers. *J Clin Bioinforma* 2011, **1**(1):28.
11. Wang XD, Peer D, Petersen B: Molecular and Cellular Therapies: New challenges and opportunities. *Mol Cell Therapies* 2013, **1**:1.
12. López E, Wang XD, Madero L, López-Pascual J, Latterich M: Functional phosphoproteomic mass spectrometry-based approaches. *Clin Transl Med* 2012, **1**:20.
13. Navin NI, Hicks J: Future medical applications of single-cell sequencing in cancer. *Genome Med* 2011, **3**(5):31.
14. Macaulay IC, Voet T: Single cell genomics: advances and future perspectives. *PLoS Genet* 2014, **10**(1):e1004126.
15. Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML, Zamani Esteki M, Van der Aa N, Mateiu L, McBride DJ, Bignell GR, McLaren S, Teague J, Butler A, Raine K, Stebbings LA, Quail MA, D'Hooghe T, Moreau Y, Futreal PA, Stratton MR, Vermeesch JR, Campbell PJ: Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 2013, **41**(12):6119-6138.
16. Sandberg R: Entering the era of single-cell transcriptomics in biology and medicine. *Nat Methods* 2014, **11**:22-24.
17. Speicher MR: Single-cell analysis: toward the clinic. *Genome Med* 2013, **5**:74.
18. Shapiro E, Biezuner T, Linnarsson S: Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 2013, **14**(9):618-630.
19. van Loo P, Voet T: Single cell analysis of cancer genomes. *Curr Opin Genet Dev* 2014, **24**:82-91.

doi:10.1186/s12967-014-0212-3

Cite this article as: Zhu et al.: Single-cell transcriptome in the identification of disease biomarkers: opportunities and challenges. *Journal of Translational Medicine* 2014 **12**:212.