Division of Molecular Neurobiology, Department of Biochemistry and Biophysics

From single-cell transcriptomics to single-molecule counting

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet öfvertygligt förvaras i Gustaf Retzius (Cesar), Retzius väg 20, Karolinska Institutet, Solna Campus

Tisdag den 19 November, 2013, kl 09.00

av
Saiful Islam

Huvudhandledare:
Dr. Sten Linnarsson
Karolinska Institutet
Molecular Neurobiology Division
Department of Medical Biochemistry and Biophysics

Birehäuser:
Professor Ernest Arenas
Karolinska Institutet
Molecular Neurobiology Division
Department of Medical Biochemistry and Biophysics

Fakultetsopponent:
Professor Piero Carninci
Leader of the Functional Genomics Technology Team
RIKEN Omics Science Center
Japan

Betygsämne:
Professor Mats Nilsson
Stockholm University
Department of Biochemistry and Biophysics

Professor Ingemar Ernberg
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Dr. Björn Högberg
Karolinska Institutet
Department of Neuroscience

Stockholm 2013
ABSTRACT

RNA-sequencing (RNA-seq) technology has been progressing so fast in the last few years and made it possible to perform transcriptome analysis at single-cell level that was even unimaginable a few years before. Nowadays, the importance of gene expression analysis at the single-cell level is increasingly appreciated for the study of complex heterogeneous tissue. Also, in order to solve the obscure and no consensus definition of cell types, the single-cell gene expression analysis approach will be important. In this thesis we described a novel approach of single-cell gene expression profiling, called Single-cell Tagged Reverse Transcription (STRT). Here we can analyse 96 single cells at the same time by using a DNA barcode to tag individual single-cell. We analysed both 41 embryonic stem cells (ES) and 44 mouse embryonic fibroblast (MEF) cells in a 96-well PCR plate using STRT and successfully separated the distinct cell types from each other by cluster analysis using only the gene expression profile. We believe that the strategy to discriminate cell types using gene expression profiling will empower the unbiased discovery and analysis of heterogeneous cell populations in both normal and disease tissue. However, the low efficiency of cDNA synthesis (~ 10%) and PCR amplification bias of this first STRT version reduced throughput and also decreased the possibility to detect genes that were expressed at lower level.

To improve the efficiency of cDNA synthesis, we significantly improved the template switching (TS) mechanism – the mechanism to make secondary oligonucleotide from RNA during first strand cDNA synthesis. Since most of the present single-cell methods use TS mechanism, the base preference of terminal transferase activity was studied in detail. We found that an NGG motif at the 3´ end of the template switching oligonucleotide (TSO) will work better than a GGG motif.

To remove the amplification bias we introduced a molecular barcode, a 5 bp short random degenerate sequence, which entirely eliminate the amplification bias. As this short random sequence uniquely labels each single molecule, their exact number can be determined. By introducing a microfluidic sample preparation, Fluidigm C1, we ensure the quality of cells that undergo sample processing until sequencing. In this updated method, called STRT-C1, we used 96 ES cells and split it as two final libraries consisting of one single strand cDNA library and one double strand cDNA obtained by amplifying the first single strand cDNA. We also used and analysed spike-in RNA in the same experiment. Both the reproducibility (correlation coefficient >99.5 within and between sample) and efficiency of reverse transcription (~48%, 5 folds more than previous version of STRT) of the improved method are excellent. Correlation coefficient for the endogenous gene within and between samples is much better at the molecule level than at the read level, especially for the genes of low abundance.

By analysing only the spike-in RNAs, we found that technical noise is minimized and we can observe true biological noise for endogenous genes. Some biological noise is intrinsic at the single-cell level but most of the genes showed only low level of noise. However, we detected ~173 genes in ES cells which shows significant noise. Our data also revealed that biologically noisy genes have a significant function to give a resonant state of the embryonic stem cell.

In conclusion, the single-cell molecule counting method makes it possible to count molecules in single-cell accurately and without any bias. So we believe our STRT-C1 version of the single-cell method is a significant step forward for all expression analysis.

ISBN 978-91-7549-313-8