Combined sequencing of mRNA and DNA from human embryonic stem cells

Florian Mertesa,⁎, Heiner Kuhlb, Wasco Wrucc, Hans Lehracha, James Adjaye c,⁎

a Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
b Next Generation Sequencing Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany
c Institute for Stem Cell Research and Regenerative Medicine, Heinrich Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany

Abstract

Combined transcriptome and whole genome sequencing of the same ultra-low input sample down to single cells is a rapidly evolving approach for the analysis of rare cells. Besides stem cells, rare cells originating from tissues like tumor or biopsies, circulating tumor cells and cells from early embryonic development are under investigation. Herein we describe a universal method applicable for the analysis of minute amounts of sample material (150 to 200 cells) derived from sub-colony structures from human embryonic stem cells. The protocol comprises the combined isolation and separate amplification of poly(A) mRNA and whole genome DNA followed by next generation sequencing. Here we present a detailed description of the method developed and an overview of the results obtained for RNA and whole genome sequencing of human embryonic stem cells, sequencing data is available in the Gene Expression Omnibus (GEO) database under accession number GSE69471.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords:
Next generation sequencing
RNA and whole-genome sequencing
Ultra-low input sequencing
Single cell
Embryonic stem cells

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69471.

2. Experimental design, materials and methods

For RNA and DNA extraction, 150 to 200 cells were collected from undifferentiated colonies by mechanical fragmentation using a StemProEZPassage Disposable Stem Cell Passaging Tool (Invitrogen, cat# 23181-010). For amplified cDNA, a customized version of the μMACS SuperAmp Kit (Miltenyi Biotec) and for WGA DNA the REPLI-g Midi Kit (Qiagen) was used. An overview of the combined processing of RNA and DNA is depicted in Fig. 1.

2.1. Amplified cDNA

Oligo-dT magnetic micro beads were applied to the lysed cell suspension for mRNA binding and transferred to low volume flow-through columns located in a magnetic field. On column cDNA synthesis was performed by applying 20 μL of reaction mixture (2 μL 10× Reverse Transcriptase Buffer (Ambion), 0.5 mM dNTPs, 1 μg T4 Gene 32 Protein (NEB), 400 U M-MLV Reverse Transcriptase (Enzymatics), 20 U RNase Inhibitor (Ambion)) at 42 °C for 60 min. The cDNA together with magnetic beads was collected by centrifugation followed by 3'-tailing according to the manufacturer. PCR amplification was performed by the addition of 76.5 μL reaction mixture (14 μL 5× Phusion HF buffer (Finnzymes), 0.5 mM dNTPs, 60 μL resuspended μMACS SuperAmp PCR mix, 2 U PhusionTaq (Finnzymes)) with the following cycling conditions: 78 °C for 30 s, 95 °C for 1 min, [98 °C for 3 s, 64 °C for 30 s, 72 °C for 2 min] × 40 cycles, 72 °C for 5 min.
Fig. 1. Workflow for combined isolation and amplification of mRNA and whole-genome DNA derived from sample material. Adapted from Mertes et al. [1], BMC Genomics, CC BY 4.0.

Fig. 2. RNA-seq and whole genome sequencing coverage. (A) Read coverage across transcript length separated by overall transcript length ranges in base pairs (kb) from below 1 kb and up to 15 kb. (B) Read coverage of genomic DNA for individual chromosomes shown as Manhattan plot. Adapted from Mertes et al. [1], BMC Genomics, CC BY 4.0.
2.2. Whole genome amplified DNA

Genomic DNA was retained from the eluate of the first wash step of the amplified cDNA procedure. DNA was ethanol precipitated by the addition of 0.1 volume of 3 M sodium acetate solution and 5 μg glycogen (Ambion), the precipitated pellet was resuspended in 10 μL of Elution Buffer (Qiagen). WGA was performed for 16 h at 30 °C according to the manufacturer.

2.3. NGS library preparation

Multiplex libraries with insert size range of 150–300 base pairs were prepared according to the Illumina TruSeq DNA Sample Preparation Guide. Sample were pooled with a ratio of 3:1:1 (wgaDNA:mRNA1:mRNA2) for cluster generation. Paired-end sequencing with 100 base pairs was performed on a single lane of an Illumina HiSeq 2000 instrument.

2.4. Data analysis

RNA-seq and genomic data was mapped to the human genome with identical parameters by Tophat and bowtie respectively. Picard was used to estimate duplicate read counts. Sequencing coverage was calculated via IGVtools from (exon) aligned BAM files with transcript window size of 25 bp for RNA-seq and 10,000 bp for genome sequencing. Individual transcript coverage calculations were based on ENSEMBL V74 with exon unions of human genes for plus and minus strand separately. Transcript coverage was calculated for transcript size intervals of 0–1 kb, 1–2 kb, 2–3 kb, 3–4 kb, 4–5 kb and 5–15 kb based on 40 equally sized bins for each transcript. Results are summarized in Fig. 2.

3. Discussion

We describe here a data set for the combined analysis of transcriptome and genome sequencing data derived from minute amounts of human embryonic stem cells. With the presented method we could show, that the preparation of material in the sub-colony range can be analyzed on the RNA and DNA level in a single approach. The results show that the method is robust as well as sensitive.

A detailed discussion of the data presented, can be found in the research article “Combined ultra-low input mRNA and whole-genome sequencing of human embryonic stem cells” [1]. The sequencing data is available in the Gene Expression Omnibus (GEO) database under accession number GSE69471.

Acknowledgments

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no 115234, resources of which are composed of financial contribution from the European Union’s Seventh Framework Program (FP7/2007-2013) and EFPIA companies’ in kind contribution. JA acknowledges support from the EU FP7 project AgedBrainSYSBIO (Grant Agreement No 305299) (http://www.agedbrainsysbio.eu), the Duesseldorf School of Oncology (funded by the Comprehensive Cancer Center Dusseldorf/Deutsche Krebshilfe and the Medical Faculty HHU Dusseldorf).

Reference

[1] F. Mertes, B. Lichtner, H. Kuhl, M. Blattner, W. Wruck, B. Timmermann, H. Lehrach, J. Adjaye, Combined ultra-low input mRNA and whole-genome sequencing from human embryonic stem cells. BMC Genomics 16 (1) (2015 Nov 12) 925.