Mutation Research 772 (2015) 55–62

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles☆

Wenge Li[a], R. Brent Calder[a], Jessica C. Mar[b], Jan Vijg[a,*]

[a] Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, United States
[b] Department of Systems & Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461, United States

ARTICLE INFO

Article history:
Received 29 December 2014
Received in revised form 7 January 2015
Accepted 8 January 2015
Available online 16 January 2015

Keywords:
Single cell transcriptogenomics
ENU
Germline mutations
Somatic mutations
Next generation sequencing
Allelic exclusion

ABSTRACT

Recently, great progress has been made in single cell genomics and transcriptomics. Here, we present an integrative method, termed single-cell transcriptogenomics (SCTG), in which whole exome sequencing and RNA-seq is performed concurrently on single cells. This methodology enables one to track germline and somatic variants directly from the genome to the transcriptome in individual cells. Mouse embryonic fibroblasts were treated with the powerful mutagen ethylnitrosourea (ENU) and subjected to SCTG. Interestingly, while germline variants were found to be transcribed in an allelically balanced fashion, a significantly different pattern of allelic exclusion was observed for ENU-mutant variants. These results suggest that the adverse effects of induced mutations, in contrast to germline variants, may be mitigated by allelically biased transcription. They also illustrate how SCTG can be instrumental in the direct assessment of phenotypic consequences of genomic variants.

1. Introduction

One of the central goals of biology is to elucidate the relationship between genotype and phenotype [1,2]. Conventionally, DNA-directed RNA transcription is examined at the bulk level, tacitly assuming sample homogeneity. This is probably valid for germline variation, which is universally present throughout the organism, or somatic mutations in clonal lineages, but it does not hold for sporadic somatic mutations across tissues. Indeed, there is abundant evidence for extensive genetic mosaicism of somatic tissues [3–5], indicating the existence of profound cell-to-cell variation. Especially in functionally complex tissues this necessitates a single cell approach to dissect and examine the consequences of genomic heterogeneity.

With the advent of massively parallel sequencing (MPS), genome-wide mutation profiles can be obtained at single nucleotide resolution. MPS has been successfully used to identify *de novo* germline mutations across human generations [6,7]. However, in contrast to *germline* variants, randomly occurring, low-abundance *somatic* mutations are very difficult to detect. When a heterogeneous tissue or cell population is examined by MPS, an individual mutant allele is almost always outnumbered by wild type alleles rendering it indistinguishable from amplification and sequencing errors. We previously demonstrated that sporadic somatic mutations can be reliably detected at the single cell level [8]. In a diploid cell, after whole genome amplification (WGA) and MPS, an authentic mutation resulting in heterozygosity can be confidently detected by its consensus presence in ~50% of the sequencing reads, whereas amplification errors and sequencing artifacts can be effectively filtered out [8].

In principle, single cell analysis has the potential to directly assess the functional consequence of DNA sequence variants at the RNA level. Accumulating evidence indicates that DNA-directed RNA transcription has profound flexibility, as illustrated by the observation of allelically biased transcription [9] and RNA editing [10]. Therefore, an integrative assay which can simultaneously analyze the genome and transcriptome of the same single cell would be critically important to assess how specific patterns of DNA sequence variants affect transcriptional profiles.

In the present study, we report the concurrent genomic and transcriptomic analysis of the same single cell after treatment with the powerful mutagen ethylnitrosourea (ENU). This assay, termed "single-cell transcriptogenomics (SCTG)", allowed us to track germline and somatic variants directly from the genome to the transcriptome. The results indicated, not unexpectedly,

☆ *Data access*: Our original sequencing data have been submitted to Sequence Read Archive (SRA), with an accession code of SRP040646.

* Corresponding author at: 1301 Morris Park Avenue, Price Center 450, Bronx, NY 10461, United States. Tel.: +1 718 678 1151; fax: +1 718 678 1016.
E-mail address: jan.vijg@einstein.yu.edu (J. Vijg).

http://dx.doi.org/10.1016/j.mrfmmm.2015.01.002

independent transcription of alleles containing germline sequence variants. However, surprisingly, alleles containing ENU-induced somatic mutations were significantly less frequently transcribed. This transcriptional bias against ENU-mutated alleles suggests a new layer of maintaining genome integrity.

2. Materials and methods

2.1. Mutagenesis

Two groups of early passage, sub-confluent MEFs derived from different C57BL/6J parents were treated with 500 μg/ml (4.3 mM) ENU (Sigma) for 30 min and cultured for 72 h [8]. Single MEFs were collected and snap-frozen in dry ice as described [8,11].

2.2. SCTG

After cell lysis, the mRNAs of single frozen MEFs were selectively pulled down by biotinylated oligo-dT peptidyl nucleic acids and streptavidin-beads using a mTRAP midi kit (Active Motif). The streptavidin beads were washed three times with increasing stringency. Subsequently, the purified beads were subjected to whole transcriptome amplification [11,12]. The exogenous polyC/G tails were cleaved at a MmeI site embedded in the primers (Supplementary Table 1). The MmeI-digested WTA amplicons were subjected to RNA-seq. Single cell genomic DNAs were ethanol-precipitated from the RNA wash-off fractions in the presence of glycogen and excess tRNA [13]. After tRNA-removal by RNase A, whole genome amplifications were performed using a Qiagen REPLI-g kit [8]. After a locus dropout test, the WGA amplicons exhibiting balanced amplification were subjected to whole exome sequencing.

2.3. Data analysis

For the WES data, the pipeline of sequence alignment and variant calling has been described in detail previously [8]. For the RNA-seq data, Illumina sequencing data were first subjected to computational trimming of residual adapters, and then aligned to the mouse mm9 reference genome using gsnap [14], followed by HTSeq-count and DESeq [15]. Detected genomic and transcriptomic variants were verified by integrative genomics viewer (IGV) validated by Sanger sequencing and compared directly.

For full details, see the extended Materials and methods in the Supplementary Materials.

3. Results

3.1. SCTG: concurrent genomic and transcriptomic variant analysis of the same cell

Fig. 1 schematically depicts the SCTG procedure. Previously, we successfully used a RT-PCR-based, unbiased global mRNA amplification procedure to append a polyC and a polyG tail on the 5′ and 3′ end of cDNA, respectively, allowing PCR amplification using a homopolymer primer with high annealing temperature [11,12]. We adapted this procedure for SCTG (Fig. 1). Briefly, after cell lysis, mRNAs were selectively pulled down for whole transcriptome amplification (WTA). Subsequently, the polyC/G tails of cDNAs were trimmed via a MmeI site embedded in the PCR primers (Supplementary Table 1). The processed cDNAs were then subjected to RNA-seq (Supplementary Table 2). Meanwhile, the genomic DNA in the wash-off fractions after mRNA pull-down was precipitated [13] and subjected to WGA using multiple displacement amplification (MDA) [8]. We used a locus drop-out (LDO) test [8] to positively select the least biased MDA amplicons (Supplementary Table 3), which were subjected to whole exome sequencing (WES).

SCTG was performed in mouse embryonic fibroblasts (MEFs) treated with 500 μg/ml ENU [8]. This dose has virtually no effect on cell survival but increased somatic mutation frequency by 35-fold [8]. For technical replication, we performed SCTG in two groups of MEFs derived from two different C57BL/6J mice. MEF1 to 4 and MEF8 were from one mouse; MEF5 to 7 and MEF9 to 10 were from another mouse. Those two groups of MEFs were treated with ENU independently.

A total of 10 cells showed satisfactory WGA (Fig. 2A) with acceptable LDO profiles (Supplementary Table 3, Fig. 2B), as well as seemingly successful WTA (Fig. 2C, Supplementary Fig. 1). Those single-cell samples were subsequently subjected to WES

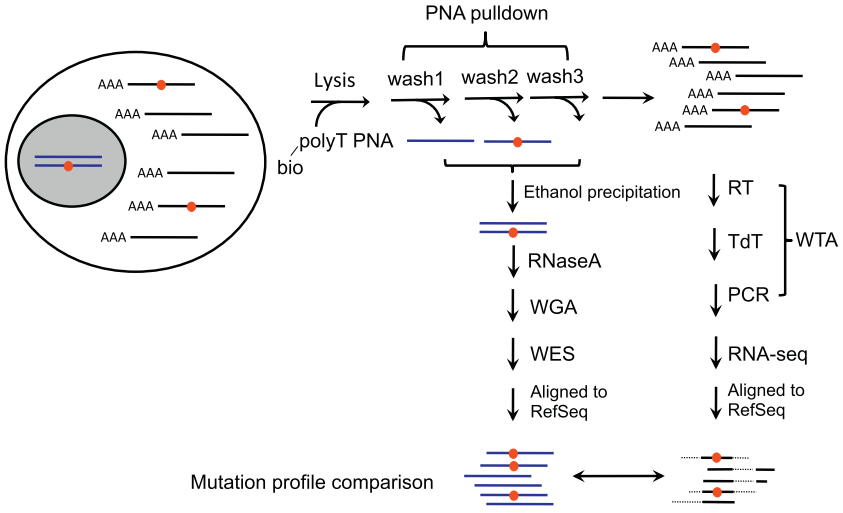

Fig. 1. Schematic illustration of single-cell transcriptogenomics (SCTG). After cell lysis, the polyA tailed (AAA) mRNAs (black lines) are selectively pulled down by oligo-dT peptidyl nucleic acid (PNA) beads. After reverse transcription (RT) and terminal deoxynucleotide transferation (TdT), global mRNA amplification is performed by PCR using polyC primers. The resultant cDNA products are subjected to RNA-seq. After alignment to the RefSeq, the cDNAs exhibit discontinued alignment patterns, with discrete exons (solid lines) linked by introns (dashed lines) that have been spliced out. The PNA wash-off fractions containing genomic DNAs (gDNAs, blue lines) are collected, and gDNAs are precipitated, pre-treated with RNaseA and subjected to whole genome amplification (WGA). The WGA products are subjected to whole exome sequencing (WES). The WES and RNA-seq data are subjected to mutation detection analysis. Direct comparison of the mutation profiles from gDNAs and cDNAs reveals the expression of a genomic mutation, such as a heterozygous point mutation (red dots).