Locus-specific DNA methylation analysis of retrotransposons in ES, somatic and cancer cells using High-Throughput Targeted Repeat Element Bisulfite Sequencing

Arundhati Bakshi  
*Louisiana State University*

Muhammad B. Ekram  
*Louisiana State University*

Joomyeong Kim  
*Louisiana State University*

Follow this and additional works at: [https://digitalcommons.lsu.edu/biosci_pubs](https://digitalcommons.lsu.edu/biosci_pubs)

**Recommended Citation**

Locus-specific DNA methylation analysis of retrotransposons in ES, somatic and cancer cells using High-Throughput Targeted Repeat Element Bisulfite Sequencing

Arundhati Bakshi, Muhammad B. Ekram, Joomyeong Kim

Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

A R T I C L E   I N F O

Article history:
Received 23 October 2014
Received in revised form 25 November 2014
Accepted 25 November 2014
Available online 29 November 2014

Keywords:
IAP LTR
DNA methylation
Retrotransposons
Endogenous retrovirus
HT-TREBS

A B S T R A C T

DNA methylation is a major epigenetic mark associated with multiple aspects of retrotransposons within the mammalian genome. In order to study DNA methylation of a large number of retrotransposons on an individual-locus basis, we have developed a new protocol termed High-Throughput Targeted Repeat Element Bisulfite Sequencing (HT-TREBS) (Ekram and Kim, 2014 [1]). We have used this technique to characterize the locus-specific patterns of DNA methylation of 4799 members of the mouse IAP LTR (Intracisternal A Particle Long Terminal Repeat) retrotransposon family in embryonic stem, somatic and Neuro2A cells (Bakshi and Kim, 2014 [2]). Here we describe in detail the sample preparation and bioinformatics analyses used for these studies. The somatic cell data may be accessed under GEO accession number GSE60007. The ES and Neuro2A data are deposited under GEO accession number GSE49222. The ES and Neuro2A data are


Experimental design, materials and methods

Sample preparation: DNA isolation and library construction

Whole brain, liver and kidney from a 1-week-old male C57BL6/N mouse were lysed overnight in 10X w/v Tail Lysis Buffer (100 mM Tris–Cl [pH 8.5], 5 mM EDTA [pH 8.0], 200 mM NaCl, 0.2% w/v sodium dodecylsulfate [SDS]) and 0.01X v/v 20 mg/mL Proteinase K (Sigma–Aldrich). The tissue-lysate products were divided into 300 μL aliquots and stored at −20°C. One 300 μL aliquot was used to isolate DNA using phenol–chloroform–isoamyl alcohol followed by ethanol precipitation, was imple-

direct link to deposited data

Raw and processed data for the studies described here may be found at the following links:


Corresponding author. Tel.: +1 225 578 7692; fax: +1 225 578 2597.
E-mail address: jkim@lsu.edu (J. Kim).

2213-5960/© 2014 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/3.0/).
DNA isolation

Sonication

End-repair and Ligation

Discard fragments <300 bp

Size selection, Quantitation and Sequencing

qPCR and PCR amplification

Bisulfite Conversion

<table>
<thead>
<tr>
<th>400 bp –&gt;</th>
<th>350 bp –&gt;</th>
<th>300 bp –&gt;</th>
<th>250 bp –&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extract fragments 330-380 bp</strong> (250-300 bp insert size)</td>
<td><strong>IAP LTR</strong></td>
<td>A</td>
<td>IAP LTR</td>
</tr>
</tbody>
</table>

Fig. 1. Library preparation for HT-TREBS. The isolated DNA is subjected to sonication to yield ~700 bp fragments and end-repaired before methylated-C Ion Torrent “A” adaptor ligation. Following one round of size selection, all fragments >300 bp are bisulfite treated and PCR amplified. The cycle number for PCR was determined for each individual library to be the one that corresponds to the midpoint of the exponential portion of the amplification curve from qPCR. Finally, the amplified library was size-selected for 250–300 bp insert size, quantified by Bioanalyzer and sequenced on the Ion Torrent PGM platform. The color coding within this figure is as follows: yellow bars indicate unique sequence, blue bars represent IAP LTRs, green and gray boxes indicate the Ion Torrent “A” and “P1” adaptors respectively and orange arrows represent the primers used for amplification.

“A” adapters (Integrated DNA Technologies) and 800 units of T4 DNA Ligase (New England Biolabs). Unligated adapters were removed using the DNA Clean and Concentration Kit (Zymo Research) with 5X v/v Binding Buffer and the resulting “A”-adapter-ligated fragments were eluted in 50 μL HPLC water. The adaptor-ligated DNA fragments were further size-selected to remove any excess adapters and DNA fragments smaller than 300 bp in length using the Agencourt AMPure XP beads (Beckman Coulter) using a DNA:bead ratio of 1:0.7 (100 μL DNA + 70 μL beads) (Fig. 1). This library was then quantified using Bioanalyzer (Agilent) and subsequently subjected to one round of bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research).

This bisulfite-converted library (1 μL) was then used to conduct quantitative real-time PCR using SYBR-Green (Life Technologies) to determine the appropriate cycle number to amplify the bisulfite-converted library for subsequent size selection and sequencing. The cycle number for PCR amplification was individually determined for each library to be the cycle number corresponding to the midpoint of the exponential portion of the amplification curve (Fig. 1). The forward primer used for PCR was complementary to the 5′ end of the “A” adaptor region (5′-CCATCTCT ATCCCCGTGTCTCCGACTCAG-3′). The reverse primer (5′-CCACTACGCC TCCGGTTCTCTATGGGCAGTCGGT GAT^CTCCCTAAAT AACTACAACCCATC-3′) was designed to bind the 24-bp region within the U3 region of the LTR which is conserved between five subtypes of IAP LTR (IAPLTR1, IAPLTR1a, IAPLTR2, IAPLTR2a, and IAPLTR2b) and is devoid of any CpG sites. The sequence at the 5′ end of the carat (^) corresponds to the “P1” adaptor which is a part of the amplification scheme used by Ion Torrent (Life Technologies) (Fig. 1). The amplified libraries were then size-selected using agarose gel extraction (MEGAquick-spin™ Total Fragment DNA Purification Kit, Intron Biotechnology) to have approximately 250–300 bp insert length flanked by the “A” and “P1” adaptors. This was achieved by excising the gel fragment corresponding to 330–380 bp in order to account for the ~80 bp combined length of the two adaptors (Fig. 1).

The size-selected library was then quantified using Bioanalyzer (Agilent Technologies). Approximately 25 μL of the size-selected library at 10 pM was used for the emulsion PCR and subsequent next-generation sequencing using the Ion Torrent Personal Genome Machine (PGM) Sequencer and Ion 318 chips (Ion Torrent, Life Technologies).

Bioinformatics analyses

We have implemented the following bioinformatics pipeline to process the raw sequence reads from the NGS platform (Fig. 2). We have used Bowtie2 [3] for mapping and BQ analyzer HT [4] for DNA methylation analyses. We have used several Unix-based command lines along with custom-made Perl scripts for the pipeline. First, we used the following steps for the construction of a custom database (below and Supplemental Material 1).

1) Download the sequence of 9282 IAP LTRs containing 330-bp LTR along with two 350-bp flanking regions in a fasta format from the Table database at UCSC genome browser.
2) Reverse complement half of the IAPLTRs in an opposite direction so that the entire set will have an identical forward direction with a custom Perl script.
3) Convert all the IAP LTRs into bisulfite-converted sequences with a custom Perl script.
4) Compile all the bisulfite-converted sequences into one large sequence with a Unix command line.
5) Construct a searchable index file with a large compiled sequence with bowtie2-build.

The execution of these steps will provide the following files: (i) a directory containing the 9282 files with the original sequences in a fasta format (Step 2) and (ii) a directory containing 6 indexable file (Step 5). These two directories will be used for the following analyses.
Second, we used the following steps for the mapping of sequence reads and subsequent DNA methylation level analyses (below and Supplemental Material 1).

1) Remove raw sequence reads smaller than 40 bp in length.
2) Map the raw reads against the custom-made database using Bowtie2 [3].
3) Remove headerlines from the sam file derived from mapping with a Unix command line.
4) Sort each line of the sam file based on IAP LTR locus and generate a directory containing each file with a custom Perl script.
5) Filter out raw reads from each IAP LTR file that are not qualified based on the insufficient overlap with the flanking unique region (greater than 10 bp in length) with a custom Perl script.
6) Execute BiQ Analyzer HT [4] with two directories, the directory containing all the sequences of IAP LTRs (Step 2 of database construction) and the directory containing all the mapped bisulfite sequence reads, with a custom Perl script.
7) Extract the necessary information (number of reads used, % methylation, standard deviation) from the output of BiQ Analyzer HT [4] with a custom Perl script.

The execution of these steps will finally derive one text file containing the DNA methylation level for the entire set of IAP LTRs, which can be imported into an excel file for further calculation and inspection (Fig. 2). This series of bioinformatics analyses require multiple Unix command lines and custom Perl scripts. We have provided these scripts and command lines as Supplemental Material 1.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2014.11.013.

Acknowledgments

We would like to thank Drs. Scott Herke and Mark Batzer for their help on NGS sequencing, and Isabel Lorenzo at Baylor College of Medicine for providing ES cells. We also thank Dr. Hana Kim, Corey Bretz, Wesley Frey and all other members of the Joo Kim Lab for their careful reading and discussion of the manuscript. This research was supported by the National Institute of Health (J.K. R01-GM066225 and R01-GM097074).

References