MethodsX 8 (2021) 101445

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

Method of targeted bisulfite massive parallel sequencing of the human LINE-1 retrotransposon promoter

Stanislav A. Vasilyev[a,*], Anton V. Markov[a], Oksana Yu. Vasilyeva[a], Ekaterina N. Tolmacheva[a], Lada A. Zatula[b], Diana V. Sharysh[a], Daria I. Zhigalina[a], Victoria V. Demeneva[a], Igor N. Lebedev[a]

[a] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*
[b] *Siberian State Medical University, Tomsk, Russia*

A B S T R A C T

The methylation index of the LINE-1 promoter is one of the most commonly used markers for assessing the global level of genome methylation in various human cells and tissues. We developed an NGS-based protocol for DNA methylation analysis of the LINE-1 retrotransposon promoter. This approach allows assessment of the DNA methylation index of 19 CpG sites in the LINE-1 promoter that have the highest tissue- or tumor-specific variability. The method provides a DNA methylation profile for analyzing either the methylation index of each CpG site independently or the mean DNA methylation index across the LINE-1 promoter. The results obtained using the developed method corresponded well to the level of methylation assessed using a commercially available kit for DNA pyrosequencing. In addition, our method provides much more information: 1) the DNA methylation profile of a significant part of the LINE-1 promoter and 2) the level of DNA methylation at individual LINE-1 loci in the genome. The method of targeted bisulfite massive parallel sequencing of the human LINE-1 retrotransposon promoter can be used in large-scale studies of the global level of genome methylation in normal human cells or tumors.

To accomplish this, we modified the targeted massive parallel sequencing method based on 16S Metagenomic Sequencing Library Preparation protocol (Illumina, USA) by:

- Introduction of the stage of bisulfite conversion of DNA.
- Development of specific primers for the LINE-1 sequence.

* Corresponding author.
E-mail address: stanislav.vasilyev@medgenetics.ru (S.A. Vasilyev).

https://doi.org/10.1016/j.mex.2021.101445

ARTICLE INFO
Method name: NGS-based LINE-1 retrotransposon methylation analysis
Keywords: LINE-1 retrotransposon, DNA methylation, Targeted bisulfite massive parallel sequencing
Article history: Received 7 May 2021; Accepted 5 July 2021; Available online 6 July 2021

Specifications table

Subject area:	Biochemistry, Genetics and Molecular Biology
More specific subject area:	Epigenetics
Method name:	NGS-based LINE-1 retrotransposon methylation analysis
Name and reference of original method:	16S Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System (Official Illumina Protocol)
Resource availability:	Reagents: • EZ DNA Methylation-Direct Kit (Zymo Research, USA) • 2x BioMaster HS-Taq PCR mix (Biolabmix, Russia) • 2x Phusion high-fidelity DNA polymerase mix (NEB, USA) • Nextera XT index kit (Illumina, USA) • AMPure XP beads (Beckman Coulter, USA) • Reagent Micro Kit v2 for MiSeq (300 cycles) (Illumina, USA) Equipment: • Qubit fluorometer (Thermo Fisher, USA) • Bioanalyzer 2100 instrument (Agilent Technologies, USA) • MiSeq (Illumina, USA)

Method details

Overview

The methylation index of the long interspersed nuclear element 1 (LINE-1) retrotransposon, which occupies approximately 17% of the human genome, is often used as a marker of the level of whole-genome methylation. Most often, LINE-1 methylation index analysis is performed by real-time quantitative methylation-specific PCR [6,11], methylation-sensitive high resolution melting analysis [9], or pyrosequencing 3-4 CpG sites in its promoter region [1,3,4,16]. We developed a targeted method for determining the index of methylation in the LINE-1 retrotransposon promoter using targeted bisulfite massive parallel sequencing. In our method, LINE-1 methylation is assessed based on the DNA methylation index of 19 CpG sites in the consensus sequence of the LINE-1 promoter that have the highest tissue- or tumor-specific variability [14]. Our method can potentially have a wide range of applications, from evaluating epigenetic abnormalities in various tissues and tumors to analyzing the epigenetic effects of mutagens and monitoring studies.

Targeted bisulfite massive parallel sequencing

Considering the bisulfite converted consensus sequence of the LINE-1 promoter (GenBank Accession: X58075.1), we designed primers to cover 19 CpG sites: forward 5'-TATTAGGGAGTGTTAGATAGTGGG-3' and reverse 5'-CCTCTAAACCAAATATAAAATATAATCTC-3' (Fig. 1a). These 19 CpG sites included the 3 CpG sites covered by the PyroMark LINE-1 kit (Qiagen, Hilden, Germany). Primers with attached Illumina adapters were synthesized, and the DNA product was amplified by PCR using a modified Illumina 16S rRNA protocol.

First, sodium bisulfite conversion of 1000 ng of total genomic DNA was performed using an EZ DNA Methylation-Direct Kit (Zymo Research, USA). First PCR amplifications were performed in 20 µl reaction mixtures. A master mix for each reaction was made using the 2x BioMaster HS-Taq PCR mix (Biolabmix, Russia), primers with attached Illumina adapters, and 60 ng of bisulfite-converted DNA. The cycling conditions were as follows: 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 57 °C for 1 min and 72 °C for 1 min. A final 7 min elongation step was performed at 72 °C.

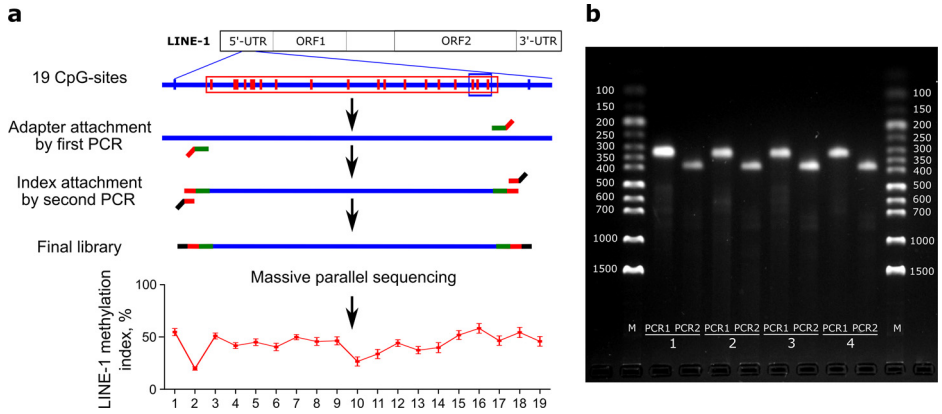

Fig. 1. Method of targeted bisulfite massive parallel sequencing of the human LINE-1 retrotransposon promoter. **a.** Schematic representation of the targeted bisulfite massive parallel sequencing of 19 CpG sites in the LINE-1 promoter. The red frame indicates the analyzed CpG sites in this study, and the blue frame indicates the most frequently analyzed CpG sites as determined using the Qiagen pyrosequencing LINE-1 kit (explanations are in the text). The schematic profile of the plot was obtained from the LINE-1 methylation index across the 5'-UTR region in analyzed samples. **b.** Gel image of the results of the amplified LINE-1 sequences for five samples after the first PCR with attached adapters and the same samples with attached indices after the second PCR. M represents Sky-High DNA marker (Biolabmix, Russia) with size in bp indicated.

A second PCR amplification, used to incorporate barcodes and sequencing adapters into the final PCR product, was performed in 25-microliter reaction mixtures using the 2x Phusion high-fidelity DNA polymerase mix (NEB, USA). As a matrix for the reaction, 2 µl (1/10 of the volume) of the unpurified product of the first PCR was used. A separate primer pair containing a unique 6-base barcode from the Nextera XT index kit was added to each well. The cycling conditions were as follows: 95 °C for 3 min, followed by 8 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. A final 5 min elongation step was performed at 72 °C. Products from both PCRs were verified by gel electrophoresis (Fig. 1b) with careful examination of the fragment sizes (PCR1 – 311 bp, PCR2 – 380 bp).

The quality and quantity of the amplified libraries were assessed by a Qubit fluorometer (Thermo Fisher, USA) and pooled together in equimolar quantities. Pooled library purified by AMPure XP beads (Beckman Coulter, USA), the size of fragments was assessed using a Bioanalyzer 2100 instrument (Agilent Technologies, USA), and the library was sequenced by MiSeq (Illumina, USA) using a Reagent Micro Kit v2 (300 cycles) in paired-end mode, with 30% PhiX added for clear cluster separation.

The short length of the LINE-1 amplified region (224 bp) made it possible to analyze more than 300 samples in a single MiSeq run, which was significantly more cost-effective than pyrosequencing.

NGS data processing

Sequenced reads were demultiplexed by standard Illumina i7 and i5 barcodes using the Genome Analysis Toolkit (GATK, v4.1.3.0) [10]. Read quality was assessed using FastQC v0.11.8, and residual adapter sequences and low-quality reads were removed with Trimmomatic v0.39 [2]. Read alignment was performed using bwa-meth v0.2.2 (https://github.com/brentp/bwa-meth) with default parameters. To avoid contamination in the library preparation step, we first mapped reads to the human genome assembly (February 2009 release, hg19/GRCh37) with masked repeat sequences, and then, unmapped reads (SAM flag “4”) were extracted using SAMtools v1.12 [8] and aligned to the consensus sequence of CpG-rich region of the LINE-1 element (GenBank Accession: X58075.1) [5] using bwa-meth. Human genome repeats were masked using BEDTools v2.26.0 [12] and RepeatMasker sequences [7], as described previously [15]. DNA methylation data in the CpG context were extracted from obtained BAM files using MethylDackel (https://github.com/dpryan79/MethylDackel). The results are presented as the methylation index, which is the ratio of the number of cytosines to the sum of cytosine and thymine in a given CpG site. The resulting methylation index characterizes the average level of

methylation of a part of the LINE-1 sequences in the genome that are amplified using the primers used and have specific CpG sites. Statistical analysis was carried out with Statistica v8.0 (StatSoft) and R statistical software (v. 3.6.1) using Pearson tests for the correlation analysis.

Method validation

The method was tested on DNA samples obtained from various human tissues: chorionic villous trophoblast cells from 42 induced and 193 spontaneous human abortions; tumor ($n = 2$) and normal tissues ($n = 2$) of the mammary gland; peripheral blood leukocytes ($n = 36$); tissues of the carotid artery ($n = 38$) and great saphenous vein ($n = 10$); and alpha-actin-2 positive (ACTA2+) cells ($n = 8$) and macrosialin-positive (CD68+) cells ($n = 6$) from carotid atherosclerotic lesions.

Commercially available control samples (in 3 replicates) from the Human Methylated & Nonmethylated DNA Set (Zymo Research, Irvine, CA, USA) were used to verify the validity of the assay.

Chorionic villous trophoblast cells were obtained by maceration of chorionic villi in 70% acetic acid. The isolated cells were washed in phosphate buffer. Genomic DNA was isolated from chorionic villous trophoblasts using the phenol-chloroform method. Artery wall tissue and whole blood samples were harvested from patients with atherosclerosis during carotid endarterectomy. ACTA2+ cells were derived from fresh frozen and immunostained samples of atherosclerotic lesions by laser-capture microdissection on a PALM MicroBeam (Carl Zeiss, Germany). DNA extraction and sodium bisulfite conversion were performed using an EZ DNA Methylation-Direct kit (Zymo Research, Irvine, CA, USA). Converted DNA from ACTA+ and CD68+ cells was further amplified using an EpiTect whole bisulfitome kit (Qiagen, Germany).

We were able to assess up to 346 samples in one sequencing run. The average output was 3333 (maximum 9076 and minimum 19) reads per demultiplexed library. Samples with more than 50% of short reads mapped to the human genome (13 out of 346) and those showing less coverage than 1000 reads per sample (85 out of 346) were excluded. Only 70% of the samples (242 of 346) reached the 1000x base coverage threshold to be considered for CpG methylation analysis [13].

The results of the LINE-1 promoter methylation index in hypomethylated (DNMT1$^{-/-}$ and DNMT3b$^{-/-}$ HCT cells) and methylated (MspI-treated HCT cells) control samples showed average LINE-1 methylation indices of 19 $\pm$ 6% and 65 $\pm$ 13%, respectively. One possible reason for the low LINE-1 methylation index in a methylated control sample is the spontaneous transition of CpG to TpG during evolution, which is one of the mechanisms for controlling the activity of retrotransposons. When analyzing the methylation index, all CpG sites where such a transition occurred is interpreted as having a zero methylation index. This may reduce the average observed methylation index. However, this method can be used to compare the methylation index between different samples and groups of samples, given similar initial evolutionary conditions for all individuals.

In addition to our method, the methylation index of the LINE-1 retrotransposon in 35 analyzed samples was determined by pyrosequencing using the PyroMark Q24 CpG LINE-1 (Qiagen, Germany) assay according to the manufacturer's protocol. The mean LINE-1 methylation index at 19 CpG sites assessed by targeted bisulfite massive parallel sequencing had a strong correlation with the mean LINE-1 methylation index at 3 CpG sites obtained using bisulfite pyrosequencing (Fig. 2a). The correlation between average methylation indices across identical CpG sites measured by different techniques was also high (Fig. 2b). However, the methylation index estimated by pyrosequencing did not correlate well with the methylation index for all individual CpG sites after targeted bisulfite massive parallel sequencing (Fig. 3). This finding indicates that our method provides not only the same information on average values as pyrosequencing but also additional information on individual CpG sites.

The developed method of targeted bisulfite massive parallel sequencing makes it possible to estimate the methylation index of the LINE-1 retrotransposon quickly and cost-effectively (Table 1). A possible scenario for using our method is to analyze large groups of heterogeneous samples (e.g., the analysis of the LINE-1 methylation index in tumors of different locations in a large number of patients). Our results indicate that despite the relatively good correlation between the LINE-1 methylation indices estimated by targeted bisulfite NGS and pyrosequencing, the correlations for a portion of the CpG sites are quite weak. This indicates that the methylation index of the three

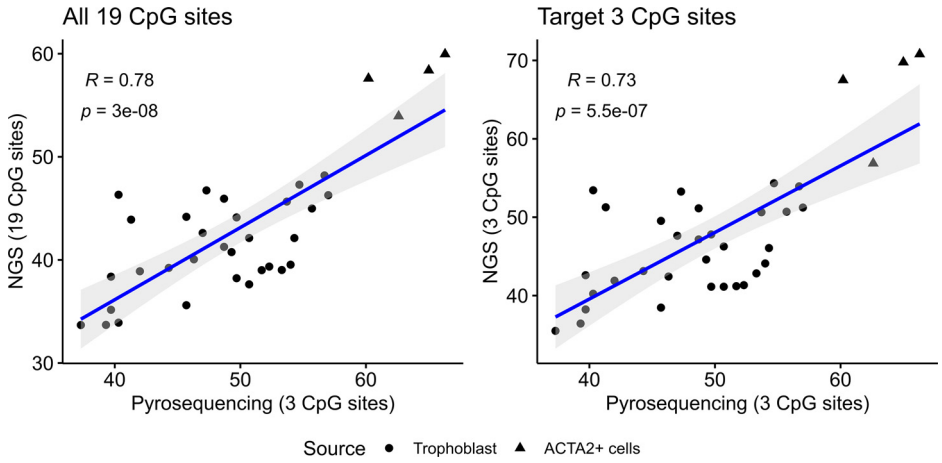

Fig. 2. Relationship between the average LINE-1 methylation index (%) measured by the introduced method (NGS) and bisulfite DNA pyrosequencing.

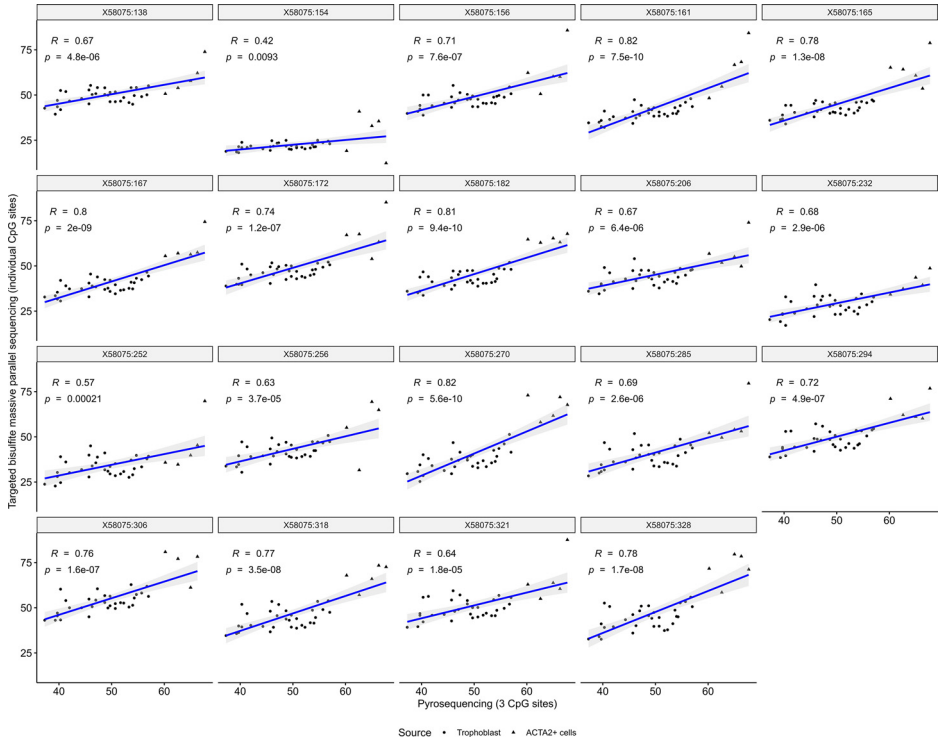

Fig. 3. Relationship between methylation indices of individual CpG sites as measured by the introduced method and average LINE-1 methylation index (%) as assessed by bisulfite DNA pyrosequencing (at positions 318, 321, and 328) trophoblast and ACTA2+ cells.

Table 1
Cost-effectiveness of the analysis of LINE-1 methylation index by targeted bisulfite NGS compared with pyrosequencing (steps after bisulfite conversion of DNA are compared only).

Pyrosequencing	Supplier	Price per 1 sample, $ (96 samples)
PyroMark Q24 CpG LINE-1	Qiagen	8.5
PyroMark Binding Buffer	Qiagen	0.01
PyroMark Denaturation Sol.	Qiagen	0.25
PyroMark Wash Buffer	Qiagen	0.25
PyroMark Annealing Buffer	Qiagen	0.01
Streptavidin Sepharose® High Performance	Sigma Aldrich	0.22
PyroMark Q24 Plate	Qiagen	0.11
PyroMark Gold Q24 Reagents	Qiagen	3.23
	Total	**12.58**

Targeted bisulfite NGS		300 samples	200 samples	100 samples
PCR 1	Biolabmix	0.21	0.21	0.21
PCR 2	NEB	2.04	2.04	2.04
AMPure XP	Beckman	0.01	0.01	0.02
Adapters	Illumina	0.35	0.35	0.35
Indexes	Illumina	0.46	0.46	0.46
Reagent Micro Kit v2 (300 cycles)	Illumina	3.23	4.85	9.69
	Total	**6.30**	**7.92**	**12.78**

CpG sites evaluated by pyrosequencing is insufficient to obtain complete information about the methylation profile of the LINE-1 promoter.

Thus, the results of our method correspond well to the methylation index assessed using the commercially available kit for pyrosequencing and provide much more information, allowing for the evaluation of the methylation profile in a large part of the LINE-1 promoter. In addition to estimating the average LINE-1 methylation index in the genome, this method potentially allows the estimation of the level of methylation in individual LINE-1 promoters at individual genome loci containing specific genetic variants. The introduced method of targeted bisulfite massive parallel sequencing of the human LINE-1 retrotransposon promoter can be used in large-scale studies of genome methylation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The molecular cytogenetic and molecular genetic studies were performed in the “Medical Genomics” Core Facility of the Tomsk National Research Medical Center of the Russian Academy of Sciences using the resources of the biocollection “Biobank of the population of northern Eurasia” of the Research Institute of Medical Genetics, Tomsk NRMC.

Funding

This study was supported by the Russian Science Foundation (project 19-74-10026).

Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the local Research Ethics Committee of the Research Institute of Medical Genetics, Tomsk NRMC (08.12.2020 / No 166).

Consent to participate

Written informed consent was obtained from the patients for their participation.

References

[1] A.B. Benitez-Trinidad, I.M. Medina-Diaz, Y.Y. Bernal-Hernandez, B.S. Barron-Vivanco, C.A. Gonzalez-Arias, J.F. Herrera-Moreno, I. Alvarado-Cruz, B. Quintanilla-Vega, A.E. Rojas-Garcia, Relationship between LINE-1 methylation pattern and pesticide exposure in urban sprayers, Food Chem. Toxicol. 113 (2018) 125–133.

[2] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.

[3] V. Bollati, S. Fabris, V. Pegoraro, D. Ronchetti, L. Mosca, G.L. Deliliers, V. Motta, P.A. Bertazzi, A. Baccarelli, A. Neri, Differential repetitive DNA methylation in multiple myeloma molecular subgroups, Carcinogenesis 30 (8) (2009) 1330–1335.

[4] Y.H. Cho, H.D. Woo, Y. Jang, V. Porter, S. Christensen, R.F. Hamilton Jr., H.W. Chung, The association of LINE-1 hypomethylation with age and centromere positive micronuclei in human lymphocytes, PLoS One 10 (7) (2015) e0133909.

[5] P.J. Crowther, J.P. Doherty, M.E. Linsenmeyer, M.R. Williamson, D.M. Woodcock, Revised genomic consensus for the hypermethylated CpG island region of the human L1 transposon and integration sites of full length L1 elements from recombinant clones made using methylation-tolerant host strains, Nucleic Acids Res. 19 (9) (1991) 2395–2401.

[6] C. Furlan, J. Polesel, L. Barzan, G. Franchin, S. Sulfaro, S. Romeo, F. Colizzi, A. Rizzo, V. Baggio, V. Giacomarra, A.P. Dei Tos, P. Boscolo-Rizzo, E. Vaccher, R. Dolcetti, L. Sigalotti, E. Fratta, Prognostic significance of LINE-1 hypomethylation in oropharyngeal squamous cell carcinoma, Clin. Epigenet. 9 (2017) 58.

[7] J. Jurka, Repbase update: a database and an electronic journal of repetitive elements, Trends Genet. 16 (9) (2000) 418–420.

[8] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, Genome Project Data Processing S, The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.

[9] J.L. Marques-Rocha, F.I. Milagro, M.L. Mansego, D.M. Mourao, J.A. Martinez, J. Bressan, LINE-1 methylation is positively associated with healthier lifestyle but inversely related to body fat mass in healthy young individuals, Epigenetics 11 (1) (2016) 49–60.

[10] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo, The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (2010) 1297–1303.

[11] A.A. Ponomaryova, E.Y. Rykova, T.L. Azhikina, A.A. Bondar, O.V. Cheremisina, E.O. Rodionov, V.V. Boyarko, P.P. Laktionov, N.V. Cherdyntseva, Long interspersed nuclear element-1 methylation status in the circulating DNA from blood of patients with malignant and chronic inflammatory lung diseases, Eur. J. Cancer Prev. 30 (2) (2021) 127–131.

[12] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (2010) 841–842.

[13] S. Roeh, T. Wiechmann, S. Sauer, M. Kodel, E.B. Binder, N. Provencal, HAM-TBS: high-accuracy methylation measurements via targeted bisulfite sequencing, Epigenet. Chromatin 11 (1) (2018) 39.

[14] A. Sharma, M.A. Jamil, N. Nuesgen, A. Dauksa, A. Gulbinas, W.A. Schulz, J. Oldenburg, O. El-Maarri, Detailed methylation map of LINE-1 5′-promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity, Mol. Genet. Genomic Med. 7 (5) (2019) e601.

[15] Z. Tang, J.P. Steranka, S. Ma, M. Grivainis, N. Rodic, C.R. Huang, I.M. Shih, T.L. Wang, J.D. Boeke, D. Fenyo, K.H. Burns, Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer, Proc. Natl. Acad. Sci. USA 114 (5) (2017) E733–E740.

[16] S.A. Vasilyev, E.N. Tolmacheva, A.A. Kashevarova, E.A. Sazhenova, I.N. Lebedev, Methylation status of LINE-1 retrotransposon in chromosomal mosaicism during early stages of human embryonic development, Mole. Biol. 49 (1) (2015) 144–152.