Gene xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Gene

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

Methodological paper

DNA methylation detection at single base resolution using targeted next generation bisulfite sequencing and cross validation using capillary sequencing

Smitha Bhat [a], Sandeep Mallya [a], Vinay Koshy Varghese [a], Pradyumna Jayaram [a], Sanjiban Chakrabarty [a], Kalpana S. Joshi [b], Tanuja M. Nesari [c], Kapaettu Satyamoorthy [a,*]

[a] Department of Biotechnology, School of Life Sciences, Manipal University, Manipal 576104, Karnataka, India
[b] Department of Pulmonary Medicine, Chest Research Foundation, Pune, Maharashtra, India
[c] Department of Dravyaguna, Tilak Ayurved Mahavidyalaya, Pune, Maharashtra, India

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form 27 August 2016
Accepted 12 September 2016
Available online xxxx

Keywords:
DNA methylation
CpG islands
Biomarker
Next generation sequencing (NGS)
Gene discovery

ABSTRACT

With a purpose of accurate and simultaneous determination of DNA methylation from multiple loci in multiple samples, here, we are demonstrating a method to aid rapid DNA methylation detection of genomic sequences. Using genomic DNA of peripheral blood from 14 healthy individuals, DNA methylation in 465 CpG sites from 12 loci of genes (*ADAM22*, *ATF2*, *BCR*, CD83, *CREBBP*, *IL12B*, *IL17RA*, *MAP2K2*, *RBM38*, *TGFBR2*, *TGFBR3*, and *WNT5A*) was analysed by targeted next generation bisulfite sequencing. Analysed region for three genes, *BCR*, *IL17RA* and *RBM38* showed an absolute mean DNA methylation of 25.6%, 89.2% and 38.9% respectively. Other nine gene loci were unmethylated and exhibited <10% absolute mean DNA methylation. Two genes, *IL17RA* and *RBM38* were technically validated using direct capillary sequencing and results were comparable with positive correlation ($P = 0.0088$ & $P < 0.0001$ respectively) in the CpG sites for DNA methylation. All CpG sites analysed from *RBM38* genes locus displayed 95% limits of agreement for DNA methylation measurements from the two methods. The present approach provides a fast and reliable DNA methylation quantitative data at single base resolution with good coverage of the CpG sites under analysis in multiple loci and samples simultaneously. Use of targeted next generation bisulfite sequencing may provide an opportunity to explore genes in the discovery panel for biomarker identification and facilitate functional validation.

1. Introduction

Addition of methyl group to cytosine residues as an epigenetic mark may lead to regulated gene expression without any alterations in the underlying DNA sequence (Moore et al., 2013; Lorincz et al., 2002). There are about 28 million CpG sites in the mammalian genome and 60–80% of these are methylated (Edwards et al., 2010; Smith and Meissner, 2013). Unmethylated CpGs are randomly distributed, but usually enriched in the form of clusters called CpG islands (Deaton and Bird, 2011). These clusters are prevalent near the gene promoter in the genome and >50% of the genes initiate transcription in the region with CpG islands (Vavouri and Lehner, 2012). Varied pattern of DNA methylation in human cell types has been reported, making it an important cellular mark for differential expression of genes related to the cell development and differentiation (Meissner et al., 2008; Schneider et al., 2010; Varley et al., 2013; Reinius et al., 2012). Today, there are variety of techniques to quantify DNA methylation based on absolute/relative measurements, targeted/non targeted enrichment, array detection and whole genome analysis (Plongthongkum et al., 2014).

Extensive research and data generation from candidate gene approaches (Ullah et al., 2015; Lin et al., 2015), encyclopedia of DNA elements (ENCODE) consortium (ENCODE Project Consortium, 2012) and genome-wide DNA methylation profiling of diseased and healthy individuals in variety of cell types and tissues (Exner et al., 2015; Glossop et al., 2016; Fernandez-Santiago et al., 2015; Yang et al., 2009), identified genes that can be potential biomarkers for the discovery of prognostics, diagnostics and therapeutics targets. Identified differentially methylated regions from different gene loci with multiple CpG sites is validated and replicated in large samples for its sensitivity and specificity before the clinical use of a biomarker in disease prediction/progression or therapeutic intervention (Tetzner et al., 2009; Van Neste et al., 2012; Boers et al., 2016). Difference in DNA methylation of genes is

Abbreviations: COBRA, Combined Bisulfite Restriction Analysis; DCS, direct capillary sequencing; ENCODE, encyclopedia of DNA elements; ESME, epigenetic sequencing methylation; IGV, integrated genome viewer; ISP, Ion spear Particles; MSRE, methylation sensitive restriction enzyme; PGM, Personal Genome Machine; RT-PCR, real-time polymerase chain reaction; TNBS, targeted next generation bisulfite sequencing.

* Corresponding author at: School of Life Sciences, Manipal University, Planetarium Complex, Manipal 576104, India.
E-mail address: ksatyamoorthy@manipal.edu (K. Satyamoorthy).

http://dx.doi.org/10.1016/j.gene.2016.09.019
0378-1119/

<5% in noncommunicable diseases when compared to the cancer related genes (sometimes >50%) in case control studies (Mikeska and Craig, 2014). Accurate detection of DNA methylation difference and quantitation at such a low resolution with good specificity and sensitivity is very important for biomarker discovery in different diseases. Recently, studies have reported various approaches for targeted bisulfite next generation sequencing of genes in the human genome at single base resolution to observe the difference in DNA methylation. These approaches provide high resolution detection and are fast, highly sensitive, and cost effective due to multiplexing and exhibit broad application in clinical and research (Vaca-Paniagua et al., 2015; Masser et al., 2013; Pietrzak et al., 2016).

In the present methodology paper, we employed targeted next generation bisulfite sequencing (TNBS) using the Ion Torrent Personal Genome Machine (PGM) Sequencer to quantify and analyse specific target regions (465 CpG sites) of 12 genes in human genome amplified using bisulfite converted DNA from peripheral blood of healthy subjects. To evaluate efficacy, direct capillary sequencing (DCS) of the PCR product from bisulfite converted DNA was performed and validated for exact quantification of DNA methylation in the CpG sites of two genes in portion of samples used for targeted NGS. DNA methylation detection approach presented here is similar to the published NGS based deep bisulfite sequencing method (Lee and Kim, 2016). We have compared and discussed the experimental, bioinformatics similarities and the modifications used in our approach to the method by Lee and Kim (2016).

2. Materials and methods

2.1. Genomic DNA bisulfite conversion for PCR and direct sequencing

To carry out TNBS and DCS, 14 healthy individuals with the mean age of 38.07 (M/F; 6/8) were recruited. All the study subjects provided written informed consent before participation; the study was approved by the institutional ethics committee of Tilak Ayurveda Mahavidhyalaya and Sheth Tarachand Ramnath Charitable Ayurveda Hospital, Pune, India. Genomic DNA from peripheral blood of all the healthy individuals recruited, was extracted using Flexigene-DNA isolation kit (Qiagen, Hilden, Germany), following manufacturer's instruction. Bisulfite specific primers for 12 gene loci were designed using online available tool Methyl Primer Express Software (version 1.0) (Li and Dahiya, 2002) with default parameters (Table 1). The schematic representation of 12 gene loci with the location of CpG sites is shown in Additional Fig. 1. One microgram of genomic DNA was used for bisulfite conversion by EZ DNA Methylation™ Kit (Zymo research, CA, USA) following manufacture's protocol and bisulfite specific PCR was performed. The PCR conditions used: (1) Initial denaturation at 95 °C for 5 min, (2) denaturation at 95 °C for 30 s, (3) annealing (Table 1) for 1 min, (4) extension at 72 °C for 1 min, 30 s, 35 cycles (steps 2 to 4), (5) final extension at 72 °C for 10 min, and (6) and 4 °C hold using Veriti® Thermal Cycler (Thermo Fisher Scientific, Waltham, USA). To verify the status of amplification, PCR products were visualized on 1.2% agarose gel under UV light and further purified in the case of non-specific amplification.

2.2. Next generation sequencing of PCR product from bisulfite converted DNA

Concentration (nM) of gene amplicons was estimated using Qubit™ 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA) and gene amplicons were pooled to a final concentration of 25 nM per sample. DNA concentration (ng/μl) of the pooled PCR products was determined and sample pools (120 ng each) were barcoded to generate barcoded library using Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher Scientific, Waltham, USA), following manufacturer's instructions. All steps were performed in 1.5 ml Eppendorf LoBind Tubes (Eppendorf, Hamburg, Germany) whenever required. Fig. 1 shows schematic representation of targeted next generation bisulfite sequencing. Briefly, 120 ng of pooled PCR product was sheared using the Ion shear™ Plus Kit and purified using Agencourt AMPure XP Beads [(1.8× sample volume) (Beckman Coulter, CA, USA)]. Purified sheared DNA was subjected to end repair, ligation of Ion Xpress Barcode and Ion P1 adapters and nick-repair as per manufacturer's protocol using Ion Xpress™ Plus Fragment Library Kit. The barcoded adapter-ligated library DNA was purified by Agencourt AMPure XP Beads (1.2× sample volume) and was subjected to size-selection between 200 bp and 300 bp using E-Gel® 2% SizeSelect™ Agarose Gels (Thermo Fisher Scientific, Waltham, USA). Size selected, unamplified library was amplified for 8 cycles using Ion Xpress™ Plus Fragment Library Kit in Veriti® Thermal Cycler (Thermo Fisher Scientific, Waltham, USA). Amplified library DNA was purified using Agencourt AMPure XP Beads (1.5× sample volume). Purified

Table 1
Primer information and PCR condition.

Gene symbol	Description	Forward & reverse primer sequence	PCR amplicon length (bp)	Genomic coordinate (Hg19)	Annealing temperature
ADAM22	ADAM metallopeptidase domain 22 (ADAM22)	F: ATGTAGATGAATATTGGTTGTTGT R: ACCTCATCCCCTAACAAAAAC	313	chr7:87,563,206–87,563,518	60 °C
ATF2	Activating transcription factor 2	F: AATTAGTGATTTGGAAAGGTGAG R:CCTTTTTCCTCTACCTTAAAAAAAC	311	chr2:176,032,485–176,032,795	56 °C
BCR	Breakpoint cluster region	F: YGTAAGGGTTATGGTTAGT R: AAAACRCTTCTCCTACTCAA	444	chr22:23,523,631–23,524,074	56 °C
CD83	CD83 molecule	F:TTTAAGTGGGATTAGGAGGG R:AAACAATACAAAACAAATCCACATCTT	376	chr6:14,117,881–14,118,256	56 °C
CREBBP	CREB binding protein	F:GTTTTGATTATTTTGGAGTAGTTGTT R:CTAAACCACAACTTTTATTCTAATAA	381	chr16:3,930,967–3,931,347	58 °C
IL12B	Interleukin 12B	F:GTYGGATAATTAGTGGGTTT R:CRTAAACAAAAACATATACCTACAC	439	chr5:158,758,451–158,758,889	58 °C
IL17RA	Interleukin 17 receptor A	F:GTTTGGGAGGGTTTTTTAGG R:CCCACCCCTTTACCTAATTCTA	338	chr22:17,564,888–17,565,225	62 °C
MAP2K2	Mitogen-activated protein kinase kinase 2	F:AAAGGYGGTTTTGTGTTGTTG R:AAACCTCCRACTAACCCCTACCCACTCACT	490	chr19:4,123,750–4,124,241	64 °C
RBM38	RNA binding motif protein 38-MINUS STRAND	F:ATAATTTTGAATTAGGGAGGTTAG R:AAAAAACCCACACCTACTTC	422	chr20:55,965,765–55,966,186	62 °C
TGFBR2	Transforming growth factor beta receptor II	F:TAGGAGTYGGATTTTTGTGTAG R:ACTCRCACAAAAAAAACCAAA	572	chr3:30,648,141–30,648,713	62 °C
TGFBR3	Transforming growth factor beta receptor II	F:TTTTTAGTGAGTGAAGGAGGGTAGT R:TCCCCAAAAAATAAACAC	421	chr1:92,351,249–92,351,670	62 °C
WNT5A	Wingless-type MMTV integration site family, member 5A	F:TAGAGATGTTTATATTAATT R:AATCRAAACRCAACTAAAAAAC	404	chr3:55,521,279–55,521,683	56 °C

Please cite this article as: Bhat, S., et al., DNA methylation detection at single base resolution using targeted next generation bisulfite sequencing and cross validation using..., Gene (2016), http://dx.doi.org/10.1016/j.gene.2016.09.019