Mitochondrion 27 (2016) 25–31

Contents lists available at ScienceDirect

Mitochondrion

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

Hydroxymethyl cytosine marks in the human mitochondrial genome are dynamic in nature

Sourav Ghosh [a,c], Shantanu Sengupta [a,c,*], Vinod Scaria [b,c,**]

[a] Genomics and Molecular Medicine, CSIR Institute of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India
[b] GN Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India
[c] Academy of Scientific and Innovative Research (AcSIR), CSIR-IGIB South Campus, Mathura Road, Delhi 110025, India

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 19 January 2016
Accepted 20 January 2016
Available online 28 January 2016

Keywords:
5-Hydroxymethylcytosine
Mitochondria
Epigenome

ABSTRACT

Apart from DNA methylation, hydroxymethylation has increasingly been studied as an important epigenetic mark. 5- hydroxymethylcytosines, though initially were thought to be an intermediary product of demethylation, recent studies suggest this to be a highly regulated process and modulated by the TET family of enzymes. Recent genome wide studies have shown that hydroxymethylcytosine marks are closely associated with the regulation of important biological processes like transcription and embryonic development. It is also known that aberrant hydroxymethylation marks have been associated with diseases like cancer. The presence of hydroxymethylcytosines in the mitochondrial genome has been earlier suggested, though the genome-scale map has not been laid out. In this present study, we have mapped and analyzed the hydroxymethylcytosine marks in the mitochondrial genome using 23 different publicly available datasets. We cross validated our data by checking for consistency across a subset of genomic regions previously annotated to hydroxymethylcytosines and show good consistency. We observe a dynamic distribution of hydroxymethylation marks in the mitochondrial genome. Unlike the methylcytosine marks, hydroxymethylcytosine marks are characterized by the lack of conservation across the samples considered, though similar cell types shared the pattern. We additionally observed that the hydroxymethylation marks are enriched in the upstream of GSS (gene start site) regions and in gene body as similar as nuclear genes. To the best of our knowledge, this is the first genome-scale map of hydroxymethyl cytosines in the human mitochondrial genome.

1. Introduction

Recent reports have suggested that the mitochondrial genome could be epigenetically modified, in a manner similar to the well characterized epigenetic modification of the nuclear genome (Manev et al., 2012). Additional evidence suggests that these epigenetic modifications could have an intricate role in regulation and functional outcomes in the cell (Iacobazzi et al., 2013). Presently available evidences also suggest the role of the mitochondrial epigenetic signatures in both physiological and pathophysiological conditions (Chandel, 2014; D'Aquila et al., 2015). In addition, mitochondrial epigenetic changes have also been explored as a major outcome of environmental interactions (Byun and Baccarelli, 2014). Methylation of cytosine bases in the DNA is one of the most popular and well studied epigenetic modification. Mechanistically, methylation of cytosine bases in the mitochondria is thought to be modulated through the DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3A and 3B (DNMT3A and 3B) gene, which is also translocated to the mitochondrial matrix (Bellizzi et al., 2013; Shock et al., 2011; Wong et al., 2013). In the recent past, we have integrated a number of orthologous datasets available in public domain to reconstruct a genome scale map of cytosine methylation in the human mitochondria (Ghosh et al., 2014).

Apart from DNA methylation at cytosine bases, another epigenetic modification that has recently attracted immense interest is hydroxymethylation of cytosine (5hmC) (Penn et al., 1972). 5hmC has been largely thought to be derived from the conversion of 5mC enzymatically catalyzed by the ten-eleven translocation (TET)-family of genes (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Thus hydroxymethylation is believed to be an intermediate step of DNA demethylation (Kohli and Zhang, 2013). However, hydroxymethylation has also been recently studied for their roles in gene regulation (Guo et al., 2011; Serandour et al., 2012). In the mammalian genome hydroxymethylation is thought to show a tissue specific pattern (Li and Liu, 2011; Nestor et al., 2012). For example, the brain tissue has been shown to have a larger frequency of hydroxymethylation compared to other tissues (Wen et al., 2014). In addition, a number of tumors have also been explored to have characteristic patterns of hydroxymethylation

* Correspondence to: S. Sengupta, Genomics and Molecular Medicine,CSIR Institute of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India.
** Correspondence to: V. Scaria, GN Ramachandran Knowledge Center for Genome Informatics, CSIR Institute of Genomics and Integrative Biology, Mathura Road, Delhi 110025, India.
E-mail addresses: shantanus@igib.res.in (S. Sengupta), vinods@igib.in (V. Scaria).

http://dx.doi.org/10.1016/j.mito.2016.01.003
1567-7249/

(Thomson et al., 2013). Several recent studies have shown that hydroxymethylation marks are predominantly present at the promoter, enhancers and gene body regions in the nuclear genome, which has positive correlation with gene expression (Gan et al., 2013; Nestor et al., 2012; Rudenko et al., 2013; Tsumagari et al., 2013). The presence of hydroxymethylation of cytosine has been previously reported in mitochondria (Shock et al., 2011).

However, the levels of hydroxymethylation across the entire mitochondrial genome have not been characterized, in the present study, we have extensively used genome scale datasets of hydroxymethyl immunoprecipitation followed by sequencing, available in public domain for 23 different cell lines and tissue types to create a genome-scale map of 5hmC across the mitochondrial genome. Our analysis suggests a dynamic pattern and profile of hydroxymethylation in the human mitochondrial genome.

2. Materials and methods

The datasets were derived from publicly available resource for next-generation sequence data, the Short Read Archive (SRA) maintained at the National Center for Biotechnology Information (NCBI). The hydroxymethylation datasets used in the present study were derived from 4 different studies (Kim et al., 2014; Madzo et al., 2014; Wang et al., 2012; Wang et al., 2013a), with corresponding accession IDs SRP013367, SRP012058, SRP015138 and SRP015342. These studies comprised of five different cell and tissue types, viz. CD34, IMR90, Human cerebellum, iPS and H9 HESC, and a total of 23 datasets (Table 1). Human reference genome (hg19 build) derived from University of California, Santa Cruz (UCSC) (Karolchik et al., 2004) database was used as the reference genome for alignments. The raw FASTQ files of all the 23 different datasets were first quality checked using FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and then quality trimmed using SolexaQA (Cox et al., 2010). Reference alignment was performed against the human genome build hg19, using bowtie2 aligner (Langmead and Salzberg, 2012) with default parameters. All the datasets were aligned with an average 90% mapping efficiency and had 38 million mapped reads in an average. In-house perl scripts were then used to extract reads specific to mitochondria from the alignments for further downstream analysis (Table 1). The alignments corresponding to the mitochondrial reads were then formatted as standard BED format using SAMtools and BEDtools (Li et al., 2009; Quinlan and Hall, 2010). The formatted reads were further used to compute the hydroxymethylation density across the mitochondrial genome using custom perl scripts. The analysis flow-diagram has been shown in Fig. 1.

For constructing the bin-wise map of hydroxymethylation density across the mitochondrial genome, the total mitochondrial genome (16,571 bp) was divided into non-overlapping, continuous 50 nucleotide bins. Further, we calculated the total number of hydroxymethylated reads overlapping each of the bins. Overlapping read counts in each 50 bp bins were then further normalized by the total number of mapped reads, mapping to the mitochondrial genome in the specific dataset. This hydroxymethylation density across the mitochondrial genome for 23 different cell and tissue types was then clustered hierarchically using R-statistical tools 'gplots and ggplot2' packages and then visualized using the Circos tool (Krzywinski et al., 2009) (Fig. 2A and B).

Since the mitochondrial genome share sequence homology with nuclear genome, which might affect the mapping of hydroxymethylated reads in the mitochondrial genome. We have simulated 50 bp synthetic reads with single base pair sliding window, throughout the hg19 mitochondrial genome. All 16,522 synthetic reads were then mapped back to the human hg19 genome, using default parameter of bowtie2 aligner. A count of 1 was assigned against the synthetic reads mapped uniquely to the mitochondria and reads mapped to other chromosomes assigned count 2, which represent a non-unique mitochondrial region. These numeric count were then plotted according to their corresponding synthetic reads, throughout the mitochondrial genome as a bar plot (Supplementary figure 1).

Further, these unique and non-unique synthetic reads were used for reconstructing the bin wise hydroxymethylation density map, only for unique mitochondrial sequences. Using the in-house Perl scripts, all the 50 bp non-overlapping bins were removed, which were overlapped (complete as well as partially) by non-unique synthetic reads. Thus, all remaining bins were completely overlapped by unique synthetic reads, which represents the unique mitochondrial sequences. Hydroxymethylation density across these unique mitochondrial regions was then plotted as a heatmap, using R-statistical tools 'gplots and ggplot2' packages (Fig. 3).

A unique-score was calculated using the above mentioned algorithm, where a score 1 represents the unique 50 bp bin and anything less than 1 score, represents non-unique 50 bp bin. These scores were calculated throughout the mitochondrial genome and plotted as a stepplot using R-statistical tool (Fig. 3).

Table 1
Summary of the hMeDIP datasets analyzed in the present study with details of raw and aligned reads.

Sample ID	Name	Raw read No.	Read No. (after QC)	Mapped reads	Total alignment %	Mapped mt reads	Author	PMID
GSM936817	H9 Embryonic stem (ES) cells	16,017,155	14,304,421	12,912,546	90.27%	146,203	Mirang Kim et al.	24087792
GSM936818	H9 ES-derived neural precursor cells	16,835,595	15,989,618	13,767,996	86.11%	125,317	Mirang Kim et al.	24087792
GSM936819	H9 ES-derived dopamine neuron cells	17,042,948	16,205,398	14,178,007	87.49%	146,669	Mirang Kim et al.	24087792
GSM909322	HUES48	46,614,652	32,696,834	27,200,497	83.19%	9000	Tao Wang et al.	23685628
GSM909323	HUES49	39,430,026	26,660,668	22,790,606	85.48%	7163	Tao Wang et al.	23685628
GSM909324	HUES53	45,901,658	38,410,398	34,042,555	88.63%	14,723	Tao Wang et al.	23685628
GSM909325	iPSIMR90	50,177,110	34,660,531	27,775,871	80.14%	9933	Tao Wang et al.	23685628
GSM909326	iPSA2	35,090,991	22,888,293	18,932,158	82.72%	14,740	Tao Wang et al.	23685628
GSM909327	iPSAG2.3	53,611,290	39,202,066	32,303,016	82.40%	7028	Tao Wang et al.	23685628
GSM909328	iPSB22	49,186,918	32,651,348	28,323,130	86.74%	5763	Tao Wang et al.	23685628
GSM909329	iPSB23	46,420,305	32,146,330	26,725,959	83.14%	9566	Tao Wang et al.	23685628
GSM909330	iPSRX35i	39,544,303	35,773,260	33,417,482	93.41%	36,256	Tao Wang et al.	23685628
GSM909331	iPSS1	38,636,965	35,380,626	34,336,877	97.05%	6806	Tao Wang et al.	23685628
GSM909332	iPSS2	41,301,319	37,459,788	36,437,649	97.27%	4213	Tao Wang et al.	23685628
GSM909333	iPSS3	47,333,331	42,499,638	41,239,589	97.04%	20,994	Tao Wang et al.	23685628
GSM909335	CRL2097duplicate	40,095,335	38,340,559	36,435,008	95.03%	17,474	Tao Wang et al.	23685628
GSM909337	GM0011duplicate	36,662,588	35,046,401	32,786,818	93.55%	24,441	Tao Wang et al.	23685628
GSM909339	IMR90duplicate	46,006,973	43,689,653	41,371,414	94.69%	42,358	Tao Wang et al.	23685628
GSM992935	CD34: d0	86,121,038	86,121,038	84,624,810	98.26%	21,689	Jozef Madzo et al.	24373966
GSM992936	CD34: d3	75,836,386	75,836,386	73,735,133	97.23%	121,953	Jozef Madzo et al.	24373966
GSM992937	CD34: d7	113,955,241	113,955,241	110,505,641	96.97%	212,501	Jozef Madzo et al.	24373966
GSM992938	CD34: d10	99,452,202	99,452,202	96,499,903	97.03%	66,075	Jozef Madzo et al.	24373966
GSM995965	Human cerebellum DNA	10,849,737	9,050,776	7,951,534	87.85%	972	Tao Wang et al.	23042784