Genomics 103 (2014) 204–210

Contents lists available at ScienceDirect

Genomics

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

Direct ChIP-bisulfite sequencing reveals a role of H3K27me3 mediating aberrant hypermethylation of promoter CpG islands in cancer cells

Fei Gao [a,*,1], Guanyu Ji [a,1], Zhaowei Gao [a], Xu Han [a], Mingzhi Ye [a], Zhimei Yuan [a], Huijuan Luo [a], Xiaojun Huang [a,e], Karthikraj Natarajan [a], Jun Wang [a,b,c,d], Huanming Yang [a], Xiuqing Zhang [a]

[a] Science & Technology Department, BGI-Shenzhen, Shenzhen, 518083, China
[b] Department of Biology, University of Copenhagen, Copenhagen, 2200, Denmark
[c] King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[d] The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, 2200, Denmark
[e] College of Life Sciences, Wuhan University, Wuhan, 430072, China

ARTICLE INFO

Article history:
Received 10 July 2013
Accepted 28 December 2013
Available online 7 January 2014

Keywords:
DNA methylation
Chromatin immunoprecipitation
CpG Island
Gastric cancer

ABSTRACT

The model describing that aberrant CpG island (CGI) methylation leads to repression of tumour suppressor genes in cancers has been influential, but it remains unclear how such aberrancy is induced. Recent studies provided clues indicating that promoter hypermethylation in cancers might be associated with PRC target genes. Here, we used ChIP-BS-seq to examine methylation of the DNA fragments precipitated by the antibodies to both H3K27me3 and H3K4me3 histone modifications. We showed that, for a set of genes highly enriched with H3K27me3 both in cancer and normal cells, CGI promoters were aberrantly hypermethylated only in cancer cells in comparison with normal cells. In contrast, such aberrant CGI hypermethylation in cancer promoters that were deficient of H3K27me3 was not notable. Furthermore, we confirmed that these genes were consistently hypermethylated in TCGA primary cancer cells. These works support the association between H3K27me3 and DNA methylation marks for specific cancer genes and will spur future work on combined histone and DNA methylation that could define cancer's epigenetic abnormalities.

1. Introduction

Epigenetic information comprises histone modifications [1] and DNA methylation [2] that can substantially influence chromatin structure and DNA accessibility [3,4]. In the last decade, the study of epigenetic mechanisms has been highlighted in cancer causation, progression and treatment as an alternative for genetic defects [5]. Especially, aberrant DNA hypermethylation of CpG island (CGI) promoters is associated with transcriptional repression of many tumor suppressor genes that can lead to tumor progression in many cancers. Although this model has been hugely influential [6], the significance of hypermethylation at CGIs in cancer has long been debated as well [7,8]. Recently, a research group has observed that genes that are hypermethylated and repressed in cancers were also repressed in pre-cancerous tissues even though their promoters are hypomethylated [9,10]. Another research group found that many genes with *de novo* promoter hypermethylation in colon cancer were among the subset of genes that were bound with both the repressing H3K27me3 and the activating H3K4me3 in embryonic stem cells and adult stem/progenitor cells [11,12]. H3K27me3 is catalyzed by the SET domain histone methyltransferase EZH2, a component protein of the Polycomb-repressive complex (PRC). It was known that genes that are targeted and repressed by PRC are also poised for activation in pluripotent cells [13]. It was also discovered by using chromatin immunoprecipitation in a previous study that binding of DNA methyltransferases (DNMTs) to several EZH2-repressed genes depended on the presence of EZH2 [14]. These observations suggested that promoter hypermethylation in cancers might be associated with PRC target genes. However, *de novo* DNA methylation without PRC occupancy or *de novo* PRC occupancy without DNA methylation did exist as well [15].

Recently, a simple and effective new method termed ChIP-BS-seq (chromatin immunoprecipitation followed by bisulfite sequencing) was developed, enabling direct examination of the methylation status of DNA sequences immunoprecipitated by ChIP for specific histone modifications [16,17]. Therefore, we took advantage of this new ChIP-BS-seq technology and applied it to examine trimethylation of histone H3 lysine 27 (H3K27me3) and histone H3 lysine 4 (H3K4me3) profiles for a normal cell line (YH lymphoblastoid) and three cancer cell lines (one cervical cancer cell line (HeLa) and two gastric cancer (GC) cell

Abbreviations: CGI, CpG island; PRC, Polycomb-repressive complex; H3K27me3, histone H3 lysine 27; TSSs, transcription start sites; COAD, colon adenocarcinoma; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma; RPKM, Reads Per Kb per Million reads.

* Corresponding author.
E-mail address: gaofei@genomics.cn (F. Gao).

[1] Both authors contributed equally to this work.

0888-7543/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.12.006

lines (BGC-823 and AGS)). Additionally, we also downloaded the ChIP-BS-seq data for a normal and two cancer cell lines from previous studies [16,17], thus, enabling examination of 7 cell lines in total.

By overall comparison between these cancer and normal cell lines, we showed that a set of genes commonly enriched with H3K27me3 marks both in cancer and normal cells presented aberrant hypermethylation of promoter CGIs in cancers, but hypomethylated state is maintained in the normal cells. Gene ontology analyses suggested that these genes were highly enriched in ion transport or cellular ion homeostasis pathway, which were reported frequently in the study of carcinogenesis and cancer metastasis.

Furthermore, we obtained data from primary cancers in the Cancer Genome Atlas (TCGA), and confirmed that some highly methylated genes in cancer cells were also significantly hypermethylated in TCGA data for primary cancers. By combining our cell line methylation data with TCGA data, we discovered new genes that possess significantly different methylation patterns between cancer and normal tissues.

2. Materials and methods

2.1. Cell acquisition

A lymphoblastoid cell line was generated from a male Han Chinese individual (YH), whose genome sequence was reported previously [18]. Two human gastric cancer cell lines (BGC-823 and AGS) were provided by Beijing Tumor Hospital; the HeLa cell line was purchased from the American Type Culture Collection (ATCC). Cells were cultured with RPMI1640 (Gibco C22400500BT) supplemented with 10% fetal bovine serum (Gibco 12657–029) in a humidified incubator with 5% CO_2 at 37 °C.

2.2. Chromatin immunoprecipitation sequencing (ChIP-seq)

Cells were precipitated by centrifugation, and the supernatant was removed. For each ChIP assay, approximately 5×10^6 cells were used. DNA and proteins were cross-linked with 1% formaldehyde in 10 ml PBS at 37 °C for 10 min, and then the cells were washed with pre-cooled PBS with 0.5% bovine serum and by PBS supplemented with protease inhibitor compound (PIC). The cells were collected by centrifugation at 850 rpm for 3 min after each wash. The cells were resuspended in 200 μl ice-cold lyses buffer (1% SDS, 10 mM EDTA, 50 mM Tris–HCl, plus PIC) and then thawed on ice for 10 min to allow cell lyses. The cell lysate was sonicated for 180 s using BioruptorTM 200 (pulses: 30 s on/30 s off) to generate chromatin fragments of size range from 100 to 700 bp. As input, one-tenth of the sonicated chromatin sample was separated. The remaining chromatin was immunoprecipitated in ChIP dilution buffer (1% Triton, 2 mM EDTA, 150 mM NaCl, 20 mM Tris–HCl) with 4 μg of antibody against H3K4me3 (Millipore; 17–614) or H3K27me3 (Millipore; 17–622) that has been pre-incubated with protein A/G magnetic beads (Invitrogen; 10003D). The immunoprecipitation reaction was incubated overnight at 4 °C and the beads were washed twice with each of the following buffers at 4 °C: RIPA buffer (10 mM Tris, 1 mM EDTA, 0.1% SDS, 0.1% Na-deoxycholate, 0.1% Triton X-100); RIPA buffer plus 0.3 M NaCl; LiCl buffer (0.25 M LiCl, 0.5% NP-40, 0.5% Na-deoxycholate) and TE buffer. The reaction tube was placed in a magnetic rack to capture the beads. The bounded DNA was eluted from the beads with elution buffer (1% SDS, 0.1 M $NaHCO_3$) at 65 °C for 2–3 h. The same procedure was performed for the input sample. The immunoprecipitated DNA was purified by phenol-chloroform extraction and precipitated with ethanol and glycogen. Successful immunoprecipitation was verified by qPCR using input as background. The obtained high-quality DNA was subjected to library preparation and sequenced on Illumina Hiseq 2000 using a standard pair-end 50 bp (PE50) sequencing protocol.

2.3. Library construction for bisulfite-sequencing of ChIP-ed DNA

The ChIP DNA (60 ng) ends was repaired in a 100 μl reaction containing 1XT4 PNK buffer, 3 units T4 DNA polymerase (ENZYMATICS; P708), 0.5 unit Klenow enzyme (ENZYMATICS; P7060), 10 units T4-PNK (ENZYMATICS; Y9040) and dNTP (0.125 mM for each) for 30 min at 20 °C. To the end repaired ChIP DNA, 15 unit Klenow fragment (3′–5′ exo, ENZYMATICS; P7010-LC) in a 50 μl reaction containing 1X blue buffer (ENZYMATICS; B011) and 0.2 mM dATP at 37 °C for 30 min to generate protruding 3′Abase. Methylated pair-end adapters were ligated to the DNA fragments using 2400 unit of rapid T4 DNA ligase (ENZYMATICS; L6030-HC) at 37 °C for 15 min. After purification, 200 ng exogenous λ-DNA fragments were added to the samples, and the sodium bisulfite conversion assay (ZYMO D5006) was performed, followed by 16 cycles of PCR amplification that was consisted of denaturation (94 °C for 15 s), annealing (60 °C for 30 s), extension (72 °C for 30 s). Then, the PCR products were size-selected on 2% agarose gel, retaining 250–350 bp DNA fragments. The purified DNA (ChIP-BS libraries) was used for cluster generation and standard PE50 sequencing using Illumina Hiseq 2000.

2.4. Processing of ChIP-seq reads

After PE50 sequencing, ChIP-seq reads were processed by the Illumina base-calling pipeline. Low quality reads that contain more than 30% of 'N' or over 10 % of the sequence with low quality per reads were omitted from the data analysis. The reads were aligned with the human reference genome (UCSC hg18) using SOAP (Short Oligonucleotide Analysis Package) 2.01 [19] with default parameters. Reads that were mapped to more than one position in the genome were filtered out. Multiple reads mapping to the same position were counted once to avoid potential bias from PCR. RSEG algorithm was applied for identification of H3K27me3- and H3K4me3-enriched regions [20]. RSEG is based on hidden Markov model (HMM) framework including the Baum–Welch training and posterior decoding, modeling the read counts with a negative binomial distribution. Subsequently, it uses a two-state HMM for segmentation of the genome into fore-ground domains and background domains. To define an enriched region of H3K27me3 or H3K4me3, default RSEG settings was used based on bin size of 500 bp, including that the posterior probability of each bin obtained by HMM decoding is larger than 0.95 and that the mean of read counts within a region is above the top 90th percentile of foreground emission distribution. The adjacent enriched bins were merged.

2.5. Processing of ChIP-BS reads

For histone modification signal, same processing was performed as above for ChIP-seq reads. For DNA methylation signal, ChIP-BS sequencing reads were aligned to the human reference genome (UCSC hg18) using an algorithm adapted from the procedure described by Lister et al. [21]. Because DNA methylation has strand specificity, all cytosines in the reference sequences ("original form") were replaced *in silico* by thymines ("alignment form") to allow alignment after bisulfite conversion. The "original forms" of the sequencing reads were transformed to cope with the BS treatment nucleotide conversion during the alignment process using the following criteria: (1) observed cytosines in the forward read of each read pair were replaced by thymines *in silico* and (2) observed guanines in the reverse read of each read pair were replaced by adenosines *in silico*. We then mapped the "alignment form" sequencing reads to the "alignment form" reference sequences using SOAP 2.01 [19] with default parameters. After mapping, the number of methylated (C) and unmethylated (T) basecalls at each CpG site within the genome was used to determine the methylation status of each sequenced cytosine within a CpG context, both on the forward strand as well as on the reverse strand. The DNA methylation level of each genomic region was defined as the ratio of supported methylated reads with the sum of methylated and