

Minireview

Reduced-representation methylation mapping

Jeffrey A Jeddloh*, John M Greally[†] and Oliver J Rando[‡]

Addresses: *Roche NimbleGen Inc., Research and Development, 500 S. Rosa Rd, Madison, WI 53719, USA. [†]Center for Epigenomics and Department of Genetics (Division of Computational Genetics), Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, NY 10461, USA. [‡]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation St, Worcester, MA 01605, USA.

Correspondence: Jeffrey A Jeddloh. Email: jeffrey.jeddloh@roche.com. Oliver J Rando. Email: Oliver.Rando@umassmed.edu

Published: 1 September 2008

Genome Biology 2008, **9**:231 (doi:10.1186/gb-2008-9-8-231)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/8/231>

© 2008 BioMed Central Ltd

Abstract

The power of massively parallel sequencing has been harnessed to map cytosine methylation patterns in the mouse genome, allowing insights into the relationship of methylation with DNA sequence, histone modifications, transcriptional activity and dynamic changes in methylation status during differentiation.

The past few decades have seen a revolution in the understanding of mechanisms of 'epigenetic' inheritance - the passing on from one generation of cells to another of information that affects phenotype but does not alter the actual DNA sequence. One of the longest known epigenetic modifications is the methylation of cytosines in DNA, which is typical of organisms with large genomes whose cells undergo many divisions over the organism's lifetime. In recent years the realization that inappropriate methylation of promoters of tumor suppressor genes may contribute to oncogenesis has sparked renewed interest in DNA methylation. Despite many years of work, however, how changes in DNA methylation status occur during normal cell differentiation and whether these changes have a role in regulating gene expression still remain unclear. One of the few unambiguously recognized facts about DNA methylation is that a methylated promoter is highly likely to be silenced, whereas hypomethylation of a promoter can be associated with active or silent genes.

As with all things DNA-centric, the genomics revolution holds great promise for understanding where cytosine methylation occurs throughout the genome, potentially suggesting biological function through 'guilt by association'. In a study published recently in *Nature*, Meissner *et al.* [1] ally the power of massively parallel DNA sequencing to a clever strategy for picking out GC-rich sequences from the

genome to produce the most comprehensive insight so far into cytosine methylation at nucleotide resolution in a large mammalian genome, that of the mouse. Such techniques will undoubtedly increase the potential for understanding the biological role of this venerable epigenetic regulator and its dysregulation in disease.

Cytosine methylation goes 'omic

Current assays for cytosine methylation are generally divided into those involving protein-mediated detection and those using chemical detection. The former allow lower-resolution discovery studies genome-wide, a useful first pass in many applications, whereas the latter allow nucleotide-resolution studies but have been refractory to scaling. Only chemical detection can provide whole-genome, single-base resolution from individual DNA molecules. This level of resolution is based on the deamination of cytosine, but not 5-methylcytosine, to uracil by bisulphite and related compounds. *In vitro* mutagenesis of DNA with bisulphite followed by resequencing thus identifies unmethylated cytosines, as they will have been replaced by thymines after replication of the converted DNA. Failure to find thymine replacement is taken as a signal that the cytosine in question was originally the site of a 5-methylcytosine. Careful analysis of the resulting sequence is necessary to rule out technical artifacts, but as almost all mammalian cytosine methylation occurs within

the context of CG dinucleotides, non-CG cytosines serve as useful controls for conversion of unmethylated Cs.

Two types of bisulphite assays have been used, exploring what we call intermolecular and intramolecular dimensions. By intermolecular, we mean the overall frequency of methylation at a given CG across a population of molecules, whereas by intramolecular we mean the coordinate methylation patterns of different CGs located *in cis* on the same molecule. The short read lengths of the study by Meissner *et al.* [1] limit the ability to detect intramolecular information, although given evidence of correlation between CGs located up to 1 kb *in cis* in the human genome [2] this may not prove a great loss. On the other hand, intermolecular information is determined by read number per cytosine, and these data are valuable when allelic methylation is being studied, such as in genomic imprinting (for a review see [3]), and can indicate heterogeneity of methylation occurring in different cells in the population being studied. In any case, we note that the ideal assay to explore both dimensions of nucleotide-resolution bisulphite sequence would be one with deep coverage consisting of long *cis* read information.

Such a platform does not yet exist, but significant insights have been obtained into the relatively small *Arabidopsis thaliana* genome using massively parallel sequencing of bisulphite-converted DNA on a short-read platform (from Illumina) [4]. The substantially greater size of the human genome makes a comparable approach daunting, especially in analyzing the data. In large part, this is because after bisulphite conversion, the four-base 'native' genome is effectively collapsed into a three-base genome, with the original Cs remaining on only some CG dinucleotides. Worse, the strands of the genome are no longer complementary, so the effective size of the converted genome doubles, making sequence mapping to 6 Gb converted mammalian genomes a serious challenge. The study of Meissner *et al.* [1] partly gets round this problem by using a clever strategy of 'reduced-representation bisulphite sequencing'.

Reduced-representation bisulphite sequencing

Meissner *et al.* extend previous work from the same group [5] which used Sanger sequencing on a restriction-enzyme-based sampling of the genome. In the new work, reduced representation of the genome was achieved by the isolation of small restriction fragments generated by the methylation-insensitive type II endonuclease *MspI*, resulting in enrichment of CG-rich regions, and thereby directing the analysis towards CpG islands and less to CG-depleted regions of the genome. This enabled comprehensive sequencing of this fraction and easier mapping of the bisulphite-converted sequences onto this more limited search space. Moreover, because of the imposed directionality of the sequencing adaptors, there was only one strand to analyze, further simplifying the mapping problem at the expense of identifying

instances of hemimethylation - a relatively small price to pay. Only around 1% of the genome was sampled in this study, however, which leaves the remaining 99% of potential methylation space unexplored.

Analysis of the sequencing data was designed to ensure that only reads with unique alignments to the reference mouse genome were picked up. This involved mapping the sequence data to a simulated bisulphite-converted unmethylated genome, converting any remaining Cs in the sequence reads to Ts for initial alignment, and eliminating reads with more than one map position. Once mapped, the original Cs were recalled to the analysis, counting the frequency at which Cs and Ts were located at the same position across reads, and thus inferring the methylation status. To avoid including sequencing errors, quality scores for sequencing base calls were utilized as a data filter, ignoring low-quality bases at potential methylation sites, an important advance on previous studies [6,7].

Bisulphite genomic representations have traditionally demonstrated bias in genomic coverage, but recent observations suggest that by avoiding a step of cloning in bacteria, current massively parallel sequencing protocols have avoided such bias, although this should be determined empirically for each platform. Any detected distortion of representation and/or coverage at a CG site when comparing native and bisulphite-converted spaces could be used to adjust the inferred DNA methylation levels from those regions of the genome. Another useful *in silico* experiment would be to assess the impact of random data assembly within the representation space. An assessment of the contribution of random data to the inferred methylation levels would allow this artifact to be accounted for analytically in a region-specific manner. Finally, a future analytical goal should be to modify the match matrices used for alignment scoring. Mapping reads to a native genome using a penalty matrix tolerant to the bisulphite-induced SNPs in the reads should allow more accurate positioning of reads and identification of sample-specific SNPs to extract optimum-quality methylation data. Certainly, moving to larger endeavors (such as the whole genome) will require such fundamental approaches to be optimized. The study by Meissner *et al.* [1] sets the stage and serves as a compass for future efforts. Encouragingly, their platform's ability to measure genomic methylation was demonstrated through the convincing detection of biologically significant variation occurring with cell differentiation.

Insights into the biological significance of cytosine methylation

Meissner *et al.* had two major biological aims: to look first at static relationships of cytosine methylation with DNA sequence features and histone modifications in embryonic stem (ES) cells; and then to test the dynamic properties of

cytosine methylation during cell differentiation. In the first case, their analysis largely confirmed previous findings, with transposable elements consistently methylated and a strong correlation between CG density and hypomethylation. However, not all transposable elements behaved the same way - autonomously transposing long interspersed nuclear elements (LINEs) and long terminal repeat (LTR) elements were methylated whatever their CG content, whereas the more CG-rich subgroup of non-autonomous short interspersed nuclear elements (SINEs) were generally unmethylated and looked comparable to nonrepetitive DNA of similar CG content.

Correlation of cytosine methylation with histone-modification patterns was confirmed, with results suggesting that histone H3 lysine 4 and lysine 9 methylation are better predictors of cytosine methylation than CG content, although trimethylation of lysine 27 on H3 did not have the same predictive power. The last finding is interesting, as biochemical studies have found the enzyme responsible for methylating lysine 27 (the Polycomb group protein EZH2) to be in a complex with DNA methyltransferases [8], suggesting that both these repressive marks - trimethylated H3K27 and methylated cytosines - might have been coincident in the genome. Interestingly, CG-depleted non-promoter regions enriched in H3 lysine 4 dimethylation displayed a greater tendency for the DNA to be unmethylated, suggesting that the cell generally marks regulatory elements with DNA hypomethylation, an idea first proposed by Adrian Bird (for review see [9]).

A question that arises when considering cytosine methylation mapping studies is how and why methylation is directed to some sequences and not others. Much has been learned from the study of methylation mutants in the mouse and *Arabidopsis* [10-12]. Cytosine methylation is catalyzed by a family of DNA methyltransferases that do not have innate sequence specificity, with one recognized exception, the Dnmt3a-Dnmt3L complex, which appears to process CGs spaced apart by 8-10 bp (a single helical turn) more efficiently [13]. Demethylation may occur either passively - by failure to remethylate after replication - or actively by demethylase activity: potential demethylases might include glycosylase [14] or base excision activity [15] replacing the methylcytosine with cytosine, and even by mutagenic deamination mediated by DNA methyltransferases [16], although this field is not free from controversy [17]. The correlation between histone modifications and DNA methylation supports ideas that methylation may be targeted by pre-existing signals such as histone modification [18], or that both histone modification and cytosine methylation could be directed by sequence-specific molecules such as transcription factors or even small RNAs (for a review see [19]). However, one observation suggests that we are still missing part of the mechanism for establishing and maintaining the methylation pattern. In both the mouse and *Arabidopsis*, DNA left unmethylated as a result of mutation

of DNA methyltransferase will become remethylated on reintroduction of the enzyme, but such remethylation occurs much more slowly in the absence of the SWI/SNF maintenance factor DDM1/LSH1 [11,20-22].

The advent of see-it-all 'omics has reopened investigation into the role of cytosine methylation in cellular differentiation - a controversial topic. There are clear differences in cytosine methylation between distinct cell types [23], but many promoters are constitutively hypomethylated even when the gene is transcriptionally silent [24]. Meissner *et al.* [1] addressed this issue by differentiating ES cells to neural lineages in culture and retesting the same loci for cytosine methylation. They found that CG-dense regions remained hypomethylated for the most part, whereas CG-depleted regions were more likely to change DNA methylation status with differentiation. The so-called 'bivalent chromatin domains' first described in ES cells [25] tended to remain constitutively cytosine hypomethylated. The authors also compared *in-vivo*-derived, minimally cultured neural precursor cells (NPCs) with NPCs derived *in vitro*, revealing substantially less methylation in the primary cells. However, multiple passaging of the *in-vivo*-derived NPCs resulted in methylation patterns comparable with the ES-cell-derived NPCs, implicating cell culture in the generation of these 'epialleles'. Intriguingly, a specific subset of CG-dense promoters tended to be more consistently susceptible to this acquisition of methylation, an observation paralleling changes observed in cancer cells [26]. The effect of tissue culture on the variability of cytosine methylation has been recognized for some time [27]; the current study points to a similar phenomenon and was able to test far more CGs than previously possible. The obvious question is whether chromatin modifications, which are shown to be correlated with cytosine methylation, are also influenced by cell culture or whether this is an observation specifically affecting cytosine methylation.

The study by Meissner *et al.* represents a breakthrough in the ability to study cytosine methylation in mammalian cells, an advance that is directly due to the introduction of massively parallel sequencing technologies, coupled with a clever experimental design that allowed comprehensive sequencing of CpG islands and mapping of degenerate sequences to the genome. With longer reads, intramolecular cytosine methylation could be explored more comprehensively, and increased sequencing throughput should some day make whole-genome methylation studies feasible for genomes as large as those of mouse and human. Meissner *et al.* note the potential for increasing genomic coverage by adding further restriction enzyme representations or by coupling massively parallel bisulphite sequencing with the sequence-capture approach [28]. With these and other advances, the potential for revealing how cytosine methylation exerts its effects in normal cells will be immense, forming the basis for understanding the changes that we see in human disease.

Acknowledgements

JMG is supported by grants from the National Institutes of Health (NIH, R01 HD044078, R01 HG004401, R21 CA122339), and the High-Q Foundation. OJR is supported by a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund, and by grants from the NIH (R01 GM079205 and NIH Roadmap grant U54 RR020839).

References

- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: **Genome-scale DNA methylation maps of pluripotent and differentiated cells.** *Nature* 2008, **454**:766-770.
- Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci USA* 1992, **89**:1827-1831.
- Ubeda F, Wilkins JF: **Imprinted genes and human disease: an evolutionary perspective.** *Adv Exp Med Biol* 2008, **626**:101-115.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: **Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning.** *Nature* 2008, **452**:215-219.
- Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** *Nucleic Acids Res* 2005, **33**:5868-5877.
- Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, Caldwell CW, Shi H: **Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing.** *Cancer Res* 2007, **67**:8511-8518.
- Korshunova Y, Maloney RK, Lakey N, Citek RW, Bacher B, Budiman A, Ordway JM, McCombie WR, Leon J, Jeddelloh JA, McPherson JD: **Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA.** *Genome Res* 2008, **18**:19-29.
- Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F: **The Polycomb group protein EZH2 directly controls DNA methylation.** *Nature* 2006, **439**:871-874.
- Suzuki MM, Bird A: **DNA methylation landscapes: provocative insights from epigenomics.** *Nature Rev Genet* 2008, **9**:465-476.
- Li E, Bestor TH, Jaenisch R: **Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.** *Cell* 1992, **69**:915-926.
- Vongs A, Kakutani T, Martienssen RA, Richards EJ: ***Arabidopsis thaliana* DNA methylation mutants.** *Science* 1993, **260**:1926-1928.
- Woo HR, Dittmer TA, Richards EJ: **Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*.** *PLoS Genet* 2008, **4**:e1000156.
- Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X: **Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation.** *Nature* 2007, **449**:248-251.
- Jost JP, Siegmund M, Sun L, Leung R: **Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase.** *J Biol Chem* 1995, **270**:9734-9739.
- Vairapandi M: **Characterization of DNA demethylation in normal and cancerous cell lines and the regulatory role of cell cycle proteins in human DNA demethylase activity.** *J Cell Biochem* 2004, **91**:572-583.
- Métivier R, Gallais R, Tiffocche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G: **Cyclical DNA methylation of a transcriptionally active promoter.** *Nature* 2008, **452**:45-50.
- Ooi SK, Bestor TH: **The colorful history of active DNA demethylation.** *Cell* 2008, **133**:1145-1148.
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH: **DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA.** *Nature* 2007, **448**:714-717.
- Morris KV: **RNA-mediated transcriptional gene silencing in human cells.** *Curr Top Microbiol Immunol* 2008, **320**:211-224.
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddelloh JA, Riddle NC, Verbsky ML, Richards EJ: ***Arabidopsis* MET1 cytosine methyltransferase mutants.** *Genetics* 2003, **163**:1109-1122.
- Kakutani T, Munakata K, Richards EJ, Hirochika H: **Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*.** *Genetics* 1999, **151**:831-838.
- Zhu H, Geiman TM, Xi S, Jiang Q, Schmidtman A, Chen T, Li E, Muegge K: **Lsh is involved in *de novo* methylation of DNA.** *EMBO J* 2006, **25**:335-345.
- Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiak E, Figueroa ME, Glass JL, Chen Q, Montagna C, Hatchwell E, Selzer RR, Richmond TA, Green RD, Melnick A, Grealley JM: **Comparative isochizomer profiling of cytosine methylation: the HELP assay.** *Genome Res* 2006, **16**:1046-1055.
- Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D: **Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.** *Nat Genet* 2007, **39**:457-466.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125**:315-326.
- Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H: **Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer.** *Nat Genet* 2007, **39**:232-236.
- Meins F, Jr., Binns A: **Epigenetic variation of cultured somatic cells: evidence for gradual changes in the requirement for factors promoting cell division.** *Proc Natl Acad Sci USA* 1977, **74**:2928-2932.
- Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA: **Direct selection of human genomic loci by microarray hybridization.** *Nature Methods* 2007, **4**:903-905.