Epigenomics and Genome Wide Methylation Profiling

Pubudu S Samarakoon BSc, MSc

Tutor, Masters of Science in Biomedical Informatics Course, Postgraduate Institute of Medicine, University of Colombo

E-mail address: saneth.samarakoon@gmail.com

Abstract

Epigenetics is the study of the changes in gene expression that are heritable and do not involve a change in the DNA sequence. DNA methylation is one of the key epigenetic mechanisms that is clearly understood. DNA methylation plays a major role in transcriptional silencing in X inactivation, genomic imprinting and tumor or cancer formation. Today the field of epigenetics has evolved to epigenomics and the focus of DNA methylation analysis has shifted to genome wide methylation analysis. With the increasing interest in the field of epigenomics, initiatives such as the NIH Roadmap Epigenomics Program were established aiming to transform biomedical research by developing new technologies and resources for comprehensive epigenomic studies. Because of high productivity and high accuracy, "high throughput DNA methylation profiling" techniques are at the heart of these initiatives. Methylation profiling is performed on chemically treated DNA fragments using bead array platforms and DNA sequences resulting from high throughput sequencing (HTS). Bead array platforms use sets of probes for the identification of the methylation status and computer algorithms identify the methylation status by mapping DNA sequences to the reference genome.

Key words: Epigenetics, DNA methylation,

Preamble

It is well documented that most of the genomic sequences are identical even in cell types that differentiate early during the embryogenesis. This means that different physical characteristics of specialized cell types may occur due mechanisms other than changes in DNA sequences. However these acquired physical characteristics propagate during cell division, so that they are mitotically and meiotically heritable. Scientists were interested in these processes from the early days of study of human genetics and the study of the phenomenon by which a genotypes gives rise to different phenotypes by mechanisms other than changes in the underlying DNA sequence, was termed as "Epigenetics" by Corad Waddington in 1942 [1]. Today the field of epigenetics has evolved to keep up with the overwhelming data and knowledge from the human genome project. With the advancement of the field, epigenetics is currently defined as "the study of the changes in gene expression that are mitotically and/or meiotically heritable and do not involve a change in the DNA sequence" [2].

With the discovery of methylation specific restriction enzymes, investigators were able to identify a direct association between DNA methylation and epigenetic mechanisms. DNA methylation is the enzymatic attachment of methyl group to the 5th carbon of the cytosine and formation of 5-methyl cytosine. In the Human Genome, most of the 5-methylcytosines occur in CpG dinucleotides. DNA methylation is carried out by DNA methyltransferases (DNMTs) - DNMT1, DNMT3A and DNMT3B [3]. DNMT1 shows significant affinity to duplex DNA where only one of the two strands are methylated (hemimethylated DNA) [3]. It is identified as the enzyme which is responsible for the maintenance of genome wide DNA methylation by copying the methylation patterns to daughter strands during DNA replication [1]. DNMT3A and DNMT3B enzymes are the de novo methyltransferases that set up DNA methylation

patterns in early development and play a major role in embryogenesis [4].

Biological Significance of DNA Methylation

Two functions have been identified for DNA methylation. They are however not entirely mutually exclusive. They involve the suppression of gene expression and maintenance of genome integrity by suppression of repetitive elements and heterochromatin [1].

DNA methylation is associated with transcriptional silencing in X inactivation and genomic imprinting [5]. Methylation contributes to the stability of gene inactivation because both X inactivation and retroviral silencing [5] can be reversed by treatment of somatic cells with demethylation agents [6]. Association between DNA methylation and X inactivation is supported by the frequent reactivation of X-liked transgenes in mouse embryo cells and in cultured cells when DNMT1 is absent or inhibited [7]. It is documented that individuals who lack DNMT3B show reduced methylation and imperfectly silence X-linked genes [8].

Changes in the chromatin structure due to methylation modulated interactions between DNA and proteins affect transcription [9]. Methylation of the regulatory sequences for transcription factors myc and AP-2 inhibitions their binding to the regulatory elements and prevents transcription [10]. Similarly interactions between the zinc finger protein CTCF and parentally imprinted Igf2/H19 insulator can be blocked by DNA methylation [11]. It is nearly 15 years since scientists recognized the significance of DNA methylation in tumors. Studies have shown that methylation patterns in tumor cells are significantly different from those in normal cells [12]. Global hypomethylation and hypermethylation of genomic DNA are the main mechanisms involved in DNA methylation and tumor formation [1].

Tumor cells may even display global hypomethylation in genomic DNA even before tumor formation. Such decreased levels of genomic methylation leads to increased and inappropriate gene expression [13]. This is shown in studies on bcl-2 in chronic lymphocytic leukemia [14]. Activation of promoters derived from transposable elements is also identified as a consequence of DNA methylation deficiency. In normal somatic cells transposable element related sequences are heavily methylated and transcriptionally silent. Alu family is the most abundant SINE (Short Interspersed Nuclear Element) in the human genome which contains many inactive functional promoters. Interestingly, artificial demethylation processes can stimulate the expression of these inactive promoters [6]. This is further supported by the transcription of strongly hypomethylated LINE-1 (Long Interspersed Nuclear Element) retrotransposons and HERV-K (human endogenous retrovirus) proviral DNA observed in urothelial cell carcinomas [15].

Hypermethylation of CpG islands is the most commonly observed alteration in DNA methylation in cancers. Transcriptional repression due to hypermethylation of promoters of tumor suppressor genes which are not methylated in normal cells is identified as one of the courses for tumor formation [16]. In colorectal, lung and breast carcinomas, cell cycle inhibitor $P16^{INK4a}$ is heavily hypermethylated [17]. Hypermethylation profiling over more than 15 tumor types (colon, stomach, pancreas, liver, kidney, lung, head and neck, breast, ovary, endometrium, kidney, bladder, brain, and leukemia and lymphomas) has shown that all the metabolic pathways are affected by promoter hypermethylation-associated silencing. This is supported by the hypermethylation associated silencing shown in genes that are responsible for following processes in above mentioned tumor cells. Those processes are cell cycle ($p16^{INK4a}$ and $p15^{INK4b}$), DNA repair (hMLH1, MGMT, and BRCA1), cell adherence and

metastasis process (CDH1, TIMP3, DAPK), p53 network (p14ARF and p73), metabolic enzymes (GSTP1), and the APC/β-catenin route (APC). Thus, cell immortalization and transformation seen in any given tumor, it is possible to find simultaneous inactivation of several pathways by aberrant DNA methylation [18]. Recent epigenetic studies related to tumor development using restriction landmark genomic scanning have shown the methylation patterns are tumor specific [19]. Based on unique hypermethylation profiles, several groups have developed microarray based expression profiling methods to classify different tumors into several classes [20].

As discussed earlier DNA methylation plays an important role in genomic imprinting. Thus epigenetic studies have become significant in characterization of syndromes that arisen from defective imprinting. Prader-Willi Syndrome (PWS), Angelman syndrome (AS), Beckwith-Wiedemann syndrome (BWS) and ICF (immunodeficiency, centromeric heterochromatin, facial anomalies) immunodeficiency syndrome are three of the well characterized syndromes related to defective genomic imprinting [1].

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) were the earliest reported imprinting disorders in humans. Imprinting centers (IC) in chromosome 15q11-q13 were identified as the centers that control resetting of parental imprinted SNRPN and NECDIN genes along with clusters of genes in chromosome 15q11-q13 region [21]. Due to this, the maternally inherited copies of these genes are virtually silent, only the paternal copies of the genes are expressed. It has been found that a MspI/HpaII restriction site at the D15S63 locus on chromosome 15q11-q13 is methylated on the maternally derived chromosome, but unmethylated on the paternally derived chromosome [22]. PWS results from the loss of paternal copies of this region. Deletion of the same region on the maternal chromosome causes Angelman syndrome (AS) [23,24].

Beckwith-Wiedemann syndrome (BWS) is also a model imprinting disorder resulting from mutations or epigenetic events involving imprinted genes at chromosome 11p15.5. There are maternally expressed CDKN1C, KCNQ1, and H19 genes and the paternally expressed IGF2 and KCNQ1OT genes in this region. Molecular pathogenesis of familial and sporadic BWS differs but loss of imprinting in IGF2, epigenetic silencing in H19 and loss of methylation (LOM) at a differentially methylated region (KvDMR1) within the *KCNQ1* gene are strongly associated with BWS [25]. A study designed to group a large series of Beckwith-Wiedemann syndrome patients based on the methylation status of H19 and KCNQ1-overlapping transcript 1 (KCNQ1OT1) has shown reduced methylation of KCNQ1OT1 in all the subgroups of familial BWS cases. This suggests that the imprinting switch mechanism is disturbed by the alteration of the methylation status of these genes [26].

With increasing demand for understanding of epigenetic mechanisms novel investigations have focused on genome wide methylation pattern identification and expression analysis. This field of study has gained a real identity by itself and has been termed "epigenomics" [1]. The NIH Roadmap Epigenomics Program is one of the major initiatives established aiming to transform biomedical research by developing comprehensive reference epigenome maps and new technologies for comprehensive epigenomic studies. The overall hypothesis of this program is that the origin of health and susceptibility to diseases are, in part, the result of epigenetic regulation of the genetic blueprint [27].

"Reference Epigenome Production Centers" which are supported by the NIH Roadmap Epigenomics Program, develop reference epigenomes of a variety of human cells. These cell

types are selected by an international committee and other networks of this program. These reference epigenome maps are helpful to identify potential therapeutic targets, improve the knowledge of disease mechanisms, provide additional insights to genetic susceptibility of disease, pursue therapeutic opportunities in stem cell based and tissue regeneration strategies, and to understand normal differentiation, development, and aging process.

Epigenomics Data Analysis and Coordination Center (EDACC), which is funded by the NIH Roadmap Epigenomics Program coordinate all the mapping centers and provide data analysis. Further EDACC collects the data related to the Roadmap Epigenomics Program generated outside of the reference Epigenome Mapping Centers. In addition, the EDACC is responsible for banking standard data and making them publicly available by implementing data pipeline for National Center for Biotechnology Information (NCBI) database resources.

"Technology Development in Epigenomics" and "Discovery of Novel Epigenetic Marks in Mammalian Cells" are the other initiatives funded by NIH Roadmap Epigenomics Program. The technology development initiative aims to revolutionize epigenetic and epigenomics profiling and capture and analyze in vivo images of epigenetic changes in cells, tissues and eventually intact organisms. Identification of stable and long term changes in the epigenetic process, as epigenetic markers, and translating them into global epigenome maps in human cells is the goal of the "Discovery of Novel Epigenetic Markers in Mammalian Cells" program. These markers are important to study epigenetic processes which are critical to the normal development and function of an organism. Information on these markers would be significant to understand disease pathogenesis and devise novel therapeutic approaches [28].

Methods of DNA Methylation Analysis

Today, with the development of the technology and the amount of data generated in molecular biology, the demand for the understanding of methylation mechanisms and its biological significance is ever increasing. Advance technologies such as "High Throughput Technologies" are at the heart of current genomic studies. As such high throughput methylation profiling has become an important aspect of the field of epigenomics. High throughput methylation profiling is a challenging task as methylation patterns are not associated with changes in the DNA sequence. Thus most of these techniques rely on changing 5-methylcytosine to a thymine through a series of chemical modification. Bisulfite treatment and PCR amplification is the commonly used method for this modification and identification of converted thymines from 5-methylcytosines occur during DNA methylation profiling.

High Throughput DNA Methylation Profiling Using Universal Bead Arrays

Genome wide DNA methylation and its various impacts are investigated using high throughput DNA methylation profiling techniques. Such investigations that are performed on bead array platforms are commonly termed "High Throughput DNA methylation profiling using universal bead arrays" [29]. High Throughput Single Nucleotide Polymorphism (SNP) genotyping [30] is also used for High Throughput DNA methylation profiling. Here methylation is detected based on high throughput genotyping of bisulfite converted genomic DNA.

During the bisulfite conversion of genomic DNA, unmethylated cytosines are converted to uracils, while the methylated cytosines remain unchanged. This technique identifies methylation status of a particular CpG site by determining whether the site is bisulfite-converted or not using several types of probes and primer extension methodologies. Here,

two pairs of probes are designed, one each for methylated state and for unmethylated state of each CpG site. They are an allele-specific oligonucleotide (ASO) probe pair and a locus-specific oligonucleotide (LSO) probe pair. ASO consists of a 3 prime end which is complementary to either the "C" or "T" allele of the targeted CpG site and anneal to the bisulfite converted genomic DNA. The 5 prime end of the ASO contains a universal PCR primer sequence which is used as a priming site in the PCR amplification process. The LSOs consist of three parts, CpG locus-specific sequence at the 5 prime end, an address sequence which is complementary to a corresponding capture sequence on the bead array in the middle and a universal PCR priming site at the 3 prime end. Pooled oligos anneal to the bisulfite converted genomic sequence and primer extension is performed. Allele specific primers extend only if the 3 prime end is complementary to the correct allele of the genomic DNA template. This is followed by the ligation of the extended ASOs to their corresponding LSOs, to create PCR templates. The ligated template sequences are then amplified by PCR using universal primers. Common primers that anneal to the ASOs that associate with the "T" (unmethylated) allele and "C" (methylated) allele are fluorescent labeled with different dyes. PCR products are then hybridized to a bead array bearing the complementary sequences. Two fluorescent dyes are then used to distinguish between methylated and unmethylated loci. Methylation status of the interrogated CpG site is then calculated as the ratio of the signal from methylation (M) prob relative to the sum of both methylated (M) and unmethylated (U) prob. $\beta = (max(M, 0)) / (|U| + |M| + 100)$. In this calculation background intensity computed from a set of negative controls is subtracted from each analytical data point and constant bias of 100 is added to regularize β when both U and M values are small [29].

The Illumina epigenetic analysis solutions provide two assay types for high throughput methylation profiling. They are "Infinium methylation" and "GoldanGate Methylation" assay techniques. For these techniques, "Human methylation27", "cancer panel1" and "custom designed" bead chip platforms are offered. Human methylation27 bead chip makes it possible to investigate genome wide methylation profiles of 12 samples simultaneously. It interrogate more than 27,000 informative CpG sites from more than 14,000 genes per sample at single site resolution. Cancer panel1 spans 1505 CpG sites from 807 genes involved in different types of cancers and 96 samples per array can be assayed simultaneously [29,31].

The accuracy of this method has been studied by using bead array platform for methylation profiling of six X-liked genes (EFNB1, ELK1, FMR1, G6PD, GPC3, and GLA) along with 371 other genes in male and female genomic DNA. Result showed that low or no methylation levels in genes from male samples and hemimethylated genes from female samples. In this study the best 18 CpG sites in X-chromosomal genes where β was < 1.0, were selected to assess the sensitivity of the methodology. Female genomic DNA with these sites were then diluted with male DNA from the same set of genes (female:male is 5:95, 10:90, 20:80, 50:50). Two independent set of mixtures were made and four replicas for each mixture were done to run in parallel. Results were statistically analyzed and the standard deviation of the β-value obtained for all the 1536 CpG sites across the four replicates was <0.06 in 99% of cases. The average slope of β versus the expected methylation level for the selected X-chromosomal sites was equal to 1. Therefore, it is possible to conclude that the methylation profiling using bead array platforms can discriminate levels of methylation (β-values) that differ by as little as 0.17 [29].

High-Throughput Sequencing and DNA Methylation
New technologies have recently been developed for high-throughput DNA sequencing (HTS) which dramatically increases throughput compared to standard Sanger sequencing on

capillary sequencers. HTS technique parallelize the sequencing process enabling to produce millions of sequences in a single run [32]. Demand for such low cost and highly productive technologies has driven the development of bisulfite sequencing techniques with the power of high-throughput DNA sequencing. Pipeline of these high-through put bisulfite sequencing techniques can be described in three main steps. They are denaturation, bisulfite treatment and PCR amplification (Figure 1)**.**

Figure 1. The bisulfite sequencing pipeline [33]
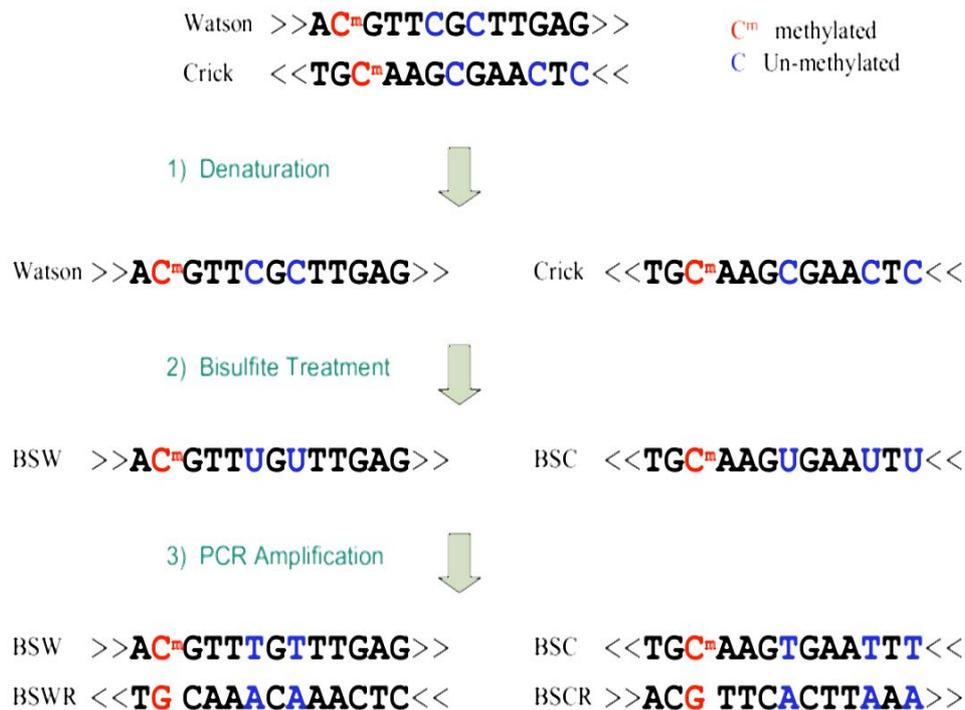
Denaturation step separates the two strands (Watson and Crick) of the double stranded DNA sequence and the bisulfite treatment converts all the unmethylated cytosines ("C" blue) to uracils ("U" blue). During the PCR amplification all the uracils are translated to thymines ("T" blue). Conversion of the unmethylated cytosines to thymines leads to the formation of four distinct bisulfite converted DNA sequences from a double stranded reference DNA fragment. Furthermore these resulting four sequences can be grouped into two groups. They are "C" poor sequences where all the unmethylated "C" are converted to "T" during the entire process, and "G" poor sequences where all the "G" are translated to "C" during the PCR amplification [33] . As a result overall "C" content of the original sequence has reduced by ~50% at the end of the process and this has a direct effect on the complexity of the genomic DNA.

The reduced complexity of the bisulfite converted sequences makes the DNA methylation profiling challenging for investigators who study genomic DNA sequences derived from bisulfite sequencing. Alignment of bisulfite converted sequences to reference genomes has become the basis of studies that have focused on such DNA methylation profiling techniques. These studies have used different approaches for mapping bisulfite converted sequences to reference genomes and identify methylated and unmethylated sites.

An approach for methylation analysis of DNA sequences is to convert all the "C" to "T" in the reference sequence *in silico* and map the bisulfite treated sequence to the converted

reference. Then the results are analyzed to count false positive bisulfite C/T mismatches where "C" of the bisulfite treated sequence is aligned with a "T" in the *in silico* converted reference sequence. This method makes it possible to identify methylated sites if the "C" poor bisulfite treated DNA strand is used in the analysis [34]. One of the limitations in this method is that it cannot be used if a "G" poor sequence is used to map with the converted reference. In "G" poor sequences, all the "C" in this sequences are being translated from "G's" during the PCR amplification process. So alignment of "C" from "G" poor bisulfite converted sequence with a "T" in the *in silico* converted reference sequence do not provide information on the methylation status. Occurrence of large number of false positives during mapping process is the other major limitation of this process. This is because mapping algorithm do not consider that a "T" in the bisulfite treated sequence can be matched with either "T" or "C" in the *in silico* converted reference [33].

Another bisulfite mapping strategy is the alignment of bisulfite treated DNA sequence to 3 reference sequences. These are original DNA reference sequence, *in silico* "C" to "T" converted sequence and *in silico* "G" to "A" converted sequence. Information on methylation status is computed by combining results from all three alignments and capture possible C/T alignments which are allowed to be mismatches. The alignment of sequence reads to reference sequences is performed using per-base probabilities derived from Gaussian mixture models (GMMs) developed to optimize the base call [35]. The drawback of this algorithm is that the number of bisulfite C/T mismatches can be identified is restricted because of the restrictions of the mismatches allowed in the alignment. In a study using this method DNA methylation in Arabidopsis was studied using 32bp bisulfite reads. 2 mismatches were allowed. So CpG islands with more than 2 methylated CpG sites were not identified. This number was further reduced by the presence of other true miss matched sites such as SNPs in the sequence [35].

General purpose Bisulfite Sequence MApping Program (BSMAP) is another computational approach to map bisulfite treated genomic sequences with the reference genomes. This program address the issue of T/C asymmetric mapping by masking "Ts" in the bisulfite reads that align with "Cs" in the reference and keeping the rest of the "Ts" in the read unchanged. The program uses a Hash table derived from the reference sequence to identify the map position of the bisulfite read. The original sequence fragment and all possible bisulfite variants of that fragment are used as keys in the hash table and corresponding coordinates in the reference genome are used as values. Each bisulfite read is fragmented in 6 reads and looked up in the hash table to identify the map position. Then bisulfite read and the correct map position of the reference are aligned. During the masking, DNA nucleotides are represented by two bits (A: 00, C: 01, G:10, T:11) and both masked read and the reference DNA are represented as binary strands. Based on the alignment, a bitwise "AND" operation is performed while masking to convert "Ts" to "Cs" in the bisulfite reads that align with "Cs" in the reference. Mismatches between masked bisulfite read and the reference are counted implementing a "bitwise exclusive OR" (XOR) operation [33]. Because of this masking, unmethylated cytosines which were converted to thymines during the bisulfite conversion are masked as cytosines.

This computational reverse bisulfite conversion allows proper identification of mismatches in the alignment between the bisulfite reads and reference genome sequences. BSMAP is implemented based on the open source software SOAP (Short Oligonucleotide Alignment Program) [36]. The other important feature of this program is that each bisulfite read and its reverse complement is aligned with both strands of the double stranded reference DNA. This

is because the reverse complement of the "Bisulfite Watson Reverse" (BSWR) strand matches with the reference Watson strand and reverse complement of the "Bisulfite Crick Reverse" (BSCR) strand matches with the reference Crick strand [33]. This is important because there is no precise way to determine the original strand that the read is derived from. So the result of four alignments allows selection of the precise match between bisulfite read and the reference sequence.

Summary

Epigenetics is the "study of the changes in gene expression that are mitotically and/or meiotically heritable and do not involve a change in the DNA sequence" [2]. DNA methylation process that add methyl group to the 5th carbon atom of the cytosine base at CpG dinucleotides without changing the nucleotide sequence is one of the key processes in epigenetic mechanisms. Transcriptional silencing in X inactivation and genomic imprinting are two important epigenetic mechanisms where DNA methylation plays a major role. It is well known that DNA hypermethylation and hypomethylation are directly associated with tumor formation. Hypomethylation leads to the inappropriate and increased levels of gene expression in tumors. Trancriptional repression that is seen in cancers is also mostly due to hypermethylation.

Today epigenetics has evolved to epigenomics and focus on the genome wide epigenetic analysis. This increasing demand for understanding of the genome wide methylation status has made high throughput techniques significant in methylation profiling. High throughput genome-wide methylation profiling techniques are developed using universal bead arrays and computational algorithms. High throughput DNA methylation profiling using universal bead arrays detect methylation status by high throughput genotyping of bisulfite converted genomic DNA. Methylation profiling of sequence reads coming from high throughput sequencing is performed by mapping bisulfite reads to the reference genome using computer algorithms.

References

1. Novik K, Nimmrich I, Genc B, Maier S, Piepenbrock C, *et al*. Epigenomics: Genome-Wide Study of Methylation Phenomena. Mol Biol 2002;4:111-128.

2. Wu C, Morris, J. Genes, Genetics, and Epigenetics: A Correspondence. Science 2001;293:1103-1105.

3. Jeffrey A Y, Nelliesh S S, Verdine G, Timothy B. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe1. ScienceDirect - Journal of Molecular Biology 1997;270:385-395.

4. Okano M, Bell D, Haber D, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99:247-257.

5. Adrian B. DNA methylation patterns and epigenetic memory. Genes and Development 2002;16:6-21.

6. Bird A. DNA methylation patterns and epigenetic memory. Genes and Development 2002;16:6-21.

7. Sado T, Fenner M, Tan S, Tam P, Shioda T, *et al* . X inactivation in the mouse embryo

deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Developmental Biology 2000;225:294-303.

8.   Miniou P, Jeanpierre M, Blanquet V, Sibella VR, Bonneau D, *et al*. Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Human Molecular Genetics 1994;3:2093-2102.

9.   Jones PL, Veenstra GCJ, Wade PA, Vermaak D, Kass SU, *et al*. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics 1998;19:187-191.

10.   Comb M, Goodman H. CpG methylation inhibits proenkephalin gene expression and binding of the transcription factor AP-2. Nucleic Acids Res 1990;18:3975-3982.

11.   Hark A, Schoenherr C, Katz D, Ingram R, Levorse J, *et al*. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 2000;405:486-9.

12.   Goelz S, Vogelstein B, Hamilton S, Feinberg A. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science 1985;228:187-190.

13.   Christman JK, Sheikhnejad G, Dizik M, Abileah S, Wainfan E. Reversibility of changes in nucleic acid methylation and gene expression induced in rat liver by severe dietary methyl deficiency. Carcinogenesis 1993;14:551-557.

14.   Hanada M, Delia D, Aiello A, Stadtmauer E, Reed J. bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. 1993;82:1820-1828.

15.   Florl A, Löwer R, Schmitz-Dräger B, Schulz W. DNA methylation and expression of LINE-1 and HERV-K provirus sequences in urothelial and renal cell carcinomas. British Journal of Cancer 1999;80:1312-1321.

16.   Baylin S, Herman J. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends in Genetics 2000;16:168-174.

17.   Merio A, Herman J, Mao L, Lee DJ, Gabriaelson E, *et al*. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nature Medicine 1995;1:686-692.

18.   Esteller M, Corn PG, Baylin SB, Herman JG. A Gene Hypermethylation Profile of Human Cancer. Cancer Research 2001;61:3225-3229.

19.   Costello J, Frühwald M, Smiraglia D, Rush J, Robertson G, *et al* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nature Genetics 2000;24(2):101-102.

20.   Golub T, Slonium D, Tamayo P, Huard C, Gaasenbeek M, *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:513-517.

21.   Ohta T, Gray T, Roghan PK, Gabriel J, Saitoh S, *et al*. Imprinting-mutation mechanisms in Prader-Willi syndrome. The American Journal of Human Genetics 1999;64:397-413.

22.   Dittrich B, Robinson WP, Knoblauch H, Schmidt K, Gabriele-Kaesbach G, *et al*. Molecular diagnosis of the Prader-Willi and Angelman syndromes by detection of parent-of-origin specific DNA methylation in 15q11-13. Human Genetics 1992;90:313-315.

23. Sahoo T, Gaudio D, German J, Shinawi M, Peters S, *et al.* Prader-Willi phenotype caused by paternal paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. Nature Genetics 2008;40:719-721.

24. Ding F, Dhar M, Jhonson D, Garnacho-Montero C, Nicholls R, *et al.* Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader-Willi syndrome mouse models. Mamm Genome 2005;16:424-431.

25. Engel J, Smallwood A, Harper M, Oshimura M, Reik W, *et al.* Epigenotype-phenotype correlations in Beckwith-Wiedemann syndrome. Journal of Medical Genetics 2000;37:921-926.

26. Bliek J, Mass SM, Ruijter JM, Hennekam RC, Alders M, *et al.* Increased tumour risk for BWS patients correlates with aberrant H19 and not KCNQ1OT1 methylation: occurrence of KCNQ1OT1 hypomethylation in familial cases of BWS. Human Molecular Genetics 2001;10:467-476.

27. Epigenomics - Overview [Internet]. The NIH Common Fund 2009;
     Available from: http://nihroadmap.nih.gov/epigenomics/
     Accessed 17 December 2009

28.   Epigenomics - Program Initiatives [Internet]. Epigenomics Common Fund 2009.
     Available from: http://nihroadmap.nih.gov/epigenomics/initiatives.asp
     Accessed 17 December 2009

29. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, *et al.* High-throughput DNA methylation profiling using universal bead arrays. Genome Research 2006;16:383-393.

30. Fan J, Oliphant A, Shen R, Kermani B, Gracia F, *et al.* Highly Parallel SNP Genotyping. Cold Spring Harbor Symposia on Quantitative Biology 2003;68:69-78.

31. Illumina inc. High-Throughput DNA Methylation Profiling with Illumina GoldenGateo Technology [Internet]. Illumina® Epigenetics Analysis. Available from:

     http://www.illumina.com/Documents/products/datasheets/datasheet_goldengate_methylation.pdf
     Accessed 17 December 2009

32. Hall N. Advanced sequencing technologies and their wider impact in microbiology. Journal of Experimental Biology 2007;210:1518-1525.

33. Yuanxin X, Wei L. BSMAP: whole genome Bisulfite Sequence MAPping program. BMC Bioinformatics 2009;10:232.

34. Meissner A, Mikkelsen T, Gu H, Wernig M, Slvachenko A, *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008;454:766-770.

35. Cokus S, Feng S, Zhang X, Merriman B, Haudenschild C, *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methyl-ation patterning. Nature 2008;452:215-219.

36. Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. Bioinformatics 2008;24:713-714.