

Epigenetics – A Science of Heritable Biological Adaptation

SHERIN U. DEVASKAR AND SANTANU RAYCHAUDHURI

Department of Pediatrics, Division of Neonatology and Developmental Biology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095

This annual review issue is dedicated to “Epigenetics” an emerging but fast growing field in the postgenomic era involving nonMendelian heritable changes in gene expression that are not mediated by alterations in Watson-Crick base-pairing of the DNA sequence (1,2). Conrad Waddington (1905–1975) coined the term “epigenetics” and defined it as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” (1,2). Both Darwin and Lamarck, the founders of evolutionary theory, predicted that “evolution may favor the development of self-guiding mechanisms, maximizing variability where and when it is most likely to yield positive changes while minimizing phenotypic variability when and where it is not needed” (1,2), supporting the general idea of nonrandom evolution. Epigenetic regulation mediates adaptation to the environment by the genome lending plasticity that translates into the presenting phenotype, particularly under “mismatched” environmental conditions (3).

Most investigations focus on demonstrating chemical modifications of the DNA sequence and/or the nuclear histone proteins that in turn alter the configuration of chromatin affecting its packaging and accessibility within the nucleus. The state of chromatin is critically important for the transcription of genes. Thus, *heterochromatin* signifies tightly wound chromatin impeding access of specific DNA sequences (cis-elements) by nuclear trans-activating proteins, a process that inhibits gene transcription. On the other hand, *euchromatin* is an active state, in which the loosely wound chromatin allows binding of the promoter region of genes by activating proteins resulting in active transcription. This simple concept, while serving as the ultimate process, is complicated by nuclear transport *via* the nuclear membrane pore or retention of critically important proteins, the complex nuclear arrangement of chromatin and proteins in a network creating loops and higher order structures, and recruitment of various enzymes and proteins to transcription initiation complexes, that are vital for either activating or repressing transcription (4,5), along with the newly evolving role for noncoding RNAs (6,7). The organization of chromatin in the nucleus allows the close proximity of the 5′- and 3′- ends of chromatin setting the stage

for factors that mediate mRNA processing at the 3′-end to be involved in transcription at the 5′-end. Further, the approximation of chromosomes lends to the process of nonrandom translocation that determines a particular cellular phenotype (4,5). This self-organization of the nucleus is thought to be responsible for the formation of nucleoli either in the presence or absence of a scaffolding network consisting of lamin and/or actin filaments (4,5).

Cytosine methylation of CpG dinucleotide clusters at the carbon 5 position found in close proximity to critically important *cis*-elements within the promoter can silence a gene. *De novo* and maintenance DNA methyl transferase enzyme isoforms (Dnmt1, 3a, 3b) responsible for methylating DNA and repressor methyl-binding domain (MBD 1, 2, 3) proteins or methyl CpG binding proteins (MeCP 1, 2) are necessary for mediating and/or modifying this reaction (8). Methylated DNA in turn attracts chromodomain binding proteins (heterochromatin protein, HP1) that maintain the heterochromatin state. What lends to further complexity is the long range chromatin interactions that occur which may not be evident at first pass when examining the promoter region of a gene where a concentration of unmethylated CpG islands tend to cluster (9). Usually CpGs in a GC rich region with excluded *Alu* repeats within a promoter are unmethylated and conserved across species while CpGs within or in close proximity to transposons or repeats are methylated (9). The mammalian one carbon metabolism provides the methyl groups for all biologic methylation reactions that in turn is dependent on methyl donors (methionine and choline) and co-factors (folic acid, vitamin B12, pyridoxal phosphate). Examples of the genome that carry methylated CpG regions as seen in the centromere, in the imprinting control regions (ICRs) which are characterized by differentially methylated regions (DMRs) are called “methylation marks” (4,8). These imprints resulting in gene silencing are seen with transposons, X-chromosome inactivation and parenterally imprinted genes causing mono-allelic gene expression (IGF 2, IGF2 receptor genes, *Peg*) (4).

The genome of the immature primordial germ cells of the developing embryo (zygote and blastomeres) undergoes extensive demethylation allowing for totipotency while erasing methylation marks thereby repressing the somatic program (10). Subsequent re-establishment of appropriate sex-dependent patterns of cytosine methylation during gametogenesis sets the stage for pluripotency that is maintained through embryonic development during many rounds of rapid cellular proliferation (mitosis) until final maturation of cell differentiation. The re-establishment of methylation imprints occurs

Correspondence: Sherin U. Devaskar, M.D., David Geffen School of Medicine at UCLA, Department of Pediatrics, Division of Neonatology and Developmental Biology, 10833 Le Conte Avenue, Room B2-375 MDCC, Los Angeles, CA 90095, Phone: (310) 825-9436, Fax: (310) 267-0154; e-mail: SDevaskar@mednet.ucla.edu

Statement of Financial Support: The authors are supported by grants from NICHD HD-25024, -33997, -41230 [S.U.D.] and -46979 [S.U.D., S.R.].

DOI: 10.1203/pdr.0b013e31805cbbd8

during late fetal development in male and postnatally in female germ cells. The male genome is highly condensed, partly through binding by protamines which are rapidly replaced by histones, a process that requires the histone chaperone (Hira) and occurs before the *S* phase. While the female genome must undergo demethylation, there are certain factors (Stella) that protect it from massive demethylation (10,11). Any derangement of these finely balanced processes can perturb the ultimate phenotype. Lessons have accrued from an increased incidence of parent-of-origin dependent gene imprinting disorders, such as Beckwith-Wiedemann and Angelman syndromes, reported in products of assisted reproductive technologies. This has led to conflicting theories related to the etiology being secondary to the technology *per se* or the innate presence of imprinting defects in gametes obtained from subfertile individuals (10). Other conditions such as Prader-Willi, Russel-Silver syndromes and Neonatal diabetes mellitus also implicate parent-of-origin differential gene methylation (12). Similarly hydatiform moles that affect the placenta are also due to disturbed expression of certain imprinted genes (10). Unlike methylation of CpGs that usually results in gene silencing, methylation of ^mCA/T repeat sequences in gene promoters has led to only a (~30–40%) decrease in gene expression (13). These chemical modifications of DNA sequences work in concert with co-valent histone modifications.

Chromatin structure is the packaging of DNA through association with histone proteins. The *nucleosome*, the basic repeating unit of chromatin consists of 147 bp of DNA wrapped around an octameric histone core containing two copies each of histones H2A, H2B, H3 and H4 and further compacted by association with the linker histone H1, additional nonhistone proteins and by higher order looping and folding of the chromatin fiber (10-30 nm) within the nucleus (14). The organization of chromatin not only limits physical access of nuclear proteins to the underlying DNA, but it is now clear that posttranslational modifications of histone proteins can alter chromatin conformation by an ATP-dependent remodeling process and play direct regulatory roles in gene expression (14). While a majority of nucleosomes in the cell are composed of the same four types of core histones, tremendous diversity in the histone/nucleosome structures is generated by a variety of posttranslational modifications such as acetylation, phosphorylation, methylation, de-imination, ubiquitylation, sumoylation, ADP-ribosylation and proline isomerization of histones (14). Acetylation, methylation, phosphorylation and ubiquitylation are implicated in activation while methylation, ubiquitylation, sumoylation, de-imination and proline isomerization in repression. However, any given modification is capable of either activating or repressing transcription depending on the specific location. An example is methylation at the lysine 36 residue of histone 3 (H3.K36) or H3.K9 when located in the coding region activates while in the promoter region represses gene expression (14).

Some charge-altering modifications such as acetylation that activates gene transcription and is mediated by histone acetyl transferases (HAT), or de-acetylation, that represses gene transcription and is mediated by histone deacetylases (HDAC), phosphorylation, *via* protein kinases that activates

genes, and de-phosphorylation, by phosphatases that deactivates genes are reversible, dynamic and associated with inducible expression of individual genes (14). Other modifications such as methylation were considered to be more stable and involved in long term maintenance of the expression status of regions in the genome, until the discovery of the demethylase class of enzymes (15,16). All these chemical modifications occur on multiple but specific amino acid residues of the N-terminal tails of histones, suggesting that histones act as signaling platforms integrating upstream signaling pathways to elicit appropriate nuclear responses such as transcriptional activation or repression of gene(s). While HATs (*e.g.*, CBP/p300) are involved in recruitment of transactivating nuclear proteins, the HDACs (multiple isoforms belonging to three major classes) in turn recruit methylases (*e.g.*, Suv39H1) or histone methyl transferases (HMT; *e.g.*, Sin3A, EZH2, SETDB1) and the chromodomain containing protein, HP1, that play a major role in gene silencing and maintaining the heterochromatin state (14).

Acetylation/de-acetylation occurs on specific lysine residues of the histone N-terminal tail, phosphorylation/de-phosphorylation on specific serine or threonine residues and methylation/demethylation on specific lysine and arginine residues. Methylation of lysines (K) can either activate or repress genes depending on the actual residue involved. For example, H3.K4 methylation activates while H3.K9 methylation deactivates gene transcription. Methylation of H3.K4 and H3.K36 is involved in elongation rather than initiation of RNA polymerase II mediated transcription, while methylation of H3.K9, H3.K27 and H4.K20 is connected to transcriptional repression. These lysine residues can be methylated in the form of mono-, di- or tri-methylation, while arginine residues can be methylated in the mono- or di- form providing further functional diversity to each site of lysine or arginine methylation (14). The discovery of demethylase enzymes such as LSD1 that shows specificity for H3.K4 and the JmjC domain containing JHDM1A-B that shows specificity for H4.K36 and JHDM2A-D supports demethylation as a process that dynamically affects gene transcription (14–16).

There is evidence for the emergence of a “combinatorial code” that H3 serine 10 phosphorylation inhibits, while lysine 14 deacetylation facilitates lysine 9 methylation which in turn silences gene transcription (17,18). Arginine methylation similar to lysine methylation, can be activating or repressing and rely on arginine methyl transferases (CARM1, PRMT4, PRMT5) however the proteins that bind methylated arginine residues and the enzymes that reverse this phenomenon are unknown (14). Arginines in H3 and H4 can be converted to citrulline by de-imination, thereby preventing methylation of arginine and dampening the activity of methylated arginines (14). The process of ubiquitylation causes a large modification in H2A and H2B that recruits proteins belonging to the Polycomb complex and is associated with transcriptional repression. Sumoylation, similar to ubiquitylation, results in large modifications in H4, H2A, H2B, antagonizes acetylation and ubiquitylation on the same lysine residue thereby causing transcriptional repression. ADP ribosylation can be mono- or poly- that is mediated by mono-ADP-ribosyltransferase

(MART) or poly-ADP-ribosylpolymerase (PARP) respectively (14), the latter is functional during DNA repair. Finally proline isomerization by FPR4 is required for the recognition and methylation of H3.K36 by Set2 methyl transferase. However, the recent discovery of *bivalent domains* that possess both activating and repressive modifications has destabilized this simplistic view of activating *versus* silencing modifications of chromatin (14).

Small noncoding RNAs can cause heterochromatin silencing as well. Co-transcriptional processing results in small RNA products that is responsible for RNA interference, an important regulatory mechanism for homology-dependent gene silencing. Posttranscriptional (PTGS; siRNA, microRNA) and transcriptional (TGS; antisense RNA) gene silencing form important epigenetic mechanisms (6,7). The first microRNA precursor involved in imprinting was *H19* in the case of IGF 2 gene expression (6,7) and the noncoding antisense transcript *Air* is expressed from a promoter in intron 2 of the IGF2 receptor gene on the paternally repressed allele while the maternal copy of the promoter is nonfunctional due to DNA methylation. Premature termination of imprinted *Air* transcription leads to loss of silencing of the IGF2 receptor gene (6,7).

While epigenetic studies have predominantly focused on chromatin and the use of specific enzyme inhibitors *in vitro* to prove their intracellular biologic role, more recent investigations have uncovered their significance *in vivo*, providing a functional role for some of these processes in the context of whole body physiology and phenotype. Examples include: a role for gene silencing of tumor suppressor genes (p53) or genes that mediate proliferation/apoptosis (Wnt signaling, *Bcl2* family) resulting in cancerous proliferation of cells leading to tumor formation (19); the role of environmental toxins containing phytoestrogens or diethyl stilbestrol, an estrogen receptor agonist, or an endocrine disruptor vinclozolin that affects chromatin structure leading to dysregulation of reproductive function (20) and persists into the fourth generation; a role for gene silencing and reprogramming during embryonic development (21); perturbations in DNA methylation with aging (10); the parent-of-origin differential methylation along with binding of nuclear factors such as CTCF (CCCTC binding factor) to the hypomethylated allele causing disordered growth (4,12); the influence of early stress affecting methylation of specific neural genes that dictate and permanently modify adult behavior (22); a role for early nutrition on gene transcription that impacts the ultimate phenotype as detected by the coat color (Agouti gene) (8) or the development of chronic diseases (diabetes/obesity) (8,23); and the confirmatory role of epigenetic regulation in trans-generational stable inheritance (20). The transgenerational persistence of phenotypic changes will have profound epidemiologic effects on health and disease.

Further gene mutations in the CREB binding protein (CBP), a histone acetyl transferase underlies the Rubenstein Taybi syndrome (22) and mutations in the MeCP2 protein are responsible for the neurodevelopmental disorder of Rett's syndrome (22), while other such related aberrations may underlie mental disorders including schizophrenia (22). Recent use of a genetically manipulated mouse model of Rett's syndrome

demonstrated that imposing normal levels of MeCP2 in neurons reversed the autism-like picture of Rett's syndrome thereby providing hope toward the future development of treatment strategies targeted at this clinical condition (24). A growing number of non-genetic disorders of growth, such as intra-uterine growth retardation, small for gestational age infants, the extremely premature infant, the product of *in vitro* fertilization, and infant of an obese or diabetic mother, have by association embedded their etiology in epigenetics since environmental cues perturb the embryonic hormonal/metabolic milieu (23,25).

As technology advances to large scale screening of methylated DNA regions by tiling array that allows complete high resolution analysis of the DNA methylome or genome-wide bisulfite sequencing to detect methyl cytosines, co-valent perturbations of chromatin by chromatin immunoprecipitation (ChIP)-on-chip, it is important to be cognizant of the limitations of these techniques. Amplification bias plays a major role in attempting to determine the methylated cytosines in the genome with a tendency to overlook the regions rich in repeats, while the ChIP-on-chip is as good as the detecting antibody. More recent technology addressing these limitations include the top down proteomics approach with mass spectroscopy (26). As more information in this field is forthcoming, it is critically important to continue to prioritize unraveling the biologic significance of these structural changes in chromatin with a "cause and effect" strategy. Many questions remain unanswered as they pertain to normal development, for instance in the face of demethylation, how are the methylation marks preserved and passed on to the next generation? Similarly, how are the histone modifications conserved and transmitted to the next generation, especially now that demethylases have been identified? How are the changes in the chromatin structure that occur in progenitor cells passed onto the daughter cells without an element of dilution? While these questions need answers, in the field of cancer, drugs targeted for clinical treatment have been developed. These include agents that reverse gene silencing as in the case of azanucleoside drugs that inhibit DNA methyltransferase enzymes and are currently approved by the food and drug administration (FDA) for myelodysplastic syndrome. Histone deacetylase inhibitors (*e.g.*, SAHA) have also been approved by the FDA for treatment of T cell cutaneous lymphoma. The future is focused on the ability to capitalize on the synergistic action of these two classes of agents against cancers not amenable to standard chemotherapy, since they only affect the abnormally proliferating cells that express the relevant silenced genes (19).

Structural chromatin changes may exist but not translate into a phenotype in the absence of the necessary environmental cues that serve as the selection pressure (*e.g.*, diet, drug, toxin, development or aging) and are required for either the survival advantage and/or adaptive expression. The best example is the differential phenotypic presentation of older monozygotic twins (10). Epigenetics is a science of heritable biologic adaptation, a concept that will become evident in the reviews presented in this issue.

REFERENCES

1. Goldberg AD, Allis CD, Bernstein E 2007 Epigenetics: A landscape takes shape. *Cell* 128:635–638
2. Rando OJ, Verstrepren KJ 2007 Timescales of genetic and epigenetic inheritance. *Cell* 128:655–668
3. Godfrey KM, Lillycrop KA, Burdge GC, Gluckman PD, Hanson MA 2007 Epigenetic mechanisms and the mismatch concept of the developmental origins of health and disease. *Pediatr Res* 61:5–10
4. Ling JQ, Hoffman AR 2007 Epigenetics of long range chromatin interactions. *Pediatr Res* 61:11–16
5. Misteli T 2007 Beyond the sequence: cellular organization of genome function. *Cell* 128:787–800
6. Saetrom P, Snove O, Rossi JJ 2007 Epigenetics and microRNAs. *Pediatr Res* 61:17–23
7. Chuang JC, Jones PA 2007 Epigenetics and microRNAs. *Pediatr Res* 61:24–29
8. Dolinoy DC, Das R, Weidman JR, Jirtle RL 2007 Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr Res* 61:30–37
9. Kouzarides T 2007 Chromatin modifications and their function. *Cell* 128:693–705
10. Poulsen P, Vaag ME, Fraga MF 2007 The epigenetic basis of twin discordance in age-related diseases. *Pediatr Res* 61:38–42
11. Surani MA, Hayashi K, Hajkova P 2007 Genetic and epigenetic regulators of pluripotency. *Cell* 128:747–762
12. Smith AC, Choufani S, Ferreira JC, Weksberg R 2007 Growth regulation, imprinted genes and chromosome 11p15.5. *Pediatr Res* 61:43–47
13. Malone CS, Miner MD, Doerr JR, Jackson JP, Jacobsen SE, Wall R, Teitell M 2001 CmC(A/T)GG DNA methylation in mature B cell lymphoma gene silencing. *Proc Natl Acad Sci USA* 98:10404–10409
14. Li B, Carey M, Workman JL 2007 The role of chromatin during transcription. *Cell* 128:707–719
15. Forneris F, Binda C, Vanoni MA, Battaglioli E, Mattevi A 2005 Human histone demethylase LSD1 reads the histone code. *J Biol Chem* 280:41360–41365
16. Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y 2006 Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439:811–816
17. Grewal SI, Elgin SC 2002 Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev* 12:178–187
18. Jenuwein T, Allis CD 2001 Translating the histone code. *Science* 293:1074–1080
19. Jones PA, Baylin SB 2007 The epigenomics of cancer. *Cell* 128:683–692
20. Skinner MK 2007 Endocrine disruptors and epigenetic transgenerational disease etiology. *Pediatr Res* 61:48–50
21. Paoloni-Giacobino AP 2007 Epigenetics in reproductive medicine. *Pediatr Res* 61:51–57
22. Feng J, Fouse S, Fan G 2007 Epigenetic regulation of neural gene expression and neuronal function. *Pediatr Res* 61:58–63
23. Simmons RA 2007 Developmental origins of beta-cell failure in type 2 diabetes: the role of epigenetic mechanisms. *Pediatr Res* 61:64–67
24. Guy J, Gan J, Selfridge J, Cobb S, Bird A 2007 Reversal of neurological defects in a mouse model of Rett Syndrome. *Science* 315:1143–1147
25. Cutfield WS, Hofman PL, Mitchell MD, Morison IM 2007 Could epigenetics play a role in the developmental origins of health and disease? *Pediatr Res* 61:68–75
26. Bernstein BE, Meissner A, Lander ES 2007 The mammalian epigenome. *Cell* 128:669–681