Epigenetic Code for Cell Fate During Development and Disease in Human

Selcen Celik Uzuner
Department of Molecular Biology and Genetics, Faculty of Science, Karadeniz Technical University, Trabzon, Türkiye

Abstract
Epigenetic reprogramming is the leading mechanism for cell differentiation in early development which gradually takes place upon zygote formation. This is governed by epigenetic modifications of genes involved in cell differentiation defined by Waddington's landscape. Somatic cells have specific gene expression profiles regulated by distinct epigenetic patterns. Therefore, they maintain their identity and specific gene profiles throughout lifetime. Although somatic cells can be induced into stem cell-like structures, the possible transformation of the cells can be associated with disruptions in cell identity leading to carcinogenesis. The epigenetic code for cell identity is the crucial player for maintaining stability and wellness of the cells during their lifespan. This review summarizes the epigenetic regulations involved in establishment of cellular fate and their abnormalities in cancer.

Keywords: Development, disease, epigenetics

Epigenetic Reprogramming
Human begins his/her journey with a single cell called zygote. Zygote includes all the genetic information from ancestries provided by maternal and paternal pronuclei. Parental genetic background is then combined by pronuclei fusion followed by cell cleavage. Embryogenesis consists of sequential rounds of cell proliferation, and cell proliferation also occurs along with cell differentiation to end up with specified cells that further form embryonic layers and organs. Epigenetic reprogramming is the main mechanism that gives cells different identities even if they have the same DNA code. This is managed by a well-organised epigenetic machinery including DNA and histone modifications.[1]

The cloning of Dolly the Sheep was the groundbreaking discovery in the late 90’s.[2,3] This discovery has opened the doors to an approach for understanding of reprogramming of differentiated cells into a new embryo. This showed that something like a ‘reverse evolution’ is technically possible in developmental biology suggesting that not only stem cells are differentiated into somatic cells, but also somatic cells can be differentiated to stem cells. This indicates that the genome of a somatic cell has the potential to return to the first stage of its life as zygote. The method used for cloning of Dolly was “somatic cell nuclear transfer” (SCNT). Soon after the number of in vitro studies using SCNT method have focused on understanding the detailed principles of epigenetic reprogramming at gene and genome level. Although the Dolly was successfully cloned and lived for 7 years, the rate of live birth in such SCNT is low. One of the main reasons is the failure of epigenetic establishment derived by donor and/or transferred nucleus and the abnormalities in epigenetic modifications in cloned embryos.
These epigenetic aberrations include i) errors in canonical and non-canonical genomic imprinting, ii) cell aging and iii) disruption in somatic cell-specific patterns of DNA and histone modifications.[4,5] Nevertheless, in vitro introduction of somatic cells into pluripotent stem cells provides explanations how a differentiated cell gains stem cell like characteristics suggesting that the cell fate is not necessarily unidirectional.

The sequential modification of specific genes either by DNA methylation or demethylation plays the critical role for the proper differentiation of cells.[6,7] Histone modifications also involve in the process of differentiation.[6,10] Therefore, the precise establishment of epigenetic patterns is a dynamic and inheritable mechanism in early development[8] and in adults it governs renewing some organs such liver.[10] Epigenetic pattern is a vital player from the very beginning of life, as cells gain their identity through epigenetic regulations. There was no life for multicellular organisms if there would no epigenetic reprogramming.

Epigenetic Modifications: The Key Players in Epigenetic Reprogramming

Epigenetics investigates inheritable but dynamic and reversible chemical modifications occurring on DNA and histones that regulate gene expression. The principles of epigenetics can provide explanations for questions that cannot be explained by classical Mendelian Genetics.

Epigenetic modifications regulate cell differentiation processes to form tissues/organs from a zygote. Although all cells in a human body technically have the same DNA sequences, each cell type has its own epigenetic patterns.

Epigenetic modifications are mainly classified into two groups, DNA modifications and histone modifications (Fig. 1). DNA modifications on cytosine includes four sub-groups i) methylation (5meC) catalysed by DNMT (DNA methyltransferase) enzymes, ii) hydroxymethylation (5hmC), iii) formylation (5fC) and iv) carboxylation (5caC) by the oxidation of 5meC with TET (Ten-eleven translocation) enzymes. The common belief is that these are not the only epigenetic modifications on DNA, but more modifications are supposed to be identified in the near future. DNA methylation was previously thought to be associated only with gene inactivation, however today we know well that the function of DNA methylation depends on the modified location within the gene such as in promoters or in gene bodies.[11,12] Methylation at promoter regions is mainly associated with gene inactivation, however the methylation at gene bodies is associated with gene activation.[11-13] The other factor influencing the function of DNA methylation is the existence of methylation within the CpG repeats or not (such in shores, shelves and non-CpG regions) throughout the genome.[11,14] (Fig. 1, left panel).

On the other hand, histones, the specific proteins composing chromatin structure, are the other critical player in epigenome. Histones can be modified by acetylation, methylation, phosphorylation, ubiquitination and SUMOylation.[15] Modifying group (methyl, acetyl etc.) is definitive for the biological function of modification. However, the type of modified histone (H2A, H2B, H3 and H4), the number of modifying groups (such di- or trimethylation) and the type of modified amino acid (such lysine or arginine) are also extremely important for the regulatory effects of histone modifications on gene expression profiles.[15] (Fig. 1, right panel). Histone modifications, in particular lysine acetylation, also specifically play critical roles in cellular activities including chromatin remodeling, cell cycle and splicing.

DNA and histone modifications do not independently function of each other and it is well-known that there is a dynamic crosstalk between the two machineries for the establishment, maintenance and regulation of gene expressions.[17,18] This fine-tuning of epigenetic modifications adds further complexity to understanding epigenetic reprogramming in normal and pathological processes.

Epigenetic modifications are dynamically regulated during early development. A wave of epigenetic programming occurs in the germ cell line to ensure that genomic imprinting is truly set. Genomic imprinting presents a model for monoallelic expression of some genes based on parent-of-origin. Genomic imprinting is regulated not only by DNA methylation but also histone modifications, particularly by H3K27me3. [19] Failure to maintain genomic imprinting is associated with a range of abnormalities in human.[20,21]

The paternal DNA methylation level is more than the maternal DNA methylation level at fertilisation. However in zygote, paternal and maternal methylation levels in pronuclei are found to be similar.[11] After fertilisation, the level of 5meC decreases by the oxidation mechanism catalysed...
by TET3 enzyme and accordingly the levels of the oxidised forms 5hmC, 5fC/5caC increase.[22] All cytosine modifications are then progressively erased until the blastula stage. TET1 and TET2 subsequently catalyse the increases in 5meC and 5hmC in blastula,[22] The understanding on reprogramming of other cytosine modifications (5hmC, 5fC and 5caC) remains more elusive than 5meC programming. After fertilisation, de novo methylation also occurs, and de novo methylation continues until implantation. After implantation of the embryo into the uterus, the embryo has its own epigenetic profiles including X chromosome inactivation[23] and tissue specific gene expression profiles.[24] This epigenome profile is established not only by DNA methylation and but also histone modifications.[25] (Fig. 2A)

One of the main objectives of epigenetic reprogramming is to form organs through the cell differentiation. This process is hypothetically explained by Waddington’s landscape.[26] (Fig. 2B) This model suggests the progressive limitation of stem cell characteristics but inducement of tissue specific patterns. The cell fate is managed by epigenetic landscape. Each organ/cell has the same DNA code but the insulin protein, for instance, is only expressed in beta cells, a very specific pancreatic cell. Therefore, cells in a human body have the same genome, but hundreds/thousands of different epigenomes.

The epigenetic profiles of cells can be altered by environmental conditions. These conditions mainly refer to lifestyle such as eating habits,[27] educational attainment,[28] socioeconomic position,[29] psychiatric status[30,31] and smoking.[28] The field of "Environmental Epigenetics" investigates the epigenetic influence of human choices and external conditions. This is quite interesting that these epigenetic changes can be reversed after causal conditions, such as smoking, have ceased.[32] Reversible behaviour of epigenetic patterns on the genes offers a more dynamic and more manageable model for reprogramming than mutations.

A Code in the Code: Cellular Memory, Cell Identity and Tissue-Specific Epigenomes

A human has a DNA (genome) that is identical in all cells, but there are numerous epigenomes that vary in different tissues (even within a tissue). This phenomenon suggests that there is another code in the DNA code, ‘A code in the code’, governed by epigenetic regulations. ‘A code in the code’ hypothesis defines the specific epigenetic profile for maintenance of cell identity, and somatic cells are therefore aware of which cell they are during their lifespan. The specific identity is established and characterised by specific gene expression and epigenomic profiles.[33–35] Somatic cells have an intelligence (a code) for managing themselves in terms of their identity. Somatic cells may be capable of "epigenetic reprogramming" even though they have already established their identity. But their ability to epigenetic reprogramming remains unclear.

Cellular memory is a phenomenon which is generally considered to be a property of immune cells or neurons. For instance, specialized immune cells called T and B cells (cytotoxic memory cells) can learn and memorize the information about antigens. Therefore, they reveal cytotoxic properties against these antigens/pathogens more easily. This pathogen knowledge can be inherited to the next generations to maintain the efficacy of the immune system cells.[36] But cellular memory of immunogenicity is not limited to specific immune cells, even fibroblasts in connective tissues have acquired a learned immunity against pathogenic microorganisms, and thus these cells have functions in the adaptive immune system.[37,38] Fibroblasts contain 1-10 receptors (TLRs, toll like receptors) that can recognize different microbial structures and activate immune system cells.[38] Fibroblasts are also involved in the repair of damaged cells by obtaining structural information from non-damaged cells in tissue damage.[39]“Cellular memory” extensively represents the ability of cells to be aware of what type of cells they are throughout their lifespans. Cellular memory is also called "transcriptional memory."[44,45] Each cell type has its own transcriptional memory represented by diverse profiles of gene expressions.[13] Memory for cell identity is shaped by epigenetic rearrangements in certain gene groups, and this memory is inherited to the next generations throughout the cell cycle.[46–48] Transcription factors are important for tissue-specific expression of genes by methylation-mediated specificity.[43] These transcription factors include for instance, HOXB13,
CDX1 and CDX2, which have affinity for binding methylated cytosines.43 On the other hand, FoxA genes play a role for tissue-specific expression in hepatocytes and the methylation pattern of these genes is the key regulator for specific expression.44 Interestingly, muscle cells have a memory of former physical activity, thus they have a high adaptation to retraining, and not surprisingly this memory is governed by epigenetic mechanisms, called ‘epi-memory’.45 Addition of neural-lineage transcription factor cocktail consisting of Ascl1, Brn2 and Myt1l converts fibroblasts to neuron cells.46 Another transcription factor, NeuroD1, is also eligible to transform microglial cells into the neurons by mediating the histone alterations.47 These suggest that reactivation of transcription factors by epigenetic regulations may alter cell fate, and somatic cells are able to be reprogrammed in vitro. But their capacity to reverse in vitro epigenetic manipulations remains unclear.

The pattern of DNA methylation is the key player to maintain tissue-specific expression in different types of cells. Lineage specific methylation governs the specification of the cells that further forms tissues and/or organs. This is also a defining factor to sub-specify cells derived in the same tissues. For instance immune system cells have distinct methylome patterns, and these patterns are associated with gene expression profiles.48,49 The cells functioning in adaptive or innate immune response reveals different methylome marks.49 Integrative analyses of omics data can provide details for understanding the cell-type specific characteristics, and this approach will further suggest predictive models for complex diseases.48 Interestingly, DNA methylation is also associated with alternative splicing in a tissue-specific manner and this association may also be associated with abnormalities in diseases.50

Epigenetic Dysregulation Involved in Diseases

A line of evidence suggests that pathological processes involve abnormal patterns of epigenetic modifications.51–53 Tumour suppressor genes and proto-oncogenes have been found to be mostly up- or downregulated by alterations in DNA methylation.54–56 For instance a melanoma associated protein is upregulated by epigenetic mechanisms in an aggressive form of breast cancer.51 Global hypomethylation is also commonly detected in different cancers,57 but some cancers such as hepatocellular carcinoma, are significant with global hypermethylation.58 Driver genes including tumour suppressor and proto-oncogenes are mostly found hypermethylated or hypomethylated in cancers, respectively.59 Gene specific changes in DNA methylation are also highly informative for classification of cancers and estimation of the original tissue that cancer develops.60 Neurodegeneration is also associated with the epigenetic changes in some genes.61 Some congenital diseases are associated with abnormalities in epigenetic code. For instance, some errors in genomic imprinting are related to Prader-Willi and Angelman syndromes.62–64 DNMT1 depletion is lethal for human.62 The lethal mutations in DNMT1 gene therefore resulted in abortion or foetus death in utero. Embryo cannot develop if DNMT1 enzyme is absent because cells can not differentiate and therefore not form organs. A cell does not value itself without any identity in multicellular organisms. Overexpression of DNMT1 gene is also associated with lethality in embryos or errors in genomic imprinting resulting in congenital abnormalities.65 Increase in DNMT1 activity probably relates to misidentification of the cells.

From Epigenetic Reprogramming to Epigenetic Targeting Drugs: A New Era in Therapy of Human Diseases

In some cases, the epigenetic changes that occur in a range of abnormalities such as cancer64 and neuropsychiatric diseases65 can be reversed by epigenetic drugs. Epigenetic targeting drugs, also known as ‘epi-drugs’, have been rather a new trend in disease treatments. These drugs, such 5’-aza-2’-deoxycytidine (Decitabine), have been widely used in cancer, and new drug candidates are in demand to discover more specific and effective drugs with minimum side-effects.66–68 Low dose Decitabine is not cytotoxic for normal peripheral blood cells,69 but are particularly cytotoxic to natural killer cells.70 The main interest in epigenetic-based alterations of the genome has focused on carcinogenesis but cancer is not the only disease that is associated with epigenetic abnormalities. The diseases related to errors in genomic imprinting highlight the importance of correct establishment of epigenetic marks in the genome and there are epigenetic therapy approaches being considered for the treatment of genomic imprinting abnormalities.71 It is important to maintain the paternal-based expression of alleles and this mechanism is governed by DNA methylation in the imprinted genes.

There are four main groups of proteins that regulate epigenetic modifications on DNA and histones; 1) writers, 2) erasers, 3) modulator proteins, and 4) mediator proteins (Fig. 3). Writer proteins are the enzymes, such as DNMTs that add methyl groups to DNA, histone acetyltransferase (HATs) and histone methyltransferase (HMTs) enzymes that add acetyl and methyl groups to histones, respectively. Eraser proteins are the enzymes such as histone deacetylases (HDACs) removing the acetyl group from histones and CpG demethylase or TET1 removing the methyl group from DNA. Writer and eraser proteins are also called “Epigenetic Modifying Proteins”. Modulatory proteins are the
proteins that directly alter the epigenome through DNA methylation, histone modification, or structural changes of chromatin. For example, chromatin-modifying and re-modelling proteins are the modulator proteins. Mediator proteins are downstream targets of epigenetic modifications, and pluripotency factors such as OCT4, NANOG and SOX2.[73] These proteins collectively play role for epigenetic reprogramming in the cells.

Epigenetic drugs are mainly classified into two groups: i) DNA methylation inhibitors and ii) histone modification inhibitors (Fig. 3). A large group of these drugs involves enzyme inhibitors targeting DNMT1, HDAC or HAT enzymes to reverse the epigenetic marks.[74] DNMT1 inhibitors, which will be used in this project, include nucleoside or non-nucleoside analogues (Fig. 4). Nucleoside analogues, such 5’-azacytidine and 5’-deoxy-2’-azacytidine mimic the structure of cytidine, integrating into newly replicated DNA instead of methylated cytosine. These induce a gradual decrease in DNA methylation by each DNA replication. But non-nucleoside analogues act directly as inhibitors of DNMT1 enzyme to lower DNA methylation levels (Fig. 4). Although some of the DNMT1 inhibitors are clinically approved and used for treatment of a range of diseases, particularly in some cancers,[75,76] the main handicap of these drugs is that they do not target the specific sites in the genome, leading to undesired reprogramming in the epigenetic landscape. To deal with this limitation gene-specific approaches have been developed, including CRISPR-Cas9 and TET-based methods that can block DNA methylation and histone modifications in vitro on the specific sites.[77,78]

The Missing Part of the Puzzle: Reprogramming Potential in Differentiated Cells

There are many studies mostly focused on how epigenetic reprogramming occurs in early development from stem cells to differentiated cells, and how induced reprogram-

Figure 3. Proteins involved in epigenetic modifications and epigenetic targeting drugs (ETDs).

Figure 4. Mechanism of action of DNMT1 inhibitors (nucleoside and non-nucleoside analogous).
39. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-