Chapter 7
Epigenetics and Medical Biotechnology

Mumtaz Anwar, Samia Rashid, and Zeenat Farooq

Abstract The field of epigenetics deals with the study of processes that lead to changes in gene expression without changing the underlying DNA sequence. These processes include DNA methylation, post-translational modifications of histones as well as non-coding RNAs. This field has undergone enormous progress over the years, especially during the last two decades with the introduction of high-throughput sequencing and computational biology. This magnitude of progress has enabled research in the area that uncovered its role in various key cellular processes and disease conditions alike. Epigenetics is increasingly being pursued as a discipline of study to understand diseases like diabetes, cancer, neurological abnormalities and physiological process like circadian rhythms, behaviour and homeostasis. This chapter discusses the role of epigenetics in various aspects of medical biotechnology. It includes sections on different techniques used in epigenetics, their application in medical biotechnology and role of epigenetics in various diseases.

Keywords Biotechnology applications · Technology · Stem cells · CRISPR · Genetic engineering

M. Anwar (✉) · Z. Farooq (✉)
Department of Pharmacology and Regenerative Medicine, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA
e-mail: mumtazan@uic.edu

S. Rashid
Government Medical College Srinagar, Srinagar, Jammu and Kashmir, India

M. Anwar et al. (eds.), *Fundamentals and Advances in Medical Biotechnology*,
https://doi.org/10.1007/978-3-030-98554-7_7

7.1 Introduction to Epigenetics

7.1.1 Epigenetics: Background

Transmission of characters in a stable manner from one cell generation to another is governed by the macromolecule DNA which contains all the information necessary for synthesis of cellular proteins and hence controls all the vital cellular processes. The sequence of amino acids in the proteins is dictated by the sequence of nucleotides within DNA. This DNA, if measured from end to end in a typical human cell, measures about 2 m. Therefore, cells have developed efficient mechanisms to compact the DNA inside the small nuclear space. This compaction is achieved by the association of negatively charged DNA with the positively charged proteins called as histones. These are small, basic proteins (~12–17 kDa in size) which bind to DNA to neutralise its charge and therefore ensure efficient packaging. Various classes of histone proteins have been identified so far [1, 2]. Various other proteins also play a role in DNA compaction but those remain less well-characterised.

The prokaryotic cells do not have the same level of orderly arrangement of DNA–protein complexes as eukaryotes. They, however, contain a less organised DNA–protein complex called "nucleoid".

During the earlier part of the century, a pre-dominant concept among the researchers was that the complexing of DNA with histones only occurs for accommodation into the tiny nucleus. However, it was later realised that the complex is much more structured and organised than initially anticipated. This organisation allows for what later came to be known as the "*differential patterns of gene expression*".

7.1.2 History and Definition of Epigenetics

The term Epigenetics was coined by Waddington in 1953 to describe "changes in the gene expression that occur independently of the changes in the sequence of DNA". These "*changes*" are propagated by covalent histone modifications, covalent DNA modifications and micro-RNAs.

The original concept behind this definition has its roots in the studies carried out much earlier in the nineteenth century in cell biology and embryology. These studies laid the foundation for our present day understanding of gene expression and transmission of characters [3].

There is a long history behind the discovery of Epigenetics and in all the years that followed since its discovery, the definition has evolved a lot. Our present definition of the term reflects the understanding that although the nature of DNA is the same in many different cell types in an organism, the difference in epigenetics has the potential to give each cell type its own identity, pattern of gene expression and the power for vertical transmission. This understanding has led to a very interesting

"working definition" of the term epigenetics as "*the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence*" [4].

7.1.3 Evolution of the Definition of Epigenetics: Insights from Research Involved

Following the discovery of chromosomes by the German Biologist Walther Flemming in 1879, strong evidence were provided by many investigators about developmental program being governed by chromosomes. The most significant demonstration was provided by Thomas Hunt Morgan in 1911 about the "genetic linkage" of several Drosophila genes to the X-chromosome [5]. Rapid progress followed in the aftermath of this discovery in creating linear chromosome maps of Drosophila in which various gens were assigned specific positions linearly on the chromosome with respect to each other [6]. The discovery that chromosomes contained both nucleic acids and proteins sparked the debate about the contribution of both these biomolecules in carrying genetic information. Furthermore, evidence from Drosophila genetics suggested the occurrence of changes in phenotype without similar changes observed in genotype which apparently supported the notion that "cytoplasm" also plays a role in dictating "developmental plans". However, this debate was silenced by the discovery of DNA as the primary carrier of genetic information. Nevertheless, with advancements in technology which permitted cutting edge research into molecular genetics and development, the role of the "non-DNA" elements in governing development became more and more apparent and obvious. Today, each of the two elements of chromatin "DNA and proteins" is known to play key respective roles in dictating "epigenetic and developmental landscapes".

Eventually, it became both convenient and useful to define epigenetics as the mechanisms which propagate heritable changes without changing the sequence of DNA itself.

Quite often, the epigenetic and non-epigenetic components are intertwined and hence a clearer description of individual epigenetic carriers is useful to understand the entire concept (Fig. 7.1).

7.1.4 Histones and Nucleosomes

Five major classes of histone proteins have been identified. These are known as H1, H2A, H2B, H3 and H4. Two copies, each of histones H2A, H2B, H3 and H4 come together and are wrapped by ~147 bp of DNA to give rise to "nucleosome". Presence of H2A, H2B, H3 and H4 inside the nucleosome core lends the name core histones to these proteins [7, 8]. Nucleosome structure is the fundamental unit of chromatin

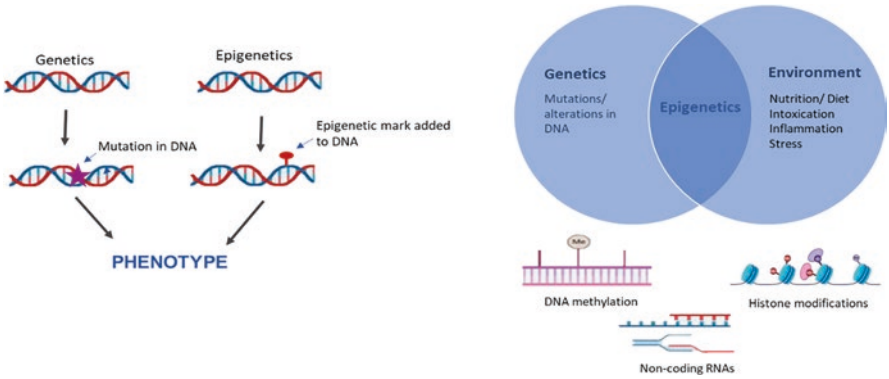


Fig. 7.1 Effect of genetic and epigenetic factors on phenotype. The figure shows the difference between the mechanism behind propagation of genetic and epigenetic factors in the effect on phenotype. (**a**) Shows that genetics affects the phenotype through mutations which alter DNA sequence, while epigenetics affects phenotype without changing DNA sequence. (**b**) Shows how genetics and environment come together to regulate epigenetics to affect phenotype. Most prominent mechanisms of epigenetic progression (DNA methylation, histone modifications and non-coding RNAs) are also depicted

organisation identified so far [9]. Histone H1 binds at the intervening DNA between adjacent nucleosomes (known as entry and exit sites) and is therefore known as the linker histone. H1 ensures further compaction of chromatin beyond nucleosomes into higher order structures like beads on a string (30 nm fibre), etc. Linker histone found in aves is known as H5 (Fig. 7.2).

7.1.5 Post-translational Modifications of Histones

In their 3-dimensional structure, histone proteins consist of a globular domain at C-terminus and an unstructured tail at their N-terminus. The globular domain is responsible for DNA–DNA and DNA–protein interactions and the tail acts as a site for different covalent modifications, collectively and famously known as histone modifications. These modification marks are mostly found in histones H3 and H4 [10]. Till date, more than 100 different types of modifications have been described, including acetylation, methylation, ubiquitination, phosphorylation, crotonylation, sumoylation [10]. In 2011, the graph of known histone modifications was raised by about 70% through identification of 67 previously uncharacterised histone modifications in a single report. Interestingly, eight of these were found to be "short-chain lysine acylations" (propionylation, butyrylation, 2-hydroxyisobutyrylation, succinylation, malonylation, glutarylation, crotonylation and β-hydroxybutyrylation) [11].

We shall describe the most prominent class of histone modifications in the next section in some more detail, which are

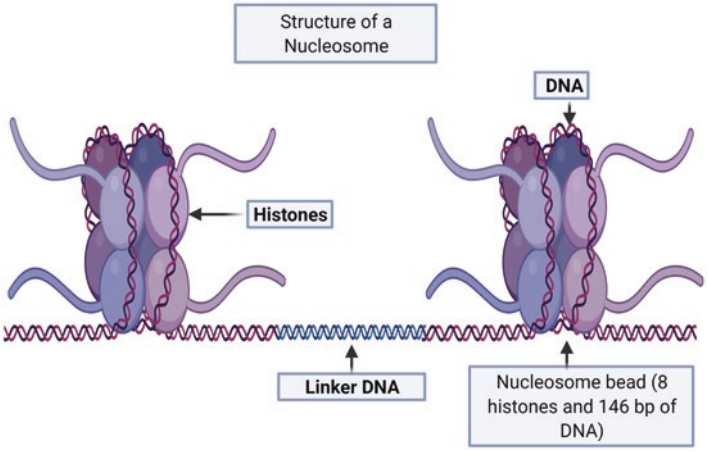


Fig. 7.2 Organisation of a nucleosome. The figure shows a nucleosome composed of two units of each core histone wrapped by 147 bp of DNA to form a nucleosome. The DNA between adjacent nucleosomes is bound by linker histone H1

- Phosphorylation of serine (S) and threonine (T) residues.
- Methylation of lysine (K) and arginine (R) residues.
- Acetylation of lysine (K) residues and.
- Ubiquitination of lysine (K) residues [11].

These modifications can act in various ways to either increase accessibility to DNA and gene expression or to decrease it.

1. Alterations in the charge of histone to change the electrostatic (negative: positive) attractions between DNA and histones, respectively. For example, acetylation neutralises the positive charge of histones, thereby decreasing DNA–histone interactions. This leads to an increase in the accessibility of DNA for processes like replication, transcription, etc.
2. Some modifications act as a signal (docking molecule) for other proteins to bind chromatin. This can influence chromatin structure and gene expression in two ways.

   (a) Modifications that act as signalling platform for proteins which promote chromatin condensation. For example, methylation on histone H3 on lysine 9 residue (H3K9) acts as a docking site for the binding of heterochromatin protein (HP1) which increases heterochromatinisation.
   (b) Modifications that act as signalling platform for proteins which promote chromatin de-condensation. For example, acetylation of H4K16 acts as a binding platform for BRD2. This protein is a transcriptional co-activator and hence enhances transcription at the de-condensed chromatin regions created by H4K16 acetylation.

We will talk about each of these modifications and their function in slightly more detail in the next section.

7.1.5.1 Histone Methylation

In this process, a methyl group is added to histones by the enzymes known as histone methyltransferases, using *S*-adenosyl methionine as the methyl group donor. Histone methylation is a neutral modification. Being non-polar in nature, methylation enhances the hydrophobic interactions inside chromatin. Methylation is the most complex histone modification because lysine can undergo three degrees of methylation (mono, di or trimethylation) and arginine can undergo two degrees (mono and dimethylation). Also, it is the most stable modification with the slowest turnover. Methylation acts as a binding site for various proteins which promote different functions, depending upon the residue being methylated and the enzyme involved. Histone methylation takes place on lysine and arginine residues on the histones H3 and H4. Methylation on H3K4, H3K36 and H3K79 plays role in gene activation and on H3K9, H3K27 and H4K20 play role in gene silencing [7, 12, 13].

7.1.5.2 Histone Acetylation

It is the most well-studied histone modification which involves addition of an acetyl group from acetyl-CoA to histones with the help of histone acetyltransferase enzymes. The negative charge on the acetyl group neutralises the positive charge on histone and hence decreases histone–DNA interactions [14]. Therefore, acetylation is involved in chromatin de-condensation and gene expression. For this reason, various transcriptional activator complexes contain histone acetylation enzymes as members of the complex like the CBP 300 and TAF II 250 [15].

7.1.5.3 Histone Phosphorylation

Phosphorylation of histones involves addition of a phosphate group on histone with the aid of enzymes known as kinases (phosphotransferases) which transfer a phosphate group from ATP to the hydroxyl side chain of the amino acid residues serine, threonine, tyrosine and histidine. Phosphorylation adds sufficient negative charge on chromatin and disrupts histone:DNA interactions to promote gene expression. In fact, various transcriptional up regulators contain phosphor-binding domains [16, 17]. However, histone phosphorylation is a versatile modification. For example, the same modification might be involved in transcriptional upregulation in one context and with condensation in another. This is best exemplified by H3Ser10 phosphorylation which performs a dual role of activation on the one hand and of chromosome condensation and segregation on another [18]. Phosphorylation status changes dynamically with the gene expression profile [19]. Phosphorylation is reversed by

another class of enzymes known as phosphatases. Phosphorylation is a very prominent post-translational protein modification mark, and it plays a role in various cellular signalling pathways by employing phosphorylation–de-phosphorylation switches like mTOR pathway, p53 signalling pathway, ISR pathway, etc. In fact, phosphorylation (including that beyond histones) is the most well-studied cellular post-translational protein modification. In non-histone proteins, phosphorylation mostly takes place on serine residues.

7.1.5.4 Histone Ubiquitination

It is a process in which ubiquitin molecules are added to lysine residues of histones. Monoubiquitination is the major form of ubiquitination in histones. However, histones H2A and H2B can also be modified by polyubiquitination. The first ubiquitinated histone to be identified was H2A [20]. H2A and H2B also hold the distinction of being the most abundantly ubiquitinated proteins in the nucleus [20, 21]. In addition, H3, H4 as well as H1 have been reported to be modified by ubiquitin but the biological function of these ubiquitinations has not been well characterised [22, 23]. Histone ubiquitinations perform several important nucleosomal functions. Chromatin immunoprecipitation (ChIP) experiments have revealed enrichment of monoubiquitinated H2A (H2Aub) in the satellite regions of genome and of H2Bub in transcriptionally active genes [24, 25].

Histone ubiquitination involves addition of the 76-amino acid protein ubiquitin to the lysine residues of histones with the help of enzymes known as ubiquitin ligases. Unlike other histone modifications, ubiquitination is the only modification in which the substrate is a protein rather than an inorganic molecule. Polyubiquitination was well known before the discovery of histone ubiquitination as a mechanism behind proteasomal protein degradation. However, histones are generally modified by mono-ubiquitination. In this case, this "*mark*" acts as a signalling platform for further modifications or downstream chromatin effector functions rather than as a "*signal*" for degradation. Histone ubiquitination mostly takes place on H2A and H2B. Histones H1, H3 and H4 have also been reported to be modified by ubiquitin, with the roles of these modifications being less well characterised [20–23] (Table 7.1).

7.1.6 DNA Modifications

In addition to histones, covalent modifications also take place on DNA itself, expanding the horizon of "*epigenetic programming*" of nuclear functions. DNA can be modified by addition of hydroxy-methyl cytosine and methylcytosine. However, hydroxymethylation remains a less significant modification, with only being identified in limited circumstances [29]. 5′ cytosine methylation involves transfer of a methyl group from *S*-adenosylmethionine to the 5′ position of cytosine residues in

Table 7.1 Outline of the major histone modifications and their effects

Modification	Charge	Effect on gene expression	Major histone modified	Major functions	References
Methylation	Neutral	Increases or decreases	H3, H4	Heterochromatin organisation, gene silencing	[26, 27]
Acetylation	Negative	Increases	H3, H4	Gene activation	[15, 28]
Phosphorylation	Negative	Increases	H2A, H3	DNA repair, gene activation	[16, 17, 19]
Ubiquitination	Neutral	Variable	H2A, H2B	Satellite DNA binding, transcriptional activation	[20–24]

A more detailed review of histone modifications has been done elsewhere

DNA and is the most well-understood DNA modification and one of the most prominent epigenetic events. This modification has been shown to play important roles in various cellular processes like genome integrity, genome imprinting, X-chromosome inactivation, development [30, 31] and diseases like CVD and chronic kidney disease [29, 32–34].

5′ cytosine methylation of DNA is carried out by two groups of enzymes.

1. *De novo methyltransferases*. This group involves the enzymes DNMT 3a and 3b. These enzymes methylate cytosine on the 5′ position, without utilisation of a template to copy the methylation pattern, hence the name. Therefore, these enzymes display high level of expression during embryogenesis and gradually diminish in adult tissues [35]. Studies performed in mice shown that mice deficient in DNMT 3b are embryonic lethal and in DNMT 3a survive for only up to 4 weeks [36].
2. *Maintenance methyltransferases*. This group involves the enzyme DNMT1 which methylates hemi-methylated DNA by utilising methylation pattern information of parent DNA strand to copy the same on daughter DNA strand following DNA replication. In this manner, it helps in the "*maintenance*" of "*DNA methylation signature*" following cell division. DNMT1 is ubiquitously expressed and is critically important for mammals as mice deficient in DMNT1 display embryonic lethality [36].

7.2 Epigenome and Diseases

7.2.1 Role of DNA Methylation in Diseases

DNA methylation is a long-term, relatively stable, epigenetic trait that contributes to maintaining the cellular homeostasis and mis-regulations in which can lead to diseases. DNA methylation predominantly takes place in the context of concentrated regions of CG dinucleotides, known as CpG islands. These islands are found

in the promoters of many genes and mostly remain unmethylated to allow gene expression (e.g., promoters of housekeeping genes like glyceraldehyde 3′ phosphate dehydrogenase (GAPDH) are rich in CpG islands in their promoters) and methylation of cytosine decreases the expression of the associated downstream genes. There are two ways in which CpG methylation reduces transcription.

1. The presence of methyl group blocks the binding of some transcription factors. For example, sp1 and sp3 transcription factors and Hif1 transcription factor in hypoxia [31, 35, 36].
2. Methyl DNA binding factors promote the binding of silencing factors. For example, MeCP2 binds methyl DNA and recruit histone deacetylases (HDACs) which promote chromatin condensation and decrease gene expression [29, 37, 38].

7.2.2 DNA Methylation and Human Diseases

Alterations in DNA methylation are responsible for a large number of human diseases, including cancer (see Table 7.2). These alterations take place as a result of the mechanisms described below.

Change in DNA methylation status. Disturbances in DNA methylation can typically lead to either overexpression of some genes like oncogenes or reduced expression of some others like tumour suppressor genes (see Fig. 7.3). A large number of mis-regulations in DNA methylation have been identified in various forms of cancer.

Cross talk between genetic and epigenetic mechanism. Another way in which CpG methylation can contribute to disease is through single nucleotide polymorphism (SNP) in CpG sites. This change links a genetic mechanism with an epigenetic mechanism and in consequence can lead to a drastic alteration of the normal function of an associated gene. For example, the gene encoding for the respiratory

Table 7.2 Changes in DNA methylation of different genes in different forms of cancer

S. No.	Gene name	Cancer type	Gene expression change	DNA methylation change
1.	P16	Renal, lung, colorectal, oral, head and neck, hepatic	Decrease	Increase
2.	P cadherin	Hepatic, breast, pancreatic, salivary gland, lung	Decrease	Increase
3.	Hmlh1 and hMSH2	Colorectal, renal	Decrease	Increase
4.	Cyclin D2	Gastric		Decrease
5.	P15	Oral carcinoma	Decrease	Increase
6.	APC	Lung, colorectal	Decrease	Increase
7.	FHIT	Lung	Decrease	Increase

Adapted with permission from Zeenat et al. (DOI: 10.5772/intechopen.97379) [40–44]

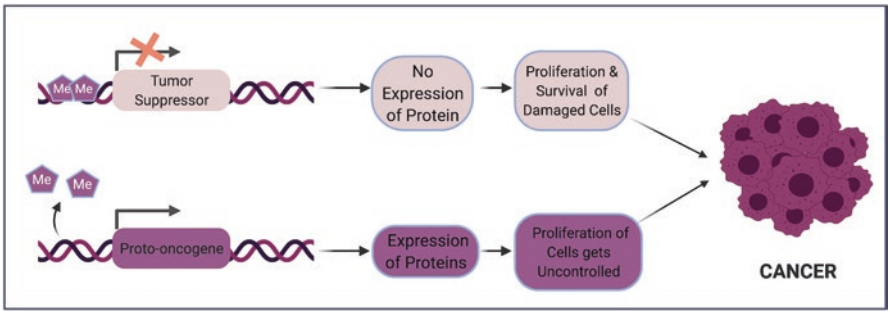


Fig. 7.3 Schematic of role of DNA methylation in cancer progression through two pathways. Hypermethylation of tumor suppressor genes allows unchecked growth and accumulation of damaged cells to generate cancer phenotype. Hypomethylation of proto-oncogenes favours uncontrolled proliferation of cells to generate cancer cell mass. Adapted and modified from Zeenat et al. (DOI: 10.5772/intechopen.97379)

chain protein NDUFB6 is a target of SNPs in a CpG site in its promoter. This gene has reduced levels of expression in type 2 diabetes. The level of expression of this gene is also found to be inversely correlated with the degree of promoter methylation in muscle biopsy samples from elderly patients [39].

7.2.3 High-Throughput Methylome Sequencing

This technique involves identification of DNA methylation pattern on a genome wide level. There are many variant techniques available to perform it but two of these techniques are the most widespread.

7.2.3.1 Bisulphite Sequencing

In essence, this technique involves treatment of DNA with a bisulphite reagent which converts non-methylated cytosine residues to uracil and leaves methylated cytosine unaltered. After performing PCR on the bisulphite converted DNA, the uracil is subsequently replaced by thymine (T) from dNTPs. Thus, by comparing the sequence of converted DNA with original DNA, the position of methylated cytosine residues on DNA can be determined. This technique is either performed on a single gene level, employing bisulphite treatment followed by PCR or at the genome wide level, employing bisulphite treatment followed by high-throughput sequencing.

Table 7.3 Shows a comparison of the variant genome sequencing techniques which employ bisulphite conversion of DNA

S. No.	Bisulphite sequencing variant	Procedural steps involved	Input DNA requirement (μg)	Coverage (%)
1.	Whole genome bilsulphite sequencing (WGBS)	Sonication of DNA, library preparation, gel-size selection, bisulphite treatment, library amplification	1–5	~95
2.	Methyl binding domain (MBD) cap sequencing	Sonication of DNA, capture 5mC by MBD, library preparation Gel-size selection, library amplification	0.2–1.0	~17.8
3.	Reduced representation bisulphite sequencing (RRBS)	Digestion with MspI, library preparation, gel-size selection Bisulphite treatment, library generation	0.01–0.03	~3.7

An outline comparison of steps involved, DNA requirement and sequence coverage has been shown

This is the only technique with a single base resolution and therefore represents the *gold standard* of studying cytosine methylation. However, the bisulphite treatment is a harsh chemical procedure which heavily degrades DNA and makes recovery difficult. A number of variants of the technique have been developed like HELP, COBRA, bisulphite pyrosequencing, etc. A comparison of some of the most commonly used variants of bisulphite sequencing is presented in the Table 7.3 [45, 46].

Methylated DNA immunoprecipitation sequencing (MeDIP seq). This technique involves immunocapture of methylated DNA fragments, utilising an antibody against 5′ methyl cytosine. In this way, DNA fragments containing methylated cytosines become enriched and can be analysed either by quantitative PCR for a single gene or by high-throughput sequencing for the entire genome. This technique is analogous to chromatin immunoprecipitation (ChIP). Although the resolution of this technique is low compared to bisulphite sequencing, it is more versatile and reproducible (Fig. 7.4).

7.2.4 Role of Methylome Sequencing in Diseases Prognosis and/or Diagnosis

Many diseases have been documented in which cytosine methylation remains a major reason for the problem. In most of these conditions, single gene methylation changes have been reported behind the problem. However, methylome sequencing could serve as an even more useful tool that yields high quality genome wide results

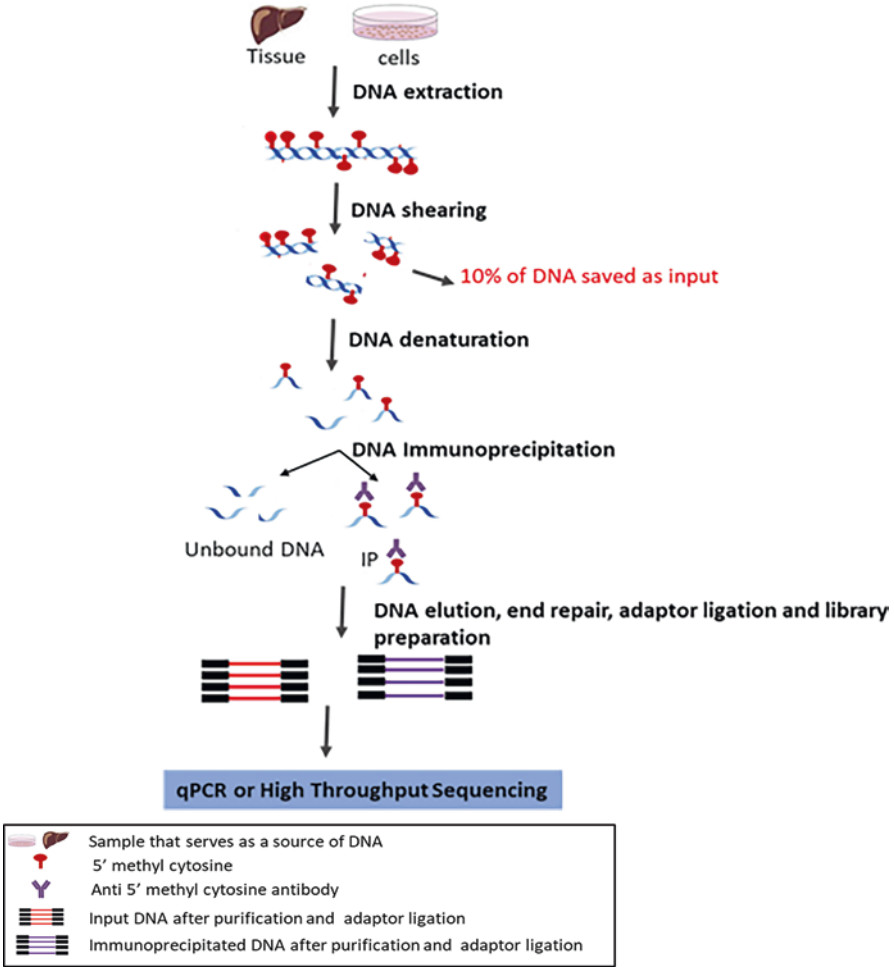


Fig. 7.4 Shows the major steps of methylated DNA immunoprecipitation. DNA is extracted from a sample followed by shearing and denaturation to generate small, single stranded DNA fragments (200–600 bp). This is followed by pull down of methylated DNA fraction with the help of antibody. Both the input DNA (a fraction of total genomic DNA used for normalisation) and immunoprecipitated (IP) DNA are purified, end-repaired (and ligated with adaptors for sequencing) and analysed after performing qPCR or sequencing

when comparing two different conditions side by side. For example, cytosine methylation of a normal versus tumour sample. This can give us an idea of the role of DNA methylation in imparting the particular disease and hence could lead to better ways of understanding the diseases and development of therapeutics [45, 47–49].

7.2.5 Chromatin Immunoprecipitation

In principle, chromatin immunoprecipitation is a technique in which chromatin is cross-linked with the help of formaldehyde followed by its isolation from the cells. This cross-linking covalently links chromatin associated proteins (like histones) to DNA. This is followed by immunoprecipitation of the sheared chromatin fragments using the antibody of interest to enrich these fragments. For example, to study methylation status of lysine residue 9 on histone H3 (known as H3K9me), crosslinked and sheared chromatin is immunoprecipitated with anti-H3K9me antibody to enrich the fragments containing this modification versus the fragments which do not. Subsequently, the level of H3K9me of an individual gene can be observed by performing quantitative PCR on that gene. This can be better understood by taking the example of the tumour suppressor gene X. In this situation, we want to study H3K9 methylation on two different tissue samples, normal and tumour for gene X. We begin by isolating chromatin from each sample and performing H3K9me ChIP. Following immunoprecipitation, we perform qPCR in both tissues for the gene X.

To study the effect of any chromatin modification on a genome wide level, Chromatin immunoprecipitation can also be followed by high-throughput sequencing (**ChIP Seq**) (Fig. 7.5).

7.2.6 Use of ChIP in Diagnostic and Prognostic Applications

In recent years, chromatin immunoprecipitation has been used in varied clinical applications for diagnostic, prognostic and research purposes. Some of the most exciting clinical applications of the technique are listed below.

1. Chromatin immunoprecipitation in cell free ChIP (cf ChIP).

   This technique relies on immunocapturing and sequencing of cell free chromatin fragments from different cells. In this technique, chromatin fragments released into the bloodstream from different cells are immunocaptured in the plasma. After washing and purification of DNA fragments isolated after digestion of chromatin, libraries are created and sequenced. Cell free (cf) chromatin maps generated for various chromatin modifications are rich in important information and such sequences are available, with plasma obtained from various volunteer donors. These maps can vary between healthy and diseased conditions and hence can serve as good molecular indicators for disease conditions. These maps can not only be used to identify differences in chromatin associated modifications (e.g., histone post-translational modifications) in different genes, but can also be used to compare similar modifications in other genomic

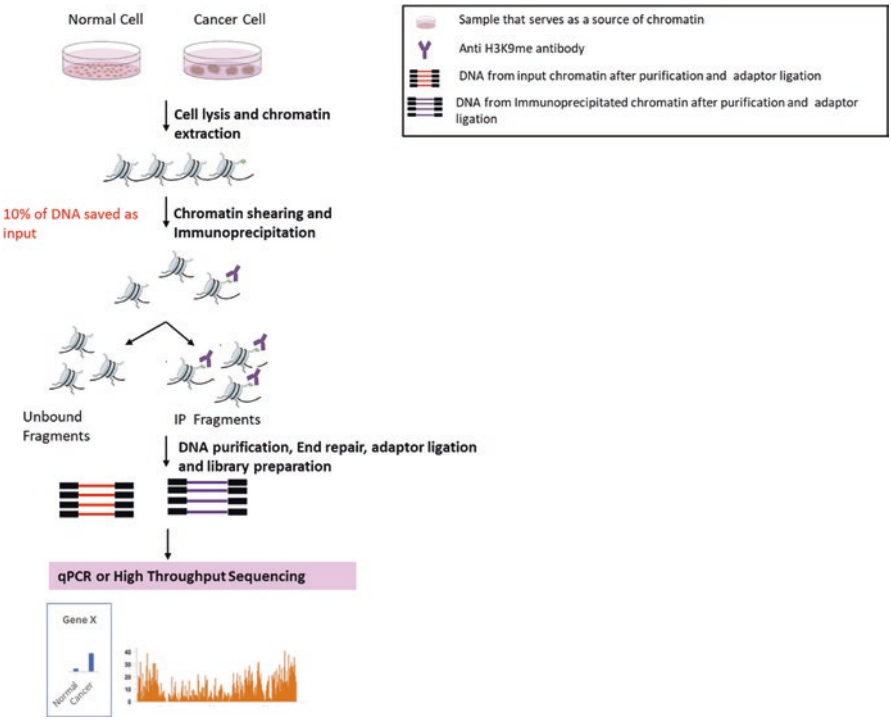
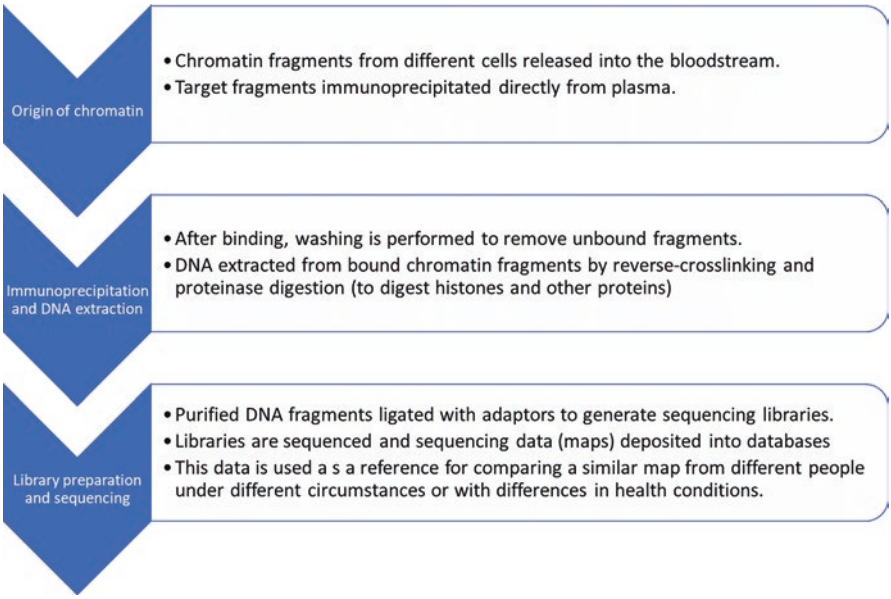


Fig. 7.5 Shows the major steps of chromatin immunoprecipitation. Chromatin is extracted from a sample followed by shearing and denaturation to generate small fragments (200–600 bp). This is followed by pull down of chromatin fragments containing the modification against which antibody is used. Both the input DNA (a fraction of total genomic DNA used for normalisation) and immunoprecipitated (IP) DNA are purified, end-repaired (and ligated with adaptors for sequencing) and analysed after performing qPCR or sequencing

regions like promoters and enhancers. Use of cell free (cf) chromatin offers a non-invasive method of patient analysis and molecular diagnosis. During the years to come, more and more **cf *ChIP seq*** libraries, corresponding to different chromatin marks would become available. This would mark a huge leap into the future of molecular diagnosis, especially epigenetic diagnosis and "*epi-therapy*" [50].



2. Study of cell type specific chromatin modifications.

   It is very useful in obtaining important information about the epigenetic status of each different cell type in tissue under different conditions (like normal versus disease). For example, studies of chromatin modifications of β-cells of pancreas in normal and diabetic individuals using ChIP can yield important information about the precise role of the said modification in β-cells in imparting and propagating type 2 diabetes [51].

7.3 Chromatin Structure and Genome Organisation

As mentioned previously, genome inside eukaryotic cells does not exist as a linear molecule but instead exists as a hierarchically packed compact structure with multiple levels of organisation. Therefore, Genome Organisation refers to the arrangement of genomic elements (DNA and proteins) in the 3-dimensional (3D) space within the nucleus. The order in which genomic elements are distributed into different chromosomes also contributes to genome organisation. Other structural genomic elements which are quintessential to this organisation are chromosome territories, compartments, topologically associating domains. The macro-scale elements termed as topologically associated domains (TADs) are demarcated by architectural proteins like cohesion and CTCF (CCCTC-binding factor) and interact through chromatin loops. The 3D genome organisation is of particular significance in

processes like replication, transcription, recombination, gametogenesis, mitosis, meiosis, development, stem cell differentiation and pluripotency maintenance [52]. In this section, we shall briefly discuss about the role of genome organisation in maintaining chromatin architecture and alterations in 3D genome in the aetiology of diseases.

7.3.1 Regulation of 3D Genome Organisation

An increase in the complexity and size of genome is observed across species, from lower to higher eukaryotes. For example, the size difference between yeast and mammalian genomes is approximately 300-fold. The most significant genomic elements which undergo a proportional increase with genome complexity are the repetitive DNA elements. In mammals, repetitive DNA and non-coding elements account for about 96% of the genomic DNA sequences (44% repetitive and 52% non-coding sequences) [53]. This leads to a proportional increase in the heterochromatic regions since repetitive DNA elements undergo hyper-recombination at higher frequencies and are therefore tightly wrapped in highly condensed, heterochromatic regions. Organisation and maintenance of such complex genomes require more extensive epigenetic programming and silencing mechanisms. Alternatively, study of these mechanisms not only provides opportunities for better understanding of genomes and their regulation but also opportunities for targeting epigenetic modifications for therapeutic purposes. This concept has immense implications in "*personalised epigenetic therapy*" [54] (Fig. 7.6).

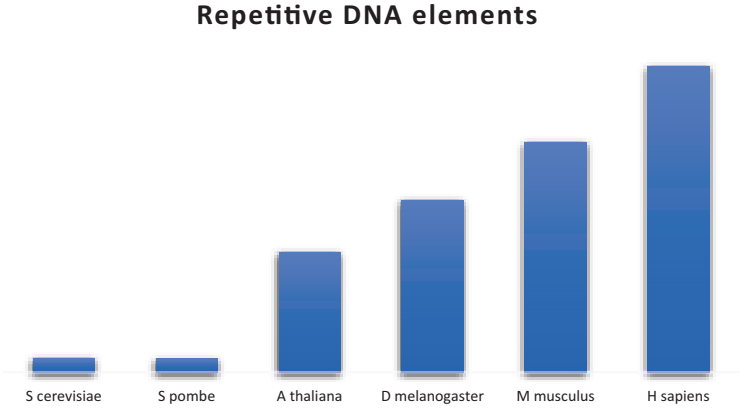


Fig. 7.6 The graph represents increase in the percentage of repetitive DNA elements across eukaryotic species, from Saccharomyces cerevisiae to humans. This increase is proportional to an increase in size and complexity in the 3D genome

7.3.2 Non-coding RNAs (ncRNA)

Non-coding RNAs are ribonucleic acid molecules that are not translated into proteins. Genomic DNA sequences from which these molecules arise are typically known as "*RNA genes*". Some of the non-coding RNAs are known for a very long time like the transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), both of which play very important roles in the process of protein translation. However, with advancements in the tools and techniques of molecular biology like "omics" and "*computational biology*" over the past two decades, various classes of non-coding RNAs have been identified with diverse range of size and functional roles. In addition to tRNAs and rRNAs, other major types of ncRNAs include small and long non-coding RNAs.

Small non-coding RNAs include a size range of 18–200 nucleotides. Major types are micro-RNAs (miRNAs), small interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), snRNAs.

miRNAs play role in downregulation of gene expression through binding to 3′ untranslated regions (3′ UTRs) of mRNAs. A single miRNA can bind and affect the expression of several mRNAs through partial sequence complementarity. miRNAs have been identified to play various roles in disease like miR 125b in type 2 diabetes, miR 206 in oestrogen receptor alpha [55], miR 135a and 135b in colorectal cancer [56].

siRNAs are formed by the RNA interference pathway and play role in heterochromatin organisation, maintenance and transcriptional gene silencing. These have been identified and well characterised in the yeast *Schizosaccharomyces pombe* [8].

piRNAs form RNA–protein complexes with Piwi proteins and play role in transcriptional gene silencing in somatic cells. In germline cells of testes, the Piwi protein–RNA complexes are linked to transcriptional gene silencing of retrotransposons. Unlike siRNAs and miRNAs, Piwi RNAs are formed in a dicer-independent manner.

snoRNAs are small RNA molecules found in the nucleolus that mainly participate in guiding chemical modifications of rRNAs and tRNAs, thereby assisting in the process of translation.

Long non-coding RNAs (lncRNAs) have a size range of approximately 1,000 to 10,000 residues. These play important roles in gene imprinting and X-chromosome inactivation. The most well-studied examples include X inactive (Xist) and homeobox (HOX) transcript anti-sense RNA (HOTAIR).

Xist In case of human and other placental mammal females containing two copies of X-chromosome, one of the copies remains inactive and heterochromatinised (called as inactive X or Xi or Barr-body) through various mechanisms. Xist RNAs are transcribed from the X-chromosome and play a prominent role in gene silencing and heterochromatinisation of Xi.

HOTAIR is a 2.2 kb long non-coding RNA, encoded by the HOTAIR gene, located within the HOXC cluster, between HOXC11 and HOXC12 on chromosome 12. It plays a role in transcriptional downregulation of HOXD gene cluster on chromosome 2. However, the sequence and function of HOTAIR are different in the two most well-studied organisms, mice and humans [57].

7.3.3 *Effect of Non-coding RNAs on Gene Silencing and Genome Organisation*

Recruitment of chromatin remodelling complexes and deposition of inhibitory histone methylation mark H3K27me3 contributes to the gene silencing effect of lncRNAs like HOTAIR [58–60]. Recruitment of RNA binding proteins by lncRNAs that interfere with binding of transcription factors to gene promoters also contributes to their function. Non-coding RNAs can also contribute to or propagate DNA methylation. In fact, recent findings have shown that siRNA and miRNAs affect transcriptional silencing through DNA methylation, histone methylation (H3K9 and H3K27), histone deacetylation and recruitment of remodelling complexes [29].

7.3.4 *Chromosome Territories, Compartments and Nuclear Lamins in Genome Organisation*

Microscopy based studies, chromosome painting, chromosome conformation capture (3C) and its variants (4C, 5C, HiC) have shown that interphase chromosomes preferentially reside in separate chromosome territories. Within each chromosome territory, position of individual genomic elements coincides with transcriptional activity, with gene-rich, transcriptionally active regions occupying the borders of chromosome territories. Also, at a mega-base scale, similar genomic regions interact with one another. Transcriptionally active regions interact with other active regions which possess activating chromatin modifications, higher gene density and chromatin accessibility. Similarly, inactive regions interact with other inactive regions. The boundaries of topologically associated domains (TADs) can also display interactions. This can be of consequence when we study genes participating in same or related functions but belonging to different chromosomes. Overall, nuclear compartments containing TAD boundaries (active regions) are frequently found in the interior nuclear space and those containing heterochromatinised, gene poor TADs occupy nuclear periphery and preferentially associate with nuclear lamina (innermost layer of nuclear membrane) through the proteins termed as nuclear lamins [61]. Nuclear lamin A, B and C are mainly responsible for this peripheral localisation of heterochromatin TADs and absence of these proteins can result in re-localisation of heterochromatin to nuclear interior [12, 52]. Heterochromatin

protein (HP1) is responsible for maintenance of peripheral, constitutive heterochromatin. More details about the role of 3D genome organisation have been reviewed elsewhere [52].

7.3.5 *Higher Order Genome Organisation and Diseases*

Mis-regulations in individual components responsible for maintaining genome organisation can result in a large number of human diseases [62] (Table 7.4).

1. CTCF has been implicated in diseases like Huntington's disease, fragile X mental retardation, Silver-Russell (SRS), Beckwith-Wiedemann syndrome (BWS), cancer and myotonic dystrophy [63, 64].
2. Mutations in cohesin are implicated in Cornelia de Lange Syndrome (CLS) and Roberts Syndrome [RS] [65, 66].
3. HP1 has been shown in various studies to be involved in different forms of cancer [67, 68].
4. Absence of architectural chromatin proteins HP1, KAP-1 or HDAC1/2 hastens ATM-mediated repair [69].
5. Alterations in relative chromosome positioning in the 3D genome results can result in problems with adipocyte differentiation and cancer [70].
6. Mis-regulations in lamin A are associated with ageing [71].
7. Changes in overall nuclear architecture can affect genome stability [72].

Table 7.4 A list of papers related to some interesting aspects of epigenetics and diseases

S. No	Title of the paper	PMID
1.	The developmental origins of Well-being	15347527
2.	Epigenetic differences Arise during the lifetime of monozygotic twins	16009939
3.	Epigenetics: Connecting environment and genotype to phenotype and disease	19493882
4.	R loops: From physiological to pathological roles	31607512
5.	Impact of genetic and epigenetic factors from early life to later disease	18803962
6.	The impact of nutrition and environmental epigenetics on human health and disease	30388784
7.	Recent developments on the role of epigenetics in obesity and metabolic disease	27408648
8.	Impact of oxidative stress during pregnancy on Foetal epigenetic patterns and early origin of vascular diseases	26024054
9.	Impact of epigenetic dietary compounds on transgenerational prevention of human diseases	24114450
10.	DNA methylation: The pivotal interaction between early-life nutrition and glucose metabolism in later life	25327140

References

1. Allison LA (2012) Fundamental molecular biology. Wiley, Hoboken
2. Lehninger AL, Nelson DL, Cox MM (2005) Lehninger principles of biochemistry. W.H. Freeman, New York
3. Felsenfeld G (2014) A brief history of epigenetics. Cold Spring Harb Perspect Biol 6(1):a018200
4. Russo VEA, Martienssen RA, Riggs AD (1996) Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press, New York
5. Schaeffer S (2018) Muller "elements" in Drosophila: how the search for the genetic basis for speciation led to the birth of comparative genomics. Genetics 210:3–13. https://doi.org/10.1534/genetics.118.301084
6. Sturtevant AH (1921 Mar) Genetic studies on Drosophila simulans. III. Autosomal genes. General discussion. Genetics 6(2):179–207
7. Farooq Z, Banday S, Pandita TK, Altaf M (2016) The many faces of histone H3K79 methylation. Mutat Res 768:46–52
8. Farooq Z, Abdullah E, Banday S, Ganai SA, Rashid R, Mushtaq A, Rashid S, Altaf M (2019) Vigilin protein Vgl1 is required for heterochromatin-mediated gene silencing in *Schizosaccharomyces pombe*. J Biol Chem 294(48):18029–18040
9. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389(6648):251–260
10. Mersfelder EL, Parthun MR (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. Nucleic Acids Res 34(9):2653–2662
11. Peterson CL, Laniel MA (2004) Histones and histone modifications. Curr Biol 14(14):R546–R551
12. Banday S, Farooq Z, Rashid R, Abdullah E, Altaf M (2016) Role of inner nuclear membrane protein complex Lem2-Nur1 in heterochromatic gene silencing. J Biol Chem 291(38):20021–20029
13. Altaf M, Saksouk N, Côté J (2007) Histone modifications in response to DNA damage. Mutat Res 618(1–2):81–90
14. Ganai SA, Farooq Z, Banday S, Altaf M (2018) In silico approaches for investigating the binding propensity of apigenin and luteolin against class I HDAC isoforms. Future Med Chem 10(16):1925–1945
15. Ganai SA, Banday S, Farooq Z, Altaf M (2016) Modulating epigenetic HAT activity for reinstating acetylation homeostasis: a promising therapeutic strategy for neurological disorders. Pharmacol Ther 166:106–122
16. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications. Nat Struct Mol Biol 14:1025–1040
17. Yun M, Wu J, Workman JL, Li B (2011) Readers of histone modifications. Cell Res 21:564–578
18. Gurley LR et al (1978) Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. Eur J Biochem 84:1–15
19. Nowak SJ, Corces VG (2000) Phosphorylation of histone H3 correlates with transcriptionally active loci. Genes Dev 14:3003–3013
20. Goldknopf IL, Taylor CW, Baum RM, Yeoman LC, Olson MO, Prestayko AW, Busch H (1975) Isolation and characterization of protein A24, a 'histone-like' non-histone chromosomal protein. J BiolChem 250:7182–7187
21. West MH, andBonner W.M. (1980) Histone H2B can be modified by the attachment of ubiquitin. Nucleic Acids Res 8:4671–4680
22. Pham AD, Sauer F (2000) Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in Drosophila. Science 289:2357–2360
23. Jones JM, Bhattacharyya A, Simkus C, Vallieres B, Veenstra TD, Zhou M (2011) The RAG1V(D)J recombinase/ubiquitin ligase promotes ubiquitylation of acetylated, phosphorylated histone3.3. ImmunolLett 136:156–162

24. Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M (2008) Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. NatCell Biol 10:483–488
25. Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM (2011) BRCA1 tumour suppression occurs via heterochromatin- mediated silencing. Nature 477:179–184
26. Jia S, Wang Y (2009) Degrees make all the difference. Epigenetics 4(5):1–4
27. Gilbert SF (2010) Developmental biology, 9th edn. Sinauer Associates, Sunderland, pp 35–37
28. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer NG, Perez-Rosado A, Calvo E, Lopez JA, Cano A, Calasanz MJ, Colomer D, Piris MA, Ahn N, Imhof A, Caldas C, Jenuwein T, Esteller M (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet 37:391–400
29. Handy DE, Castro R, Loscalzo J (2011) Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation 123(19):2145–2156
30. Bestor TH (2000) The DNA methyltransferases of mammals. Hum Mol Genet 9:2395–2402
31. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16:6–21
32. Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. Neuron 53(6):857–869
33. Fontecha-Barriuso M, Martin-Sanchez D, Ruiz-Andres O, Poveda J, Sanchez-Niño MD, Valiño-Rivas L, Ruiz-Ortega M, Ortiz A, Sanz AB (2018) Targeting epigenetic DNA and histone modifications to treat kidney disease. Nephrol Dial Transplant 33(11):1875–1886
34. Robertson KD (2001) DNA methylation, methyltransferases, and cancer. Oncogene 20(24):3139–3155
35. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19(1):92–105
36. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99(3):247–257
37. Deplus R, Brenner C, Burgers WA, Putmans P, Kouzarides T, de Launoit Y, Fuks F (2002) Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucleic Acids Res 30(17):3831–3838
38. Kim JK, Samaranayake M, Pradhan S (2009) Epigenetic mechanisms in mammals. Cell Mol Life Sci 66(4):596–612
39. Ling C, Poulsen P, Simonsson S, Rönn T, Holmkvist J, Almgren P, Hagert P, Nilsson E, Mabey AG, Nilsson P, Vaag A, Groop L (2007) Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Invest 117(11):3427–3435
40. Farooq Z, Shah A, Tauseef M, Rather RA, Anwar M. Evolution of epigenome as the blueprint for carcinogenesis. https://doi.org/10.5772/intechopen.97379
41. Anwar M, Nanda N, Kochhar R et al (2013) Mutations & expression of APC & β-catenin in sporadic colorectal tumors: a mutational "hotspot" for tumorigenesis. J Gastroenterol Hepatol 28:665–666
42. Anwar M, Kochhar R, Singh R et al (2016) Frequent activation of the β-catenin gene in sporadic colorectal carcinomas: a mutational & expression analysis. Mol Carcinog 55(11):1627–1638
43. Malhotra P, Anwar M, Kochhar R, Ahmad S, Vaiphei K, Mahmood S (2014) Promoter methylation and immunohistochemical expression of hMLH1 and hMSH2 in sporadic colorectal cancer: a study from India. Tumour Biol 35(4):3679–3687
44. Malhotra P, Anwar M, Nanda N, Kochhar R, Wig JD, Vaiphei K, Mahmood S (2013) Alterations in K-ras, APC and p53-multiple genetic pathway in colorectal cancer among Indians. Tumour Biol 34(3):1901–1911
45. Stirzaker C, Taberlay PC, Statham AL, Clark SJ (2014) Mining cancer methylomes: prospects and challenges. Trends Genet 30(2):75–84

46. Chatterjee A, Rodger EJ, Stockwell PA, Le Mée G, Morison IM (2017) Generating Multiple Base-resolution DNA Methylomes using reduced representation bisulfite sequencing. Methods Mol Biol 1537:279–298
47. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28(10):1057–1068
48. Tang J, Fang F, Miller DF, Pilrose JM, Matei D, Huang TH, Nephew KP (2015) Global DNA methylation profiling technologies and the ovarian cancer methylome. Methods Mol Biol 1238:653–675
49. Sun YV, Hu YJ (2016) Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Adv Genet 93:147–190
50. Sadeh R, Sharkia I, Fialkoff G, Rahat A, Gutin J, Chappleboim A, Nitzan M, Fox-Fisher I, Neiman D, Meler G, Kamari Z, Yaish D, Peretz T, Hubert A, Cohen JE, Salah A, Temper M, Grinshpun A, Maoz M, Abu-Gazala S, Ben Ya'acov A, Shteyer E, Safadi R, Kaplan T, Shemer R, Planer D, Galun E, Glaser B, Zick A, Dor Y, Friedman N (2021) ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. Nat Biotechnol 39(5):586–598. https://doi.org/10.1038/s41587-020-00775-6. Epub 2021 Jan 11. Erratum in: Nat Biotechnol. 2021 Jan 21;: Erratum in: Nat Biotechnol. 2021 Mar 1;: PMID: 33432199; PMCID: PMC7610786
51. Mito M, Kadota M, Tanaka K, Furuta Y, Abe K, Iwasaki S, Nakagawa S (2018) Cell type-specific survey of epigenetic modifications by tandem chromatin immunoprecipitation sequencing. Sci Rep 8(1):1143
52. Zheng H, Xie W (2019) The role of 3D genome organization in development and cell differentiation. Nat Rev Mol Cell Biol 20(9):535–550
53. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al (2001) The sequence of the human genome. Science 291(5507):1304–1351. https://doi.org/10.1126/science.1058040. Erratum in: Science 2001 Jun 5;292(5523):1838. PMID: 11181995
54. West M, Ginsburg GS, Huang AT, Nevins JR (2006) Embracing the complexity of genomic data for personalized medicine. Genome Res 16(5):559–566
55. Adams BD, Furneaux H, White BA (2007) The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. Mol Endocrinol 21(5):1132–1147
56. Anwar M, Kochhar R, Kaul D, Bhatia A, Mahmood S (2018) Apc gene regulation by Mir-135A and Mir-135B in sporadic colorectal cancer: in silico approach and expression analysis. Gastroenterology 154(6):S-349. https://doi.org/10.1016/S0016-5085(18)31496-3
57. Schorderet P, Duboule D (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. PLoS Genet 7(5):e1002071
58. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323
59. Wagschal A, Sutherland HG, Woodfine K, Henckel A, Chebli K, Schulz R, Oakey RJ, Bickmore WA, Feil R (2008) G9a histone methyltransferase contributes to imprinting in the mouse placenta. Mol Cell Biol 28(3):1104–1113
60. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322(5908):1717–1720
61. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453(7197):948–951
62. Anania C, Lupiáñez DG (2020) Order and disorder: abnormal 3D chromatin organization in human disease. Brief Funct Genomics 19(2):128–138
63. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. Annu Rev Neurosci 30:575–621
64. Witcher M, Emerson BM (2009) Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. Mol Cell 34(3):271–284

65. Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, Jukofsky L, Wasserman N, Bottani A, Morris CA, Nowaczyk MJ, Toriello H, Bamshad MJ, Carey JC, Rappaport E, Kawauchi S, Lander AD, Calof AL, Li HH, Devoto M, Jackson LG (2004) Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster nipped-B. Nat Genet 36(6):631–635
66. Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T (2004) NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet 36(6):636–641
67. Dialynas GK, Vitalini MW, Wallrath LL (2008) Linking heterochromatin protein 1 (HP1) to cancer progression. Mutat Res 647(1–2):13–20
68. Ruginis T, Taglia L, Matusiak D, Lee BS, Benya RV (2006) Consequence of gastrin-releasing peptide receptor activation in a human colon cancer cell line: a proteomic approach. J Proteome Res 5(6):1460–1468
69. Goodarzi AA, Noon AT, Jeggo PA (2009) The impact of heterochromatin on DSB repair. Biochem Soc Trans 37(pt 3):569–576
70. Parada LA, McQueen PG, Munson PJ, Misteli T (2002) Conservation of relative chromosome positioning in normal and cancer cells. Curr Biol 12(19):1692–1697
71. Scaffidi P, Misteli T (2006) Lamin A-dependent nuclear defects in human aging. Science 312(5776):1059–1063
72. Oberdoerffer P, Sinclair DA (2007) The role of nuclear architecture in genomic instability and ageing. Nat Rev Mol Cell Biol 8(9):692–702