We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 2

Regulation of Mammalian Gene Expression

Partha P. Mitra

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72848

Abstract

Regulation of mammalian gene expression has been an ever growing subject in the field of Biology and the biomedical science research. In the last several decades, extensive amount of research together with the implementation of the latest technologies revealed that the whole process is regulated at the multiple stages with a series of interconnected complex biochemical and molecular pathways. Unearthing this complexity in one hand helps us in understanding the concerted effort put by the respective cellular machinery to regulate the whole process, and on the other hand, it provides a new insight about the development of several diseases where gene expressions play a pivotal role. Discussions here focus on the involvement of transcription factors or cofactors and the linkage of the transcription network with the signal transduction pathways. Besides proteins as a regulator, the role of the nucleic acids such as miRNA, chromosomal conformation and the modification of DNA bases or core histone proteins, in gene expression has also been explored. The purpose of this chapter is to provide the big picture of the diverse regulatory network and the phenomenal complexity of the regulation of gene expression.

Keywords: transcription factor, gene expression, structure-function relation

1. Introduction

Over the past few decades, extensive amounts of research have been carried out to understand the regulation of mammalian gene expression. Studies were originally started with bacteriophage, yeast and other lower order eukaryotes, and the acquired knowledge was later implicated to understand the mammalian systems, including the human cells. Several milestone discoveries in early days helped scientists to draw the very basic picture of gene expression, which includes lysogenic to lytic phase transition of the bacteriophage lambda (λ), inducible gene regulation in bacteria (lac operon system) and the sequential gene expression during

INTECH
open science | open minds

© 2018 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

early development of Drosophila embryo. All those studies clearly demonstrated that gene expression is an outcome of a concerted participation, triggered by intracellular or extracellular stimuli, of several intracellular protein factors and cofactors. At the same time, it was assumed that the scenario will be more complicated for the higher order eukaryotes simply due to presence of multistep regulatory processes with the involvement of more factors and cofactors. Until recently, a substantial amount of studies clearly depicted that the regulation of mammalian gene expression is more complicated than we thought ever before. This complex regulatory process comprises of a sequential or simultaneous involvement of at least four major steps and those are as follows:

A. transcriptional

B. posttranscriptional

C. epigenetic

D. translational

Due to the lack of space, rather than going in to the intricate details, a broad overview of each step would be provided taking the example of a few very well-studied systems.

Transcriptional control: Studies on transcriptional regulation are perhaps the most investigated segment in understanding the complexity of mammalian gene expression. Before going into the detail, discussion about the regulatory mechanisms, we should have a very clear idea about the process of transcription. Broadly, it is a process where the enzyme RNA polymerase (RNA Pol) decodes the genetic information, in the form of RNA that stored in the chromosomal DNA. The transcription machinery produces five types of RNAs, which includes messenger RNA (mRNA) contributes between 1 and 2% of the total transcripts, ribosomal RNA (rRNA) that covers more than 80% of the total transcripts, transfer RNA (tRNA, required for translation), the recently discovered micro RNA (miRNA) and small interfering RNA (siRNA). Among the several subtypes of RNA produced at any time, only mRNA translates into proteins. All different types of RNA molecules, synthesized by 5′ to 3′ movement of polymerases, are not produced by a single type of enzyme. For example, mRNA is transcribed by RNA PolII, whereas rRNA is produced by RNA PolI, and obviously, the regulation of transcription to synthesize each subtypes of RNA is significantly diverse and complicated. Our discussion here is mostly focused on the regulation of RNA PolII driven transcription, which has been investigated most rigorously.

Almost 24% of the human genes contain an evolutionary conserve DNA sequence element (5′TATAAA3′) in the core promoter or a variant of it called as TATA box or also known as Goldberg-Hogness box located 25–35 bp upstream of the transcription start site. The advantage of this AT rich sequence is that it facilitates the unwinding of the promoter DNA upon binding of the specific protein TBP (TATA binding protein) which is part of the TFIID complex. This huge multiprotein TFIID complex is specifically playing a very important role because it is associated with CDK7, −8, or −9 which are required for the phosphorylation of the C-terminal domain (CTD) of RNA PolII. TFIIA, which is another core factor, stabilizes the TFIID-DNA complex. Once the binding is stabilized, then RNA-PolII recognized the protein

complex and recruited to make the PIC (pre-initiation complex). Among the other core factors, TFIIH plays a very important role in the transcription initiation because this multifunctional protein comprises of DNA-dependent ATPase, helicase and protein kinase activities [1].

Recent bioinformatics studies on human genome indicated that about 80% of the genes are transcribed from the promoter where TATA box does not exist. Such TATA-less genes are characterized by the presence of multiple promoters and transcription start sites and generate several transcripts. However, the question that remains unanswered here was how the transcription starts in this class of promoter? Or, is there any other role of TBP here? Earlier studies [2] indicated that the transcription factor TFIID and TBP were also involved in the initiation of transcription. The involvement of TFIID is conceivable because the associated CDK's are required for the CTD phosphorylation of RNA PolII but the function of TBP was not clear. Recent studies [3] on unicellular eukaryote showed that the DNA binding domain of TBP was not required for the transcription, which implied that TBP does other essential functions which could be a subject of further studies.

Transcriptions in mammalian cells are regulated at multiple stages and several protein factors and cofactors are involved at each stage. In general, a transcriptionally active gene is controlled by a stretch of DNA sequence mostly situated at the upstream of the transcription start site (−500 bp to −1000 bp) defined as promoters which is a docking site of several proteins termed as transcription factors (TF). Mammalian cells synthesize around 3000 transcription factors [4], and each one harbors a specific DNA sequence binding motif.

Transcription factors (TF) are the fundamental regulators of eukaryotic transcription. Therefore, to understand the complexity of transcription, we should have a very clear idea about correlation between the structural diversities and the functional activities of these proteins. TFs can be subdivided into two major categories based on the mechanisms by which they control the gene expression. TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH are known as basal transcription factors because they are required to form a complex with RNA PolII, known as the PIC, for the transcription of the majority of the mammalian genes irrespective of the nature of cells or tissue types [5].The PIC is a huge multiprotein complex with multiple functions that include binding of the DNA sequence at the transcription start site, recruitment of RNA PolII, creating the bubble by changing the helical structure of the DNA so the polymerase can move after phosphorylation of the CTD.

Ubiquitous TFs are a class of DNA proteins bind to the promoter proximal region of a vast range of mammalian genes after recognizing a unique and conserved DNA sequence. Transcription factors such as Sp1 binds to the 5′-GGGCGG-3′ and AP1 binds to the 5′-TGA(G/C)TCA-3′ across the species. In general, ubiquitous transcription factors such as AP1 and SP1 are engaged in two major functions. One is the DNA binding and other is the recruitment of associated factors to initiate the transcription and in this context, structure-function relation plays a very important role. In case of TF, DNA binding introduces an inevitable change in the three-dimensional structure of the protein, which makes it interactive to the other cofactor proteins. This altered structure promotes to make a functionally active multiprotein complex that is essential in establishing the link with the extra and intracellular signal transduction pathways and at the same time, passing this signal effectively to the transactivation domain for transcription initiation.

Structurally, TFs can be subdivided into four major categories such as (1) helix-turn-helix proteins, (2) zinc finger proteins, (3) leucine zipper proteins and (4) helix-loop-helix proteins. Significant amount of studies were conducted to understand the impact of the three-dimensional structure with the functional activity of the protein. Each structural motif contributes to the (A) binding of the protein to the DNA, (B) homo- or heterodimer formation and (C) subsequent transactivation. To understand the structure-function relationship, we could further discuss citing examples of zinc finger proteins.

Mammalian cells produce about 1000 different types of zinc finger proteins and a significant part of them work as a transcription factor. The very well-studied Sp1 protein contains three zinc finger domains at the C-terminal end of the protein, which are responsible for the DNA binding activity of this protein. To make a higher order multiprotein complex, SP1 interacts with a variety of proteins which is often mediated by the zinc finger domains. For example, SP1 interacts with the CyclinD1 and the retinoblastoma protein pRB to regulate the transcription of the human keratin4 gene in squamous epithelium cells [6]. Our studies with zinc finger transcription factor HiNF-P very clearly demonstrated that the zinc finger domains are responsible for the DNA binding as a well as for the interaction with negative cell cycle regulator protein p57/kip2 (the zinc finger domains third and fourth domain from the N-terminal end of HiNF-P are required, (**Figure 1**). Our studies demonstrated that the HiNF-P-p57/Kip2 interaction is required for the downregulation of H4 gene transcription and the HiNF-P-NPAT/220 association, which is discussed further, not mediated through Zn finger domain, is required for the transcription activation [7].

Transcription factors that belong to the class of helix-turn-helix proteins, leucine zipper proteins and helix-loop-helix proteins are evolutionary conserve group of proteins and are responsible for the expression of the genes associated with cellular differentiation, lineage commitments and

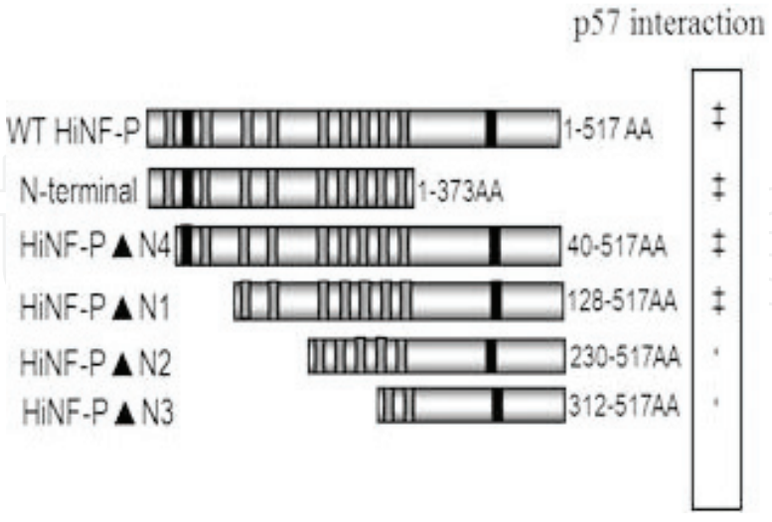


Figure 1. Zn-finger domain of HiNF-P is responsible for the interaction with p57. Zn fingers are marked as vertical gray rectangles. N-terminal deletion mutants of HiNF-P were co-expressed with p57 to perform co-immunoprecipitation experiments. Numbers of amino acids present in each deletion mutant are mentioned at the end of the c-terminal domain. Interaction with p57 of each deletion mutants is mentioned in a separate table at the right-hand side of the figure with +ve (interacting) or –ve (noninteracting).

organogenesis. The conserve helical structure contributes to the binding of the major groove of the DNA and the dimerization. In addition to form homodimer, they also form heterodimer and often found that this heterodimer is the functionally active transcription promoting complex. For example, the leucine zipper transcription factor c-Fos cannot bind the DNA unless it forms a heterodimer with another leucine zipper protein c-Jun and interestingly, this heterodimer formation enhances the binding efficiency around 30-fold.

Therefore, it can be concluded that the mammalian gene expression is primarily regulated by the general and a set of ubiquitous transcription factors. However, the next level of regulation begins with the binding of a set of gene selective transcription factor to the promoter proximal region. Most of the cases, these gene-selective transcription factors are connected to the extra- or intracellular signal transduction pathways, which act as master regulator to switch ON or OFF the gene expression. This fact can be illustrated further by taking a very well-studied cell cycle regulated transcription of human histone H4 gene.

Transcription of H4 gene upregulates several fold at the onset of S-phase of the mammalian cell cycle in order to package the newly synthesized DNA. The one part of the proximal promoter, close to the transcription start site (Site-II) of the H4 gene contains the binding site of three major gene specific transcription factor HiNF-M/IFR-2, HiNF-D/CDPcut and HiNF-P (**Figure 2**). HiNF-M/IRF2 is a downstream target of master transcription factor E2F, which also regulates the expression of cell cycle check point controlling cyclins and CDKs. On the other hand, HiNF-P binds with its cofactor NPAT/p220, which is a direct subject of CyclinE/CDK2 that controls the G1/S transition, whereas HiFND/CDPcut is a multimeric protein with homeodomain protein CDPcut participates in DNA binding. Ectopic expression of HiNF-P and HiNF-M activates the H4 gene transcription but HiNF-D/CDPcut downregulates the transcriptional activity, which is an indication that transcription can be positively or negatively regulated depending upon the relative abundance of these factors at this region of the promoter.

The other part of the proximal promoter located further from the transcription start site (site-I) is the binding location for the ubiquitous transcription factors such as AP1 and SP1. Further studies

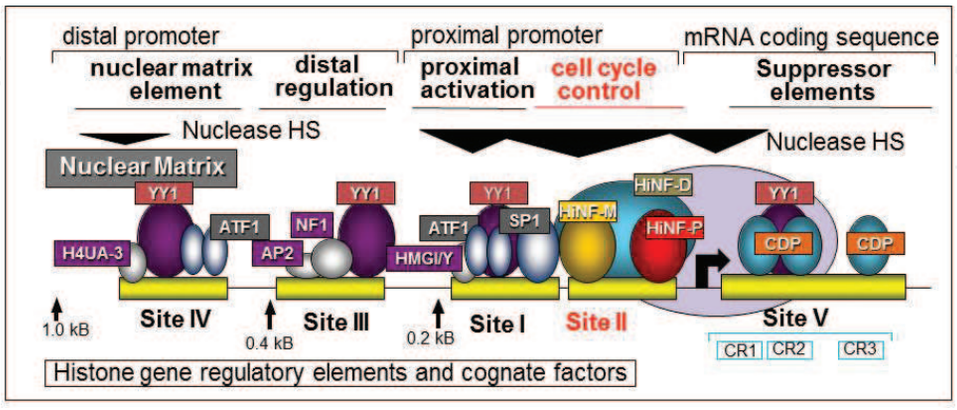


Figure 2. Human H4 gene promoter and the transcription factor binding sites. H4 gene is regulated at the G1/S cell cycle check point by a series of basic and the gene elective transcription factors binding over span of 1.0 kb DNA sequence. The Site-II DNA sequence is considered as the cell cycle element because it is the binding site of the cell cycle regulatory transcription factors as mentioned in the text.

provided clear evidence that the binding of the gene-specific transcription factors at site-II was conditional. A complete loss of HiNF-M/IFR-2, HiNF-D/CDPcut and HiNF-P binding was noticed when cells switch over to differentiation where the H4 gene transcription is shut down completely, but the bindings of AP1, SP1 were observed to be unaffected under this condition. Now the question is how histone H4 gene transcription is connected to the cell cycle check point? Growth factor-dependent signal activated CyclinE/CDK2 complex phosphorylates many essential proteins including NPAT/p220. Upon phosphorylation, NPAT/p220 binds to HiNF-P and makes the functionally active complex, which binds to the HiNF-P binding element at site-II and activates the transcription. At the late S phase, the CyclinE/CDK2 complex becomes inactivated, which in turn fails to phosphorylate the NPAT/p220-HiNF-P complex (**Figure 3**). Therefore, the H4 gene transcription model reveals, in a very simple way, how growth factor-dependent signal transduction pathway controls the gene expression. In order to keep our discussion very focused,

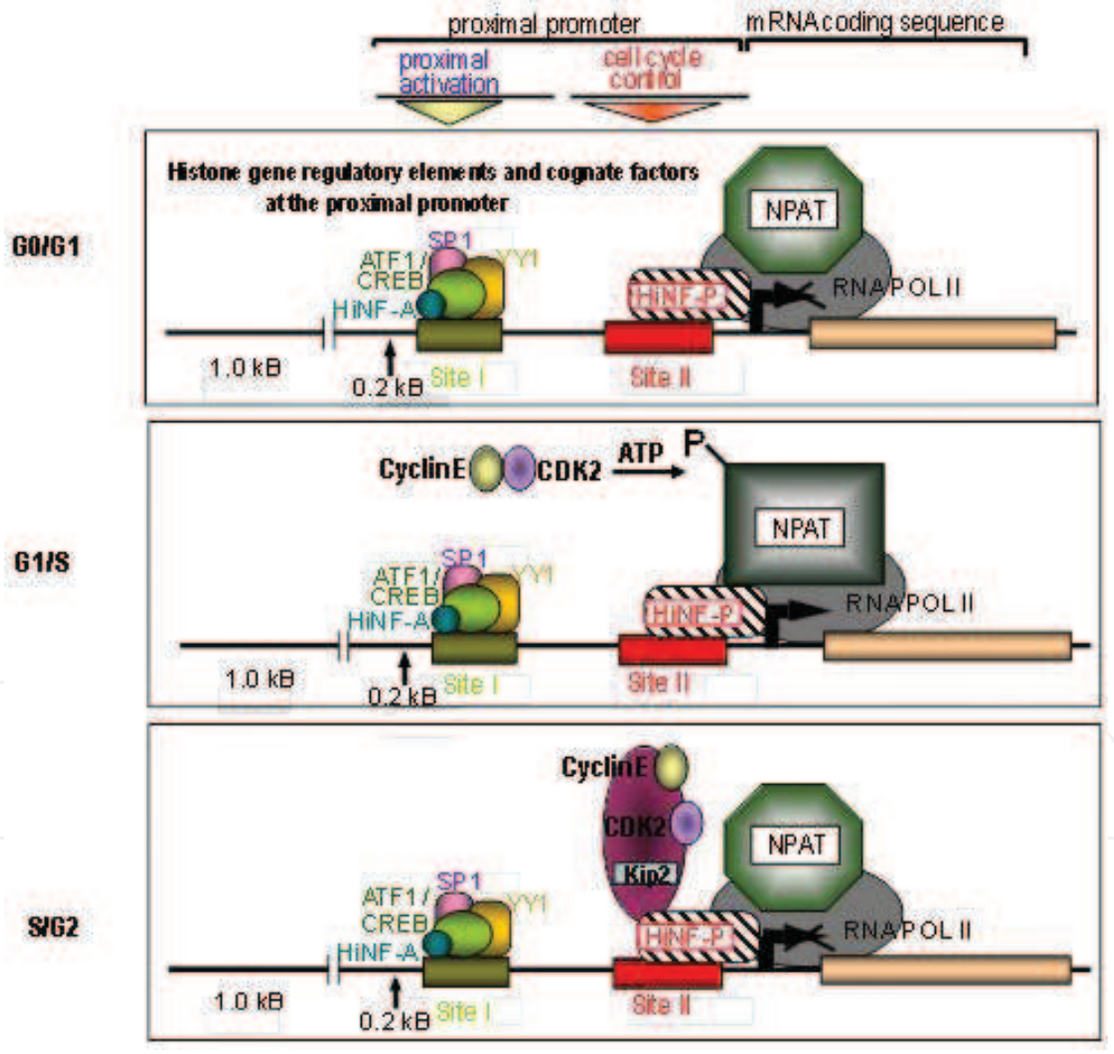


Figure 3. Regulation of human h4 gene transcription at the G1/S transition occurs through transcription complex formation. At the onset of G1/S transition, in association with other cognate Site-II specific binding factor, CyclinE/CDK2 mediated activation of NPAT-HiNF-P complex by phosphorylation is a prime driver of the transcription. At the end of S phase, the complex is functionally inhibited due to the inactivation of the CyclinE/CDK2 complex by the cellular kinase inhibitor p57/kip2.

the involvement of other cognate factors in other sites of this promoter is excluded. However, several important questions are yet to be answered regarding this H4 gene transcription regulation and perhaps one of them is how all three site-II specific factors act in a coordinate fashion to regulate the transcription [8–12].

So far our discussion focused on the effect of promoter and the associated factors or cofactors in the regulation of transcription. Recent studies revealed that besides the promoter, DNA sequence element located several megabases up or downstream of the transcription start site, termed as enhancers, also play a very important role in the regulation. The effect of enhancer on gene expression was revealed long time ago when researchers were trying to understand the massive transcription upregulation of the β-globin gene. However, the mechanism through which the enhancer controls the gene expression was very elusive. The most obvious question was how these cis-acting elements, located so far form the coding region, could control the transcription of a specific gene? The hypothesis that was put forward to explain the regulatory role of enhancers pointed towards the three-dimensional chromatin looping. Mammalian genome has been considered as a series topologically associated domain (TAD) comprises several megabases of DNA connected through intergenic sequence. Genome-wide Chromosomal Conformation Capture (3C) experiments, a recently developed method to estimate the looping in chromosome, indicated that proteins such as CTCF and cohesion are responsible for the TAD formation. Within a TAD, though promoter and enhancer are separated by megabases but due to the loop formation mediated by CTCF and cohesion, the enhancer comes closer to the promoter [13, 14].

The contribution of chromosome folding, which brings the enhancer in the close proximity to the promoter were very well demonstrated in one of the recent studies on transcriptional regulation of mouse c-MYB gene. This gene encodes a transcription factor that activates several downstream genes to support cell proliferation. However, at the onset of differentiation, the transcription of this gene is turned OFF completely. Interestingly, transcription of c-MYB is attenuated at the first intron where a CTCF binding exist, and the enhancer elements are located 36 kilobase (kb), 68, 81 and 108 kb upstream of the transcription start site. When cells are actively proliferating, the three-dimensional conformation of chromosome is changed to make an active transcriptional hub where all those enhancer elements are brought in close proximity of the conserved CTCF binding site located within the intron. On the other hand, during differentiation, the 3D structure of the chromosome is perturbed which in turn destabilizes the formation of active transcriptional hub and downregulates the transcription up to several folds (**Figure 4**).

Posttranscriptional regulation: In order to understand the posttranscriptional regulation, we should have a much updated vision about the movement of RNA Pol through the gene bodies, which is an integral part of the transcription and was a subject discussion for a long period of time. Crystallographic studies of RNA PolII provide very important information about the structural aspects of transcription. Several recent high resolution crystallographic studies in this field indicated that two major transcription bubble fork forms at the upstream and the downstream of the DNA associated with the polymerase. The transcription bubble is a small amount of unwound double stranded DNA (11 bp), which is exposed to polymerase to synthesize the

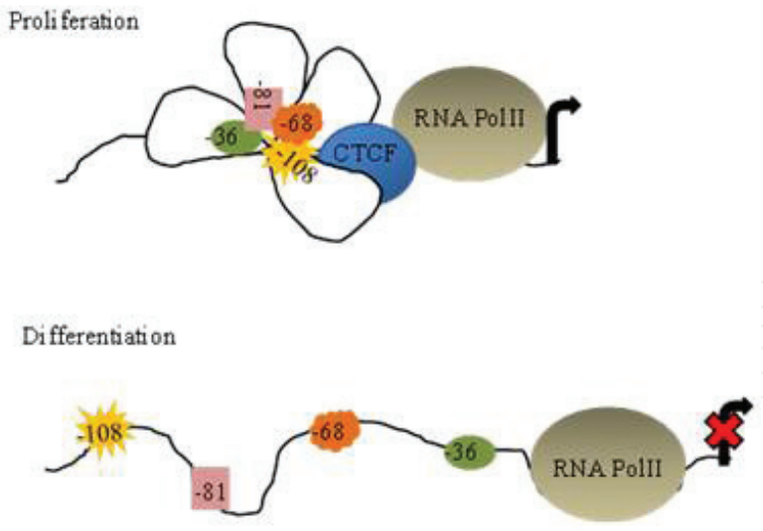


Figure 4. Enhancer-dependent regulation of mouse c-MYB gene. Transcription is upregulated several folds during proliferation when enhancer elements, located several kilobases away, fold to make an active transcription hub mediated through CTCF binding at the intronic region. During differentiation, loss of CTCF binding unfolds the structure and downregulates the transcription.

nascent RNA. The upstream fork forms a more open conformation and participates in DNA annealing and the synthesis of RNA transcripts; on the other hand, the downstream fork forms a rigid or closed domain with the non-template strand. This synchronized shift of open to close conformation allows the polymerase to translocate through the gene bodies [15].

Mammalian RNA Pol is a multi-subunit protein and the C-terminal domain (CTD), which is the biggest subunit of this protein, has several heptapeptide repeats (YSPTSPS). Phosphorylation of amino acid at the Serine-2 (Ser-2), Ser-5 and Ser-7 is very crucial for polymerase to start the transcription and several cofactors such as TFIIH (responsible for ser-5 phosphorylation) and cyclin-dependent kinase-7 (CDK7) or CDK9 (responsible for Ser-2 phosphorylation) mediate those phosphorylations.

Genome-wide chromatin immunoprecipitation (ChIP) experiments using antibodies against phospho-RNA PolII, followed by massive parallel sequencing, opened a new window about our understanding in transcriptional regulation particularly focusing on the movement and the distribution of RNA polymerase throughout the gene bodies. Several such studies indicated about 70% of the actively transcribing mammalian gene, the peak of the RNA PolII binding is located at the transcription start site and the availability of the polymerase along the gene body tapers off as we move along to the 3′ direction. Transcription starts after Ser-5 residues are phosphorylated predominantly by the cofactor TFIIH, whereas Ser-2 phosphorylation is insignificant [16]. For the remaining 30% of the gene, the major peak is located several bases downstream of the putative transcription start site indicating the polymerase stalls, though it initiated the transcription at the start site. In the presence of appropriate signal, the transcription complex recruits the P-TEFb complex, which is a heterodimer of CyclinT1 and CDK9. When RNA polII is stalled, it is associated with two major protein complexes those are DRB sensitivity inducing factor (DSIF) and negative elongation factor (NELF). Upon

recruitment of P-TEFb complex, the CDK9 phosphorylates the CTD of RNA PolII and the NELF and DSIF. The phospho-NELF is dissociated and DSIF continues with the polymerase and the process continues until the signal dies away. The function of P-TEFb complex is also regulated by 7SK snRNP-a small nuclear ribonucleoprotein associated with a core noncoding RNA bound with RNA binding protein HXIM (HXIM1 and 2). P-TEFb complexed with 7SK snRNP and HXIM1 are considered to functionally inactive and mammalian bromodomain protein Brd4 and human immunodeficiency virus Tat protein can replace the 7SK snRNP and make the functionally active P-TEFb complex [17, 18].

Discoveries of this pausing mediated regulation raised a few fundamental questions about the regulatory process. (A) Why nature devised this kind of additional regulatory system and (B) what are those genes that belong to this category?

Depending upon the nature of expression, mammalian genes can be classified into two categories. The genes that belong to the constitutive active class are expressing themselves in a continuous fashion. Most of the genes, that encode proteins to carry out biochemical and the metabolic pathways, transcribe almost continuously. However, cells possess another class of genes which are expressed under certain conditions. Expressions of those genes are restricted because abandoned expression may cause abnormalities in the biochemical or molecular pathways that control the natural activities of mammalian cells. This phenomenon was first noticed in the expression of the Drosophila heat-shock gene expression where the proteins are expressed under the exposure of a particular stress condition such as low or high temperature, UV exposure and starvation or under hypoxia (less oxygen tension). Detailed study underpinned that the polymerase synthesizes 25–30 nucleotide (nt) short RNA before it pauses and stays there at least 10 min before it moves. However, during heat shock condition, the RNA polII stays only less than ~4 s. In mammalian cells, most of the developmentally and immediate response genes are regulated following this mechanism. Now, how these immediate response genes in the mammalian system are regulated by this complex mechanism can be discussed further by taking an example of a well-studied system [19].

The human oncogenic transcription factor c-MYB (a counter part of the mouse gene described earlier) is responsible for the development of certain types of breast cancer and leukemia. Overexpression of this protein promotes uncontrolled cell proliferation by activating a bunch of genes that drive the proliferation. In human breast epithelial cells, the transcription of this gene is absolutely regulated by the hormone estrogen. In the absence of estrogen, the transcription is paused at the 1.7 kb downstream of the transcription start site and generates a short transcript with a stretch of poly adenylated (poly A) tail and in the presence of estrogen the polymerase resumes the transcription beyond the pausing site (**Figure 5**). The analysis of DNA sequence indicated that the nascent RNA generated under this condition has the potential to form a secondary hair pin structure, which is considered as a docking site of the P-TEFb complex. Estrogen receptors (ESR1) are a group of nuclear proteins that have ligand (estrogen)-binding domain as well as DNA binding and transactivation domain. Genome-wide ESR1 binding studies and our independent investigations identified a solo ESR1-binding domain close to the upstream of the transcription pausing site. Our in depth studies to understand the underlying mechanism of ESR1-mediated overcoming of the pausing revealed

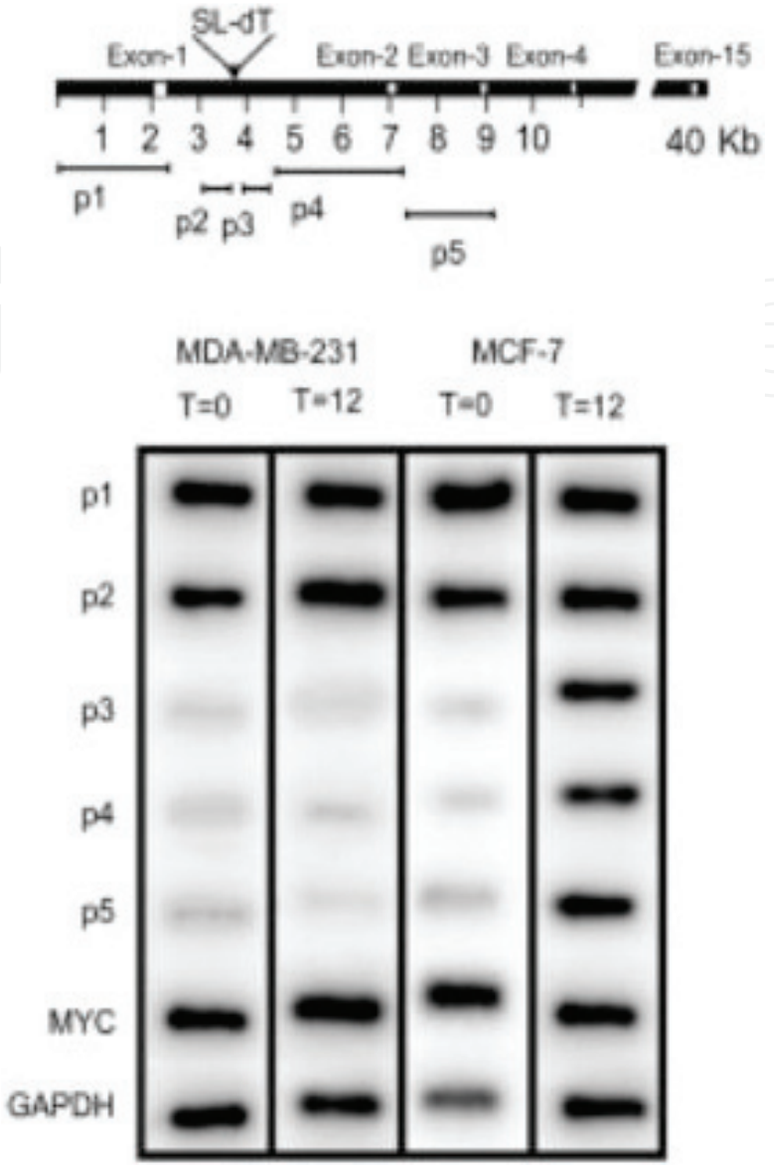


Figure 5. Estrogen-dependent transcription of c-MYB in breast cancer cells. A schematic diagram of human c-MYB gene showing exons and introns and the pausing site (SL-dT, Upper Panel). Nuclear run-on experiment showed that the transcription continues beyond the pausing site in the presence of estrogen in MCF7 cells but not in MDA-MB-231 cells, which do not express MYB (lower panel). Probes used in this hybridization were marked as P1, P2, P3, P4 and P5, and time of estrogen treatment was mentioned as T0 (control) and T12 (12 h).

that the ESR1 makes a tripartite complex with CyclinT1 and CDK9 and thus recruited the P-TEFb complex at the docking site. This recruited CDK9 phosphorylates the Ser-2 residue of the CTD of PolII and drives the transcription (**Figure 6**). Understanding the transcription regulation of oncogenic proteins also has significance in the field of cancer drug discovery. For example, the CDK9, which is playing such a pivotal role in the transcription of *c-MYB,* is a targetable molecule to develop a novel anticancer drug, and several such CDK9 inhibitors are currently under clinical trials to test their efficacies [20–22].

Epigenetic regulation: By definition, epigenetic modification is an inheritable process of regulation of gene expression without changing the DNA base pairs. The modifications take place

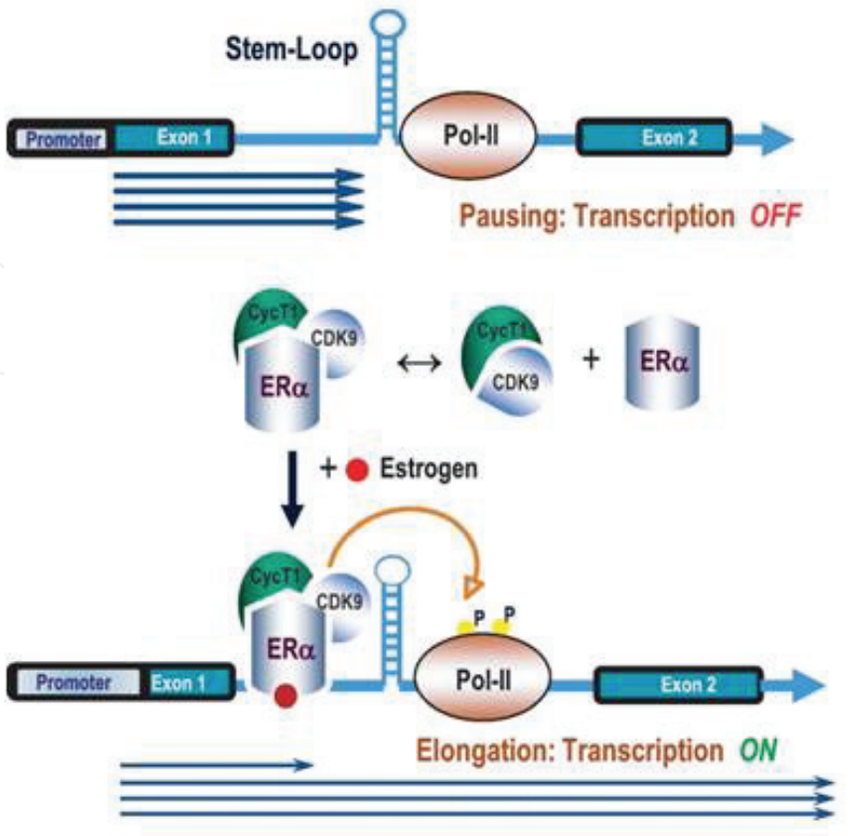


Figure 6. Estrogen-dependent transcription of *c-MYB* in breast cancer cells. In the absence of estrogen, the transcription is attenuated at the pausing site (upper panel). ERα forms a complex with CyclinT1 and CDK9 and in the presence of ligand; it binds DNA close to the pausing site, and the transcription resumes with subsequent phosphorylation of the CTD of the RNA PolII.

by enzyme-mediated inclusion or removal of the methyl groups in the nucleotides of the double stranded DNA or modifications of histone proteins by, for example, acetylation or deacetylation. Changes in the DNA bases or modification of the core histone proteins allow a particular portion of the chromatin accessible for the transcription complex or the repressor proteins to control the gene expression. Enzymes, those are responsible for the modification of DNA or histone proteins have been considered recently in a subject of in depth research in the context of their function and drug development in several diseases. Attempts are underway to develop novel therapeutics against diseases like cancer where the abnormal gene expression, caused by epigenetic modifications, contributes to the uncontrolled cell proliferation.

As discussed earlier, the epigenetic modifications can be subdivided into two different categories such as DNA modification and histone proteins modification. Traditionally, DNA modifications such as methylation happens when the enzyme DNA methyltransferases transfers the methyl group of a donor such as S-adenosylmethionine to a cytosine base and in most of the cases, it happens at the CpG (where the cytosine is connected to the guanosine by phosphate bonding) dinucleotide residues. Most of the CpG methylation in mammalian genome occurs at the outside of a stretch of elevated C- and G-rich region of the DNA called as CpG island and in case of human genome, this stretch is around 1 kb long and overlaps with the promoter

region of the 60–70% of the gene. Therefore, CpG methylation in the genome acts as a landmark for the transcription complex to locate site of the chromosomes ready for transcription. DNA methylation at the promoter site contributes significantly in the gene expression as it can be understood that this modification acts like a 'mask', which attenuates the access of the transcription factors. However, it was also very well demonstrated that this methylation is a dynamic and a completely reversible process, which further emphasizes the fact that gene expression can be controlled by manipulating the methylation of the promoter sequence. It was also observed that DNA methylation is abandoned in lots of genes which are permanently silent; on the other hand, significant DNA methylations were observed in several actively transcribing genes [23, 24]. Therefore, further research is needed to understand the significance of DNA methylation in mammalian gene expression.

DNA methylation and its relation to the gene expression have been found to be strongly correlated with the development of diseases such as cancer. A large scale meta-analysis of the methylation profiles of target genes, which includes oncogenes and tumor suppressor proteins in several cancer tissues such as breast, colon and lung indicated that the promoter methylation patterns are significantly different in those tissues in comparison to their normal counter parts. The above statement can be illustrated further by using the very well-studied WNT-β-catenin pathway which is one of the most frequently dysregulated in renal cancer. The proto-oncogene β-catenin, which is the downstream target of WNT pathway, activates the expression of several proteins that promotes tumorigenesis such as proto-oncogene c-MYC and CyclinD1. Expressions of several key regulators that negatively regulate the WNT-β-catenin pathway are controlled by the promoter methylation, which eventually drives to the uncontrolled synthesis of β-catenin and the activation of the downstream target genes. For example, in case of renal carcinoma, expression of several WNT-inhibitor factors (WIF1, at least four, Dickkopf (DKK1 or 2) and IGFBP1 (insulin-like growth factor binding protein 1) are downregulated by promoter methylation [25]. In their studies, Moarii et al. [26] showed a significant amount of modification of the promoter methylation in the cancer tissues targeting to the transcription factor. Expression of several genes that are reported to be associated with the cell cycle (p16INK4a, p15INK4b, p14ARF), DNA repair (hMLH1, MGMT), apoptosis (DAPK), tumor suppression (p53) are downregulated or modified due to promoter methylation. For example, in case of p53, in vitro promoter methylation studies indicated that the DNA methylations can downregulate more than 90% of the mRNA expression. A further support of these data came along when the analysis of p53 expression in correlation to the promoter methylation was studied in vivo in patient samples. Several such studies indicated that the aberrant DNA hypermethylation of the p53promoter strongly correlates with the attenuated expression of this gene in a significant portion of the primary hepatocellular carcinoma, breast cancer, acute lymphoblastic leukemia (ALL) and chronic lymphocytic leukemia (CLL) patients. An increased expression of DNA methyltransferase (DNMT) activities have been noticed in several cancer cells which encouraged scientist to proposed the hypothesis that this enhanced activities hypermethylate the promoter of the tumor suppressor genes such as p53, which eventually promotes tumor development [27]; and therefore, a promising approach would be to develop an inhibitor against DNMTs to upregulate the expression of the tumor suppressor genes. Two of such inhibitors, azacytidine and decitabine, have been considered as the most successful drugs though their applications

are restricted due to the toxic side effect. However, this outcome encouraged researchers to develop new drugs with less toxic side effects, and currently few of them are under clinical trials with promising results [28].

Posttranslational modification of histone proteins is one of the very well-studied epigenetic modifications. The mammalian chromosomes are compacted into the nucleus by forming the primary and several higher order structures by the building block nucleosomes. Each nucleosome comprises a histone octamer (four core histones H2A, H2B, H3, and H4 in duplicates) surrounded by 146 base pair of DNA where the amino terminal (N-terminal) part of histone protein protrudes out of the histone-DNA assembly. The N-terminal modification of the core histone proteins is very common and those modifications are acetylation, phosphorylation, methylation, sumoylation and ubiquitination. All those modifications are unique in a sense because each one of them introduces a specific change in the secondary and higher order structure of the chromatins which in turn contributes to the gene expression. For example, histone acetylation occurs at the lysine residue by the enzyme histone acetyl transferases (HATs) and it is associated with the transcription activation. However, the histone deacetylases (HDACs) remove the acetyl group and thereby suppresses the transcription. Dynamic regulation of acetylation and deacetylations of chromosomes have shown to play a very important role in the regulation of gene expression and propagation of disease. Acetylation of histone by HATs has been shown to open the chromatin structure, which allows the transcription factor to access that region. Research in last more than one decade established a strong connection between HATs and HDACs with diseases such as cancer. Extensive research in this field revealed that the malfunction of enzymes related to these activities can cause aberrant cell proliferation and differentiation. Recent studies established a very strong correlation between histone acetylation and deacetylation with the development of several types of cancer such as in hematopoietic malignancies and observed that HAT or HDACs are the common target of mutagenesis. Due to their significant role in disease development, HAT and HDACs are considered as important targets of drug development. One of the HDAC inhibitors suberoylanilide hydroxamic acid (SAHA, marketed by Merck as Zolinza) has been shown to be remarkably effective in the treatment of cutaneous T-cell lymphoma (CTCL). Another HDAC inhibitor, Panobinostat (marketed by Novartis) has been approved to treat multiple myeloma and currently under clinical trials to develop as a drug to treat ovarian and certain types of blood cancer. Similarly, modification of histones by methylation also contributes to the transcription. The methylation marks on chromosomes are recognized by the transcription factor or cofactors to locate the region of the chromosome ready for transcription. The amount of methylation also contributes to the activation and the repression of genes. For example, monomethylation of histone H3 lysine-4 form the N-terminal end (H3K4) is associated with both activation and repression, whereas trimethylated H3K4 is only associated with the repression (**Figure 7**).

Translational regulation: So far, we have discussed the regulation of mammalian gene expression where proteins such as enzymes or transcription factors play the key role in turning/switching ON or OFF the gene expression. In the beginning of early 1990s, researchers discovered a short stretch of non-coding RNA, known as miRNA, highly conserved across the species, regulates gene expression at the level of translation. Since then, list of miRNAs has been piling up each year and currently more than 2000 miRNAs were reported with their functional association in gene expression, cell proliferation and differentiation. The analysis of human genome

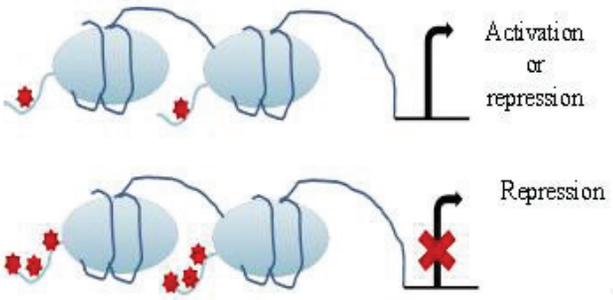


Figure 7. Methylation of histone is a mark of transcriptional activity. Transcription can still be active under monomethylated condition (black stars) of core histone protein but completely or permanently shuts down when histones are trimethylated.

sequence indicated that genes encode miRNA are located either in the intergenic (between two genes) or in the intragenic (located within the gene) region. The intergenic miRNA are transcribed by the independent promoter, whereas intragenic one is transcribed by the same gene-specific promoter. Both miRNAs are synthesized in the form of pre-miRNA that are several kilobases long and later are processed in the nucleus and in cytoplasm to generate a short hair pin, functionally active form.

Extensive studies in the last decade or more have generated a significant amount of evidence describing the mechanism of action of those miRNAs. Nevertheless, it is a subject of ongoing

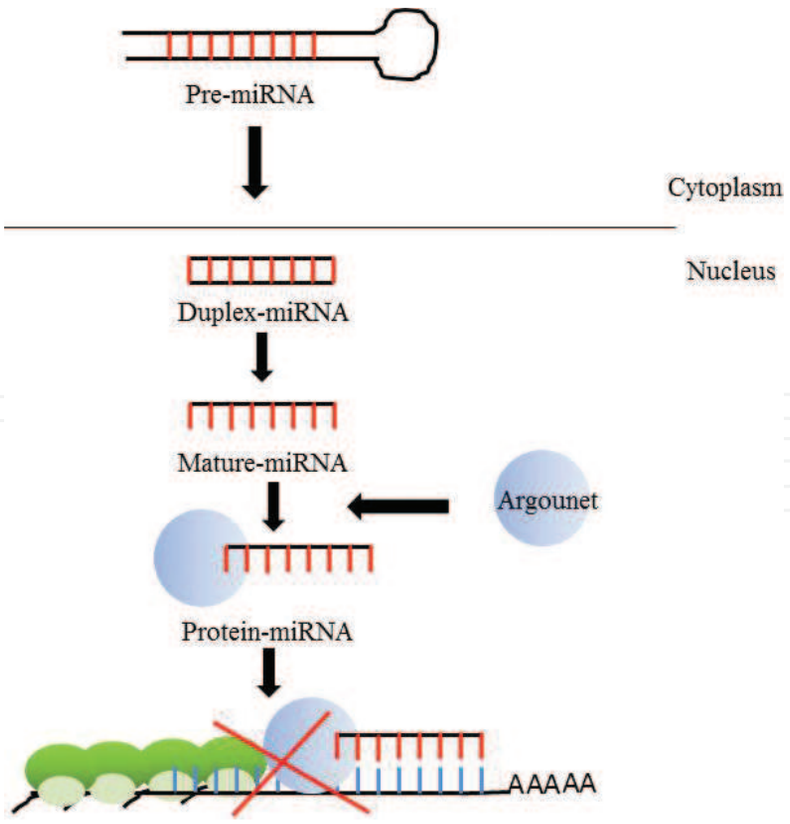


Figure 8. A model of miRNA-mediated regulation of gene expression. Several base pair long pre-miRNA is processed to short duplex miRNA. The sense strand forms a complex with RNA binding protein Argonaut. The RNA strand hybridizes with the 3′UTR of the target gene and the complex blocks the movement of the ribosomes.

study to understand how miRNAs are controlling gene expression. The posttranscriptional regulation is perhaps one of the most established mechanisms for the miRNA mediated control of gene expression. Several observations suggest that miRNA form a complex with the protein Argonaut, which is a highly conserved RNA binding protein. The specificity of the base pairing between miRNA-mRNA follows the Watson-Crick law where the 5′ proximal end of the miRNA forms a 2–8 bp of double-stranded RNA with the 3′ untranslated region (3′UTR) of the target mRNA. This miRNA-mRNA hybridization initiates several processes simultaneously to inhibit the gene expression. For example, the secondary structure form due to the hybridization causes premature termination and slowed elongation of translation, and at the same time, it stimulates the ribosomal drop off. The miRNA-protein complex recruits several factors and co-factors including the endonucleases to degrade the template RNA. On the other hand, Argonaut competes with the 5′mRNA CAP binding protein and elongation factor to prevent transcription initiation. Besides downregulating the gene expression, miRNA mediated upregulation has also been reported (**Figure 8**). However, the mechanism by which the activation occurs is not clear, but it has been proposed that the miRNA-protein complex perhaps inactivates the other miRNA that downregulates the gene expression. Understanding of the function of miRNAs and the complexity of their function was further revealed by the fact that a single miRNA has been shown to be acting like an activator or a repressor. The miR-145 upregulates the expression of the gene myocardin, which encodes a transcription factor that requires muscle cell differentiation. However, the expression of the Rho-associated coiled-coil contain ing protein kinase 1(ROCK1) is downregulated by the same miRNA-145 during osteosarcoma [29].

2. Conclusion

The complexity of the mammalian gene expression is more than what has ever been previously conceived due to the continuous accumulation of information over the period of last several decades. Researchers are trying to understand the molecular basis of every step associated with this process by utilizing cutting edge technologies and thus generating an enormous amount of data, which will take perhaps several decades to validate. The author's effort here is to provide a broad overview of the regulation of gene expression in the mammalian cells. It is hoped that this content will motivate readers to put their efforts to explore further the phenomenal complexity underlying the entire process and translate that knowledge in developing new therapeutics.

Acknowledgements

The data and the models presented here were generated based on the published data from the laboratory of Prof. Gary Stein, Chairman, Department of Biochemistry, University of Vermont and Prof. Thomas J. Gonda, University of South Australia. The generous support of Prof. Erik Thompson, Queensland University of Technology, was also appreciated for providing necessary resources to write the manuscript.

Author details

Partha P. Mitra

Address all correspondence to: p.mitra@uq.edu.au

School—Biomedical Sciences, Queensland University of Technology, Australia

References

[1] Dvir A, Conaway RC, Conaway JWA. Role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. Proceedings of the National Academy of Sciences of the United States of America. 1997;**94**(17):9006-9010

[2] Pugh BF, Tjian R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes & Development. 1991;**5**(11):1935-1945

[3] Kamenova I, Warfield L, Hahn S. Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. Molecular and Cellular Biology. 2014;**34**(15):2929-2943

[4] Baatar D, Jones MK, Pai R, Kawanaka H, Szabo IL, Moon WS, et al. Selective cyclooxygenase-2 blocker delays healing of esophageal ulcers in rats and inhibits ulceration-triggered c-Met/hepatocyte growth factor receptor induction and extracellular signal-regulated kinase 2 activation. The American Journal of Pathology. 2002;**160**(3):963-972

[5] Allen BL, Taatjes DJ. The mediator complex: A central integrator of transcription. Nature Reviews Molecular Cell Biology. 2015;**16**(3):155-166

[6] Opitz OG, Rustgi AK. Interaction between Sp1 and cell cycle regulatory proteins is important in transactivation of a differentiation-related gene. Cancer Research. 2000;**60**(11):2825-2830

[7] Mitra P, Ghule PN, van der Deen M, Medina R, Xie RL, Holmes WF, et al. CDK inhibitors selectively diminish cell cycle controlled activation of the histone H4 gene promoter by p220NPAT and HiNF-P. Journal of Cellular Physiology. 2009;**219**(2):438-448

[8] van Wijnen AJ, van den Ent FM, Lian JB, Stein JL, Stein GS. Overlapping and CpG methylation-sensitive protein-DNA interactions at the histone H4 transcriptional cell cycle domain: Distinctions between two human H4 gene promoters. Molecular and Cellular Biology. 1992;**12**(7):3273-3287

[9] Aziz F, van Wijnen AJ, Vaughan PS, Wu S, Shakoori AR, Lian JB, et al. The integrated activities of IRF-2 (HiNF-M), CDP/cut (HiNF-D) and H4TF-2 (HiNF-P) regulate transcription of a cell cycle controlled human histone H4 gene: Mechanistic differences between distinct H4 genes. Molecular Biology Reports. 1998;**25**(1):1-12

[10] Vaughan PS, Aziz F, van Wijnen AJ, Wu S, Harada H, Taniguchi T, et al. Activation of a cell-cycle-regulated histone gene by the oncogenic transcription factor IRF-2. Nature. 1995;**377**(6547):362-365

[11] Mitra P, Xie RL, Medina R, Hovhannisyan H, Zaidi SK, Wei Y, et al. Identification of HiNF-P, a key activator of cell cycle-controlled histone H4 genes at the onset of S phase. Molecular and Cellular Biology. 2003;**23**(22):8110-8123

[12] Miele A, Braastad CD, Holmes WF, Mitra P, Medina R, Xie R, et al. HiNF-P directly links the cyclin E/CDK2/p220NPAT pathway to histone H4 gene regulation at the G1/S phase cell cycle transition. Molecular and Cellular Biology. 2005;**25**(14):6140-6153

[13] Ghirlando R, Felsenfeld G. CTCF: Making the right connections. Genes & Development. 2016;**30**(8):881-891

[14] Ong CT, Corces VG. CTCF: An architectural protein bridging genome topology and function. Nature Reviews Genetics. 2014;**15**(4):234-246

[15] Barnes CO, Calero M, Malik I, Graham BW, Spahr H, Lin G, et al. Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. Molecular Cell. 2015;**59**(2):258-269

[16] Ghamari A, van de Corput MP, Thongjuea S, van Cappellen WA, van Ijcken W, van Haren J, et al. In vivo live imaging of RNA polymerase II transcription factories in primary cells. Genes & Development. 2013;**27**(7):767-777

[17] Zaborowska J, Egloff S, Murphy S. The pol II CTD: New twists in the tail. Nature Structural & Molecular Biology. 2016;**23**(9):771-777

[18] Chen R, Yik JH, Lew QJ, Chao SH. Brd4 and HEXIM1: Multiple roles in P-TEFb regulation and cancer. BioMed Research International. 2014;**2014**:232870

[19] Lis JT, Mason P, Peng J, Price DH, Werner J. P-TEFb kinase recruitment and function at heat shock loci. Genes & Development. 2000;**14**(7):792-803

[20] Drabsch Y, Hugo H, Zhang R, Dowhan DH, Miao YR, Gewirtz AM, et al. Mechanism of and requirement for estrogen-regulated MYB expression in estrogen-receptor-positive breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2007;**104**(34):13762-13767

[21] Mitra P, Pereira LA, Drabsch Y, Ramsay RG, Gonda TJ. Estrogen receptor-alpha recruits P-TEFb to overcome transcriptional pausing in intron 1 of the MYB gene. Nucleic Acids Research. 2012;**40**(13):5988-6000

[22] Mitra P, Yang RM, Sutton J, Ramsay RG, Gonda TJ. CDK9 inhibitors selectively target estrogen receptor-positive breast cancer cells through combined inhibition of MYB and MCL-1 expression. Oncotarget. 2016;**7**(8):9069-9083

[23] Graff J, Kim D, Dobbin MM, Tsai LH. Epigenetic regulation of gene expression in physiological and pathological brain processes. Physiological Reviews. 2011;**91**(2):603-649

[24] Illingworth RS, Bird AP. CpG islands—'a rough guide'. FEBS Letters. 2009;**583**(11):1713-1720

[25] Bennett LB, Taylor KH, Arthur GL, Rahmatpanah FB, Hooshmand SI, Caldwell CW. Epigenetic regulation of WNT signaling in chronic lymphocytic leukemia. Epigenomics. 2010;**2**(1):53-70

[26] Moarii M, Boeva V, Vert JP, Reyal F. Changes in correlation between promoter methylation and gene expression in cancer. BMC Genomics. 2015;**16**:873

[27] Saldana-Meyer R, Recillas-Targa F. Transcriptional and epigenetic regulation of the p53 tumor suppressor gene. Epigenetics. 2011;**6**(9):1068-1077

[28] Gnyszka A, Jastrzebski Z, Flis S. DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer Research. 2013;**33**(8):2989-2996

[29] Valinezhad Orang A, Safaralizadeh R, Kazemzadeh-Bavili M. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. International Journal of Genomics. 2014;**2014**:970607

IntechOpen