

# 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](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Modulation of Gene Expression by Gene Architecture and Promoter Structure

---

Aditya Kumar and Manju Bansal

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.76051>

---

## Abstract

Regulation of gene expression is achieved by the presence of *cis* regulatory elements; these signatures are interspersed in the noncoding region and also situated in the coding region of the genome. These elements orchestrate the gene expression process by regulating the different steps involved in the flow of genetic information. Transcription (DNA to RNA) and translation (RNA to Protein) are controlled at different levels by different regulatory elements present in the genome. Current chapter describes the structural and functional elements present in the coding and noncoding region of the genome. Further we discuss role of regulatory elements in regulation of gene expression in prokaryotes and eukaryotes. Finally, we also discuss DNA structural properties of regulatory regions and their role in gene expression. Identification and characterization of *cis* regulatory elements would be useful to engineer the regulation of gene expression.

**Keywords:** DNA structural properties, gene architecture, gene expression, promoter structure

---

## 1. Introduction

Genome, the blue print of life, is essentially comprised of coding (genes) and noncoding (regulatory regions and other repetitive sequences) DNA. Genetic information is embedded in the form of coding regions (genes) that encode proteins. This flow of information from gene to proteins is a multistep pathway *viz.* transcription that is synthesis of RNA from the DNA and continues with the translation which is protein synthesis from RNA. Control of this flow of information is crucial for fate of the cell and this phenomenon is known as the regulation of the gene expression. The function of the cell is determined by the amount and type of the

RNA and protein molecules that is achieved by the regulation of the gene expression. There are various steps involved in this flow of information process such as chromatin domain organization, transcription (initiation, elongation and termination), post-transcriptional modification, RNA export (exclusive for eukaryotes), translation and mRNA degradation. Among all these different regulated stages of gene expression transcription initiation is the most utilized point of regulation. Transcription event is coupled with the translation process in the case of prokaryotes due to availability of ribosomes in the same compartment (due to lack of nucleus). However, transcription process is far more complicated in case of eukaryotes due to involvement of additional steps that are RNA splicing and RNA export. These additional steps provide extra stages for the regulation of gene expression process in eukaryotes.

Regulation of gene expression is achieved by harnessing the regulatory elements, located in the noncoding as well as coding regions of the genome. Current chapter focuses on the different structural and functional elements present in the coding regions (genes) and noncoding regions (regulatory regions), which are utilized by the cell to regulate the gene expression process.

## 2. Gene architecture

### 2.1. Noncoding elements of the genes

Genes are the repositories for primary information content of inheritance in genome and their expression determines the phenotypes, which in turn decides future of the cell in multicellular organisms. Functioning of gene products *viz.* mRNA (messenger RNA) and ncRNA (non-coding RNA) is modulated by complex gene regulatory networks. Eukaryotic genomes are mostly comprised of compositional properties (repetitive sequences, codon usage bias, mutational information, etc.) and functional signals (TATA box, Inr-element, cap signal, Kozak sequence, GT-AG splicing sites, etc.) [1]. Processing of the transcript is an important phase in the gene expression process, which also provides additional level of regulation in eukaryotes. Transcription and translation events are coupled in prokaryotes due to the availability of ribosomes to the mRNA while transcript undergoes several levels of processing in nucleus and finally processed transcripts are exported to the cytoplasm for translation in eukaryotes. Complexity in the gene structure results into the phenotypic diversity and this complexity arises from the occurrence and arrangement of the noncoding elements interspersing the coding region. Gene expression diversification is achieved by the presence of trailer sequences known as untranslated region (UTR) and intervening noncoding sequences known as introns [2]. These elements exert several direct and indirect functions.

### 2.2. Untranslated regions

UTRs are the trailer sequences located at the 5' and 3' end of the coding region which are the part of the transcribed mRNA but remain untranslated. Presence of alternative promoters or more than one transcription start site result into multiple 5' UTRs which in turn controls the gene expression in several ways [3–6]. G quadruplex or G4 structure is a predominant secondary structure situated in the guanine rich 5' UTRs which in turn hinders the translation

process [7–9]. Highly and constitutively expressed genes are associated with short and poor in guanine base 5' UTR in order to facilitate the translation process [10]. Sequential unwinding of natural stem loop structures located in the 5' UTR in some mRNA is found to be associated with efficient translation [11–13].

IRES (internal ribosome entry sites), located usually upstream of the initiation codon (in the 5' UTR) are responsible for the translation initiation in a cap independent mechanism by recruiting ribosome near the initiation site [5, 14–16]. The IRES mediated translational regulation occurs under certain stress conditions such as cellular stress, nutritional stress, mitotic stress etc. [17–19]. Conserved upstream open reading frames (uORFs) located in the 5' UTR are also found to regulate protein translation, which are followed by main start codon (AUG) in the downstream [20–22]. Antibiotic resistance in the pathogenic bacteria is also found to be associated with uORFs mediated regulation [23]. In a recent study, fusion of uORF in the upstream of the auto-activated immune receptor gene developed the resistance to the plant diseases in Arabidopsis and rice [24].

Apart from these regulatory regions located in the 5' UTR, the 3' UTR is also rich in regulatory sequences located at the end of the coding sequence or gene. The conserved motif/s associated with 3' UTR play crucial roles in gene expression at the posttranscriptional level. The 3' UTR perform various regulatory functions, which are providing stability to the mRNA by polyadenylation, transcript cleavage, serve the binding site for microRNAs etc. Different isoforms of mRNA are derived from the alternative splicing and polyadenylation with alternative 3' UTR. The varying expression levels and spatiotemporal localization for the same protein is achieved by differing 3' UTR sequence in human [25–27]. AU rich elements (AREs) which are 50 to 150 nucleotide long and associated with multiple copies of pentanucleotide AUUUA regulate gene expression by stabilizing the mRNA [28, 29]. The abundance of AREs in the 3' UTR of wide range of gene families indicates significant role in the gene regulation process [30]. MicroRNA response elements (MREs) are mostly located in the 3' UTR region where single stranded 22 nucleotide long microRNA binds to regulate the expression of mRNA [31]. Poly(A) tail is stretch of adenosine (around 250 nucleotide) attached at the 3' end of the RNA by adenylation process. The poly(A) binding proteins (PABP), specific class of regulatory proteins (nuclear and cytoplasmic) binds to the poly(A) tail and perform different regulatory functions like stability of mRNA, export and decay of the mRNA. These proteins play vital role in gene regulation [32–35].

### 2.3. Intronic regions

An intron is a noncoding DNA sequence that is transcribed but not translated; it is removed during the processing of pre-mRNA (precursor mRNA) to final mature RNA also known as RNA splicing. There are four different types of introns based on different splicing mechanisms. Spliceosomal introns are the foremost discovered and well characterized introns, which are excised by spliceosome, a ribonucleoprotein complex [36, 37]. Group I introns, widely present in mRNA, rRNA and tRNA of variety of organisms including algae, fungi, lower eukaryotes and few bacteria [38–42]. Similarly, group II introns are large autocatalytic ribozymes widely present in the mitochondria, chloroplast, plants, fungi, yeast and many bacteria, play major role in genome evolution [43–46]. The tRNA introns widely present in all domains of life are

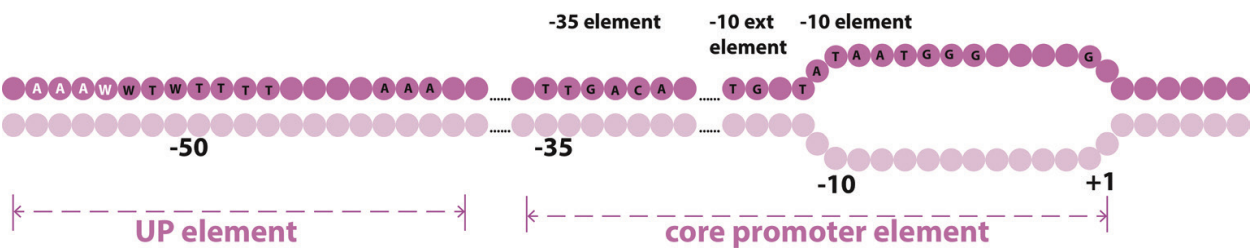
exceptionally different as enzymes are involved in the removal of intron and in the joining of the two halves [47–49]. Gene regulation is modulated to a great extent by count or number, length and position of the introns and they have several direct and indirect biological functions [50]. Multiple protein isoforms of the same gene are derived from the regulated alternative splicing process in eukaryotes [51–54]. Introns modulate gene expression either by the presence of transcriptional regulatory elements or by intron mediated enhancements [55–57]. Introns also regulate the gene expression by mediating the chromatin assembly (chromatin structure modulation) and controlling the mRNA export [58–61]. Apart from these direct biological functions, introns also exert indirect influence, for example position and length of the intron in the gene have potential role in the regulation of the expression level of the transcript [62–65].

### 3. Promoter structure

#### 3.1. Different promoter elements

Promoters are stretch of genomic sequences where assembly of transcription machinery (RNAP and other accessory proteins) takes place prior to initiation of transcription. Although prokaryotic and eukaryotic polymerase shares functional similarity, promoter architecture differs in complexity [66]. Single type of RNA polymerase along with the specific  $\sigma$  factors recognizes promoter elements in prokaryotes [67]. Where  $-10$  and  $-35$  elements located in the upstream of the transcription start sites (TSSs) are recognized by different domains specific  $\sigma$  factors while UP element, an AT rich sequence situated from  $-40$  to  $-60$  is recognized by CTDs of  $\alpha$  subunit of RNAP (**Figure 1**). An extension of extended  $-10$  element,  $-15$  element (TGnT) has been also proposed as new element situated from  $-15$  to  $-12$ . It has been found that  $-15$  element determines the overall promoter strength by complementing the weak  $-10$  element.

On the other hand, complexity of promoter architecture in eukaryotes increases from yeast to mammals. Different types of RNA polymerases (normally three) are responsible for the generation of variety of RNA such as ribosomal RNA, messenger RNA (mRNA) and tRNA. As in case of bacterial RNAP, archaeal RNA polymerase and eukaryotic RNA Pol II (responsible for transcribing mRNA) also require specific factors and promoter elements to initiate transcription at specific sites in the genome. Eukaryotic promoters can be broadly classified in to three categories such as core, distal and proximal. The core promoter (approximately 50 nucleotide



**Figure 1.** Prokaryotic promoter structure: UP element and core promoter elements ( $-35$ ,  $-10$  extended and  $-10$  element).



sequence) is a platform where assembly of RNA polymerase and associated general transcription factors (GTFs), collectively referred as pre- initiation complex (PIC) takes place [68, 69]. Various promoter elements (**Table 1**) in the vicinity of the transcription start site; upstream and downstream regions are recognized by Pol II and other factors, such as TATA box, are recognized by TATA-binding protein (TBP), the B recognition element (BRE) by TFIIB and other elements by TBP-associated factors (TAFs) [70] (**Figure 2**). Apart from these, core promoter regions also consist of Inr element and may also contain downstream elements like downstream promoter element (DPE), motif ten element (MTE) (in humans) [71].

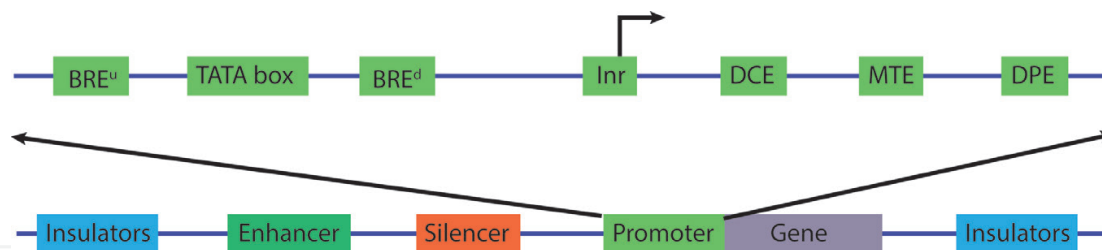
Proximal promoters are located in the immediate upstream (up to a few hundred base pairs) of core promoter, are comprised of GC box, CAAT box, *cis*-regulatory modules (CRM) etc. CpG islands are stretch of short DNA sequences, which are rich in GC content located in the upstream of house keeping and other regulated gene promoters [72, 73]. Proximal promoters mostly work as tethering element for distal promoters instead of acting as direct activators. On the other hand, distal promoters work from long distance. Enhancers, silencers and insulators are present in the distal promoter regions. Enhancers, also known as “promoters of promoter” mainly control specificity of gene expression by deploying unique enhancers in deferent cell types [74]. Multiple enhancers associated with single gene and single enhancer activating multiple genes provides additional level of diversity in phenotypes. In contrast to other regulators, enhancers exert their effects over tens of kilobases of DNA [75, 76]. Silencers are sequence specific elements where negative transcription factors bind to down regulate the gene expression [77]. Insulators are also referred to as boundary elements which block the effect of transcriptional activity of neighboring genes [77, 78].

### 3.2. Promoter structure and nucleosome dynamics

The locations and strengths of transcription factor and RNAP binding sites, also known as *cis*-regulatory elements and list of all nucleosome-binding sites are collectively defining the promoter structure. Nucleosomes are not only involved in the packaging of DNA but also bring order to the eukaryotic genome by regulating replication and transcription [79, 80]. Nucleosomes provide the first line of defense to avoid the unwanted transcription initiation.

Name	Location (relative to TSS at +1)	Associated factor/s
BRE <sup>a</sup>	Upstream of the TSS	TFIIB
TATA box	-30/-31 to -23/-24	TBP
BRE <sup>d</sup>	Downstream of the TATA box	TFIIB
Inr	-2 to +4/+5	TAF1 & TAF2
DCE	+6 to +11, +16 to +21, +30 to +34	TAF1
MTE	+18 to +29	TAF6 & TAF9
DPE	+28 to +33	TAF6 & TAF9

**Table 1.** List of core promoter elements and factors associated with them [72, 73].



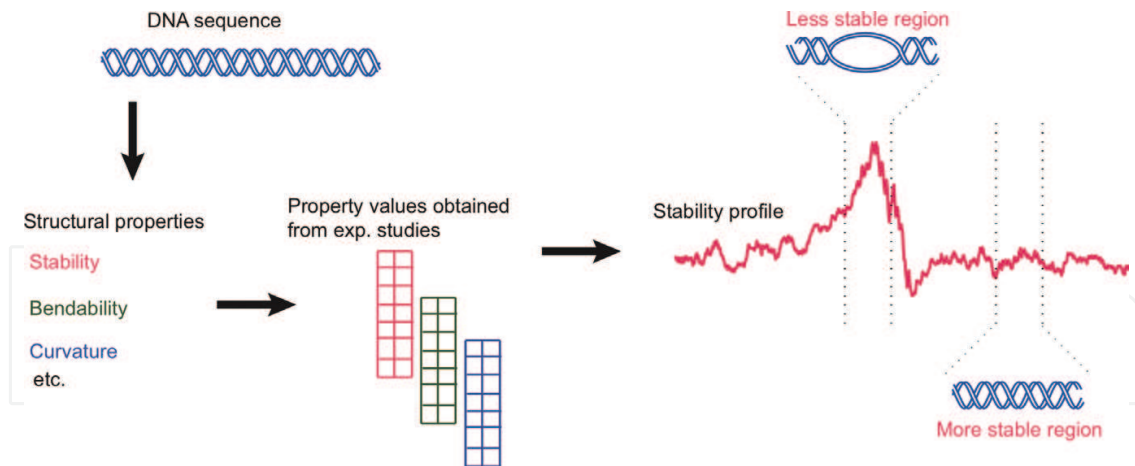
**Figure 2.** The different types of gene regulatory elements in eukaryotes.

Gene promoters involved in active transcription require accessibility to the DNA by RNAP machinery and associated factors, which is facilitated by nucleosome free region (NFR) or nucleosome-depleted region (NDR) [81, 82]. Nucleosome positioning is the probability of finding nucleosome at given genomic location relative to the surrounding locations while nucleosome occupancy refers to the average number nucleosomes present at the given genomic location in a given population of cells [83, 84]. Cellular gene expression is the final outcome of nucleosome dynamics, which itself depends on a complex interplay between nucleosome positioning and occupancy [85–87].

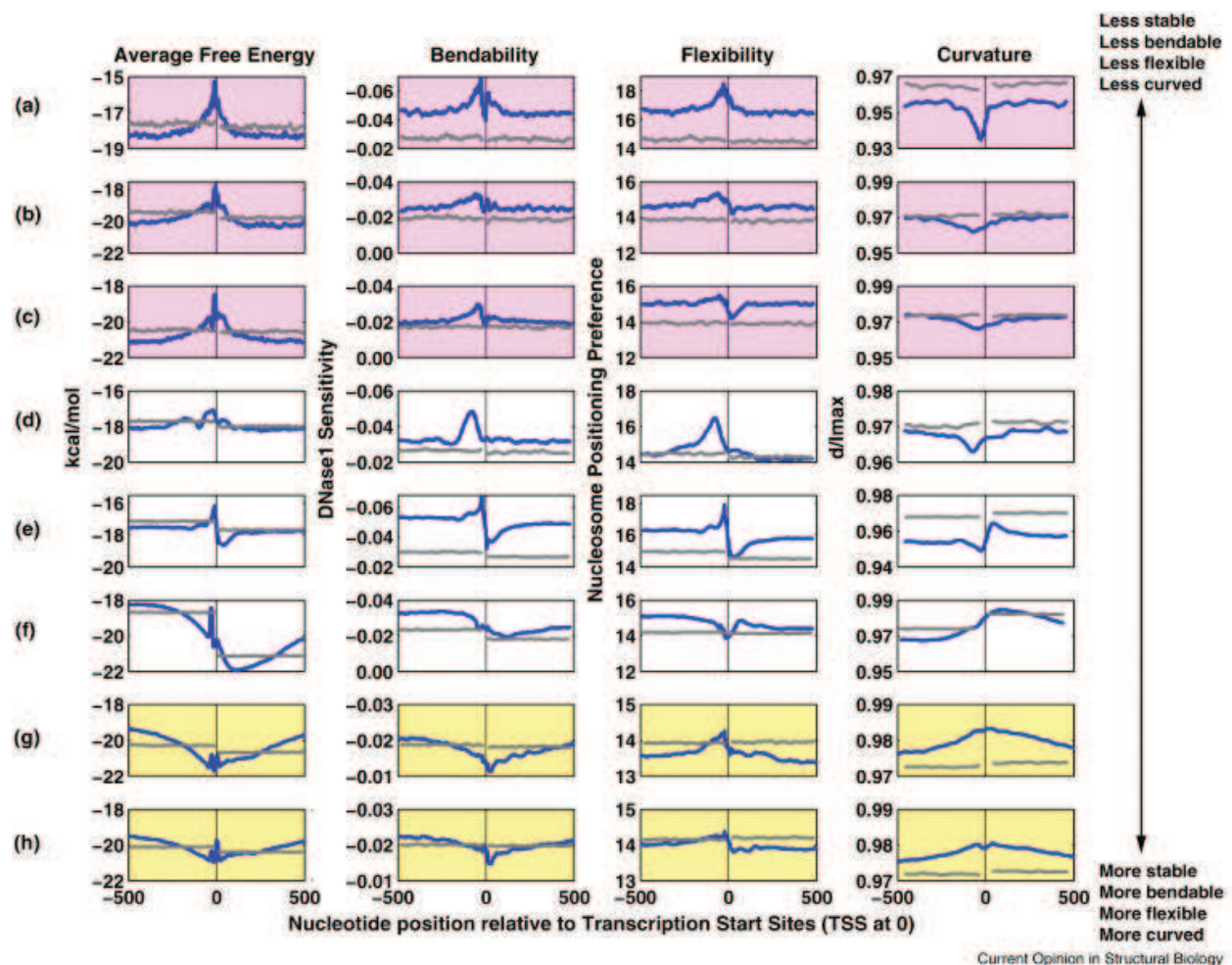
### 3.3. DNA structural properties of promoter regions

DNA sequence not only determines the distinct or base specific interactions but also determines the overall conformational shape, which is recognized by different proteins in case of non-base specific interactions [88]. The higher DNA binding specificity is achieved by combining different readout mechanisms by DNA binding proteins, with DNA shape playing an important role in gene regulation and genome organization [89]. The DNA sequence dependent structural properties can be roughly divided into two categories, conformational and physiochemical [90]. Conformational properties represent the static DNA structure, which are influenced by geometry of base pair steps described by translational (shift, slide and rise) and rotational (tilt, roll and twist) parameters [91]. These also determine variation in the major and minor groove dimensions, which are crucial for DNA protein interactions. The physiochemical properties refer to the dynamic DNA structural properties such as persistence length, stress induced duplex destabilization, DNA duplex stability, protein induced bendability and intrinsic curvature etc.

Structural properties of given DNA sequence can be calculated using different di, tri tetra nucleotide models reported in experimental as well as theoretical studies. These models provide property values (lookup tables) for different oligonucleotides and using these values and appropriate length (sliding window), a given DNA sequence can be converted into a series of numerical values referred to as a structural property profile. These profiles of given DNA sequence show variation in the structural property over the different regions of the sequence (**Figure 3**). An average structural property profile is calculated by taking mean of the feature value over all positions by aligning the different sequences [92]. DNA structural features such as low stability, protein induced bendability and intrinsic curvature are consistently observed in the prokaryotic and eukaryotic promoters (**Figure 4**) [93–96]. Promoter regions of different categories of transcripts (primary, internal, antisense and noncoding RNA) present in prokaryotic transcriptome show distinctly different DNA



**Figure 3.** Schematic illustration showing DNA structural properties profile (example shown for DNA duplex stability) using values for di, tri, tetra nucleotide etc. obtained from experimental studies. Stability profile shows variation depending on the DNA sequence.



**Figure 4.** DNA sequence dependent structural properties of the promoter regions (-500 to +500 with respect to transcription start site at 0 position). Profiles of four structural properties (DNA duplex stability, DNase I sensitivity, Nucleosome Positioning Preference and intrinsic curvature) are shown of eight model systems: (a) *H. pylori*, (b) *E. coli*, (c) *K. pneumoniae*, (d) *S. cerevisiae*, (e) *C. elegans*, (f) rice, (g) mouse and (h) human. Figure taken (with permission) from Bansal et al., [96].



structural features [97]. Moreover, promoter regions of orthologous genes show conserved DNA structural properties in prokaryotes and plants [98–100]. These findings suggest that the DNA structural properties of promoter regions are conserved across the various classes of organisms.

## 4. Modulation of gene expression

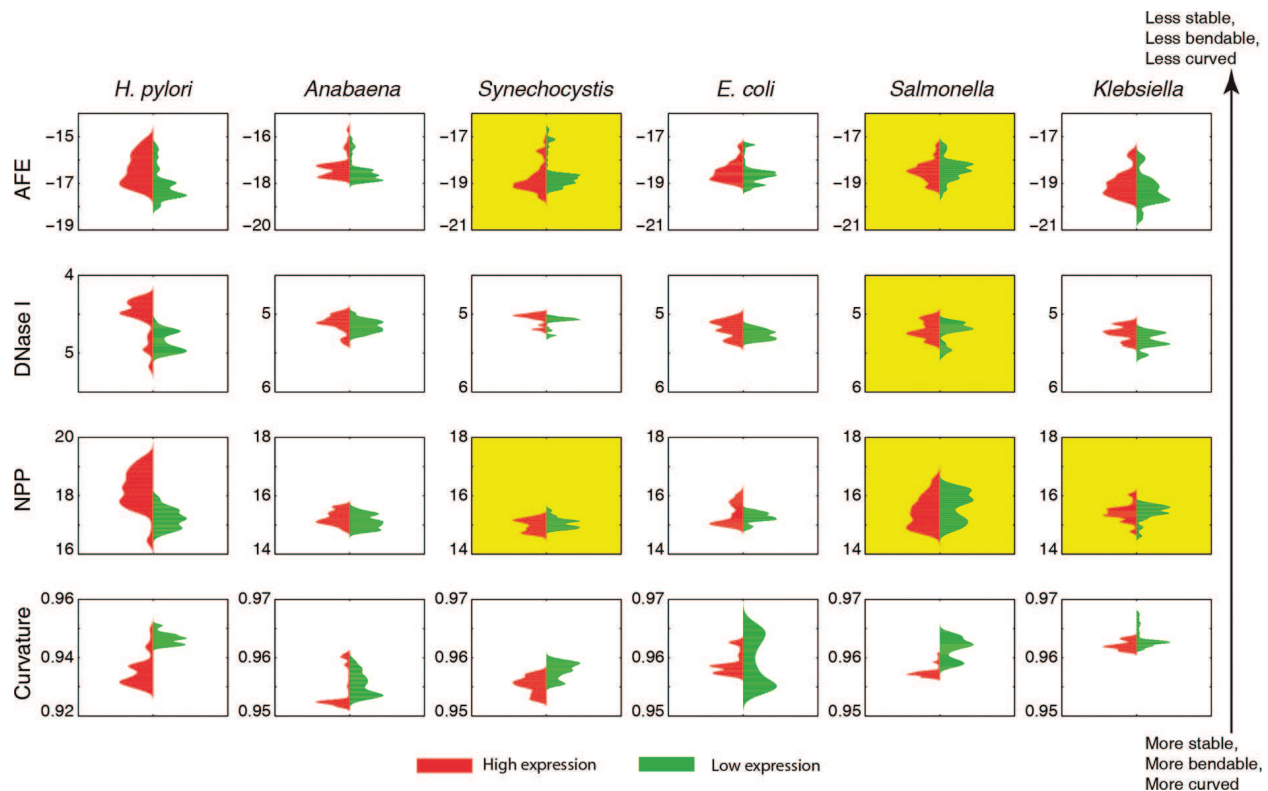
The activity of RNAP and RNAPII in prokaryotes and eukaryotes respectively is tightly regulated to ensure proper level of gene expression. Transcription factors (TFs), proteins that bind to specific regulatory sequences (*cis*-regulatory elements or CRE) are the key regulators of transcription [101]. The complex gene regulation in eukaryotes is a consequence of the large number of transcription factors available and localization of *cis*-regulatory elements.

### 4.1. Gene expression noise and its regulation

A variation in the copy number of mRNA or protein molecules for a given gene in cell is referred as gene expression noise. It is largely under the control of regulatory DNA since it is linked with the promoter structure. TATA box with variable strength, transcription factor binding sites count, strength and their position in the promoter and nucleosome binding sites in the regulatory region have enormous effect on gene expression noise in eukaryotes [102]. Though transcriptional regulation is quite well understood at molecular level, very little is known about gene expression noise in the case of prokaryotes. Transcription factors and inducer molecules play a major role in gene regulation process. Additionally, genome condensation assisted by nucleoid associated proteins and DNA supercoiling also play a vital role in gene regulation in bacteria. Gene expression noise is essential for achieving phenotypic heterogeneity and it has been found to be universal in nature.

### 4.2. DNA structural properties and their role in gene expression

Nucleosome organization in the genome has been found to be closely associated with the gene expression and its variability [82, 84, 85]. Genes with dissimilar expression levels tend to have sequences with different structural features in order to attain the required nucleosome organization [103–105]. Plasticity of gene expression, also known as gene expression variability is crucial for cell survival, is closely linked with the DNA structural properties of promoter region in *S. cerevisiae*. Promoters of genes with low plasticity (less responsive) are less stable, less bendable and lower nucleosome occupancy compared to the promoters of genes with high plasticity (high responsive) [106, 107]. A recent study in six different prokaryotes with variable genomic GC content (ranging from 39–58%) shows good correlation between DNA structural properties of promoter regions and gene expression. It has been found that promoter regions associated with high gene expression are less stable, less bendable and more curved as compare to the genes associated with low gene expression as seen from **Figure 5**. Intrinsic curvature was found to be most significant property which is distinctly present in the promoter regions associated with high gene expression as compared to those with low



**Figure 5.** Violin plot of four DNA structural property values in the promoter regions (-100 to 0 nucleotide with respect to TSS at 0) associated with high and low gene expression in six different prokaryotes. The x-axis shows the probability density while y-axis represents the DNA structural feature value. Plots with yellow background indicate the cases which failed to reject the null hypothesis using two sample KS test at the level of significance of  $P = 0.01$  (Figure taken from Kumar & Bansal, [97]).

gene expression across all organisms [97]. Hence estimation and characterization of DNA structural features of promoter regions could be very informative in analyzing the expression of associated gene.

## 5. Conclusions

The growing plethora of genomic information in the form of whole genome sequences requires its annotation to make sense of it. Mere delineation of coding sequences (gene identification) is not enough to get complete understanding of functional genomics since regulation of gene expression orchestrates the fate of cells. Gene expression regulation depends on different regulatory elements localized in the noncoding and coding region of the genome. Identification and characterization of these regulatory elements is the next level of challenge in the genome annotation process. Studies on DNA structural features of the regulatory regions show quite promising results toward achieving this goal. Moreover, DNA structural properties based characterization of regulatory regions is more sensitive and precise as compared to sequence-based approaches and most importantly it is universal in nature, applicable to all domains of life. Accumulating evidence shows a close relationship between gene expression and structural properties of promoter DNA; furthermore, this information can be used to engineer the regulatory sequences to modulate gene expression.

## Acknowledgements

MB is Senior Scientist, Indian National Science Academy (INSA) and the recipient of J. C. Bose National Fellowship of Department of Science and Technology (DST), India.

## Conflict of interest

None declared.

## Author details

Aditya Kumar<sup>1</sup> and Manju Bansal<sup>2\*</sup>

\*Address all correspondence to: mb@iisc.ac.in

1 Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam, India

2 Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, India

## References

- [1] Huang Y, Chen SY, Deng F. Well-characterized sequence features of eukaryote genomes and implications for ab initio gene prediction. *Computational and Structural Biotechnology Journal*. 2016;**14**:298-303. DOI: 10.1016/j.csbj.2016.07.002
- [2] Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences*. 2012;**69**(21):3613-3634. DOI: 10.1007/s00018-012-0990-9
- [3] Smith L, Brannan RA, Hanby AM, Shaaban AM, Verghese ET, Peter MB, Pollock S, Satheesha S, Szyrkiewicz M, Speirs V, et al. Differential regulation of oestrogen receptor beta isoforms by 5' untranslated regions in cancer. *Journal of Cellular and Molecular Medicine*. 2010;**14**(8):2172-2184. DOI: 10.1111/j.1582-4934.2009.00867.x
- [4] Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;**106**(18):7507-7512. DOI: 10.1073/pnas.0810916106
- [5] Lepppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews. Molecular Cell Biology*. 2018;**19**(3):158-174. DOI: 10.1038/nrm.2017.103

- [6] Srivastava AK, Lu Y, Zinta G, Lang Z, Zhu JK. UTR-dependent control of gene expression in plants. *Trends in Plant Science*. 2018;**23**(3):248-259. DOI: 10.1016/j.tplants.2017.11.003
- [7] Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Research*. 2012;**40**(11):4727-4741. DOI: 10.1093/nar/gks068
- [8] Serikawa T, Spanos C, von Hacht A, Budisa N, Rappsilber J, Kurreck J. Comprehensive identification of proteins binding to RNA G-quadruplex motifs in the 5' UTR of tumor-associated mRNAs. *Biochimie*. 2018;**144**:169-184. DOI: 10.1016/j.biochi.2017.11.003
- [9] Cammas A, Dubrac A, Morel B, Lamaa A, Touriol C, Teulade-Fichou MP, Prats H, Millevoi S. Stabilization of the G-quadruplex at the VEGF IRES represses cap-independent translation. *RNA Biology*. 2015;**12**(3):320-329. DOI: 10.1080/15476286.2015.1017236
- [10] Pickering BM, Willis AE. The implications of structured 5' untranslated regions on translation and disease. *Seminars in Cell & Developmental Biology*. 2005;**16**(1):39-47. DOI: 10.1016/j.semcdb.2004.11.006
- [11] Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN. Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Molecular and Cellular Biology*. 2007;**27**(13):4685-4697. DOI: 10.1128/MCB.02138-06
- [12] Kapp LD, Lorsch JR. The molecular mechanics of eukaryotic translation. *Annual Review of Biochemistry*. 2004;**73**:657-704. DOI: 10.1146/annurev.biochem.73.030403.080419
- [13] Jackowiak P, Hojka-Osinska A, Gasiorek K, Stelmaszczuk M, Gudanis D, Gdaniec Z, Figlerowicz M. Effects of G-quadruplex topology on translational inhibition by tRNA fragments in mammalian and plant systems in vitro. *The International Journal of Biochemistry & Cell Biology*. 2017;**92**:148-154. DOI: 10.1016/j.biocel.2017.10.001
- [14] Yamamoto H, Unbehaun A, Spahn CMT. Ribosomal chamber music: Toward an understanding of IRES mechanisms. *Trends in Biochemical Sciences*. 2017;**42**(8):655-668. DOI: 10.1016/j.tibs.2017.06.002
- [15] Johnson AG, Grosely R, Petrov AN, Puglisi JD. Dynamics of IRES-mediated translation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2017;**372**(1716). DOI: 10.1098/rstb.2016.0177
- [16] Terenin IM, Smirnova VV, Andreev DE, Dmitriev SE, Shatsky IN. A researcher's guide to the galaxy of IRESs. *Cellular and Molecular Life Sciences*. 2017;**74**(8):1431-1455. DOI: 10.1007/s00018-016-2409-5
- [17] Komar AA, Hatzoglou M. Internal ribosome entry sites in cellular mRNAs: Mystery of their existence. *The Journal of Biological Chemistry*. 2005;**280**(25):23425-23428. DOI: 10.1074/jbc.R400041200
- [18] Nevins TA, Harder ZM, Korneluk RG, Holcik M. Distinct regulation of internal ribosome entry site-mediated translation following cellular stress is mediated by apoptotic

- fragments of eIF4G translation initiation factor family members eIF4GI and p97/DAP5/NAT1. *The Journal of Biological Chemistry*. 2003;**278**(6):3572-3579. DOI: 10.1074/jbc.M206781200
- [19] Sajjanar B, Deb R, Raina SK, Pawar S, Brahmane MP, Nirmale AV, Kurade NP, Manjunathareddy GB, Bal SK, Singh NP. Untranslated regions (UTRs) orchestrate translation reprogramming in cellular stress responses. *Journal of Thermal Biology*. 2017;**65**:69-75. DOI: 10.1016/j.jtherbio.2017.02.006
- [20] Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: Molecular switches in (patho)physiology. *BioEssays*. 2010;**32**(10):885-893. DOI: 10.1002/bies.201000037
- [21] Barbosa C, Peixeiro I, Romao L. Gene expression regulation by upstream open reading frames and human disease. *PLoS Genetics*. 2013;**9**(8):e1003529. DOI: 10.1371/journal.pgen.1003529
- [22] Wethmar K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdisciplinary Reviews. RNA*. 2014;**5**(6):765-778. DOI: 10.1002/wrna.1245
- [23] Dar D, Sorek R. Regulation of antibiotic-resistance by non-coding RNAs in bacteria. *Current Opinion in Microbiology*. 2017;**36**:111-117. DOI: 10.1016/j.mib.2017.02.005
- [24] Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, Wang S, Dong X. uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature*. 2017;**545**(7655):491-494. DOI: 10.1038/nature22372
- [25] Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*. 2008;**320**(5883):1643-1647. DOI: 10.1126/science.1155390
- [26] Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biology*. 2012;**9**(5):563-576. DOI: 10.4161/rna.20231
- [27] Dickson AM, Wilusz J. Polyadenylation: Alternative lifestyles of the A-rich (and famous?). *The EMBO Journal*. 2010;**29**(9):1473-1474. DOI: 10.1038/emboj.2010.67
- [28] Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O. Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biology*. 2008;**6**(4):e92. DOI: 10.1371/journal.pbio.0060092
- [29] Chen CY, Shyu AB. AU-rich elements: Characterization and importance in mRNA degradation. *Trends in Biochemical Sciences*. 1995;**20**(11):465-470
- [30] von Roretz C, Di Marco S, Mazroui R, Gallouzi IE. Turnover of AU-rich-containing mRNAs during stress: A matter of survival. *Wiley Interdisciplinary Reviews. RNA*. 2011;**2**(3):336-347. DOI: 10.1002/wrna.55
- [31] Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'xxUTR evolution. *Cell*. 2005;**123**(6):1133-1146. DOI: 10.1016/j.cell.2005.11.023



- [32] Gorgoni B, Richardson WA, Burgess HM, Anderson RC, Wilkie GS, Gautier P, Martins JP, Brook M, Sheets MD, Gray NK. Poly(A)-binding proteins are functionally distinct and have essential roles during vertebrate development. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;**108**(19):7844-7849. DOI: 10.1073/pnas.1017664108
- [33] Wilkie GS, Gautier P, Lawson D, Gray NK. Embryonic poly(A)-binding protein stimulates translation in germ cells. *Molecular and Cellular Biology*. 2005;**25**(5):2060-2071. DOI: 10.1128/MCB.25.5.2060-2071.2005
- [34] Smith RW, Blee TK, Gray NK. Poly(A)-binding proteins are required for diverse biological processes in metazoans. *Biochemical Society Transactions*. 2014;**42**(4):1229-1237. DOI: 10.1042/BST20140111
- [35] Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F. Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay. *Molecular and Cellular Biology*. 2013;**33**(23):4718-4731. DOI: 10.1128/MCB.00887-13
- [36] Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. *Cold Spring Harbor Perspectives in Biology*. 2014;**14**(6). DOI: 10.1101/cshperspect.a016071
- [37] Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. *Biology Direct*. 2012;**7**:11. DOI: 10.1186/1745-6150-7-11
- [38] Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. *Trends in Genetics*. 2005;**21**(2):111-119. DOI: 10.1016/j.tig.2004.12.007
- [39] Cech TR. Self-splicing of group I introns. *Annual Review of Biochemistry*. 1990;**59**(1): 543-568. DOI: 10.1146/annurev.bi.59.070190.002551
- [40] Nielsen H, Johansen SD. Group I introns: Moving in new directions. *RNA Biology*. 2009;**6**(4):375-383
- [41] Hedberg A, Johansen SD. Nuclear group I introns in self-splicing and beyond. *Mobile DNA*. 2013;**4**(1):17. DOI: 10.1186/1759-8753-4-17
- [42] Raghavan R, Minnick MF. Group I introns and inteins: Disparate origins but convergent parasitic strategies. *Journal of Bacteriology*. 2009;**191**(20):6193-6202. DOI: 10.1128/JB.00675-09
- [43] Pyle AM. Group II intron self-splicing. *Annual Review of Biophysics*. 2016;**45**:183-205. DOI: 10.1146/annurev-biophys-062215-011149
- [44] Zhao C, Pyle AM. Structural insights into the mechanism of group II intron splicing. *Trends in Biochemical Sciences*. 2017;**42**(6):470-482. DOI: 10.1016/j.tibs.2017.03.007
- [45] Zhao C, Pyle AM. The group II intron maturase: A reverse transcriptase and splicing factor go hand in hand. *Current Opinion in Structural Biology*. 2017;**47**:30-39. DOI: 10.1016/j.sbi.2017.05.002
- [46] Novikova O, Belfort M. Mobile group II introns as ancestral eukaryotic elements. *Trends in Genetics*. 2017;**33**(11):773-783. DOI: 10.1016/j.tig.2017.07.009

- [47] Fujishima K, Kanai A. tRNA gene diversity in the three domains of life. *Frontiers in Genetics*. 2014;**5**:142. DOI: 10.3389/fgene.2014.00142
- [48] Yoshihisa T. Handling tRNA introns, archaeal way and eukaryotic way. *Frontiers in Genetics*. 2014;**5**:213. DOI: 10.3389/fgene.2014.00213
- [49] Lopes RR, Kessler AC, Polycarpo C, Alfonzo JD. Cutting, dicing, healing and sealing: The molecular surgery of tRNA. *Wiley Interdisciplinary Reviews. RNA*. 2015;**6**(3):337-349. DOI: 10.1002/wrna.1279
- [50] Jo BS, Choi SS. Introns: The functional benefits of introns in genomes. *Genomics & Informatics*. 2015;**13**(4):112-118. DOI: 10.5808/GI.2015.13.4.112
- [51] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 2008;**40**(12):1413-1415. DOI: 10.1038/ng.259
- [52] Mastrangelo AM, Marone D, Laido G, De Leonardis AM, De Vita P. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Science*. 2012;**185-186**:40-49. DOI: 10.1016/j.plantsci.2011.09.006
- [53] Barbazuk WB, Fu Y, McGinnis KM. Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. *Genome Research*. 2008;**18**(9):1381-1392. DOI: 10.1101/gr.053678.106
- [54] Zhang C, Yang H, Yang H. Evolutionary character of alternative splicing in plants. *Bioinformatics and Biology Insights*. 2015;**9**(Suppl 1):47-52. DOI: 10.4137/BBI.S33716
- [55] Shaul O. How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology*. 2017;**91**(Pt B):145-155. DOI: 10.1016/j.biocel.2017.06.016
- [56] Rose AB. Intron-mediated regulation of gene expression. *Current Topics in Microbiology and Immunology*. 2008;**326**:277-290
- [57] Gallegos JE, Rose AB. The enduring mystery of intron-mediated enhancement. *Plant Science*. 2015;**237**:8-15. DOI: 10.1016/j.plantsci.2015.04.017
- [58] Brown SJ, Stoilov P, Xing Y. Chromatin and epigenetic regulation of pre-mRNA processing. *Human Molecular Genetics*. 2012;**21**(R1):R90-R96. DOI: 10.1093/hmg/dds353
- [59] Valencia P, Dias AP, Reed R. Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;**105**(9):3386-3391. DOI: 10.1073/pnas.0800250105
- [60] Carrillo Oesterreich F, Bieberstein N, Neugebauer KM. Pause locally, splice globally. *Trends in Cell Biology*. 2011;**21**(6):328-335. DOI: 10.1016/j.tcb.2011.03.002
- [61] Schwartz S, Meshorer E, Ast G. Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*. 2009;**16**(9):990-995. DOI: 10.1038/nsmb.1659
- [62] Bradnam KR, Korf I. Longer first introns are a general property of eukaryotic gene structure. *PLoS One*. 2008;**3**(8):e3093. DOI: 10.1371/journal.pone.0003093

- [63] Carmel L, Rogozin IB, Wolf YI, Koonin EV. Evolutionarily conserved genes preferentially accumulate introns. *Genome Research*. 2007;**17**(7):1045-1050. DOI: 10.1101/gr.5978207
- [64] Park SG, Hannenhalli S, Choi SS. Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics*. 2014;**15**:526. DOI: 10.1186/1471-2164-15-526
- [65] Zhang Q, Li H, Zhao XQ, Xue H, Zheng Y, Meng H, Jia Y, Bo SL. The evolution mechanism of intron length. *Genomics*. 2016;**108**(2):47-55. DOI: 10.1016/j.ygeno.2016.07.004
- [66] Decker KB, Hinton DM. Transcription regulation at the core: Similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annual Review of Microbiology*. 2013;**67**:113-139. DOI: 10.1146/annurev-micro-092412-155756
- [67] Browning DF, Busby SJ. The regulation of bacterial transcription initiation. *Nature Reviews. Microbiology*. 2004;**2**(1):57-65. DOI: 10.1038/nrmicro787
- [68] Kadonaga JT. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology*. 2012;**1**(1):40-51. DOI: 10.1002/wdev.21
- [69] Gupta K, Sari-Ak D, Haffke M, Trowitzsch S, Berger I. Zooming in on transcription Preinitiation. *Journal of Molecular Biology*. 2016;**428**(12):2581-2591. DOI: 10.1016/j.jmb.2016.04.003
- [70] Roy AL, Singer DS. Core promoters in transcription: Old problem, new insights. *Trends in Biochemical Sciences*. 2015;**40**(3):165-171. DOI: 10.1016/j.tibs.2015.01.007
- [71] Kadonaga JT. The DPE, a core promoter element for transcription by RNA polymerase II. *Experimental & Molecular Medicine*. 2002;**34**(4):259-264. DOI: 10.1038/emm.2002.36
- [72] Vilar JM, Saiz L. DNA looping in gene regulation: From the assembly of macromolecular complexes to the control of transcriptional noise. *Current Opinion in Genetics & Development*. 2005;**15**(2):136-144. DOI: 10.1016/j.gde.2005.02.005
- [73] Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes & Development*. 2011;**25**(10):1010-1022. DOI: 10.1101/gad.2037511
- [74] Kim TK, Shiekhatter R. Architectural and functional commonalities between enhancers and promoters. *Cell*. 2015;**162**(5):948-959. DOI: 10.1016/j.cell.2015.08.008
- [75] Krivega I, Dean A. Enhancer and promoter interactions-long distance calls. *Current Opinion in Genetics & Development*. 2012;**22**(2):79-85. DOI: 10.1016/j.gde.2011.11.001
- [76] Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nature Reviews. Genetics*. 2013;**14**(4):288-295. DOI: 10.1038/nrg3458
- [77] Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*. 2006;**7**:29-59. DOI: 10.1146/annurev.genom.7.080505.115623
- [78] Ali T, Renkawitz R, Bartkuhn M. Insulators and domains of gene expression. *Current Opinion in Genetics & Development*. 2016;**37**:17-26. DOI: 10.1016/j.gde.2015.11.009

- [79] Iyer VR. Nucleosome positioning: Bringing order to the eukaryotic genome. *Trends in Cell Biology*. 2012;**22**(5):250-256. DOI: 10.1016/j.tcb.2012.02.004
- [80] Ballare C, Zaurin R, Vicent GP, Beato M. More help than hindrance: Nucleosomes aid transcriptional regulation. *Nucleus*. 2013;**4**(3):189-194. DOI: 10.4161/nucl.25108
- [81] Rando OJ, Ahmad K. Rules and regulation in the primary structure of chromatin. *Current Opinion in Cell Biology*. 2007;**19**(3):250-256. DOI: 10.1016/j.ceb.2007.04.006
- [82] Jiang C, Pugh BF. Nucleosome positioning and gene regulation: Advances through genomics. *Nature Reviews. Genetics*. 2009;**10**(3):161-172. DOI: 10.1038/nrg2522
- [83] Struhl K, Segal E. Determinants of nucleosome positioning. *Nature Structural & Molecular Biology*. 2013;**20**(3):267-273. DOI: 10.1038/nsmb.2506
- [84] Teif VB. Nucleosome positioning: Resources and tools online. *Briefings in Bioinformatics*. 2016;**17**(5):745-757. DOI: 10.1093/bib/bbv086
- [85] Lai WKM, Pugh BF. Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature Reviews. Molecular Cell Biology*. 2017;**18**(9):548-562. DOI: 10.1038/nrm.2017.47
- [86] Liu MJ, Seddon AE, Tsai ZT, Major IT, Floer M, Howe GA, Shiu SH. Determinants of nucleosome positioning and their influence on plant gene expression. *Genome Research*. 2015;**25**(8):1182-1195. DOI: 10.1101/gr.188680.114
- [87] Shukla A, Bhargava P. Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. *Biochimica et Biophysica Acta*. 2017. DOI: 10.1016/j.bbagr.2017.11.008
- [88] Harteis S, Schneider S. Making the bend: DNA tertiary structure and protein-DNA interactions. *International Journal of Molecular Sciences*. 2014;**15**(7):12335-12363. DOI: 10.3390/ijms150712335
- [89] Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry*. 2010;**79**:233-269. DOI: 10.1146/annurev-biochem-060408-091030
- [90] Meysman P, Marchal K, Engelen K. DNA structural properties in the classification of genomic transcription regulation elements. *Bioinformatics and Biology Insights*. 2012;**6**:155-168. DOI: 10.4137/BBI.S9426
- [91] Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, et al. A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology*. 2001;**313**(1):229-237. DOI: 10.1006/jmbi.2001.4987
- [92] Yella VR, Kumar A, Bansal M. DNA structure and promoter engineering. In: Singh V, Dhar PK, editors. *Systems and Synthetic Biology*. Dordrecht: Springer Netherlands; 2015. pp. 241-254
- [93] Kanhere A, Bansal M. Structural properties of promoters: Similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Research*. 2005;**33**(10):3165-3175. DOI: 10.1093/nar/gki627



- [94] Pedersen AG, Jensen LJ, Brunak S, Staerfeldt HH, Ussery DW. A DNA structural atlas for *Escherichia coli*. *Journal of Molecular Biology*. 2000;**299**(4):907-930. DOI: 10.1006/jmbi.2000.3787
- [95] Rangannan V, Bansal M. High-quality annotation of promoter regions for 913 bacterial genomes. *Bioinformatics*. 2010;**26**(24):3043-3050. DOI: 10.1093/bioinformatics/btq577
- [96] Bansal M, Kumar A, Yella VR. Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. *Current Opinion in Structural Biology*. 2014;**25**:77-85. DOI: 10.1016/j.sbi.2014.01.007
- [97] Kumar A, Bansal M. Unveiling DNA structural features of promoters associated with various types of TSSs in prokaryotic transcriptomes and their role in gene expression. *DNA Research*. 2017;**24**(1):25-35. DOI: 10.1093/dnares/dsw045
- [98] Kumar A, Bansal M. Characterization of structural and free energy properties of promoters associated with primary and operon TSS in *Helicobacter pylori* genome and their orthologs. *Journal of Biosciences*. 2012;**37**(3):423-431
- [99] Kumar A, Manivelan V, Bansal M. Structural features of DNA are conserved in the promoter region of orthologous genes across different strains of *Helicobacter pylori*. *FEMS Microbiology Letters*. 2016;**363**(18). DOI: 10.1093/femsle/fnw207
- [100] Morey C, Mookherjee S, Rajasekaran G, Bansal M. DNA free energy-based promoter prediction and comparative analysis of Arabidopsis and rice genomes. *Plant Physiology*. 2011;**156**(3):1300-1315. DOI: 10.1104/pp.110.167809
- [101] Wittkopp PJ, Kalay G. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews. Genetics*. 2011;**13**(1):59-69. DOI: 10.1038/nrg3095
- [102] Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. *Annual Review of Biophysics*. 2013;**42**:469-491. DOI: 10.1146/annurev-biophys-083012-130401
- [103] Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research*. 2008;**18**(7):1073-1083. DOI: 10.1101/gr.078261.108
- [104] Choi JK, Kim YJ. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nature Genetics*. 2009;**41**(4):498-503. DOI: 10.1038/ng.319
- [105] Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biology*. 2010;**8**(7):e1000414. DOI: 10.1371/journal.pbio.1000414
- [106] Yella VR, Bansal M. DNA structural features and architecture of promoter regions play a role in gene responsiveness of *S. cerevisiae*. *Journal of Bioinformatics and Computational Biology*. 2013;**11**(6):1343001. DOI: 10.1142/S0219720013430014
- [107] Yella VR, Kumar A, Bansal M. Identification of putative promoters in 48 eukaryotic genomes on the basis of DNA free energy. *Scientific Reports*. 2018;**8**(1):4520. DOI: 10.1038/s41598-018-22129-8



