

Role of epigenetics in expression of recombinant proteins from mammalian cells

Optimizing productivity from recombinant Chinese hamster ovary cells requires understanding of the transcriptional and translational regulatory processes. Several 'omics approaches have been utilized to characterize these processes. Still, many questions remain unanswered as to the mechanisms underlying transgene expression. As a result, bioprocess development still remains highly variable, and the cell selection process must be repeated for every new product synthesized. It has become clear that epigenetic processes play a role in recombinant protein expression and that gaining a better understanding of these processes will aid in process development and cell-line selection. In this perspective, we highlight our current knowledge of the influence of various epigenetic factors that control recombinant protein expression.

The early biopharmaceutical industry relied on *Escherichia coli* and yeast for recombinant protein expression, as these hosts require short culture duration, permit facile genetic manipulation and can be cultured in inexpensive media [1]. Biomolecules like insulin and growth hormones were successfully manufactured using this model. In recent years, the product catalog of biopharmaceuticals has evolved significantly [2]. Better understanding of disease pathways has led to the discovery and development of many drug targets. Today's biotherapeutics include monoclonal antibodies (primarily chimeric, humanized and human), Fc-fusion proteins, hormones, cytokines and blood products. Production of these sophisticated biomolecules requires the post-transcriptional and metabolic machinery only available in mammalian cells. Due to their ability to adapt to culture conditions and ease of maintenance, Chinese hamster ovary (CHO) cells have emerged as the workhorse of the biopharmaceutical industry, with nearly 70% of all therapeutic proteins expressed in the CHO cells [3]. Recently, CHO cells have been investigated for the production of nonprotein pharmaceuticals such as hepa-

rin, a glycosaminoglycan [4]. More than two decades of bioprocess research and development has led to vast improvements in the final protein yields. There have been great advances in cloning techniques, vector design, transfection technologies and selection strategies. Automation has allowed for dramatic increases in cell-selection throughput. Bioprocess optimizations such as media design, rational feed strategies, improvements in process conditions and new bioreactor technologies have all contributed to the steady rise in volumetric productivity. Currently, the highest reported yield from mammalian cell cultures is >10 g/l [5]. While these titers are quite impressive, they are still approximately fivefold lower than can be obtained from yeast or bacteria. Thus, there is demand for improving yield while maintaining product quality attributes for recombinant protein pharmaceuticals.

The dramatic advances in our understanding of **chromatin** structure and epigenetic mechanisms have afforded new opportunities for improvements in protein productivity [6]. In eukaryotic cells, the genome is packaged into chromatin by wrapping the negatively charged DNA around positively charged histones [7,8].

Hussain Dahodwala¹ & Susan T Sharfstein^{*1}

¹SUNY Polytechnic Institute, Colleges of Nanoscale Science & Engineering, NFE, 257 Fuller Road, Albany, NY 12203, USA

*Author for correspondence:

Tel.: +1 518 4378820

ssharfstein@sunycnse.com

FUTURE
SCIENCE

part of

fsg

Key terms

Chromatin: Chromatin is a complex of DNA and proteins that forms chromosomes within the nucleus of eukaryotic cells. The state of the chromatin influences gene expression.

Epigenetics: The study of changes in gene expression caused by chemical modifications to the DNA and the gene-associated proteins. These modifications are reversible but may also be heritable.

Recombineering: Recombineering is the engineering of bacterial artificial chromosome clones by homologous recombination. It is performed inside *Escherichia coli* cells that have been engineered to permit homologous recombination.

Within each cell, nearly two meters of DNA is packed into a volume of 100 μm^3 . It is widely believed this compacting is not a random event [9], but rather a specific, higher order organization is used to regulate the transcription of genes [8].

Epigenetic mechanisms & their influence on gene expression

Gene expression is a complex process involving numerous steps with many levels of regulation as shown in Figure 1 [10]. Transcription, translation and subsequent protein modification are the steps by which a cell transfers genetic information from an endogenous gene or transgene, resulting in the production of proteins. The term ‘epigenetics’, which literally means ‘outside conventional genetics’, is now used to describe the study of stable alterations in gene expression. The US NIH in their recent epigenomics initiative [11] stated that ‘epigenetics refers to changes in gene activity and expression and also stable, long-term, alterations in the transcriptional potential of a cell that are not necessarily heritable’. The main epigenetic processes that influence transgene expression patterns are transgene localization, cytosine methylation of the DNA, post-translational modification (PTM) of histones, modification of transcriptional proteins, resulting in altered DNA–protein interactions, and RNA-based mechanisms.

Transgene localization

A critical determinant of foreign gene expression is the state of the surrounding chromatin, as determined by histone modifications and promoter methylation [12]. The repressed state of the chromatin is reversible, but this status is transmitted to daughter cells. Strategies to overcome the positional effects of the adjacent chromatin have been reviewed extensively [6]. In particular, cis-acting locus control

regions have been employed to open the chromatin and enhance transcription factor activity. Other strategies include employing STabilizing AntiRepressor elements and other ‘antirepressor’ elements flanking the vectors, which counteract chromatin-associated repression effects [13]. The use of matrix attachment regions (MARs) elements in vectors has also been investigated. MARs help generate and maintain an open chromatin domain that is favorable to transcription [14].

Chromosomal vectors

Yeast & bacterial artificial chromosomes

Another strategy to alleviate the positional effects on expression is the use of vectors that contain all the elements required to create their own chromatin environment. One such strategy is the employment of vectors with large cloning capacities that can accommodate an entire mammalian genetic locus [15]. These vectors employ predefined open chromatin loci with their own chromatin environment. Therefore, these vectors are not affected by the surrounding chromatin of their integration site. The most popular vectors with large cloning capacity are yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs). YACs can accommodate up to 2 Mb of DNA, making them interesting candidates for expression vectors. However, manipulation of YACs is very labor-intensive and time consuming [16]. BACs, circular plasmids maintained in *E. coli*, are derived from the F-factor (a bacterial episome with its own origin of replication that can integrate into the bacterial chromosome) and are able to accommodate up to 350 kb of DNA. BACs are easy to manipulate, and yields of purified BAC DNA are reasonable and can be transfected into mammalian cells using conventional methods [17,18]. BACs can accommodate almost all of the elements that help overcome positioning effects and augment expression of a gene of interest [19]. Thus, BACs can be considered as complete expression units. There has been some success in protein expression using BACs [19–21], and with improvements in recombineering [22,23], we can expect to see more examples using BACs as expression vectors applied to recombinant protein production in mammalian cells.

Mammalian artificial chromosomes

Several mammalian artificial chromosome systems have been reported, with the most notable being the satellite DNA-based artificial chromosome (ACE) marketed by Chromos Molecular Systems [24]. The advantage of artificial chromosomes is that they can carry all the elements to modulate the epigenetic environment. Such a vector would be independent of site of

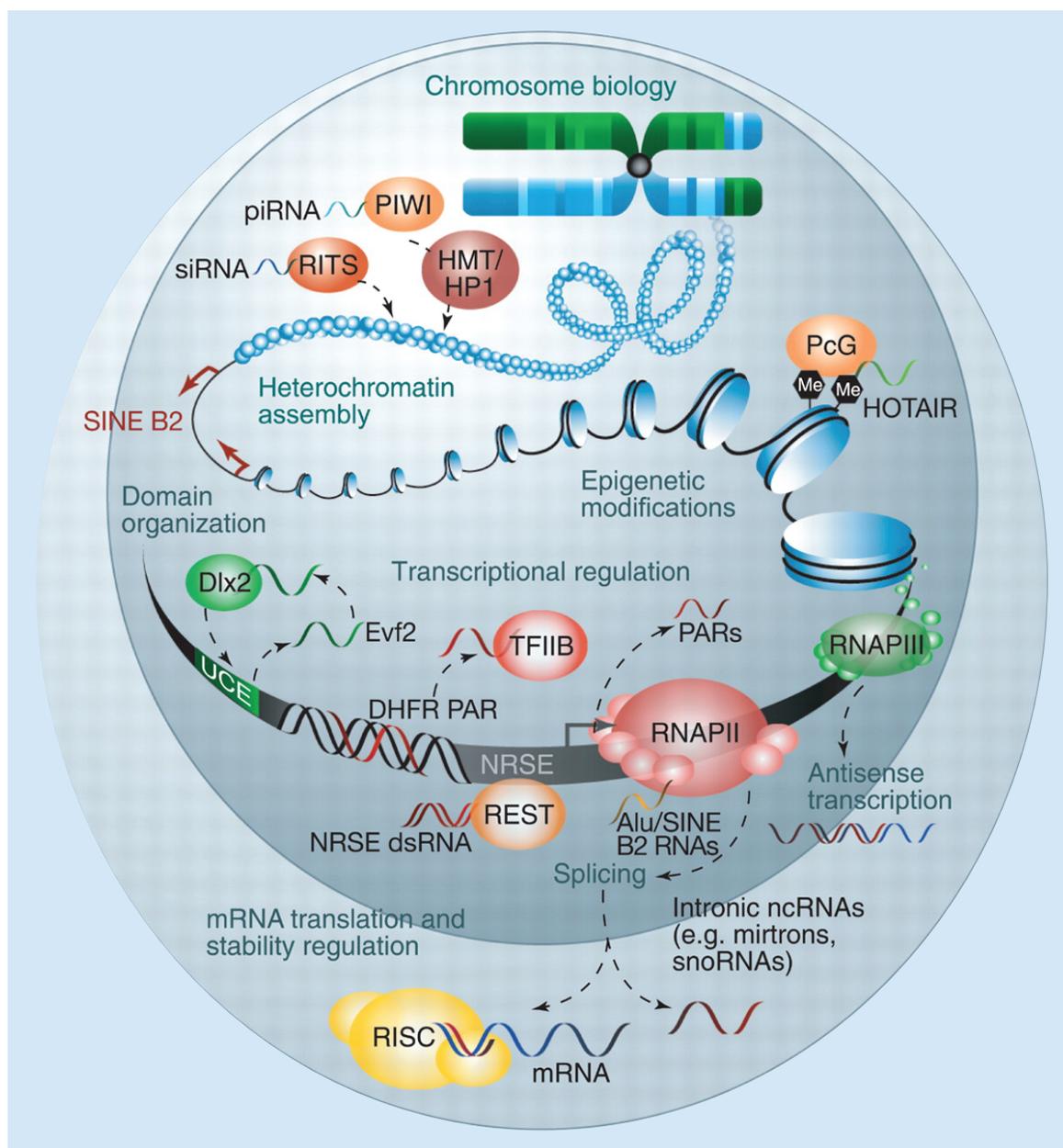


Figure 1. Overview of eukaryotic gene expression control elements. Regulation of eukaryotic gene expression by (A) chromatin organization, (B) heterochromatin assembly by histone protein interactions, (C) DNA methylation, (D) transcriptionally active ultra-conserved elements (UCEs), (E) transcriptional factors and regulators, (F) influence of RNA pol complexes, (G) mRNA stability, (H) RISC complexes that degrade mRNA, (I) siRNA silencing, (J) ncRNA mediated silencing.

ncRNA: Noncoding RNA; RISC: RNA-induced silencing complex; RNAPII: RNA polymerase II; RNAPIII: RNA polymerase III; SINE: Short interspersed element.

Reproduced with permission from [10].

insertion and also have the added advantage of being nonviral [25]. Their propagation would be straightforward as they replicate autonomously alongside host chromosomes [26]. Many different cell lines have been successfully transfected using this technology, resulting in increased productivity and shorter selection times for higher producers [27]. Unfortunately, Chromos was

unable to create a successful commercial venture based on this technology; however, it still retains significant potential.

Hot spot targeting

Another strategy for improving the localization of the transgene consists of identifying so called 'hot spots'

in the host cell genome that allow reproducible, high expression of the transgene. With the availability of CHO genome, CHO-specific hot spots can be rationally investigated. Once an optimal hot spot has been identified, the expression vector can be 'knocked into' this predefined genomic region. Several recent tools for chromosome editing (described below) can be used for this targeted integration.

Zinc-finger nucleases

Zinc-finger nucleases (ZFNs) are artificial restriction enzymes that facilitate site-directed DNA insertion by inducing double-strand breaks at specific nucleotides and then allowing for insertion of transgenes by nonhomologous end joining [28]. It is a well characterized technique and gene-specific applications can be obtained from Sigma-Aldrich (MO, USA). Its applicability and robustness has been demonstrated in a variety of cell types and applications [29]. ZFNs have been employed to generate a more efficient platform cell line by targeted knockout of the *GS* gene from CHOK1 SV cell lines [30]. Multiple gene knockouts have been performed in a single cell line using ZFN as well [31]. Such approaches will be immensely important as the new 'omics data helps to identify redundant genes or genes implicated in epigenetic silencing.

TALENs & TALEs

A newer entry into the genome editing field, transcription activator-like effector nucleases (TALENs) have the potential for revolutionizing transgene engineering. TALENs are engineered nucleases that can be made with high specificity for any given DNA sequence [32]. This technology has made it possible to customize genome editing to any desired hot spot with high precision [33]. Targeting of the DNA via TALENs depends on matching the amino acid sequence of the TALEN to recognize the desired DNA sequence and therefore, is more modular than the context-dependent ZFNs [34].

TALEs (transcription activator-like effectors) are proteins secreted by *Xanthomonas* bacteria. These TALEs can bind promoter sequences in plant and activate the expression of genes that help aid the bacteria. Bioinformatic analysis of TALEs shows that the proteins contain a central repeat domain consisting of a variable number of approximately 34 amino acid repeats. There appears to be a one-to-one correspondence between the identity of two critical amino acids in each repeat and each DNA base in the target sequence. This allows the design of specific proteins that can target any given promoter. The engineered TALEs also have a transcription factor recruiting domain. Due to their ease of design, high efficiency of

genome editing and lower off-target effects [35], TALEs have been used to amplify endogenous gene expression, via binding to promoters and recruiting transcription factors to the site of interest, in a number of systems, such as yeast [36], plants [37,38] and various mammalian cells, including stem cells [39–42]. Recently, TALENs were engineered to overcome methylated cytosine targeting that was previously inaccessible [43]. Thus, TALENs have the potential to insert transgenes into any desired genomic loci, circumventing epigenetic mechanisms. The imminent use of TALE-directed gene activation in mammalian cell production systems will allow tuning of gene expression levels that will broadly enable synthetic biology and biotechnology.

The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system was discovered as a microbial nuclease system that is responsible for defense against invading phages and plasmids [44]. The CRISPR/Cas9 system is composed of two noncoding RNAs (ncRNAs), the pre-CRISPR RNA (crRNA) array and transactivating crRNA (tracrRNA) that are transcribed from the CRISPR locus. Upon maturation, the complex directs Cas9 to the target DNA where the endonuclease mediates cleavage of target DNA. The system works through RNA–DNA recognition and subsequent cleavage of target via the Cas9 endonuclease. It is possible to achieve precise deletion, insertion and sequence replacement with this mechanism. Decades of research into this method has led to a new type of genome editing tool [45]. In molecular biology applications, it has been established that a short guide RNA (sgRNA) made by fusing crRNA with tracrRNA can be made to function to a similar level as the crRNA and tracrRNA complex [46]. Therefore only two components, Cas9 and sgRNA, are necessary for the CRISPR/Cas system to function. Since small ncRNA mediate the recognition of target sequence and Cas9 cleavage of DNA, therefore, the system design is flexible, simpler and more economical compared with other constructs like TALENs.

In a study of pluripotent stem cells, CRISPR outperformed TALENs when introducing gene knockouts into clones [35]. sgRNAs expressed using plant RNA polymerase III promoters, such as U6 and U3, can be used to improve knock-in efficiency in human pluripotent stem cells. However, many more instances of off target effects occur with the CRISPR system when compared with TALENs. It is speculated that the availability of a Web-based resource, along with a searchable, genome-wide database [47,48] will help in improve the design of sgRNAs leading to better success.

CHO codon-optimized Cas9 has been utilized successfully for knockout of *FUT8* (alpha-

fucosyltransferase activity) and *COSMC* (C1GALT1-specific chaperone 1) genes in the CHO genome [49]. These genes are responsible for fucosylation of N-linked glycans and initiation of O-linked glycosylation, respectively [50,51]. By combining the CRISPR/Cas9 technology and the CRISPy bioinformatics tool, this system holds the potential to be a powerful technique in CHO cells.

Chromatin modification

DNA methylation

In recombinant CHO cells, gene expression is generally driven by the strong, viral CMV promoter, which is known to be negatively regulated by methylation [52]. **DNA methylation** causes loss of gene transcriptional activity. The methylation of mammalian genomic DNA is catalyzed by DNA methyltransferases (DNMTs). Methylation of genomic DNA recruits the methylated-DNA-binding proteins, which are known to recruit other chromatin remodeling factors and influence gene expression [53]. However, such modifications are reversible [54]. DNA demethylation has a prominent role in enhancing gene expression [55]. 5-Azacytidine is a well-studied methyltransferase inhibitor, known to activate gene expression [56–58]. Using **bisulfite sequencing**, methylation of specific nucleotides has been implicated in production instability, caused by transcriptional silencing of the cytomegalovirus (CMV) promoter [59]. Addition of 5-azacytidine inhibits DNMTs by binding to them and inhibiting their activity, thereby increasing transgene expression [52].

The CHO genome and transcriptome sequences were reported in 2011 [60,61], with the sequence of the Chinese hamster following in 2013 [62]. This data enables researchers to conduct genome-scale analyses investigating regulatory cellular mechanisms. However, the genomic data do not shed any light on the dynamic nature of epigenetic modifications, nor does it provide any information on the incorporation or epigenetic modification of transgenes. The CMV promoter has a large CpG island that can be differentially methylated under different culture conditions [63,64]. Promoter and genome-wide methylation changes according to culture condition, altering DNA expression [65]. To this end, a **CpG islands** (CGI) microarray has been developed to characterize the methylation status of DNA after butyrate treatment [66]. Such an array could be employed to aid in selection of clones that show inherent globally undermethylated DNA and to understand the global changes in methylation under different culture conditions and after treatment to modify epigenetic characteristics (e.g., butyrate and 5-azacytidine).

Key Terms

DNA methylation: DNA methylation is a biochemical process in which a methyl group is added to cytosine in genomic DNA by replacing H5 in cytosine with a methyl group. It is carried out by DNA methyltransferases.

Bisulfite sequencing: Bisulfite modification of DNA is used to study DNA methylation. Bisulfite treatment converts unmethylated cytosines to uracil but leaves 5-methylcytosine residues unaffected. The methylated and unmethylated cytosines can then be mapped by sequencing.

CpG islands: CpG islands (CGI) are short, interspersed DNA sequences that differ from the average genomic pattern by being CG-rich and subject to cytosine methylation.

Hydroxymethylation: Hydroxymethylation is a recently discovered mechanism of epigenetic control in which the H5 in cytosine is replaced by a hydroxymethyl group.

Inhibition of the CMV promoter leads to inactivation and downregulation of gene expression. By treatment with DNMT inhibitors it is possible to derepress CMV promoters [59]. These lessons have inspired the design and engineering of a synthetic promoter [67]. The synthetic promoter lacks CGI and is less susceptible to gene silencing. Furthermore, since CGI are known to interfere with DNA stabilizing elements, the lack of CpG in the synthetic promoters makes it more amenable to use with the ubiquitous chromatin opening elements and MARs elements.

DNA methylation is an important epigenetic mechanism in the NS0 cell lines, which are also widely used in the production of recombinant antibodies. Methylation of gene *Hsd17b7* is responsible for the cholesterol auxotrophy of NS0 cells [68]. By using DNMT inhibitors, *Hsd17b7* transcription was restored, generating high-frequency cholesterol-independent variants. The induction of cholesterol independence by altering DNA methylation pattern highlights the role of epigenetics in the metabolic adaptation of NS0 cells for bioprocessing applications. With the sequencing of the CHO genome, similar strategies may be employed to modulate a variety of biological processes in CHO cells as well.

Recently, cytosine **hydroxymethylation** was identified as another important epigenetic modification of DNA in mammalian cells [69]. Similar to methylation, hydroxymethylation is defined as the addition of a hydroxymethyl group at the C5-position in cytosine (5-hydroxymethylcytosine [5hmC]). Cytosine hydroxymethylation is also involved in gene regulation, with disruption of hydroxymethylation causing distorted cell function and leading to cancer in some systems [69–71]. However, 5hmC has not, to date, been identified in CHO cells. It will be interesting to characterize the existence and relevance of this novel

Key terms

Histone code: The histone code is a hypothesis that the transcription of genetic information encoded in DNA is regulated by chemical modifications to histone proteins.

Chromatin immunoprecipitation: Chromatin immunoprecipitation (ChIP) is a method for studying *in vivo* interactions between DNA and proteins. ChIP can be used to determine transcription factor binding to specific genomic sequences.

LncRNA: Long noncoding RNAs (lncRNAs) are RNA molecules >200 nucleotides that do not encode proteins. Their expression is developmentally regulated, and lncRNAs can be tissue and cell type specific.

modification and its role in transgene expression in mammalian cells.

Histone modifications

The basic unit of DNA packaging in eukaryotic cells consists of two units of the following histones: H2A, H2B, H3 and H4, forming an octamer. DNA is spooled around this core [72]. Various modifications to the histones can influence the DNA wrapping and thereby alter the gene expression. PTMs to histones influence recruitment of various factors. Together these modifications have an impact on DNA polymerase interactions with the DNA and also causes changes to the overall stability of the chromatin. Important histone modifications include lysine acetylation and lysine and arginine methylation. Other known histone modifications include phosphorylation, ubiquitination and sumoylation [73–75]. Many of these modifications impact transcription [76], recombination [77], DNA repair and replication [78,79] and chromosomal organization [74,80].

Many different combinatorial patterns of histones can exist. For example, Histone H3 contains 19 lysines. Each of the 19 lysines can be methylated to various degrees. The overall state of the histones at the transgenic promoter is dictated by the **histone code** in the flanking chromatin region [81]. Histone structure can be manipulated using histone deacetylase inhibitors such as sodium butyrate and sodium valproate. Sodium butyrate can enhance gene expression [82], but not uniformly. However, the cost, apoptosis induction and need for removal have discouraged butyrate use in large bioreactors. Interestingly, DNA elements like MARs, ubiquitous chromatin opening elements and other insulators promote the deposition of active histone marks (e.g., acetylation) and prevent methylation events on the histones along the transgene [83]. Therefore, the histone architecture can be influenced by engineering of the DNA elements used in transfection, showing crosstalk

between the transgene insertion mechanisms, site of integration and chromatin state.

The field of chromatin organization is still developing, and newer histone variants have recently discovered. Recent studies demonstrated that the histone variant H2ABbd, when incorporated in place of H2A, renders the chromatin in a more relaxed state and augments gene expression [84]. Discoveries like these hold the potential for much excitement. Histone PTMs are key epigenetic regulators in chromatin-based processes. Increasing evidence suggests that vast combinations of PTMs exist within chromatin histones. Identification of these changes are difficult as they must be carried out by antibody recognition and **chromatin immunoprecipitation** (ChIP). Recently, efforts to develop quantitative, antibody-free, ChIP-less methods have yielded some success [85]. Utilizing a combinatorial histone peptide microarray, unique chromatin states were quantitatively profiled by mass spectrometry to reveal interconnections between nucleosomal histone PTMs. With the recent sequencing of the CHO genome, we anticipate that these tools may be applied to CHO cells, identifying novel CHO histones and their role in chromatin state, permitting more insight into CHO nuclear architecture and pointing toward new strategies in engineering chromatin architecture.

RNA-based epigenetic control mechanisms

lncRNA

ncRNA is commonly used to define RNA that does not encode a protein. Long ncRNA (**lncRNA**) are transcripts longer than 200 nucleotides. Typically these transcripts have little or no protein-coding capacity. Long ncRNAs have various biological functions. By silencing other mRNA, they modulate transcription and influence protein activity [86]. There is much speculation about the abundance of ncRNA, but considering that the coding region of the genome represents only a small portion, the lncRNA is thought to be very abundant [87]. lncRNA can regulate gene expression at many levels (**Figure 2**). These include transcriptional and epigenetic regulation. lncRNA can influence DNA methylation, histone modifications, recruit protein complexes and remodel chromatin. Furthermore, the transcription of lncRNAs can be influenced through epigenetic modifications at their promoters and/or enhancers [88,89].

There is increasing evidence that lncRNA plays an important role in maintaining the DNA methylation marks along specific genes. Many lncRNAs have been identified that exert their influence on DNA methylation by binding to and targeting DNMTs to

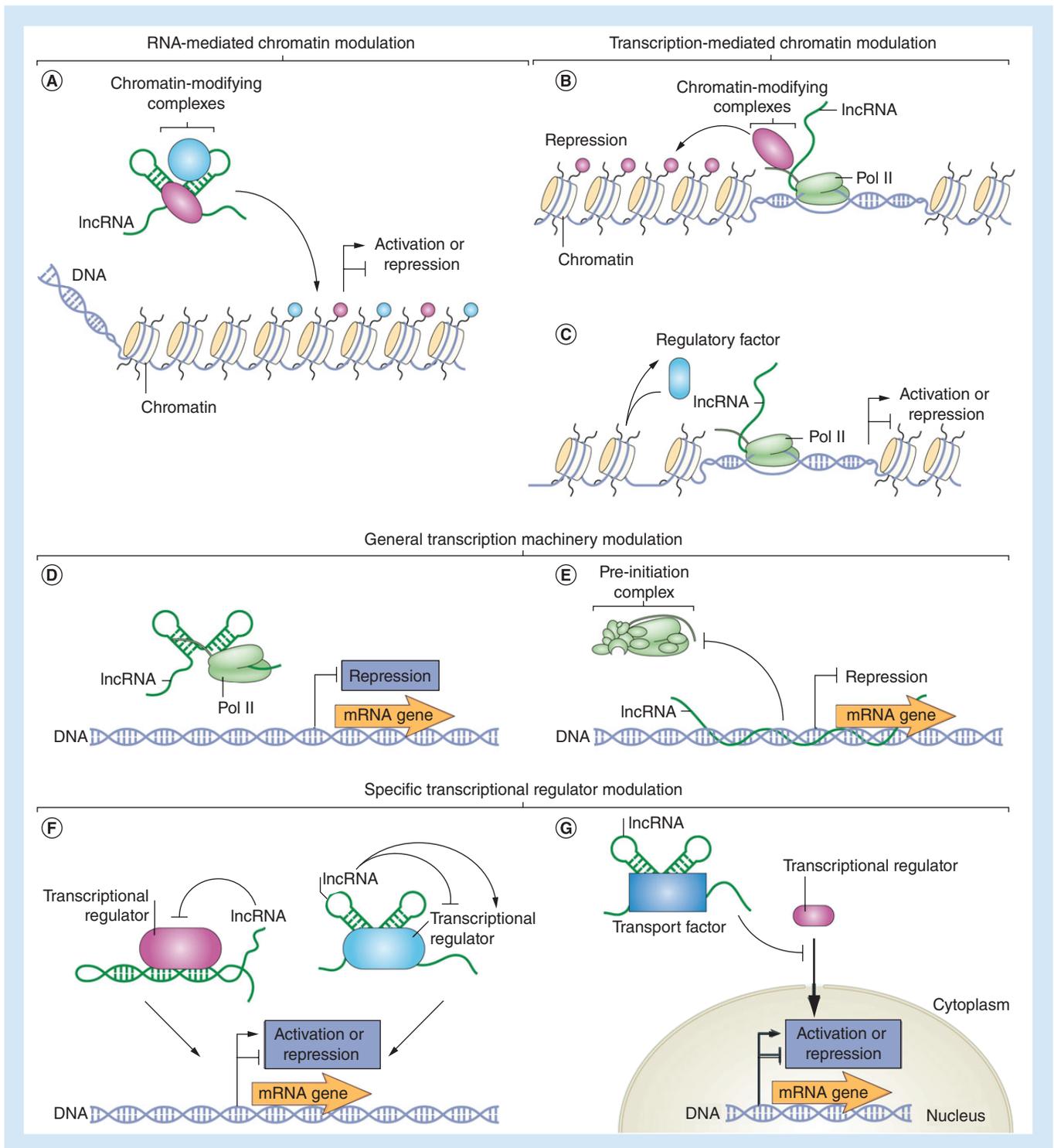


Figure 2. RNA-based epigenetic control mechanisms. Long noncoding RNAs (lncRNAs) can modulate chromatin through (A) transcription-independent and (B & C) transcription-dependent mechanisms. Depending on the nature of the factors that bind during remodeling, gene expression is activated or repressed. (D–G) lncRNAs can modulate both the general transcription machinery and specific regulatory factors. (G) lncRNAs can also regulate gene expression by binding specific transport factors to inhibit the nuclear localization of specific transcription factors. Reproduced with permission from [90].

specific gene loci. The lncRNA *Kcnq1ot1* and *tsix* are known to mediate DNMT activity by interaction with *DNMT1* and *DNMT3a*, respectively [91,92]. In another example, downregulation of lncRNA *ecCEBPA* lead to repression of *CEBPA* transcripts. The loss of lncRNA correlated with increased methylation along the *CEBPA* promoter, leading to repression of transcription [93]. Hence, lncRNA seems to influence both DNA methylation and demethylation via interaction with a range of targets.

lncRNA can also influence gene expression by influencing histone modifications via a number of routes. *Lnc-Jade* indirectly influences H4 acetylation. It positively influences the transcription of protein JADE, which is a cofactor in histone acetylation [94]. There is also evidence to suggest that *Lnc-Jade* is involved in activating gene expression. lncRNA *HOTAIR* has a well-studied and important role in the epigenetic regulation of gene expression. It interacts with the LSD1 complex and is involved in the targeting of PRC2 and LSD1 to chromatin, which promotes histone H3K27 methylation and H3K4 demethylation [95].

lncRNA can also serve as a scaffold to recruit protein complexes and influence the targeting of chromatin modifying complexes to specific target regions. Such complexes establish silencing marks such as DNA methylation and repressive histones. Components of the SWItch/Sucrose NonFermentable (SWI/SNF) chromatin-remodeling complexes also associate with lncRNA-binding proteins. ATP-dependent SWI/SNF proteins comprise a class of enzymes that regulate DNA accessibility through disruption of nucleosome–DNA contacts, movement of nucleosomes along the DNA and removal and exchange of nucleosomes. Protein–lncRNA complexes that associate with SWI/SNF remodelers seem to interfere with RNA polymerase II [96].

In a bioprocessing related application, the lncRNA antisense *IκBα* is overexpressed in some higher producing CHO cells [97]. Further overexpression of the lncRNA antisense to *IκBα* led to improvement in productivity from parental cell clones [97]. This phenomenon has opened up new strategies for host cell engineering. Long ncRNA that targets chromatin remodeling [98] via recruitment of silencing complexes could be targeted for optimizing productivity from clones.

There are many other specific examples of lncRNAs directly and indirectly influencing gene expression, which have been reviewed extensively elsewhere [99,100]. We now know that lncRNA–chromatin-remodeling enzyme interactions are responsible for transcriptional regulation in different cell

types via epigenetic mechanisms, during development and disease [101]. From the above examples, we know that lncRNA can control DNA methylation, demethylation, histone acetylation and methylation, and even recruit chromatin modifying enzymes to specific genomic sites. Many mechanisms have been observed that implicate the lncRNA contribution to all major epigenetic modifications and 3D organization of the chromatin. More and more examples of gene regulation via lncRNA are being elucidated, and a lncRNA database [102] provides comprehensive annotations of eukaryotic lncRNAs [103]. However, much more research is needed to comprehensively describe molecular mechanisms and precise modes of interactions. Nonetheless, lncRNA provides an exciting avenue in bioprocess research.

miRNA

MicroRNAs (miRNAs) are naturally occurring, noncoding, small RNAs that regulate gene expression post-transcriptionally, primarily by interfering with mRNA translation [104]. Therefore, they have a significant role in regulation of expression of enzymes that influence DNA and chromatin modification. The epigenome affects the miRNAs and changes in miRNA expression regulate the epigenome [105]. In CHO cells, only a few miRNAs are known to have an influence on the epigenetic mechanisms. In particular, miR29 reverts methylation by targeting the methyltransferases [106]; the promoters of miRNAs miR148 and miR152 are silenced by *DNMT1*-dependent methylation. *DNMT1* in turn is repressed by miRNAs [107]. From a bioprocessing perspective, glucose depletion activates miR-466h-5p expression by histone deacetylation along the miR-466h-5p promoter [108]. This miRNA has a role in induction of apoptosis, and therefore can be a suitable candidate for engineering. Our understanding of the role of miRNA activity in the context of productivity from mammalian cells is evolving rapidly [109]. As the database of available miRNA transcripts increases [110,111], it will improve the feasibility of miRNA as tools for bioprocessing.

Epigenetic engineering of rRNA

Modulating the gene expression of nontransgenic elements in a cell can open up pathways to higher productivity. In recombinant cells, loss of protein production can also be a function of silencing rRNA [112]. Increased activity of *TIP5* has been implicated in rRNA downregulation [113]. Downregulation of rRNA can lead to loss of protein production in a cell. Knockdown of *TIP5* decreased the DNA methylation along the promoter of ribosomal genes and

increased protein production [114]. Epigenetic mediated modification to rRNA genes opens up new approaches for cellular engineering and may provide new insights into the mechanisms of translational processes in recombinant protein producing cell lines.

New methods to characterize epigenetic influences on protein production

A variety of methods are applied to the study of epigenetic processes, and the past decade has witnessed an exponential increase in novel approaches to elucidate the molecular mysteries of epigenetic inheritance. Most of the elucidation of the chromatin modification and histone protein architecture has been performed by ChIP and DNA methylation studies. However, with the advances in genomics, a variety of new and exciting methods have been developed to decode the mechanisms by which modifications to histones, chromatin and DNA can affect gene expression.

ChIP-seq

Different cells undergoing the selection procedure en route to a higher productivity phenotype will develop unique epigenomes. Methodological limitations of older DNA methylation studies constrained analysis to CGI and promoter regions, largely ignoring the remainder of the genome. Genome-wide histone modification maps based on chromatin immunoprecipitation sequencing (ChIP-seq) [115] should further dissect the machinery determining these marks and their regulatory roles in different cells. Given the complex interplay between DNA methylation, histone modification and subsequent gene regulation, we anticipate discovery of new, coordinated changes in epigenetic patterns that drive productivity. ChIP-seq (Figure 3) combines the powerful methods of ChIP and next-generation sequencing (NGS) to accurately elucidate interactions between proteins, DNA and RNA. With the advent of powerful software tools and even more sensitive detection methods, there will be more reliable information and more importantly, better detection and analyses of distal regulatory regions, which are distant in sequence but brought close in 3D space by DNA bending [116].

PiCh

Another variation of ChIP has been reported as proteomics of isolated chromatin segments (PiCh). PiCh is a powerful method of chromatin isolation that allows identification of proteins bound to the chromatin. The crosslinked chromatin is purified by using magnetic DNA probes. Proteomic techniques

like mass spectrometry can be utilized to elucidate the proteins bound to the chromatin. However, development of suitable probes for the clearing of complexes remains challenging [118].

Insertional ChIP

In an attempt to study the *in vivo* interaction of the DNA protein complexes, insertional ChIP (iChIP) was developed [119]. In this approach, a specific DNA sequence (LexA with a FLAG tag) was inserted upstream of the genomic region of interest (usually a promoter or insulator region). After cross-linking and lysis, the chromatin–protein complex can be recovered by immunoprecipitation and analyzed using proteomic techniques.

Nucleosome occupancy and methylome sequencing

The methylation of individual nucleotides along a promoter does not definitively predict the chromatin architecture and transcriptional activity of the promoter. With the aid of nucleosome occupancy and methylome sequencing we can determine both the DNA methylation status as well as the organization of the nucleotide around the histone core [120]. The assay utilizes the methyltransferase (M.CviPI GpC) that methylates GpC dinucleotides not associated with nucleosomes or transcription factors. Sequencing of the DNA determines which base pairs are wrapped around nucleosomes. This information is then compared with DNA methylation status. By combining this information we can identify the nucleotides that are methylated and also their organization around the core histones.

Oxidative bisulfite sequencing

Oxidative bisulfite sequencing is able to distinguish between cytosine, 5-methylcytosine (5mC) and 5hmC at single base resolution [121]. In this method, 5hmC is oxidized to 5-formylcytosine (5fC) using chemical reagents. After oxidation, bisulfite sequencing is performed and the cytosines and 5fCs are converted to read as thymines while leaving the 5mCs to continue to be read as cytosines. At the same time, a round of bisulfite sequencing is performed on the same DNA sample without the oxidation step. Comparison of the sequences of the two samples will reveal the locations of the 5hmCs.

Key term

Next-generation sequencing: Describes non-Sanger sequencing methods. These methods are high throughput and involve many parallel reads at a time, but require sophisticated bioinformatic analysis.

This is the first method that allows genome-wide detection of 5hmC.

Nanofluidic epigenetic analysis: single chromatin analysis at the nanoscale & nanofluidic DNA methylation detection

This method employs nanofluidics to sort and analyze single, methylated DNA molecules in real time [122]. In this method, DNA is labeled with a red fluorescent marker, while Methyl-CpG-binding domain protein 1, which binds to methylated DNA, is labeled green. After mixing, the device sorts the molecules based on color, sending methylated DNA to one channel and unmethylated DNA into another. Since the methylated DNA fragments are sorted, it is possible to look for the exact locations of those methyl marks using methyl-sensitive restriction enzyme mapping or bisulfite PCR.

Single molecule real-time sequencing

Determining the methylation status along the promoter region of the transgene is essential to gain insight into the molecular nature of transcriptional regulation in higher producing cells. However, the promoter regions have a high GC content, which makes NGS challenging, due to introduction of CG template bias during the amplification step [123–125]. Single-molecule real-time (SMRT) sequencing is a technology recently developed by Pacific Biosciences (CA, USA). SMRT sequencing is based on real-time imaging of fluorescently tagged nucleotides. Since there is no amplification step, the method minimizes the error caused by mismatches [126,127]. SMRT sequencing utilizes the zero-mode waveguide (ZMW) [128]. A ZMW is defined as an optical waveguide that guides light energy into a volume that is small in all dimensions compared with the wavelength of the light. Thus, the observation volume is 20 zeptoliters (20×10^{-21} liters) in a chamber small enough to visualize the individual nucleotide incorporation. In each well, a single DNA polymerase enzyme is tethered at the bottom with a single molecule of DNA as a template. Each of the four DNA bases is attached to one of four different fluorescent dyes. When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved and diffuses out of the observation area of the ZMW. The detector identifies this loss of signal and makes a base call based on the fluorescence signal.

Different modifications to the nucleotide bases in the sequence affect the kinetics of DNA polymerase. The rate of change of fluorescence (rather than the color or magnitude of the change) is an indication of the modification in the incorporated base [129]. Hence,

SMRT sequencing allows for direct detection of DNA methylation. This method is expected to compensate for the major drawbacks of NGS of high-GC content genomes.

Conclusion & future perspective

Many different regulatory factors govern the flow of information from DNA to protein. Epigenetic analysis attempts to unite these phenomena and provide new insights into the mechanisms underlying these factors. Epigenetics still remains a nascent, challenging field of research because of the difficulties in experimental design, time required for sample preparations, the low amounts of material generated and challenges in data analysis. However, newer techniques and interdisciplinary approaches are quickly emerging in the field. The discovery of 5hmC changes our understanding of the DNA sequence complexities. One of the interesting questions is whether hydroxyl methylation is an independent mechanism [69,130] or an intermediate in the demethylation stage [131]. Studies addressing this avenue are expected in the near future.

To date, manipulations of epigenetic mechanisms have been carried out by addition of reagents. Various groups have reported the use of 5-azacytidine as a DNMT inhibitor to positively drive transgene expression. However, treatment with methyltransferase inhibitors has produced mixed results in cell culture [132–134]. Loss of viability due to toxic effect hinders any rise in titers. Furthermore, it is cost prohibitive to use reagents in large scale bioreactors. Using novel genome engineering tools, it will be possible to design knock-out strategies to inhibit DNMT and histone deacetylase activity. The use of RNA elements such as ncRNA and RNAi to specifically target epigenetic regulators along the transgene holds much promise. As observed in the case of ribosomal engineering [114], nontransgenic elements can be targeted using genome engineering to increase productivity.

Probing histones for PTMs is a challenging job. Many different PTMs have been observed on histones, and the presence of multiple markers makes it difficult to elucidate the significance of a particular modification. As cell regulation is thought to be a function of combinatorial modifications on the histones, single marker analysis does not yield significant clues. The advent of powerful analytical tools makes it possible to perform mass spectrometry on histone proteins, elucidate the patterns of histone marks and correlate them to the state of the chromatin [85]. The massive amounts of data that is generated from these efforts will require novel analysis tools to improve our understanding of the various factors and their influence on protein expression [135]. Now that the CHO sequence informa-

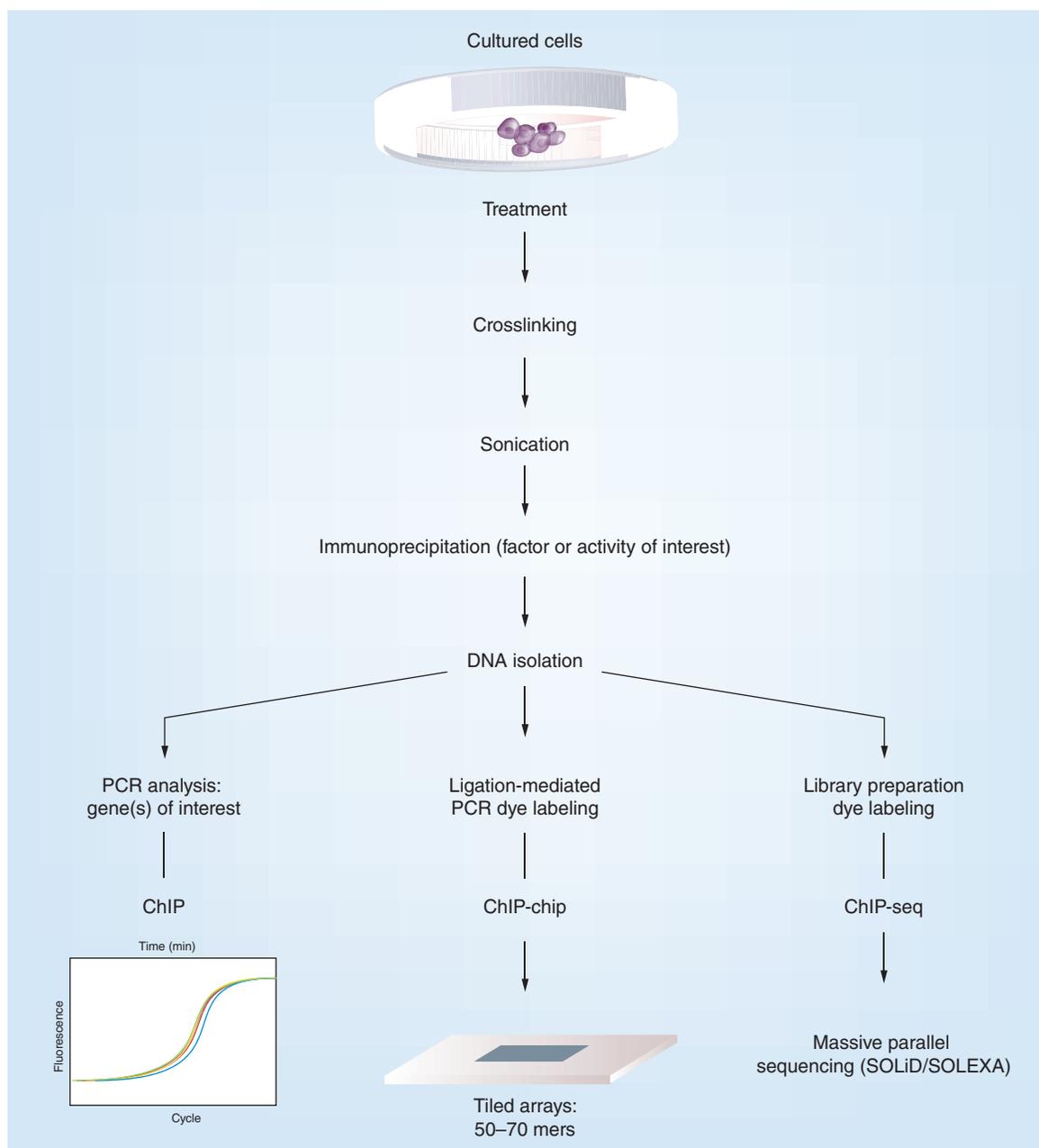


Figure 3. Chromatin immunoprecipitation sequencing procedure. DNA is extracted post crosslinking and is immunoprecipitated with protein of interest. DNA thus obtained can be analysed by PCR, microarray or sequenced. ChIP: Chromatin immunoprecipitation; seq: Sequencing; SOLiD: Sequencing by Oligonucleotide Ligation and Detection.

Reproduced with permission from [117].

tion is available, it will be possible to investigate the epigenome rationally. The elucidation of the interactions between chromatin and transcriptional factors will continue to increase our understanding of the gene expression. Systemic approaches, together with technological advances such as disposable bioreactors and microbioreactors, are expected to lead to increased quality and quantity of biopharmaceuticals, as well as reduced product development times.

Financial & competing interests disclosure

This work was supported in part by a grant from the US National Science Foundation, CBET-0967821. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed..

No writing assistance was utilized in the production of this manuscript.

Executive summary

Background

- Chinese hamster ovary (CHO) cells are the mainstay of recombinant protein production in the biopharmaceutical industry.
- Regulation of factors influencing transgene expression is not completely understood.
- Epigenetic factors and mechanisms that influence epigenetic regulation in the cell are known to influence protein expression from transgenes.

Epigenetic mechanisms & their influence on gene expression

- Some well-characterized epigenetic regulatory factors that have an impact on the production of transgene in cells:
- Positional effects: the position of a gene within the genome can influence its expression. The state of the chromatin along the transgene insert is influenced by DNA methylation and histone modifications.
 - Chromosomal vectors: cloning vectors listed below contain all the regulators necessary to influence the state of the chromatin and thereby reduce silencing due to positional effects;
 - Yeast artificial chromosomes and bacterial artificial chromosomes: yeast artificial chromosomes can accommodate up to 2 Mb of DNA. They are considered complete vectors that contain all the elements to overcome positional effect silencing and augment gene expression;
 - Mammalian artificial chromosomes: these vectors replicate autonomously alongside host chromosomes. They are promising candidates for transfection and hold great potential in increasing productivity and reducing screening times for higher producers;
 - Hot spot targeting: with the aid of precise genome editing tools; it is possible to target the transgene toward transcriptional hot spots;
 - Zinc-finger nucleases: zinc finger nucleases are engineered DNA binding proteins that can cause DNA breaks and facilitate gene inserts at specific locations;
 - Transcription activator-like effector nucleases (TALENs): the newly-developed TALENs are made up of a nonspecific DNA-cleaving nuclease and a DNA-binding domain that can be specifically engineered to target a given sequence. TALENs are poised to have profound impacts on biological research;
 - CRISPR/Cas9: a newly developed genome editing technique inspired by a microbial defense system. It utilizes the RNA–DNA recognition to bind a target in the desired region of the genome and allows its cleavage by Cas9 endonuclease.
- Chromatin modification: eukaryotic chromatin can exist as euchromatin or heterochromatin. The accessibility of chromatin is mainly driven by:
 - DNA methylation;
 - Histone modifications.
- RNA-based epigenetic control mechanisms: RNA is no longer just an intermediate in the central dogma. RNA transcripts can play an important role in governing rate of transcription of a given gene.
 - Lone noncoding RNAs are a class of RNAs that do not code for proteins. They are different from miRNAs as these transcripts are greater than 200 nucleotides. noncoding RNAs may comprise a regulatory network that, including transcription factors, are responsible for finely controlling gene expression in complex eukaryotes;
 - MicroRNA: microRNAs are naturally occurring, noncoding small RNAs that regulate of gene expression by binding to their targets and silencing them. Increasing evidence of microRNA-regulated gene expression has been observed in CHO cells;
 - Epigenetic engineering of rRNA: newly observed phenomenon of nongenic elements that control gene expression has opened up alternate avenues for improving transgene expression.

New methods to characterize epigenetic influence on protein production

- Many new methods have been developed to characterize existing epigenetic markers and also identify new regulators in a high throughput manner:
 - Chromatin immunoprecipitation sequencing;
 - Proteomics of isolated chromatin segment;
 - iChIP;
 - Nucleosome occupancy and methylome sequencing;
 - Oxidative bisulfite sequencing;
 - Nanofluidic epigenetic analysis: single chromatin analysis at the nanoscale and nanofluidic DNA methylation detection;
 - Single-molecule real-time sequencing.

Executive summary (cont.)**Future perspective**

- In the light of the CHO genome, characterization of epigenetic mechanisms will give an insight into the underlying regulatory processes in the cells.
- Newer techniques will help elucidate the known regulators in a high-throughput, genome-scale manner.
- Identifications of newer regulatory mechanism are coming to light and inspiring engineering efforts to exploit epigenetic mechanisms to further augment productivity.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- Hu W-S, Aunins JG. Large-scale mammalian cell culture. *Curr. Opin. Biotechnol.* 8(2), 148–153 (1997).
- Zhu J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* 30(5), 1158–1170 (2012).
- Butler M, Meneses-Acosta A. Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Appl. Microbiol. Biotechnol.* 96(4), 885–894 (2012).
- Baik JY, Linhardt RJ, Sharfstein S. Toward a bioengineered heparin: challenges and strategies for metabolic engineering of mammalian cells. *Bioengineered* 3(4), 227–231 (2012).
- Kim JY, Kim Y-G, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* 93(3), 917–930 (2012).
- Kwaks THJ, Otte AP. Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol.* 24(3), 137–142 (2006).
- Bradbury EM, Van Holde KE. Chromatin. In: *Series in Molecular Biology*. Springer-Verlag, NY, USA, 540 (1988).
- Van Holde K, Zlatanova J. What determines the folding of the chromatin fiber? *Proc. Natl Acad. Sci. USA* 93(20), 10548–10555 (1996).
- Griffith JD. Many ways to loop DNA. *J. Biol. Chem.* 288(41), 29724–29735 (2013).
- Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science* 319(5871), 1787–1789 (2008).
- Bernstein BE, Stamatoyannopoulos JA, Costello JF *et al.* The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 28(10), 1045–1048 (2010).
- **The Roadmap Epigenomics Program attempts to understand the importance of the human epigenome and how it varies across tissues, individuals and disease states. This effort further highlights the importance of epigenetics in cell regulation.**
- Mette MF, van der Winden J, Matzke MA, Matzke AJ. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. *EMBO J.* 18(1), 241–248 (1999).
- Kwaks THJ. Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat. Biotechnol.* 21, 553–558 (2003).
- Harraghy N, Buceta M, Regamey A, Girod P-A, Mermod N. Using matrix attachment regions to improve recombinant protein production. *Methods Mol. Biol.* 801, 93–110 (2012).
- Zeng C, Kouprina N, Zhu B *et al.* Large-insert BAC/YAC libraries for selective re-isolation of genomic regions by homologous recombination in yeast. *Genomics* 77(1–2), 27–34 (2001).
- Kunert R, Casanova E. Recent advances in recombinant protein production: BAC-based expression vectors, the bigger the better. *Bioengineered* 4(4), 258–261 (2013).
- Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* 2(10), 769–779 (2001).
- Yang XW, Model P, Heintz N. Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol.* 15(9), 859–865 (1997).
- Blaas L, Musteanu M, Eferl R, Bauer A, Casanova E. Bacterial artificial chromosomes improve recombinant protein production in mammalian cells. *BMC Biotechnol.* 9(1), 3 (2009).
- Blaas L, Musteanu M, Grabner B, Eferl R, Bauer A, Casanova E. The use of bacterial artificial chromosomes for recombinant protein production in mammalian cell lines. *Methods Mol. Biol.* 824, 581–593 (2012).
- Bian Q, Belmont AS. BAC TG-EMBED: one-step method for high-level, copy-number-dependent, position-independent transgene expression. *Nucleic Acids Res.* 38(11), e127 (2010).
- Mader A, Prewein B, Zboray K, Casanova E, Kunert R. Exploration of BAC versus plasmid expression vectors in recombinant CHO cells. *Appl. Microbiol. Biotechnol.* 97(9), 4049–4054 (2013).
- Ciotta G, Hofemeister H, Maresca M *et al.* Recombineering BAC transgenes for protein tagging. *Methods* 53(2), 113–119 (2011).
- Kennard ML. Engineered mammalian chromosomes in cellular protein production: future prospects. *Methods Mol. Biol.* 738, 217–238 (2011).
- **Describes the utility and advantages of the use of mammalian chromosome vectors to improve protein production and reduce epigenetic silencing.**
- Lindenbaum M. A mammalian artificial chromosome engineering system (ACE system) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucleic Acids Res.* 32(21), e172–e172 (2004).

- 26 Csonka E, Cserpan I, Fodor K *et al.* Novel generation of human satellite DNA-based artificial chromosomes in mammalian cells. *J. Cell Sci.* 113(18), 3207–3216 (2000).
- 27 Kennard ML, Goosney DL, Monteith D *et al.* The generation of stable, high MAb expressing CHO cell lines based on the artificial chromosome expression (ACE) technology. *Biotechnol. Bioeng.* 104(3), 540–553 (2009).
- 28 Cathomen T, Keith Joung J. Zinc-finger nucleases: the next generation emerges. *Mol. Ther.* 16(7), 1200–1207 (2008).
- 29 Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11(9), 636–646 (2010).
- 30 Fan L, Kadura I, Krebs LE, Hatfield CC, Shaw MM, Frye CC. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* 109(4), 1007–1015 (2012).
- 31 Liu P-Q, Chan EM, Cost GJ *et al.* Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol. Bioeng.* 106(1), 97–105 (2010).
- 32 Sun N, Zhao H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol. Bioeng.* 110(7), 1811–1821 (2013).
- **A comprehensive review of the development of transcription activator-like effector nuclease technology.**
- 33 Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* 14(1), 49–55 (2013).
- 34 Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 326(5959), 1501 (2009).
- 35 Ding Q, Regan SN, Xia Y, Ostrom LA, Cowan CA, Musunuru K. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* 12(4), 393–394 (2013).
- 36 Blount BA, Weenink T, Vasylychko S, Ellis T. Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS ONE* 7(3), e33279 (2012).
- 37 Cermak T, Doyle EL, Christian M *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39(12), e82 (2011).
- 38 Morbitzer R, Römer P, Boch J, Lahaye T. Regulation of selected genome loci using *de novo*-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl Acad. Sci. USA* 107(50), 21617–21622 (2010).
- 39 Perez-Pinera P, Ousterout DG, Brunger JM *et al.* Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat. Methods* 10(3), 239–242 (2013).
- 40 Li Y, Moore R, Guinn M, Bleris L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.* 2, 897 (2012).
- 41 Tremblay JP, Chapdelaine P, Coulombe Z, Rousseau J. Transcription activator-like effector proteins induce the expression of the frataxin gene. *Hum. Gene Ther.* 23(8), 883–890 (2012).
- 42 Bultmann S, Morbitzer R, Schmidt CS *et al.* Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.* 40(12), 5368–5377 (2012).
- 43 Valton J, Dupuy A, Daboussi F *et al.* Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol. Chem.* 287(46), 38427–38432 (2012).
- 44 Deltcheva E, Chylinski K, Sharma CM *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* 471(7340), 602–607 (2011).
- 45 Cong L, Ran FA, Cox D *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121), 819–823 (2013).
- 46 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096), 816–821 (2012).
- 47 CRISPR resources. www.genome-engineering.org/crispr/?page_id=41
- 48 DRSC. Finding CRISPRs. www.flyrnai.org/crispr/
- 49 Ronda C, Pedersen LE, Hansen HG *et al.* Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. *Biotechnol. Bioeng.* 111(8), 1604–1616 (2014).
- 50 Miyoshi E, Noda K, Yamaguchi Y *et al.* The alpha1–6-fucosyltransferase gene and its biological significance. *Biochim. Biophys. Acta* 1473(1), 9–20 (1999).
- 51 Wang Y, Ju T, Ding X *et al.* Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc. Natl Acad. Sci. USA* 107(20), 9228–9233 (2010).
- 52 Yang Y, Mariati, Chusainov J, Yap MGS. DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J. Biotechnol.* 147(3–4), 180–185 (2010).
- 53 Osterlehner A, Simmeth S, Göpfert U. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnol. Bioeng.* 108(11), 2670–2681 (2011).
- 54 Turek-Plewa J, Jagodziński PP. The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell. Mol. Biol. Lett.* 10(4), 631–647 (2005).
- 55 Razin A, Cedar H. DNA methylation and gene expression. *Microbiol. Rev.* 55(3), 451–458 (1991).
- 56 Raj K, Mufti GJ. Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. *Ther. Clin. Risk Manag.* 2(4), 377–388 (2006).
- 57 Stallings RL, Crawford BD, Tobey RA, Tesmer J, Hildebrand CE. 5-Azacytidine-induced conversion to cadmium resistance correlates with early S phase replication of inactive metallothionein genes in synchronized CHO cells. *Somat. Cell Mol. Genet.* 12(5), 423–432 (1986).

- 58 Harris M. High-frequency induction by 5-azacytidine of proline independence in CHO-K1 cells. *Somat. Cell Mol. Genet.* 10(6), 615–624 (1984).
- 59 Kim M, O'Callaghan PM, Droms KA, James DC. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.* 108(10), 2434–2446 (2011).
- 60 Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 29(8), 735–741 (2011).
- **Sequencing of the Chinese hamster ovary genome.**
- 61 Becker J, Hackl M, Rupp O *et al.* Unraveling the Chinese hamster ovary cell line transcriptome by next-generation sequencing. *J. Biotechnol.* 156(3), 227–235 (2011).
- 62 Brinkrolf K, Rupp O, Laux H *et al.* Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* 31(8), 694–695 (2013).
- 63 Williams S, Mustoe T, Mulcahy T *et al.* CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol.* 5(1), 17 (2005).
- 64 Ndlovu MN, Denis H, Fuks F. Exposing the DNA methylome iceberg. *Trends Biochem. Sci.* 36(7), 381–387 (2011).
- 65 Illingworth RS, Bird AP. CpG islands – ‘a rough guide’. *FEBS Lett.* 583(11), 1713–1720 (2009).
- 66 Wippermann A, Klausning S, Rupp O *et al.* Establishment of a CpG island microarray for analyses of genome-wide DNA methylation in Chinese hamster ovary cells. *Appl. Microbiol. Biotechnol.* 98(2), 579–589 (2014).
- 67 Brown AJ, Sweeney B, Mainwaring DO, James DC. Synthetic promoters for CHO cell engineering. *Biotechnol. Bioeng.* 111(8), 1638–1647 (2014).
- **Research article that investigates the rational design of synthetic promoter to increase production and improve cell line stability in Chinese hamster ovary cells.**
- 68 Seth G, McIvor RS, Hu W-S. 17Beta-hydroxysteroid dehydrogenase type 7 (Hsd17b7) reverts cholesterol auxotrophy in NS0 cells. *J. Biotechnol.* 121(2), 241–252 (2006).
- 69 Wossidlo M, Nakamura T, Lepikhov K *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2, 241 (2011).
- 70 Pfeifer GP, Kadam S, Jin S-G. 5-hydroxymethylcytosine and its potential roles in development and cancer. *Epigenetics Chromatin* 6(1), 10 (2013).
- 71 Booth MJ, Branco MR, Ficz G *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336(6083), 934–937 (2012).
- 72 Bártová E, Krejčí J, Harnicarová A, Galiová G, Kozubek S. Histone modifications and nuclear architecture: a review. *J. Histochem. Cytochem.* 56(8), 711–721 (2008).
- 73 Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8(12), 983–994 (2007).
- 74 Gelato KA, Fischle W. Role of histone modifications in defining chromatin structure and function. *Biol. Chem.* 389(4), 353–363 (2008).
- 75 Burgess DJ. Epigenetics: asymmetric complexity of the histone code. *Nat. Rev. Genet.* 13(11), 756–756 (2012).
- 76 Birney E, Stamatoyannopoulos JA, Dutta A *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146), 799–816 (2007).
- 77 Krivtsov A V, Feng Z, Lemieux ME *et al.* H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* 14(5), 355–368 (2008).
- 78 Groth A, Rocha W, Verreault A, Almouzni G. Chromatin challenges during DNA replication and repair. *Cell* 128(4), 721–733 (2007).
- 79 Rossetto D, Truman AW, Kron SJ, Côté J. Epigenetic modifications in double-strand break DNA damage signaling and repair. *Clin. Cancer Res.* 16(18), 4543–4552 (2010).
- 80 Koch CM, Andrews RM, Flicek P *et al.* The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* 17(6), 691–707 (2007).
- 81 Yan C, Boyd DD. Histone H3 acetylation and H3 K4 methylation define distinct chromatin regions permissive for transgene expression. *Mol. Cell. Biol.* 26(17), 6357–6371 (2006).
- 82 Jiang Z, Sharfstein ST. Sodium butyrate stimulates monoclonal antibody over-expression in CHO cells by improving gene accessibility. *Biotechnol. Bioeng.* 100(1), 189–194 (2008).
- 83 Majocchi S, Arironovska E, Mermod N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res.* 42(1), 193–204 (2014).
- 84 Gautier T, Abbott DW, Molla A, Verdel A, Ausio J, Dimitrov S. Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep.* 5(7), 715–720 (2004).
- 85 Su Z, Boersma MD, Lee J-H *et al.* ChIP-less analysis of chromatin states. *Epigenetics Chromatin* 7(1), 7 (2014).
- 86 Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23(13), 1494–1504 (2009).
- 87 Carninci P, Kasukawa T, Katayama S *et al.* The transcriptional landscape of the mammalian genome. *Science* 309(5740), 1559–1563 (2005).
- 88 Panning B, Jaenisch R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* 10(16), 1991–2002 (1996).
- 89 Warnecke PM, Biniszkiwicz D, Jaenisch R, Frommer M, Clark SJ. Sequence-specific methylation of the mouse H19 gene in embryonic cells deficient in the Dnmt-1 gene. *Dev. Genet.* 22(2), 111–121 (1998).
- 90 Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14(11), 699–712 (2013).

- 91 Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol. Cell.* 21(5), 617–628 (2006).
- 92 Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development* 137(15), 2493–2499 (2010).
- 93 Di Ruscio A, Ebraldize AK, Benoukraf T *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503(7476), 371–376 (2013).
- 94 Wan G, Hu X, Liu Y *et al.* A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. *EMBO J.* 32(21), 2833–2847 (2013).
- 95 Liu X, Liu Z, Sun M, Liu J, Wang Z, De W. The long non-coding RNA HOTAIR indicates a poor prognosis and promotes metastasis in non-small cell lung cancer. *BMC Cancer* 13(1), 464 (2013).
- 96 Keller C, Kulasegaran-Shylini R, Shimada Y, Hotz H-R, Bühler M. Noncoding RNAs prevent spreading of a repressive histone mark. *Nat. Struct. Mol. Biol.* 20(8), 994–1000 (2013).
- 97 Tabuchi H. Novel strategy for a high-yielding mAb-producing CHO strain (overexpression of non-coding RNA enhanced proliferation and improved mAb yield). *BMC Proc.* 7(Suppl. 6), O3 (2013).
- **Research article on use of overexpression of ncRNA to improve protein production.**
- 98 Rinn JL, Kertesz M, Wang JK *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129(7), 1311–1323 (2007).
- 99 Li L-C. Chromatin remodeling by the small RNA machinery in mammalian cells. *Epigenetics* 9(1), 45–52 (2014).
- 100 Backofen R, Vogel T. Biological and bioinformatical approaches to study crosstalk of long-non-coding RNAs and chromatin-modifying proteins. *Cell Tissue Res.* 356(3), 507–526 (2014).
- 101 Li P-F, Chen S-C, Xia T *et al.* Non-coding RNAs and gastric cancer. *World J. Gastroenterol.* 20(18), 5411–5419 (2014).
- 102 Incrnadb.
www.lncrnadb.org/
- 103 Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNadb: a reference database for long noncoding RNAs. *Nucleic Acids Res.* 39, D146–D151 (2011).
- 104 Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8(2), 93–103 (2007).
- 105 Iorio M V, Piovan C, Croce CM. Interplay between microRNAs and the epigenetic machinery: an intricate network. *Biochim. Biophys. Acta* 1799(10–12), 694–701 (2010).
- 106 Fabbri M, Garzon R, Cimmino A *et al.* MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl Acad. Sci. USA* 104(40), 15805–15810 (2007).
- 107 Xu Q, Jiang Y, Yin Y *et al.* A regulatory circuit of miR-148a/152 and DNMT1 in modulating cell transformation and tumor angiogenesis through IGF-IR and IRS1. *J. Mol. Cell Biol.* 5(1), 3–13 (2013).
- 108 Druz A, Betenbaugh M, Shiloach J. Glucose depletion activates mmu-miR-466h-5p expression through oxidative stress and inhibition of histone deacetylation. *Nucleic Acids Res.* 40(15), 7291–7302 (2012).
- 109 Jadhav V, Hackl M, Druz A *et al.* CHO microRNA engineering is growing up: recent successes and future challenges. *Biotechnol. Adv.* 31(8), 1501–1513 (2013).
- 110 Stiefel F, Hackl M, Grilliari J, Hesse F. Identification of process relevant miRNA in CHO cell lines – process profiling reveals interesting targets for cell line engineering. *BMC Proc.* 7(Suppl. 6), P81 (2013).
- 111 Hackl M, Borth N, Grillari DJ. The CHO miRNA Transcriptome. *MicroRNAs as Tools Biopharm. Prod.* 49–64 (2012).
- 112 Santoro R, Li J, Grummt I. The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat. Genet.* 32(3), 393–396 (2002).
- 113 Santoro R. The silence of the ribosomal RNA genes. *Cell. Mol. Life Sci.* 62(18), 2067–20679 (2005).
- 114 Santoro R, Lienemann P, Fussenegger M. Epigenetic engineering of ribosomal RNA genes enhances protein production. *PLoS ONE* 4(8), e6653 (2009).
- **Research article that employs the use of knockout engineering to reduce the number of silent rRNA to augment production.**
- 115 Valouev A, Johnson DS, Sundquist A *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* 5(9), 829–834 (2008).
- 116 Bailey T, Krajewski P, Ladunga I *et al.* Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput. Biol.* 9(11), e1003326 (2013).
- 117 Pike JW, Meyer MB. The vitamin D receptor: new paradigms for the regulation of gene expression by 1,25-dihydroxyvitamin D(3). *Endocrinol. Metab. Clin. North Am.* 39(2), 255–269 (2010).
- 118 Rusk N. Reverse ChIP. *Nat. Methods* 6(3), 187–187 (2009).
- 119 Hoshino A, Fujii H. Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. *J. Biosci. Bioeng.* 108(5), 446–449 (2009).
- 120 Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 22(12), 2497–506 (2012).
- 121 Booth MJ, Ost TWB, Beraldi D *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* 8(10), 1841–18451 (2013).
- 122 Cipriani BR, Zhao R, Murphy PJ *et al.* Single molecule epigenetic analysis in a nanofluidic channel. *Anal. Chem.* 82(6), 2480–2487 (2010).
- 123 Aird D, Ross MG, Chen W-S *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12(2), R18 (2011).

- 124 Niu B, Fu L, Sun S, Li W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* 11, 187 (2010).
- 125 Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 36(16), e105 (2008).
- 126 Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N *et al.* Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany – NEJM. *N. Engl. J. Med.* 365, 709–717 (2011).
- 127 Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum. Mol. Genet.* 19(R2), R227–R240 (2010).
- 128 Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299(5607), 682–686 (2003).
- 129 Flusberg BA, Webster DR, Lee JH *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7(6), 461–465 (2010).
- 130 Iqbal K, Jin S-G, Pfeifer GP, Szabó PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* 108(9), 3642–3647 (2011).
- 131 Ooi SKT, O'Donnell AH, Bestor TH. Mammalian cytosine methylation at a glance. *J. Cell Sci.* 122(Pt 16), 2787–2791 (2009).
- 132 Klug M, Heinz S, Gebhard C *et al.* Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels. *Genome Biol.* 11(6), R63 (2010).
- 133 Grassi G, Maccaroni P, Meyer R *et al.* Inhibitors of DNA methylation and histone deacetylation activate cytomegalovirus promoter-controlled reporter gene expression in human glioblastoma cell line U87. *Carcinogenesis* 24(10), 1625–1635 (2003).
- 134 Kumpatla SP, Teng W, Buchholz WG, Hall TC. Epigenetic transcriptional silencing and 5-azacytidine-mediated reactivation of a complex transgene in rice. *Plant Physiol.* 115(2), 361–373 (1997).
- 135 Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11(7), 476–486 (2010).