The context of gene expression regulation
Johan H. Gibcus and Job Dekker*

Address: Program in Systems Biology; Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA, USA

* Corresponding author: Job Dekker (Job.Dekker@umassmed.edu)

F1000 Biology Reports 2012, 4:8 (doi:10.3410/B4-8)

Abstract
Recent advances in sequencing technologies have uncovered a world of RNAs that do not code for proteins, known as non-protein coding RNAs, that play important roles in gene regulation. Along with histone modifications and transcription factors, non-coding RNA is part of a layer of transcriptional control on top of the DNA code. This layer of components and their interactions specifically enables (or disables) the modulation of three-dimensional folding of chromatin to create a context for transcriptional regulation that underlies cell-specific transcription. In this perspective, we propose a structural and functional hierarchy, in which the DNA code, proteins and non-coding RNAs act as context creators to fold chromosomes and regulate genes.

Introduction
Transcriptional regulation of eukaryotic genes has classically been viewed as the interaction of elements in the immediate vicinity of the transcription start site (promoters) with upstream elements (enhancers) [1,2]. Mutations in conserved upstream promoter DNA sequences affect gene expression directly [3,4]. However, transcriptional regulation is not only determined by DNA sequence but involves additional layers of control that include nucleosome positioning, DNA binding regulatory proteins such as transcription factors, histone modifications and non-coding RNA [5-7]. Moreover, the development of high throughput chromosome conformation capture techniques [8] has shown that the three-dimensional organization of genes and upstream elements affects transcriptional activity [9-11]. Here, we present how the combination of these factors can provide a “context” for the regulation of gene expression.

Putting the genome into context
The concept of context
Gene expression regulation in eukaryotes includes interactions between promoters and enhancers, but our understanding of the mechanisms that drive these interactions, or that determine their specificity, is far from complete. At the basis of transcription lies the DNA code that directly determines the composition and location of DNA elements and provides specific recognition sites for DNA binding proteins. The binding of transcription factors and recruitment of complexes that modify histones create an environment that allows for element interaction and initiation of gene transcription. However, predicting the location of promoters and enhancers based solely on histone modifications and transcription factor binding relies on complicated models that are still suboptimal [12-14]. Moreover, recent findings on the involvement of non-coding RNA in transcriptional regulation further imply a more complicated reality [15,16]. Is there a combination of histone modifications which is sufficient to predict the position of regulatory elements? Are specific non-coding RNA transcripts involved in recruiting proteins to regulatory elements?

It is now highly debated whether chromatin modifications comprise a “code” similar to DNA (reviewed by Henikoff and Shilatifard [17]). The crosstalk between histones, transcription factors and non-coding RNA suggest that they interact to form a highly interwoven level of organization [16]. In the context model, transcriptional
regulation is subdivided into three levels of interactions: the DNA level, the local chromatin level, and the three-dimensional folding of the genome. The first level, the DNA code, forms an interaction platform by providing protein binding sites for transcription factors that, together with non-coding RNAs and histone modifications, form the next layer of gene regulation. This layer of interactions enables specific long-range interactions that result in a three-dimensional folding of the chromatin. This higher order organization provides transcriptional context that can either facilitate or block the initiation of transcription (Figure 1). Importantly, changes at any level are not necessarily unidirectional. Compaction of higher order structures, for example, will influence the accessibility of DNA and binding of transcription factors, providing a likely feedback mechanism. Below we describe each of the layers of context in more detail.

**Level 1: the DNA code**
The binding of (core) transcription factors critically depends on the recognition of specific DNA sequences, known as DNA motifs. High throughput techniques, such as ChIP-seq and enhanced yeast one-hybrid, which visualizes the interaction of a transcription factor with a bait DNA sequence, are now employed to uncover transcription factor-DNA and DNA-transcription factor interactions, respectively [18,19]. Such studies show that transcription factor-binding events are abundant and can occur at large genomic distances from genes. Similarly, conserved non-coding sequences can be found throughout the genome. Their conservation implies that they have an important function – they may affect the binding affinity of factors, or encode non-coding RNAs [20,21]. Interestingly, deletion of these sequences with unknown function can influence gene expression of genes located hundreds of Kb away, implying that long-range looping of DNA brings the sequences into contact with the genes they regulate [22].

**Level 2: context creators**
This level of interactions involves histone positioning and modifications, repertoires of transcription factors and non-coding RNA and the interplay between them. This level is perhaps the most complex, and it has become the center of attention in recent years.

Although nucleosome positioning, histone modifications, non-coding RNA and transcription factor binding are useful descriptors for genomic elements, they do not seem to define regulatory elements by themselves [17]. Histone modifications have been used to classify upstream regions as promoters (i.e. H3K4me3) and enhancers (i.e. H3K4me1, H3K27ac) [13], but this may not reflect the complete picture. Recent developments in DNA sequencing have led to the appreciation of the importance of non-coding RNA as a regulatory component in the genome [15,23,24]. Although the extent of non-coding RNA involvement is still debated [25], more and more examples of the involvement of non-coding RNA in transcriptional regulation appear in the current literature. Both long and short non-coding RNAs have been identified at regulatory elements [26,27]. Their function is still mostly unknown, and their nomenclature is purely descriptive, based on their site of occurrence (e.g. PASR for Promoter-Associated Short RNA or eRNA for enhancer-RNA) [23,28]. It is possible that non-coding RNA can act as a fast and flexible intermediate to recruit histone-modifying complexes to DNA elements [29-32]. Several reports on long non-coding RNA have shown their involvement in chromatin remodeling, affecting differentiation and disease [33] (reviewed by Huarte and Rinn [34] and Hung and Chang [35]). Long non-coding RNAs have also been found to be involved in transcriptional repression via polycomb proteins, which are known to maintain cell identity by repressing developmental regulators in certain cell types [15,16,31]. Although long
non-coding RNA are now widely studied, small non-coding RNA and antisense RNA have also been implicated in polycomb-mediated transcriptional gene silencing [36-38]. The combinatorial complexity at this level of chromatin regulation and structure is further modulated by feedback and feedforward signals between histone modifications, non-coding RNA and transcription factors.

**Level 3: context**

The three-dimensional folding of DNA is the final context that allows for gene transcription to initiate. The folding of DNA into higher order structures is not a random event, and it has long been thought to affect gene transcription [39-42]. At the nuclear level, chromosomes occupy specific nuclear territories (reviewed by Cremer and Cremer [43]). Chromatin interactions within and between broad zones of chromosomes lead to nuclear compartments where active genes tend to co-locate, near the center of the nucleus, and inactive genes cluster near the nuclear periphery. This indicates a strong correlation, though not necessarily causation, between nuclear positioning and gene activity (reviewed by Geyer et al. [44]). At the finest scale, precise DNA folding or “looping” interactions between gene promoters and their distal regulatory elements can be found. DNA looping (level 3) is guided by long-range interactions between DNA sequence elements (level 1), which can be mediated by interacting context creators (level 2). An illustration of the interplay between multiple levels of regulation that leads to a context for gene expression can be found in the regulation of the HOXA locus. Here, chromosomal looping brings the non-coding RNA HOXTIP in close proximity to HOXA genes. HOXTIP recruits the histone 3 lysine 4 modifying complex MLL. HOTTIP chromosomal looping brings the non-coding RNA and transcription factors.

**Summary**

Recent technical advances in DNA sequencing have enabled genome-wide analysis at each of the three levels: (1) genome sequencing to identify conserved regulatory elements; (2) ChIP-seq [45], DeepCAGE [46], RNA-seq [47], NET-seq [48], Chirp [49] and CHART [50] to identify chromatin modification, transcription factor binding and (non-coding) RNA expression and localization; and (3) 3C combined with deep sequencing (e.g., Hi-C and 3-seq) [8,51,52] to probe three-dimensional folding of the genome. Combined, these tools allow an integrated systems approach towards a more complete understanding of the context in which the genome is regulated.

**Competing interests**

The authors declare they have no competing interests.

**Acknowledgements**

Supported by a Rubicon grant from the Netherlands Organisation for Scientific Research and a Dutch Cancer Society Fellowship to Johan H. Gibcus. Supported by grants from the National Institutes of Health, National Human Genome Research Institute (HG003143 and HG003143-06S1), and a W.M Keck Foundation distinguished young scholar in medical research grant to Job Dekker.

**References**


F1000 Factor 12
Evaluated by David Schatz 13 Jan 2003, Michael Meisterernst 17 Mar 2003

F1000 Factor 6
Evaluated by John Stamatoeyannopoulos 16 May 2011


F1000 Factor 16


F1000 Factor 6
Evaluated by Kenneth Zaret 11 Apr 2007


F1000 Factor 6
Evaluated by 1 King Jordan and Ahsan Huda 21 Oct 2011


F1000 Factor 18


F1000 Factor 20

F1000 Factor 8
Evaluated by Ulf Pettersson 29 Sep 2008


F1000 Factor 8
Evaluated by Andrew D Sharrocks 19 Jun 2006


F1000 Factor 14


F1000 Factor 8
Evaluated by Alain Ghysen 18 Dec 2001


F1000 Factor 15


F1000 Factor 17


F1000 Factor 19


F1000 Factor 8
Evaluated by Andrew D Sharrocks 20 May 2010


F1000 Factor 8
Evaluated by Michael Cole and Mandi Paik 07 Mar 2011


F1000 Factor 8
Evaluated by Leonie Ringrose 01 Aug 2011


44. Cremer T, Cremer M: Gene regulation in the third dimension.


