The spectacular landscape of chromatin and ncRNAs under the Tico sunlight

H. Th. Marc Timmers & Laszlo Tora

The biannual Abcam meeting on Chromatin: Structure & Function held last November covered many aspects of chromatin regulation in health and disease. Important discussion points were the dynamic aspects of chromatin and the ever-increasing involvement of non-coding RNAs in chromatin and epigenetic mechanisms.

The chromatin meetings organized by Tony Kouzarides and Abcam have a tradition of presenting the latest results and concepts in chromatin research in a relaxed Caribbean setting. The 2009 event—held from 16 to 19 November—was no different, as it landed on the Papagayo beach on the Pacific coast of Costa Rica. Attendees could discover the country's impressive natural beauty and geology and experience the friendliness of its Tico population. Many new advances in our understanding of chromatin structure and function marked this exciting meeting. Here, we discuss the insight provided into how long non-coding RNAs (ncRNAs) could be involved in chromatin function and how intrinsically dynamic cellular processes can establish chromatin or epigenetic marking.

Involvement of ncRNAs in chromatin

In the plenary lecture, John Mattick (U. Queensland) provided an overview of RNA-directed phenomena that mediate changes in chromatin, chromosomes and nuclear architecture during differentiation and development of animals and plants (Mercer *et al*, 2009). Several large-scale genomic analyses have been crucial to demonstrate that 90–99% of a mammalian genome is transcribed into RNA at some time in life, whereas less than 2% encodes proteins.

The literature is still dominated by short ncRNAs—such as miRNAs, piRNAs, siRNAs, snoRNAs and snRNAs—but there is an increasing number of reports of long transcripts ranging from 200 bp up to more than 100 kb that do not encode proteins and act to regulate genes through chromatin

(Table 1). Ramin Shiekhattar (Wistar Institute) and John Rinn (Broad Institute) reported separately the genome-wide identification of long ncRNAs, which they called long non-coding (Ln)RNAs and long intergenic non-coding (linc)RNAs, respectively. To find LnRNAs, Rinn searched for chromatin signatures of a short region in which histone H3 Lys4 was trimethylated (H3K4me3), corresponding to active gene promoters, and a larger region of H3K36me3, which corresponds to pol II transcribed regions. His group found about 5,000 lincRNAs, a few of which were reported to associate with Polycomb group (PcG) proteins, forming the repressive PRC2 complex (Mercer *et al*, 2009). By using *in vivo* RNA

immunoprecipitations against PRC2 components and DNA microarray detection, as many as 20% of the new lincRNAs were found to associate with PRC2, suggesting a more global mechanism (Khalil *et al*, 2009). The Rinn group proposes that lincRNAs impart specificity in the deposition of H3K27me3 to repress transcription through triple helix formation, RNA–protein complexes with specific transcription factors and RNA processing proteins, or by facilitating DNA looping through RNA–protein interactions. Importantly, lincRNAs seem to mediate gene repression *in trans*, but *cis*-acting models cannot be ruled out (Khalil *et al*, 2009). In contrast to this are the LnRNAs identified by Shiekhattar and co-workers through the use of systematic, large-scale RNA sequencing and genome annotation. From 30% of the genome, they identified about 3,000 LnRNAs, which seem to contribute to an enhancer-like activity to co-regulate genes within 100 kb. The extent of overlap between the lincRNA and LnRNA families is unclear, but mammalian genomes could encode up to 10,000 of these ncRNAs. Whether they all have chromatin-dependent and/or epigenetic-dependent regulatory roles remains to be elucidated.

The ncRNAs produced from the *Saccharomyces cerevisiae* genome can be divided into stable unannotated transcripts (SUTs) and cryptic unstable transcripts (CUTs), which initiate mostly from nucleosome-free regions associated with divergent mRNA promoters (Xu *et al*, 2009). Antonin Morillon (Institut Curie, Paris) and Claude Thermes (CGM, Gif-sur-Yvette) used RNAseq

Table 1 | Non-coding RNAs that regulate transcription or translation through chromatin-associated pathways

RNA	Size	Mechanism of action	Described in
siRNA	~21–25 nt	CDGS and PTGS	Plants, animals, fungi
miRNA	~21–25 nt	CDGS and PTGS	Plants, animals
piRNA	~21–31 nt	CDGS and PTGS	Animals
snoRNA	~60–300 nt	Site-specific modification of nucleotides in target RNAs	Plants, animals, fungi
TSSa RNA	~20–90 nt	Regulation of chromatin and transcription?	Animals
CUTs°	~300–500 nt	Transcriptional silencing?	Fungi
SUTs°	~500–1500 nt		
lincRNA	~200–5000 nt	CDGS	Animals
LnRNA°	~100–1000 nt	CDGA	Animals

Housekeeping spliceosomal snRNAs and tRNAs, which are involved in splicing and translation, are not shown.

° Ramin Shiekhattar and Antonin Morillon, personal communication during the meeting.

CDGA, chromatin-dependent gene activation; CDGS, chromatin-dependent gene silencing; CUTs, cryptic unstable transcripts; lincRNA, long intergenic non-coding RNA; LnRNA, long non-coding RNA; PTGS, post-transcriptional gene silencing; SUTs, stable unannotated transcripts; TSSa, transcription start site-associated.

to identify about 1,100 ncRNAs that they called XUTs (Xrn1p-sensitive unstable transcripts), as they are increased in yeast that lack the Xrn1p 5'–3' exonuclease. Many belong to the SUT family, but they also found novel ncRNAs that are transcribed in antisense orientation with respect to adjacent mRNA genes. They identified a non-coding transcript responsible for the repression of the nearby *TIR1* gene, which involves the Set1C/COMPASS methyltransferase and deposition of H3K4me2/3. The data supports a model in which certain yeast promoters are controlled by cryptic or non-coding transcripts that induce chromatin modifications.

Many new advances in our understanding of chromatin structure and function marked this exciting meeting

The EMSY protein is an important link between BRCA2 and the development of sporadic breast and ovarian cancer. EMSY is a nuclear protein, which seems to be involved in DNA-damage response and chromatin remodelling. Tony Kouzarides (Gurdon Institute) reported that EMSY-dependent chromatin pathways involve miRNAs and can lead to breast cancer. His group discovered that EMSY can regulate the expression of three miRNAs that are upregulated in metastatic cancer cells. EMSY is

recruited by a transcription factor to the miRNA promoters to repress their transcription. Interestingly, this repression involves the action of a histone demethylase enzyme.

Putting Polycomb in place

The PcG proteins, which are crucial for establishing and maintaining epigenetic control, are divided among the repressive PRC1 and PRC2 complexes. By virtue of its RING domain proteins (Ring1B and Bmi1), PRC1 mediates histone H2A ubiquitination. Methylation of H3K27 by the Ezh1/2 lysine methyltransferases of PRC2 has an important role in PRC1 recruitment, but how PRC2 binds to specific loci has remained enigmatic for a long time. Caroline Woo from Kingston's lab (MGH, Boston) presented the identification of a 1.8 kb region in the human HOXD cluster with binding sites for the YY1 protein. Bmi1 (PRC1) and Eed (PRC2) proteins were found to be associated with this evolutionarily conserved region located between the HOXD11 and HOXD12 genes. In a reporter gene assay, this region can autonomously repress transcription in human embryonic stem (ES) cells, for which it is dependent on Bmi1, Eed and RYBP (Woo *et al*, 2010). The crucial test will now be whether this Polycomb response element also confers PcG responsiveness in animals.

Kristian Helin (BRIC, U. Copenhagen) reported that Jarid2 (Jumonji, AT-rich interactive domain 2) protein is part of a subpopulation of PRC2, both in murine ES and human differentiated cell types.

Jarid2 is the founding member of the Jumonji-domain-containing histone lysine demethylases (KDMs), but it lacks enzymatic activity owing to the absence of key residues in the Fe²⁺-binding module. A small region of Jarid2 (aa 147–165) is essential for PRC2 interaction and transcriptional repression. Its ARID domain is required for PRC2 association to target genes and has been found to bind to both A/T-rich and G/C-rich DNA sequences. ChIP-seq data showed that the genomic distributions of Jarid2 and Suz12 (PRC2) are virtually overlapping. A model emerges of the catalytically inactive Jarid2 as a direct binder of DNA, which recruits PRC2 to Polycomb-regulated genes (Pasini *et al*, 2010). Two other groups reported similar findings in a recent issue of *Cell* (Peng *et al*, 2009; Shen *et al*, 2009). All groups agree that only a fraction of PRC2 complexes contain Jarid2 and *vice versa*. This leaves open the possibility of other recruiters for PRC2. Relevant to this is the analysis of lincRNA complexes as discussed above. However, many questions remain unanswered: do lincRNAs and Jarid2 coexist (and collaborate?) in the same PRC2 complex, and do they represent protein- or RNA-based routes to recruit PRC2 for the establishment of PcG-mediated silencing of chromatin? Does Jarid2 also exert chromatin functions separate from PcG proteins? Are Jarid2 and/or lincRNAs sufficient for PRC2 recruitment to specific loci? Or is there still a role for DNA sequence-specific binding proteins?

Chromatin is jumping up and down

Eukaryotic genomes encode various SWI2/SNF2-like ATPases, which assemble into chromatin remodelling complexes. These complexes—such as SWI/SNF, ISWI, INO80 and NURD—hydrolyse ATP to alter DNA-nucleosome contacts and can also exchange histones on nucleosomal templates. Although they share certain subunits with SWI/SNF complexes, vertebrate BAF complexes—which contain Brg1 or Brm—generally function to repress genes at a distance. Lena Ho, from the Crabtree lab (Stanford U.), showed that ES cells have a family of BAF complexes—known as esBAF—which are essential for development and maintenance of pluripotency and contain subunits such as Brg1, BAF155 and BAF47. Genome-wide localization data suggest that esBAF regulates the chromatin structure globally to ensure access of the pluripotency factors to their targets and, thus, esBAF is an important

'executer' of the transcriptional network that regulates pluripotency.

Yuri Moshkin, from the Verrijzer lab (Erasmus U.), presented a comprehensive genome-wide binding survey of different *Drosophila* remodelling complexes (SWI/SNF, ISWI and NURD) and their impact on chromatin structure and gene expression. He reported that these remodellers create distinct classes of locally open or closed chromatin, thereby allowing specific transitions in histone density and DNA accessibility. The fly genome seems to be densely populated by distinct ATP-dependent remodelling complexes, which display only limited overlap. The depletion of individual remodellers causes global, but distinct, changes in chromatin structure, providing a molecular basis for the unique control of gene expression and biological programmes prompted by different classes of remodellers.

...large-scale genomic analyses have [... demonstrated] that 90–99% of a mammalian genome is transcribed into RNA at some time in life, whereas less than 2% encodes proteins

The theme of ATP-dependent remodellers was continued by Patrick Varga-Weisz (Babraham Institute) who identified ATAD2 as an interacting partner of the REST (repressor element 1-silencing transcription) factor, involved in the repression of neuronal genes. ATAD2—which contains a double AAA ATPase domain and a bromo-domain—is bound firmly to chromatin and ChIP-seq revealed a strong overlap of ATAD2 and REST binding sites. Although REST and ATAD2 stimulate each other's interaction with chromatin, ATAD2 seems to antagonize REST-mediated repression.

DNA sequence has been shown recently to be an important determinant for nucleosome positioning. Jim Broach and co-workers (Princeton U.) discussed genome-wide nucleosome maps obtained from yeast growing in glycerol or in glucose. Although the gene expression patterns differ widely between these conditions, there is surprisingly little difference in nucleosome positioning at the promoters of differentially expressed genes (Zawadzki *et al.*, 2009). However, promoters containing a TATA box are more likely to undergo changes in

nucleosome positioning, which supports previous models suggesting that TATA-box-containing promoters are highly regulated chromatin structures.

Spreading the 'histone code' word

Many participants touched on the regulation of histone modifications through the crosstalk between histone tails and modifying complexes. Gaining a deeper insight into this process is essential to appreciate the physiological implications of the 'histone code', as the sheer number of combinations of individual modifications is immense. Danny Reinberg (Howard Hughes Medical Institute) discussed the essential function of PR-SET7-mediated monomethylation of H4K20 during mitosis and its importance in chromosome compaction. This modification is found randomly throughout the genome, but it can be silenced by acetylation of the neighbouring H4K16. H4K20me1 provides binding to the L3MBTL1 protein and also increases chromosome compaction, but whether this is a direct effect or mediated by other proteins remains unknown.

Ali Shilatifard (Stowers Institute) discussed the division of labour between the MLL and hSet1A/B complexes, which are involved in H3K4 methylation in mammalian cells. The hSet1A/B complexes seem to be responsible for global levels of H3K4me3, whereas the MLL1/2—menin—and possibly the MLL3/4—complexes are gene-specific (Wang *et al.*, 2009). The generation of specific small-molecule inhibitors for each of the different MLL and hSet1A/B complexes would be of great help to address this issue. Shilatifard mentioned briefly the identification of lead compounds specific for MLL1 and the yeast Set1p enzymes, although further chemistry will be required because the IC₅₀ of these compounds now range around 100 µM. Mark Fiedler, from the Bienz laboratory (MRC, Cambridge, UK), described allosteric effects in the binding of modified chromatin. The HD1 domain of BCL9/Legless binds to the plant homeo-domain of the Pygopus protein and alters its pocket, allowing it to bind to Thr3 of H3 and leading to an increased affinity for H3K4me2/3. He also showed that the *Drosophila* Pygopus/Legless complex requires the binding to H3R2me2a together with H3K4me2 to achieve affinities comparable with the orthologous mammalian complex.

The future of chromatin and epigenetic regulation in all kinds of cellular processes could not be any brighter...

Maria-Elena Torres-Padilla (IGBMC, Illkirch) discussed the role of H3 variants in the mouse embryo after fertilization; methylation of H3 establishes a parental asymmetry and is predominant in the maternal pronucleus. She investigated the incorporation of differentially modified H3 variants in the parental chromatin. H3.3 localizes specifically to paternal pericentromeric chromatin during S phase and acts to regulate transcription from repeat regions. Surprisingly, among the lysines analysed, only Lys27 of H3.3 is necessary for early embryogenesis, which indicates a novel function for Lys27 methylation in the establishment of early embryonic heterochromatin.

The impressive natural beauty of Costa Rica was matched by the great chromatin landscape sketched by the participants of this meeting. More than a decade has passed since chromatin stepped out of its cave and into the daylight of gene expression control, thanks to the discovery of chromatin modifying and remodelling activities as transcriptional co-activators. ncRNAs are now also stepping out of the shade to join chromatin in the sunlight. The future of chromatin and epigenetic regulation in all kinds of cellular processes could not be any brighter, and one can hardly wait for the next Chromatin meeting under the Caribbean sun.

REFERENCES

- Khalil AM *et al.* (2009) *Proc Natl Acad Sci USA* **106**: 11667–11672
- Mercer TR, Dingler ME, Mattick JS (2009) *Nat Rev Genet* **10**: 155–159
- Pasini D *et al.* (2010) *Nature* doi:10.1038/nature08788
- Peng JC *et al.* (2009) *Cell* **139**: 1290–1302
- Shen X *et al.* (2009) *Cell* **139**: 1303–1314
- Wang P *et al.* (2009) *Mol Cell Biol* **29**: 6074–6085
- Woo CJ *et al.* (2010) *Cell* **140**: 99–110
- Xu Z *et al.* (2009) *Nature* **457**: 1033–1037
- Zawadzki KA, Morozov AV, Broach JR (2009) *Mol Biol Cell* **20**: 3503–3513

Marc Timmers is at the University Medical Center Utrecht, Utrecht, The Netherlands.

E-mail: h.t.m.timmers@umcutrecht.nl

Laszlo Tora is at IGBMC, UMR 7104 CNRS, University of Strasbourg, Illkirch, France.

Published online 12 February 2010

EMBO reports (2010) **11**, 147–149.
doi:10.1038/embor.2010.26