npg

RESEARCH HIGHLIGHT

Non-coding transcription SETs up regulation

An abundance of long non-coding RNA (lncRNA) present in most species from yeast to human are involved in transcriptional regulation, dosage compensation and imprinting. This underscores the importance of lncRNA as functional RNA despite the fact that they do not produce proteins. Two recent papers in *Cell* have demonstrated that transcription of the non-conserved lncRNAs, but not the RNAs themselves, is necessary to introduce co-transcriptional regulatory histone marks to regulate gene expression.

Transcription over a chromatin template is modulated by a number of post-translational modifications. Histone acetylation is a key activating mark that facilitates access to the DNA not only by impairing the ability of nucleosomes to bind the DNA tightly, but also by recruiting factors that aid the movement of the nucleosomes *in cis* or *in trans*. Therefore, in order to achieve a tight control over gene expression, it is imperative that the acetylation marks must be removed by histone lysine deacetylases (KDACs). While promoter acetylation is removed by the targeted recruitment of KDACs by repressor proteins, acetylation over the gene body is removed in a co-transcriptional manner. Studies from several labs have identified two RNA polymerase II (RNAPII)-associated KDAC complexes [1] that act on the 5′ (SET3 complex) and 3′ (Rpd3S complex) ends of gene bodies (Figure 1A), maintaining nucleosomes in a hypoacetylated state [2, 3]. The specificity of targeting these deacetylases to specific regions of the gene body is achieved by the differential distribution of H3K4 di-methylation (H3K4me2) and the H3K36 methylation marks over the gene body (Figure 1B). These methylation marks are deposited over nucleosomes at a specified distance from the promoter owing to the RNAPII phosphorylation kinetics [4], which recruits the histone lysine methyltransferase (KMT) Set1 complex to methylate H3K4 at the 5′ ends of genes [5] and Set2 to methylate H3K36 at the 3′ ends of genes [6]. These methylation marks are recognized by specific methyl-lysine binding domains in the Set3 and Rpd3S complexes, resulting in targeted activation of the deacetylase. Thus, co-transcriptional deacetylation of nucleosomes over gene bodies regulates RNAPII elongation, thereby controlling gene expression. Interestingly, the loss of Set2-mediated H3K36 methylation gives rise to a variety of transcripts, both sense and antisense, which are initiated from within the gene bodies. A majority of these transcripts do not code for proteins, which gives rise to the question as to what role these internally initiated transcripts play in the cell? Furthermore, does this spurious initiation of transcripts also occur upon the loss of Set3?

Buratowski and colleagues elegantly answered the latter question using yeast as a model system. Since the loss of Set3 produces minimal change in yeast gene expression in glucose-containing media [7], Kim *et al.* [8] subjected the wild-type yeast and *set3* deletion mutant to a series of carbon source changes to determine effects of Set3 on gene expression. The Set3 complex, including the deacetylases Hos2 and Hst1 (Figure 1A), was found to repress transcription in a number of genes that demonstrated elevated levels of H3K4me2 over the promoter regions. Interestingly, two-thirds of these genes were previously shown to have overlapping ncRNAs that initiated from a distal promoter [7, 9, 10]. In the presence of the repressive carbon source, ncRNA production from the distal promoter results in the co-transcriptional addition of H3K4me2 over nucleosomes at the proximal promoter. This mark recruits the Set3 complex, deacetylating these nucleosomes, thereby causing delayed expression or repression of the downstream genes. In genes that do not have upstream overlapping transcription, the Set3 complex was shown to suppress ncRNA arising from the 5′ ends of genes (Figure 1C). This is similar to the role played by Set2 at the 3′ ends of genes [3]. Interestingly, Buratowski and colleagues also found genes that were activated by the Set3 deacetylase complex, due to the repression of antisense ncRNA. Genes containing such anti-correlative antisense transcripts are usually involved in the responses to environmental changes or stress conditions. Could the ncRNA-mediated regulation of gene expression play a key role in modulating such responses?

To answer this question, Amon and colleagues, in a study published in *Cell* [11], used the classical genetic model of mating-type control of sporulation in yeast. Upon nutrient deprivation, diploid yeast cells undergo meiosis resulting in the production of four haploid gametes housed in stress-resistant spores. The entire sporulation program is under the control of a transcription factor, Ime1. Gene expression from the *IME1* locus
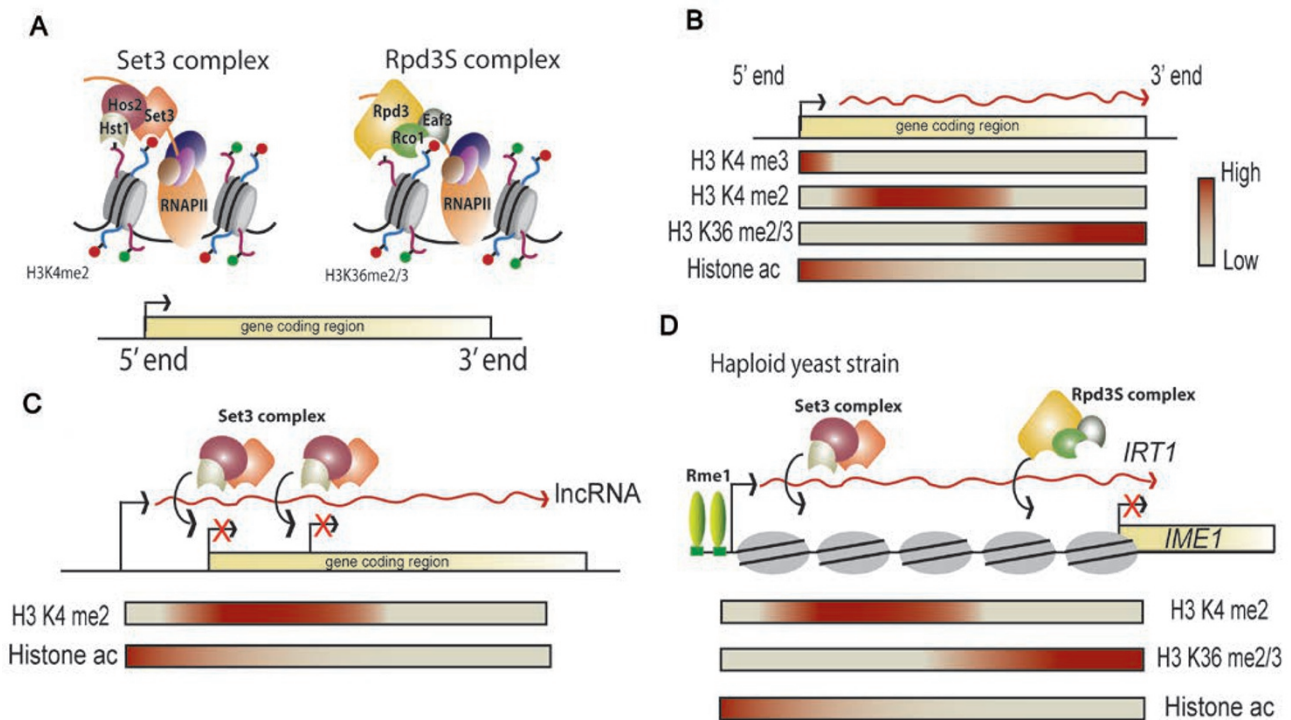
Figure 1 Long non-coding RNA (lncRNA) transcription recruits co-transcriptional histone modifying factors to repress downstream genes. **(A)** RNA polymerase II (RNAPII)-mediated recruitment of the Set3 and Rpd3S complexes at the 5′ and 3′ ends of genes, respectively. Set3 recognizes H3K4 di-methylation (H3K4me2), while the Eaf3 subunit of Rpd3S recognizes the H3K36 di- and tri-methylation marks (H3K36me2/3) to activate the deacetylase of each of the complexes. Methylation is denoted by the red dot, while acetylation is denoted by the green dot. **(B)** Transcription-dependent distribution of H3K4 and H3K36 methylation over the coding region of genes. Histone acetylation (Histone ac) anti-correlates with H3K4me2 and the H3K36me2/3 distribution due to the recruitment of the specific deacetylases. A key to the color code is provided. **(C)** Expression of lncRNA from an upstream promoter results in distribution of H3K4me2 marks over the promoter of the downstream gene, resulting in the recruitment of the Set3 deacetylase complex and the repression of gene expression. Internally initiated transcripts are also suppressed by a similar mechanism. **(D)** Transcription factor Rme1-dependent transcription of the lncRNA *IRT1* in the haploid yeast strain establishes a gradient of H3K4me2 and H3K36me2/3 marks over the promoter of the *IME1* gene, resulting in the Set3- and Rpd3S-mediated deacetylation, setting up a repressive chromatin structure and preventing transcription initiation.

is a tightly controlled process, occurring only in diploid cells sensing starvation. To prevent lethal cell division from occurring in haploid cells in response to starvation, the *IME1* gene is known to be inhibited by the Rme1 transcription factor. *RME1* gene transcription is inhibited in diploid cells by the a1-α2 repressor complex, which is absent in each of the haploid mating types. Precisely how a transcription factor, known to activate gene transcription, is involved in the repression of a gene ~2 kb downstream was not clearly understood. van Werven *et al.* [11] tackled this question in a systematic manner beginning with the identification of a stable un-annotated transcript (SUT), *IRT1* (*IME1* regulatory transcript 1) regulated by Rme1. The expression of *IRT1* anti-correlated with that of *IME1*, and was reduced upon induction of sporulation in diploid cells. Interestingly, *IRT1* inhibited *IME1* transcription *in cis*, suggesting that the mechanism was unlike the ncRNA-mediated silencing of the *HOXD* locus in humans. The *IRT1* transcript prevented binding of a *IME1*-specific activator by establishing a repressive chromatin structure at the *IME1* promoter, abolishing the nucleosome free region (NFR). Loss of the KMT Set2 and the KDAC Set3 in combination with each other resulted in the activation of the *IME1* gene in the presence of the *IRT1* lncRNA, suggesting that both these proteins were necessary for *IRT1*-mediated repression. Consistent with this observation, loss of Set2 and Set3 resulted in the recruitment of the *IME1*-specific transcription factor and NFR formation. Set2-mediated H3K36 methylation has recently been shown to prevent histone exchange over gene bodies [12]. Thus, H3K36 methylation

over the *IME1* promoter may not only result in deacetylation of nucleosomes, but also in decreased histone exchange, leading to the abrogation of the NFR. Indeed, artificial mis-targeting of the Set2 protein to gene promoters has been shown to result in repression of transcription [13]. Finally, Amon and colleagues also found that the *IME4* gene, necessary for facilitating entry into the sporulation program along with the transcription factor Ime1, was regulated in a similar manner by an antisense transcript (*IME4-AS*). Abrogation of *IRT1* and *IME4-AS* transcripts resulted in the initiation of lethal meiosis in haploid yeast cells, underscoring the importance of the ncRNA-mediated regulation in the sporulation gene expression program.

lncRNA-mediated transcription regulation is known to occur through a variety of different mechanisms. As mentioned earlier, silencing of the human *HOXD* locus is achieved by the *HOTAIR* (Hox transcript antisense RNA) ncRNA that recruits the Polycomb repressive complex and results in remodeled chromatin. X-chromosome inactivation and dosage compensation also proceed through a similar mechanism where ncRNAs recruits a chromatin remodeling complex.

ncRNAs have also been shown to recruit proteins to specific promoters, antagonizing the function of transcriptional coactivators p300/CBP or act without an intermediary protein complex to prevent TFIID binding through the formation of a triplex with the DNA at the promoter regions (reviewed in [14]). One of the key features distinguishing these mechanisms from the two papers from the Buratowski and Amon labs is that they act *in trans*. These papers in *Cell* highlight the fact that the process of transcribing ncRNAs itself can regulate gene expression from overlapping targets by utilizing the Set2- and Set3-mediated chromatin repression mechanisms [2, 3, 12, 15]. Interestingly, despite the fact that a majority of ncRNAs have relative short half-lives, this mechanism would not be limited by stability issues since the ncRNA itself is not involved in the process of repression. It remains to be seen whether the ncRNA transcription-dependent setting of repressive chromatin is conserved in higher organisms.

Swaminathan Venkatesh[1],
Jerry L Workman[1]

[1]*Stowers Institute for Medical Research, Kansas City, MO 64110, USA*

Correspondence: Jerry L Workman
E-mail: JLW@stowers.org

References

1   Govind CK, Qiu H, Ginsburg DS, *et al. Mol Cell* 2010; **39**:234-246.
2   Kim T, Buratowski S. *Cell* 2009; **137**:259-272.
3   Carrozza MJ, Li B, Florens L, *et al. Cell* 2005; **123**:581-592.
4   Bartkowiak B, Mackellar AL, Greenleaf AL. *Genet Res Int* 2011; **2011**:623718.
5   Ng HH, Robert F, Young RA, *et al. Mol Cell* 2003; **11**:709-719.
6   Li B, Howe L, Anderson S, *et al. J Biol Chem* 2003; **278**:8897-8903.
7   Lenstra TL, Benschop JJ, Kim T, *et al. Mol Cell* 2011; **42**:536-549.
8   Kim T, Xu Z, Clauder-Munster S, *et al. Cell* 2012; **150**:1158-1169.
9   Xu Z, Wei W, Gagneur J, *et al. Nature* 2009; **457**:1033-1037.
10 van Dijk EL, Chen CL, d'Aubenton-Carafa Y, *et al. Nature* 2011; **475**:114-117.
11 van Werven FJ, Neuert G, Hendrick N, *et al. Cell* 2012; **150**:1170-1181.
12 Venkatesh S, Smolle M, Li H, *et al. Nature* 2012; **489**:452-455.
13 Strahl BD, Grant PA, Briggs SD, *et al. Mol Cell Biol* 2002; **22**:1298-1306.
14 Mercer TR, Dinger ME, Mattick JS. *Nat Rev Genet* 2009; **10**:155-159.
15 Smolle M, Venkatesh S, Gogol MM, *et al. Nat Struct Mol Biol* 2012; **19**:884-892.