Biochimica et Biophysica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagrm

Pervasive lncRNA binding by epigenetic modifying complexes — The challenges ahead☆

Juan G. Betancur

Laboratory for Developmental Genetics, RIKEN-IMS, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form 1 October 2015
Accepted 8 October 2015
Available online xxxx

Keywords:
lncRNA
Polycomb
RNA-binding protein
DNA methylation
Chromatin remodeling
Histone modification

ABSTRACT

Epigenetic modifying factors are fundamental regulators of chromatin structure and gene expression during development and differentiation through the induction of chemical modifications on histones, DNA or *via* remodeling of the chromatin structure. Protein complexes involved in these three processes contain non-canonical RNA-binding components that interact with long non-coding RNAs, in many cases in the absence of any sequence or structural signatures. However, there is growing evidence of the role of such protein–lncRNA interactions in the regulation of the epigenetic landscape *in vivo*. This review summarizes the growing number of epigenetic modifying factors described to interact with lncRNAs in mouse and human, and then discusses the challenges that lay ahead in understanding lncRNAs as part of the intricate networks of epigenetic regulation. A combination of protein and RNA-centric approaches is required for this purpose.
This article is part of a Special Issue entitled: Clues to long noncoding RNA taxonomy, edited by Dr. Tetsuro Hirose and Dr. Shinichi Nakagawa.

1. Introduction

A large fraction of eukaryotic genomes is transcribed, but a major part of the total transcriptome does not encode any proteins. RNAs that fall in this category are collectively known as non-coding RNAs (ncRNA) [1]. Many ncRNAs fall within one of the well known classes of small ncRNAs (miRNAs, piRNAs, siRNAs, among others), while ncRNAs that are longer than ~200 nt are classified as long non-coding RNAs (lncRNAs). The only common feature to all lncRNAs is their size, and instead members of this class show large variations in stability, intracellular localization, cell type-specific expression pattern and function [2–4]. Long non-coding RNA genes evolve at a faster rate than protein coding genes and therefore many are likely to lack any biological function and some might even disappear before they acquire any [5–7]. Therefore, the current challenge is to identify lncRNAs that are physiologically relevant. Most functional lncRNAs do not act alone and instead form ribonucleoprotein complexes (RNPs) to exert their functions; therefore, identification of the protein partners of lncRNAs and their RNA-binding regions (RBRs), is crucial to understand lncRNA functions.

DNA in eukaryotic organisms is highly compacted into a dynamic structure known as chromatin and regulatory events that occur on chromatin without altering the DNA sequence are known collectively as epigenetics. Epigenetic modifying factors are broadly classified into histone modifying, DNA modifying and chromatin remodeling factors. Regulation at this level controls the degree of compaction and thus the accessibility of transcription factors and the transcriptional machinery. This happens through a combination of chemical modifications on histone tails and the DNA as well as active remodeling of the chromatin structure [8,9]. There is growing biochemical and genetic evidence that these complexes are protein partners of a subset of the mammalian lncRNA repertoire and that virtually all subclasses of epigenetic modifying factors have at least one member known to interact with lncRNAs (Table 1, Fig. 1A). These observations hint at RNA–protein interactions being a general element in the regulation of chromatin structure and function, for example in the recruitment of epigenetic factors to their target loci as well as in the regulation of their enzymatic activity.

In this review I summarize the growing number of epigenetic modifying factors described to interact with lncRNAs in mouse and human and then discuss the challenges that lay ahead in understanding lncRNAs as part of the intricate networks of epigenetic regulation.

2. lncRNAs and histone modifying factors

2.1. Polycomb Group (PcG) repressive complexes have multiple RNA-binding subunits

The four core histones have a central structured region and an unstructured tail susceptible to post-translational modifications catalyzed by enzyme complexes collectively known as histone modifying complexes. Polycomb Group (PcG) proteins deposit histone marks

☆ This article is part of a Special Issue entitled: Clues to long noncoding RNA taxonomy, edited by Dr. Tetsuro Hirose and Dr. Shinichi Nakagawa.

E-mail address: juan.betancurmedina@riken.jp.

http://dx.doi.org/10.1016/j.bbagrm.2015.10.009
1874-9399/

Table 1
RNA-binding subunits of epigenetic modifying complexes and their RBRs.

Type of epigenetic modification	Complex	RNA-binding protein	RNA-binding region (aminoacid positions)	Binding specificity *in vitro*	Direct RNA binding *in vitro* / *in vivo*[2]	
Histone modification (histone methyl-transferase)	PRC2	EZH2	342–368	Promiscuous binding[1]. Affinity partly dependent on RNA length.	Yes / Likely	(24–26,28,29,38)
		SUZ12	Unknown	Promiscuous binding[1]. Affinity partly dependent on RNA length.	Yes / NA	(24,25,27)
		JARID2	332–358	Unknown.	Yes / Likely	(25,40,45)
Histone modification (histone ubiquitin-transferase)	PRC1	CBX4,5,7,8	Chromodomain	Apparently non-sequence specific.	Yes / NA	(49,50)
		SCML2	256–330	Apparently non-sequence specific Also binds dsDNA and nucleosomes.	Yes / ND	(52)
		PHC1	775–860	Apparently non-sequence specific.	Yes / NA	(48)
Histone modification (histone methyl-transferase)	MLL1	WDR5	Binding site defined by Y228, L240, K250, and F266	Unknown.	Yes / NA	(68)
	–	G9A	Unknown	Unknown.	NA / NA	(56,58)
Histone modification (histone de-methylase)	LSD1/coREST	LSD1	~174–382	Moderate sequence specificity for TERRA RNA motifs.	Yes / NA	(54)
DNA methylation	DNMT1	DNMT1	1081–1616	Apparently non-sequence specific. Structured RNAs (only ~22 nt long RNAs tested). Low affinity for dsDNA.	Yes / ND	(70)
Chromatin remodeling	BAF	BRG1	462–728 (binding to Evf2)	Apparently non-sequence specific.	Yes / NA	(75)
			774–1310 (binding to Mhrt)	Binds to chromatin and lncRNA through the same domain.	Yes / NA	(76)
		BAF155 / BAF170	Unknown	Unknown.	Yes / NA	(75)

[1] Assembled PRC2 shows promiscuous RNA binding as defined in (39,47).
[2] Interaction is considered direct *in vitro* when assessed with purified proteins by EMSA, UV cross-linking, or mutation/deletion of the RBR. Interaction is considered likely to be direct *in vivo* when assessed by CLIP-like methods (cross-linked RNA-containing complexes migrate in SDS-PAGE at a molecular weight similar to the molecular weight of the protein). RNA immunoprecipitation, where there is no evidence that binding does not occur through a co-precipitated factor, is not considered evidence of direct RNA binding. ND, not formally demonstrated. NA, not assessed.

that result in transcriptional repression of target genes. The two most representative members of this group are the Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2) which monoubiquitinate lysine 119 of histone H2A (H2AK119ub) and methylate lysine 27 of histone H3 (H3K27me), respectively. In mammals, PRC1 consists of a heterogeneous group of complexes that have RING1A or RING1B as their core and catalytic component. Canonical PRC1 also contains Polycomb Group RING finger (PCGF) 2/4, Polyhomeotic (PHC) 1/2/3, plus one chromobox protein (CBX) that recognizes and binds to H3K27me3 deposited by PRC2 [10–12]. This observation led to the idea that PcG recruitment is hierarchical, with PRC1 recruited downstream of PRC2 function, although this view was recently challenged. RING/YY1 binding protein (RYBP) or its homolog YAF2 competes with CBX proteins for binding to RING1A/B to form variant PRC1 together with any of the PCGF proteins (PCGF1–6) [13,14]. Variant PRC1 binds to target sites in the absence of H3K27me3 and its enzymatic product, H2AK119ub, engages PRC2 in a fashion reciprocal to recruitment of canonical PRC1 by PRC2 [15–17].

Compared to PRC1, PRC2 is more homogeneous. Its core components are the catalytic subunit EZH2, SUZ12 and EED. In addition, other proteins including RBAP46/48, JARID2, PCL1/2/3 and AEBP2, associate with the complex at sub-stoichiometric ratios [18]. In particular, AEBP2–JARID2-containing PRC2 interacts with H2AK119ub and this binding enhances the methyltransferase activity of PRC2 [16].

In addition to their interactions with histones, PcG proteins also bind RNA. The first description of this phenomenon stemmed from the finding that PcG proteins are required for X chromosome inactivation (XCI), a process in which one of the two X chromosomes in cells of female mammals is silenced to equalize gene expression from the X chromosome between males and females. X-inactive-specific transcript (Xist) is transcribed from the X inactivation center (XIC) of the inactive X (Xi) and coats the entire chromosome. Because recruitment of PcG proteins is dependent on the expression of Xist [19–22], and initial immuno-FISH experiments suggested strong colocalization of PcG proteins and the RNA [21,23], the prevailing model became that of PcG proteins being recruited by Xist to the Xi.

In the case of PRC2, evidence of direct RNA binding came from *in vitro* experiments that identified both EZH2 and SUZ12 as RNA-binding proteins (Table 1 and Fig. 1B) [24–29]. Both proteins as well as assembled PRC2 bind to Xist fragments derived from a region at the 5′ end of the transcript called repeat-A, which is required for chromosomal silencing [30]. However, some observations challenge the view that PRC2 recruitment involves direct PcG–Xist binding. Notably, various groups showed that an Xist transgene lacking repeat-A can also induce PRC2 recruitment, although with reduced efficiency [19,31], and a recent report showed spatial separation between Xist and PcG proteins, revealed by super resolution microscopy [32]. This review will focus on the RNA-binding roles of PRC2. However, this field is in active discussion and detailed reviews of the role of Xist during XCI, including alternative views, can be found elsewhere [33,34]. No evidence of direct binding of PRC1 to Xist has been reported.

Besides its role in XCI, PRC2 has also been shown to interact with hundreds of other lncRNAs *in vivo* [29], notably Hotair, which is transcribed from the *HoxC* locus and acts in *trans* to repress transcription of genes in the *HoxD* locus [35] as well as hundreds of other genes genome wide [36,37].

In contrast to the roles of PRC2 that involve contacts with specific RNAs (*e.g.*, with Xist during XCI and with Hotair to regulate *Hox* gene expression), *in vivo* PRC2 also binds nascent transcripts genome wide [38–40]. This interaction together with binding of other subunits of the complex to modified histones [41–43] allows the complex to sense the transcriptional and surrounding epigenetic state of the loci where it binds. One model of the RNA-mediated recruitment and activity of PRC2 proposes that EZH2 binding to RNA and the presence of active epigenetic marks render PRC2 catalytically inactive [25,40,41], while in the absence of transcription or in the presence of H3K27me3 the complex spreads and maintains H3K27me3 at target loci [38,39,42,44].

Underscoring the important role of RNA in the function of PRC2, an additional protein that associates with the complex, JARID2, was also shown to bind to RNAs (Table 1 and Fig. 1B) [25,45,46]. Because JARID2 binding to PRC2 is facilitated by RNAs [45] and because JARID2 can bind to chromatin independent of the core PRC2 complex [46] it is

Please cite this article as: J.G. Betancur, Pervasive lncRNA binding by epigenetic modifying complexes — The challenges ahead, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbagrm.2015.10.009