Cell Tissue Res (2014) 356:507–526
DOI 10.1007/s00441-014-1885-x

MINI REVIEW

Biological and bioinformatical approaches to study crosstalk of long-non-coding RNAs and chromatin-modifying proteins

Rolf Backofen · Tanja Vogel

Received: 16 February 2014 / Accepted: 27 March 2014 / Published online: 13 May 2014

Abstract Long-non-coding RNA (lncRNA) regulates gene expression through transcriptional and epigenetic regulation as well as alternative splicing in the nucleus. In addition, regulation is achieved at the levels of mRNA translation, storage and degradation in the cytoplasm. During recent years, several studies have described the interaction of lncRNAs with enzymes that confer so-called epigenetic modifications, such as DNA methylation, histone modifications and chromatin structure or remodelling. LncRNA interaction with chromatin-modifying enzymes (CME) is an emerging field that confers another layer of complexity in transcriptional regulation. Given that CME–lncRNA interactions have been identified in many biological processes, ranging from development to disease, comprehensive understanding of underlying mechanisms is important to inspire basic and translational research in the future. In this review, we highlight recent findings to extend our understanding about the functional interdependencies between lncRNAs and CMEs that activate or repress gene expression. We focus on recent highlights of molecular and functional roles for CME–lncRNAs and provide an interdisciplinary overview of recent technical and methodological developments that have improved biological and bioinformatical approaches for detection and functional studies of CME–lncRNA interaction.

Keywords Chromatin-modifying enzymes · Histone methylation · lncRNA · Histone acetylation · PCG TRXG

R. Backofen
Institute of Computer Science, Albert-Ludwigs-University, Freiburg, Germany

T. Vogel (✉)
Institute of Anatomy and Cell Biology, Albert-Ludwigs-University, Freiburg, Germany
e-mail: tanja.vogel@anat.uni-freiburg.de

General remarks

Long-non-coding RNA (lncRNA) can regulate gene expression at many levels. These include transcriptional and epigenetic regulation as well as alternative splicing in the nucleus, but also regulation of mRNA translation, storage and degradation in the cytoplasm. Several different mechanisms have been elucidated (Mercer and Mattick 2013; Rinn and Chang 2012; Wang and Chang 2011): LncRNA production can serve as a signal, and the process of its transcription already has a regulatory effect, alongside further regulatory roles of the transcribed lncRNA. In this regulatory setting, lncRNA can serve as a scaffold to assemble protein complexes. They also have the property to function as decoys by titrating proteins, such as transcription factors, away from the chromatin. This function as decoys has also been attributed to miRNA (Hansen et al. 2013). Finally, lncRNA can act as a guide to localise chromatin-modifying enzymes (CME) to a specific target region. Besides the regulatory roles affecting transcription, there is a substantial portion of lncRNAs that is localised in the cytoplasm. In this location, they exert post-transcriptional regulatory effects (Mercer and Mattick 2013; Yoon et al. 2013). As one example, lncRNA can influence alternative splicing as has been shown for the lncRNA *Malat1*, which binds to SR-proteins and thereby acts as a decoy for this class of splicing factors. LncRNAs also influence mRNA decay by stabilising or destabilising mRNA. In these cases, the regulatory effect is mediated via RNA:RNA interaction. One example is the *Bace1-antisense* transcript that forms a duplex with its sense *Bace1*-counterpart (Faghihi et al. 2008). Through this base pairing, the latter is stabilised by protection against RNase cleavage. Partial base-pairing between the lncRNAs *1/2-sbsRNAs* with Alu-repeat-containing mRNAs recruits the STAUFEN1 protein and thus promotes the decay of these transcripts (Gong and Maquat 2011). LncRNA binding to its direct target RNA also leads to

translational activation or inhibition, as e.g. in the case of *lncRNA-p21*, which inhibits the translation of *JunB* and *Ctnnb* mRNAs (Yoon et al. 2012). Another mechanism of action is observed, e.g. during X-inactivation. Here, the complementary lncRNAs *Xist* and *Tsix* bind to each other to form a duplex, which is recognised and processed into small RNAs through Dicer. These small RNAs are involved in the repression of *Xist* transcription (Ogawa et al. 2008). A similar method of action has been described for *borderline* lncRNAs, that are produced from the boundary between eu- and heterochromatin (Keller et al. 2013).

Several lncRNAs and small RNAs are implicated in transcriptional control and exert this function in close functional interaction with CME. As data on small RNAs have been recently reviewed (Li 2013), they will not be a major topic of this overview. Here, we will focus on lncRNA–CME interactions. The functional interaction between lncRNAs and CMEs can be two-fold: On the one hand, transcription of lncRNAs itself can be influenced through epigenetic modifications at their respective promoters and/or enhancers (Wang et al. 2011). On the other hand, CME exert RNA-binding activity or act in complexes with partners that are able to recruit lncRNAs. It is of note that several lncRNAs have the ability to interact with a variety of different CME, in overlapping or distinct regions. Thus, lncRNAs are able to regulate a chromatin state by bringing different partners together. This implies that lncRNAs are involved in coordinating chromatin modification at specific loci. Loci specificity is mediated by sequence specific features such as the sequence itself but also through the secondary structure of the DNA.

As CME–lncRNA interaction is an emerging field with several new and important findings acquired in recent years, we will highlight these recent findings to extend our understanding about the functional interdependencies. In addition to the review of recent highlights of molecular and functional roles for CME–lncRNAs, we will also provide recent technical and methodological developments that have improved laboratory approaches for detection and functional study of CME–lncRNA interaction.

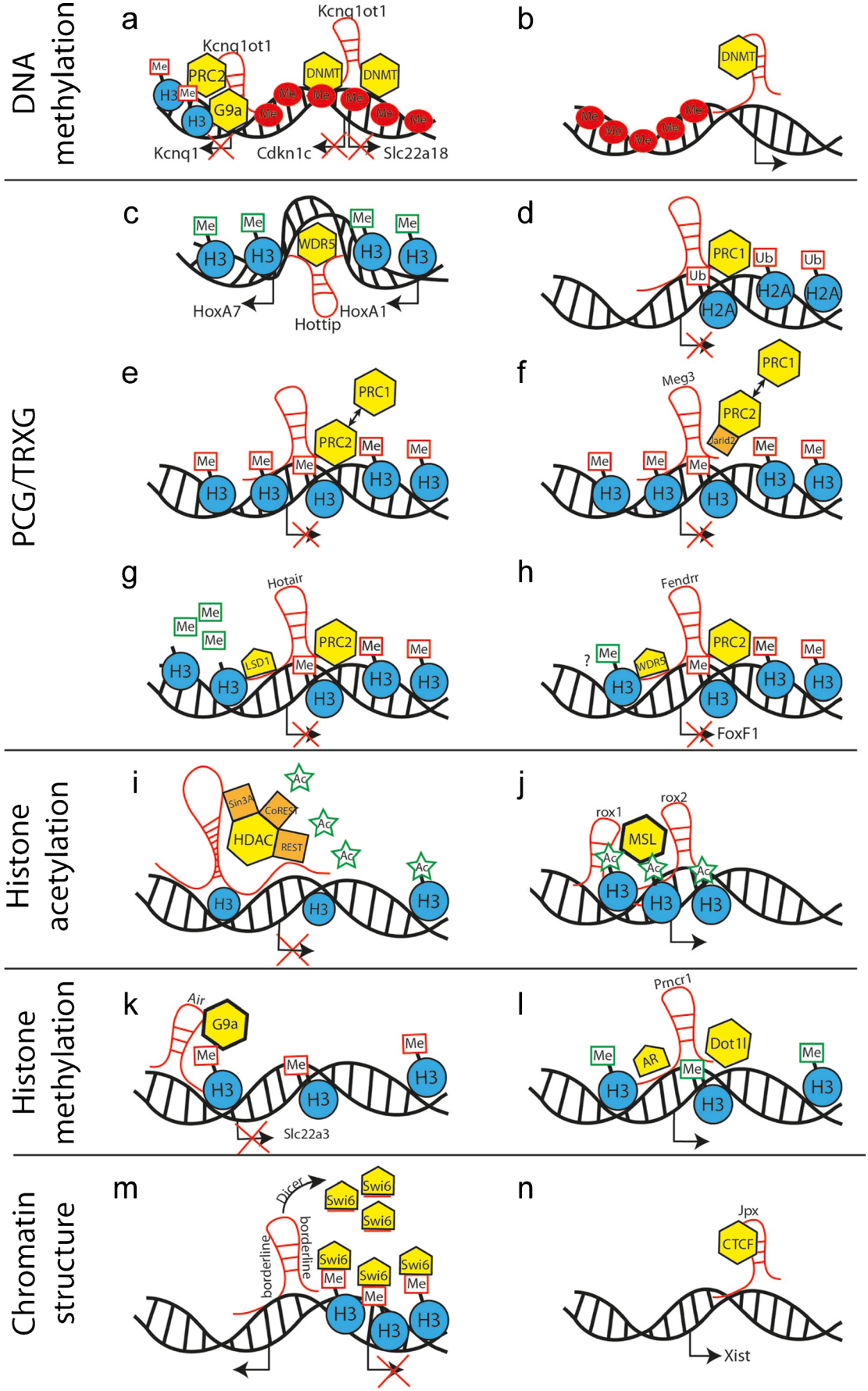
LncRNAs affecting DNA methylation

The expression of ncRNAs, such as siRNAs and miRNAs, can be under control of DNA methylation itself, and this is specifically true for lncRNAs (Panning and Jaenisch 1996; Warnecke et al. 1998). There are several examples of siRNAs and miRNAs regulating transcription by interaction with DNA methyltransferases (DNMTs). Thereby, siRNAs and miRNAs bind and target DNMT3A and 3B to specific genomic loci (Benetti et al. 2008; Braconi et al. 2010; Chavali et al. 2012; Fabbri et al. 2007; Ji et al. 2013). Data on lncRNAs that bind to DNMTs are limited but will probably increase substantially in the near future. Among the few examples so far identified are the non-coding sense–antisense transcripts *Xist* and *Tsix* that control mammalian X-inactivation. *Tsix* has been shown to specifically bind to DMNT3A, but not to the other DMNTs. This interaction mediates de novo methylation at the *Xist* promoter, sustaining repression of *Xist* and an active X chromosome (Sun et al. 2006). Another example is the lncRNA *Kcnq1ot1* that has been shown to directly interact with DNMT1 (Mohammad et al. 2010). This interaction is required for the maintenance of DNA methylation at imprinted gene loci, such as *Cdkn1c* and *Slc22a18*. DNA methylation at these loci is lost upon conditional deletion of *Kcnq1ot1* (Mohammad et al. 2012).

In addition to DNMT1, the *Kcnq1ot1* lncRNA interacts with the euchromatic histone lysine N-methyltransferase 2 (EHMT2/G9a), Polycomb repression complex (PRC) 1 and PRC2 to silence the imprinted *Kcnq1* transcriptional unit (Fig. 1a) (Mohammad et al. 2010). The last example shows that the functional network of a specific lncRNA can comprise a diversity of different CMEs. This implies that multi-level approaches are necessary to study the specific functions of an lncRNA, which can be rather diverse.

Whereas *Kcnq1ot1*-and *Tsix*-DNMT interaction mediates DNA methylation and transcriptional repression, lncRNA–

Fig. 1 Examples of lncRNA–CME interactions. **a**, **b** LncRNA and DNA ▶ methylation. **a** *Kcnq1ot1* directly interacts with DNMT1 and maintains DNA methylation at promoters of *Cdkn1c* and *Slc22a18. Kcnq1qt1* also interacts with EHMT2/G9a, PRC1 and PRC2 to silence the imprinted *Kcnq1* gene. **b** lncRNAs act as a scaffold for DNMTs whereby they prevent direct recruitment of DNMTs and subsequent DNA methylation, resulting in transcriptional activation. **c**–**h** LncRNA and PCG or TRXG interaction. **c** *Hottip* maintains the activity of the *Hoxa* genes through binding to TRXG-complex member WDR5, which mediates activating H3K4me. **d** PRC1 mediates transcriptional repressive ubiquitination of H2A via the RING proteins, of which RING1B interacts with lncRNAs. **e** PRC2 core components bind to lncRNA and mediate repressive H3K27me. They can recruit PRC1 complexes for additional H2Aub but can also act independently. **f** The lncRNA *Meg3* targets the repressive PRC2 to specific loci in a complex with JARID2. **g** *Hotair* binds PRC2 and LSD1 in distinct binding domains and thereby mediates H3K27me3 alongside the H3K4 demethylation to silence transcription. **h** *Fendrr* lncRNA binds to PRC2 complexes and TRXG/MLL–complex member WDR5. Whether it activates transcription is unclear but it represses FoxF1 transcription through H3K27me. **i**, **j** LncRNA and histone acetylation. **i** REST, CoREST, SIN3A and associated HDACs bind lncRNAs, e.g. after focal ischemia, and mediate transcriptional silencing. **j** MSL complex requires presence of rox1 and rox2 for enhanced transcription through H4K16 acetylation of the single male X chromosome in Drosophila. **k**, **l** LncRNA and histone methylation. **k** *Air* binds the *Slc22a3* promoter, recruits EHMt2/G9a and mediates repressive H3K9me2 and me3. **l** *Prncr1* recruits AR and transcriptional activator DOT1l that mediates H3K79me. **m**, **n** LncRNA and chromatin remodelling, structure and expression boundaries **m** *Borderline RNAs* are processed by Dicer1 into small RNAs that bind SWI6 and prevent spreading of H3K9–methylated heterochromatin. **n** *Jpx* binds to CTCF and thereby titrates CTCF away from the *Xist* promoter. It thereby prevents silencing of *Xist*


DNA methylation
PCG/TRXG
Histone acetylation
Histone methylation
Chromatin structure
a
Kcnq1ot1
Kcnq1ot1
PRC2
G9a
DNMT
DNMT
Me
H3
Kcnq1
Cdkn1c
Slc22a18
b
DNMT
c
WDR5
HoxA7
Hottip
HoxA1
d
PRC1
Ub
H2A
e
PRC1
PRC2
f
Meg3
PRC1
PRC2
Jarid2
g
Hotair
LSD1
PRC2
h
Fendrr
?
WDR5
PRC2
FoxF1
i
Sin3A
CoREST
HDAC
REST
Ac
j
rox1
rox2
MSL
k
Air
G9a
Slc22a3
l
Prncr1
AR
Dot1l
m
Dicer
Swi6
borderline
n
Jpx
CTCF
Xist

DNMT interaction can also result in loss of DNA methylation and transcriptional activation. The *Cebpa* genomic locus produces diverse transcripts of which a long, non-polyadenylated, nuclear RNA, *ecCebpa*, also binds DNMT1 and interferes with its methylation activity at the *Cebpa* locus (Di Ruscio et al. 2013). Further analyses using RNA-immunoprecipitation (RIP) followed by massive parallel sequencing revealed more than 6,000 potential RNA–DNMT1 interactions. The data do not point to a specifically recognised sequence but indicate that DNMT1 binding is influenced by RNA secondary structure. The observations support a model in which the RNA acts as a scaffold for DNMT1 whereby it prevents direct recruitment of DNMT1 and subsequent DNA methylation (Fig. 1b). The data also provide evidence that lncRNA binding to DNA methylating enzymes is not restricted to a small subset but is observed for a large number of lncRNAs (Di Ruscio et al. 2013). Thus, lncRNAs will be important components to establish this critical epigenetic hallmark.

Not only DNMTs are associated with lncRNA. Several proteins that are involved in binding of methylated DNA also have RNA-binding ability. Methyl-CpG-binding domain (MBD) 2 and methyl-CpG-binding protein 2 (MECP2) are able to bind methylated DNA as well as RNA (Jeffery and Nakielny 2004). Only a few of the associated RNAs, however, have been identified so far. Known lncRNAs that associate with MECP2 are, e.g. *Malat1* and *Rncr3* (Maxwell et al. 2013), but, so far, the functional meaning of this interaction is elusive. As reviewed in more detail below, MBD1 binds the *H19* lncRNA in the context of imprinted genomic loci. This complex is able to recruit enzymes conferring repressive histone marks to the associated chromatin and to interfere with transcriptional activation (Monnier et al. 2013). It might be possible that MBD2 and MECP2 act in a similar manner; however, this hypothesis has to be addressed in future experiments.

LncRNA, PCG and TRXG interaction

The Polycomb group (PCG) of proteins are important regulators of transcription and they influence a variety of cellular processes. They were discovered as essential factors implicated in spatial regulation of *Hox* gene transcription during developmental body axis segmentation. In this context, PCG proteins act as antagonists to Trithorax (TRXG/MLL) proteins that activate *Hox* transcription through the histone H3 methylation activity on lysine 4 (H3K4me), mediated by distinct SET-domain proteins within the TRXG-complexes.

Gene silencing through PCG complexes is a well-described process. A large portion of data describes the influence of lncRNAs on PCG function. PCG proteins are widely expressed and it is important to understand how regulation at specific loci is achieved, i.e. how specific genes are selected as PCG targets whereas others are unaffected. While the genome of *Drosophila melanogaster* contains specific DNA-recognition sites that are bound by PCG complexes (reviewed in Ringrose and Paro 2007), these targeting sequences are missing in mammals. Further, core components of PRC2 do not contain direct DNA-binding domains and targeting to chromatin will be achieved through interaction with cofactors. Among these are proteins as well as chromatin-associated lncRNAs.

Hox genes are assembled in clusters. They are transcribed in colinear patterns between their genomic position and the expression domain along the body axis, meaning in an orchestrated sequence from anterior to posterior and proximal to distal. The *Hoxa* genomic locus (with *Hoxa1* located at the 3′ end and *Hoxa13* at the 5′ end) is flanked by two lncRNAs, *Hottip* at the 5′ end and *Hotairm1* at the 3′ end of the cluster (Wang et al. 2011). *Hottip* is required to maintain the activitiy of the *Hoxa* genes located at the 5′ end through binding to TRXG-complexes. This has been shown through RIP by using an antibody against TRXG-complex member WDR5, which precipitated *Hottip*. Thus, *Hottip* recruits H3K4me to the active *Hoxa* genes at the 5′ end. Its potency declines with distance from its own expression site, e.g. *Hoxa13*, is much more influenced than *Hoxa7*. This long-distance effect over multiple *Hoxa* genes (A13–A7) located at the 5′ end, is achieved through chromosomal looping. Thereby, *Hottip* comes into the vicinity of more distant genes such as *Hoxa7*, although *Hottip* itself is not implicated in mediating the 3D-chromatin conformation (Fig. 1c).

Hottip–WDR5 interaction is one of the few known examples of transcriptional activation through TRXG–lncRNA interaction. The majority of data about CME–lncRNA association comes from the counteracting PCG family.

PCG complexes are grossly separated into two main distinct complexes, namely PRC1 and PRC2. PRC1 complexes are diverse in their composition. Core components of PRC1 are CBX, PCGF, RING, and PHC proteins. Each of these proteins comprises a family of a multitude of different members. Accordingly, several data revealed that PRC1 exists in distinct sub-complexes (Gao et al. 2012; Morey et al. 2013; Vogel et al. 2006). PRC2 cores are less diverse and contain EZH2, EED, and SUZ12. In a simplified view, PRC1 and 2 have two different functions: whereas PRC2 mediates histone H3 methylation at lysine 27 methylation (H3K27me), PRC1 binds to this histone modification via the chromodomain of CBX proteins and mediates ubiquitination of histone H2A through the ligase activity of RING1B. Recent data, however, indicate another way of coupling PRC2 and PRC1 activity, as the PRC2 core component EED binds PRC1 proteins RING1A and B, and BMI1 (Cao et al. 2014). EED is thus involved in recruiting PRC1 to PRC2-occupied chromatin independent from H3K27 methylation.

In addition to having a joint function, both complexes can also act independently from each other (Schoeftner et al. 2006; Tavares et al. 2012; Wu et al. 2013b), indicating that the crosstalk between PRC1 and PRC2 will be much more sophisticated than described so far.

PRC1 and PRC2 are both associated with lncRNAs (Guttman et al. 2011). Known protein components that interact with lncRNAs are RING1B within PRC1 (Fig. 1d) (Guttman et al. 2011; Schoeftner et al. 2006) and EHZ2, EED and SUZ12 within PRC2 (Fig. 1e) (e.g. Khalil et al. 2009; Rapicavoli et al. 2011; Wu et al. 2013a). Whether other components that constitute PRC1 bind lncRNAs as well is, to the best of our knowledge, not yet reported. One example of a well-studied lncRNA is *Xist* which is necessary to initiate mammalian X-inactivation during early development. *Xist* expression leads to H3K27me3, monomethylation of histone H4 at lysine 20 (H4K20me1) and ubiquitination of histone H2A at lysine 119 (H2Aub) through recruitment of PRC1 and PRC2. This *Xist*-mediated silencing characterises the initiation of X-inactivation, while other epigenetic mechanisms are involved in the maintenance of the inactive state. These include DNA methylation and hypoacetylation of histone H4. Although PRC2-mediated H3K27 methylation might serve as a platform for PRC1 recruitment, *Xist* colocalises with the PRC1 complex member RING1B. This indicates a PRC2-independent mechanism for maintenance (Schoeftner et al. 2006), but not for initiation (Leeb and Wutz 2007) of the inactive state.

PRC2 components SUZ12 (Kanhere et al. 2010) and EHZ2 (Zhao et al. 2010) have the potential to directly associate with *Xist*, whereas we are not aware of comparable data about RING1B–*Xist* interaction. However, the evidence that *Xist* is directly bound by PCG proteins is under debate (Brockdorff 2011). The reason is that these interaction data were generated by RIP (see below), which determines both direct and indirect targets. Robust methods for determining direct targets using UV crosslinking as an additional step (see below) were yet not employed in studies of PRC2–*Xist* interaction. Thus, it is still possible that the interaction between *Xist* and PCG complexes is indirect and mediated by PCG-associated cofactors. An indirect association is supported by recent data showing that the lncRNA *Meg3* is implicated in targeting the PRC2 to genes that confer differentiation and need to be silenced in pluripotent ESCs (Kaneko et al. 2013). Here, the recruitment to chromatin depends on the presence of JARID2, which bridges PRC2–lncRNA interaction (Fig. 1f). Thus, specificity of PCG silencing at specific loci can be mediated through cell- or stage-specific presence of either cofactors and/or lncRNAs. This modular way of using lncRNAs to tether PRC2 towards chromatin, directly or indirectly, gives a multitude of possibilities for cell-type and sequence-specific gene silencing through the PCG complexes.

Several other recent data report on further lncRNA–PCG interactions. The lncRNA *Ancr* (anti-differentiation ncRNA) interferes with osteoblast differentiation by associating with EZH2 and thereby establishing H3K27me at the *Runx2* promoter (Zhu and Xu 2013). Up-regulated in bladder cancer 1 (*linc-UBC1*) is immunoprecipitated through EHZ2 and SUZ12 antibodies. This lncRNA is upregulated in bladder cancer and correlates with poor survival of patients as it promotes cancer cell proliferation, motility and metastasis (He et al. 2013). Loss of RAS-association domain family member 1A (RASSF1A) expression is a hallmark of different tumours. Its silencing is dependent on its lncRNA counterpart *Anrassf1*, that forms an lncRNA–DNA hybrid. This sequence-specific structure recruits SUZ12 that mediates H3K27me (Beckedorff et al. 2013) and leads to increased cancer cell proliferation. Again, mostly RIP led to the observation of lncRNA-PRC2 association and much more research is needed to reveal direct binding and/or involvement of other bridging factors.

Another prominent PRC-associated lncRNA is *Hotair* that has various interacting partners, which confer repressive chromatin modifications. Among these are the PRC2 and the complex of Lysine-specific demethylase 1 (LSD1), repressor element-1 silencing transcription factor (REST) and REST corepressor 1 (CoREST) (Tsai et al. 2010). Interestingly, the binding domains for PRC2 on one side and LSD1/REST/CoREST on the other are distinct from each other. Thus, *Hotair* can serve as coordinator of transcriptional repression by recruiting enzymes that mediate H3K27me3 alongside the LSD1-associated H3K4 demethylation activity (Fig. 1g). So far unexplored are the underlying kinetics, the question of sequential binding and, if that is the case, in which sequence it happens. Such physical parameters might influence target site selection and binding, and might also be subject to cell-type-specific variations. *Hotair* is bound by an EZH2-EED heterodimer. The responsible binding motif within *Hotair* is strikingly different to the A-repeat that is recognised in the *Xist* transcript (Wu et al. 2013a). This finding implies that association of PRC proteins with lncRNAs is probably diverse, mediated by differently composed complexes, association with cofactors and/or different binding domains.

Genome-wide association studies identified the lncRNA *Anril* (antisense Non-coding RNA in the INK4 Locus) within a genomic region that confers susceptibility for e.g. cardiovascular diseases, atherosclerosis, several cancers, glaucoma, Alzheimer's disease and diabetes. In *cis*, *Anril* regulates expression of the tumour suppressors *Cdkn2a/b (*Ink4b/Arf/Ink4a). *Anril* has multiple splice variants and also produces circular molecules. However, only a few of these molecules have been assigned with specific functions (reviewed in Congrains et al. 2013). *Anril* exerts transcriptional control through binding to PCG proteins from both complexes, PRC1 and PRC2 (Kotake et al. 2011; Yap et al. 2010). Whereas PRC1 association is mediated through CBX7,

PRC2 binding was shown for SUZ12. *Anril* binding of CBX7 is competitive with the association of CBX7 with H3K27me3 (Yap et al. 2010). This indicates that *Anril* might act as a coordinator of PRC1 function: it influences the extent of PRC1 that is associated with chromatin to mediate sustained gene repression via H2Aub. Further coordinative function might be associated with the binding of both PRCs. Of note in this context is that *Anril*–SUZ12 interaction affects *Cdkn2b* transcription. And contrasting this observation, *Anril* knockdown in CBX7-dependent transcriptional control interfered only with *Cdkn2a* expression, but not with *Cdkn2b*. The molecular reasons for these observations are so far unknown. One could hypothesise that it might be associated with the different cell systems that have been used in the studies, but the answer of this question is still open.

Another recent study also revealed *Anril*-mediated transcriptional regulation in *trans* via PRC complexes in the context of atherosclerosis (Holdt et al. 2013). The comprehensive analysis of PCG proteins revealed *Anril* precipitation through antibodies against CBX7, RING1B, RYBP, YY1, and the PRC2 components EED, SUZ12, JARID2, and RBAP46. *Anril* target genes in *trans* contain Alu-sequences within their promoter region that are recognised by an Alu-motive within *Anril,* indicating how *Anril*, and subsequently PCG complexes, are recruited to regulatory sequences in *trans*.

The *Fendrr* lncRNA is involved in embryonic development of the lateral plate mesoderm and can bind to PRC2 complexes to mediate H3K27me at specific promoters such as *Foxf1* in *cis*, and *Pitx2* in *trans* (Grote et al. 2013). Interestingly, *Fendrr* also binds WDR5, one core component of PCG-antagonists assembled in TRXG/MLL complexes (Fig. 1h). Thus, in principle, *Fendrr* can mediate transcriptional repression as well as activation. It might also be involved in balancing access of PCG or TRXG proteins to a specific locus, by mediating recruitment of PCG proteins and sequestering TRXG proteins. The finding that loss of *Fendrr* results in less H3K27me at specific loci, whereas differences in H3K4me have so far not been observed, would support this hypothesis. Although a lot more research is needed to understand the function of these bivalent lncRNAs, these initial findings indicate the coordinating potential of lncRNAs with regard to developmental processes that are under PCG and TRXG control.

LncRNAs influencing histone modifications

Histone acetylation

Trichostatin A (TSA) interferes with histone deacetylation and alongside relocalises splicing factors within the nuclear compartment. TSA induces association of lncRNAs with specific splicing factors. Interestingly, lncRNAs are the only RNA-subtype that showed splicing factor association after TSA treatment, with *Malat1* and *Neat1* being the most abundant (Schor et al. 2012). The recruitment of splicing factors after TSA treatment is a strong indication that chromatin modification on histones induces not only the expression of lncRNAs but also influences their binding to splicing factors and recruitment to splice sites. Whether this process solely depends on histone deacetylation or whether the splicing factors themselves are modified through acetylation is not yet resolved. It is possible that TSA treatment resulted in acetylation of proteins other than histones. Such proteins could be splicing factors and/or cofactors implicated in lncRNA recruitment.

Alongside this research, a recently discovered lncRNA, *lnc-Jade*, indirectly influences H4 acetylation (Wan et al. 2013). It associates with the transcription factor breast cancer 1 (BRCA1) and the histone acetyltransferase (HAT) complex of E1A binding protein p300/CREB binding protein (p300/CBP) at the *Jade1* promoter. Thereby it activates *Jade1* transcription. JADE1 itself is a cofactor in the human acetylase binding to ORC1 (HBO1) complex that mediates histone H4 acetylation at positions K5, K8, and K12. Accordingly, overexpression of *lnc-Jade* results in increased genome-wide histone H4 acetylation. It is of note that *lnc-Jade* expression itself is embedded within the signalling pathways that are active in DNA damage response. This is because Ataxia-telangiectasia mutated (ATM) and nuclear factor 'kappa-light-chain-enhancer' of activated B-cells (NF-κB) regulate the expression of *lnc-Jade* (Wan et al. 2013).

Hypoxia and focal ischemia induce DNA damage, and expression of lncRNAs is altered significantly in these conditions. Apart from general changes of the expression of lncRNAs, selective lncRNA–protein interactions are induced (Dharap et al. 2013). REST corepressors CoREST and SIN3A showed increased binding to lncRNAs in rat brains after focal ischemia (Fig. 1i). Amongst a plethora of interacting proteins of CoREST and SIN3A are proteins implicated in chromatin modifications such as histone deacetylases (HDACs), MECP2, retinoblastoma binding protein 4 (Rbbp4), and several members of the SWItch/Sucrose NonFermentable (SWI/SNF) family. Identification of target sequences of specific lncRNAs as well as further analyses of the functional meaning of CoREST/SIN3A interaction with specific lncRNAs will increase our knowledge about implication of lncRNA interaction with chromatin modifiers in normal and diseased brain.

One recently reported example showed that the transcription of *lncRNA–Let* (low expression in tumour) is under specific control of HDAC3. In hypoxic cancer cells, HDAC3 is activated by the hypoxia-induced factor (HIF1α) and promotes histone H3 and H4 deacetylation at the *lncRNA–Let* promoter. This results in decreased expression of *lncRNA–Let. LncRNA–Let* itself suppresses expression of HIF1α and is therefore implicated in a positive feedback loop to increase HIF1α response under hypoxic conditions. The association of

misregulation within this molecular network and hypoxia-induced metastasis (Yang et al. 2013a) is of clinical importance.

Early X-inactivation at the blastocyst stage is sensitive to HDAC inhibition as this treatment decreased *Xist* expression (Oliveira et al. 2013). Thus, *Xist* expression is under control of acetylated histones. Histone H4 hypoacetylation and H3/H4 hypomethylation are implicated in this process (Pullirsch et al. 2010). However, mechanistic insight into whether and how these repressive marks are integrated into direct *Xist* function awaits further reports.

In contrast to mammals in which X chromosome dosage compensation is achieved through inactivation of one allele, dosage compensation in *Drosophila* is mediated through enhanced transcription of the single male chromosome. This is achieved by the male-specific lethal (MSL) protein complex, which is composed of five different proteins, MSL1–3, MOF, and MLE, and it additionally contains the two lncRNAs, RNA on the X1 and 2 (*roX1* and *2*). This complex mediates histone H4 acetylation at lysine 16 (H4K16) and thereby promotes transcription. *Rox1* and *2* are required for efficient dosage compensation and provide a binding platform for MSL complex assembly (Fig. 1j). Different binding affinities to specific regions of the lncRNAs might be implicated in creation of a flexible hub through which the CMEs are recruited to low as well as high affinity binding sites dispersed along the X chromosome (Ilik et al. 2013).

Histone methylation

Apart from *Xist* and *Kcnq1ot1*, the *Air* lncRNA is involved in silencing within an imprinted region in *cis*. Among the silenced genes within this region are *Slc22a3* and *Igf2r* that are putatively targeted by *Air* using alternative molecular mechanisms (Fig. 1k). *Air* directly covers the *Slc22a3* promoter region and recruits the histone H3 methyltransferase EHMT2/G9a that mediates repressive H3K9 di- and tri-methylation on the paternal allele (Nagano et al. 2008). Neither *Air* nor EHMT2/G9a are present on the *Igf2r* promoter, although the latter is included within the *Air*-silenced DNA region. It remains to be determined whether the presence at a distant promoter creates an environment sufficient to silence over a hundred kilobases, or whether a different molecular mechanism is used at the *Igf2r*.

Although a member of a protein family that can bind to methylated DNA, MBD1 also binds to unmethylated DNA and to the *H19* lncRNA (Monnier et al. 2013). *H19* is an imprinted lncRNA and its expression is regulated through the imprinting control region (ICR). The ICR is a differentially methylated region. Interaction between *H19* and MBD1 was observed in mouse embryonic fibroblasts (MEFs). Although *H19* binds to EZH2 in bladder cancer cells (Luo et al. 2013), this was not observed in MEFs. Thus, temporal and cell-specific variability in the association of a specific lncRNA with different CME is another layer of complexity within transcriptional regulation. MBD1 not only binds to DNA but also interacts with histone H3 lysine 9 (H3K9) methyltransferases, e.g. KMT1A/SUV39H1 and KMT1E/SETDB1 (Fujita et al. 2003; Sarraf and Stancheva 2004). Thus, through interaction with *H19* the repressive H3K9 methylation is recruited to the imprinted region, e.g. into a chromatin region that is already marked through differential DNA methylation. This observation might indicate a fine-tuning in transcriptional regulation for which lncRNAs mediate the specificity of the genomic locus. Another interpretation is that lncRNAs function as mediators to link DNA methylation to histone methylation.

The human prostate cancer-associated non-coding RNA1 (*Prncr1*) binds to several proteins, among which are the histones H3 monomethylated at position lysine 4 (H3K4me1) and H4K16ac. Both modifications are localised at active enhancers. Further interacting proteins are the disruptor of telomeric silencer 1-like (DOT1L) and the androgen receptor (AR). DOT1L mediates the mainly transcriptional enhancing histone H3 di-and trimethylation at position 79 (H3K79me2/3). This network of *Prncr1*-interacting molecules is thought to mediate AR-dependent transcription. Via the interaction with modified histones at specific enhancer regions, *Prncr1* recruits AR and transcriptional activator DOT1L to AR-regulated target sites (Fig. 1l) (Yang et al. 2013c).

A novel study from *Plasmodium falciparum* might indicate further correlation of histone methylation and lncRNA-modulated gene transcription (Jiang et al. 2013). Here, the authors report the expression of erythrocyte membrane protein 1 that is encoded by multiple *var* genes. *Var* genes are expressed one at a time but how this timely expression is controlled is still under investigation. Active transcription of *var* genes is associated with expression of a corresponding lncRNA and low states of histone H3 trimethylation at position 36 (H3K36me3). In this context, it is unlikely that the responsible CME, PfSETvs, directly binds to the lncRNA, as the enzyme is enriched at the promoter when the *var* gene and the corresponding lncRNA are inactive. It is tempting to speculate that active transcription of both lncRNA and *var* gene might depend on other CMEs. These could be targeted to the locus via the lncRNA and prevent recruitment of silencing PfSETvs.

Recent data from *S. cerevisea* indicate an indirect link between histone H3K9 methylation and lncRNAs (Marina et al. 2013). The authors identified the Seb1-protein that binds to lncRNAs transcribed from the pericentromeric region. Seb1 also associates with the SHREC protein complex, which has HDAC activity. Loss of histone tail acetylation may promote H3K9 methylation and stable silencing of the pericentromere.

However, the molecular mechanism upon which histone methylation follows loss of the acetylation mark has not been clarified in detail so far.

LncRNAs involved in chromatin remodelling, chromatin structure and expression boundaries

LncRNAs have the ability to guide proteins to specific genomic loci. One interesting example has been observed in plants, where lncRNAs seem to generate a scaffold on which protein complexes assemble. Such complexes establish silencing marks such as DNA methylation and repressive histone marks (Haag and Pikaard 2011). Components of the chromatin-remodelling SWI/SNF 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 (Hargreaves and Crabtree 2011). Protein–lncRNA complexes that associate with SWI/SNF remodelers seem to interfere with RNA polymerase II (RNAPolII) recruitment (Zhu et al. 2013). The molecular mechanism, however, that comprehensively describes the effect on RNAPolII transcription awaits elucidation. Further, it is not yet clear whether SWI/SNF–lncRNA association does solely result in transcriptional repression or whether it could also positively influence transcription.

Mediator is a large multi-protein complex that integrates and interprets various signals that influence transcription. It can be considered as a gigantic hub that is used to infer specific activating modifications, e.g. affecting the transcriptional pre-initiation complex (Malik and Roeder 2010). The Mediator complex exerts this function also through a plethora of diverse interacting molecules, among which are CMEs such as p300, GCN5 or G5a. Through its kinase module, Mediator has specific activity towards the phosphorylation of histone H3 at serine 10 (H3S10p). H3S10p is generally associated with transcriptional activation. Yet, we still lack precise knowledge how Mediator–CME interaction translates mechanistically to gene transcription. However, Mediator plays a direct role in chromatin looping. It is thus important for interpretation of chromatin structure and to mediate this over a long distance into a transcriptional signal at a specific promoter region. Recent data show that chromatin looping associated with H3S10p through Mediator is tightly coupled to the presence of activating lncRNAs (*lncRNA-a*), namely *lncRNA-a3* and *-a7* (Lai et al. 2013). Depletion of these lncRNAs results in loss of higher order chromatin structure and decreased levels of H3S10p at specific target gene loci, such as snail homolog 1 (*Snai1*), Aurora kinase A (*Aurka*) and T cell acute lymphocytic leukemia 1 (*Tal1*). This indicates a crucial function of lncRNAs in organising higher order chromatin structure and in bridging chromatin modifying enzymatic activities in *cis* over larger distances.

Heterochromatin and euchromatin are distinct in their histone modifications as the former has high levels of H3K9 methylation. This serves as a binding platform of HP1 protein that spreads over silenced regions. In yeast, like *S. pombe*, small interfering RNAs (siRNAs) mediate centromeric heterochromatin formation. These siRNAs are bound by Argonaut 1 (AGO1) and guide the RNA-induced transcriptional silencing (RITS) complex to the pericentromeric region. The RITS complex recruits the H3K9 methyltransferase Clr4 that places the heterochromatic mark. SWI6 is the HP1 homologue of yeast and binds methylated H3K9. By this, heterochromatin spreads across some genomic distances. Recently, it has been shown that lncRNAs are also involved in this process. They prevent further spreading of heterochromatin (Keller et al. 2013). In contrast to centromeric siRNAs, distinct lncRNAs are expressed from the transition region between hetero- to euchromatin and are therefore called *borderline RNAs*. Although Dicer1 also processes these lncRNAs into small RNAs, these do not load onto AGO1 but instead bind to SWI6. These processed lnRNAs might therefore be implicated in preventing SWI6-mediated spreading of H3K9-methylated heterochromatin (Fig. 1m).

Crosstalk of lncRNA and chromatin is an important mechanism to define boundaries of region-wide transcriptional activation or repression (Wang et al. 2011), which is e.g. observed in imprinted regions (Mohammad et al. 2012). One mechanism is the presence of CTCF and its binding sites in flanking regions. The zinc finger protein CTCF binds DNA and has pleiotropic functions such as transcriptional regulation, insulation, X-inactivation and imprinting (for review, see Phillips and Corces 2009). In addition to DNA binding, recent data also demonstrate RNA-binding properties of CTCF that are involved in X-inactivation. In this context, CTCF binds the lncRNA *Jpx* and this event competes with CTCF binding to the P2 *Xist* promoter (Sun et al. 2013c). *Jpx*-mediated eviction of CTCF leads to *Xist* expression alongside X-inactivation (Fig. 1n). This example shows that lncRNAs are important influencers of the chromatin structure with their capability to recruit or evict proteins with chromatin-organising capacity.

CTCF binding is not the only molecular mechanism through which e.g. genomic boundaries can be defined. Recent data suggest that enhancer specific histone methylations, e.g. histone H3 acetylation at lysine 27 (H3K27ac) and H3K4me1 prevent lncRNA-mediated gene silencing. So far, this insulator function of modified histones has been shown for *Xist*- and *Kcnq1ot1*-dependent silencing (Mohammad et al. 2012), but much more research is needed to elucidate this process in detail. This finding indicates that, on the one hand, lncRNAs function to organise the chromatin. On the

other hand, a specific chromatin structure is also a prerequisite to regulate lncRNA functions.

Taken together, recent data emphasise the important role of lncRNA–CME interactions for transcriptional regulation in different cell types, during development and disease (Tables 1, 2). Thereby, lncRNA contribute to all major epigenetic modifications and 3D–organisation of the chromatin. Although data are increasing, our review indicates that much more research is needed to comprehensively describe molecular mechanisms and precise modes of interactions. Within the following, we will provide an overview about current techniques and methods to decipher lncRNA function.

Table 1 Summary of CME–lncRNA associations as discussed in detail in this review

	CME	*lncRNA*	Chromatin mark	Effect on transcription	Reference
DNA methylation	DMNT3a	*Tsix*	DNAme	Repressive	Sun et al. 2006
	DMNT1	*Kcnq1ot1*	DNAme	Repressive	Mohammad et al. 2010
		ecCebpa		Activating	Di Ruscio et al. 2013
	MECP2	*Malat1*		n.a.	Maxwell et al. 2013
		Rncr3		n.a.	Maxwell et al. 2013
	MBD1	*H19*		Repressive	Monnier et al. 2013
TRX	WDR5	*Hottip*	H3K4me	Activating	Wang et al. 2011
		Fendrr		n.a.	Grote et al. 2013
PCG/PRC1	RING1B	*Anril*	H2Aub	Repressive	Holdt et al. 2013
		Xist		Repressive	Schoeftner et al. 2006
	CBX7	*Anril*	H2Aub	Repressive	Yap et al., Holdt et al. 2013
PCG/PRC2	EED	*Anril*	H3K37me	Repressive	Holdt et al. 2013
		Hotair		Repressive	Wu et al. 2013a
	EZH2	*Kncq1ot1*	H3K37me	Repressive	Mohammad et al. 2010
		Xist		Repressive	Zhao et al. 2010
		Ancr		Repressive	Zhu et al. 2013
		linc–UBC1		Repressive	He et al. 2013
		Hotair		Repressive	Wu et al. 2013a
		Fendrr		Repressive	Grote et al. 2013
		H19		Repressive	Luo et al. 2013
	SUZ12	*Anril*	H3K37me	Repressive	Kotake et al. 2011, Holdt et al. 2013
		Xist		Repressive	Kanhere et al. 2010
		linc–UBC1		Repressive	He et al. 2013
		Anrassf1		Repressive	Beckedorff et al. 2013
		Fendrr		Repressive	Grote et al. 2013
PcG associated proteins	RYBP	*Anril*		n.a.	Holdt et al. 2013
	YY1	*Anril*		n.a.	Holdt et al. 2013
	JARID2	*Anril*	H3K27me	Repressive	Holdt et al. 2013
		Meg3		Repressive	Kaneko et al. 2013
	RBAP46	*Anril*		n.a.	Holdt et al. 2013
Histone acetylation	p300/CBP	*lnc–Jade*		Activating	Wan et al. 2013
	MSL	*Rox1*	H4K16ac	Activating	Ilik et al. 2013
		Rox2		Activating	Ilik et al. 2013
Histone methylation	EHMT2/G9a	*Kcnq1ot1*	H3K9me	Repressive	Mohammad et al. 2010
		Air		Repressive	Nagano et al. 2008
	LSD1	*Hotair*	H3K4me	Repressive	Tsai et al. 2010
	DOT1L	*Prncr1*	H3K79me	Activating	Yang et al. 2013a, b, c
Chromatin structure/remodelling	SWI6	*borderline*		Activating	Keller et al. 2013
	CTCF	*Jpx*		Activating	Sun et al. 2013c
	Mediator	*lncRNA–a*		Activating	Lai et al. 2013

Table 2 Summary of CME–lncRNA associations extracted from the lncRNA database (http://www.lncrnadb.org), but beyond the scope of this review

	CME	*lncRNA*	Chromatin mark	Effect on transcription	Reference
DNA methylation	MECP2	*Evf–2, Dlx6as1*	methylated DNA binding	Activating	Bond et al. 2009
PCG	PRC1	*adapt33*	H2Aub	Repressive	Guttman et al. 2011
		Hotair			Guttman et al. 2011
		linc1257		Repressive	Guttman et al. 2011
		linc1610			Guttman et al. 2011
		Neat			Guttman et al. 2011
		Rian			Guttman et al. 2011
		snhg3			Guttman et al. 2011
	PRC2	*adapt33*	H3K27me	Repressive	Guttman et al. 2011
		antiPeg11			Zhao et al. 2010
		Coldair			Heo and Sung 2011
		gtl2–as, Meg3–as			Zhao et al. 2010
		linc1257		Repressive	Guttman et al. 2011
		linc1610			Guttman et al. 2011
		malat1–as			Guttman et al. 2011
		Meg3			Zhao et al. 2010
		Neat			Guttman et al. 2011
		nespas			Zhao et al. 2010
		Rian			Zhao et al. 2010, Guttman et al. 2011
		snhg3			Guttman et al. 2011
		Tug1			Khalil et al. 2009
	CBX3	*linc1547*	H3K9me binding		Guttman et al. 2011
		Rian			Guttman et al. 2011
	CBX1	*adapt33*	H3K9me binding	Repressive	Guttman et al. 2011
		Hotair			Guttman et al. 2011
		linc1257		Repressive	Guttman et al. 2011
Histone methylation	ESET/SETDB1	*adapt33*	H3K9me3	Repressive	Guttman et al. 2011
		linc1609			Guttman et al. 2011
		Neat			Guttman et al. 2011
	SUV39H1	*adapt33*	H3K9me3	Repressive	Guttman et al. 2011
		linc1257		Repressive	Guttman et al. 2011
		linc1547			Guttman et al. 2011
		Neat			Guttman et al. 2011
		snhg3			Guttman et al. 2011
	KDM5B/JARID1B	*adapt33*	H3K4me	Repressive	Guttman et al. 2011
		linc1257		Repressive	Guttman et al. 2011
		Neat			Guttman et al. 2011
		Rian			Guttman et al. 2011
		snhg3			Guttman et al. 2011
	KDM5C/JARID1C	*linc1609*	H3K4me		Guttman et al. 2011
		linc1610			Guttman et al. 2011
		Rian			Guttman et al. 2011
	SETD8	*adapt33*	H4K20me 1	Repressive	Guttman et al. 2011
		linc1257		Repressive	Guttman et al. 2011
	KMT2A/MLL1	*Mistral*	H3K4me		Bertani et al. 2011
chromatin structure	CTCF	*linocr*			Lefevre et al. 2008

Table 2 (continued)

	CME	*lncRNA*	Chromatin mark	Effect on transcription	Reference
unclassified	n.a.	*CAR Intergenic 10*	H3K4me2	Activating	Mondal et al. 2010

Bioinformatic approaches for the analyses of ncRNAs and their interactions

Detection of lncRNAs

The first step in the analysis of lncRNAs is to detect their respective genes in the genome. While lncRNAs were first detected by tiling arrays, RNA-seq is now a huge source to discover new and to investigate known lncRNAs (see Ilott and Ponting 2013; Garber et al. 2011 for recent reviews). The main problem of the analyses is to reconstruct the transcripts from the short sequencing reads. Thereby, a non-trivial process to assemble the mapped reads into longer transcripts follows mapping of the short reads to the genome. Alternative splicing and incomplete coverage of the whole length of the transcript are the main obstacles. In many applications, however, it is possible to rely on standard approaches for solving this problem such as Scripture (Guttman et al. 2010) and Cufflinks (Trapnell et al. 2010).

LncRNA genes can also be detected by the analysis of chromatin modifications (see, e.g., Hon et al. 2009 for a review). This was exemplified by Guttman et al. (2009), who used chromatin maps for H3K4me3 and H3K36me3, which mark promoters and transcribed regions, respectively (Mikkelsen et al. 2007). Subsequently, H3K4me3 and H3K36me3 enriched regions were selected for further investigation. The criterion was a size of at least 5 kb without overlap with genes that were coding for proteins or known miRNAs.

Common to both, the RNA–seq as well as the chromatin map based approaches, is that identified regions might still code for an unknown protein and thus do not resemble a ncRNA. There are different possibilities to exclude this. First, it is possible to use sequence–based statistical measurements for the coding potential such as the codon substitution frequency, e.g. the CSF (Clamp et al. 2007) or the CNCI method (Sun et al. 2013a). RNAcode is an approach that additionally considers information about evolutionary conservation (Washietl et al. 2011). A very elegant recent method is to use additional experimental data to evaluate coding potential such as ribosome occupancy that is revealed through ribosome profiling (Guttman et al. 2013).

A common problem in the detection of lncRNAs is that the precise transcript boundaries are mostly unknown. The main reason for imprecise transcript definition, which includes fragmentation and inexact transcript boundaries, is the low expression level of lncRNAs (Derrien et al. 2012). Partially defined transcripts do not pose a major problem for association studies as discussed below, since the approaches used in these studies often rely only on the quantification of a partial transcript. However, imprecise transcripts can hinder the further functional annotation of ncRNAs.

LncRNA associations: in situ hybridisation and co-expression analysis

The next important step in the analysis of identified lncRNAs is to associate these RNAs with pathways, cell types, proteins or diseases. Direct association with specific cell types or even with specific proteins is often determined by in situ hybridisation (ISH) (see e.g. Dinger et al. 2008; Mercer et al. 2008). As shown recently this is not restricted to fresh tissues but can be employed on fixed samples as well (Chisholm et al. 2012). In this study, ISH approaches were used to study the association of three *Hox*-locus lncRNAs (*Hotair*, *ncHoxA1*, and *ncHoxD4*) with the expression of the PRC2 member EZH2 in different formalin-fixed paraffin-embedded breast cancer tissues. Using a 283-lesion tissue microarray, this study revealed a significant correlation (in 40–70 % of the lesions) between the expression of these lncRNAs and EZH2. Furthermore, they investigated the correlation of lncRNA expression with pathologic parameters such as metastasis, estrogen receptor status and others. While it was known that *Hotair* expression correlates with metastasis and death (Gupta et al. 2010), the association with the other pathological parameters mentioned above was unknown.

ISH is an excellent tool to investigate the association between selected lncRNAs and proteins. There is, however, the need for a systematic association analysis that can be used for large numbers of lncRNAs simultaneously. A recent and also commonly used approach is *guilt by association* (Cabili et al. 2011; Dinger et al. 2008; Guttman et al. 2009; Ramos et al. 2013). *Guilt by association* detects co-regulated genes or gene groups based on expression profiles for different tissues, conditions or time points. In more detail, the method presented in Guttman et al. (2009) combines the information about coexpression with a *gene set enrichment analysis* (GSEA) (Subramanian et al. 2005) to determine sets of genes that are associated with given lncRNAs. GSEA is based on the definition of biological relevant gene sets. For each of these given genes represented in set "S", GSEA determines whether the genes in "S" specifically correlate with a certain phenotype. For that purpose, a Pearson correlation coefficient between the

lncRNA-associated genes with the phenotype is calculated, and genes are sorted according to the correlation coefficient. Each gene in "S" is marked in a sorted list. This list is used to calculate an enrichment score ES(S), which is a running sum starting from the highest correlating gene. Whenever a gene is in "S", the score is increased by the correlation coefficient; otherwise, the coefficient decreases (see Fig. 2). The maximum value of this score is called the enrichment score ES(S) for the gene set "S". This step is followed by a statistical evaluation of ES(S) to determine significant associations. In the *guilt by association* approach, correlation to the phenotype was replaced by the correlation to the expression of a single lncRNA, thus determining functional gene sets associated with this lncRNA in expression (Guttman et al. 2009).

Another method that relies on coexpression networks (Liao et al. 2011a) assigns putative functions for 340 lncRNAs based on combinations of network properties (such as module sharing and association with hubs) and genomic adjacency. Typical functions associated are tissue development, cellular transport and metabolic processes. The method is also available via a web service (Liao et al. 2011b). A more recent approach combines links generated by coexpression of lncRNAs and proteins with protein–protein interactions into a single bi-colored network (Ji et al. 2013). On this common network, putative functions are associated by a method that can be described as 'function flow'. Examine the toy example in Fig. 2, where the lncRNA inherits the functional annotation of group A (proliferation). In the "function flow" approach, this would be done by propagating a "small portion" of the functional annotation for each gene in group A via the coexpression links. The annotation of group A (and not of B) would accumulate at the lncRNA as it is linked to many genes in the group A but not in B. Protein–protein interactions make this network denser and thus improve the efficiency of this method for function propagation. The *NONCODE* database (Xie et al. 2014) uses the latter two methods (Ji et al. 2013; Liao et al. 2011a) to assign potential functions especially to lncRNAs. *NONCODE* currently contains roughly 210,000 entries corresponding to lncRNAs, which were retrieved from Ensemble, Refseq or curated literature.

Concerning disease associations, the *lncRNAdisease* database currently lists 480 manually curated disease–lncRNA associations, corresponding to 166 diseases. For most of the lncRNAs (478 out of 480), information of the associated binding partner is collected (Chen et al. 2013). The authors also developed an in silico approach to detect new putative lncRNA–disease associations. This is based on the observation that miRNAs in close genomic vicinity are often associated with similar diseases. By carrying over this approach to lncRNAs, this study identified 33 new putative disease–lncRNA associations.

Besides studying general associations of lncRNA to proteins, more direct experimental approaches are available when considering transcriptional regulation of lncRNAs by transcription factors. Here, the current method of choice is to determine direct DNA-binding sites of transcription factors via CHIP-seq experiments. Since CHIP-seq experiments have an inherently high false positive rate due to unspecific binding, a good approach is to combine CHIP-seq experiments with in silico promoter analysis. As an example, Idogawa et al. (2014) used CHIP-seq experiments to identify roughly 42000 peaks for the protein p53 all over the genome. On this initial peak set, they searched for known binding motifs for p53.

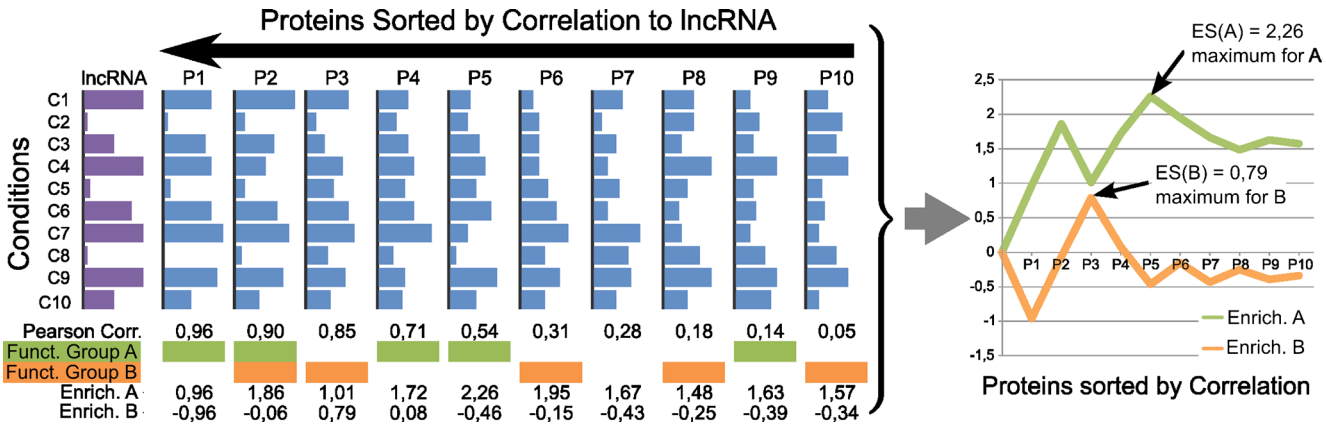


Fig. 2 Toy example for the *guilt by association* approach. Given a specific lncRNA, we have expression data for ten conditions (tissue, time point, etc.), covering the expression levels for one lncRNA and ten proteins. The expression profile across all conditions is used to calculate Pearson correlation coefficients between all proteins and the lncRNA. The proteins are sorted according to the correlation to the lncRNAs. Furthermore, we have two functional gene lists: A, consisting of P1, P2, P4, P5 and P9, and B, consisting of P2, P3, P6, P8 and P10. The proteins of group A occur more to the left, i.e. they have a higher tendency to correlate with the lncRNA expression. This is expressed by the enrichment, which is a running sum of correlation coefficients. Starting from the protein with the highest correlation, the coefficients are added if they are contained in the functional group, and subtracted from the running sum if not. The diagram to the right displays the rows denoted by *Enrich. A/B*. The ES score for a group is the maximal value for the enrichment. In this toy example, it is 2.26 for A, and 0.79 for B. Thus, proteins of A are more likely to be associated with the lncRNA, and the lncRNA would inherit the function associated with group A

Additional filtering of the remaining peaks according to their genomic position (upstream region of lncRNAs, vicinity to TSS etc.) reduced this set finally to 857 lncRNAs putatively regulated by p53. From this set, the author could experimentally validate 23 lncRNAs to be upregulated by p53. This example shows the importance of a thorough bioinformatics analysis of CHIP-seq experiments. CHIPbase combines the information of 543 CHIP-seq experiments with predictions of transcription factor binding sites (TFBS) and further in silico analysis (Yang et al. 2013b). Another tool in this direction is PscanChip (Zambelli et al. 2013).

LncRNA annotation

Once a list of lncRNAs associated with a protein complex, tissue, cell condition or function has been determined, an important step is to annotate these lncRNAs. For the structured lncRNAs or lncRNAs containing conserved structural elements, one possible annotation tool is RFAM (RNA Families Database; Burge et al. 2013). RFAM is the largest database of ncRNA families and classes, which are grouped together according to sequence and structure properties. Families are defined by evolutionary relationship, whereas RNA classes comprise ncRNAs that share functional similarities without being homologous. Well-known examples of RNA classes are miRNAs or snoRNAs. Although RFAM lists 2208 families, it contains only a few entries for long non-coding RNAs.

When searching for the classification "*lncRNA*", RFAM lists only 225 families, which often cover only well-conserved structural elements of these RNAs. When disambiguating different exons and regions of the same lncRNA, then only 88 lncRNA families are left.

For that reason, there is a need for additional classifiers that detect common sequence and structure properties for a set of lncRNAs grouped by biological functions. Examples for such groups would be either the entity of lncRNAs bound by PCG or all lncRNAs co-expressed with certain chromatin modifiers. This has been exemplified in Glazko et al. (2012), where sequence-structure features were used in combination with a support-vector-machine (SVM) to classify lncRNAs into PRC2-bound and non-PCR2-bound RNAs. As features, they used 397 so-called RSSPs (RNA sequence-structure pattern) to capture the ncRNA structure. These RSSPs were initially collected in the Structator approach (Meyer et al. 2011). RSSPs are RFAM models of highly structured families, and the scores for all these models are thus a good descriptor of the sequence-structure properties of an ncRNA. Concerning sequential features, three types are considered. One type consists of binding sites of transcription factors taken from the Jaspar database (Mathelier et al. 2014). The rationale of choosing this feature is that many ncRNAs are transcribed from promoter and enhancer regions and thus should be enriched in these motifs. The second type considers short oligonucleotide (k-mer) frequencies. Here, overrepresented k-mers might represent binding sites of RNA-binding proteins, and the optimal length derived by the approach (namely k=6) is compatible with the usual length expected for these motifs. Finally, as an additional feature, they used the repeat structure of an ncRNA to distinguish PRC2-bound from PRC2-non-bound ncRNAs. The repetitiveness of an ncRNA can be measured by the Kolmogorrow Complexity of the ncRNA-sequence.

The approach described above uses clustering according to features derived from sequence and predicted structure. For structured ncRNAs, this approach has already been introduced (Torarinsson et al. 2007; Will et al. 2007) and has developed into a standard to detect new classes of structured ncRNAs for various biological topics (see, e.g., Heyne et al. 2012; Lange et al. 2013; Parker et al. 2011; Tseng et al. 2009). For lncRNAs, it is not clear whether the majority are structured or un-structured, which limits the applicability of the sequence-structure based approaches. There is, however, another source of information that can be used to cluster similar RNAs, which stems from family-specific processing patterns and their traces in RNA-seq data (Findeiss et al. 2011). This has already been used for miRNAs, where e.g. the miRDeep approach successfully identifies new miRNA candidates by searching for the characteristic read profile covering the mature miRNA and its complement miRNA* (see e.g. Gan et al. 2008). General tools for clustering and annotating ncRNAs are DARIO (Fasold et al. 2011), DeepBlockAlign (Langenberger et al. 2012), CoRAL (Leung et al. 2013) or ALPS (Erhard and Zimmer 2010). Although these approaches are currently mainly used for small ncRNAs, it is conceivable that we will see improved classification of lncRNAs based on the combination of different features including the processing pattern.

LncRNA interactions: RIP, CLIP-seq, CHIRP-seq and CLASH-seq

The previously described association and annotation studies do not provide mechanistic explanations for lncRNA associations with other molecules nor do they differentiate direct from indirect interactions. For that purpose, one needs approaches to directly analyse physical interactions. Basically, we have three major classes of interactions: RNA–protein, RNA–RNA and RNA–DNA.

RNA–protein interactions are determined in vitro by the SELEX approach (Systematic Evolution of Ligands by Exponential Enrichment; see Stoltenburg et al. 2007 for a review), usually resulting in a small set of strong binders. However, great success has been made recently in determining RNA–protein interactions in vivo on a genome-wide level based on microarrays or high-throughput sequencing. Most of the existing experimental techniques use an antibody against

the protein of interest to enrich RNAs bound to this protein. The enrichment step is either affinity-based, or requires the formation of a covalent bond using a cross-linking step before the purification. Furthermore, reducing the length of the bound RNA molecule by sonification or by RNase treatment is common to all protocols.

RIP (Gilbert and Svejstrup 2006) is an affinity-based method that detects direct or indirect interaction of transcripts with a given protein. The sequences that are bound by the protein are determined by different methods, namely PCR, microarray (RIP-CHIP) (Tenenbaum et al. 2000), or high-throughput sequencing (RIP-seq) (Zhao et al. 2010). The advantage of RIP-seq (and to a lesser extent also RIP-chip) is a relatively good quantification of the overall effect of an RNA-binding protein on a transcript. Both RIP-CHIP and RIP-seq have been used to determine the lncRNAs associated with PCG proteins (Khalil et al. 2009; Zhao et al. 2010) or with the epigenetic activator Mixed lineage leukemia 1 (MLL1) (Bertani et al. 2011).

The main disadvantage of the RIP-based approach is that it cannot differentiate direct from indirect interactions. The latter would comprise cases where an RNA molecule is not directly bound by the protein of interest. Instead, a whole protein complex that contains the protein under investigation binds it. Another disadvantage is that the resulting read length is too large for determining the actual binding site. Both problems are solved when using an additional cross-linking step as done in the CLIP approach (Ule et al. 2003). There are currently four different CLIP-seq protocols that combine CLIP with high-throughput sequencing, termed HITS-CLIP (Licatalosi et al. 2008), CRAC (Granneman et al. 2009), PAR-CLIP (Hafner et al. 2010) and iCLIP (Konig et al. 2010) (see Milek et al. 2012 for an excellent review of the details). The main idea is to cross-link the RNA to the protein in vivo, thus avoiding additional interactions that may happen after cell lysis. Even more importantly, the cross-linking step allows using more stringent purification protocols. Furthermore, again due to the cross-linking, an RNase treatment allows the reduction of the average length of the enriched sequences to about 50 nucleotides, which improves the determination of the actual binding site.

After applying quality controls and mapping the reads to the genome, the bioinformatics analysis of RIP-seq and CLIP-seq data usually starts with the reduction of the false positive rate, which results again from unspecific interactions. For that purpose, so-called peak callers have to be used. Because of the similarity to the CHIP-seq protocol one could imagine using standard CHIP-seq peak callers like MACS (Zhang et al. 2008) also for RIP-seq data. However, the bimodal distribution of reads due to DNA double-strands is usually not observed in RNA-protein interaction data. For that purpose, specialized peak callers like piranha (Uren et al. 2012) or PARalyzer (Corcoran et al. 2011) have to be used.

There is, however, a significant yet underestimated problem of missing binding sites (i.e. false negatives), especially when using published CLIP-seq data. The reason is very simple. The number of reads recovered by a CLIP-seq experiment greatly depends on the expression level for the associated transcript. This implies that a CLIP-seq experiment cannot be used to detect interactions for low or non-expressed genes. Even for high expressed genes, some binding sites might be missing due to limited mapability of the short reads in low entropy regions of the genome. It also implies that one needs in theory a separate CLIP-seq experiment for each tissue or condition.

For that reason, it is of common interest to determine binding motifs from existing CLIP-seq data. There are a multitude of tools to discover sequence motifs for transcription-factor binding sites, with popular examples being MEME and DREME (Bailey et al. 2009), MatrixREDUCE (Foat et al. 2006) and DRIMust (Leibovich et al. 2013). Although these approaches are often used to analyse RNA–protein interaction data from RIP-seq or CLIP-seq experiments (see, e.g., Sanford et al. 2009; Gupta et al. 2013; Wang et al. 2014), it has become clear that one needs to consider the secondary structure of the RNA in addition to its sequence (Hiller et al. 2007; Kazan et al. 2010). Structural features were first introduced in the process of binding motif detection by BioBayesNet (Nikolajewa et al. 2007) for transcription factors and by MEMERIS (Hiller et al. 2006) for RNA-binding proteins. The vast amount of data, however, provided by CLIP-seq and RIP-seq experiments required the development of new approaches. Currently, the two best performing tools are RNAcontext and GraphProt. RNAcontext (Kazan et al. 2010) combines a standard biophysical model for sequence affinity with a model for the structural context. GraphProt (Maticzka et al. 2014) uses an advanced machine-learning model based on so-called graph-kernels, which allow an efficient representation of complex graphs encoding the sequence and structure properties of binding sites. On a cross-validation test on 24 CLIP-seq datasets, GraphProt outperformed RNAcontext as well as a sequence-based method (MatrixReduce), which in addition showed the greatest variation in quality. For some proteins, MatrixReduce performed comparably to the two structure-based approaches, but completely failed for others. This indicates that different proteins have different structural preferences. In addition, it could be shown that GraphProt can effectively discover missing binding sites.

For RNA–RNA interactions and RNA–DNA interactions, only a few experimental high-throughput approaches are available. RNA–DNA interactions can be detected with the CHIRP-seq method. CHIRP-seq (Chu et al. 2011) is again based on a cross-link between the RNP containing the RNA and its target DNA. After sonification, cross-linked target DNA is enriched via a purification step based on biotin-

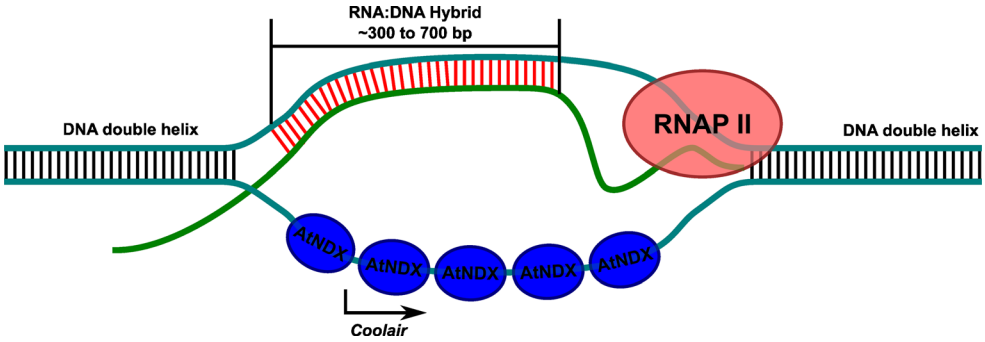


Fig. 3 R-loop formation repressing the expression of the *Coolair* lncRNA in *Arabidopis* (adopted from Sun et al. 2013b). The R-loop formation consists of an RNA:DNA hybrid within an open DNA bubble. R-loop can occur when nascent RNA invades the DNA double helix. In the case of *Coolair*, data indicates that the protein AtNDX binds to the opposite DNA single strand, thus stabilizing the R-loop formation. This in turn results in a repression of the *Coolair* lncRNA

labelled oligonucleotide probes for the RNA of interest. Streptavidin with magnetic beads is then used to enrich the RNA-bound oligonucleotides. Since these are cross-linked with the DNA, this also leads to an enrichment of the bound DNA. CHIRP-seq was developed by the Chang laboratory to generate a genome-wide map of the DNA interactions of three lncRNAs, namely the *Drosophila melanogaster rox2* RNA, the human telomerase RNA *Terc* and the *Hotair* lncRNA. Bioinformatics analysis of binding sites was again performed using sequence motif detection tools, in this case, MEME. One has to mention that, for this kind of data, it is more complicated to include structural information, which implies that sequence-based motif detection tools might currently be the best choice. The reason is that it is necessary to model the DNA–RNA interaction for including structural information. There are, however, two completely different structural models known for DNA:RNA hybrids. An R-loop is formed when RNA hybridises with a single-stranded region of the DNA. A current model is that R-loops may occur in regions of transcription-induced supercoiling, where it could repress gene expression (Aguilera and Garcia-Muse 2012; Drolet 2006; Drolet et al. 2003; Huertas and Aguilera 2003; Sun et al. 2013b). An example is displayed in Fig. 3. While the prediction of such formations might partially be possible using specialised mixed DNA–RNA parameters in a standard RNA-folding approach (Lorenz et al. 2012), this is much harder for the second known class, namely triple-helix formations. It was shown that promoter-associated RNAs form a DNA:DNA:RNA triple-helix at the binding site of DNA methyltransferase DNMT3b, suggesting that this mechanism may be more common in epigenetic regulation (Schmitz et al. 2010). However, there are currently no accepted structural models for triple helices that could be used to predict such binding sites. Triplexator (Buske et al. 2012), which is currently the only computational tool to predict triple-helix formation genome-wide, is henceforth based purely on sequence information. For that reason, the specificity is probably not high enough in most application scenarios. However, the situation might change since there are now two techniques for analyses available that determine R-loop formation on a genome-wide level, namely DRIP-seq and DRIVE-seq (Ginno et al. 2012).

For the last type of interaction, namely RNA–RNA (see Fig. 4), there have been very successful approaches to predicting these interactions for small RNAs, i.e. miRNAs in the eukaryotic system (reviewed in, e.g., Pasquinelli 2012)

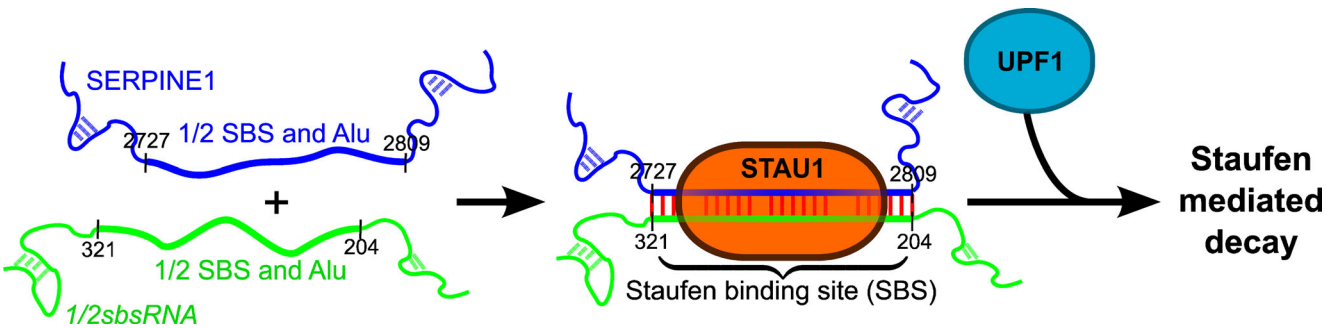


Fig. 4 Example of post-transcriptional regulation by an lncRNA–mRNA interaction. STAUFEN 1 binds to double-stranded RNA at so-called Staufen binding site (SBS). The lncRNA *1/2sbsRNA* contains a half Staufen binding site, which also happens to be an Alu-element. SERPINE 1 contains a partially complementary Alu-element, which is also a half SBS. When these two sites together form a duplex by imperfect base pairing, they form a full SBS and recruit STAU1. This leads, in an UPF1-dependent way, to Staufen-mediated decay (Gong and Maquat 2011)

and small RNAs in bacteria (reviewed in Backofen and Hess 2010). Popular miRNA target prediction tools are PITA (Kertesz et al. 2007), Pictar (Lall et al. 2006), MiRanda (Enright et al. 2003) and TargetScan (Friedman et al. 2009). For bacterial sRNAs, these are IntaRNA (Busch et al. 2008), TargetRNA (Tjaden et al. 2006), CopraRNA (Wright et al. 2013) and RNApredator (Eggenhofer et al. 2011). Due to the length of lncRNAs and incomplete structural models, the specificity of these prediction approaches would be much too low for the reliable determination of lncRNA targets. Variants of a recent experimental high-throughput method termed CLASH-seq (Helwak et al. 2013) might help to identify RNA targets of lncRNA. The CLASH-seq protocol is a variant of CLIP-seq that was used to determine miRNA targets more precisely. Here, after immunoprecipitation of the cross-linked AGO complex, the miRNA and its targets are ligated. After sequencing, the ligated interacting hybrids are directly detected as chimeric sequencing reads. It remains to be elucidated whether RNA–RNA interaction prediction tools like IntaRNA and RNApredator can be used to improve the analysis of CLASH-seq data, similar to the use of TF binding site prediction tools in the analysis of CHIP-seq data. Here, the tools were successfully used to pin down the actual binding site in CHIP-seq data, as described above in section "LncRNA associations: in situ hybridisation and co-expression analysis".

References

Aguilera A, Garcia-Muse T (2012) R loops: from transcription byproducts to threats to genome stability. Mol Cell 46:115–124

Backofen R, Hess WR (2010) Computational prediction of sRNAs and their targets in bacteria. RNA Biol 7:33–42

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208

Beckedorff FC, Ayupe AC, Crocci-Souza R, Amaral MS, Nakaya HI, Soltys DT, Menck CFM, Reis EM, Verjovski-Almeida S (2013) The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. PLoS Genet 9:e1003705

Benetti R, Gonzalo S, Jaco I, Muñoz P, Gonzalez S, Schoeftner S, Murchison E, Andl T, Chen T, Klatt P, Li E, Serrano M, Millar S, Hannon G, Blasco MA (2008) A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nature structural &#38. Mol Biol 15:268–279

Bertani S, Sauer S, Bolotin E, Sauer F (2011) The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. Mol Cell 43: 1040–1046

Bond AM, Vangompel MJW, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. Nat Neurosci 12:1020–1027

Braconi C, Huang N, Patel T (2010) MicroRNA-dependent regulation of DNA methyltransferase-1 and tumor suppressor gene expression by interleukin-6 in human malignant cholangiocytes. Hepatology (Baltimore, Md) 51:881–890

Brockdorff N (2011) Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. Development 138: 5057–5065

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226–D232

Busch A, Richter AS, Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24:2849–2856

Buske FA, Bauer DC, Mattick JS, Bailey TL (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res 22:1372–1381

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25:1915–1927

Cao Q, Wang X, Zhao M, Yang R, Malik R, Qiao Y, Poliakov A, Yocum AK, Li Y, Chen W, Cao X, Jiang X, Dahiya A, Harris C, Feng FY, Kalantry S, Qin ZS, Dhanasekaran SM, Chinnaiyan AM (2014) The central role of EED in the orchestration of polycomb group complexes. Nat Commun 5:3127

Chavali V, Tyagi SC, Mishra PK (2012) MicroRNA-133a regulates DNA methylation in diabetic cardiomyocytes. Biochem Biophys Res Commun 425:668–672

Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res 41:D983–D986

Chisholm KM, Wan Y, Li R, Montgomery KD, Chang HY, West RB (2012) Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. PLoS ONE 7:e47998

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. Mol Cell 44:667–678

Clamp M, Fry B, Kamal M, Xie X, Cuff J, Lin MF, Kellis M, Lindblad-Toh K, Lander ES (2007) Distinguishing protein-coding and non-coding genes in the human genome. Proc Natl Acad Sci USA 104: 19428–19433

Congrains A, Kamide K, Ohishi M, Rakugi H (2013) ANRIL: Molecular Mechanisms and Implications in Human Health. Int J Mol Sci 14: 1278–1292

Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22:1775–1789

Dharap A, Pokrzywa C, Vemuganti R (2013) Increased binding of stroke-induced long non-coding RNAs to the transcriptional corepressors Sin3A and coREST. ASN Neuro 5:283–289

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, Figueroa ME, De Figueiredo Pontes LL, Alberich-Jorda M, Zhang P, Wu M, D'Alò F, Melnick A, Leone G, Ebralidze KK, Pradhan S, Rinn JL, Tenen DG (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. Nature 1–10

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS (2008) Long

noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 18:1433–1445
Drolet M (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol Microbiol 59:723–730
Drolet M, Broccoli S, Rallu F, Hraiky C, Fortin C, Masse E, Baaklini I (2003) The problem of hypernegative supercoiling and R-loop formation in transcription. Front Biosci J Virtual Libr 8:d210–d221
Eggenhofer F, Tafer H, Stadler PF, Hofacker IL (2011) RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res 39:W149–W154
Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS (2003) MicroRNA targets in Drosophila. Genome Biol 5:R1
Erhard F, Zimmer R (2010) Classification of ncRNAs using position and size information in deep sequencing data. Bioinformatics 26:i426–i432
Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison CD, Chan KK, Marcucci G, Calin GA, Huebner K, Croce CM (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci USA 104:15805–15810
Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed–forward regulation of beta-secretase. Nat Med 14:723–730
Fasold M, Langenberger D, Binder H, Stadler PF, Hoffmann S (2011) DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 39:W112–W117
Findeiss S, Langenberger D, Stadler PF, Hoffmann S (2011) Traces of post-transcriptional RNA modifications in deep sequencing data. Biol Chem 392:305–313
Foat BC, Morozov AV, Bussemaker HJ (2006) Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. Bioinformatics 22:e141–e149
Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19:92–105
Fujita N, Watanabe S, Ichimura T, Tsuruzoe S, Shinkai Y, Tachibana M, Chiba T, Nakao M (2003) Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1–HP1 heterochromatic complex for DNA methylation-based transcriptional repression. J Biol Chem 278:24132–24138
Gan J, Shaw G, Tropea JE, Waugh DS, Court DL, Ji X (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. Mol Microbiol 67:143–154
Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45:344–356
Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8:469–477
Gilbert, C, Svejstrup, JQ (2006) RNA immunoprecipitation for determining RNA–protein associations in vivo. Current protocols in molecular biology / edited by Frederick M Ausubel [et al.] Chapter 27: Unit 27 24
Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell 45:814–825
Glazko GV, Zybailov BL, Rogozin IB (2012) Computational prediction of polycomb-associated long non-coding RNAs. PLoS ONE 7:e44878
Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature 470:284–288
Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci USA 106:9613–9618
Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, Herrmann BG (2013) The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev Cell 24:206–214
Gupta SK, Kosti I, Plaut G, Pivko A, Tkacz ID, Cohen-Chalamish S, Biswas DK, Wachtel C, Waldman Ben-Asher H, Carmi S, Glaser F, Mandel-Gutfreund Y, Michaeli S (2013) The hnRNP F/H homologue of Trypanosoma brucei is differentially expressed in the two life cycle stages of the parasite and regulates splicing and mRNA stability. Nucleic Acids Res 41:6577–6594
Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464:1071–1076
Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458:223–227
Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28:503–510
Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477:295–300
Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154:240–251
Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol 12:483–492
Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495:384–388
Hargreaves DC, Crabtree GR (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res 21:396–420
He W, Cai Q, Sun F, Zhong G, Wang P, Liu H, Luo J, Yu H, Huang J, Lin T (2013) linc-UBC1 physically associates with polycomb repressive complex 2 (PRC2) and acts as a negative prognostic factor for lymph node metastasis and survival in bladder cancer. Biochim Biophys Acta 1832:1528–1537
Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665
Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 331:76–79
Heyne S, Costa F, Rose D, Backofen R (2012) GraphClust: alignment-free structural clustering of local RNA secondary structures. Bioinformatics 28:i224–i232
Hiller M, Pudimat R, Busch A, Backofen R (2006) Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. Nucleic Acids Res 34:e117

Hiller M, Zhang Z, Backofen R, Stamm S (2007) Pre-mRNA secondary structures influence exon recognition. PLoS Genet 3:e204

Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, Finstermeier K, Stahringer A, Wilfert W, Beutner F, Gielen S, Schuler G, Gäbel G, Bergert H, Bechmann I, Stadler PF, Thiery J, Teupser D (2013) Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet 9:e1003588

Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. Hum Mol Genet 18:R195–R201

Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12:711–721

Idogawa, M, Ohashi, T, Sasaki, Y, Maruyama, R, Kashima, L, Suzuki, H, Tokino, T (2014) Identification and analysis of large intergenic non-coding RNAs regulated by p53 family members through a genome-wide analysis of p53-binding sites. Hum Mol Genetics (in press)

Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R, Chang HY, Akhtar A (2013) Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Mol Cell 51: 156–173

Ilott NE, Ponting CP (2013) Predicting long non-coding RNAs using RNA sequencing. Methods 63:50–59

Jeffery L, Nakielny S (2004) Components of the DNA methylation system of chromatin control are RNA-binding proteins. J Biol Chem 279:49479–49487

Ji W, Yang L, Yuan J, Yang L, Zhang M, Qi D, Duan X, Xuan A, Zhang W, Lu J, Zhuang Z, Zeng G (2013) MicroRNA-152 targets DNA methyltransferase 1 in NiS-transformed cells via a feedback mechanism. Carcinogenesis 34:446–453

Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, Peng W, Wei G, Jing Q, Wakabayashi Y, Bansal A, Luo Y, Ribeiro JMC, Scherf A, Aravind L, Zhu J, Zhao K, Miller LH (2013) PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. Nature 499: 223–227

Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D (2013) Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. Mol Cell 1–11

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell 38:675–688

Kazan H, Ray D, Chan ET, Hughes TR, Morris Q (2010) RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. PLoS Comput Biol 6:e1000832

Keller C, Kulasegaran-Shylini R, Shimada Y, Hotz H-R, Bühler M (2013) Noncoding RNAs prevent spreading of a repressive histone mark. Nature Structural &#38. Mol Biol 20:994–1000

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39: 1278–1284

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci USA 106:11667–11672

Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–915

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y (2011) Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene 30:1956–1962

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494:497–501

Lall S, Grun D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P, Kao HL, Gunsalus KC, Pachter L, Piano F, Rajewsky N (2006) A genome-wide map of conserved microRNA targets in C. elegans. Curr Biol 16:460–471

Lange SJ, Alkhnbashi OS, Rose D, Will S, Backofen R (2013) CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. Nucleic Acids Res 41:8034–8044

Langenberger D, Pundhir S, Ekstrom CT, Stadler PF, Hoffmann S, Gorodkin J (2012) DeepBlockAlign: a tool for aligning RNA-seq profiles of read block patterns. Bioinformatics 28:17–24

Leeb M, Wutz A (2007) Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol 178:219–229

Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C (2008) The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. Mol Cell 32:129–139

Leibovich L, Paz I, Yakhini Z, Mandel-Gutfreund Y (2013) DRIMust: a web server for discovering rank imbalanced motifs using suffix trees. Nucleic Acids Res 41:W174–W179

Leung YY, Ryvkin P, Ungar LH, Gregory BD, Wang LS (2013) CoRAL: predicting non-coding RNAs from small RNA-sequencing data. Nucleic Acids Res 41:e137

Li, LC (2013) Chromatin remodeling by the small RNA machinery in mammalian cells. Epigenetics 9 (in press)

Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbo G, Wu Z, Zhao Y (2011a) Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res 39:3864–3878

Liao Q, Xiao H, Bu D, Xie C, Miao R, Luo H, Zhao G, Yu K, Zhao H, Skogerbo G, Chen R, Wu Z, Liu C, Zhao Y (2011b) ncFANs: a web server for functional annotation of long non-coding RNAs. Nucleic Acids Res 39:W118–W124

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X, Darnell JC, Darnell RB (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

Lorenz R, Hofacker IL, Bernhart SH (2012) Folding RNA/DNA hybrid duplexes. Bioinformatics 28:2530–2531

Luo M, Li Z, Wang W, Zeng Y, Liu Z, Qiu J (2013) Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. Cancer Lett 333:213–221

Malik S, Roeder RG (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11:761–772

Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD (2013) A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. Genes Dev 27:1851–1856

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 42:D142–D147

Maticzka D, Lange SJ, Costa F, Backofen R (2014) GraphProt: modeling binding preferences of RNA-binding proteins. Genome Biol 15:R17

Maxwell SS, Pelka GJ, Tam PP, El-Osta A (2013) Chromatin context and ncRNA highlight targets of MeCP2 in brain. RNA Biol 10:1741–1759

Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20:300–307

Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci USA 105:716–721

Meyer F, Kurtz S, Backofen R, Will S, Beckstette M (2011) Structator: fast index-based search for RNA sequence-structure patterns. BMC Bioinforma 12:214

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560

Milek M, Wyler E, Landthaler M (2012) Transcriptome-wide analysis of protein–RNA interactions using high-throughput sequencing. Semin Cell Dev Biol 23:206–212

Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C (2010) Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development 137:2493–2499

Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, Gyllensten U, Kanduri C (2012) Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. Development 139:2792–2803

Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C (2010) Characterization of the RNA content of chromatin. Genome Res 20: 899–907

Monnier P, Martinet C, Pontis J, Stancheva I, Ait-Si-Ali S, Dandolo L (2013) H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. Proc Natl Acad Sci USA 110:20693–20698

Morey L, Aloia L, Cozzuto L, Benitah SA, Di Croce L (2013) RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Rep 3:60–69

Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322:1717–1720

Nikolajewa S, Pudimat R, Hiller M, Platzer M, Backofen R (2007) BioBayesNet: a web server for feature extraction and Bayesian network modeling of biological sequence data. Nucleic Acids Res 35:W688–W693

Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA interference and X-inactivation pathways. Science 320:1336–1341

Oliveira CS, Saraiva NZ, Cruz MHC, Mazeti B, Oliveira LZ, Lopes FL, Garcia JM (2013) HDAC inhibition decreases XIST expression on female IVP bovine blastocysts. Reproduction (Cambridge, England) 145:9–17

Panning B, Jaenisch R (1996) DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev 10:1991–2002

Parker BJ, Moltke I, Roth A, Washietl S, Wen J, Kellis M, Breaker R, Pedersen JS (2011) New families of human regulatory RNA structures identified by comparative analysis of vertebrate genomes. Genome Res 21:1929–1943

Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat Rev Genet 13: 271–282

Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. Cell 137:1194–1211

Pullirsch D, Härtel R, Kishimoto H, Leeb M, Steiner G, Wutz A (2010) The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. Development 137:935–943

Ramos AD, Diaz A, Nellore A, Delgado RN, Park KY, Gonzales-Roybal G, Oldham MC, Song JS, Lim DA (2013) Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell 12:616–628

Rapicavoli NA, Poth EM, Zhu H, Blackshaw S (2011) The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. Neural Dev 6:32

Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134:223–232

Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166

Sanford JR, Wang X, Mort M, Vanduyn N, Cooper DN, Mooney SD, Edenberg HJ, Liu Y (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res 19: 381–394

Sarraf SA, Stancheva I (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol Cell 15:595–605

Schmitz KM, Mayer C, Postepska A, Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev 24:2264–2269

Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A (2006) Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. EMBO J 25:3110–3122

Schor IE, Llères D, Risso GJ, Pawellek A, Ule J, Lamond AI, Kornblihtt AR (2012) Perturbation of chromatin structure globally affects localization and recruitment of splicing factors. PLoS ONE 7: e48084

Stoltenburg R, Reinemann C, Strehlitz B (2007) SELEX– a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol Eng 24:381–403

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550

Sun BK, Deaton AM, Lee JT (2006) A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell 21:617–628

Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C (2013a) R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science 340:619–621

Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT (2013b) Jpx RNA activates Xist by evicting CTCF. Cell 153:1537–1551

Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y (2013c) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 41:e166

Tavares L, Dimitrova E, Oxley D, Webster J, Poot R, Demmers J, Bezstarosti K, Taylor S, Ura H, Koide H, Wutz A, Vidal M, Elderkin S, Brockdorff N (2012) RYBP–PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell 148:664–678

Tenenbaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci USA 97:14085–14090

Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G (2006) Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res 34:2791–2802

Torarinsson E, Havgaard JH, Gorodkin J (2007) Multiple structural alignment and clustering of RNA sequences. Bioinformatics 23: 926–932

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511–515

Tsai M-C, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329:689–693

Tseng HH, Weinberg Z, Gore J, Breaker RR, Ruzzo WL (2009) Finding non-coding RNAs through genome-scale clustering. J Bioinforma Comput Biol 7:373–388

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302: 1212–1215

Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD (2012) Site identification in high-throughput RNA–protein interaction data. Bioinformatics 28:3013–3020

Vogel T, Stoykova A, Gruss P (2006) Differential expression of polycomb repression complex 1 (PRC1) members in the developing mouse brain reveals multiple complexes. Dev Dyn 235:2574–2585

Wan G, Hu X, Liu Y, Han C, Sood AK, Calin GA, Zhang X, Lu X (2013) A novel non-coding RNA lncRNA–JADE connects DNA damage signalling to histone H4 acetylation. EMBO J 32:2833–2847

Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell 43:904–914

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472:120–124

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505: 117–120

Warnecke PM, Biniszkiewicz D, Jaenisch R, Frommer M, Clark SJ (1998) Sequence-specific methylation of the mouse H19 gene in embryonic cells deficient in the Dnmt-1 gene. Dev Genet 22:111–121

Washietl S, Findeiss S, Muller SA, Kalkhof S, von Bergen M, Hofacker IL, Stadler PF, Goldman N (2011) RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. RNA 17:578–594

Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol 3:e65

Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J (2013) Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci USA 110: E3487–E3496

Wu X, Johansen JV, Helin K (2013a) Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell 49:1134–1146

Wu L, Murat P, Matak-Vinkovic D, Murrell A, Balasubramanian S (2013b) Binding Interactions between Long Noncoding RNA HOTAIR and PRC2 Proteins. Biochemistry 52:9519–9527

Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y (2014) NONCODEv4: exploring the world of long non-coding RNA genes. Nucleic Acids Res 42:D98–D103

Yang F, Huo X, Yuan S, Zhang L, Zhou W, Wang F, Sun S (2013a) Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. Mol Cell 49:1083–1096

Yang JH, Li JH, Jiang S, Zhou H, Qu LH (2013b) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res 41: D177–D187

Yang L, Lin C, Jin C, Yang JC, Tanasa B, Li W, Merkurjev D, Ohgi KA, Meng D, Zhang J, Evans CP, Rosenfeld MG (2013c) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature 500:598–602

Yoon JH, Abdelmohsen K, Gorospe M (2013) Posttranscriptional gene regulation by long noncoding RNA. J Mol Biol 425:3723–3730

Yap KL, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou M-M (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol Cell 38:662–674

Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M (2012) LincRNA-p21 suppresses target mRNA translation. Mol Cell 47:648–655

Zambelli F, Pesole G, Pavesi G (2013) PscanChIP: Finding over-represented transcription factor-binding site motifs and their correlations in sequences from ChIP-Seq experiments. Nucleic Acids Res 41:W535–W543

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40:939–953

Zhu L, Xu P-C (2013) Downregulated LncRNA–ANCR promotes osteoblast differentiation by targeting EZH2 and regulating Runx2 expression. Biochem Biophys Res Commun 432:612–617

Zhu Y, Rowley MJ, Böhmdorfer G, Wierzbicki AT (2013) A SWI/SNF chromatin-remodeling complex acts in noncoding RNA-mediated transcriptional silencing. Mol Cell 49:298–309