Chapter 13

Chromatin Regulation at Parental Gene Promoters by Pseudogene Sense lncRNAs

Stefan Schoeftner, Michele Scarola, and Roberta Benetti

Abstract

There is accumulating evidence that pseudogenes can produce functionally relevant lncRNAs in a tightly controlled manner. This class of transcripts has been demonstrated to play an important role in development and disease, by controlling parental gene expression. Classically, pseudogene derived lncRNAs compete with parental transcripts for miRNAs or factors that control parental mRNA metabolisms. Recently, pseudogene lncRNAs were demonstrated to take over the control of classic chromatin modifying enzymes and alter parental gene promoter activity or genome wide gene expression. Here, we discuss a new mechanism of parental gene expression controlled by the *mOct4P4* lncRNA, a sense transcript derived from the murine *Oct4 pseudogene 4*. *mOct4P4* lncRNA specifically interacts with the RNA binding protein FUS and the Histone Methyltransferase SUV39H1 to target heterochromatin formation at the parental *Oct4* promoter *in trans*. In addition, we will address key issues for the functional dissection of epigenetic control of parental gene promoters by pseudogene lncRNAs.

Key words *Oct4*, *Oct4* pseudogenes, SUV39H1, FUS, *mOct4*, *mOct4P4*, *hOCT4P3*, lncRNA

1 Pseudogene Derived RNAs

Pseudogenes refer to sequences in the genome regions that contain defective copies of genes. However, increasing evidence demonstrate that pseudogenes can hold important biological functions.

Pseudogenes are classified by their mechanisms of origin into four groups. Unitary pseudogenes have formed by the acquisition of mutations of the original gene without involving duplication events; duplicated pseudogenes (also named "unprocessed pseudogenes") originate from a gene duplication event paired with the acquisition of mutations; processed pseudogenes originate from retrotransposition events involving mRNA; finally, a rare class of polymorphic pseudogenes do not show disabling mutations in a limited number of individuals (*see* refs. 1–3 and chapter 5 of this book). The GENCODE project, which has the goal to annotate all evidence-based gene features, currently lists a total of 14,763

Laura Poliseno (ed.), *Pseudogenes: Functions and Protocols*, Methods in Molecular Biology, vol. 2324,
https://doi.org/10.1007/978-1-0716-1503-4_13,

pseudogenes, where processed pseudogenes represent 72% of all pseudogenes (release 34, Sisu-*GENCODE pseudogenes*, this book). Importantly, approximately 15% of pseudogenes provide evidence for transcription, indicating the positioning of a significant number of pseudogenes in the vicinity of regulatory elements [4]. Transcribed pseudogenes are reported to produce sense and antisense transcripts that can act as functional ncRNAs with acquired, new biological function and critically impact on development and disease [4, 5]. Remarkably, pseudogenes derived transcripts can also hold coding potential for peptide synthesis [6, 7]. However, unraveling the functional relevance of these peptides needs further investigation.

Here, we briefly introduce different pathways that involve sense pseudogene lncRNAs in the control of parental gene expression and focus in detail on target site specific chromatin modification at promoter of the murine *Oct4* gene triggered *in trans* by the *Oct4* pseudogene 4 (*mOct4P4*) lncRNA.

2 Mechanisms of Parental Gene Expression Control by Sense Pseudogene lncRNAs

In many cases, sequence similarity provides the basis for pseudogene lncRNA mediated control of parental gene expression. Mechanisms of antisense pseudogene lncRNA function are discussed in a separate chapter (Chapter 14 by Lister et al.).

LncRNA generated by sense transcription of pseudogenes can alter parental gene expression by competing with the parental transcript for miRNAs, thus increasing the expression of the parental gene. The pioneering work on the function of pseudogene lncRNA as competing endogenous RNAs (ceRNAs) demonstrated the tumor suppressive impact of the *PTENP1* lncRNA by sponging miRNAs that target the 3′UTR region of the haploinsufficient tumor suppressor PTEN, resulting in reduced cancer cell proliferation [8, 9]. In addition, sequence similarity to parental mRNAs allows offspring pseudogene derived lncRNA to sequester protein factors originally directed to control the metabolism of the respective parental mRNA [10, 11]. Furthermore, pseudogene transcripts can impact on parental gene expression control by contributing an RNA substrate for the production of endo-siRNAs. This can be governed by antisense pseudogene transcription, but also by transcribing sense and antisense RNA from inverted repeat pseudogene sequences, as reported for endo-siRNAs targeting the GTPase-activating protein for Ran (Ran-GAP) [12, 13]. Further, recent reports suggest that a subset of pseudogenes is subjected to translation of truncated or mutated proteins with novel biological function [6, 7]. Finally, pseudogene-derived sense lncRNAs, such as DUXAP8 and DUXAP10, have been shown to interact with epigenetic writers [14–16]. However, DUXAP lncRNAs rather

appear to act as scaffold for epigenetic regulatory complexes that have a role in overall epigenome regulation, but do not selectively target the parental gene.

3 *Oct4* Pseudogenes Display Controlled Expression and Have Acquired Multiple Biological Functions

Oct4 is a member of the POU family of transcription factors, controls early embryonic differentiation and is essential to maintain mESC pluripotency and self-renewal [17–19]. During evolution, the murine and human Pou5f1/POU5F1 genes, that encode *OCT4*, gave rise to five processed murine (*Pou5F1P1-Pou5F1P5*) and eight processed human pseudogenes (POU5F1P1-POU5F1P8), hereinafter referred to as *mOct4P1–5* and *hOCT4P1–8* pseudogenes, with validated lncRNA expression [20–27]. Here, we further concentrate on mouse *Oct4* pseudogenes (Fig. 1). Individual segments of mouse *Oct4* pseudogene transcripts display sequence conservation levels ranging from 66% to 100% with respect to the paternal *Oct4* mRNA (Fig. 1). Interestingly, *mOct4P4* shows homology to the entire mature *Oct4* mRNA

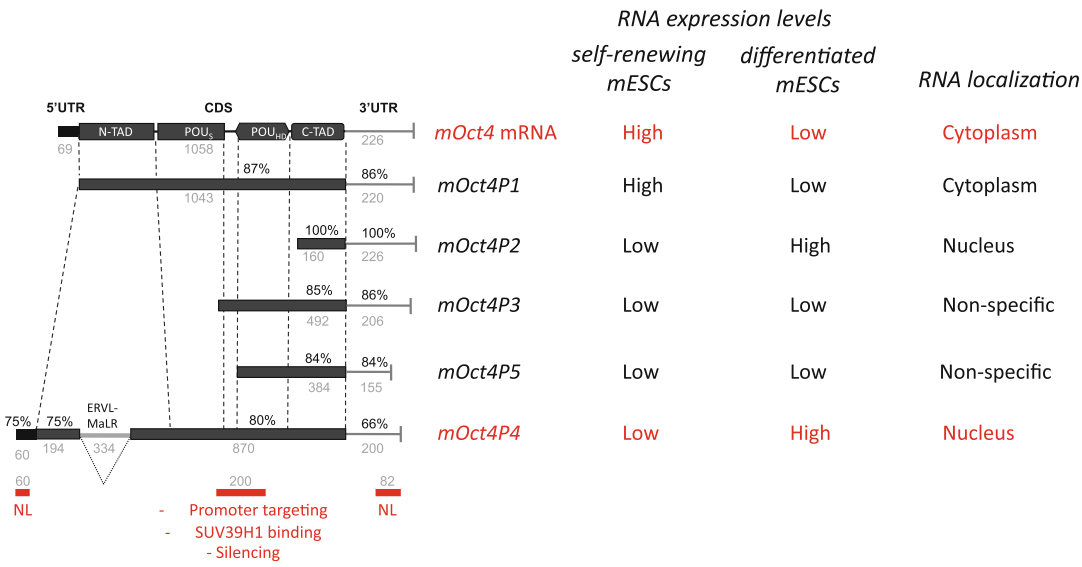

Fig. 1 Murine *Oct4* pseudogenes and functionally relevant *mOct4P4* lncRNA regions. (*left*) Gray numbers: length (bp). Black numbers: percentage of sequence homology of *Oct4* pseudogenes compared to the corresponding *Oct4* regions. ERVL-MaLR: spliced region of *mOct4P4* with homology to the LTR of ERVL-MaLR retrotransposons. OCT4 protein domains: N-terminal trans-activating domain (N-TAD); POU-specific DNA-binding domain (POU_S); DNA-binding homeodomain (POU_{HD}); C-terminal trans-activating domain (C-TAD). Position of *mOct4P4* lncRNA regions with functional relevance (red bars): nuclear localization of lncRNA (NL); 200 nucleotide region essential for *Oct4* promoter targeting, SUV39H1 binding and *Oct4* promoter silencing. (*right*) Transcript expression in self-renewing and differentiated mESCs

but has acquired a 334 nt insertion with homology to the LTR element of the ERVL-MaLRs retrotransposon family. This segment is however eliminated by RNA splicing and has no relevance for *mOct4P4* function [24].

Identified murine *Oct4* pseudogene derived lncRNAs show a defined pattern of expression during mouse embryonic stem cells (mESC) differentiation and specific cytoplasmatic or nuclear localization (Fig. 1) [24]. This supports evidence for the acquisition of new biological function during evolution [24]. In line with this, a set of *Oct4/OCT4* pseudogene lncRNAs alter parental gene expression by acting as classic ceRNAs, or may recruit the histone methyltransferase (HMTase) Ezh2 to the *OCT4* promoter, guided by pairing of antisense pseudogene RNA with *Oct4* transcripts [28–31].

4 *mOct4P4* lncRNA-Mediated Chromatin Regulation, a Novel Paradigm of Parental Gene Expression Control *In Trans*

4.1 *Structural Features of* mOct4P4 *lncRNA and Biological Relevance*

The murine *Oct4* pseudogene lncRNA *mOct4P4* uses a novel mechanism for parental gene control that is based on chromatin modulation *in trans*. *mOct4P4* lncRNA originates from a retrotransposition of *Oct4* mRNA, including 5′ and 3′UTR elements, into the X chromosome. Interestingly, *mOct4P4* is efficiently upregulated during mouse embryonic stem cell (mESC) differentiation, a biological process that is defined by the rapid repression of the transcription factor *Oct4* [24]. Transient depletion of *mOct4P4* lncRNA during mESC differentiation or in terminally differentiated cells results in the failure of *Oct4* repression and partial reestablishment of stem cell features, such as telomerase dependent telomere elongation [24]. In contrast, ectopic expression of *mOct4P4* lncRNA in self-renewing mESCs results in reduced *Oct4* expression and expression of early markers of mESC differentiation. Localization of the *Oct4* gene on chromosome 17 and exclusive sense transcription of the *mOct4P4* gene from the X chromosomes indicate a *trans*-acting silencing mechanism that does not involve RNA:RNA duplex-based targeting of the parental gene.

4.2 mOct4P4 *lncRNA Has Acquired New Elements with Defined Function*

A classic deletion analysis on ectopically expressed *mOct4P4* lncRNA identified sequence elements that are critical for *mOct4P4* function (Fig. 1): pseudogene-specific motifs in the 5′ and 3′ UTR regions that mediate nuclear localization, as well as a 200 nucleotide element located at a central region of *mOct4P4* lncRNA, essential for *Oct4* promoter targeting and chromatin modification (green in Fig. 2 [32]. Importantly, artificial, minimal *mOct4P4* lncRNA versions comprising the 200-nucleotide region plus 5′ and 3′ UTR elements are sufficient to trigger heterochromatin formation at the *Oct4* promoter (Fig. 1, 2; [32]).

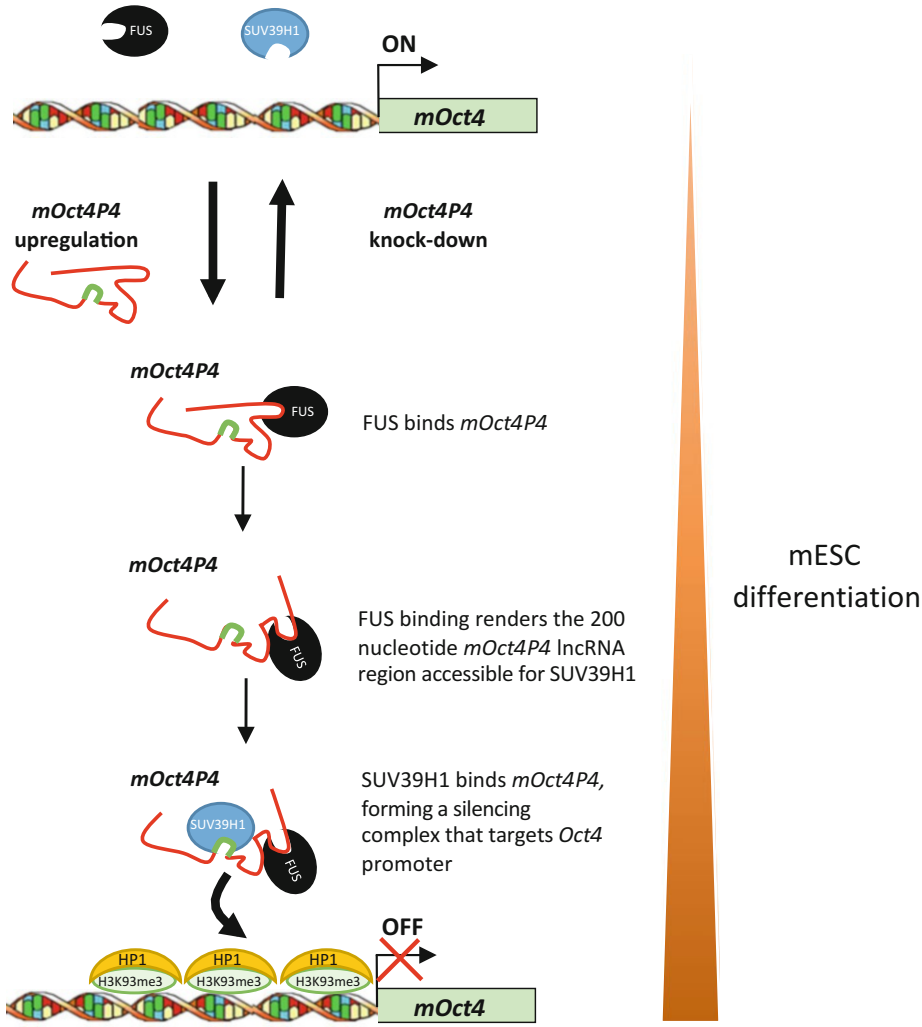

Fig. 2 FUS dependent licensing of targeted silencing of *Oct4* by *mOct4P4*. FUS binds to *mOct4P4* lncRNA to render the 200 nt region, depicted in green, accessible to SUV39H1 binding. The resulting *mOct4P4* lncRNA–SUV39H1 complex holds targeting specificity for the promoter of parental *Oct4*, leading to local formation of heterochromatin *in trans*

4.3 Chromatin Modulation of Parental Oct4 *Promoter Triggered by* mOct4P4 *lncRNA*

The use of the MS2 RNA tagging system to generate ectopically expressed, chimeric *mOct4P4* revealed major insights on the role of *mOct4P4* lncRNA in triggering alterations of chromatin structure at the promoter of the parental *Oct4* gene [24]. mESCs expressing the flag-tagged MS2 page coat protein were stably transfected with *mOct4P4* lncRNA, tagged at the 3′end with 24 repeats of the MS2 RNA stem loop motif. Anti-flag-MS2 ChIP allowed to locate the chimeric *mOct4P4*-MS2 stem loop lncRNA at conserved elements of the parental *Oct4* gene promoter. This was paralleled by recruitment of the repressive SUV39H1 HMTase, which induced local trimethylation of histone H3 lysine 9 (H3K9me3) and subsequent

recruitment of Heterochromatin Protein 1 (HP1). Thus, *mOct4P4* lncRNA directed the establishment of a classic, repressive chromatin signature *in trans*. Depletion of the chimeric lncRNA or SUV39H1 abrogated *Oct4* silencing, demonstrating that the *mOct4P4* lncRNA has a critical role in triggering heterochromatinization *in trans*.

4.4 RNA:Protein Interaction Regulates mOct4P4 *lncRNA* Function

RNA immunoprecipitation (RIP) experiments revealed that binding of SUV39H1 to the 200 nucleotide *mOct4P4* lncRNA region is essential for *Oct4* promoter targeting and gene silencing, thus identifying this region as central regulator of parental gene silencing [32]. *mOct4P4*-MS2 stem loop lncRNA IP experiments in flag-MS2 protein expressing cells coupled with mass spectrometry identified the RNA binding protein FUS as novel *mOct4P4* lncRNA interactor. FUS binding to the full-length *mOct4P4* lncRNA is essential to allow the interaction of SUV39H1 HMTase with the critical 200 nucleotide region [32]. Remarkably, FUS is dispensable for *Oct4* silencing by the minimal functional *mOct4P4* version. This suggests a role for FUS in modelling endogenous *mOct4P4* lncRNA structure to allow region (Fig. 2; [32]).

Together, this demonstrates that pseudogene evolution can lead to the development of multistep regulatory mechanisms that may have a critical role in fine-tuning lncRNA function or limit pseudogene lncRNA function to a defined developmental window or physiological context. Along these lines, PRMT1-dependent arginine methylation of FUS was recently shown to prevent its interaction with the CCND1 gene promoter-associated noncoding RNA–D (pncRNA-D), thereby blocking the repression of the Histone acetylation activity of the CBP/p300 HAT complex [33, 34]. Identifying *mOct4P4* lncRNA binding proteins and addressing posttranslational modifications of FUS represent interesting approaches to link pseudogene lncRNA function to central signaling cascades or pathways in development and disease.

5 Mechanisms Underlying Target Specificity for Pseudogene lncRNAs That Modify Chromatin Structure *In Trans*

Chromatin regulation represents a central feature of lncRNA function. The lncRNA *Xist* has a critical role in recruiting Polycomb repressive complexes to the future inactive X chromosome, representing a hallmark model for lncRNA mediated chromatin regulation [35–37]. DUXAP pseudogene derived lncRNAs have been shown to be relevant for chromatin modulation *in trans* by interacting with epigenetic writers. DUXAP10 interacts with the histone demethylase Lysine specific demethylase 1 (LSD1) to repress the tumorsuppressor LATS2 and RRAD [15]. DUXAP8 lncRNA recruits LSD1 and Polycomb group protein EZH2 to repress the

tumor suppressors EGR1 and RHOB in lung cancer cells [16]. However, DUXAP lncRNAs appears to act as general scaffold for epigenetic regulatory complexes without targeting the parental DUXAP gene. To date, the *mOct4P4* lncRNA represent an exclusive example for sense transcribed lncRNA with target specificity toward the parental *Oct4* gene *in trans*. However, the precise mechanism ensuring the translocation of the *mOct4P4* transcript from its site of production on the X chromosome to the *Oct4* promoter on Chr. 17 is still elusive. Here, we propose possible scenarios for such a mechanism.

5.1 mOct4P4 *lncRNA Recruitment by Factors Deposited at the* Oct4 *Promoter*

Classic transient luciferase reporter experiments revealed that ectopic *mOct4P4* lncRNA expression reduces the activity of a subcloned 3 kb *Oct4* promoter [24]. This suggests that the *Oct4* promoter may contain sequence elements loaded with proteins that may recruit *mOct4P4* lncRNA and SUV39H1. Such scenario is reminiscent of dosage compensation of X linked gene expression in *D. melanogaster*, which involves the X-linked roX1 and roX2 lncRNAs and the chromatin activating male sex lethal (MSL) complex. The MSL complex binds to "high-affinity" chromatin entry sites, subsequently roX1 and roX2 lncRNAs coordinate the spreading of MSL complex along the entire X chromosome, upregulating X-linked gene expression in males [38–40]. In the context of mESC differentiation, strong upregulation of the *mOct4P4* gene may mediate enhanced *mOct4P4* lncRNA – SUV39H1 complex formation, allowing complex recruitment by factors already deposited at parental *Oct4* gene promoter (Fig. 3a). Subsequent imposition of H3K9me3 creates a high-affinity binding-site for HP1, which in turn interacts with SUV39H1, creating a self-enforcing loop that can stabilize and maintain heterochromatin at the parental *Oct4* promoter [41]. A deletion analysis focusing on the *Oct4* promoter region may provide relevant information on sequence elements bound by potential *mOct4P4* lncRNA recruitment factors. Time-course ChIP experiments with inducible *mOct4P4* lncRNA expression models will be useful to map possible *mOct4P4* lncRNA entry sites along *Oct4* promoter regions.

5.2 *Long-Distance Chromatin Interactions Between* Oct4 *and* mOct4P4 *Alleles*

Precise multidimensional organization of chromatin in the nucleus is central to the establishment of specific gene expression programs during development and disease [42, 43]. LncRNAs, specialized proteins such as CTCF and cohesin play key roles in controlling intrachromosomal formation of topologically associated domains (TADs) or enhancer-promoter interactions [44]. Interchromosomal contacts with a role in gene expression control can be subdivided into nonhomologous chromosome contacts (NHCC) and homologous chromosome contacts between two alleles of the same gene, also known as "transvection" [42, 45]. Transcription of ribosomal genes located on five different human chromosomes in

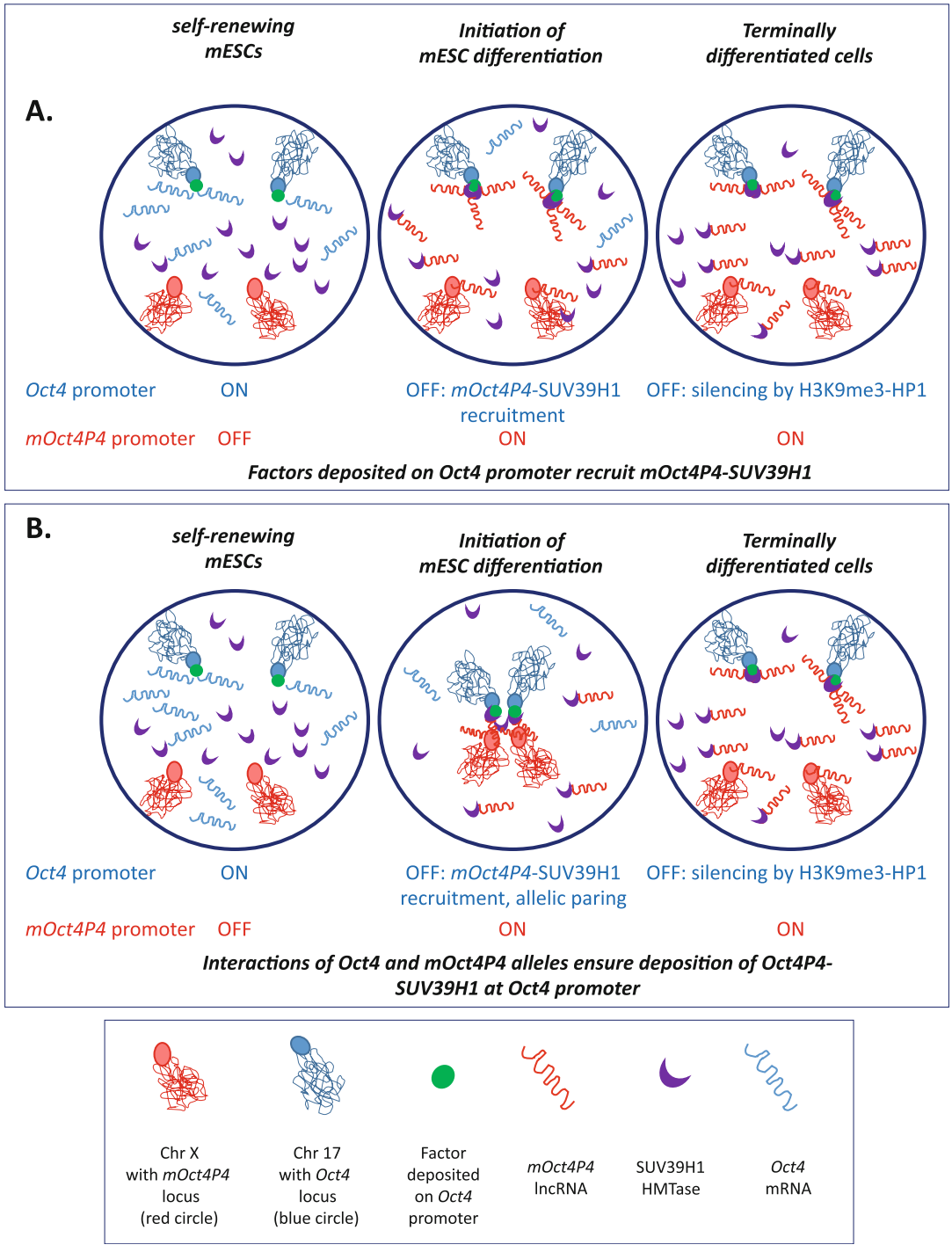

Fig. 3 Models for target site-specific action of *mOct4P4* lncRNA. (**a**) In self-renewing mESCs, factors for lncRNA recruitment may be deposited at the promoter of the parental *Oct4* gene (green circles). Parental *Oct4* is expressed at high levels; SUV39H1 is constitutively expressed. At the onset of mESC differentiation, *mOct4P4* lncRNA is upregulated, binds the SUV39H1 HMTase and may be recruited by the factors deposited

the nucleolus, and the so-called olfactosome (a subnuclear structure that hosts and coordinates numerous olfactory receptor (OR) genes) represent classic examples for NHCC [46, 47]. Transvection has been first documented in pioneering studies on the Ubx locus in *Drosophila melanogaster* [45]. In vertebrates, the allelic pairing of the X-chromosome inactivation center in female embryonic stem cells prior to X chromosome silencing, and the reported allelic pairing of the *Oct4* locus represent mechanism similar to classic transvection in *D. melanogaster* [48, 49]. In particular, the transient pairing of *Oct4* alleles during mESC differentiation and embryonic development was reported to precede efficient *Oct4* downregulation, which is required for cell differentiation [49]. Importantly, allelic pairing was found to be dependent on promoter/enhancer regions that also respond to ectopically expressed *mOct4P4* lncRNAs in reporter assays [24]. This raises the possibility that the *mOct4P4* lncRNA - SUV39H1 complex may work on paired *Oct4* alleles. It might also be possible that such pairing includes *Oct4* pseudogene alleles (Fig. 3b). This would allow direct loading of *mOct4P4* lncRNA—SUV39H1 complexes to the *Oct4* promoter, facilitating efficient silencing of parental *Oct4* gene expression (Fig. 3b). Future studies on *mOct4P4* lncRNA function should comprise approaches that address 3-dimensional gene positioning of pseudogenes, parental genes and pseudogene lncRNAs.

6 Fast Strategies to Address the Biological Relevance of Pseudogene-Encoded Sense lncRNAs

Investigating the impact of pseudogene-derived lncRNAs in modifying chromatin structure at parental gene promoters or across the genome requires non-standard experimental procedures. Key issues in successfully addressing pseudogene lncRNA function comprise selection of a biological model system with traceable physiological

Fig. 3 (continued) at the promoter of parental *Oct4* gene *in trans*. This triggers local H3K9me3 and HP1 binding. In terminally differentiated cells, a H3K9me3-HP1-SUV39H1-*mOct4P4* lncRNA self-reinforcing loop ensures the maintenance of epigenetic silencing. (**b**) In self-renewing mESCs, *Oct4* (Chr. 17) and *mOct4P4* alleles (Chr. X) do not pair; *Oct4* and SUV39H1 are expressed; the *mOct4P4* locus is silent. At the onset of mESC differentiation, *Oct4* loci show allelic pairing [50] and *mOct4P4* lncRNA expression is strongly upregulated. A potential allelic pairing of *mOct4P4* alleles with parental *Oct4* alleles may represent an ideal situation to locate *de novo* formed *mOct4P4* lncRNA—SUV39H1 complexes to the promoters of parental *Oct4* alleles. Factors located at *Oct4* promoters (green circles) may support this mechanism. The SUV39H1 HMTase induces local H3K9me3 and HP1 recruitment to the *Oct4* promoters. At later stages of mESC differentiation, allelic pairing is disrupted (ref. 50) and *Oct4* silencing may be maintained by a H3K9me3-HP1-SUV39H1-*mOct4P4* lncRNA self-reinforcing loop

changes that are paralleled by altered expression of parental gene and pseudogenes, as well as experimental methods that allow to define pseudogene lncRNA localization and expression level.

6.1 Individuating a Model System That Allows to Address the Biological Relevance of Pseudogene-Derived lncRNAs

The availability of a strong biological model system that allows to validate the biological relevance of pseudogene lncRNAs is of primary importance. The parental gene should be a key regulator of a biological process and its expression should undergo alteration during such event, thus allowing to perform gain and loss of function experiments. *Oct4* is a central self-renewal transcription factor that is dramatically downregulated during experimentally triggered in vitro differentiation of mESCs. In turn, mESCs are easy to culture and available in unlimited quantity. Furthermore, alterations in differentiation status are characterized by morphological changes and altered marker gene expression.

6.2 Using Publicly Available Genomics Data and Genome Browsers to Get First Information on Pseudogenes

Current genome browsers such as ENSEMBL offer a valuable resource to identify annotated pseudogenes, their chromosomal localization and presence of regulatory elements in the region of interest. In addition, consulting publicly available RNA-seq and ChIP-seq data sets allows to study cell type-specific lncRNA expression and alterations of chromatin status at putative promoters and pseudogene bodies, both strong indicators for the acquisition of new biological function by pseudogenes.

6.3 Design of Specific Primer Sets for PCR Amplification

Protein-coding genes can give rise to multiple pseudogenes. This raises a critical need for the design of highly specific PCR primer pairs that discriminate between parental and pseudogene transcripts. The subcloning of obtained PCR products followed by DNA sequencing is essential to validate the identity of PCR amplicons. Transient pseudogene lncRNA loss and gain of function experiments are helpful in validating specific pseudogene lncRNA amplification.

6.4 Characterization of Key Features of the Pseudogene lncRNA

Quantitative real-time PCR strategies allow to define central features of pseudogene lncRNAs. This method can be applied to quantify alterations in lncRNA expression in the experimental model system of choice. Intercellular lncRNA localization can be evaluated by RT-PCR on cytoplasmatic, nuclear or chromatin fractions prepared from experimental cells. A restricted amplification of lncRNAs from nuclear/chromatin fractions may point toward chromatin regulation; amplification from cytoplasmatic fractions may rather suggests a role as competing endogenous RNA. The use of oligo-dT primers for cDNA synthesis gives a good indication for significant 3′ polyadenylation. Finally, 5′ ends of transcripts can be validated by 5′RACE.

6.5 Finding First Evidence for Pseudogene lncRNA Mediated Control of Parental Gene Expression and Function

Classic lncRNA gain and loss of function experiments give first evidence for a functional relevance. Ectopic expression of a full-length pseudogene lncRNA under the control of a strong promoter, either in transient or stable conditions, need to result in a significant alteration of parental gene expression, paralleled by physiological alterations of the chosen cell model system. In order to exclude off-target effects, the altered biological status of model cells should be rescuable by ectopic expression (or siRNA mediated depletion) of the parental gene. Pseudogene loss of function experiments can be performed by transfecting experimental cells with pseudogene lncRNA specific siRNAs, as demonstrated for the *mOct4P4* lncRNA. In all lncRNA gain and loss of functions, alterations of the expression of other relevant pseudogenes should be excluded. Optionally, CRISPR/Cas9 based technology can be used to generate loss of function models [32].

7 Dissecting Chromatin Regulation by Pseudogene lncRNAs Using the MS2 Tagging System

RNA hairpin sequences of the MS2 bacteriophage are frequently used to study mRNAs in eukaryotic cells [50]. The high-affinity interaction between the MS2 RNA stem loop and the MS2 phage coat protein allows the purification of chimeric transcripts containing the MS2 RNA stem loop motif fused to the RNA of interest. Immunoprecipitating the MS2 protein allows to follow the docking of pseudogene lncRNA to chromatin, but also enables the analysis of lncRNA interacting proteins [32]. The MS2 system requires a careful setup and optimization. In a first step, experimental cell lines stably expressing flag-epitope tagged MS2 fused to a nuclear localization signal need to be established [24, 51]. In a subsequent step, flag-MS2 expressing cells are used to generate cell clones that stably express the chimeric pseudogene lncRNA tagged with a certain number of repeats of the MS2 RNA stem loop motif [24, 51]. It is important to notice that repeat numbers need to be optimized. Multiple clones expressing the MS2 RNA stem loop tagged lncRNA and the flag-tagged MS2 protein need to demonstrate correct localization of the ectopically expressed lncRNA and the expected effect on parental gene expression. Adequate control cell lines lacking the chimeric transcript need to be established in parallel.

7.1 Identification and Validation of Target Chromatin Sites by ChIP

The use of anti-flag antibodies allows to evaluate the localization of chimeric lncRNA-MS2 RNA stem loop molecules at target chromatin sites by ChIP. Quantitative real-time PCR using primers specific for regulatory elements of the parental gene and unrelated control genes can reveal a possible interaction of the lncRNA of interest with the parental gene. In parallel, classic posttranslational

histone tail modifications defining the transcriptional status of a gene can be assessed. Priority should be given to chromatin marks related to (1) signatures of constitutive heterochromatin (H3K9me3, HP1); (2) Polycomb group (PcG) protein dependent gene silencing (H3K27me3, H2AK119u1); (3) Trithorax (TrX) protein-dependent activation of transcription (H3K4me3); and (4) histone H3 and H4 acetylation levels. The observed enrichment of histone modifications can already give an important indication on possible epigenetic writers that can be subsequently validated by additional ChIP experiments. Importantly, this experimental strategy can be extended to ChIP-seq, in order address a possible role of pseudogene lncRNAs in sculpturing the whole epigenome.

7.2 Identification and Validation of Protein Interactors by RIP

The application of RIP allows to identify lncRNA interacting proteins, using a candidate approach. This approach resulted in the identification of SUV39H1 as *mOct4P4* lncRNA interactor [24]. Similar to the ChIP approach, RIP experiments can be upgraded to the proteome level. Immunoprecipitation of chimeric lncRNA-MS2 RNA stem loop–protein complexes using the MS2 system, followed by mass spectrometry on RIP eluates, can identify functionally relevant lncRNA interactors. This strategy enabled the identification of FUS as a novel interactor of *mOct4P4* that is required to mediate access for SUV39H1 to a critical *mOct4P4* lncRNA [32]. A critical step in RIP experiments is to validate the specificity of protein candidates toward pseudogene lncRNAs. Corresponding control experiments include RT-PCR on RIP RNA eluates using PCR primers specifically amplifying abundant but unrelated lncRNAs or the mature parental mRNA. The lack of binding of the protein of interest to the parental mRNA is a strong indicator for the acquisition of new biological function by the pseudogene lncRNA.

7.3 Identification of Functional Domains of Pseudogene lncRNA by Deletion Analysis

Although functionally relevant pseudogene lncRNAs can acquire highly relevant biological function, sequence differences with the parental mRNA may only be marginal. Consequently, classic deletion analysis represents a powerful approach to link biologically relevant features of pseudogene lncRNA to specific lncRNA regions [24]. A panel of expression constructs containing truncated pseudogene lncRNA versions tagged with the MS2 RNA stem loop motif can be introduced into experimental cell lines stably expressing the flag-tagged MS2 protein. Subsequently, the impact of deletions on previously established pseudogene lncRNA properties can be evaluated. Using this strategies, *mOct4P4* lncRNA sequence elements essential for nuclear localization, recruitment of the SUV39H1 HMTase, parental gene silencing and targeting of the pseudogene lncRNA to the parental *Oct4* promoter were identified [24]. Importantly, data from this type of experiments can lead to the construction of a “minimal sufficient” pseudogene lncRNA

version that hosts sequence information required to cover all features of pseudogene lncRNA function. In the case of *mOct4P4*, this strategy resulted in the concentration of lncRNA function to a 200 nt sequence required for SUV39H1 binding and parental gene promoter targeting [32].

8 Future Outlook

mOct4P4 lncRNA represents a novel example for sense pseudogene lncRNA mediated alterations of chromatin structure at the promoter of a parental gene. Importantly, conservation of all functionally relevant *mOct4P4* features by human *hOCT4P3* pseudogene lncRNA underlines the relevance of this type of epigenetic gene regulation [32]. This also suggests that pseudogene lncRNA dependent chromatin regulation of parental gene expression may represent a more widespread regulatory process. Detailed bioinformatics analysis focusing on the paired analysis of chromatin and gene expression patterns of parental genes and "offspring" pseudogenes may reveal additional "candidate couples" for further analysis. RNA structure analyses are of central importance in order to understand how even small sequence changes in pseudogene lncRNAs can translate into new and highly specific biological function. Finally, it will be essential to study pseudogene lncRNA action in the context of nuclear organization. The application of methods such as DNA-FISH on parental and pseudogene loci, chromosome conformation capture techniques, enChIP, and SPRITE are all expected to provide insights into possible inter-/intrachromosomal communication between pseudogenes and parental genes in vertebrate cells [42].

Acknowledgments

S.S. is supported by the Italian Association for Cancer Research (AIRC)—Investigator Grants 18381 and 23074 and by the European Union, European Regional Development Fund and Interreg V-A Italia-Austria 2014–2020 (project codes ITAT1096-P and ITAT1050). R.B. is supported by the Italian Association for Cancer Research (AIRC)—Investigator Grant 17756.

References

1. Pei B, Sisu C, Frankish A et al (2012) The GENCODE pseudogene resource. Genome Biol. https://doi.org/10.1186/gb-2012-13-9-r51
2. Frankish A, Diekhans M, Ferreira AM et al (2019) GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. https://doi.org/10.1093/nar/gky955
3. Frankish A, Harrow J (2014) Gencode pseudogenes. Methods Mol Biol. https://doi.org/10.1007/978-1-4939-0835-6_10

4. Cheetham SW, Faulkner GJ, Dinger ME (2020) Overcoming challenges and dogmas to understand the functions of pseudogenes. Nat Rev Genet
5. Chen X, Wan L, Wang W et al (2020) Re-recognition of pseudogenes: from molecular to clinical applications. Theranostics 10 (4):1479–1499
6. Ji Z, Song R, Regev A, Struhl K (2015) Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. elife. https://doi.org/10.7554/eLife.08890
7. Brosch M, Saunders GI, Frankish A et al (2011) Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. Genome Res. https://doi.org/10.1101/gr.114272.110
8. Poliseno L, Salmena L, Zhang J et al (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature. https://doi.org/10.1038/nature09144
9. Salmena L, Poliseno L, Tay Y et al (2011) A ceRNA hypothesis: the rosetta stone of a hidden RNA language? Cell 146:353–358
10. Chiefari E, Iiritano S, Paonessa F et al (2010) Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes. Nat Commun 1. https://doi.org/10.1038/ncomms1040
11. Han YJ, Ma SF, Yourek G et al (2011) A transcribed pseudogene of MYLK promotes cell proliferation. FASEB J. https://doi.org/10.1096/fj.10-177808
12. Tam OH, Aravin AA, Stein P et al (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. https://doi.org/10.1038/nature06904
13. Watanabe T, Totoki Y, Toyoda A et al (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. https://doi.org/10.1038/nature06908
14. Guo X, Lin M, Rockowitz S et al (2014) Characterization of human pseudogene-derived non-coding RNAs for functional potential. PLoS One. https://doi.org/10.1371/journal.pone.0093972
15. Wei CC, Nie FQ, Jiang LL et al (2017) The pseudogene DUXAP10 promotes an aggressive phenotype through binding with LSD1 and repressing LATS2 and RRAD in non small cell lung cancer. Oncotarget. https://doi.org/10.18632/oncotarget.14125
16. Sun M, qi NF, Zang C et al (2017) The pseudogene DUXAP8 promotes non-small-cell lung cancer cell proliferation and invasion by epigenetically silencing EGR1 and RHOB. Mol Ther. https://doi.org/10.1016/j.ymthe.2016.12.018
17. Schoeftner S, Scarola M, Comisso E et al (2013) An Oct4-pRb axis, controlled by miR-335, integrates stem cell self-renewal and cell cycle control. Stem Cells. https://doi.org/10.1002/stem.1315
18. Nichols J, Zevnik B, Anastassiadis K et al (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. https://doi.org/10.1016/S0092-8674(00)81769-9
19. Boyer LA, Tong IL, Cole MF et al (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. https://doi.org/10.1016/j.cell.2005.08.020
20. Pain D, Chirn GW, Strassel C, Kemp DM (2005) Multiple retropseudogenes from pluripotent cell-specific gene expression indicates a potential signature for novel gene identification. J Biol Chem. https://doi.org/10.1074/jbc.C400587200
21. Liedtke S, Enczmann J, Waclawczyk S et al (2007) Oct4 and its pseudogenes confuse stem cell research. Cell Stem Cell
22. Lin H, Shabbir A, Molnar M, Lee T (2007) Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene. Biochem Biophys Res Commun. https://doi.org/10.1016/j.bbrc.2007.01.106
23. Panagopoulos I, Möller E, Collin A, Mertens F (2008) The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. Oncol Rep. https://doi.org/10.3892/or_00000105
24. Scarola M, Comisso E, Pascolo R et al (2015) Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun 6. https://doi.org/10.1038/ncomms8631
25. Suo G, Han J, Wang X et al (2005) Oct4 pseudogenes are transcribed in cancers. Biochem Biophys Res Commun. https://doi.org/10.1016/j.bbrc.2005.09.157
26. Poursani EM, Soltani BM, Mowla SJ (2016) Differential expression of OCT4 pseudogenes in pluripotent and tumor cell lines. Cell J. https://doi.org/10.22074/cellj.2016.3984
27. Xu G, Yang L, Zhang W, Wei X (2015) All the tested human somatic cells express both Oct4A and its pseudogenes but express Oct4A at much lower levels compared with its pseudogenes and human embryonic stem cells. Stem

Cells Dev. https://doi.org/10.1089/scd.2014.0552
28. Bai M, Yuan M, Liao H et al (2015) OCT4 pseudogene 5 upregulates OCT4 expression to promote proliferation by competing with miR-145 in endometrial carcinoma. Oncol Rep. https://doi.org/10.3892/or.2015.3763
29. Wang L, Guo ZY, Zhang R et al (2013) Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. Carcinogenesis. https://doi.org/10.1093/carcin/bgt139
30. Hawkins PG, Morris KV (2010) Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. https://doi.org/10.4161/trns.1.3.13332
31. Lettnin AP, Wagner EF, Carrett-Dias M et al (2019) Silencing the OCT4-PG1 pseudogene reduces OCT-4 protein levels and changes characteristics of the multidrug resistance phenotype in chronic myeloid leukemia. Mol Biol Rep. https://doi.org/10.1007/s11033-019-04639-4
32. Scarola M et al COMMSBIO-19-1824-T. Commun Biol
33. Wang X, Arai S, Song X et al (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. https://doi.org/10.1038/nature06992
34. Cui W, Yoneda R, Ueda N, Kurokawa R (2018) Arginine methylation of translocated in liposarcoma (TLS) inhibits its binding to long noncoding RNA, abrogating TLS-mediated repression of CBP/p300 activity. J Biol Chem. https://doi.org/10.1074/jbc.RA117.000598
35. de Napoles M, Mermoud JE, Wakao R et al (2004) Polycomb group proteins ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev Cell. https://doi.org/10.1016/j.devcel.2004.10.005
36. Plath K, Fang J, Mlynarczyk-Evans SK et al (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science (80). https://doi.org/10.1126/science.1084274
37. Silva J, Mak W, Zvetkova I et al (2003) Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell. https://doi.org/10.1016/S1534-5807(03)00068-6
38. Lucchesi JC, Kuroda MI (2015) Dosage compensation in drosophila. Cold Spring Harb Perspect Biol. https://doi.org/10.1101/cshperspect.a019398
39. Kelley RL, Meller VH, Gordadze PR et al (1999) Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell. https://doi.org/10.1016/S0092-8674(00)81979-0
40. Meller VH, Wu KH, Roman G et al (1997) roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell. https://doi.org/10.1016/S0092-8674(00)81885-1
41. Lachner M, O'Carroll D, Rea S et al (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. https://doi.org/10.1038/35065132
42. Maass PG, Barutcu AR, Rinn JL (2019) Interchromosomal interactions: a genomic love story of kissing chromosomes. J Cell Biol
43. Engreitz JM, Ollikainen N, Guttman M (2016) Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat Rev Mol Cell Biol
44. Dekker J, Mirny L (2016) The 3D genome as moderator of chromosomal communication. Cell 164(6):1110–1121
45. Duncan IW (2002) Transvection effects in Drosophila. Annu Rev Genet. https://doi.org/10.1146/annurev.genet.36.060402.100441
46. McStay B (2016) Nucleolar organizer regions: genomic 'dark matter' requiring illumination. Genes Dev
47. Monahan K, Lomvardas S (2015) Monoallelic expression of olfactory receptors. Annu Rev Cell Dev Biol. https://doi.org/10.1146/annurev-cellbio-100814-125308
48. Bacher CP, Guggiari M, Brors B et al (2006) Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat Cell Biol. https://doi.org/10.1038/ncb1365
49. Hogan MS, Parfitt DE, Zepeda-Mendoza CJ et al (2015) Transient pairing of homologous Oct4 alleles accompanies the onset of embryonic stem cell differentiation. Cell Stem Cell 16:275–288. https://doi.org/10.1016/j.stem.2015.02.001
50. Bertrand E, Chartrand P, Schaefer M et al (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell. https://doi.org/10.1016/S1097-2765(00)80143-4
51. Kula A, Guerra J, Knezevich A et al (2011) Characterization of the HIV-1 RNA associated proteome identifies Matrin 3 as a nuclear cofactor of Rev function. Retrovirology. https://doi.org/10.1186/1742-4690-8-60