



Published in final edited form as:

Trends Cell Biol. 2014 October ; 24(10): 594–602. doi:10.1016/j.tcb.2014.06.003.

Physiological roles of long noncoding RNAs: Insights from knockout mice

Lingjie Li and Howard Y. Chang

Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305

Abstract

Long noncoding RNAs (lncRNAs) are a pervasive and recently recognized class of genes. lncRNAs have been proposed to modulate gene expression and nuclear architecture, but their physiological functions are still largely unclear. Several recent efforts to inactivate lncRNA genes in mouse models have shed light on their functions. Different genetic strategies yielded specific lessons about the roles of lncRNA transcription, the lncRNA transcript itself and underlying sequence elements. Current results indicate a function for lncRNAs in organ development, immunity, organismal viability and a connection to human diseases.

Keywords

lncRNA; genetic study; gene targeting; knockout; phenotype

lncRNA discovery and characteristics

The Central Dogma posits that protein is the major player for regulating cellular functions, while RNA serves as the messenger to transmit genetic information from DNA to its encoded protein. However, over the last decade, many versatile functions of RNA have been discovered, and regulatory RNAs are gaining widespread attention in biomedical research. In fact, the mammalian genome is pervasively transcribed (e.g. a total of 62% and 75% of the human genome is found to be covered by either processed or primary transcripts), whereas only 2% of the genome encodes for proteins, indicating the majority of genomic DNA are transcribed into noncoding RNAs (ncRNAs) [1]. Interestingly, it has been suggested that the expansion of a proportion of the noncoding genome, i.e. genomic regions which are transcribed into ncRNA, contributes to the complexity of the organism during the evolution of developmental processes [2]. Recently studies demonstrate that ncRNAs play important roles in regulating various biological processes, including development, cell cycle regulation, response to stress, and cancer metastasis (reviewed in [3, 4]).

© 2014 Elsevier Ltd. All rights reserved

howchang@stanford.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The term ncRNA includes a wide variety of RNA species, including, but not restricted to, “classic” ncRNAs that regulate mRNA splicing and translation (i.e. small nuclear RNAs, ribosomal RNAs, transfer RNAs); short 21-28nt regulatory RNAs involved in transcriptional and posttranscriptional gene silencing (i.e. microRNA, small interfering RNAs, Piwi-associated (pi)RNAs); as well as a large number of newly identified and less understood long noncoding RNAs (lncRNAs) [5-7]. In this review, we focus on lncRNAs and their *in vivo* functions as assessed by targeted genetic inactivation.

lncRNAs are currently defined as RNA transcripts longer than 200 nucleotides that apparently do not function by encoding proteins [5]. Based on the genome location and relationship to protein coding genes or small RNAs, lncRNAs can be classified into the following: long intergenic noncoding RNA, transcripts overlapping with annotated genes (in sense, antisense, or within intronic sequences), small RNA host or precursor transcripts, divergent transcripts, promoter-associated transcripts, enhancer RNA, and transcribed pseudogenes [5, 7]. Most lncRNAs are expressed at lower levels and are often localized to the nucleus, preferentially in the chromatin and nuclear RNA fractions, while others localize to the cytoplasm [1, 8]. Genome-wide analysis has revealed that tens of thousands of lncRNA are expressed. Expression of lncRNAs is more tissue- or cell type-specific than protein-coding genes, suggesting they have distinct functions in different biological processes [1, 8, 9]. The primary sequence of lncRNAs is less conserved relative to that of protein-coding genes. Nevertheless, there is still greater conservation in promoter and exon regions of lncRNAs than intron or untranscribed intergenic regions. Interestingly, lncRNA genomic position is conserved across organisms, suggesting syntenic lncRNAs can have conserved functions, as well as indicating that functional structures can be maintained under weaker selective constraints during evolution [1, 10, 11].

Although only a small number of lncRNAs have been well-characterized, it is evident that they participate in a variety of molecular and cellular processes, including: chromosome dosage-compensation, imprinting, epigenetic regulation, cell cycle regulation, intracellular trafficking, reprogramming and stem cell differentiation. Moreover, emerging studies have revealed that lncRNAs can regulate multiple developmental processes and human diseases [4, 12].

lncRNAs not only contain specific sequence information, but also possess structural plasticity. They can directly interact with both DNA and RNA through base pairing, and bind to protein partners through specific structural motifs. This unique feature confers on lncRNAs the ability to regulate gene expression and function via diverse mechanisms. Biochemical and molecular studies have suggested that lncRNAs can act as an important interface between chromatin modification complexes and the genome, by acting as guides, scaffolds, decoys (**Glossary**), and direct links between higher order chromosome structure and gene expression (reviewed in [5]). Several lncRNAs can directly interact with and recruit chromatin complexes that mediate repressive or activating chromatin modifications, thereby altering the chromatin state and gene expression. Additional lncRNAs appear to localize to nuclear territories and the cytoplasm for other chromatin-templated processes and post-transcriptional regulation (reviewed in [13]). Here we focus on emerging genetic

studies that shed light on the *in vivo* functions of lncRNAs (**Table 1**), and highlight opportunities and challenges in this field.

lncRNA Knockouts: Different lessons from different strategies

Many protein-coding genes have been discovered in classical genetic studies, where coding mutations, truncations, or deletions in their genomic loci correlate with developmental defects or disease. However, only a few lncRNAs have been discovered by classical genetics (exampled in **Box1**), despite the fact that a majority of the genome is transcribed into noncoding RNAs. Some possible explanations may include: (i) lncRNAs have lower requirements for sequence conservation to maintain function and are impervious to frame shifts or stop codons, possibly making them less sensitive to certain genetic perturbations such as small indels or base substitutions; (ii) Some lncRNAs are thought to coordinate gene regulatory networks and contribute to organism complexity. Subtle phenotypic changes caused by mutation in lncRNAs may not be easily detected by forward genetic screening; and (iii) Many lncRNA loci may have functions redundant with other pathways or lack function.

According to their genome location, lncRNAs can be annotated into different classes. For lncRNAs overlapping with other genes (e.g. protein coding genes, microRNAs, etc), it is still challenging to use a conventional targeting method to disrupt a lncRNA itself without affecting overlapping genes. Many of the reported cases are lncRNAs transcribed from an intergenic region. In addition, further characterization, including delineation of transcript isoforms, coding potential analysis, syntenic conservation, and expression specificity, will be essential before picking lncRNA candidates for targeting.

For protein-coding genes, several targeting strategies can be used for a loss of function study, such as in frame insertion of a stop codon, exon replacement or insertion leading to a frame shift, whole transcript deletion, and truncation or mutation of functional domains. However, several of these strategies are not feasible for lncRNA targeting since the sequence is not translated, and the functional RNA domains are less characterized. Therefore, the whole transcript must be prevented from being transcribed. Currently different strategies for lncRNA gene targeting have yielded different insights about the lncRNA locus (**Figure 1**).

Complete deletion of the entire lncRNA gene

The conventional strategy requires the deletion of the complete gene sequence. The entire sequence from transcription start site to stop site is removed by homologous recombination. The null mutant can be generated either constitutively (directly removing the whole sequence from germ line) or conditionally (targeted region is flanked by loxp sites and can be deleted via tissue specific expressed *Cre* recombinase). A related strategy is to replace the lncRNA exons with a reporter gene cDNA. This strategy can generate a lncRNA reporter allele (in the heterozygous or homozygous state) to reveal *in vivo* expression patterns and lncRNA-null mutants (in the homozygous state). The *LacZ* reporter gene with leading KOZAK sequence (signal of translation initiation) is introduced to replace the lncRNA whole sequence, while the endogenous lncRNA promoter is kept intact. This approach is

straightforward and leaves no residual transcript, but may risk disrupting the local genomic context and disturbing neighboring gene expression, because some *cis* regulatory DNA elements embedded in the lncRNA locus may also be removed [9, 14]. The resulting phenotype suggests a role(s) for the lncRNA transcript as well as its underlying sequence; additional studies such as molecular studies of the lncRNA are useful in interpreting the results.

lncRNA promoter deletion

A lncRNA promoter can be identified by mapping the 5' ends of lncRNAs by CAGE or 5' RACE and searching for promoter-associated histone H3 lysine 4 trimethylation marks (reviewed in [5]). Promoter deletion can be focal (a few hundred base pairs) and thus minimally perturb the genomic locus. Since promoter deletion prevents lncRNA transcription, the resulting phenotype may be due to the lack of lncRNA transcription and/or the lncRNA product itself. Additional studies are necessary to disentangle these two possibilities; for instance, rescuing the phenotype by ectopic expression of the lncRNA. One potential caveat of this strategy is that, like many mRNA genes, lncRNA genes may have alternative promoters and hence retain expression of one or more isoforms after loss of a single promoter. Additionally, many lncRNAs are transcribed in a divergent manner in close proximity to protein-coding mRNA genes and promoter deletion may affect both genes in the divergent pair and complicate interpretation.

Integration of a premature polyadenylation cassette

Another strategy is to integrate a strong transcriptional stop signal at the start of the transcript unit; consequently, resulting in early termination of the transcript from its promoter. A triple repeat of the SV40 polyadenylation (polyA) signal or the chicken β -globin polyA signal is usually inserted into the first exon of a targeted gene [15-17]. As RNA polymerase transcribes across the lncRNA locus, polyA sites will cause premature cleavage and polyadenylation of the lncRNA transcript, likely leading to its decay. Barlow and colleagues used this elegant approach to show that the act of antisense transcription, but not the lncRNA product, across another gene's promoter was the basis of a regulatory switch [18]. This approach can produce significant (but incomplete) lncRNA depletion and less long-range genomic disruption for a loss-of-function study. To be noted, the transcriptional start sites need to be carefully examined to ensure there are no alternative transcripts expressed downstream of the stop cassette. The polyA addition does not prevent the act of lncRNA transcription, although it is expected that RNA polymerase II will disengage a few kilobases downstream of transcript cleavage. For some very abundant lncRNAs, premature termination may still result in detectable residual RNA. In addition, the local chromatin changes caused by the insertion of transcriptional stop signals need to be evaluated as well.

Diverse lncRNA biology revealed by genetic inactivation

Growing evidence indicates that lncRNAs exert important functions through various mechanisms. However, most of the molecular functions are deduced via *in vitro* studies in cell culture, while the physiological significance of these lncRNAs remains elusive. Notably, many lncRNAs are transcribed in unique spatial and temporal patterns during

development, making *in vitro* modeling difficult. In addition, conventional RNA interference strategies may not be feasible to elucidate the functions of some lncRNAs, due to an insufficient knockdown of nuclear, highly abundant, or structured lncRNA transcripts and off-target effects. To overcome these challenges, genetic ablation of lncRNAs in animal models is required for *in vivo* functional analysis, and will bring more insights into the developmental role of specific lncRNAs. These insights include, but are not limited to, understanding their diverse regulatory roles in complex tissues (e.g. regulating lineage commitment, tissue homeostasis, and organ morphogenesis), their interlinked impact on multiple organs, and their association with disease progression and treatment.

To date, there have been only a few lncRNA knockout animals generated for genetic study, leaving many of the other lncRNA *in vivo* functions less understood. Recent *in vivo* insights are emerging for lncRNAs commonly associated with disease through the development of single knockout mouse models as well as the development of a recent large-scale phenotype analysis with multiple lncRNA knockout mouse models [19]. These studies are revealing important functions of lncRNAs in mammalian development and physiology, and are providing a framework and resource for large-scale genetic study of lncRNAs in the future. We discuss the findings revealed from these studies that suggest lncRNAs play important roles in organ development, organismal viability, immunity and their association with disease progression.

Organ development and patterning

Hotair controls skeletal patterning and gene repression

HOTAIR (HOX transcript antisense RNA) is a lncRNA transcribed from the HOXC locus, that can repress *HoxD* genes expression *in trans* to control cell positional identity [20]. HOTAIR interacts with PRC2 and LSD1 complexes and recruits them to specific target genes, and promotes H3K27 methylation and H3K4 demethylation for gene silencing [20-22]. HOTAIR overexpression is associated with the progression and metastasis of several types of human cancer [23-25]. However, the *in vivo* function of HOTAIR during development is less understood.

Similar to the human homolog, mouse *Hotair* is transcribed from the syntenic location in the *HoxC* locus and is expressed in the distal and posterior anatomic sites, including the posterior trunk and limb bud [20, 26, 27]. Recently, a targeted *Hotair* knockout animal was generated by homologous recombination [27]. *Hotair* deletion did not affect viability, but did lead to homeotic transformation of spinal vertebrae and malformation of metacarpal-carpal bones in the limb skeleton (**Figure 2**). Targeted disruption of *Hotair* led to derepression of hundreds of genes, including several *HoxD* and imprinted genes. *Hotair* knockout embryos showed anterior expansion of the *Hoxd10* and *Hoxd11* expression domain in the posterior trunk, and altered *Dlk1* and mesenchymal gene expression in the developing wrist, further supporting the identified phenotype. In contrast, *HoxC* gene expression was not significantly affected in *Hotair* mutant embryos. Further analysis confirmed the interaction of *Hotair* with PRC2 and *Lsd1* *in vivo*; and loss of *Hotair* led to a gain of H3K4me3 and a loss of H3K27me3 at target genes. The transformation of bony elements in one anatomic site to those characteristic of another suggests that *Hotair* is

important for positional identity through its function in chromatin modification and gene regulation.

Interestingly, deletion of the entire *HoxC* locus, including the *Hotair* gene, causes perinatal lethality [26, 28], but does not show the skeletal transformation observed in *Hotair* or individual *HoxC* gene deletion. Deletion of *HoxC* locus revealed little change in *HoxD* expression or chromatin state [26]. There are nine *HoxC* genes and several lncRNAs and microRNAs in the *HoxC* locus. The mild phenotype observed upon *HoxC* deletion may indicate a compensatory mechanism from cross-regulation in the HOX family, or other lncRNAs with functions antagonistic to *Hotair* [27]. The comparison between *HoxC* deletion and targeted *Hotair* knockout emphasize the importance for fine-scale analysis and multiple alleles to define lncRNA function *in vivo*.

Fendrr is a critical regulator of cardiopulmonary development

Fendrr (Foxf1 adjacent non-coding developmental regulatory RNA) is a ~2.4kb long noncoding RNA transcribed divergently from the transcription factor coding gene Foxf1. Recently, two studies using a mouse knockout model identified *Fendrr* as a potential regulator of the development of the heart, body wall, lung, and gastrointestinal tract [15, 19] (**Figure 2**).

In the first study, *Fendrr* was identified as a lncRNA specifically transcribed in the posterior mesoderm of mid-gestation mouse embryos [15]. *In situ* hybridization showed that *Fendrr* is expressed in the caudal end of the lateral plate mesoderm (LPM), which gives rise to the heart and structures of the ventral body wall. Homologous gene targeting generated a *Fendrr*-null mutant, in which the first exon was replaced with a strong transcriptional stop signal. The *Fendrr* knockout embryo is lethal with prominent omphalocele and heart defects, due to reduced body wall and cardiac hypoplasia. This embryonic phenotype can be rescued by re-expression with a BAC clone containing the *Fendrr* gene; confirming the phenotype is due to loss of *Fendrr* transcripts instead of compromised genetic integrity. Loss of *Fendrr* results in an increased expression of a subset of cardiac transcriptional regulators Gata6 and NKX2-5, accompanied by an increase in H3K4me3 levels at their promoter. *Fendrr* binds to both the PRC2 and TrxG/MLL complexes *in vivo*, suggesting it modulates the balance between repressive and activating marks on mesoderm-related genes. Specifically, *Fendrr* can bind to target genomic sites via formation of a dsDNA:RNA triplex structure and recruit chromatin modification complexes.

The second *Fendrr*^{-/-} mouse model was generated in a functional study of 18 lncRNA knockout strains, in which the lncRNA genes were replaced with a *LacZ* reporter cassette [19]. In contrast to the observation discussed above, RNA-seq and *LacZ* staining revealed a higher expression in lung, trachea and gastrointestinal tract. *Fendrr*^{-/-} mouse is perinatal lethal with multiple defects in the lung, heart and gastrointestinal tract. Detailed analysis revealed that the respiratory failure at birth was due to defects in lung maturation and vascularization. In addition, thinning of the mesenchymal layer of the mucosa, external smooth muscle layer in the esophagus and intraventricular septal heart defects were observed in *Fendrr*^{-/-} developing embryos. Furthermore, differential RNA-seq analysis in lung tissue showed that muscle differentiation and contraction gene sets were significantly

misregulated in *Fendrr*^{-/-} lungs, which is in agreement with the identified defects in lung vasculature. In addition, the expression of neighboring genes within the *Fendrr* locus was not perturbed in the mutant animals, suggesting that the *Fendrr* gene does not act as a local enhancer.

These two studies showed different expression patterns of *Fendrr* and developmental defects in knockout animals. One possible explanation for this discrepancy may be from the distinct targeting strategies for the *Fendrr* locus. Despite these differences, both studies confirm the lethal phenotype and suggest that most of the defects are due to differentiation of mesenchyme-derived tissues across several organ systems, notably in the pulmonary system. The recent discovery of a human genetic syndrome with pulmonary atresia with mutations in the *FENDRR* locus provides strong evidence for a conserved and medically important role for this lncRNA [29].

Organismal viability and growth control

Many lncRNAs are involved in epigenetic regulation of gene expression during developmental process, such as X chromosome inactivation, genomic imprinting, etc., which are essential for organism viability and growth control [30], and their genetic deletion results in developmental lethality or growth defects. For example, knockout of lncRNAs *Xist*, *Tsix* or *Jpx* causes embryonic lethality and aberrant X-linked gene expression [31-37]; while deletion of lncRNA *Airn* or *Kcnq1ot1* results in growth deficiency and loss of imprinting at specific loci [18, 38-42].

Recently, a larger investigation of multiple knockout mouse models revealed the important role of lncRNAs in regulating organism viability and many developmental processes. The authors screened the lncRNA database with stringent bioinformatic analysis (e.g. lack of protein-coding capacity, lack of overlap with protein-coding gene or other non-lncRNA annotations) and identified 18 lncRNA candidates for targeted deletion in mouse. They used VelociGene technology [43], a high-throughput approach for gene targeting and expression analysis, to generate knockout mouse strains, in which each lncRNA gene was replaced by a *LacZ* reporter cassette. These new strains have advantages for both expression analysis and loss-of-function study of specific lncRNAs.

Five of the 18 lncRNA mutants possessed discernable developmental defects or lethality, indicating a high rate of discovery of essential functions. Three mutant strains (*Fendrr*, *Peril* and *Mdgt*) exhibit peri- and postnatal lethality, two of which also exhibit incomplete penetrance and growth defects in the survivors. The other two mutant strains (linc-*Brn1b* and linc-*Pint*) have growth defects. Notably, *Linc-Brn1b*^{-/-} mice demonstrate defects in the cerebral cortex, especially in the development of upper layer II/III-IV neurons (**Figure 2**). Several other lncRNA candidates from the screen also showed brain specific expression, suggesting brain or behavioral defects in these mutants need to be further explored as well as the viability analysis.

Immunity

Natural variation in NeST controls T cell immunity and pathogen resistance

The importance of lncRNAs can also be revealed by forward genetics. The *in vivo* function of the lncRNA NeST (nettoie Salmonella pas Theiler's [cleanup Salmonella not Theiler's]) was discovered through positional cloning of a mouse pathogen susceptibility locus [44]. This locus on mouse chromosome 11 is in proximity to interferon gamma and other cytokine genes, and is important for pathogen resistance to Theiler's virus and Salmonella bacteria as a Mendelian trait. Intriguingly, the human orthologous sequence is a major risk locus for multiple autoimmune diseases and pathogen susceptibility, as defined by genome-wide association studies [45, 46]. The formal genetic proof came from transgenic expression of the NeST cDNA in T-cells, which rescued all of the phenotypes and conferred pathogen resistance to the deficient mouse strain. Furthermore, NeST was found to be an enhancer-like lncRNA important for the rapid and potent activation of the interferon gamma gene in CD8⁺ T-cells, a critical determinant of immune activity. NeST binds to the MLL-trithorax complex subunit WDR5, and the genetic variation in NeST status modulates the active chromatin state on the interferon gamma locus. An independent study of human NEST also revealed its role in the induction of cytokines by human T-cells [47]. These studies thus expand the roles of lncRNAs into immune function, pathogen resistance, and likely autoimmunity.

Connections to Human Disease

Xist: X chromosome inactivation and tumorigenesis

The lncRNA Xist (X-inactive specific transcript) is a well-studied key regulator of X chromosome inactivation (XCI) in mammals [30]. There are two X chromosomes (XX) in females as compared to one in males (XY). To ensure equal dosage of X-linked gene expression between two sexes, mammals inactivate one of the female X chromosomes. During this process, the lncRNA Xist is transcribed from the inactive chromosome (Xi), spreads across the entire chromosome, and leads to recruitment of H3K27me3 modification and X-linked gene silencing [30]. Targeted deletion of Xist in embryonic stem cells or mice causes loss of XCI and embryonic lethality in females [32, 33] (**Figure 2**). Once XCI is established, Xist deletion *in vitro* does not cause X reactivation in some somatic cells. These observations led to the conventional view that Xist is required for the initiation of XCI, but not for the maintenance of XCI. However, Xist is continually expressed in adult females.

Recently, an Xist conditional knockout model was generated, whereby Xist was conditionally deleted in hematopoietic stem cells at embryonic day 10.5, after the establishment of XCI [31]. Deletion of Xist in the blood compartment resulted in female-specific premature death and significant disease, including primary myelofibrosis, leukemia, histiocytic sarcoma, and vasculitis. These defects resemble mixed myeloproliferative neoplasm and myelodysplastic syndrome (MPN/MDS) (**Figure 2**). Bone marrow transplantation experiments indicated that hematopoietic cells, rather than stromal cells, were the origin of the disease. Meanwhile, Xist-deficient hematopoietic stem cells show defects in maturation and maintenance. Subsequent expression analysis showed that Xist-

deleted cells had increased expression of X-linked genes, suggesting that *Xist* is also required for maintenance of XCI *in vivo*. In addition, genome-wide expression changes were observed with enrichment of genes involved in the regulation of the cell cycle, hematopoiesis, and immunodeficiency, including the previously identified oncogenes and tumor suppressor genes in MDS and MPN. This finding establishes a link between X reactivation, autosomal expression changes and tumorigenesis. Furthermore, it demonstrates that *Xist* is not only required to maintain XCI, but also suppresses cancer progression *in vivo*. Aberrant *XIST* expression has been observed in many types of human cancer, including breast cancer, ovarian cancer, cervical cancer and lymphoma [48, 49]. The genetic study of *Xist* deficiency will help to elucidate the causality and mechanism of X reactivation-associated tumorigenesis.

Malat1 is dispensable for viability but required for cancer progression

Malat1 (also known as Neat2) is an abundant nuclear lncRNA that decorates a nuclear structure termed nuclear speckles. Previous studies have suggested roles for Malat1 in the regulation of cancer progression, alternative splicing, and cell cycle control (reviewed in [50]); hence much interest awaited the *Malat1* knockout. Three groups have established *Malat1* knockout animal models [51-53]. Surprisingly, no obvious developmental phenotype or morphological abnormalities were found in all studies. In addition, neither nuclear speckle structure and component distribution, nor pre-mRNA splicing were significantly affected in *Malat1* depleted cells. Nevertheless, detailed analysis revealed mild expression changes in the *Malat1* locus, including neighboring lncRNA *Neat1*. In one knock-out study, reduction of *Neat1* expression and paraspeckle formation were observed in particular tissues, including embryonic fibroblasts, intestine and colon [53]; while in another study, *Neat1* and other adjacent genes were modestly up-regulated in the brain cortex and liver upon *Malat1* deletion [52]; thus indicating Malat1 may cis-regulate gene transcription in a cell type-dependent manner. *Malat1* knockout mice did not display a strong phenotype under physiological conditions, which may be due to functionally redundant or compensatory mechanisms during development. Future studies may reveal a phenotype in organ homeostasis over time or under stress and pathological conditions, as was revealed for many microRNA knockouts [54, 55]. Notably, human MALAT1 expression is correlated with cancer progression, and MALAT1 knockout in cancer cells or MALAT1 inhibition by antisense oligonucleotides strongly inhibited cancer progression in xenograft models [56]. The tumor-specific requirement for Malat1 makes it a promising target for cancer therapy.

Challenges and unsolved problems

The lncRNA knockout models discussed above indicate that many lncRNAs serve essential roles for development and pathogenesis *in vivo*. However, there are thousands of lncRNAs, and most of them are expressed in a tissue-specific manner and in low abundance. In comparison to protein coding genes, lncRNA characterization and mechanism are less understood, bringing more challenges to lncRNA genetic studies. Below we discuss several issues that need to be considered for making lncRNA mutants, characterizing their phenotypes, and performing functional studies.

LncRNA knockout and local genomic integrity

Considering the complex interleaved genomic structures with mRNAs and regulatory elements, targeted disruption of lncRNA raises concern for off target effects on the integrity of the local genomic context. lncRNA may contain *cis* regulatory elements, or act as enhancer-like RNA regulating neighboring gene expression, and such effects can be disrupted upon lncRNA deletion. In addition, transcriptional interference of lncRNA on adjacent genes, either by altering transcription factor binding or local chromatin modification, has been identified (reviewed in [57]). Therefore, the secondary effect on local genomic integrity needs to be well considered before stating the cause of a phenotype is due to the lncRNA itself. Notably, a genetic rescue assay via re-expression of the lncRNA in knockout animals will help to discriminate such effects.

Functional redundancy and penetration

There are several lncRNA knockout animals showing subtle or no phenotype during development, while their *in vitro* study revealed an essential function in the cell. This discrepancy could be attributed to functional redundancy with other transcripts, or to a compensatory mechanism occurring during development in knockout animals. The unique challenge for the lncRNA field is that lncRNAs with redundant function may have limited sequence similarity (e.g. the roX RNAs in *Drosophila* [58, 59]); therefore, finding a functional homolog to generate double or triple knockouts is not simple. Alternatively, such lncRNAs may have little function in a well-controlled laboratory environment. It is believed that lncRNAs can coordinate or fine-tune multiple gene expression programs; therefore, subtle changes may need to be carefully inspected, or the phenotype may need to only be present under special conditions, e.g. in cancer progression or other pathological stress.

Functional conservation across species

Although lncRNAs have less sequence similarity across species, the genomic synteny and structural information may still be maintained, which suggest their functional conservation during evolution [11]. For example, comparison between zebrafish and mammalian lncRNAs revealed a short highly conserved region in the orthologs, and importantly, the loss-of-function phenotypes in zebrafish can be rescued by mammalian orthologs, indicating conserved functions can be retained during evolution [10]. Recent mouse knockouts reveal critical functions of lncRNAs in development and viability, and provide a platform for investigating their functional conservation in different species and for modelling human diseases. Alternatively, functional lncRNAs may arise quickly in evolution and contribute to organism complexity [60]. Species-specific roles for lncRNAs, for example in brain development, cognition and behavior, may require novel models to understand their functions *in vivo*.

Concluding remarks

In the past few years, many lncRNAs have been identified with important cellular functions and diverse regulatory mechanisms. However, the physiological roles of lncRNAs *in vivo* are just beginning to be explored. Recent studies using genetic ablation in animal models indicate that multiple lncRNAs function in development and disease pathogenesis.

Considering the diverse regulatory mechanisms of lncRNA, different strategies for targeting lncRNA loci can reveal different requirements of lncRNA transcription, lncRNA product, or underlying regulatory DNA elements in biological functions. Future studies should employ and contrast multiple alleles to distinguish between these roles, and develop targeting strategies with higher specificity and efficiency. The advent of chemically modified antisense oligonucleotides that can inhibit lncRNA function *in vivo* represents an exciting strategy to modulate lncRNAs for therapeutic benefit [61-63]. Finally, new single cell and epigenomic strategies that can be applied to primary cells from knockout animals should provide valuable insights into the molecular mechanisms of lncRNAs.

Box 1

lncRNAs discovered in classical genetics

H19: Over 20 years ago, one of the first lncRNAs, H19 was discovered. This gene was initially isolated from differential expression screening in several systems (e.g. α -foetoprotein response in liver; myogenic differentiation and embryonic stem cell differentiation). H19 is a 2.3kb, imprinted maternally expressed transcript, and abundantly expressed during embryonic development. H19 is involved in regulating gene expression in the imprinted gene network, and contributes to growth control in development. Mice lacking *H19* are viable, but display an overgrowth phenotype. In humans, mutations in *H19* locus are associated with Beckwith-Wiedemann Syndrome and Wilms tumorigenesis. Misexpression of H19 causes tumorigenesis. In addition, H19 also functions as the precursor of microRNA miR-675. The detailed working mechanism of lncRNA H19 remains to be elucidated (Reviewed in [65]).

Xist (X-inactive specific transcript): Xist was initially discovered in the early 1990s, when researchers screened for genes whose expression escaped X-inactivation. Human and mouse *Xist* was found mapped to the region of the X-inactivation center and was expressed exclusively from the inactive X-chromosome. *XIST/Xist* produces a 17-20kb long noncoding RNA and is essential for the initiation and spread of X chromosome inactivation (reviewed in [30, 66], more details in the main text).

roX(RNA on the X): The roX RNAs were initially discovered in a differential expression screen in the *Drosophila* brain from two sexes. The roX RNAs were expressed by the male X chromosome, and were required for sex dosage compensation in *Drosophila*. In contrast to mammals, *Drosophila* achieves dosage compensation by hyper-transcribing the single X chromosome in males rather than silencing one X chromosome in females. The roX RNAs are essential components of the dosage compensation complex (including MSL [Male-Specific Lethal] proteins 1, -2, -3, histone H4 lysine16 acetyltransferase MOF, and RNA/DNA helicase MLE). This ribonucleoprotein complex coats the X chromosome, where MOF acetylates histone H4K16, thereby increasing transcriptional output of X-chromosomal genes in males. There are two roX RNAs in *Drosophila*, and they differ in both size and sequence (roX1, 3.7kb; roX2, 0.6kb). Genetic analysis showed that the functions of the *roX* genes are

redundant but required for the association of the MSL proteins with the X chromosome (reviewed in[67]).

Acknowledgments

We thank members of the Chang lab and J. Rinn for discussions. We apologize to colleagues whose work could not be cited due to space constraints. Supported by National Institute of Health R01-CA118750 and R01-ES023168 (to H.Y.C.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Glossary

BAC	Bacterial artificial chromosome. It is a vector used to clone DNA fragments (100-300kb insert size) from another species and can be replicated in bacterial cells.
CAGE	Cap-analysis gene expression. It is a technique used to map and measure expression levels of transcription start sites by sequencing 5' ends of capped transcripts.
Decoy	The notion that lncRNAs can interact with DNA binding proteins to prevent their binding to DNA recognition elements.
Dosage compensation	Mechanisms involved in equalizing the expression of X-linked genes between two sexes. In mammals, dosage compensation is achieved by inactivation of one of the two X chromosomes in females, while in <i>Drosophila</i> it operates by hyper-activating the single male X chromosome.
Guide	Here it refers to lncRNAs that can recruit chromatin-associated proteins to specific genomic regions, either <i>in cis</i> (on neighboring genes) or <i>in trans</i> (on distantly located genes).
Homeotic transformation	The transformation of one body part into another arising from mutation or misexpression of genes during development.
Nuclear speckles	The nuclear domain found in the interchromatin region of mammalian cells. They are enriched in pre-mRNA splicing factors, and appear as 20-50 irregularly shaped structures.
Ortholog	Genes in different species that have evolved from a common ancestral gene by specification.
Paraspeckles	A relatively new class of subnuclear bodies. Usually, there are 10-20 paraspeckles presented in the interchromatin region, and often located adjacent to nuclear speckles. They are composed of ribonucleoproteins.
RACE	Rapid amplification of cDNA ends. It is a technique used to obtain the full-length cDNA when the sequence is only partially known. RACE utilizes RT-PCR to amplify both the 5' end and the 3' ends of

genes yielding the information valuable for mapping the transcriptional start sites, identification of splicing variants, etc.

Scaffold	Here it refers to lncRNAs that can bring together multiple proteins to form ribonucleoprotein complexes.
Synten	The presence of regions of chromosomes with preserved co-localized genes in different species.

References

1. Djebali S, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101–108. [PubMed: 22955620]
2. Mattick JS. RNA regulation: a new genetics? *Nat Rev Genet*. 2004; 5:316–323. [PubMed: 15131654]
3. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol*. 2012; 9:703–719. [PubMed: 22664915]
4. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014; 15:7–21. [PubMed: 24296535]
5. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem*. 2012; 81:145–166. [PubMed: 22663078]
6. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Molecular cell*. 2011; 43:904–914. [PubMed: 21925379]
7. Kung JT, et al. Long noncoding RNAs: past, present, and future. *Genetics*. 2013; 193:651–669. [PubMed: 23463798]
8. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012; 22:1775–1789. [PubMed: 22955988]
9. Cabili MN, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011; 25:1915–1927. [PubMed: 21890647]
10. Ulitsky I, et al. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*. 2011; 147:1537–1550. [PubMed: 22196729]
11. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013; 154:26–46. [PubMed: 23827673]
12. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol*. 2011; 21:354–361. [PubMed: 21550244]
13. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013; 152:1298–1307. [PubMed: 23498938]
14. Sigova AA, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A*. 2013; 110:2876–2881. [PubMed: 23382218]
15. Grote P, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell*. 2013; 24:206–214. [PubMed: 23369715]
16. Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol*. 2013; 10:1579–1585. [PubMed: 24036695]
17. Gutschner T, et al. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res*. 2011; 21:1944–1954. [PubMed: 21844124]
18. Latos PA, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*. 2012; 338:1469–1472. [PubMed: 23239737]
19. Sauvageau M, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*. 2013; 2:e01749. [PubMed: 24381249]

20. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007; 129:1311–1323. [PubMed: 17604720]
21. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010; 329:689–693. [PubMed: 20616235]
22. Chu C, et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell*. 2011; 44:667–678. [PubMed: 21963238]
23. Gupta RA, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:1071–1076. [PubMed: 20393566]
24. Kim K, et al. HOTAIR is a negative prognostic factor and exhibits pro-oncogenic activity in pancreatic cancer. *Oncogene*. 2013; 32:1616–1625. [PubMed: 22614017]
25. Kogo R, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res*. 2011; 71:6320–6326. [PubMed: 21862635]
26. Schorderet P, Duboule D. Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet*. 2011; 7:e1002071. [PubMed: 21637793]
27. Li L, et al. Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell reports*. 2013; 5:3–12. [PubMed: 24075995]
28. Suemori H, Noguchi S. Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev Biol*. 2000; 220:333–342. [PubMed: 10753520]
29. Szafranski P, et al. Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder. *Genome Res*. 2013; 23:23–33. [PubMed: 23034409]
30. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell*. 2013; 152:1308–1323. [PubMed: 23498939]
31. Yildirim E, et al. Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell*. 2013; 152:727–742. [PubMed: 23415223]
32. Marahrens Y, et al. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev*. 1997; 11:156–166. [PubMed: 9009199]
33. Penny GD, et al. Requirement for Xist in X chromosome inactivation. *Nature*. 1996; 379:131–137. [PubMed: 8538762]
34. Tian D, et al. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*. 2010; 143:390–403. [PubMed: 21029862]
35. Sun S, et al. Jpx RNA activates Xist by evicting CTCF. *Cell*. 2013; 153:1537–1551. [PubMed: 23791181]
36. Lee JT. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell*. 2000; 103:17–27. [PubMed: 11051544]
37. Sado T, et al. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development*. 2001; 128:1275–1286. [PubMed: 11262229]
38. Sleutels F, et al. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*. 2002; 415:810–813. [PubMed: 11845212]
39. Nagano T, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*. 2008; 322:1717–1720. [PubMed: 18988810]
40. Mancini-Dinardo D, et al. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev*. 2006; 20:1268–1282. [PubMed: 16702402]
41. Pandey RR, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell*. 2008; 32:232–246. [PubMed: 18951091]
42. Fitzpatrick GV, et al. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet*. 2002; 32:426–431. [PubMed: 12410230]
43. Valenzuela DM, et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol*. 2003; 21:652–659. [PubMed: 12730667]
44. Gomez JA, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell*. 2013; 152:743–754. [PubMed: 23415224]

45. Latiano A, et al. Investigation of multiple susceptibility loci for inflammatory bowel disease in an Italian cohort of patients. *PLoS one*. 2011; 6:e22688. [PubMed: 21818367]
46. Silverberg MS, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet*. 2009; 41:216–220. [PubMed: 19122664]
47. Collier SP, et al. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of Ifng by Th1 cells. *J Immunol*. 2012; 189:2084–2088. [PubMed: 22851706]
48. Kawakami T, et al. Characterization of loss-of-inactive X in Klinefelter syndrome and female-derived cancer cells. *Oncogene*. 2004; 23:6163–6169. [PubMed: 15195139]
49. McDonald HL, et al. Involvement of the X chromosome in non-Hodgkin lymphoma. *Genes, chromosomes & cancer*. 2000; 28:246–257. [PubMed: 10862030]
50. Gutschner T, et al. MALAT1 -- a paradigm for long noncoding RNA function in cancer. *J Mol Med (Berl)*. 2013; 91:791–801. [PubMed: 23529762]
51. Eissmann M, et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol*. 2012; 9:1076–1087. [PubMed: 22858678]
52. Zhang B, et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell reports*. 2012; 2:111–123. [PubMed: 22840402]
53. Nakagawa S, et al. Malat1 is not an essential component of nuclear speckles in mice. *Rna*. 2012; 18:1487–1499. [PubMed: 22718948]
54. Miska EA, et al. Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet*. 2007; 3:e215. [PubMed: 18085825]
55. Abbott AL. Uncovering new functions for microRNAs in *Caenorhabditis elegans*. *Curr Biol*. 2011; 21:R668–671. [PubMed: 21920301]
56. Gutschner T, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013; 73:1180–1189. [PubMed: 23243023]
57. Kornienko AE, et al. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol*. 2013; 11:59. [PubMed: 23721193]
58. Meller VH, Rattner BP. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *Embo J*. 2002; 21:1084–1091. [PubMed: 11867536]
59. Franke A, Baker BS. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol Cell*. 1999; 4:117–122. [PubMed: 10445033]
60. Liu G, et al. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle*. 2013; 12:2061–2072. [PubMed: 23759593]
61. Modarresi F, et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol*. 2012; 30:453–459. [PubMed: 22446693]
62. Gutschner T, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013; 73:1180–1189. [PubMed: 23243023]
63. Wheeler TM, et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature*. 2012; 488:111–115. [PubMed: 22859208]
64. Nakagawa S, et al. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *The Journal of cell biology*. 2011; 193:31–39. [PubMed: 21444682]
65. Gabory A, et al. The H19 locus: role of an imprinted non-coding RNA in growth and development. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2010; 32:473–480.
66. Lyon MF. Some milestones in the history of X-chromosome inactivation. *Annual review of genetics*. 1992; 26:16–28.
67. Ilik I, Akhtar A. roX RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol*. 2009; 6:113–121. [PubMed: 19229132]

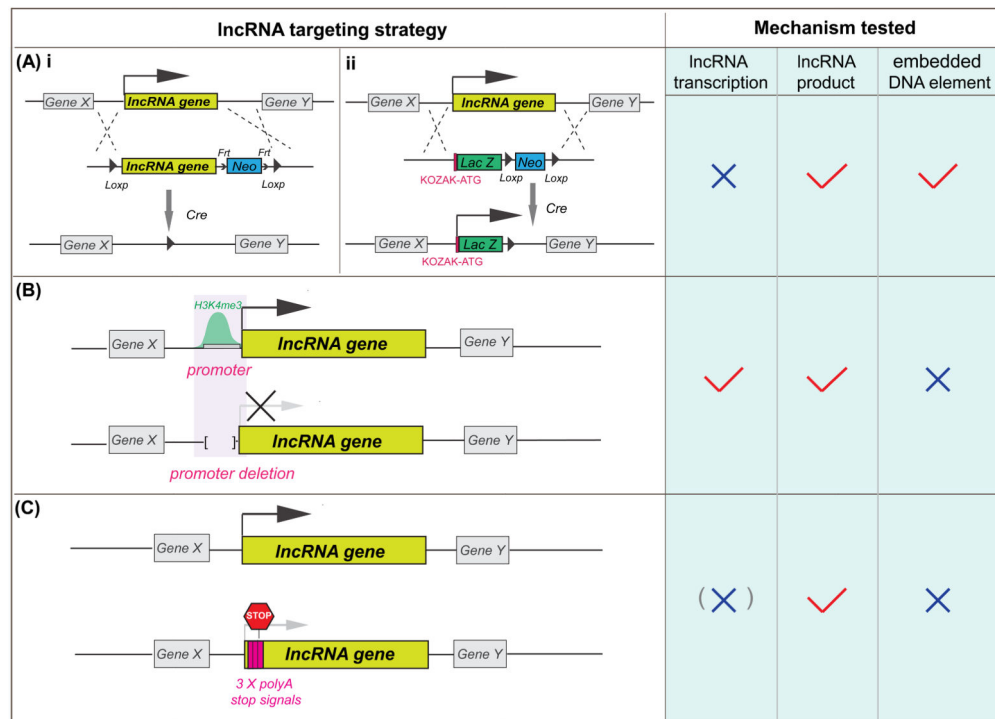


Figure 1. Genetic strategies to study lncRNA loci

(A) Complete deletion (i) or replacement (ii), (B) promoter deletion, or (C) premature termination can reveal different requirements of lncRNA transcription, RNA product, or underlying sequences in biological function. denotes the strategy tests the indicated mechanism; X denotes the strategy does not interrogate the indicated mechanism (“(X)”, under debate).

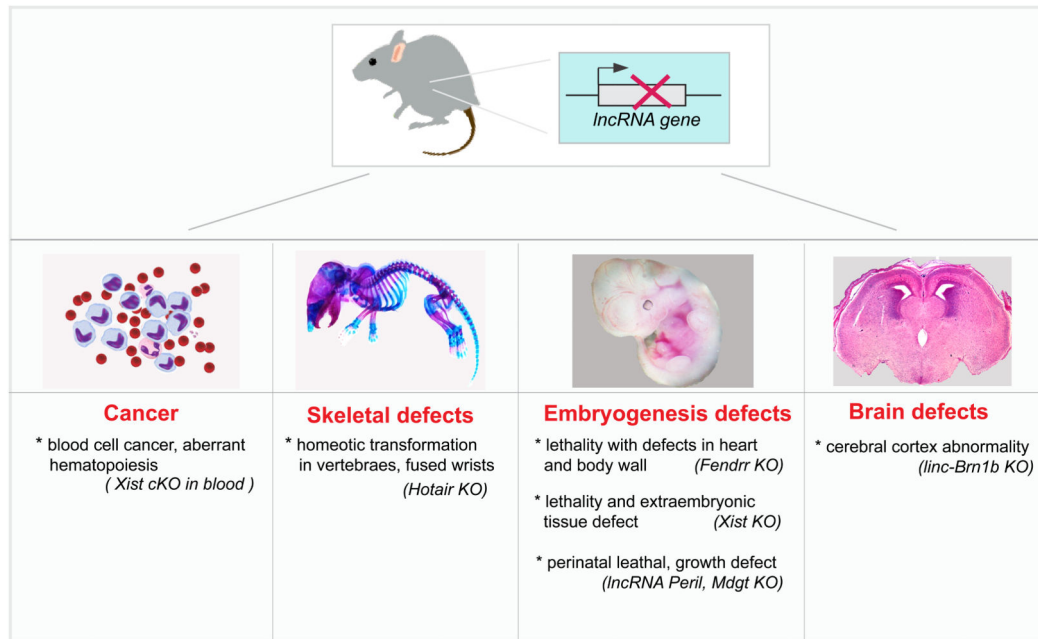


Figure 2. Highlights of lncRNA functions *in vivo*

LncRNA knockout (KO) results in multiple defects in mouse development and cancer progression. The diagram shows representative phenotypes and several knockout examples are listed below. cKO, conditional knockout.

Table 1

lncRNA genetic inactivation, resulting phenotype and related mechanism.

lncRNA	<i>In vivo</i> Function	Localization	Genetic inactivation methods	Phenotype	Mechanism	Refs
<i>Xist</i>	Dosage compensation hematopoiesis,	Nucleus	Entire gene deletion	Embryonic lethality Blood cell cancer, aberrant hematopoiesis	<i>Cis</i> acting, PRC2 related, X chromosome inactivation	[31-33]
<i>Tsix</i>	Dosage compensation	Nucleus	Promoter deletion, premature termination	Embryonic lethality	<i>Cis</i> acting, <i>Xist</i> repression	[36, 37]
<i>Jpx</i>	Dosage compensation	Nucleus	Gene deletion of 5' end region including promoter and exons	Embryonic lethality	<i>Trans</i> acting, <i>Xist</i> activation interacts with CTCF	[34, 35]
<i>Kcnq1ot1</i>	Genomic imprinting	Nucleus	Promoter deletion, premature termination by polyA insertion	Growth defect	Interacts with G9a and PRC2 complex, gene silencing in <i>cis</i>	[40-42]
<i>Airn</i>	Genomic imprinting	Nucleus	Premature termination by polyA insertion	Growth defect	Gene silencing in <i>cis</i> , interacts with G9a, transcriptional interference	[18, 38, 39]
<i>Hotair</i>	Skeletal morphogenesis	Nucleus	Entire gene deletion	Homeotic transformation in vertebrae, fusion wrists	<i>Trans</i> acting, PRC2 and LSD1 associated gene silencing (e.g. HoxD)	[27]
<i>Fendrr</i>	Organ development	Nucleus	Premature termination by polyA insertion Gene replacement by <i>LacZ</i> cassette	Embryonic lethality, heart and body wall defects Perinatal lethal, multi organ defects in lung, heart and gastrointestinal tract.	Bind to PRC2 and TrxG complex, <i>trans</i> acting	[15, 19]
<i>Malat1</i>	Cancer progression	Nucleus (Speckle)	<i>LacZ</i> -polyA insertion Entire gene deletion	Normal development	Nuclear speckle organization, associate with splicing factors	[51-53]
<i>Neat1</i>	Nuclear structure	Nucleus (Paraspeckle)	<i>LacZ</i> -polyA insertion	Normal development Loss of paraspeckles	Association with paraspeckle proteins	[64]
<i>Peril</i>	Organism viability	Nucleus	Gene replacement by <i>LacZ</i> cassette	Perinatal lethal	Not known	[19]
<i>Mdgt</i>	Organism viability and growth	unknown	Gene replacement by <i>LacZ</i> cassette	Perinatal lethal and growth defect	Not known	[19]
<i>Linc-Brn1b</i>	Brain development	Predominant in nucleus	Gene replacement by <i>LacZ</i> cassette	Cerebral cortex abnormality	Not known	[19]
<i>NeST</i>	Immune function, pathogen resistance	Nucleus	Natural polymorphisms confer distinct RNA expression and alleles	Resistance to Theiler's virus or Salmonella infection.	Bind to WDR5, <i>cis</i> activation (<i>lfn</i> g locus), enhancer like RNA	[44]